A Study on Stable Feature Representations for Artificial Olfactory System

by Taoping Liu

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Steven Su

University of Technology Sydney
Faculty of Engineering and Information Technology

March 2021

Certificate of Original Authorship

I, *Taoping Liu* declare that this thesis, is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *School of Biomedical Engineering*, *Faculty of Engineering and Information Technology* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

SIGNATURE: ______________________

[Taoping Liu]

DATE: 1st March, 2021

DEDICATION

To Mom and Dad

To My Wife, Chaoran

Acknowledgments

I would like to express my most sincere gratitude to my principal supervisor, A/Prof. Steven Su, for his continued support and wise guidance throughout my Ph.D. study. His knowledge, expertise, understanding, and insights greatly expanded my knowledge and skills in many areas. His patience and optimism helped me though the ups and downs of the Ph.D. life. Without his persistent help, the goal of this research would not have been achieved.

I would also like to show my great appreciation to my co-supervisor, Dr. Jun Li, for his solid support and skilful guidance. His expertise, intelligence, and encouragement enlarged my vision to a broader field and steered me to dive deeper into my research.

I am grateful to the UTS NOS.E Team members, Prof. Shari L. Forbes (University of Quebec in Trois-Rivières, Canada), A/Prof. Steven Su, Dr. Maiken Ueland, Dr. Wentian Zhang, and Amber Brown, for their selfless help. It is a great honour for me to be part of the team.

I am also grateful to the BeFAQT Team, especially to Prof. Ren Ping Liu, Mr. Peter Loneragan, A/Prof. Steven Su, A/Prof. Qiang Wu, A/Prof. Jian Zhang, Dr. Xu Wang, Mr. Erik Poole (Sydney Fish Market, Australia), and Dr. Zongjian Zhang for giving me the opportunity to further research in this amazing project.

I would like to thank all stuff in the School of Biomedical Engineering, especially to Prof. Joanne Tipper and Prof. Gyorgy Hutvagner, for all the suggestions, advice, and help they have given me. Special thanks to Prof. Gyorgy Hutvagner for his effort in letting us feel a sense of connectedness and belongingness, especially in this tough year.

I would like to acknowledge the assistance from my colleagues, Dr. Tao Zhang, Dr. Lin Ye, Dr. Wenhui Chen, Dr. Hairong Yu, Yao Huang, Hanjie Wu, Yujiao Wu, Palayil Baby Jephil, *etc*..

I would like to recognize the help that I received from my friends Dr. Miao Zhang, Li Wang, Chen Wang, Yanhao Zhang, Dr. Ye Shi, Dr. Huan Yu, Dr. Wenwei Mo, Kairui Guo, Lingmeng Li, and Mingjian Wang.

I must also thank Prof. David Huang (Western Sydney University, Australia) and

Mrs. Sarah Huang for their warm help.

My deepest gratitude goes to my parents for their immeasurable support and encouragement throughout my Ph.D. life.

I would like to sincerely thank the love of my wife, Chaoran, for her accompany and encouragement.

Finally, I would like to thank the sponsors for the scholarships, including China Scholarship Council (CSC), University of Technology Sydney (UTS), and Food Agility Cooperative Research Centre (Food Agility CRC).

Abstract

The electronic nose (e-nose) is an artificial olfactory system, consisting of a gas sensor array, a control system, and algorithms, designed to detect and identify the single or mixtures of odours. Owing to the boom of gas sensor technology and machine learning algorithms, e-noses have found wide applications in many different fields. However, the performance of e-noses in real applications has been challenged due to the 3D issue (i.e. the discreteness issue, the drift issue, and the disturbance issue). This thesis mainly focuses on the discreteness issue, especially the instability in feature representations caused by sensor noises.

A kernel regularization modelling-based method is proposed to provide stable representations for target odours. This method regards the e-nose as a whole system. The estimated parameters of the system are applied as features to represent the odour. The use of a smooth and stable kernel in the regularization term helps to overcome the ill-posed problem existing in deconvolution. The performance of the proposed method is verified by the experiment of classifying six target perfumes measured under a relatively stable environment.

However, the above-proposed method requires an accurate setting of the initial time (time of gas-on). In order to avoid a laborious searching of this time point, an improved method based on multiscale wavelet kernel regularization is proposed. The multiscale wavelet kernel inherits the advantage from wavelet function in approximating arbitrary signals and equips the method with the ability in resistance to random noise. The performance of the proposed method is verified by the experiment of classifying four target whiskies. The training and testing samples were obtained from two environments. Accordingly, a framework of "feature extraction – domain adaptation" is proposed in the experiment.

Aiming at the instability of traditional transient-state features extracted from the noise-contaminated signal, a novel kernel Tikhonov regularization-based numerical differentiation algorithm is proposed. The proposed method can improve such features by directly estimating accurate high-order derivatives from noisy signals. The performance

of the proposed method is verified by the experiment of identifying four target whiskies. The training and testing set were collected from two different environments. Moreover, samples of two new whiskies were added to the testing set as disturbances. Accordingly, a framework of "feature extraction – domain adaptation – one-class classification" is proposed in this experiment.

In addition, on the basis of stable feature representations, we also proposed two frameworks for food freshness monitoring and evaluation. Monitoring is realized by one-class classification. A single hidden Markov model (HMM) trained only by fresh samples is applied to track the change in freshness. For freshness evaluation, an HMM-based decoding algorithm is proposed to cluster the freshness level when the whole life-span data of meat stored in a specific storage condition is available. Then, HMM for each freshness level is trained and applied in parallel as freshness evaluation models to classify the tested meat samples.

List of Publications

Related to the Thesis :

[1] T. Liu, W. Zhang, L. Ye, M. Ueland, S. L. Forbes, and S. W. Su, *A novel multi-odour identification by electronic nose using non-parametric modelling-based feature extraction and time-series classification*, Sensors and Actuators B: Chemical, 298 (2019), pp. 126690.

[2] T. Liu, W. Zhang, M. Yuwono, M. Zhang, M. Ueland, S. L. Forbes, and S. W. Su, *A data-driven meat freshness monitoring and evaluation method using rapid centroid estimation and hidden Markov models*, Sensors and Actuators B: Chemical, 311 (2020), pp. 127868.

[3] T. Liu, W. Zhang, J. Li, M. Ueland, S. L. Forbes, W. X. Zheng, and S. W. Su, *A multiscale wavelet kernel regularization-based feature extraction method for electronic nose*, Submitted to IEEE Transactions on Systems, Man, and Cybernetics: Systems. (*Under Review*)

[4] T. Liu, W. Zhang, L. Wang, M. Ueland, S. L. Forbes, W. X. Zheng, and S. W. Su, *Numerical differentiation from noisy signals: a kernel regularization method to improve transient-state features for electronic nose*, Submitted to IEEE Transactions on Industrial Electronics. (*Under Review*)

Others :

[5] T. Liu, W. Zhang, P. McLean, M. Ueland, S. L. Forbes, and S. W. Su, *Electronic nose-based odor classification using genetic algorithms and fuzzy support vector machines*, International Journal of Fuzzy Systems, 20 (2018), pp. 1309-1320.

[6] W. Zhang, T. Liu, M. Zhang, Y. Zhang, H. Li, M. Ueland, S. L. Forbes, R. X. Wang, and S. W. Su, *NOS.E: a new fast response electronic nose health*

monitoring system, in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 4977-4980.

[7] W. ZHANG, T. LIU, L. YE, M. UELAND, S. L. FORBES, AND S. W. SU, *A novel data pre-processing method for odour detection and identification system*, Sensors and Actuators A: Physical, 287 (2019), pp. 113-120.

[8] W. ZHANG, T. LIU, M. UELAND, S. L. FORBES, R. X. WANG, AND S. W. SU, *Design of an efficient electronic nose system for odour analysis and assessment*, Measurement, 165 (2020), pp. 108089.

[9] M. ZHANG, H. LI, S. PAN, T. LIU, AND S. SU, *One-shot neural architecture search via novelty driven sampling*, in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence Yokohama, Japan: International Joint Conference on Artificial Intelligence Organization, 2020, pp. 3188-3194.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Motivation and Scope

Olfaction, the sense of smell through which odours (chemical substances) in the vicinity are perceived, plays a vital role for animals in foraging, breeding, and avoiding dangerous situations [1]. Insects release pheromones and detect olfactory cues to indicate food sources (e.g. ants' trail pheromones), seek mates (e.g. moths' sex pheromones), and call for a colony defense (e.g. bees' alarm pheromones); many mammals leave odours of urine or other secretions to mark their territory (e.g. bears' scent marking) or ward off predators (e.g. skunks' spray) [16, 25, 49, 59, 71, 93].

Humans, however, do not use the sense of smell like other mammals but rely heavily on senses like vision to acquire information and guide behaviors [83]. Most people only emphasize their sense of smell in limited scenarios such as searching for a fragrant scent when picking mangoes at the grocery store or making a decision over whether to throw away the leftover in the refrigerator. In most cases, information deriving from olfaction directly links to brain regions processing emotions and memories, thus frequently influencing people unconsciously [83, 161]. As a result, the power of smell and olfaction have been underestimated by people for a long time [151, 158].

In fact, smells can offer people information far more than traits for scent preference. Differences of fragrance can be decisive factors for the label of classification of wines or perfume; scent from food and food products can be a vital indicator for their freshness, ripeness, and quality; some odours emitted from human bodies can be olfactory marks of suffering from certain diseases; foul odours coming from gas stoves can be the alarm for hazardous gas leak.

Thus, biological olfaction has been increasingly investigated, appreciated and employed in different areas. Studies suggest that humans, with about 400 types of olfactory receptors, are able to distinguish more than one trillion different odours [18, 122, 167]. People receiving professional training for and/or having talent for distinguishing scents can make a career being perfumers, sommeliers, brewers, winemakers, and distillers. In some special fields, people also seek help from animals with more powerful olfactory system. Dogs, for example, are not only valued in the criminal investigation, but also trained for diagnosing human diseases such as cancer, malaria, and diabetes [57, 61, 184]. Recently, researchers even attempt to train dogs to detect the possible infection of coronavirus disease 2019 (COVID-19) [60, 75].

Powerful as they are, the biological olfactory systems of either humans or other animals have their natural deficiencies: i) The perception of smells is a subjective sensation [27], which varies individually; (ii) It is time and money consuming to train novice employees or other animals to become competent for certain tasks [63]; iii) The biological olfactory system is prone to fatigue and would be affected by some diseases [64, 138]; iv) Some target smells (odours) may contain toxic substances, infectious bacteria, or viruses.

Aware of the power of smells and olfaction, researchers started to devote themselves in developing artificial olfactory system. The artificial olfactory system, also known as electronic nose or e-nose for short, is a device designed to imitate the function of the human olfactory system to recognize single or mixed odours [52, 63, 70, 130, 139, 143]. An e-nose is composed of two major parts, i.e. hardware and algorithms. Hardware can be further classified as a gas sensor array and a control system, which includes a sensor chamber, an air intake and exhaust system, a power module, a sensor heating module, and a communication module. The air intake and exhaust systems is designed to mimic the breathing function (i.e. sniffing odours by inhalation and cleaning odour residuals by exhalation). Gas sensors perform as olfactory sensory neurons. Other components perform as supporting cells to maintain the olfactory function. Algorithms can be further classified as signal processing algorithms, feature extraction algorithms, and pattern recognition algorithms. The function of algorithms performs as olfactory bulbs and the brain. An intuitive comparison between human olfactory system and artificial olfactory system is illustrated in Fig. 1.1.

Existing studies have made great contributions to the technology of artificial olfactory

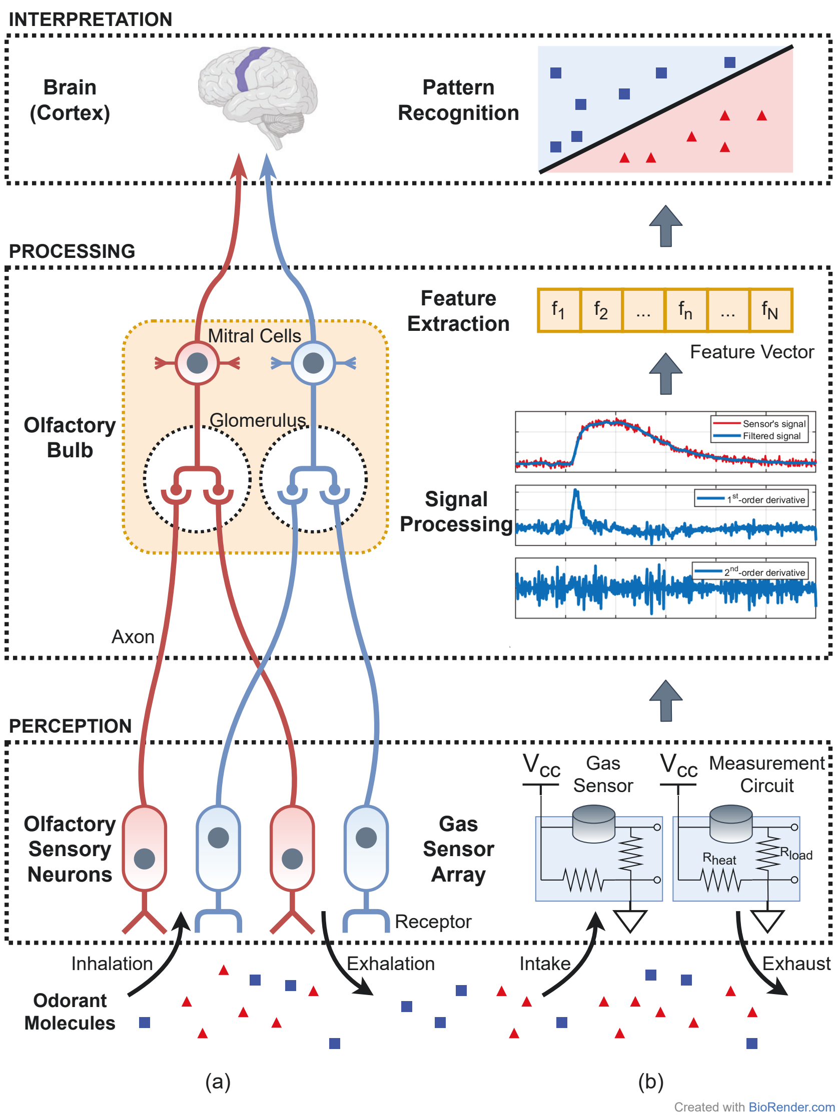


Figure 1.1: An intuitive comparison between two olfactory systems (a) Human olfactory system. Odorant molecules entered the nasal cavity by inhalation to touch olfactory sensory neurons. Once the molecules are caught, they will interact with specific reactors on the surface of olfactory sensory neurons. Then, action potentials are activated and transmitted via the neuron axons to the olfactory bulb. In the olfactory bulb, the potentials are further processed by glomerulus and mitral cells, and finally interpreted by the brain [70, 188]. (b) Artificial olfactory system. Odorant molecules are sniffed in to the sensor chamber by air intake/exhaust system. Once the odorant molecules are perceived by the sensor, the sensor's resistance will change. Then, this change is reflected by the voltage over the load resistor. After analog-to-digital (A/D) conversion, the sensor's reading is further processed by signal processing and feature extraction algorithms, and finally interpreted by pattern recognition algorithms.

system in both hardware and algorithms. Hardware design focuses on two aspects: i) Research on gas sensors focuses on the updating of sensing materials and/or designing of transduction mechanisms. It contributes to different types of gas sensors, such as metal oxide semiconductors (MOS/MOX) [76], metal oxide semiconductor field-effect transistors (MOSFET) [45], photoionization detectors (PID) [142], surface acoustic wave (SAW) sensors [5], conducting polymer sensors [117], and quartz crystal microbalance (QCM) sensors [32]. ii) Research on control system mainly includes designing heating module (constant heating temperature or periodically changing temperature) [44, 47, 54, 201, 206], air intake module (constant flow rate or periodically switching flow rate) [152, 181], sensor chamber structures (single chamber or multiple chambers) [135, 173, 174], control module (automatic control or manual control) [214, 216], and communication module (wired or wireless communication) [30]. The algorithm design includes sensor-decided signal processing and feature extraction algorithm design and task-decided pattern recognition algorithm design. Algorithms are the key object of this research. The major challenges regarding the algorithm will be further introduced in Section 1.3.

This thesis mainly focuses on the improvement in algorithms for MOS-type gas sensor array-based e-noses. The instruments adopted in experiments are our self-designed portable e-noses, i.e. UTS NOS.E II and III. UTS NOS.E II and III belong to wireless automatic e-nose with single sensor chamber, constant sensor heating temperature, and constant flow rate. The MOS-type sensors installed are commercially available, which is manufactured by Figaro Engineering Inc. (Japan).

1.2 Applications of Electronic Nose

The e-nose technology has been found wide applications in many different fields. Representative application fields are introduced as follows.

1.2.1 Auxiliary Disease Diagnosis

Human body emits hundreds to thousands of odorous or odorless volatile organic compounds (VOCs) produced by daily metabolism [155, 171]. Once pathological processes (e.g. metabolic disorders and infections) occur, a person's VOCs will change. These changes include the variation in the proportion of compounds under normal conditions or the synthesis of new substances [155]. VOCs can be used as markers to indicate individual's physical condition. For example, if a person suffers from diabetes, his body either cannot produce enough insulin or cannot use insulin effectively. The glucose in the blood cannot be broken down normally to provide energy. Then, the body turns to break down fatty acids for energy instead [67]. A high level of acetone, one kind of VOC that causes a fruity smell in breath or urine, is generated during this process [98]. The fruity smell is hence regarded as a sign for patients with diabetes.

The potential of e-nose technology has been investigated for several diseases over the past decades. Breath and urine are two main sources of VOCs for disease analysis. For breath analysis, e-noses have been applied to the diagnosis of respiratory diseases, such as distinguishing lung cancer from other lung diseases [35, 88], predicting treatment effect on lung cancer patients [28], evaluating the development of chronic obstructive pulmonary disease (COPD) [46], and detecting early lung cancer from high-risk smokers [116]. With the development of the COVID-19 pandemic, researchers also attempted to give an e-nose solution to detect this infectious disease [185]. For urine odour analysis, e-noses appear in the diagnosis such as prostate cancer [6], urinary tract infections [126], and diabetes [157]. For other sources of VOCs, the e-nose has been found in applications, such as detecting tuberculosis (TB) from sputum samples [43, 163], identifying wound infections from swab cultures [162], and checking possible use of illegal drugs from skin odours [176].

As most researchers describe, the major advantage of e-noses is the non-invasive way for disease diagnosis. But, it should be noted that e-noses only regard VOCs as a mixture without specifying what exact compounds contained [52, 63, 130, 143]. This is caused by the characteristic of gas sensors. Accordingly, it is better to regard the e-nose as an auxiliary disease diagnostic tool.

1.2.2 Food Safety Control

Food and food products usually generate unpleasant smells in the process of spoilage. These smells are actually VOCs caused by bacterial, molds, yeasts, or fungi during their metabolism [73, 180]. For example, the concentration of VOCs, such as dimethyl sulfide, trimethylamine, acetic acid, and methane-thiol, gradually increase as milk spoilage progresses [108]. The commonly used evaluation methods in the food industry include sensory analysis (e.g. human sensory on aroma, flavor, appearance, texture, or sound [39]), chemical analysis (e.g. total volatile basic nitrogen [80]), physical analysis (e.g. Warner-Bratzler shear force [31]), and micorbiological analysis (e.g. microbial total viable count [48]).

Compared with the above methods, the e-nose can provide a time-saving and economical non-destructive evaluation for food and food products. For this purpose, e-noses have been found in evaluating the quality, freshness, or shelf life of almost all kinds of foods and food products, including fruits and vegetables [55, 56], fish and sea foods [100, 123], poultry [186], meats and sausages [140, 146, 182, 183], eggs [200], dairy [89], grains [124], cooked foods [156], and alcohol and beverages [50, 110].

Moreover, e-noses are also applied in food fraud detection, such as identifying real and fake liquors [134], detecting adulteration in meat mince [165], and finding the use of preservatives [205].

1.2.3 Hazardous Gas Detection and Localization

Hazardous gases include toxic, corrosive, explosive, or oxidizing products or wastes generated from industrial facilities, and flammable gases used in households. A long term exposure to hazardous gases is a great threat to both health and life. For example, inhalation of low concentrated carbon monoxide (the product of incomplete combustion of hydrocarbons) causes symptoms like dizziness, headache, or nausea, while prolonged exposure can lead to irreversible injury or even death [133]. Therefore, hazardous gas leaks should be detected and eliminated as early as possible.

E-noses are more capable to work in such dangerous situations. The application of e-noses can be classified into two categories, i.e. passive sensing and active sensing. Passive sensing refers to collecting the on-site air in a sampling bag (e.g. Tedlar bag) for off-line gas analysis [20, 149] or placing the e-nose in a fixed position for on-line gas monitoring [24, 95]. For active sensing, gas sensors are placed on moving platforms, such as robots, vehicles, and flying drones [2, 3, 41, 118, 190]. The moving platform is able to locate the source of leakage by tracking along the odour concentration gradient. Based on the gradient, several searching methods have been proposed, such as chemotaxis (i.e. comparing the concentration difference between two gas sensors and moving towards the one with larger reading [145]), bio-inspired methods (i.e. imitating moving strategies of animals in tracking odour plumes, such as silkworm moths [19] and dung beetles [144]), swarm robotic methods (i.e. relying on the collaboration among a group of odour tracking robots [114, 160]), probability maps (i.e. modelling the odour distribution as a probability distribution and locating the source by inferences [14, 42]), and reinforcement learning (i.e. learning to find the source of leakage from environment information [178]).

1.3 Algorithmic Challenges and Solving Strategy

Similar to the receptors of the human olfactory system, gas sensors are also designed to be non-odour-specific [130, 210]. That is, the gas sensor can respond to more than one chemical stimulus. The purpose of algorithms is to find the differences as patterns among the response curves to represent different odours. However, in real applications, finding such differences has been challenged due to uncertainty in odour measurement processes. This uncertainty is usually attributed to three causes, which can be summarized as the 3D issue [210, 213], i.e. discreteness, drift, and disturbance.

1.3.1 Discreteness Issue

Discreteness issue refers to the difference when e-noses with the same gas sensor array measure the same odour under the same environmental conditions. Concentration variation of the target odour is a main cause of the difference since gas sensors produce different response curves under different concentration levels. Accordingly, the concentration of the target odour is usually controlled in one experiment. But, it should be noted that even if the odour concentration is the same, two same gas sensors' readings can still be different. The inevitable variability in the manufacturing process causes the discrepancy between the parameters of two identical sensors [187, 211]. For MOS-type sensors, the discrepancy can be described as two aspects [211, 213]: i) Baseline discrepancy caused by the fluctuation of a sensor's resistance value R_0 to fresh air; and ii) Sensitivity discrepancy caused by the fluctuation of a sensor's rate constant K_{Gas} to target gases. Another cause of the discreteness issue is the sensor noise. MOS-type sensors should work at about 400°C [121, 150, 164]. The thermal agitation of electrons inside the sensor causes Johnson-Nyquist noise (thermal noise), which is increased under this high working temperature and then reflected by the sensor's reading [10, 164]. Under the influence of noises, the same feature extraction algorithm will obtain completely different feature representations from similar response curves. A representative example of the discreteness issue is illustrated in Fig. 1.2.

1.3.2 Drift Issue

Drift issue refers to the difference caused by the change in the state of gas sensors. We classify the drift issue into two categories: i) Irreversible drift (or long-term drift) refers to a permanent change in the state of sensors, such as aging (e.g. the diffusion

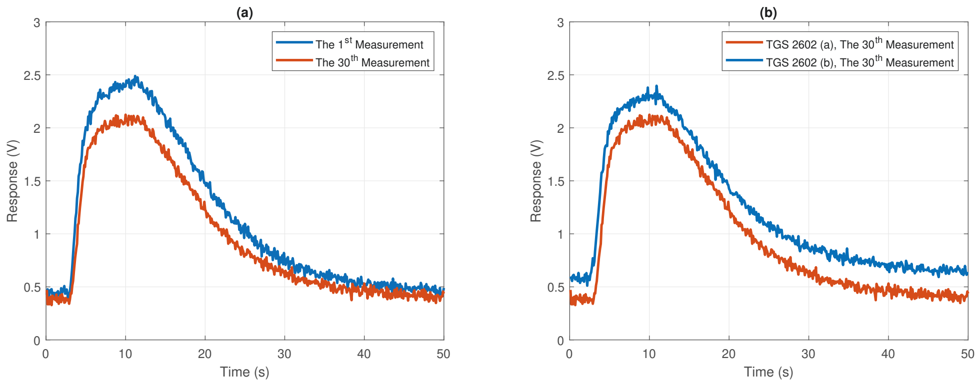


Figure 1.2: Representative example of the discreteness issue. The example displayed is responses of TGS 2602 sensors to the Chanel Chance perfume sample. (a) Odour concentration variation. The concentration of the target odour decreases as the number of measurements increases. Concentration variation causes the difference in response curves. (b) Sensor manufacturing discrepancy. Two identical sensors produce different response curves under the same environmental conditions.

or the rearrangement of metal atoms onto sensing film surface over long period of time [121]) and poisoning (e.g. the formation of strong and irreversible bonds between sensing materials and chemical pollutants [172]). ii) Reversible drift (or short-term drift) refers to a temporary change in the state of sensors, such as the variation in environmental conditions (e.g. humidity and temperature) and memory effect (e.g. the residue of previous gases). A representative example of the drift issue is illustrated in Fig. 1.3.

1.3.3 Disturbance Issue

The disturbance issue refers to the influence of non-target odours on the target odour recognition. Under the disturbance condition, the response of the sensor is actually a comprehensive result of both the target odour and non-target odours. The response curve is hence completely different from the one that only responds to the target odour itself. The cause of the disturbance issue can be classified into three categories: i) Continuous disturbance refers to a relatively constant influence of non-target odours. In this case, non-target odours can be considered as "background smells". For example, the ambient air at the Sydney Fish Market (SFM) is usually full of a "fishy smell". All target odour samples collected at the SFM contained disturbance information of such smell. ii) Contingent disturbance refers to a possible influence of non-target odours. In the process of sample collection, the target odour might be polluted contingently by

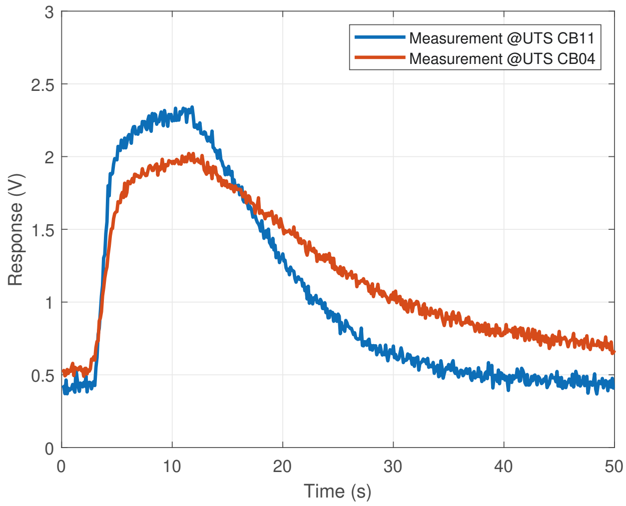


Figure 1.3: Representative example of the drift issue. The example displayed is responses of TGS 2602 gas sensor to the Chanel Chance perfume sample. The blue and red lines represent the samples measured at UTS CB11 and UTS CB04, respectively. The change in temperature and humidity causes the difference in response curves.

non-target odours. For example, due to the influence of intermittent wind, some target odour samples collected during the Australian bushfires in 2019 could be contaminated by smoke. iii) The target odour itself can change to non-target odour. For example, in food freshness monitoring, the smell released by the fresh food is regarded as the target odour. However, the target odour gradually becomes a non-target odour as the progress of spoilage process. An illustrative example of the disturbance issue is shown in Fig. 1.4.

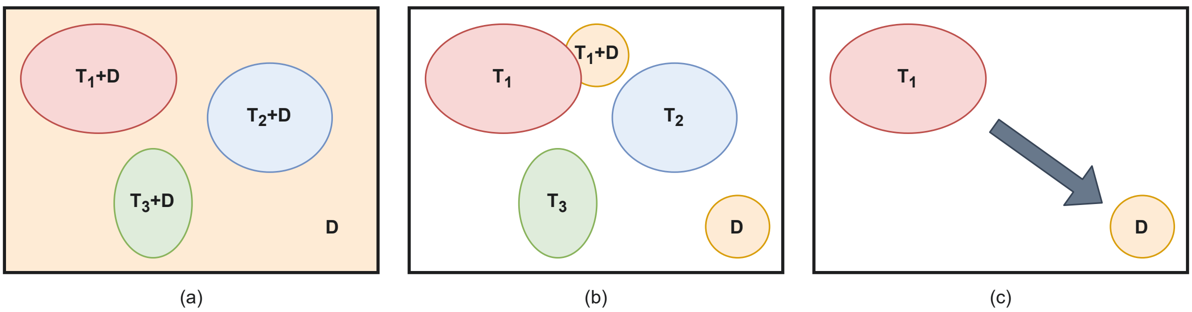


Figure 1.4: Illustrative example of the disturbance issue. (a) Continuous disturbance. (b) Contingent disturbance. (c) Changing from target odour to non-target odour. T_i denotes the i^{th} target odour. D denotes disturbance. $T_i + D$ represents the mixture of target and disturbance odour.

1.3.4 Solving Strategy

The solving strategy is determined by the cause of each problem. Accordingly, the 3D issue can be solved as follows:

i) The discreteness issue caused by:

- **Odour concentration variation**: For MOS-type sensors, the sensor's reading has a non-linear relationship with the odour concentration [17, 189]. If the concentration fluctuates within a small range, the relationship can be approximate as linear [17]. Accordingly, the concentration of the target odour should be controlled in the experiment. Then, the difference caused by the concentration variation can be eliminated or alleviated by linear transformations, such as normalization, standardization, fractional conductance, *etc*. [127].

- **Sensor manufacturing discrepancy**: The parameter difference caused by manufacturing discrepancy fluctuates within a small range. Accordingly, this influence can also be considered as linear, and then be alleviated by linear transformations [211, 213].

- **Sensor noise**: Sensor noise affects the stability of the feature representation. There are two ways to improve feature extraction algorithms: i) Applying denoising algorithms, such as filtering and regression, to improve the quality of the sensor's reading [105, 214, 215]; ii) Designing feature extraction algorithms with the ability in resistance to random noise [33, 106, 193].

ii) The drift issue caused by:

- **Irreversible drift**: From the perspective of feature space, drift is actually a problem of inconsistency between the feature distributions of samples obtained under different conditions. Based on the accurate and stable feature representations, ensemble learning can be applied. Base models are established for each feature distribution. Then, a possible model for the future feature distribution can be predicted from all previous models [104, 170, 172].

- **Reversible drift**: Transfer learning is a class of algorithms proposed to solve the problem of inconsistent feature distribution. Hence, it can be applied to solve the drift issue [102, 198, 209, 212, 213].

iii) The disturbance issue caused by:

- **Continuous disturbance**: If the influence of "background smell" is constant, the disturbance issue can be categorized into the drift issue. Hence, it can be solved by transfer learning.
- **Contingent disturbance**: From the perspective of algorithms, the disturbance issue can be considered as the influence of new classes of odour samples on the pre-trained model. If the new odour is not the pattern we are interested in, it can be removed as an outlier via novelty detection [107, 208]. Conversely, if the new odour is a pattern that we might be interested in, it can be added to the existing model as a new class via incremental learning [53].
- **The changing from target odour to non-target odour**: This problem is common in monitoring tasks. Monitoring aims to track the change of features over time. For most monitoring tasks, samples from normal conditions (target odours) are much easier to obtain than ones from abnormal conditions (non-target odours). Accordingly, monitoring can be regarded as a novelty detection task.

Solving the discreteness issue is the first step for the others since an accurate and stable feature representation will facilitate the solution to the drift and the disturbance issues. Accordingly, our strategy for solving the 3D issue is illustrated in Fig. 1.5.

1.4 Thesis Contribution

As mentioned previously, a good feature representation is a prerequisite for solving the drift and disturbance issue faced in real applications of e-noses. Accordingly, this thesis conducts research on algorithms that can provide accurate and stable feature representations for MOS-type gas sensor array-based e-nose. Based on such features, aiming at the 3D issue, this thesis also presents problem-solving frameworks in specific applications. The major contributions of this thesis are summarized as follows:

i) Proposition of feature extraction algorithms:

- **Non-parametric modelling-based feature extraction algorithm**

  This method regards the e-nose as a dynamic system. A unit step signal is assumed to be the ideal system input, which denotes the receiving of the target odour. The gas sensor's reading is considered as the system output. Parameters of this dynamic

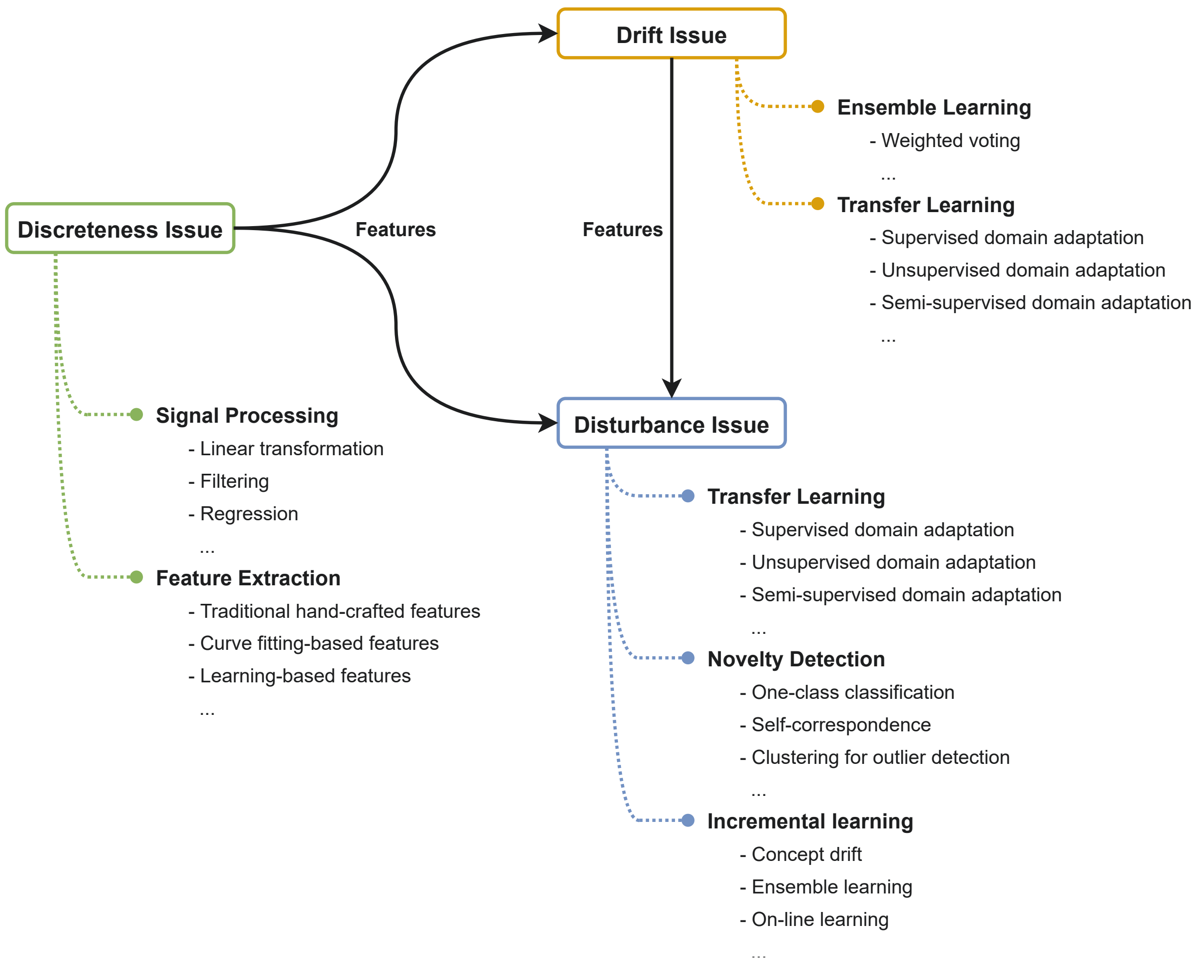


Figure 1.5: The solving strategy to the 3D issue.

system can be regarded as a representation of odour since other interfering factors of the system (e.g. relative stable environment, fixed operation parameters, and regulated concentration) have been controlled in the experiment. Accordingly, odour identification can be regarded as system identification. Here, a non-parametric modelling method, i.e. the finite impulse response (FIR) model, is applied to avoid the problem in pre-setting the order of the model. In addition, the use of a smooth and stable kernel function overcomes the ill-posed problem in calculating system parameters, which ensures a stable feature extraction.

- **Multiscale wavelet kernel regularization-based feature extraction algorithm**

  The major shortage of the above-proposed method is the requirement of an accurate setting of initial time (time of gas-on). However, in real applications, it is laborious to find such time point when the sensor's reading is contaminated by noises. Hence, an improved feature extraction algorithm based on multiscale wavelet kernel regularization is introduced to solve the problem. The multiscale wavelet kernel inherits the advantage from wavelet function in approximating arbitrary signals. Accordingly, finding the initial time can be avoided. Moreover, the use of multiscale wavelet-based regularization endues the FIR model a good ability in resistance to random noises.

ii) Proposition of signal processing algorithm:

- **Numerical differentiation algorithm using kernel-based Tikhonov regularization**

  As the simplest feature extraction algorithm, traditional hand-crafted features have been widely used in e-noses. Among the hand-crafted features, most of them are designed based on the 1^{st}- and 2^{nd}-order derivatives of the gas sensor's response curve. These features are called transient-state features. However, calculating the derivative (numerical differentiation) of a noisy signal is an ill-posed problem since it has a tendency to amplify the effect of noise. To tackle this problem, a novel numerical differentiation algorithm is proposed, which uses kernel-based Tikhonov regularization. The proposed algorithm is able to directly estimate high-order derivatives from the noise-contaminated signal without applying any smoothing technique (e.g. filtering and regression). After derivative estimation, accurate and stable transient-state features can be obtained.

iii) Proposition of problem-solving framework in real applications:

- **A framework for odour identification based on domain adaptation and one-class classification**

  In many odour identification tasks, the training and testing samples are collected under two different environmental conditions. To simplify the problem, we assume that the temperature, humidity, and background smell in each environment are relatively constant. Based on this assumption, the drift issue (i.e. reversible drift) can be alleviated by a domain adaptation algorithm, domain regularized component analysis (DRCA) [209]. Then, the disturbance issue (i.e. contingent disturbance) can be solved by a hidden Markov model (HMM)-based one-class classification method. The proposed framework is verified in a real application scenario displayed at the CeBIT (Centrum für Büroautomation, Informationstechnologie und

Telekommunikation) Australia 2019 Expo by successfully identifying four targets and two disturbance whiskies.

- **A framework for food freshness monitoring and evaluation based on hidden Markov models**

  A data-driven freshness monitoring and evaluation method is needed in food industry. To simplify the problem, we assume that all training and testing samples are collected under the same environmental conditions i.e. the temperature, humidity, and background smell are relatively constant. Based on this assumption, we proposed two frameworks for different tasks: i) To enable monitoring regardless the storage condition, we propose a monitoring method based on a single HMM trained only by fresh samples. Monitoring is realized by measuring the distance (i.e. likelihood) between the fresh sample and the tested sample. ii) When the whole life-span data of meat stored in a specific storage condition is available, a meat freshness evaluation method based on multiple HMMs is proposed. The freshness level is uncovered by a decoding algorithm (i.e. the Viterbi algorithm). Then, HMM for each level is trained and applied in parallel as a classifier to evaluate the freshness. The proposed frameworks are verified by monitoring and evaluating the freshness of three different meat products.

1.5 Thesis Outline

Arrangement of the thesis is as follows:

CHAPTER 1

This chapter presents an introduction of motivation and scope, possible application areas, faced challenges, contributions, and outline of the thesis.

CHAPTER 2

This chapter presents a literature review related to improving the feature quality for MOS-type sensor array-based e-noses, including signal processing algorithms and feature extraction algorithms.

CHAPTER 3

This chapter presents a novel feature extraction algorithm. We apply the impulse response of a kernel-based FIR model as the feature to represent the gas sensor's reading. Based on the normalized mean square error and information entropy, hyper-parameters of the kernel function are optimized to prevent the FIR model from overfitting. Based on these features, odour identification is realized by applying Gaussian mixture density hidden Markov models (GMM-HMMs). Also, hyper-parameter selection for GMM-HMM is realized according to the Bayesian information criterion (BIC) index and cross-validation. The dataset used to verify the algorithm is six different perfumes measured by the self-designed electronic nose system, UTS NOS.E II. Firstly, we validate the performance of the proposed feature extraction method in resistance to noise and compare it with other existed features. The modelling-based feature reaches the highest performance even without applying any filtering or smoothing techniques. Then, we compare the proposed combination of feature extraction and classification algorithms with other approaches. The proposed method outperforms other approaches reaching 93.56% in sensitivity and 98.71% in specificity.

The work of this chapter has been published in:

- T. Liu, W. Zhang, L. Ye, M. Ueland, S. L. Forbes, and S. W. Su, *A novel multi-odour identification by electronic nose using non-parametric modelling-based feature extraction and time-series classification*, Sensors and Actuators B: Chemical, 298 (2019), pp. 126690.

Chapter 4

Aiming to the shortage of the above proposed feature extraction algorithm, this chapter presents an improvement method. The impulse response is estimated from a FIR model constrained by a multiscale wavelet kernel regularization matrix. The kernel regularization matrix equips the proposed feature extraction method with an ability in resistance to random noise. A numerical experiment proves that, compared with single-scale kernel regularization, the use of multiscale wavelet kernel helps to achieve more stable and accurate impulse response estimation. Then, a field experiment is conducted to demonstrate the performance of the proposed features. This experiment aims to identify four different whiskeys measured by a self-designed e-nose, UTS NOS.E III. Under the framework of "feature extraction – domain adaptation", the classification result based on the proposed features outperforms those using other considered features. The accuracy of whiskey identification reaches 92.00%.

The work of this chapter has been submitted for publication:

- T. Liu, W. Zhang, J. Li, M. Ueland, S. L. Forbes, W. X. Zheng, and S. W. Su, *A multiscale wavelet kernel regularization-based feature extraction method for electronic nose*, Submitted to IEEE Transactions on Systems, Man, and Cybernetics: Systems. (*Under Review*)

CHAPTER 5

As the simplest feature extraction algorithm, traditional hand-crafted transient-state features have been widely used in the area of e-noses. However, the influence of noise in the calculation of numerical differentiation leads to the inaccuracy and instability in extracting these features. To tackle this issue, a novel numerical differentiation algorithm using kernel-based Tikhonov regularization is proposed in this chapter. The proposed method can provide accurate and stable transient-state features by directly estimating high-order derivatives from the noise-contaminated sensor's reading. The feature representation is a prerequisite for good performance of e-noses. Nevertheless, it should be noted that this performance in real applications can still be affected by other factors such as sensor drift and unknown odour disturbance. This issue can be addressed by applying a framework of domain adaptation and one-class classification. The proposed method and the adopted framework have been verified in a field experiment, which identifies the odour of four targets and two disturbance whiskies measured by UTS NOS.E III. The classification accuracy reaches 86.36%.

The work of this chapter has been submitted for publication:

- T. LIU, W. ZHANG, L. WANG, M. UELAND, S. L. FORBES, W. X. ZHENG, AND S. W. SU, *Numerical differentiation from noisy signals: a kernel regularization method to improve transient-state features for electronic nose*, Submitted to IEEE Transactions on Industrial Electronics. (*Under Review*)

CHAPTER 6

This chapter presents the use of UTS NOSE II with a data-driven approach, HMMs, to detect meat spoilage. The data-driven approach we proposed include: i) A training algorithm using segmental rapid centroid estimation (RCE) to increase the overall stability of the HMM optimization algorithm; ii) A monitoring algorithm using a single HMM trained only by fresh samples to track the change of meat freshness in real-time when the storage condition is not specified; iii) An HMM-based decoding algorithm to cluster the freshness level when the whole life-span data of meat stored in a specific storage condition is available. Then, HMM for each freshness level is trained and applied in parallel as freshness evaluation models to classify the tested meat samples. In freshness monitoring, case studies on fish, beef, and chicken demonstrate the effectiveness and potential of this method. In freshness evaluation, we observed sensitivity and specificity of 96.32%±2.83% and 99.07%±0.69%, 99.09%±2.40% and 99.82%±0.48%, as well as, 99.35%±0.27% and 98.31%±0.71% for fish, beef, and chicken, respectively.

The work of this chapter has been published in:

- T. LIU, W. ZHANG, M. YUWONO, M. ZHANG, M. UELAND, S. L. FORBES, AND S. W. SU, *A data-driven meat freshness monitoring and evaluation method using rapid centroid estimation and hidden Markov models*, Sensors and Actuators B: Chemical, 311 (2020), pp. 127868.

CHAPTER 7

This chapter summarizes the work of the paper and presents prospects for future research.

Chapter 2

Literature Review

2.1 Introduction

Feature extraction is a bridge connecting gas sensor technology and pattern recognition techniques. It converts the gas sensor array's reading into a simple representation that can be processed efficiently by pattern recognition algorithms. A discriminant and stable feature representation is also a prerequisite for good performance of e-noses. There are two ways to improve feature extraction algorithms: i) The indirect method is to apply signal processing algorithms to enhance the quality of the original response for subsequent feature extraction; ii) The direct method is to design a feature extraction algorithm with anti-noise ability for the response curve.

A typical response curve of MOS-type sensors is shown in Fig. 2.1. In the area of artificial olfaction, the entire response curve is called one sniff [91]. One sniff represents one sample (or instance) in pattern recognition algorithms. It can be clearly seen from the figure that a sniff consists of three stages: i) The baseline stage represents the sensor's response in fresh air; ii) The absorption (or response) stage represents the interaction between the sensing material and odour stimuli; iii) The desorption (or recovery) stage represents washing the residual odour on the surface of the sensing material with fresh air. Generally, MOS-type sensors take a longer time in the desorption stage than in the absorption stage. Feature extraction algorithms are usually performed in the first two stages.

In this chapter, we review the two fundamental techniques that have been applied

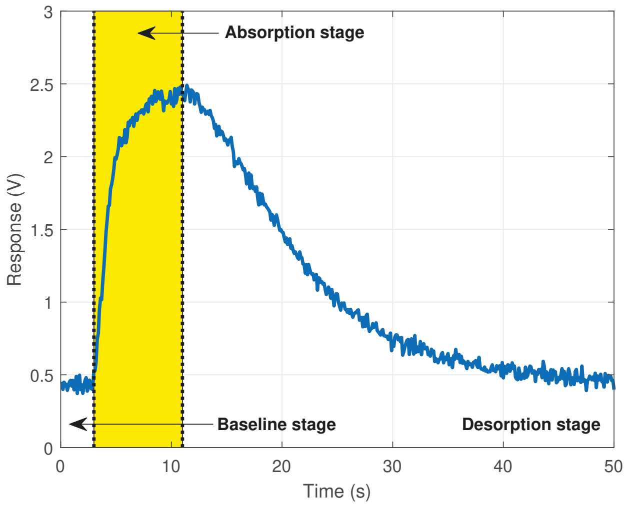


Figure 2.1: Typical response curve of MOS-type gas sensors. The example displayed is responses of TGS 2602 gas sensor to the Chanel Chance perfume sample.

in MOS-type gas sensor array-based e-noses. Section 2.2 presents signal processing algorithms. Section 2.3 presents feature extraction algorithms. The algorithms reviewed in this section are the main source of comparison algorithms for our proposed methods.

2.2 Review of Signal Processing Algorithms

2.2.1 Linear Transformation Algorithms

Linear transformations are applied to reduce the influence of discreteness issue caused by odour concentration variation and sensor manufacturing discrepancy [211]. The most widely applied linear transformation algorithms include differential response (2.1), fractional response (2.2), normalization (2.3), and standardization (2.4) [106, 107, 127, 162, 210, 211]:

$$\hat{y}(t) = y(t) - \min y(t) \tag{2.1}$$

$$\hat{y}(t) = \frac{y(t) - \min y(t)}{\min y(t)} \tag{2.2}$$

$$\hat{y}(t) = \frac{y(t) - \min y(t)}{\max y(t) - \min y(t)} \tag{2.3}$$

$$\hat{y}(t) = \frac{y(t) - \operatorname{mean} y(t)}{\operatorname{std} y(t)} \tag{2.4}$$

where $y(t)$ and $\hat{y}(t)$ represent the sensor's reading before and after transformation, respectively. The performance of above linear transformations is shown in Fig. 2.2.

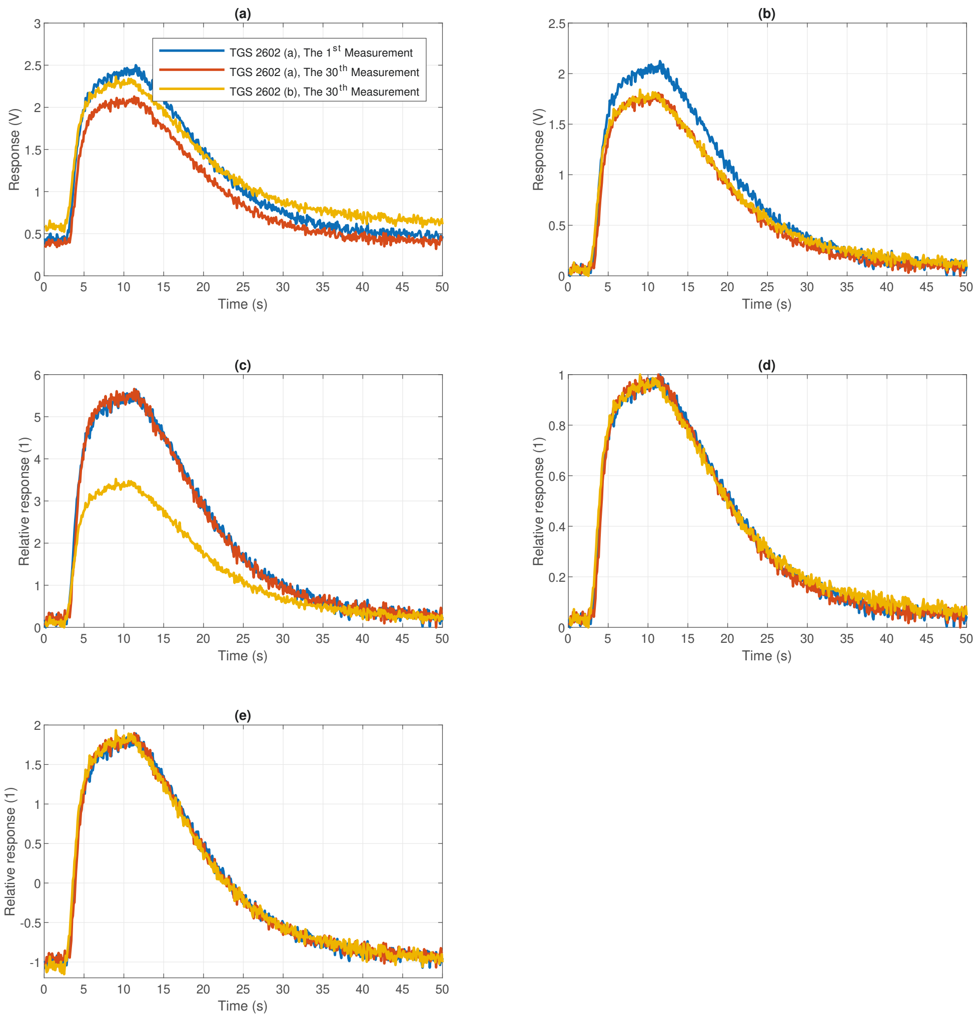


Figure 2.2: Representative example of linear transformations. The example displayed is responses of TGS 2602 sensors to the Chanel Chance perfume sample. (a) Original response curves. (b) Differential conductance. (c) Fractional conductance. (d) Normalization. (e) Standardization

The choice of linear transformation algorithm depends on the particular task in real applications. Differential response and fractional response are called baseline manipulation methods, which can align the baseline and retain information related to the response (concentration). Normalization and standardization perform better than baseline manipulation methods in reducing the effect of the discreteness issue caused by concentration variation and sensor manufacturing discrepancy. However, the concentration-related information has been totally eliminated after such transformations. Normalization is a globally scaling method, which can be influenced by outliers. Standardization is not be affected by outliers, but it cannot align the baseline.

2.2.2 Smoothing Algorithms

Smoothing algorithms are applied to reduce the influence of discreteness issue caused by sensor noise. The most widely used smoothing techniques include filtering and regression. Representative smoothing algorithms are introduced in the following parts.

2.2.2.1 Wavelet Transformation Domain Filter

Recall that the thermal agitation of electrons inside the sensor causes thermal noise in sensor's reading. Thermal noise in an ideal resistor can be considered as Gaussian white noise [10]. The wavelet transformation domain filter (WTDF), as one of wavelet filters designed to tackle the white noise, has been adopted to process e-nose signals [105, 106]. The WTDF assumes that, in a noise-contaminated signal, the components with high correlation should be passed through the filter [191]. A simple implement of WTDF is shown in Algorithm 1.

2.2.2.2 Polynomial Regression

Polynomial regression is the simplest but commonly used smoothing technique in signal processing. It uses a linear combination of basic functions $\{\phi_m(t)|m \in M\}$ to model a smooth approximation of signal $\boldsymbol{y} = [y_{[1]}, \ldots, y_{[N]}] \in \mathbb{R}^N$. If the input variables are defined as $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{N\times M}$, where $\boldsymbol{x}_n = [\phi_1(n), \ldots, \phi_M(t)]^\intercal \in \mathbb{R}^M$, the combination parameters $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_M] \in \mathbb{R}^M$ can be obtained by [68, 141, 218]:

$$
\begin{aligned}
\boldsymbol{\omega} &= \arg\min_{\boldsymbol{\omega}} J(\boldsymbol{\omega}) \\
&= \arg\min_{\boldsymbol{\omega}} \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\omega}\|_2^2 \qquad (2.5a) \\
&= (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\boldsymbol{y} \qquad (2.5b)
\end{aligned}
$$

where (2.5b) is a closed-form solution obtained by applying the least square (LS).

Algorithm 1: Wavelet Transformation Domain Filter

Input: Signal $\boldsymbol{y} \in \mathbb{R}^N$, scale $M \in \mathbb{Z}^+$, mother wavelet $\psi(t)$, and threshold $P_{\text{thresh}} > 0$.

Output: Filtered signal $\hat{\boldsymbol{y}} \in \mathbb{R}^N$.

Initialization: Filtered wavelet coefficient matrix $\mathbf{W}_F \leftarrow \mathbf{0}_{M\times N}$.

Calculate wavelet coefficient matrix $\mathbf{W} \leftarrow [\boldsymbol{w}_{1,n}^\top, \ldots, \boldsymbol{w}_{M,n}^\top] \in \mathbb{R}^{M\times N}$, where $\boldsymbol{w}_{m,n}^\top \in \mathbb{R}^N$ is the discrete wavelet transformation of $\boldsymbol{y}$ on level m by $\psi_{m,n}(t)$.

for $m = 1, \ldots, M-1$ **do**

    $P_w \leftarrow \sum_{n\in N} \boldsymbol{w}_{m,n} \circ \boldsymbol{w}_{m,n}$.   ▷ ∘: Hadamard product

    while $P_w > P_{\text{tresh}}$ **do**

        $\boldsymbol{c} \leftarrow \boldsymbol{w}_{m,n} \circ \boldsymbol{w}_{m+1,n}$.

        $P_c \leftarrow \sum_{n\in N} \boldsymbol{c} \circ \boldsymbol{c}$.

        $\hat{\boldsymbol{c}} \leftarrow \sqrt{\frac{P_w}{P_c}} \cdot \boldsymbol{c}$.

        for $n = 1, \ldots, N$ **do**

            if $|\hat{c}_{[n]}| > |w_{m,n[n]}|$ **then**

                $\mathbf{W}_{F[m,n]} \leftarrow w_{m,n[n]}$.

                $w_{m,n[n]} \leftarrow 0$.

            end

        end

        $P_w \leftarrow \sum_{n\in N} \boldsymbol{w}_{m,n} \circ \boldsymbol{w}_{m,n}$.

    end

end

Reconstruct signal $\hat{\boldsymbol{y}}$ from $\mathbf{W}_F$ via the inverse discrete wavelet transformation.

return $\hat{\boldsymbol{y}}$.

When the optimal parameters are obtained, the smoothed signal can be calculated as:

$$\hat{\boldsymbol{y}} = \mathbf{X}\boldsymbol{\omega}. \tag{2.6}$$

Improvement of polynomial regression is made from two aspects:

i) Add regularization terms to (2.5a) to avoid model overfitting, such as ridge regression (2.7) [68], LASSO (least absolute shrinkage and selection operator) regression (2.8) [141], and elastic net regression (2.9) [218]:

$$J(\boldsymbol{\omega}) = \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\omega}\|_2^2 + \alpha\|\boldsymbol{\omega}\|_2^2 \tag{2.7}$$

$$J(\boldsymbol{\omega}) = \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\omega}\|_2^2 + \beta\|\boldsymbol{\omega}\|_1 \tag{2.8}$$

$$J(\boldsymbol{\omega}) = \|\boldsymbol{y} - \mathbf{X}\boldsymbol{\omega}\|_2^2 + \alpha\|\boldsymbol{\omega}\|_2^2 + \beta\|\boldsymbol{\omega}\|_1 \tag{2.9}$$

where $\|\boldsymbol{\omega}\|_2^2$ and $\|\boldsymbol{\omega}\|_1$ represents $L2$ and $L1$ regularization term, respectively. α and β denotes trade-off parameters.

A closed-form solution can also be obtained for (2.7) via the regularized least square (ReLS) since $L2$ is a differentiable convex term. However, (2.8) lacks closed-form solutions since $L1$ term is not differentiable. A numerical solution can be obtained by using least angle regression (LARS) [85]. For elastic net regularization, (2.9) can be converted to the basic form of LASSO [196], and then solved as LASSO regression.

ii) Change the global regression to local regression with moving average, such as locally weighted regression (LWR) [26]. A simple implementation of LWR is shown in Algorithm 2.

Algorithm 2: Locally Weighted Regression (with Gaussian weight)

Input: Signal $\boldsymbol{y} \in \mathbb{R}^N$, input variable $\mathbf{X} \in \mathbb{R}^{N \times M}$, and hyper-parameter $\tau \in \mathbb{R}$.
Output: Smoothed signal $\hat{\boldsymbol{y}} \in \mathbb{R}^N$.
Initialization: Smoothed signal $\hat{\boldsymbol{y}} \leftarrow \mathbf{0}$, index $\boldsymbol{t} \leftarrow [1, \ldots, N]^\intercal \in \mathbb{R}^N$, and $\mathbf{1}_N \leftarrow [1, \ldots, 1]^\intercal \in \mathbb{R}^N$.
for $n = 1, \ldots, N$ **do**
    $\boldsymbol{a} \leftarrow \exp\left\{\frac{-(n \cdot \mathbf{1}_N - \boldsymbol{t})^{\circ 2}}{\tau}\right\}$ ▷ Gaussian weight
    $\mathbf{A} \leftarrow \mathrm{diag}(\boldsymbol{a})$
    $\boldsymbol{\omega} \leftarrow (\mathbf{X}^\intercal \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{A} \boldsymbol{y}$
    $y_{[n]} \leftarrow \mathbf{X}_{[n,:]} \boldsymbol{\omega}$
end
return $\hat{\boldsymbol{y}}$.

2.2.2.3 Tikhonov Regularization-Based Smoothing

Tikhonov regularization is considered to be an extension of ridge regression. It adds constraint on the d^{th}-order derivatives of the unknown vectors (parameters). Accordingly, ridge regression can be considered as a 0^{th}-order Tikhonov regularization.

Let $\boldsymbol{q}_1 = [1, -1]$ be the FIR approximation of the 1^{st}-order derivative operator. Based on this, a d^{th}-order ($d \geqslant 2$) derivative operator can be constructed as [147]:

$$\boldsymbol{q}_d = \boldsymbol{q}_{d-1} \circledast \boldsymbol{q}_1 \tag{2.10}$$

where $\circledast$ denotes discrete convolution. From (2.1), differentiation matrix can be constructed with zero-padding as:

$$\mathbf{Q}_d = \begin{bmatrix} q_{d[1]} & q_{d[2]} & \cdots & q_{d[d+1]} & 0 & \cdots & 0 \\ 0 & q_{d[1]} & \cdots & q_{d[d]} & q_{d[d+1]} & \cdots & 0 \\ 0 & 0 & \cdots & q_{d[d-1]} & q_{d[d]} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & q_{d[d+1]} \end{bmatrix} \in \mathbb{R}^{N\times N}. \tag{2.11}$$

Accordingly, the estimation of a d^{th}-order Tikhonov regularization is given by [147]:

$$\begin{aligned} \hat{\boldsymbol{y}} &= \arg\min_{\hat{\boldsymbol{y}}} J(\hat{\boldsymbol{y}}) \\ &= \arg\min_{\hat{\boldsymbol{y}}} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|_2^2 + \lambda \|\mathbf{Q}_d \hat{\boldsymbol{y}}\|_2^2 \\ &= (\mathbf{I}_N + \lambda \mathbf{Q}_d^\top \mathbf{Q}_d)^{-1} \boldsymbol{y} \end{aligned} \tag{2.12}$$

where $\mathbf{Q}_d \hat{\boldsymbol{y}}$ denotes the d^{th}-order derivative of $\hat{\boldsymbol{y}}$. $\mathbf{I}_N$ is a N-by-N identity matrix. λ denotes the trade-off parameter.

2.2.2.4 Support Vector Regression

Support vector regression (SVR) is a classical algorithm in nonlinear regression. It finds a line to fit a nonlinear function with a flexible error bound in a feature space mapped from the original input space.

SVR problem can be formed as [7, 159]:

$$\begin{aligned} \min_{\boldsymbol{\omega}, b, e_n, \hat{e}_n} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 + \lambda \sum_{n=1}^{N} (e_n + \hat{e}_n) \\ \text{s.t.} \quad & \boldsymbol{\omega}^\top \phi(\boldsymbol{x}_n) + b - y_{[n]} \leqslant \epsilon + e_n, \\ & y_{[n]} - \boldsymbol{\omega}^\top \phi(\boldsymbol{x}_n) - b \leqslant \epsilon + \hat{e}_n, \\ & e_n \geqslant 0, \hat{e}_n \geqslant 0, n = 1, \ldots, N \end{aligned} \tag{2.13}$$

where $\boldsymbol{\omega}$ and b are parameters of regression model to be estimated. e_n and $\hat{e}_n$ denote slack variables. ϵ represents the bound of acceptable deviation. λ is a trade-off parameter. $\phi(\cdot)$ is a mapping from the input space to feature space.

The optimization of (2.13) can be simplified by converting it to its dual problem as [7, 159]:

$$
\begin{aligned}
\max_{\boldsymbol{\alpha},\hat{\boldsymbol{\alpha}}} \quad & \sum_{n=1}^{N} y_{[n]}(\hat{\alpha}_n-\alpha_n)-\epsilon(\hat{\alpha}_n+\alpha_n) \\
& -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(\hat{\alpha}_n-\alpha_n)(\hat{\alpha}_m-\alpha_m)\boldsymbol{x}_n^{\mathsf{T}}\boldsymbol{x}_m \\
\text{s.t.} \quad & \sum_{m=1}^{N}(\hat{\alpha}_n-\alpha_n)=0, \\
& 0\leqslant \hat{\alpha}_n,\alpha_n\leqslant\lambda.
\end{aligned}
\tag{2.14}
$$

Then, (2.14) can be solved by a efficient optimization algorithm, i.e. sequential minimal optimization (SMO) [207]. The estimated model is in the form of [7, 159]:

$$
\hat{y}(\boldsymbol{x})=\sum_{n=1}^{N}(\hat{\alpha}_n-\alpha_n)\kappa\langle\boldsymbol{x}_n,\boldsymbol{x}\rangle+b \tag{2.15}
$$

where $\kappa\langle\cdot,\cdot\rangle$ denotes kernel function representing the inner product $\phi(\cdot)^{\mathsf{T}}\phi(\cdot)$.

Here, we list some widely applied kernel functions, such as radial basis function (RBF) kernel, Laplace kernel, polyomial kernel, and sigmoid kernel [125]:

$$
\kappa\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle=\exp\left\{-\frac{\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2}{2\theta_1^2}\right\} \tag{2.16}
$$

$$
\kappa\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle=\exp\left\{-\frac{|\boldsymbol{x}_1-\boldsymbol{x}_2|}{\theta_1}\right\} \tag{2.17}
$$

$$
\kappa\langle\boldsymbol{x},\boldsymbol{y}\rangle=(\boldsymbol{x}_1^{\mathsf{T}}\boldsymbol{x}_2)^d \tag{2.18}
$$

$$
\kappa\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle=\tanh(\theta_1\boldsymbol{x}_1^{\mathsf{T}}\boldsymbol{x}_2+\theta_2) \tag{2.19}
$$

where θ_i and d are hyper-parameters of the considered kernel function.

2.3 Review of Feature Extraction Algorithms

According to the recent research outcomes, we can roughly divide the existed feature extraction approaches into four categories: i) Traditional hand-crafted features; ii) Curve fitting-based features; iii) Learning-based features; and iv) Frequency-domain features.

2.3.1 Traditional Hand-Crafted Features

Traditional hand-crafted features refer to piecemeal features extracted directly from the sensor's response curve [193]. These characteristics reflect some physical information about the interaction between the odour stimulus and the sensing material, such as concentration, speed, acceleration, *etc.*.

Hand-crafted features are designed from the basic elements shown in Fig. 2.3. Based on this, the constructed features are listed in Table 2.1. It should be noted that we cannot list all hand-crafted features in this table. The one designed from the structural characteristics of the response curves can be classified as this category. We can further classify the hand-crafted features into steady-state features and transient-state features. As we can see from the table, transient-state features contribute most of the hand-crafted features.

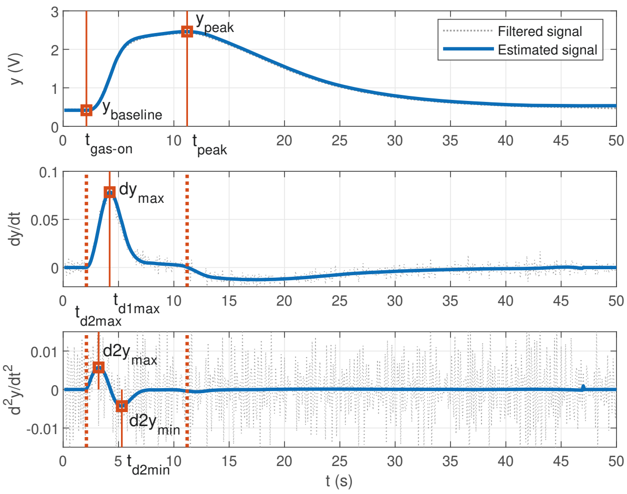


Figure 2.3: Basic elements for hand-crafted features. The example displayed is responses of TGS 2602 sensors to the Chanel Chance perfume sample. The blue solid line is the estimation result of our proposed method Chapter 5. The gray dashed line is the filtering result of WTDF.

Traditional hand-crafted features are easy to be influenced by the discreteness issue. The steady-state feature is sensitive to sensor manufacturing discrepancy and odour concentration variation. Once normalization or standardization is adopted to reduce such influences, the discriminant information it has provided will be totally eliminated, as illustrated in Fig. 2.2(d) and (e). The transient-state features designed from derivatives are sensitive to noise since numerical differentiation has a tendency to amplify the influence of noise [120]. As illustrated in Fig. 2.3, although the noise in the filtered signal has been suppressed to a small level, we can still have an large influence in the 2^{nd}-order derivative. The stability and accuracy in feature extraction are significantly challenged under such conditions. Despite such shortages, traditional features are still

Table 2.1: Representative traditional hand-crafted features

Feature	Construction	Description	Category
f_1	$y_{\text{peak}} - y_{\text{base}}$	Peak value	Steady-state feature
f_2	$t_{\text{peak}} - t_{\text{gas-on}}$	Response time	Transient-state feature
f_3	$\lvert dy_{\max}\rvert$	Maximum 1^{st}-order derivative	
f_4	$t_{\text{d1max}} - t_{\text{gas-on}}$	Time interval between t_{d1max} and $t_{\text{gas-on}}$	
f_5	$t_{\text{peak}} - t_{\text{d1max}}$	Time interval between t_{peak} and t_{d1max}	
f_6	$d2y_{\max}$	Maximum 2^{nd}-order derivative	
f_7	$t_{\text{d2max}} - t_{\text{gas-on}}$	Time interval between t_{d2max} and $t_{\text{gas-on}}$	
f_8	$t_{\text{peak}} - t_{\text{d2max}}$	Time interval between t_{peak} and t_{d2max}	
f_9	$d2y_{\min}$	Minimum 2^{nd}-order derivative	
f_{10}	$t_{\text{d2min}} - t_{\text{gas-on}}$	Time interval between t_{d2min} and $t_{\text{gas-on}}$	
f_{11}	$t_{\text{peak}} - t_{\text{d2min}}$	Time interval between t_{peak} and t_{d2min}	
f_{12}	$t_{\text{d2min}} - t_{\text{d2max}}$	Time interval between t_{d2min} and t_{d2max}	

popular in many applications, such as gas sensor array optimization and flow rate optimization [135, 162], due to their simplicity and physical significance.

2.3.2 Curve Fitting-Based Features

To tackle the influence of sensor noise, curve fitting-based feature extraction algorithms are proposed. Curve fitting methods approximate the discrete and noised sensor's response by using continuous and smooth analytical functions. The coefficients of curve fitting functions (e.g. polynomial regression, exponential regression, and sigmoidal regression [36, 193, 194]) are applied as a feature vector to represent the response curve. We take the polynomial regression as an example. The polynomial curve fitting-based features can be extracted as illustrated in Fig. 2.4.

In comparison with the traditional method, features are not required to be designed, as the coefficients of curve fitting indicate the structural characteristics of the response curve. However, this method only provides a limit number of features. The number of features is decided by the degree of the curve fitting function. If the degree is too high, the fitting function will be overfitting. As illustrated in Fig. 2.4(c) and (d), a large value is generated for the largest degree term, while the rests are close to zero. In this case, the high-degree fitting function mainly focuses on the detail of the response curve but ignores its general characteristics. The ability of the feature in generalization is hence reduced. However, if the degree is too low, the limited features shall not discriminant

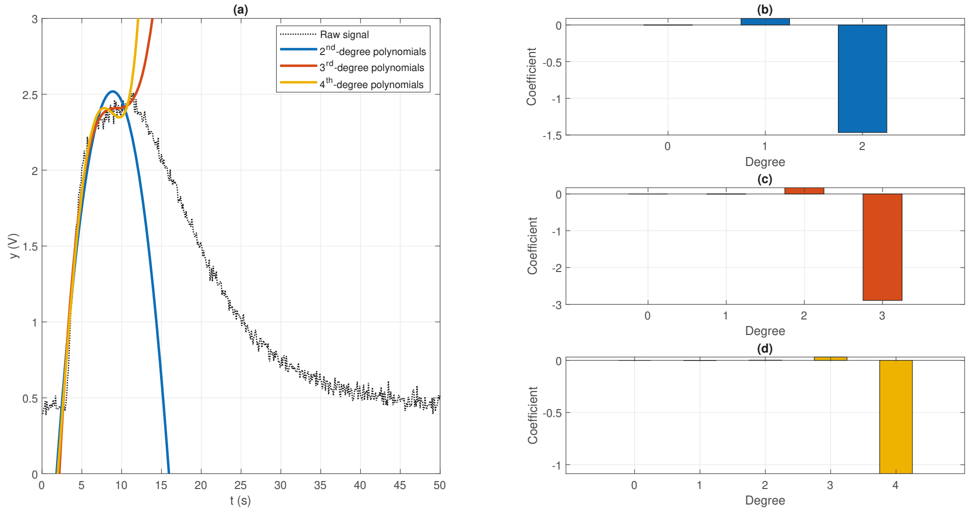


Figure 2.4: Representative example of polynomial curve fitting-based feature extraction. The example displayed is responses of TGS 2602 sensors to the Chanel Chance perfume sample. (a) Curve fitting in the absorption stage. (b) Coefficients of 2^{nd}-degree polynomials. (c) Coefficients of 3^{rd}-degree polynomials. (d) Coefficients of 4^{th}-degree polynomials.

two similar response curves. It should be noted that, once the signal is contaminated by noise, it is not trivial to find the optimal curve fitting degree that should be pre-set before coefficient estimation [79].

2.3.3 Learning-Based Features

A modern approach for extracting odour features is feature learning. Feature learning allows the algorithm to discover the representations needed for classification automatically. Artificial neural networks (ANNs) are usually selected to deal with this task. Feature learning methods can be either supervised or unsupervised. Supervised feature learning (e.g. time-delay neural network [40, 148], long short-term memory (LSTM) networks [166], convolution neural networks (CNN) [96]) allows the algorithm to both learn the features and use them to perform classification. Unsupervised feature learning (e.g. sparse auto-encoder (SAE) [217], restrict Boltzmann machine (RBM) [90, 112]) is specifically attractive to data that is unintuitive and difficult to interpret. Besides the ANNs, features can also be learnt from sparse coding, such as dictionary learning [66]. A representative example of learning-based feature extraction is illustrated in Fig. 2.5.

Learning-based features are convenient to use. One can directly feed the sensor

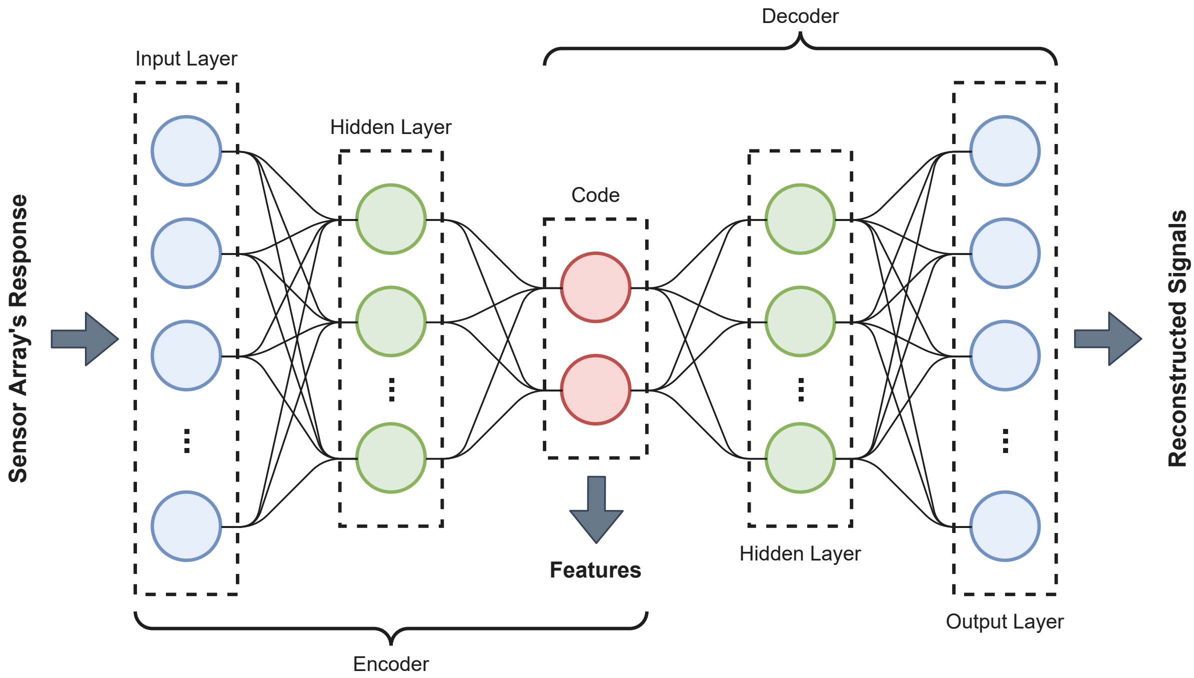


Figure 2.5: Representative example of SAE-based feature learning.

array's responses to a neural network without applying any pre-processing technique. However, a large number of training data is necessary to ensure an accurate learnt representation.

2.3.4 Frequency-Domain Features

Frequency-domain features can be extracted when sensor's responses are modulated by a periodic signal. Recall that there are two ways to realize modulation: i) Design e-nose systems with a temperature modulation module [47, 54, 201, 206], as shown in Fig. 2.6(a). This module aims to heat the sensor with a periodically changing temperature. ii) Design e-nose systems with a temperature modulation module [152, 181], as shown in Fig. 2.6(b). This module aims to switch the gas input periodically. Under such conditions, frequency-domain features can be extracted. Wavelet-based features constitute a major part of this category, such as the spectrum, energy, and entropy of wavelet (package) decomposition [69, 152]. Besides, the use of other frequency-domain analysis tools, like the Fourier, the short-time Fourier, and the Gabor transformation, is also reported [65, 69].

These features provide a wealth of discriminate information compared with their time-domain counterparts. However, these features require a special design of the modulation

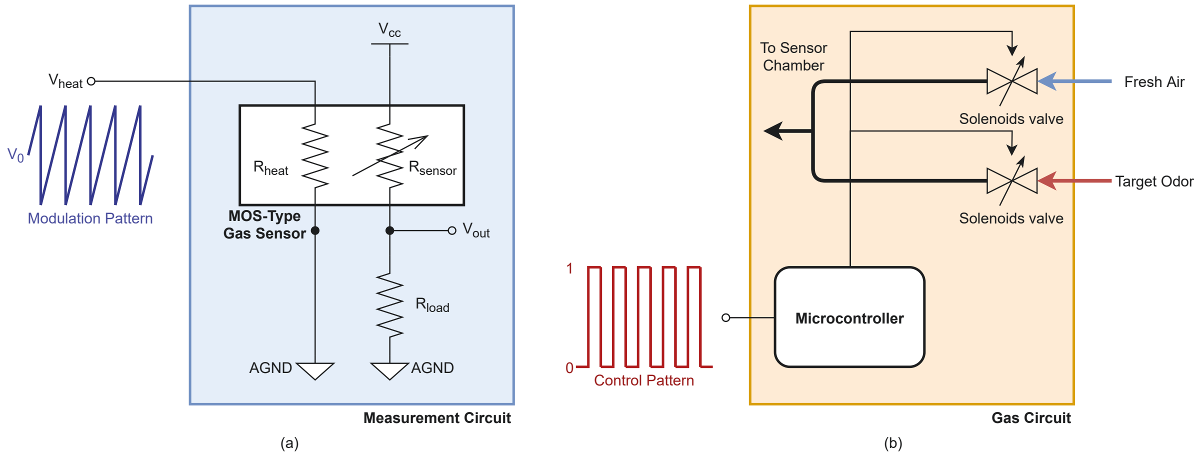


Figure 2.6: Illustrative example of modulation system. (a) Temperature modulation module. (b) Flow rate modulation module

system, which will also increase the manufacturing cost of the e-nose. Moreover, for temperature modulation, the performance of the most commercially available MOS-type gas sensor will be influenced if the pulse width modulation (PWM) signal is used to drive these sensors. MOS-type gas sensors need to work at a specific temperature to get the best performance, therefore if PWM signal is used to control the power, getting the appropriate heating temperature for the gas sensors is difficult to obtain. This method is not recommended by the commercial gas sensor manufactures, unless they add the related description in the technical data sheet (e.g. TGS2444, TGS8100).

2.3.5 Other Features

2.3.5.1 Linguistic-Level Features

The result of the first-level pattern recognition algorithm based on the above-mentioned features is used as the input of the second-level pattern recognition algorithm. This kind of two-level pattern recognition algorithms is also known as decision-level fusion. The result of the first-level pattern recognition can be regarded as linguistic-level features [192].

2.3.5.2 Comprehensive Features

Comprehensive features refer to the new features obtained by the fusion or combination of the above-mentioned features at the feature level [77]. The widely used feature fusion or combination algorithms include principle component analysis (PCA), kernel PCA (KPCA), genetic programming (GP), and SAE.

CHAPTER 3

FEATURE EXTRACTION BASED ON NON-PARAMETRIC MODELLING

3.1 Introduction

For a specific application, when the designing of the sensor array and hardware structure is completed, feature extraction and classification methods matter in improving the performance of the e-nose system. A useful feature extraction method will facilitate the subsequent classification procedure. Recall that we have classified the existed feature extraction algorithms into four categories in the previous chapter: traditional hand-crafted features, curve fitting-based features, learning-based features, and frequency-domain features.

The traditional hand-crafted features are deliberately designed piecemeal representations with the physical meaning [199, 215]. Their design requires experts' experience. Among the traditional features, the steady-state feature is most widely used for its simplicity in computing [199, 208]. However, the information provided by this feature would be totally eliminated when normalization is adopted to reduce the effect of concentration fluctuation and sensor discrepancy [106]. Derivative contributes the rest traditional hand-crafted features [107]. However, the stability is significantly challenged since numerical differential operator would amplify the influence of sensor noises. The curve fitting-based features is hence proposed to alleviates the influence of noise to some extent. These methods apply the parameter of curve fitting function or parametric modelling as features, such as polynomial fitting, exponential fitting, and autoregressive model [112]. However, once the signal is contaminated by noise, it is not trivial to find the optimal curve fitting or modelling order that should be pre-set before parameter estimation [79]. In recent years, it has been proposed that learning-based approaches can directly learn the representation from the response curve without being affected by noises, such as sparse coefficient of dictionary learning, encoding of auto-encoders, and hidden layer's output of deep neural networks [202]. However, a large number of training data is necessary to ensure an accurate learnt representation. When the sensor array is operated under a periodic heating voltage (i.e. temperature modulation [65]) or a periodic switching airflow (i.e. airflow modulation [152]), the sensor's response will contain frequency information. Under such conditions, frequency-domain features can be extracted. Wavelet-based features constitute a major part of this category, such as the spectrum, energy, and entropy of wavelet (package) decomposition [69, 152]. Besides, the use of other frequency-domain analysis tools, like the Fourier, the short-time Fourier, and the Gabor transformation, is also reported [65, 69]. However, the shortage is the requirement of a specially designed e-nose platform. Since our e-nose is not equipped with a temperature or airflow modulation system, the frequency-domain features would only contain trivial information about noises. Hence,we do not consider this category anymore in our thesis.

Apart from the approaches mentioned above, features can also be extracted based on system modelling. From the perspective of system identification, the e-nose device is regarded as a black-box system. In this study, it is assumed the input of the gas sensor, the stimulation of the target odour, can be approximately treated as a step input. However, the actual system output is subjected to divergence from the ideal input by interfering factors [33], such as the properties of gas molecules, gas circuit, type of sensor and its installation site, temperature, and relative humidity. Different interfering factors provide different model parameters to the black-box system. When environmental and device operation parameters are set to the same, the only interfering factor is the properties of gas molecules. Therefore, odour identification can be achieved by realising system identification. It is reasonable to consider the solution of system identification as features for target odours. Inspired by the above hypothesis, our work is developed accordingly. In addition, due to the fact that the complexity of the parametric model is bounded even if the amount of data is unbounded. In this study, to well improve the flexibility, we developed a non-parametric modelling-based approach.

The primary contributions of this chapter are summarized as follows:

- We proposed a feature extraction method according to a non-parametric modelling-based approach, namely, kernel-based model estimation, and developed the corresponding parameters optimization algorithm.

- To take full advantage of the discriminant information contained in the features, we combined the proposed feature extraction method with GMM-HMMs. Also, a parameter selection method for GMM-HMMs is provided.

- We developed an automatic initial time detection method based on the estimated model and achieved automatic odour identification with the proposed method.

The proposed feature extraction method shows excellent ability in noise suppression. Moreover, the combination method improves the performance of the self-designed e-nose system. Compared with traditional feature extraction approaches, the proposed method provides non-hand-design features. Compared with curve-fitting approaches, the proposed method extracts more structural characteristics of the original curve. Compared with feature-learning approaches, the proposed method achieves higher classification accuracy.

The rest of the chapter is structured as follows: Section 3.2 is devoted to describing the whole methodology of modelling-based feature extraction. Experiment setup is introduced in Section 3.3. Then, the experiment results and discussion will be presented in Section 3.4. Finally, the conclusion of this chapter will be outlined in Section 3.5.

3.2 Methodology

The response of e-nose to target odours is comprised of three stages: baseline stage, absorption stage and desorption stage. In this study, features were extracted from the data acquired in first two stages, as the third stage requires a long time for the chemical sensors to desorb the odour molecules and let the sensor's response return to the baseline. Recall that the parameters of the dynamic model can be applied as features for target odours. It performs as an improvement of conventional curve fitting-based feature extraction. Compared with curve fitting-based methods, dynamical modelling takes both the current input and the previous behaviour of the system into consideration, and therefore, can extract more features. But, it should be noted that traditional dynamical modelling approaches (e.g. ARX model [40]) are usually parametric models. A specified order of the system should be decided before estimating the parameters; that is, the complexity of the parametric model is usually bounded. Actually, it is difficult to identify the exact order when the input is a step signal, and the observation is polluted with noise. Non-parametric modelling approach, e.g. FIR model, can be applied to solve this problem. The estimation of the FIR model does not need to consider the order of the system. However, finding the impulse responses from the noised observation is an ill-posed inverse problem. Therefore, in this section, a newly developed non-parametric modelling approach [131, 132, 195], was introduced, which adds a regularization term with a well-selected kernel function. The kernel-based regularization helps prevent overfitting caused by the noise-polluted observations and obtain a more accurate and robust estimation.

3.2.1 Kernel-Based Model Estimation

Let $\tilde{v}_s(t)$ be the normalized actual response of the s^{th} sensor and $\tilde{u}_s(t)$ be its corresponding ideal step input. Then, a noiseless output, $\hat{v}_s(t|\theta_s)$, can be obtained by calculating the convolution between the D^{th}-order finite impulse responses, $\theta_s(t)$, and $\tilde{u}_s(t)$ as:

$$\begin{aligned}\hat{v}_s(t|\theta_s) &= (\theta_s * \tilde{u}_s)(t)\\ &= \sum_{\tau=1}^{D} \theta_s(\tau)\tilde{u}_s(t-\tau)\\ &= \tilde{\boldsymbol{x}}_t^{\mathsf{T}}\boldsymbol{\theta}_s, \quad t \in T\end{aligned} \tag{3.1}$$

where $\tilde{\boldsymbol{x}}_t = [\tilde{u}_s(t-1), \tilde{u}_s(t-2), \ldots, \tilde{u}_s(t-D)]^{\mathsf{T}}$ is a vector presentation of $\tilde{u}_s(t-\tau)$.

The actual normalized response can also be vectorized as: $\tilde{V}_s = [\tilde{v}_s(1), \tilde{v}_s(2), \ldots \tilde{v}_s(T)]^{\mathsf{T}}$. Accordingly, the finite impulse responses of the s^{th} sensor can be written in the matrix form as:

$$\tilde{V}_s = \tilde{\mathbf{X}}_s\boldsymbol{\theta}_s + \tilde{E}_s \tag{3.2}$$

where $\tilde{\mathbf{X}}_s = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_T]^{\mathsf{T}}$, and $\tilde{E}_s$ denotes the disturbance vector of the finite responses model.

In order to avoid any ill-conditioned solution and to obtain a better finite impulse responses model, a newly regularization method was introduced. By applying a Mercer Kernel, the $\boldsymbol{\theta}_s$ can be smoothed by projecting it into a Reproducing Kernel Hilbert Space (RKHS). $\boldsymbol{\theta}_s$ can be calculated by minimizing the following cost function as:

$$\boldsymbol{\theta}_s = \arg\min_{\boldsymbol{\theta}_s}(\tilde{V}_s - \tilde{\mathbf{X}}_s\boldsymbol{\theta}_s)^{\mathsf{T}}(\tilde{V}_s - \tilde{\mathbf{X}}_s\boldsymbol{\theta}_s) + \boldsymbol{\theta}_s^{\mathsf{T}}\mathbf{K}^{-1}\boldsymbol{\theta}_s \tag{3.3}$$

where the first squared error term implies the modelling disturbance, while the last term is a regularizer, which denotes the squared Euclidean norm in RKHS. $\mathbf{K}$ indicates a D-by-D sized kernel matrix.

The impulse response is regarded as a function that decays exponentially at a certain rate. When this finite impulse response model is obtained, the stable spline (SS) kernel, which belongs to the local stationary kernel of amplitude modulation [195], can achieve the desired effect. The element k_{ij} in SS kernel matrix, $\mathbf{K}$, is defined as:

$$k_{ij} = \frac{c}{2}\Big(exp\{-\beta(i+j) - \beta max\{i,j\}\} - \frac{1}{3}exp\{3\beta max\{i,j\}\}\Big), \quad c, \beta > 0 \tag{3.4}$$

where c and β are parameters for spline function. Parameter selection for both c and β will be discussed later.

The cost function (3.3) can be solved by applying the ReLS as:

$$\boldsymbol{\theta}_s = (\mathbf{K}\tilde{\mathbf{X}}_s^{\intercal}\tilde{\mathbf{X}}_s + \mathbf{I}_D)^{-1}\mathbf{K}\tilde{\mathbf{X}}_s^{\intercal}\tilde{V}_s \tag{3.5}$$

where $\mathbf{I}_D$ denotes an D-dimensional identity matrix.

As for a stable system, the amplitude of its impulse responses will gradually approach to zero. Therefore the order of finite impulse responses, D, can be decided at which the amplitude is close to zero.

3.2.2 Parameter Optimization

In SS kernel, the parameter c and β should be determined in advance. From the previous study [23], an experiential selection only for parameter β is given by $0.9 < \beta < 1$. In this paper, an optimization method for both parameters was proposed by maximising the following equation as:

$$\begin{aligned}(c^*, \beta^*) = \underset{c,\beta}{\arg\max}\, \frac{1}{M \cdot S} \sum_{m \in M} \sum_{s \in S} \Bigg(1 - \\ \frac{\|\tilde{v}_s^{(m)}(t) - \hat{v}_s^{(m)}(t|\boldsymbol{\theta}_s, c, \beta)\|_2^2}{\|\tilde{v}_s^{(m)}(t) - \frac{1}{T}\sum_{t \in T}\tilde{v}_s^{(m)}(t)\|_2^2} + \gamma \cdot \frac{Ent(\boldsymbol{\theta}_s|c,\beta)}{\log D}\Bigg)\end{aligned} \tag{3.6}$$

where the number of sensors and training samples are indicated by S and M respectively. $\tilde{v}_s^{(m)}(t)$ denotes the normalized responses of the s^{th} sensor in the m^{th} test, while $\hat{v}_s^{(m)}(t|\boldsymbol{\theta}_s, c, \beta)$ represents its estimation. $Ent(\boldsymbol{\theta}_s|c,\beta)$ is defined as:

$$Ent(\boldsymbol{\theta}_s|c,\beta) = -\sum_{d \in D} P_s(d) \log P_s(d) \tag{3.7}$$

where P_s is expressed as:

$$P_s(d) = \frac{|\theta_s(d|c,\beta)|}{\sum_{d\in D} |\theta_s(d|c,\beta)|}. \tag{3.8}$$

The second term of (3.6) is the normalized mean square error (NMSE), which indicates the goodness of curve fitting. A lower NMSE shows a better curve-fitting result. The third term denotes the normalized information entropy (IE) of $\boldsymbol{\theta}_s$. The finite impulse responses model will be much smoother when the normalized IE reaches a higher value. $\boldsymbol{\theta}_s$ with a smaller normalized IE will possess a higher risk of overfitting. γ is a hyper-parameter ranged from 0 to 1 controlling the trade-off between the NMSE and the normalized IE. A default selection of γ is 0.5. It means the two terms are of the same importance in parameter optimization.

3.3 Field Experiment Setup

3.3.1 Material Preparation and Instrument Settings

Six perfumes were investigated as target odours in this study, which are Chance (Chanel), Gabrielle (Chanel), Miss Dior (Dior), Red Rose (Jo Malone), Pacific Merchants (Hollister) and Princess (Vera Wang). For each perfume, the 2-mL liquid was stored in a sealed 10-mL-volume headspace crimp top vial.

The e-nose used in this experiment is a self-designed and developed device, called UTS NOS.E. This device contains an air-intake system, power system, a sensor chamber to install sensor array, a controller unit, an analogue-to-digital (A/D) converter, a Wi-Fi communication module, and a Pad-based control and configuration panel. All of the modules are housed inside a hand-held case. Five commercially available MOS-type gas sensors are adopted in perfume identification. The sensitive characteristics of the five gas sensors are shown in Table 3.1.

The ambient temperature and humidity for experiments were around 25℃ and 50% RH respectively. Non-bubbling method [129] was adopted to prevent odour vapour from extraordinary fluctuations. The vapour of target odour was injected into the e-nose chamber through the gas input port by sampling needle, and then clean air was pumped to wash odorant molecules from the sensor array [105]. Then, the A/D converts analogue signals into digital signals at a 1-Hz sampling frequency. After one measurement, the Wi-Fi communication module sends digital signals directly to PC. The schematic diagram of the experimental setup is shown in Fig. 3.1. The representative example of the measurement circuit is shown in Fig. 3.2.

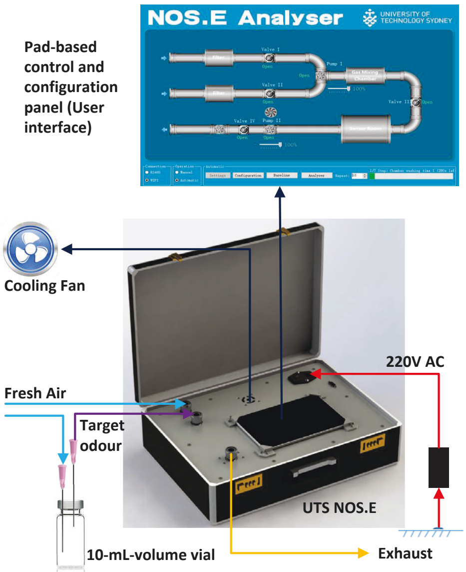


Figure 3.1: Schematic diagram of experimental setup

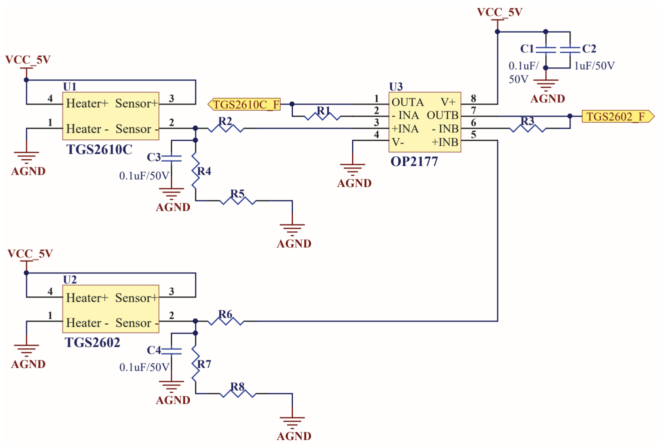


Figure 3.2: Representative example of the measurement circuit. Inside the e-nose, two pairs of sensors are connected to two dual-channel amplifiers (OP2177) respectively, and the single rest sensor is connected to one single-channel amplifier (OP1177). The representative example shows the measurement circuit using a dual-channel amplifier. The heater voltage (VCC_5V) is applied to the integrated heater to make the gas sensor work at a specific temperature that is most suitable for gas sensing. The VCC_5V DC voltage is also applied as sensor circuit voltage to allow measurement of the voltage across the load resistors (R4, R5 and R7, R8) which are connected in series with the sensor. Once the sensor detected the target gas, the measurement voltage is changed due to the variation of the sensor resistance. The measurement voltage is sent to the non-inverting amplifier (OP2177) before transmitting to the analogue to digital module of the microcontroller for further processing.

Table 3.1: Gas sensors used for perfume identification

No.	Sensor Name	Sensitivity Characteristics*
1	TGS 2610C	Iso-butane, Propane, Hydrogen, Ethanol, ect.
2	TGS 2602	VOCs, Ammonia, H2S, etc.
3	TGS 2620	Ethanol, Hydrogen, Iso-butan, CO, Methane, etc.
4	TGS 2603	Trimethylamine, Methyl mercaptan, etc.
5	TGS 2602**	VOCs, Ammonia, H2S, etc.

* The response of sensors is non-specific. They are also sensitive to other gases, which are not list in this table;
** Sensor 2 and Sensor 5 are same type of sensor, but they were installed in different places.

After setting the operating parameters of NOS.E II, the experiment is carried out automatically. One single measurement incorporates three main steps:

- Wash the chamber by exposing all gas sensors to dry clean air for 30 seconds to obtain the baseline;
- Extract the target perfume vapour into the sensor chamber for 60 seconds;
- Clean the chamber for 500 seconds.

Each perfume was tested 300 times giving a total of 1800 samples. The specific information of samples is listed in Table 3.2.

Table 3.2: Perfume Information

Label	Product Information	Sample Size
1	Chance (Chanel)	300
2	Gabrielle (Chanel)	300
3	Miss Dior (Dior)	300
4	Red Rose (Jo Malone)	300
5	Pacific Merchants (Hollister)	300
6	Princess (Vera Wang)	300

3.3.2 Experiment Process

3.3.2.1 Data Pre-Processing

Normalization (2.3) was introduced in raw responses to reduce sensor drift and compensate for the fluctuation in concentration. In a traditional e-nose system, filtering methods are implemented in the signal pre-processing, because the time-domain features, especially the derivatives-based features, extracted from a non-smooth response curve will result in lower classification accuracy in subsequent pattern recognition procedure. In this study, due to the enhanced performance of the proposed feature extraction algorithm in resistance to noise, filtering operations were not necessary for the data pre-processing procedure. The performance of resistance to noise will be discussed in Section 3.4.

3.3.2.2 Classification Based on GMM-HMM

$\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_S\}$ is a D-by-S matrix, which contains S observation sequences. This matrix represents one measurement of target odour, where $\boldsymbol{\theta}_s$ denotes the impulse response of the s^{th} sensor with a length of D. It is difficult for algorithms like support vector machine (SVM), logistic regression (LR), and other classifiers to handle this dataset, due to the high dimension of the reshaped feature vector (1-by-$D \cdot S$), which is prone to the small-sample-size problem. Therefore, a time-series classification algorithm, hidden Markov models (HMMs), is introduced in this section.

HMM is an efficient tool in dealing with time series and can cater for multi-observation scenarios. Baum *et al*. first developed the HMM at the end of the 1960s [11]. Due to its expressive structure of probabilistic networks and the availability of algorithms in calculating model parameters, HMM has been dominantly applied in time-series classification. HMM is a special case of a Bayesian network, as illustrated in Fig. 3.3.

This model can be described in terms of the joint probability distribution over hidden state sequence, $Q_{1:D}$, and observation sequence, $O_{1:D}$ as follows:

$$P(Q_{1:D}, O_{1:D}) = P(Q_1)P(O_1|Q_1)\sum_{d=2}^{D} P(Q_d|Q_{d-1})P(O_d|Q_d) \tag{3.9}$$

where $P(Q_1)$ denotes the initial state probability. $P(Q_d|Q_{d-1})$ represents transition probability and $P(O_d|Q_d)$ emission probability.

However, in a real application, observation sequences are usually continuous. In order to avoid quantization error, we can apply Gaussian Mixture density (GMMs) [92]

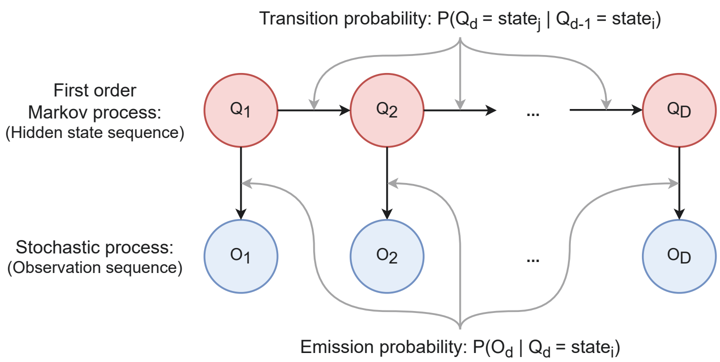


Figure 3.3: The basic structure of hidden Markov Models. HMM contains two processes: the unobserved discrete first-order Markov process, where the current state, Q_d, only depends on the previous state, Q_{d-1}, and the observed stochastic process, where the current observation, O_d, is only dependent on the current hidden state, Q_d.

to instead the emission probability in traditional discrete HMM as follows:

$$p(O_d|S_d = state_j) = \sum_{q \in Q} c_{jq} \mathcal{N}(O_d, \mu_{jq}, \Sigma_{jq}) \tag{3.10}$$

where c_{jq} denotes the mixture coefficient for the q^{th} mixture in state j, constrained by $c_{jq} \geqslant 0$ and $\sum_{q \in Q} c_{jq} = 1$. $\mathcal{N}$ represents Gaussian distribution with mean vector, μ_{jq}, and covariance matrix, Σ_{jq}, for the q^{th} mixture in state j. This type of HMM is called Gaussian mixture density hidden Markov model (GMM-HMM).

Equation 3.10 can be trained using the Baum-Welch algorithm. We use λ to represent the trained GMM-HMM. In odour identification, the finite impulse responses, $\mathbf{\Theta}$, are regarded as observation sequences. In the training procedure, the model λ_n was learned for the n^{th} target odour from the corresponding training set of finite impulse responses. In the test procedure, the observed sequences of an unknown odour were evaluated by calculating the likelihood $\log P\{\mathbf{\Theta}|\lambda_n\}$ of each trained model using the Forward-Backward algorithm. Odour identification can be achieved by selecting the model with the highest log-likelihood. The training and testing procedure of GMM-HMM-based odour identification are illustrated in Fig. 3.4(a) and Fig. 3.4(b) respectively. Further discussion on the Baum-Welch algorithm and Forward-Backward algorithm will not be discussed in detail in this paper. The basic theories of two algorithms for multiple observation sequences and GMM-HMM can be found in [99, 137].

In GMM-HMM, the number of hidden states P and Gaussian mixtures Q are hyper-parameters. However, P is difficult to determine in advance [175]. It will be discussed later. Q can be optimized in advance by employing the Bayesian information criterion (BIC) [175]. Theoretically, a GMM can approximate any signal to a certain precision

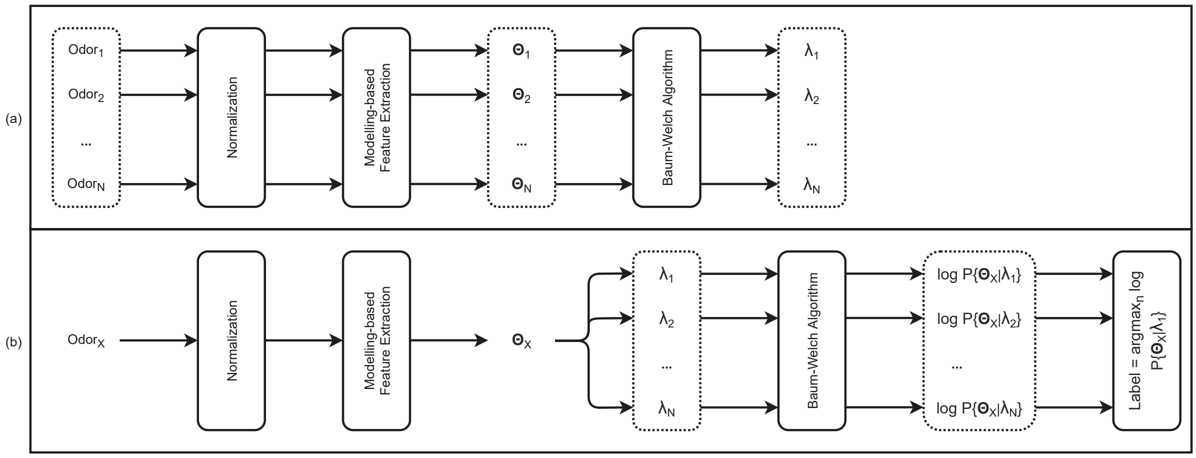


Figure 3.4: Flowchart of GMM-HMM-based odour classification. (a) Training procedure: Model parameters for both feature extraction and classification algorithms should be decided first. Next, send the normalized sensor array's responses to the feature extraction models and obtain finite impulse responses. Then, the classification model is learned based on these finite impulse responses. This procedure is repeated for every perfume class. (b) Testing procedure: sensor array's responses to an unknown odour is first normalized and then processed by the feature extraction method. Then, log-likelihood of the extracted feature sequence is evaluated by each trained classification model, λ_n. The label of the model that reaches the highest value is regarded as the label of the unknown odour.

when provided with enough mixture components. However, too many Gaussian components may lead to overfitting. The optimal Q is the one to get the lowest value of the BIC.

3.4 Results and Discussion

3.4.1 Automatic Initial Time Detection

To separate the baseline stage and the absorption stage to apply ideal step input, the detection of the initial time is important in the proposed feature extraction method. The initial time (or referred to as the time of gas-on in some literature), t_g, indicates the actual start time of the chemical sensor in response to target odours. t_g is also a critical time point that many other feature extraction methods are relying on. It is unreasonable to simply regard the initial time as the time when gas is injected to the e-nose chamber, since this time differs greatly due to the sensor's sensitivity to the concentration of odour molecules and its installation site. A traditional method assumes the time point where the response reaches $x\%$ of the peak value to be the initial time. x is ranged from 1 to 10. In order to achieve automatic odour identification, in this research, an initial time detection method can be realized by calculating the kernel-based estimated model, $\hat{v}_s(t|\theta_s)$, as expressed in (3.1). If the detected time, t_g, is before the real sensor initial time, the kernel-based estimation will contain negative values, as expressed in Fig. 3.5.

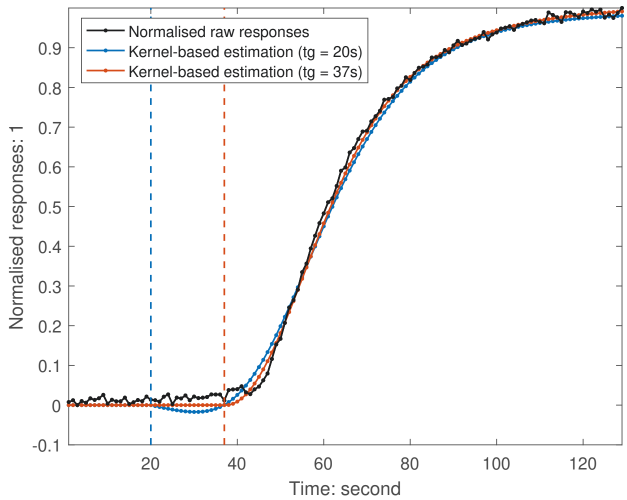


Figure 3.5: Representative example of sensor initial time detection. The real initial time is 37s for this sensor. When the selected initial time is set to 20s, its kernel-based estimation generates negative values on the interval $(20, 37)$.

Theoretically, the negative response value indicates the sensor starts response before any stimulation, which is unreasonable. Therefore, we can increase the t_g until no negative value appears in the kernel-based estimation. The pseudo-code of this process is shown in Algorithm 3.

3.4.2 Result of Parameter Selection

In SS kernel function, parameters β and c can be optimized according to (3.6). The goodness of fit is decided by both of these parameters, while c provides an additional decision to the complexity of the model. To make the parameter optimization process visual, the result is visualized in Fig. 3.6.

As mentioned in the previous section, IE has adopted to constraint the model complexity and reduce the risk of overfitting. To prove the effectiveness of IE, a representative comparison of finite impulse responses with different c is illustrated in Fig. 3.7. The

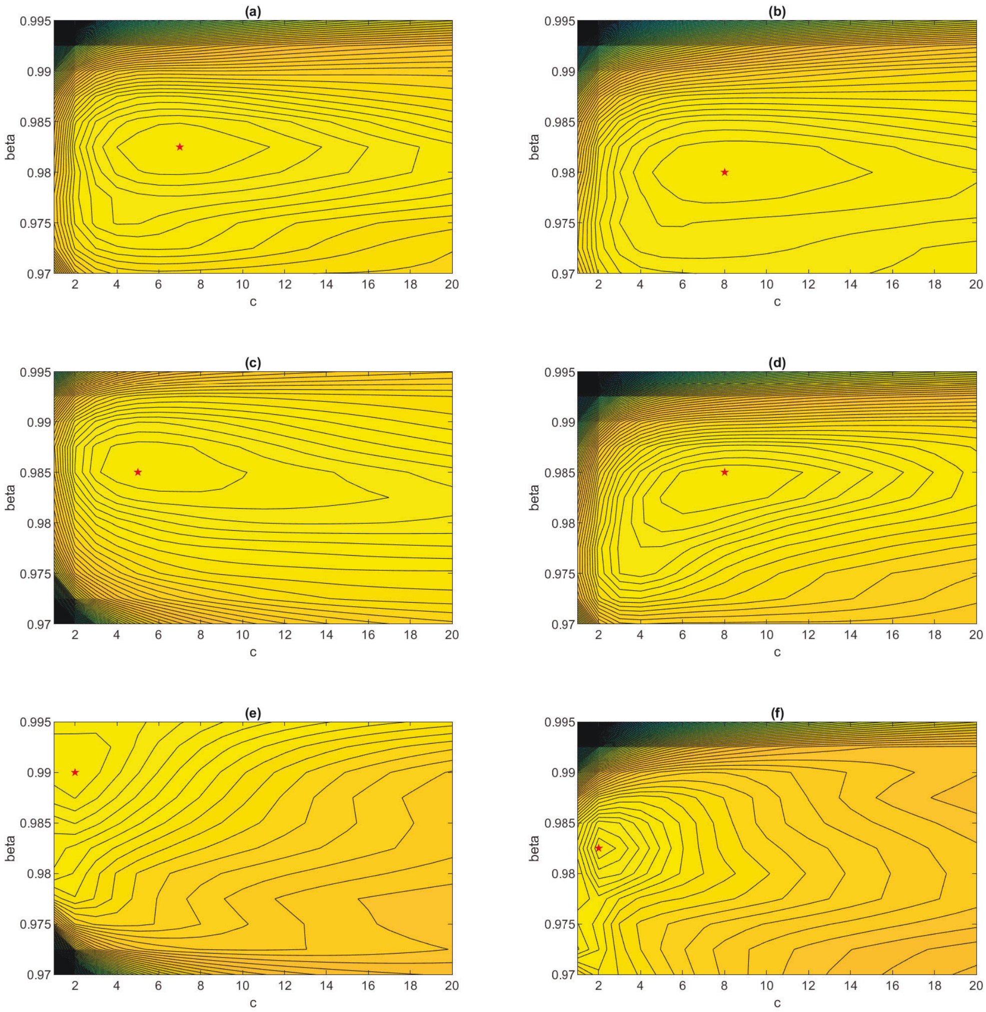


Figure 3.6: Illustration of SS kernel parameter optimization. (a) Contour plot of parameter optimization for Perfume 1. $\beta = 0.9825$, $c = 7$. (b) Contour plot of parameter optimization for Perfume 2. $\beta = 0.9800$, $c = 8$. (c) Contour plot of parameter optimization for Perfume 3. $\beta = 0.9850$, $c = 5$. (d) Contour plot of parameter optimization for Perfume 4. $\beta = 0.9850$, $c = 8$. (e) Contour plot of parameter optimization for Perfume 5. $\beta = 0.9900$, $c = 2$. (f) Contour plot of parameter optimization for Perfume 6. $\beta = 0.9825$, $c = 2$.

Algorithm 3: Initial Time Detection

Input: Signal $\tilde{v}_s$.
Output: Initial time t_g.

Initialization: $t_g \leftarrow 1$
while 1 **do**
  $\hat{v}_s \leftarrow$ Kernel Based Estimation via $\tilde{v}_s$ and t_g
  if Each element of $\hat{v}_s$ is not negative **then**
    return t_g
  else
    $t_g \leftarrow t_g + 1$
  end
end

finite impulse responses with a small c obtain a higher entropy value and are much smoother than that with a large c.

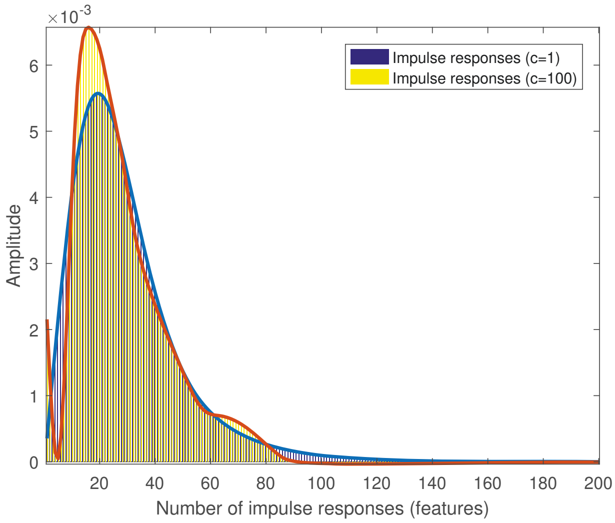


Figure 3.7: Representative example of the finite impulse responses to one sensor. The number of impulse responses, D, is selected as 200, at which the amplitude is much close to zero. When c is set to 1, the IE equals 4.1932. The IE decreases to 4.1096, when c increases to 100. However, the NMSE reduces from 0.12% to 0.06%.

In GMM-HMM, the number of Gaussian mixtures, Q, can be decided by Eq. (12) before model training. The optimization result is illustrated in Fig. 3.8.

As aforementioned, the number of hidden states, P, in GMM-HMM is difficult to determine in advance. However, the selection of P will finally affect the classification accuracy. If P is less than the optimal value, the low model complexity restricts its

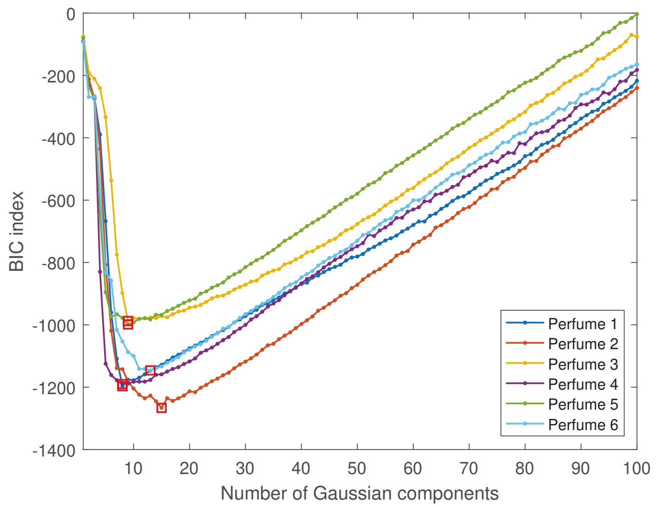


Figure 3.8: BIC for Gaussian mixtures. The results are Perfume 1: $Q = 8$, Perfume 2: $Q = 15$, Perfume 3: $Q = 9$, Perfume 4: $Q = 8$, Perfume 5: $Q = 9$, Perfume 6: $Q = 13$.

application in time-series modelling. But, when P is beyond the optimal one, the trained classifier will be overfitting.

An alternative approach to select this parameter is calculating the corresponding classification accuracy. In the multiclass classification, the number of positive samples is less than that of negative ones. Positive samples are expected to be identified as much as possible since a positive sample is more important than the negative one. Therefore, the sensitivity is applied instead of the accuracy to measure the performance of the classifier in the multiclass classification. The sensitivity is defined as M_{TP}/M_P, where M_{TP} and M_P represents the number of correctly classified positive samples and the total positive samples respectively. The sensitivity is obtained by applying five-fold cross-validation (5-CV). In the 5-CV, the 300 samples for each perfume class is randomly re-ranked and separated into five subsets. Each subset contains 60 samples. In one validation procedure, one of the five subsets is used for testing, while the rest four subsets are applied for training. The validation procedure is then repeated for five times. The 5-CV ensures that each sample for one class is used for model training and model testing only once. Finally, the five results are combined to yield one sensitivity. The classification results of six perfume classes are shown in Fig. 3.9. When the number of hidden states, P, was set to three, the classification sensitivity reaches the highest median with the lowest interquartile range in each perfume testing result. Therefore, the optimal number of hidden states is selected as three in this experiment.

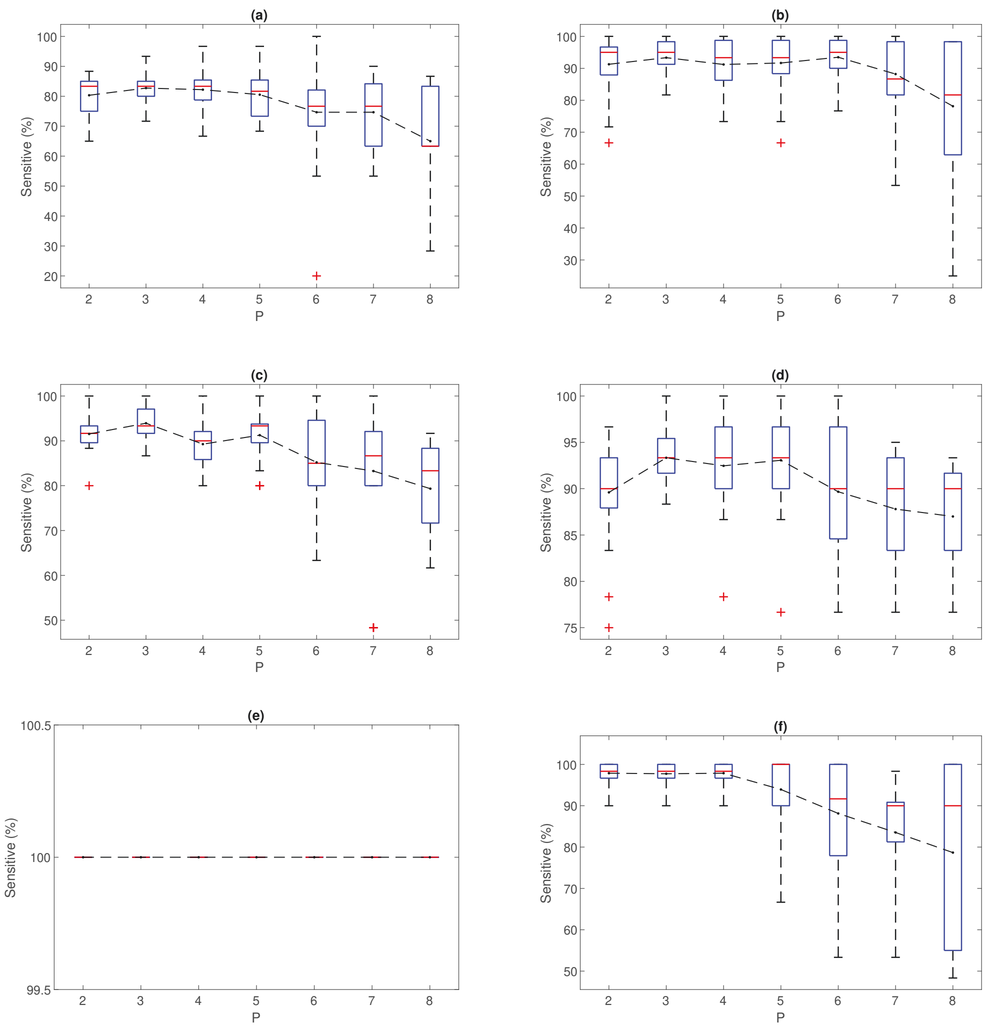


Figure 3.9: Boxplot of classification sensitivity

3.4.3 Comparison of Resistance to Noise

In this section, the capacity of the proposed feature extraction method in resistance to noise is discussed in comparing with traditional features. Gaussian noise was added to the raw responses. In the pre-processing procedure of traditional feature extraction method, as the raw output of the sensor is nonnegative, we adopted normalization to reduce the effect of sensor drift and concentration fluctuation, and then, applied different filtering and fitting techniques to smooth the response curve, including mean filter (MF) [169], WTDF (as shown in Algorithm 1) and polynomial regression (as shown in (2.5a)). For the proposed non-parametric modelling-based method, we only applied normalization in the pre-processing procedure. The traditional features selected in this experiment are mainly transient-state features and time intervals since the steady-state features become less discriminant after normalization. The detailed information of these features are (see Table 2.1): f_2, f_3, f_4, f_6, f_7, f_9, f_{10}, and the integral from gas-on to peak. The same classifier, GMM-HMM, was adopted to identify perfume classes. Also, the parameters of GMM-HMM classifier for each considered feature set is optimized using the method mentioned above. The classification results are evaluated by averaged sensitivity (A. Sen.) and averaged specificity (A. Spe.) of the six perfume classes respectively. The sensitivity quantifies the performance of a classifier in correctly identifying positive samples, while the specificity quantifies the ability of correctly classifying negative samples. The sensitivity (Sen.) is defined as mentioned in the previous section, while the specificity (Spe.) is calculated as M_{TN}/M_N. M_{TN} indicates the number of correctly classified negative samples and M_N the total number of negative samples. All classification results were calculated by applying the 5-CV. The experiment result is shown in Table 3.3.

Due to the fact that the derivative-based features are extremely sensitive to the noises, the raw traditional time-domain features perform worst in all considered methods. After applying filters, the sensitivities and specificities of traditional features become acceptable. The result of applying WTDF to traditional features is much better than that of applying MF. Since WTDF is capable of reducing Gaussian noises in an entire frequency band, while MF can only be regarded as a low-pass filter. Polynomial curve fitting is a typical smoothing method, which is still sensitive to noise. Since it performs well in a high signal-to-noise Ratio (SNR) condition but performs worse in a low SNR condition. The proposed method achieves the highest sensitivity and specificity in all considered SNR conditions even without applying any pre-processing technique.

The reason is that the kernel-based non-parametric modelling approach introduced

Table 3.3: Comparison result of resistance to noise

Method	Classification result (%)							
	SNR=50dB		SNR=40dB		SNR=30dB		SNR=20dB	
	A. Sen.	A. Spe.	A. Sen.	A. Spe.	A. Sen.	A. Spe.	A. Sen.	A. Spe.
Proposed method	**91.78**	**98.36**	**91.28**	**98.26**	**89.39**	**97.88**	**78.39**	**95.68**
Raw signal + GMM-HMM	72.67	94.53	70.72	94.14	67.72	93.54	63.28	92.66
MF* + Traditional features + GMM-HMM	78.50	95.70	75.72	95.14	71.44	94.29	70.17	94.03
WTDF** + Traditional features + GMM-HMM	91.06	98.21	90.06	98.01	83.22	96.64	74.28	94.63
Polynomial regression*** + Traditional features + GMM-HMM	90.22	98.04	90.00	98.00	80.83	96.47	70.28	94.29

* The parameters of MF are set as: sliding window length = 5, lag factor = 1;
** WTDF + Traditional features + GMM-HMM. The parameters of WTDF are selected as: wavelet factor = 'sym', level = 6;
*** The degree of polynomial curve fitting is set to 3, according to [194].

the so-called Tikhonov regularization, in which the evaluation functionals of the regularization term are bounded. This penalizes the high-frequency components of the identified finite impulse responses. In addition, the SS kernel also provides a strong condition of continuity between two impulses, which further smooths the result. Furthermore, the well-selected parameters of kernel function also prevent the finite impulse responses from overfitting. Therefore, the proposed method has an excellent ability in resistance to noise.

3.4.4 Comparison of Different Feature Extraction Methods

Typical feature extraction methods from different categories as mentioned in the introduction are analysed and compared in this section, including traditional hand-crafted features (in Section 2.3.1), polynomial curve-fitting features (in Section 2.3.2), and SAE neural network learned features (in Section 2.3.3). Meanwhile, directly applying raw responses to the time-series classifier was also compared. Before extracting such features, WTDF was applied as a pre-processing tool in above four compared algorithms due to its effectiveness in denoising as verified in the previous section. The classifier considered in this experiment were optimized GMM-HMMs. The parameter optimization method is mentioned in the previous section. All considered methods were evaluated by both sensitivity and specificity, which are obtained according to the 5-CV. The result is listed in Table 3.4.

The proposed method outperforms other feature extraction methods and achieves 93.56% averaged sensitivity, and 98.71% averaged specificity respectively. The result of directly applying raw responses in GMM-HMM is not acceptable since the length of time-series is not always the same. The proposed method solves this problem by transforming the raw responses from the time domain to feature domain with the same dimension. The classification result of traditional features is close to that of the proposed method. However, it needs a pre-processing technique to ensure the quality of the extracted features. Also, the design and selection of these features require additional expert knowledge. Parameters of the polynomial function is a kind of curve-fitting-based feature extraction method. The degree of the polynomial was three, which means that we can only extract four features (including the parameter of the constant term) from each sensor. A higher degree polynomial will lead to overfitting since the high-degree curve corresponds too closely to the original signal. It limits the extraction of the discriminant information contained in these structure parameters. In this situation, the proposed method can provide more informative structure parameters when comparing with the

Table 3.4: Comparison result of resistance to noise

Method	Classification result (%)													
	Perfume 1		Perfume 2		Perfume 3		Perfume 4		Perfume 5		Perfume 6		Averaged	
	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.	Sen.	Spe.
Proposed method	84.33	95.60	**95.00**	**98.80**	**90.67**	**98.93**	91.33	**98.93**	**100**	**100**	**100**	**100**	**93.56**	**98.71**
Raw + WTDF + GMM-HMM	50.67	85.00	54.33	97.40	85.67	97.80	83.33	94.13	**100**	**100**	97.67	**100**	78.61	95.72
WTDF + Traditional features + GMM-HMM	84.67	**97.00**	82.67	97.47	89.33	98.60	**94.33**	96.00	**100**	**100**	95.33	99.60	91.06	98.11
WTDF + Polynomial features + GMM-HMM	73.67	90.53	80.67	98.67	85.67	95.47	87.00	97.07	98.00	**100**	82.33	99.73	84.56	96.91
WTDF + SAE + GMM-HMM	**85.00**	87.00	44.33	95.00	86.67	93.00	56.67	97.20	**100**	**100**	33.67	89.07	67.72	93.54

curve-fitting features. SAE is a representative algorithm for feature learning method. The SAE compresses sensor array's responses from the input layer into a comprehensive expression from the hidden layer and then recovers the comprehensive expression to signals that closely matches the original input. This comprehensive outputs from the hidden layer are regarded as features for target odours. In this experiment, the number of hidden neurons were selected based on a genetic algorithm (GA). However, the optimized SAE performs worst among all considered methods. The reason is that the small sample size limits the training of an accuracy model.

The proposed feature extraction method is able to extract more discriminant and stable features from the raw responses than other considered algorithms, which makes it be an ideal feature extraction method for MOS-type gas sensor array-based e-noses.

3.4.5 Comparison of Different Multiclass odour Identification Methods

In this section, the proposed method is compared with other different combinations of feature extraction and classification algorithms, including proposed features combining with Discrete HMM (DHMM) and K-Nearest Neighbours (KNN), as well as traditional features combining with SVMs under different multiclass classification strategies and Softmax regression. For the combinations with traditional features, additional WTCF-based pre-processing procedure is added. The comparison results were assessed by sensitivity, specificity and the area under the ROC curve (AUC). The confusion matrices and assessment results for each considered method are detailed in Table 3.5-3.10 respectively. The ROC plot is illustrated in Fig. 3.10.

Table 3.5: Classification result of proposed method

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**253**	15	16	16	0	0	84.33		95.60	
2	15	**285**	0	0	0	0	95.00		98.80	
3	25	3	**272**	0	0	0	90.67	93.56	98.93	98.71
4	26	0	0	**274**	0	0	91.33	±6.06	100	±1.62
5	0	0	0	0	**300**	0	100		100	
6	0	0	0	0	0	**300**	100		100	

Table 3.6: Classification result of modelling-based features with DHMM

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**279**	12	1	2	0	6	93.00		86.20	
2	44	**245**	2	0	0	9	81.67		96.47	
3	54	15	**212**	9	0	10	70.67	83.06	99.47	96.61
4	56	0	1	**242**	1	0	80.67	±8.97	99.27	±5.25
5	15	2	1	0	**282**	0	94.00		99.93	
6	38	24	3	0	0	**235**	78.33		98.33	

Table 3.7: Classification result of modelling-based features with KNN

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**213**	16	25	46	0	0	71.00		96.73	
2	27	**273**	0	0	0	0	91.00		97.07	
3	21	21	**252**	6	0	0	84.00	90.28	98.33	98.06
4	0	0	0	**300**	0	0	100	±11.15	96.47	±1.54
5	0	0	0	0	**296**	4	98.67		100	
6	1	7	0	1	0	**291**	97.00		99.73	

Table 3.8: Classification result of traditional features with SVM (1 v 1)

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**274**	0	0	8	18	0	91.33		99.07	
2	8	**272**	0	4	14	0	90.67		99.93	
3	0	0	**287**	0	13	0	95.67	93.06	99.87	98.52
4	4	1	0	**265**	30	0	88.33	±4.16	99.00	±2.54
5	0	0	0	0	**300**	0	100		93.40	
6	0	0	0	1	22	**277**	93.33		99.87	

Table 3.9: Classification result of traditional features with SVM (1 v Rest)

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**239**	3	0	18	38	2	79.67		99.93	
2	0	**276**	2	0	21	1	92.00		99.73	
3	0	0	**277**	0	23	1	92.33	89.61	99.87	97.97
4	2	2	0	**249**	46	1	83.00	±7.27	98.87	±4.08
5	0	0	0	0	**300**	0	100		89.67	
6	0	0	0	0	28	**272**	90.67		99.73	

Table 3.10: Classification result of traditional features with Softmax regression

Label	Classified as						Sen. (%)	A. Sen. (%)	Spe. (%)	A. Spe. (%)
	1	2	3	4	5	6				
1	**215**	40	5	38	0	2	71.67		96.40	
2	30	**251**	17	0	0	2	83.67		95.73	
3	22	0	**276**	0	0	2	92.00	88.78	98.13	97.76
4	2	22	3	**269**	0	4	89.67	±10.03	96.93	±1.69
5	0	0	0	0	**300**	0	100		100	
6	0	2	3	8	0	**287**	95.67		99.33	

The proposed method achieves the highest averaged sensitivity and averaged specificity among all considered methods. It performs best in identifying Perfume 2, 3, 5 and 6 according to the AUC. Particularly, the proposed method acts as a perfect identification tool for perfume 5 and 6, since its AUC equals one in such classes. The number of hidden states of DHMM was optimized to three according to the cross-validation. The specification of observation symbols was set to 100. DHMM reaches the worst result in dealing with the proposed features, due to the quantization error in the procedure of finite impulse responses sequences discretization. The quantization error will decrease as the specification of observation symbols increase; however, the computational overhead will also increase. The parameter, k, for KNN is selected as 120 according to the BIC index. The finite impulse responses sequences ensure the requirement of time alignment in using KNN-based time-series classification, and therefore, KNN achieves an acceptable result. This method performs best in identifying Perfume 4 since its AUC ranks first in such class. Nevertheless, this method is still inferior to the proposed method in comparing both averaged sensitivity and averaged specificity of other perfume classes. For the combination of traditional features, the method of SVM under 1-v-1 strategy achieves a good result, and close to the performance of the proposed method. The parameters for all SVMs are optimized according to GA. However, SVM requires 15 binary classifiers to address this perfumes identification task, while GMM-HMMs only needs six classifiers to complete the same work. Moreover, if a new class of perfume is added to the sample library, another six SVMs should be trained. In this situation, GMM-HMM has a unique advantage that any new odour sample can be simply trained and directly added to the existing pattern library. SVM under 1-v-Rest strategy greatly reduces the number of trained classifiers, since only six binary classifiers are trained for each considered class, with the samples of that class as positives and the rest as negatives. However, the negative samples in the training set are five times the positives, which introduces a new problem of data imbalance. Each binary SVM tends to correctly classify the negative samples to reduce the error of the cost function. It is reflected by its reduced sensitivity. The traditional features with SVM-based methods performs best in identifying Perfume 1 according to the AUC. Softmax regression can solve the multiclass problem by training one classifier. However, this method is a linear classifier, which is not an ideal choice in dealing with non-linearly separable problems. Therefore, the proposed method can be applied in e-nose to make an accurate and robust decision for odour identification.

3.5 Conclusion

In this chapter, a new odour identification method is developed based on kernel-based model estimation and GMM-HMM. An effective parameter selection method is also provided to optimize the proposed algorithms. The extracted features are finite impulse responses containing discriminant information of different odours, which can be identified by GMM-HMM. When the kernel-based model estimation method is used to extract the features of the raw sensor array's responses, the classification sensitivity and specificity is higher than those of other considered features, including traditional hand-crafted features, polynomial features and SAE features. The experimental result demonstrates that the proposed feature extraction method is able to provide a wealth of discriminant structure information of the original response curve. Moreover, it is possible to skip the smoothing procedure before feature extraction, since this method has a better performance in resistance to noise than traditional time-domain feature extraction methods combining with additional filtering or smoothing tools. The classification result of proposed features using GMM-HMM also achieves the highest accuracy compared

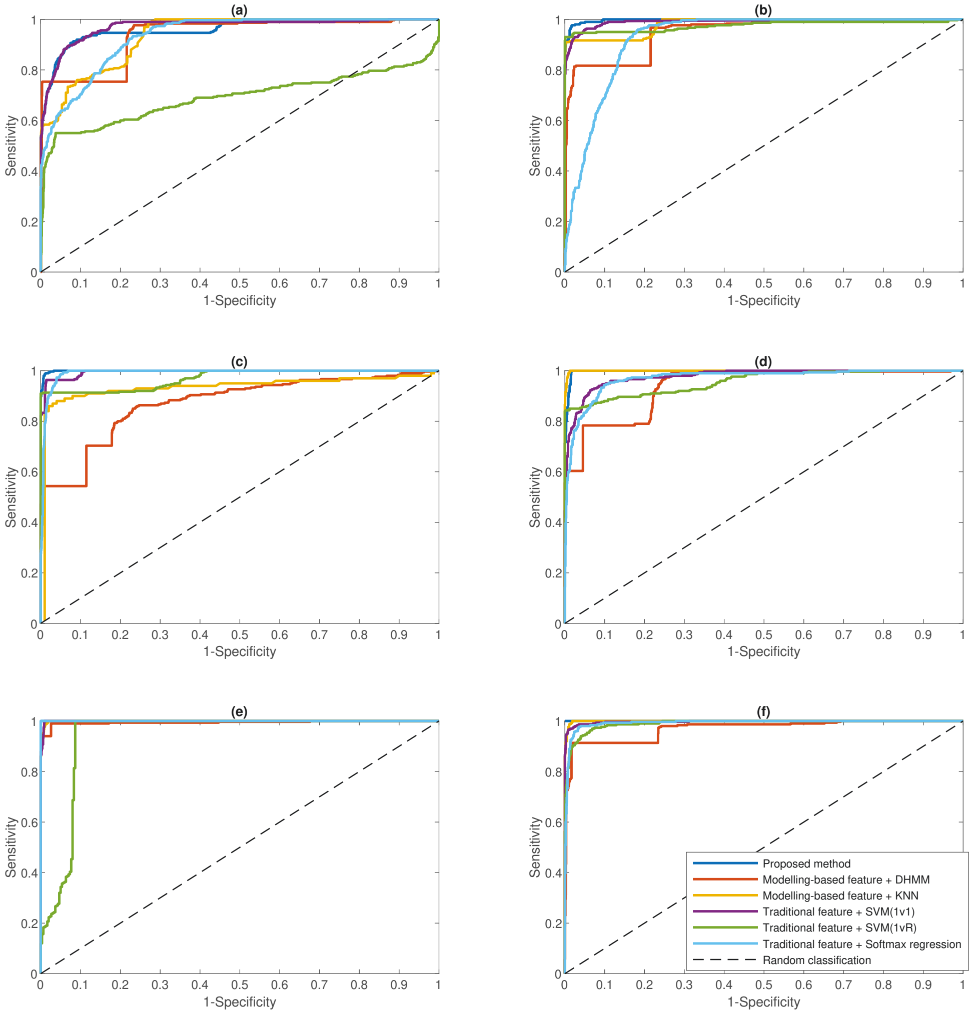


Figure 3.10: ROC curves for considered methods. (a) Perfume 1: AUC4 > AUC1 > AUC3 > AUC2 > AUC6 > AUC5; (b) Perfume 2: AUC1 > AUC4 > AUC3 > AUC5 > AUC2 > AUC6; (c) Perfume 3: AUC1 > AUC4 > AUC6 > AUC5 > AUC3 > AUC2; (d) Perfume 4: AUC3 > AUC1 > AUC4 > AUC6 > AUC5 > AUC2; (e) Perfume 5: AUC1 = AUC6 > AUC4 > AUC3 > AUC2 > AUC5; (f) Perfume 6: AUC1 > AUC3 > AUC4 > AUC6 > AUC5 > AUC2.

with DHMM, KNN, SVM and Softmax regression, and is less computationally expensive in dealing with multiclass classification task. In addition, the automatic initial time detection method makes automatic gas identification possible. The results of this study demonstrate that the proposed modelling-based feature extraction method combining with GMM-HMM based classification is more effective and applicable in identifying different odours by e-nose.

CHAPTER 4

Feature Extraction Based on Multiscale Wavelet Kernel Regularization

4.1 Introduction

Recall that the modelling-based feature extraction algorithm highly relies on the selection of kernel function. In the previous chapter, we set the kernel as the stable spline (SS) kernel. The performance of SS kernel-based modelling depends on an accurate setting of the initial time. If the given initial time is incorrect, no matter how to optimize the hyper-parameters of the kernel function, a stable feature representation cannot be obtained. Accordingly, we proposed a initial time detection algorithm (as shown in Algorithm 3) to solve this problem. In this chapter, we give another solution, which is applying a more powerful kernel – multiscale wavelet kernel.

The multiscale wavelet kernel derived from the concept of multi-resolution analysis has been proposed and used for many years for its rich and flexible kernel expansions in representing highly complex functions [111]. The multiscale wavelet kernel usually requires closed-form orthonormal wavelets since a closed-form expression is easy to compute. The simplest orthonormal wavelets with closed-form expressions are the Shannon and the Haar [84]. Recently, two new wavelets, alleged as the type I and the type II raised-cosine wavelets [177], have been designed by adding a constraint on the raised-cosine spectrum, which enlarges the family of closed-form orthonormal wavelets. Subsequently, multiscale kernels based on such wavelets are also developed [111]. Despite the computation cost, the multiscale kernel can also be constructed from the orthonormal wavelet without closed-form expressions, such as the Daubechies wavelet with $p \geqslant 2$ vanishing moment [109]. In this case, the kernel function can be calculated using the approximated mother wavelet. In real applications, one can approximate the wavelet function at the finest scale, then the function at a coarser scale can be simply obtained by down-sampling.

From the view of usage, most of the research applies the multiscale wavelet kernels in the feature domain to deal with nonlinear problems, i.e. $\kappa\langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)}\rangle$, where κ denotes kernel function, and $\mathbf{x}^{(i)}$ as well as $\mathbf{x}^{(j)}$ represent feature representations (vectors) of different observations (samples). In this usage, multiscale wavelet kernels have been reported in hybrid with support vector machines (SVMs), extreme learning machines (ELMs), principle component analysis (PCA), etc. to tackle nonlinear classification problems [82]. Multiscale wavelet kernels are also used in support vector regression (SVR) to deal with the nonlinear regression [51] or nonlinear system identification problems [111].

In this study, the multiscale wavelet kernel is introduced as a Tikhonov regularization term to equip the modelling-based feature extraction algorithm with an ability in resistance to random noises.

The primary contribution of this chapter are summarized as follows:

- A multiscale wavelet kernel regularization-based modelling method is applied to extract features, which avoids the difficulty of pre-setting the initial time encountered in the previous parametric modelling- or traditional curve fitting-based methods.

- We give a numerical experiment to show that the usage of multiscale wavelet kernel regularization outperforms single-scale kernels in resistance to noise.

- We also compare the proposed feature representations with other widely used features in a field experiment displayed on the CeBIT Australia 2019 exhibition. The training and testing samples are obtained from different places. Accordingly, drift issue are considered in this experiment.

The rest of the chapter is structured as follows: Section 4.2 is devoted to describing the whole methodology of multiscale wavelet kernel regularization-based feature extraction. To verify the proposed method, a numerical experiment is conducted and discussed in Section 4.3. Then, a field experiment is conducted and discussed in Section 4.4. The task of the field experiment is to identify four different target whiskies. Finally, Section 4.5 gives a short conclusion of this chapter.

4.2 Methodology

Modelling-based feature extraction methods regard the e-nose as a dynamic system, where a step signal is assumed to be the ideal system input denoting the receiving of target odours (as illustrated in Fig. 4.1), and the sensor's reading to be the output. Parameters of this dynamic system can be regarded as a representation of odour since other interfering factors of the system (e.g. relative stable environment, fixed operation parameters, and regulated concentration) have been well regulated in the experiment [106].

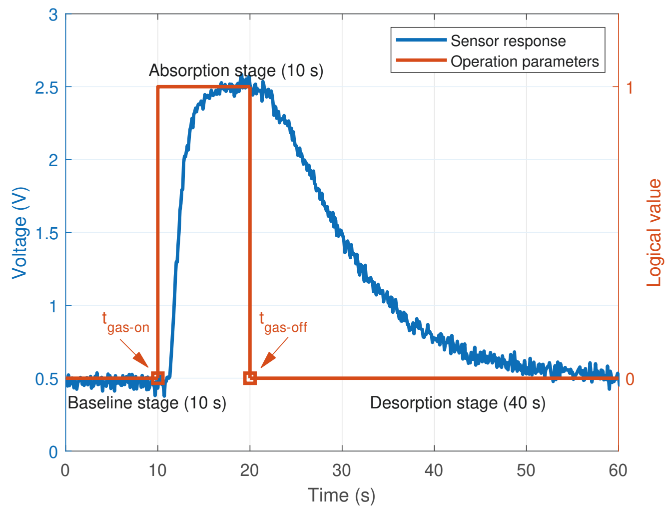


Figure 4.1: Representative example of sensor's response. The example displayed here is TGS2600's response to one sample of Ardbeg whisky. The operating parameters representing the presence or absence of the target odour divide the sensor response into a baseline stage, an absorption stage, and a desorption stage. Feature extraction is usually performed at the first two stages. The curve of operation parameters is regarded as the ideal input of the e-nose system from the view of modelling-based feature extraction.

It is difficult to pre-set an exact order for the dynamic system before parameter estimation. To avoid this problem, we apply a non-parametric method, i.e. the FIR model. However, it should be noted that the estimation of the FIR model is an ill-posed problem caused by the simple step input and the existence of large noise in the output. Hence, Tikhonov regularization technique is adopted to solve this problem [120]. Moreover, a multiscale wavelet kernel is also applied because of the advantages of wavelet kernel inherited from the wavelet function in approximating arbitrary signals. Unlike the kernel-based learning, which applies the kernel in observations to solve the nonlinear problem, the proposed method applies the kernel in the regularization term to give a bound in the estimation of the FIR model.

4.2.1 Multiscale Wavelet Kernel

Recent studies pointed out that multiscale kernel is more flexible and capable over conventional single-scale kernel, i.e. $\kappa\langle t,\tau\rangle = \psi(t-\tau)$, in modelling complex functions [111]. The multiscale wavelet kernel is constructed in the notion of multi-resolution analysis (MRA).

The MRA of a functional space $L^2(\mathbb{R})$ consists of a sequence of nested subspaces: $\{0\} \subset \dots V_{m+1} \subset V_m \subset V_{m-1} \subset \cdots \subset L^2(\mathbb{R})$, where V_m, the span of orthonormal bases $\{\phi_{m,n}(t) = (\mathrm{D^m T^n}\phi)(t)\}_{n\in\mathbb{Z}}$, is called the approximation space at scale m [84]. $\phi(t)$ represents the scaling function. Dilation operator D and time translation operator T are defined by $(\mathrm{D^m} f)(t) = 2^{-m/2} f(2^{-m}t)$ and $(\mathrm{T^n} f)(t) = f(t-n)$, respectively. For each scale m, a detail space W_m, the span of orthonormal wavelet functions $\{\psi_{m,n}(t) = (\mathrm{D^m T^n}\psi)(t) = 2^{-m/2}\psi(2^{-m}t-n)\}_{n\in\mathbb{Z}}$, is defined as the orthogonal complement of V_m in V_{m-1}, i.e. $V_{m-1} = V_m \oplus W_m$, where $\oplus$ represents the orthogonal direct sum. Hence, $\{W_m\}_{m\in\mathbb{Z}}$ are pairwise orthogonal. Accordingly, the functional space can be orthogonally decomposed as $L^2(\mathbb{R}) = \overline{\oplus_{m\in\mathbb{Z}} W_m}$ [84].

The orthogonal projection operator Q_m to detail space W_m can be precisely defined as:

$$\mathrm{Q}_m = \sum_n \psi_{m,n}\psi^*_{m,n} \tag{4.1}$$

where $*$ is the Hermitian adjoint.

Hence, the detail of a function $f(t) \in L^2(\mathbb{R})$ in the space W_m is represented as:

$$\begin{aligned}(\mathrm{Q}_m f)(t) &= \sum_n \psi_{m,n}(t)\psi^*_{m,n} f \\ &= \sum_n \psi_{m,n}(t) \int \psi_{m,n}(\tau) f(\tau) d\tau \\ &= \int \kappa_m\langle t,\tau\rangle f(\tau) d\tau \end{aligned} \tag{4.2}$$

where $\kappa_m\langle t,\tau\rangle = \sum_n \psi_{m,n}(t)\psi_{m,n}(\tau)$ is defined as the kernel function at scale m.

Since $\{W_m\}_{m\in\mathbb{Z}}$ are pairwise orthogonal, $\sum_m \mathrm{Q}_m$ is also an orthogonal projection operator [84, 111]. Accordingly, the multiscale wavelet kernel can be constructed as:

$$\kappa\langle t,\tau\rangle = \sum_m \kappa_m\langle t,\tau\rangle = \sum_m\sum_n \psi_{m,n}(t)\psi_{m,n}(\tau). \tag{4.3}$$

Note that the construction of the multiscale wavelet kernel is based on the orthogonal projection operator, implying that the basis generated by the mother wavelet ψ should be orthonormal [84, 111]. The existing closed form orthonormal wavelets include Haar (Haa), Shannon (Sha), and two types of raised-cosine wavelets (I and II) [111]. Their mother wavelets are defined, respectively, as follows:

$$\psi^{(\mathrm{Haa})}(t) = \begin{cases} 1, & 0 \leqslant t < 0.5 \\ -1, & 0.5 \leqslant t < 1 \\ 0, & otherwise \end{cases} \tag{4.4}$$

$$\psi^{(\mathrm{sha})}(t) = 2\,\mathrm{sinc}(2t) - \mathrm{sinc}(t) \tag{4.5}$$

$$\begin{aligned} \psi^{(\mathrm{I})}(t+\frac{1}{2}) = & \frac{\sin\pi(1+\alpha)t - 4\alpha t\cos\pi(1-\alpha)t}{\pi t[(4\alpha t)^2-1]} \\ & - \frac{\sin 2\pi(1-\alpha)t - 8\alpha t\cos 2\pi(1+\alpha)t}{\pi t[(8\alpha t)^2-1]} \end{aligned} \tag{4.6}$$

$$\psi^{(\mathrm{II})}(t+\frac{1}{2}) = \frac{2\cos 2\pi\alpha t}{1+4\alpha t}\mathrm{sinc}(2\pi t) - \frac{\cos 2\pi\alpha t}{1+2\alpha t}\mathrm{sinc}(\pi t) \tag{4.7}$$

where the hyper-parameter α is called the roll-off factor ranged from 0 to 1/3 [177].

However, most of orthonormal wavelets cannot be expressed in closed form, e.g. the Daubechies wavelet with $p \geqslant 2$ vanishing moment. In this case, the kernel matrix $\mathbf{K}$ can be calculated using the approximated mother wavelet. In real applications, we can approximate the wavelet function at the finest scale, then the one at a coarser scale can be simply obtained by down-sampling.

4.2.2 Feature Extraction Using Kernel-Based Modelling

Let $\mathbf{y} = [y_1, y_2, \cdots, y_N]^\intercal \in \mathbb{R}^N$ be the standardized discrete sensor's response. $\mathbf{x} = [x_1, x_2, \cdots, x_N]^\intercal \in \mathbb{R}^N$ is the assumed ideal step input representing the receiving of the stimulus from the target odour. The sampling interval is assumed to be one for convenience. The model represented by M impulse responses $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_M]^\intercal \in \mathbb{R}^M$ is in the form of:

$$y(n) = (\theta \circledast x)(n) + \epsilon(n) = \sum_{m=1}^{M} \theta(m)x(n-m) + \epsilon(n) \tag{4.8}$$

where $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_N]^\intercal \in \mathbb{R}^N$ represents the disturbance of the system, and $\circledast$ denotes discrete convolution.

The model (4.8) can be expressed compactly as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{4.9}$$

where, the Toeplitz matrix $\mathbf{X}$ is constructed from the step input $\mathbf{x}$ as:

$$\begin{aligned}\mathbf{X} &= \left[(\mathrm{T}^0 x)(n), (\mathrm{T}^1 x)(n), \cdots, (\mathrm{T}^{\text{M-1}} x)(n)\right] \\ &= \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-M+1} \end{bmatrix} \in \mathbb{R}^{N\times M} \end{aligned} \tag{4.10}$$

with T being the time translation operator with zero-padding as defined in the previous section.

However, deconvolution from (4.9) is an ill-posed problem, since the step input provides little information and the observation contains noise. Hence, a regularization term is added in the cost function (4.11a) providing an additional constraint. Since the kernel function is positive-definite, the features of a given sensor's response, i.e. the impulse response $\boldsymbol{\theta}$, can be estimated using regularized least square as (4.11b):

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 + \sigma^2 \|\boldsymbol{\theta}\|_{\mathbf{K}^{-1}}^2 \tag{4.11a}$$

$$= \arg\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\intercal (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \sigma^2 \boldsymbol{\theta}^\intercal \mathbf{K}^{-1} \boldsymbol{\theta}$$

$$= \mathbf{K}\mathbf{X}^\intercal (\mathbf{X}\mathbf{K}\mathbf{X}^\intercal + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \tag{4.11b}$$

where σ^2 is a trade-off parameter denoting the variance of the system disturbance $\boldsymbol{\epsilon}$, $\mathbf{I} \in \mathbb{R}^{N\times N}$ represents the identity matrix, and $\mathbf{K} = [\kappa\langle i,j\rangle]_{i,j\in M} \in \mathbb{R}^{M\times M}$ is the kernel matrix.

In Bayesian statistics, (4.11a) is also called Maximum a posteriori (MAP) estimation of $\boldsymbol{\theta}$. The regularization term specifies an observation sample-independent prior distribution over $\boldsymbol{\theta}$. Unlike kernel-based learning applying the kernel function to observation samples, the i and j in the kernel matrix $\mathbf{K}$ represent position indexes of the i^{th} and j^{th} impulse response, respectively. Hence, the kernel matrix $\mathbf{K}$ can be calculated without considering the observations. The multiscale wavelet kernel in (4.3) can be rewritten in the form of matrix as:

$$\mathbf{K} = \sum_m \sum_{n\in N} \psi_{m,n}(\mathbf{P}) \circ \psi_{m,n}(\mathbf{P}^\intercal) \tag{4.12}$$

where $\mathbf{P} = [1,2,\cdots,M]^\intercal \cdot [1,1,\cdots,1] \in \mathbb{R}^{M\times M}$ is the dot product of the vector of nature number from 1 to M and the all-ones vector, denoting the position of each impulse response. $\circ$ is known as Hadamard product performing element-wise multiplication. $\psi_{m,n}(\mathbf{P})$ can also be realized using element-wise operation.

The kernel matrix $\mathbf{K}(\boldsymbol{\omega})$ is parameterized by the bound of scale factor $\boldsymbol{\omega} = [m_{\text{upper}}, m_{\text{lower}}]^\intercal$. Tuning such hyper-parameter in the prior distribution is called the empirical Bayes method, which maximize the marginal likelihood $P(\mathbf{y}|\boldsymbol{\omega})$. This is equivalent to minimizing the following equation:

$$\boldsymbol{\omega}^* = \arg\min_{\boldsymbol{\omega}} \mathbf{y}^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega})\mathbf{y} + \log\det \boldsymbol{\Sigma}(\boldsymbol{\omega}) \tag{4.13}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\omega}) = \mathbf{X}\mathbf{K}(\boldsymbol{\omega})\mathbf{X}^\intercal + \sigma^2\mathbf{I}$. Although (4.13) is not convex, it can be solved by using heuristic optimization such as the DIRECT algorithm [74], or sequential convex optimization by converting the problem to multiple kernel learning [22].

A simple implementation of the proposed modelling-based feature extraction method is presented in Algorithm 4.

Algorithm 4: Modelling-based feature extraction

Input: Sensor's response from the first two stages $\mathbf{y} \in \mathbb{R}^N$, Number of inpulse response M, trade-off parameter σ^2, and scale factor bound $\boldsymbol{\omega} = [m_{\text{upper}}, m_{\text{lower}}]^\intercal$.

Output: Feature sequence $\boldsymbol{\theta}$.

Initialization: $\mathbf{x} \leftarrow \mathbf{1}_N^\intercal$.
$\mathbf{X} \leftarrow [\mathrm{T}^0\mathbf{x}, \mathrm{T}^1\mathbf{x}, \cdots, \mathrm{T}^{\text{M-1}}\mathbf{x}]$.
Initialization: $\mathbf{v} \leftarrow [1, 2, \cdots, M]^\intercal$.
for $m = m_{\text{lower}}, m_{\text{lower}}+1, \cdots, m_{\text{upper}}$ **do**
  for $n = 1, 2, \cdots, N$ **do**
    $\mathbf{K}_{m,n} \leftarrow \psi_{m,n}(\mathbf{v}\mathbf{1}_M^\intercal) \circ \psi_{m,n}^\intercal(\mathbf{v}\mathbf{1}_M^\intercal)$.
    $\mathbf{K} \leftarrow \mathbf{K} + \mathbf{K}_{m,n}$.
  end
end
$\boldsymbol{\theta} \leftarrow \mathbf{K}\mathbf{X}^\intercal(\mathbf{X}\mathbf{K}\mathbf{X}^\intercal + \sigma^2\mathbf{I})^{-1}\mathbf{y}$.
return $\boldsymbol{\theta}$.

4.3 Numerical Experiment

The purpose of this numerical experiment is to verify the effectiveness of multiscale wavelet kernel-based regularization in resistance to noise. For curve fitting or modelling-based feature extraction methods, the more accurate the reconstructed signal from the features is, the more robust in resistance to noise the features will be.

4.3.1 Experiment Setup

The simulated signal $y^*(t)$ containing the baseline and the absorption stage is performed as the ground truth. Recall that the thermal agitation of electrons inside the sensor causes thermal noise in sensor's reading. Thermal noise in an ideal resistor can be considered as Gaussian white noise. Accordingly, Gaussian white noise is added to a simulated sensor's response as the observation. The signal-to-noise ratio (SNR) is ranged from 10 dB to 40 dB. Features (i.e. the impulse responses) are extracted from the noisy observation using different kernels. Then, the signal $\widetilde{y}(t)$ is reconstructed from the impulse response and compared with the simulated response $y^*(t)$ by measuring the normalized root mean square error (NRMSE) as follows. The smaller the NRMSE, the more stable the method.

$$\epsilon_{\text{NRMSE}} = \frac{||\widetilde{y}(t) - y^*(t)||_2^2}{||y^*(t) - \text{mean}\, y^*(t)||_2^2}. \tag{4.14}$$

A theoretical Figaro TGS MOS-sensor's resistance value change with gas concentration is as follows:

$$R_{\text{sensor}}(v) = R_0(1 + k[v])^{-\beta} \tag{4.15}$$

where R_0 denotes the sensor's resistance value to fresh air. k represents the rate constant of an odour. $[v]$ is the odour concentration in mol·dm^{-3}. $\beta \in [0.25, 0.55]$ is the sensor's exponent constant. According to the measurement circuit (as shown in Fig. 4.2), the simulated sensor's response across load resistor R_{load} (as shown in Fig. 4.3) can be calculated as:

$$V_{\text{out}} = \frac{V_{\text{cc}} R_{\text{load}}}{R_{\text{sensor}} + R_{\text{load}}} \tag{4.16}$$

4.3.2 Comparison of Different Kernel-Based Modelling Methods

The kernels selected for comparison are widely used, including the radial basis function (RBF) kernel, the stable spline (SS) kernel, the tuned correlated (TC) kernel, the Mexican hat wavelet (Mex) kernel, the Morlet wavelet (Mor) kernel, and the Shannon wavelet (Sha) kernel. By defining $\max\{\cdot,\cdot\}$, $\min\{\cdot,\cdot\}$, $\cos(\cdot)$, $\text{sinc}(\cdot)$, and $\exp\{\cdot\}$ as element-wise

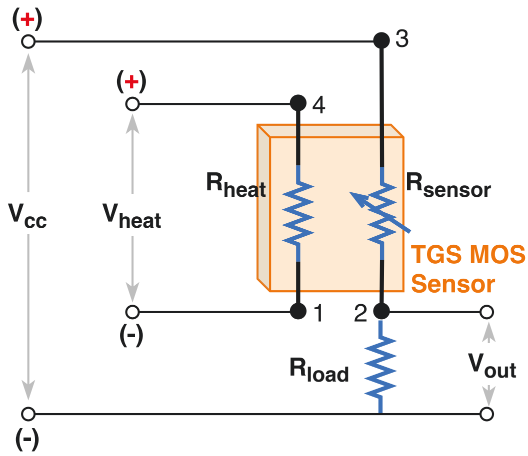


Figure 4.2: Basic measuring circuit of Figaro TGS MOS sensor. The integrated heater (R_{heat}) provides an optimal temperature for the sensing element (R_{sensor}). The sensor's reading is the voltage across the load resistor (R_{load}) connected in series with the sensor. Both of the V_{cc} and V_{heat} are constant voltage sources.

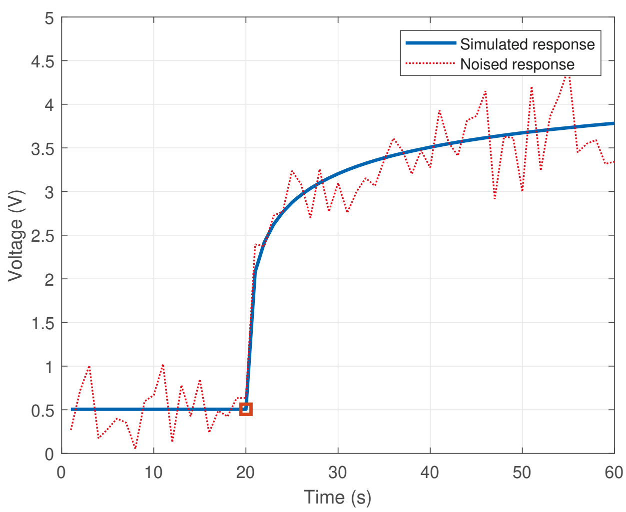


Figure 4.3: Simulated TGS MOS sensor's response. The blue line denotes the simulated response, and the red dashed line denotes the noised observation. The features are extracted from the noised observation. The experiment parameters were set as: $V_{\text{cc}} = 5$ V, $R_{\text{load}} = 0.45$ kΩ, $R_0 = 4$ kΩ, $k = 100$, and $\beta = 0.4$. The concentration of odour is ranged from 0 to 40 mol·dm^{-3}, and the initial time $t_{\text{gas-on}} = 20$ s.

operations, the kernel matrices can be calculated as:

$$\mathbf{K}^{(\mathrm{RBF})} = \exp\left\{-\frac{1}{2\alpha^2}\left(\mathbf{P}-\mathbf{P}^\intercal\right)\circ\left(\mathbf{P}-\mathbf{P}^\intercal\right)\right\} \tag{4.17}$$

$$\mathbf{K}^{(\mathrm{SS})} = \frac{1}{2}\alpha\beta^{\max\{\mathbf{P},\mathbf{P}^\intercal\}}\circ\left(\beta^{\min\{\mathbf{P},\mathbf{P}^\intercal\}} - \frac{1}{3}\beta^{\max\{\mathbf{P},\mathbf{P}^\intercal\}}\right) \tag{4.18}$$

$$\mathbf{K}^{(\mathrm{TC})} = \exp\left\{-\alpha(\mathbf{P}+\mathbf{P}^\intercal)\right\}\circ\exp\left\{-\alpha|\mathbf{P}-\mathbf{P}^\intercal|\right\} \tag{4.19}$$

$$\mathbf{K}^{(\mathrm{Mor})} = \cos\left(1.75\times\frac{\mathbf{P}-\mathbf{P}^\intercal}{\alpha}\right)\circ\exp\left\{-\frac{1}{2\alpha^2}\left(\mathbf{P}-\mathbf{P}^\intercal\right)\circ\left(\mathbf{P}-\mathbf{P}^\intercal\right)\right\}$$

$$\mathbf{K}^{(\mathrm{Sha})} = 2\,\mathrm{sinc}\left(2\left(\frac{\mathbf{P}-\mathbf{P}^\intercal}{\alpha}\right)\right) - \mathrm{sinc}\left(\frac{\mathbf{P}-\mathbf{P}^\intercal}{\alpha}\right) \tag{4.20}$$

where α and β are hyper-parameters for each kernel.

Here, the mother wavelet selected for constructing multiscale kernel is the Shannon wavelet (4.5). Hyper-parameters for each kernel matrix at different noise level are optimized according to (4.13) by using the DIRECT algorithm.

Each experiment is repeated for 100 times. The statistical result is shown in Table 4.1. On the condition of high SNR (>25 dB), the NRMSE of the reconstructed signal using the multiscale wavelet kernel is commensurate with those using single-scale kernels. The multiscale wavelet kernel gradually outperforms others with the increase of the noise level. The result of numerical experiment demonstrates the use of multiscale wavelet kernel-based regularization increases the stability of the modelling-based features.

4.4 Field Experiment

The purpose of this experiment is to demonstrate the effectiveness of our feature extraction method in odour identification. This experiment was displayed on CeBIT (Centrum für Büroautomation, Informationstechnologie und Telekommunikation) Australia 2019 exhibition.

4.4.1 Experiment Setup

4.4.1.1 Material Preparation and Instrument Settings

This experiment investigates four brands of alcoholic drinks, including two single malt whiskies and two blended whiskies. We put 5 ml of each whiskey into a sealed 20-ml headspace crimped top vial as a test material. Then, measure each material by a self-designed e-nose. When the number of measurements for a material exceeds 20 times, we discard the test material, and prepare the same one in a new top vial. The record of

Table 4.1: Simulated signal reconstruction by kernel-based modelling with different kernels

Kernel	NRMSE						
	SNR=40 dB	SNR=35 dB	SNR=30 dB	SNR=25 dB	SNR=20 dB	SNR=15 dB	SNR=10 dB
Multiscale	**0.0113**	**0.0148**	**0.0223**	**0.0376**	**0.0602**	**0.0763**	**0.0983**
	±0.0006	**±0.0009**	**±0.0015**	**±0.0028**	**±0.0057**	**±0.0060**	**±0.0102**
RBF	0.0117	0.0155	0.0239	0.0403	0.0709	0.1114	0.1479
	±0.0006	±0.0009	±0.0016	±0.0032	±0.0047	±0.0106	±0.0151
SS	0.0688	0.0696	0.0723	0.0814	0.1134	0.2190	0.3097
	±0.0013	±0.0023	±0.0044	±0.0112	±0.0307	±0.1022	±0.1609
TC	0.0747	0.0748	0.0750	0.0761	0.0796	0.0909	0.1293
	±0.0002	±0.0004	±0.0010	±0.0013	±0.0029	±0.0066	±0.0130
Morlet	0.0117	0.0154	0.0239	0.0404	0.0709	0.0954	0.1355
	±0.0005	±0.0009	±0.0018	±0.0036	±0.0047	±0.0071	±0.0129
Shannon	0.0117	0.0154	0.0244	0.0402	0.0717	0.1267	0.2297
	±0.0006	±0.0012	±0.0015	±0.0028	±0.0060	±0.0087	±0.0152

one measurement is regarded as one sample. Three different datasets are used in our field experiment. The training set was collected on different days (12-27/Oct./2019) at different labs of UTS (UTS Labs). The environment temperature and humidity at each lab are relatively constant. Since the sensor installed in the e-nose would be influenced by the environment change, a calibration set was collected at the first day (29/Oct./2019) at an exhibition hall of the Sydney Convention and Exhibition Centre (ICC Sydney) to overcome this influence. The testing set was collected randomly at the same hall during the CeBIT Australia 2019 exhibition (29-31/Oct./2019). The information of the dataset is listed in Table 4.2.

Table 4.2: Whisky information (four targets)

Lable	Product Information	Sample Size		
		Training Set	Calibration Set	Testing Set
1	Ardbeg 10 Year Old (Single Malt, ABV* 46%)	240	20	40
2	Macallan 12 Year Old (Single Malt, ABV 43%)	240	20	57
3	Johnnie Walker Red Label (Blended, ABV 40%)	240	20	62
4	Chivas Regal 12 Year Old (Blended, ABV 40%)	240	20	41

* ABV stands for alcohol by volume.

A self-designed e-nose, UTS NOS.E III, installed with four Figaro MOS-type gas sensors (i.e. TGS2600, TGS2602, TGS2603, and TGS2620) is adopted in our experiment. More details about the e-nose can be found in [214]. Before taking the first measurement on the day, the e-nose should run (initialize) in the ambient air for 60 min with all valves and pumps open. The duration of one measurement is 60 s. The time interval between two measurements is also set as 60 s, during which all valves and pumps are opened for washing. The figure of experiment setup is displayed in Fig. 4.4.

4.4.1.2 Experiment Process

The diagram of experiment process is shown in Fig. 4.5. Some basic elements in the process are explained as follows:

- **Normalization**: According to previous linear assumptions for both sensor discrepancy and concentration fluctuations, normalization (2.3), is applied to the sensor's

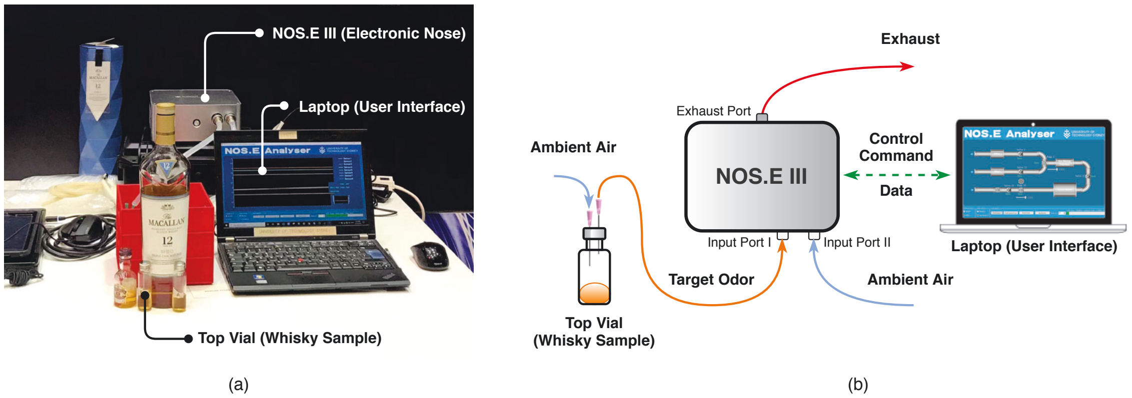


Figure 4.4: Experiment setup. (a) Field experiment on CeBIT Australia 2019. (b) Diagram of material measurement. Firstly, ambient air is drawn into the sealed sensor chamber through the input port II for 10 s to obtain a baseline for each sensor. Next, the target odour is inhaled by a sampling needle through the input port I for 10 s. Then, the sensor chamber is cleaned by the ambient air for 40 s. The flow rate is kept constant at 1.1 L/min. The sampling rate of the e-nose is set as 10 Hz. After one measurement, the collected data is sent to a laptop for further processing via the Wi-Fi.

response to reduce the effect of discreteness.

- **Drift calibration**: Sensor drift would cause feature distribution inconsistency in different environments. Hence, a recently developed domain adaptation algorithm, domain regularized component analysis (DRCA) [209], is adopted. This method introduces a regularization in the target domain, which allows the training with few calibration samples. The projection (calibration) matrix is given by:

$$\mathbf{A} = \arg\max_{\mathbf{A}} \frac{\mathrm{tr}(\mathbf{A}^{\intercal}(\Theta_{\mathscr{S}}\Theta_{\mathscr{S}}^{\intercal} + \eta\Theta_{\mathscr{T}}\Theta_{\mathscr{T}}^{\intercal})\mathbf{A})}{\mathrm{tr}(\mathbf{A}^{\intercal}(\boldsymbol{\mu}_{\mathscr{S}} - \boldsymbol{\mu}_{\mathscr{T}})(\boldsymbol{\mu}_{\mathscr{S}} - \boldsymbol{\mu}_{\mathscr{T}})^{\intercal}\mathbf{A})} \tag{4.21}$$

where $\mathbf{A} \in \mathbb{R}^{M\times d}$ denotes the projection matrix. M and d represent the dimension of the feature vector before and after domain adaption, respectively. $\Theta_{\mathscr{S}} = [\boldsymbol{\theta}_{\mathscr{S}}^{(1)},\ldots,\boldsymbol{\theta}_{\mathscr{S}}^{(N_{\mathscr{S}})}] \in \mathbb{R}^{M\times N_{\mathscr{S}}}$ is the training set in the source domain at UTS Labs, and $\Theta_{\mathscr{T}} = [\boldsymbol{\theta}_{\mathscr{T}}^{(1)},\ldots,\boldsymbol{\theta}_{\mathscr{T}}^{(N_{\mathscr{T}})}] \in \mathbb{R}^{M\times N_{\mathscr{T}}}$ its corresponding feature set used for calibration in the target domain at ICC Sydney. $N_{\mathscr{S}}$ and $N_{\mathscr{T}}$ represent the sample size of training set and calibration set, respectively. $\boldsymbol{\mu}_{\mathscr{S}}$ and $\boldsymbol{\mu}_{\mathscr{T}}$ are centroids of $\Theta_{\mathscr{S}}$ and $\Theta_{\mathscr{T}}$, respectively. η is a trade-off parameter to alleviate the data imbalance between training set and calibration set. Accordingly, the representation of a given sample $\boldsymbol{\theta}$ in the new feature space is given by $\mathbf{A}^{\intercal}\boldsymbol{\theta}$.

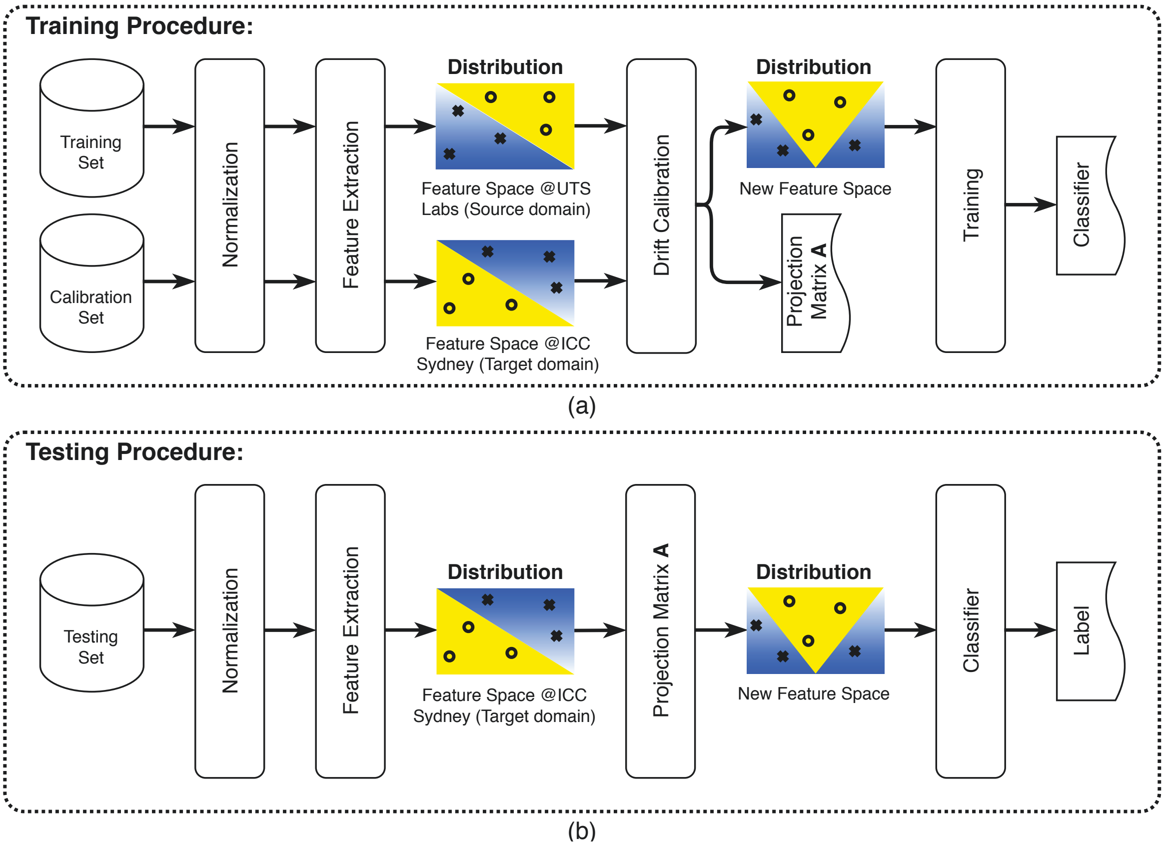


Figure 4.5: Flow diagram of experiment process. (a) Training procedure. (b) Testing procedure.

- **Classification**: A good feature representation will facilitate subsequent pattern recognition algorithms. In this experiment, a simple classifier, k-nearest neighbors (KNN), is adopted to identify whiskies. The classifier is build on the new feature space after domain adaption. The hyper-parameters for both DRCA and KNN are optimized by maximizing the classification accuracy of the five-fold cross validation (5-CV). In the 5-CV, both the training set and the calibration set are divided into five subsets and composed as five pairs. Four pairs are used for training, and the rest one pair for validation. This procedure is then repeated for five times. When optimized parameters are achieved, all training and calibration samples are used to generate the final classifier for testing.

4.4.2 Comparison of Different Feature Extraction Methods

In this section, the proposed feature extraction method is compared with the representative method in other categories mentioned in Section 2.3, including handcrafted features, polynomial features, and SAE-based features. In addition, the performance of time series classification, i.e. directly classifying sensor array's raw response without feature extraction, is applied as the baseline.

The hyper-parameters used for the proposed method are as follows: multiscale Shannon kernel with the optimized scale factor $m \in [-10, 10]$ and the number of impulse response $M = 20$. In the case of handcrafted method, only derivative-based features are compared since the responses-related features provide no information after normalization. The derivative-based features [107, 215] are (see Table 2.1): f_2, f_3, f_4, f_6, f_7, f_9, and f_{10}. Moreover, WTDF is adopted to denoise the response curve before extracting the handcrafted features. In the case of the curve fitting-based method, the order of polynomial function is selected as 4. The SAE used in this experiment is a three-layer net with 64 nodes in the hidden layer. This parameter is selected according to the highest accuracy of the cross validation. In case of time series classification, the distance metric used in the KNN is dynamic time warping. The result is summarized in Table 4.3. The detailed result (confusion matrix) for the field experiment is illustrated in Fig. 4.6.

The result of the time series classification (Fig. 4.6(e)) shows the similarity of the response curves to some extent. From this view, the sensor array's response to Whisky 1 is significantly different from that to other whiskies. This is also observed by the highest sensitivity and relative low false rate of label 1 in each considered feature extraction method. All of the considered features are able to detect the difference of whisky 4 from whisky 2, although their time series cannot be classified. Whisky 2 and Whisky 3 are the most difficult to distinguish, even though their composition and ABV are different. The polynomial curve fitting-based feature performs worst among all considered feature extraction methods. Compared with the handcrafted features, the accuracy via proposed method is much better, even without applying any additional filtering or smoothing technique. Both of the SAE and the proposed method achieve good performance in whisky identification. The proposed method performs better than SAE in identifying whisky 3.

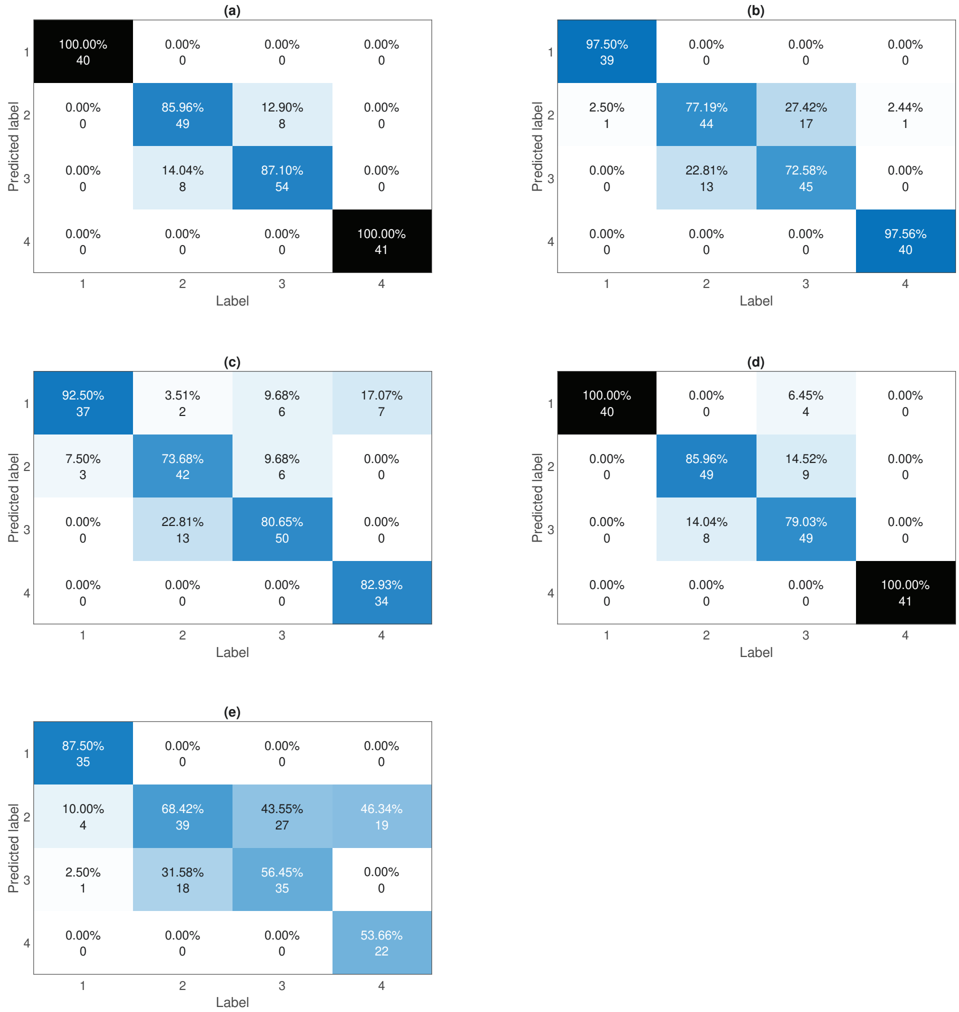


Figure 4.6: Confusion matrix for the field experiment (the testing procedure). (a) Proposed features. Sensitivity (Sen.) for each label: 100%, 85.96%, 87.10%, and 100%; Specificity (Spe.) for each label: 100%, 94.41%, 94.20%, and 100%. (b) Handcrafted features. Sen.: 97.50%, 77.19%, 72.58%, and 97.56%; Spe.: 99.38%, 90.91%, 87.68%, and 99.37%. (c) Polynomial curve fitting-based features. Sen.: 92.50%, 73.68%, 80.65%, and 82.93%; Spe.: 98.12%, 89.51%, 91.30%, and 95.60%. (d) Sparse auto-encoder features. Sen.: 100%, 85.96%, 79.03%, and 100%; Spe.: 100%, 94.41%, 90.58%, and 100%. (e) Time series (no feature extraction). Sen.: 87.50%, 68.42%, 56.45%, and 53.66%; Spe.: 96.88%, 87.41%, 80.43%, and 88.05%.

Table 4.3: Classification Results Via Different Feature Representations

Method	Hyper-parameters			Accuracy (%)	
	DRCA		KNN	Training Procedure (5-CV)	Testing Procedure
	η	d	k	(Training Set + Calibration Set)	(Testing Set)
Proposed Features	1	18	9	**95.38 ± 1.47**	**92.00**
Handcrafted Features	1	10	6	88.65 ± 2.03	84.00
Polynomial Curve Fitting-Based Features	1	5	8	87.40 ± 1.15	81.50
Sparse Auto-Encoder Features	3	16	10	93.27 ± 2.47	89.50
Time series (No Feature Extraction)	5	69	3	69.62 ± 1.64	65.50

4.4.3 Comparison in Resistance to Noises

The proposed feature extraction is able to resist random noise. To verify this, we add additional white Gaussian noise (SNR ranges from 50dB to 20dB) to the sensor array's record obtained from the above field experiment. Then, the training and testing procedure (as shown in Fig. 4.5) is operated on the noise-contaminated datasets. All hyper-parameters for each considered method are also re-optimized at different SNR levels accordingly. The filtering algorithm is only applied before extracting handcrafted features. The result of 100 repeated experiments for each SNR level is shown in Fig. 4.7.

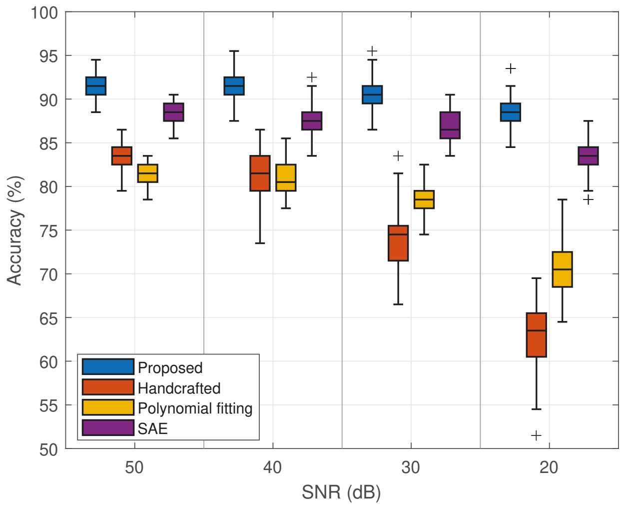


Figure 4.7: Boxplot of different features in resistance to Gaussian white noises

The classification accuracy via the proposed features falls slowly with the increase of noise level in comparing with other considered features. The multiscale wavelet kernel regularization helps stabilize the extracted features. SAE also performs well in resisting random noise. The anti-noise ability of SAE comes from its sparse and abstract representation. However, in order to obtain this ability, SAE needs to increase the number of hidden layers or hidden neurons, which also increases the computational cost. Although the handcrafted features are extracted after filtering, this method performs worst among all the considered methods. The derivative operation (the differentiation operator) acts as a high-pass filter which has a tendency of amplifying the noise to an undesirable level. The higher the order of the derivative, the more the result affected by noise.

4.5 Conclusion

In this chapter, we propose a kernel regularization-based feature extraction method for tackling the discreteness problem faced in the real application of e-nose technology. The success of kernel methods highly relies on the capacity of kernel function. Here, a multiscale wavelet kernel derived from the orthogonal projection operator is applied as the regularization matrix in solving the inverse problem - estimating impulse response (i.e. the features for target odours) from noisy sensor's response. A numerical experiment is given to verify the stability of the proposed method. Moreover, we also demonstrate the proposed method in a real application scenario. Under the framework of transfer learning, the proposed feature extraction method is applied in identifying the odours of four different whiskies measured by a self-designed e-nose with four commercial available MOS gas sensors. The obtained results outperform the classification using other considered features by reaching an accuracy of 92.00%. The results have shown a good potential of applying the proposed feature representations in the area of e-nose. The weakness of this method is the computational cost in constructing the kernel matrix when the observation sequence is very long. This is why we introduce element-wise operations (matrix operations) in kernels. Moreover, since the kernel matrix is independent of the observation, we can calculate it off-line. In order to solve this problem, our future work will focus on improvement through segmented estimation and real-time estimation.

CHAPTER 5

Numerical Differentiation Based on Kernel Regularization to Improve Transient-State Features

5.1 Introduction

Traditional hand-crafted features (i.e. steady-state and transient-state features) are considered to be most susceptible to the discreteness issue. The steady-state feature is sensitive to sensor manufacturing discrepancy and odour concentration variation. Once normalization is adopted to reduce such influences, the discriminant information it has provided will be totally eliminated. The transient-state features designed from derivatives are sensitive to noise since numerical differentiation has a tendency to amplify the influence of noise [120]. To tackle these shortages, in recent years, alternative feature extraction methods have been proposed as reviewed in Section 2.3. However, traditional features are still popular in many tasks such as sensor array optimization and flow rate optimization [135, 162], because they are easy to extract and have physical meaning.

In order to obtain accurate transient-state features, numerical differentiation methods should be adopted. We roughly divide the existing numerical differentiation (differential estimation) algorithms based on noisy signals into two groups: the indirect way and the direct way.

For the noise-contaminated observation y of a continuously differentiable function

g, the way of indirectly calculating g' from y is to construct a less noisy or smooth approximation $f \approx g$, and then compute $f' \approx g'$ using finite difference formulas (e.g. the forward and backward differences). Most of existing methods to construct f fall into following categories: filtering, regression, and Tikhonov regularization. i) Filtering methods (e.g. finite impulse response (FIR) filter, Fourier or wavelet filter) approximate f by reducing the noise in y. Although the noise in the filtered signal f has been suppressed to a small level δ, such that $\|g - f\|_2 < \delta$, we can still have an arbitrary large $\|g' - f'\|_2$. The finite difference formula has a tendency to amplify the noise. The higher the order we calculate, the less accurate the derivative is. ii) Regression methods are applied to model a smooth approximation f. In this category, the simplest algorithm is polynomial regression. Progresses of this linear method are made from two aspects: introducing regularization technique to avoid overfitting (e.g. ridge regression and lasso regression), and introducing kernel method to extend the regression to nonlinear area (e.g. the kernel ridge regression (KRR) and support vector regression (SVR)). Moreover, ensemble strategies (e.g. multiple kernel learning and ensemble learning) are also introduced to improve the general performance regression algorithms. iii) Another extension of ridge regression is Tikhonov regularization, whose re-construction error (error term) $\|y - f\|_2$ in the cost function is constrained by the upper bound of the d^{th}-order derivative $\|f^{(d)}\|_2$. A recent developed segment-wise regularization framework improves the boundary smoothness of the approximation f. Based on this, a real-time improvement of Tikhonov regularization can be achieved to effectively process long sequences.

Instead of constructing f, an alternative way is directly estimating f' from y. We further classify the direct way into two categories: Tikhonov regularization and Fourier truncation method. i) By introducing a Volterra integral equation $\mathrm{T}_1 f' = \int_0^t f'(\eta)d\eta$ [179], the error term in the cost function can be re-written as $||\hat{y} - \mathrm{T}_1 f'||_2$, where $\hat{y} = y - f(0)$ [87]. Then, Tikhonov regularization is applied to ensure an accurate estimation of f'. ii) Another direct way is to reformulate operator T_1 by truncated singular value expansion (TSVE). This setting coincides with estimating the differentiation the truncated Fourier series of y term-by term [12]. The Fourier truncation methods are widely applied to compute numerical differentiation for periodic signals.

In this chapter, we propose a novel numerical differentiation algorithm based on the Tikhonov regularization and kernel method to ensure an accurate and stable transient-state feature extraction for e-noses. The proposed algorithm is able to directly estimate high-order derivatives (i.e. 2^{nd}-order derivative) from the noise-contaminated signal without applying any pre-processing technique. The kernel method is applied to the regularization term as priors regarding the signal, which equips the proposed method with an ability to resist random noise.

The primary contribution of this chapter are summarized as follows:

- A new derivative estimation algorithm is proposed and verified by a numerical experiment.
- A framework is proposed to tackle the 3D issue encountered in real applications for e-noses.
- We also compare the improved transient-state features with other widely used features in a field experiment displayed on the CeBIT Australia 2019 exhibition. The training and testing samples are obtained from different places. We also added two new classes as disturbances. Accordingly, both drift and disturbance issue are considered in this experiment.

The rest of the chapter is structured as follows: Section 5.2 is devoted to describing the whole methodology of numerical differentiation algorithm based on kernel Tikhonov regularization. To verify the proposed method, a numerical experiment is conducted and discussed in Section 5.3. Then, a field experiment is conducted and discussed in Section 5.4. The task of the field experiment is to identify four target whiskies and two disturbances. Finally, Section 5.5 gives a short conclusion of this chapter.

5.2 Methodology

5.2.1 Derivative Estimation for Electronic Nose Signals

We start the derivation of the proposed method from a special condition, i.e. $g^{(d)}(0) = 0$, where $d = 1, \ldots, D$. This setting is similar to the characteristics of e-nose signals, that is, e-nose signals usually respond from a constant baseline.

Let $g[n]$ be the discrete-time sampling of a D times differentiable function $g(t)$. A sampling of $g[n]$ with length N can be vectorized as $\boldsymbol{g} = [g_{[0]}, \ldots, g_{[N-1]}]^\top \in \mathbb{R}^N$. It's corresponding discrete-time observation (i.e. gas sensor's reading) $\boldsymbol{y} = [y_{[0]}, \ldots, y_{[N-1]}]^\top \in \mathbb{R}^N$ is assumed to be contaminated by additive Gaussian noise $\boldsymbol{\epsilon} = [\epsilon_{[0]}, \ldots, \epsilon_{[N-1]}]^\top \in \mathbb{R}^N \sim \mathcal{N}(0, \sigma^2)$, such that $\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{\epsilon}$. Let $u_d(t)$ be the d^{th}-order numerical derivative of $g(t)$, i.e. $u_d(t) = g^{(d)}(t)$. Correspondingly, $u_d[n]$ can be represented as $\boldsymbol{u}_d = [u_{d[0]}, \ldots, u_{d[N-1]}]^\top \in \mathbb{R}^N$.

Under the constraint of $g^{(d)}(0) = 0$, the relationship between $g(t)$ and $u_D(t)$ is as follows:

$$\overbrace{\int_0^t \int_0^x \cdots \int_0^\xi}^{D} u_D(\eta)\, \underbrace{d\eta\, d\xi \cdots dx}_{D} = g(t) - g(0). \tag{5.1}$$

By defining the integration operator as:

$$\phi_D(t) = \overbrace{\int_0^t \int_0^x \cdots \int_0^\xi}^{D-1} 1\, \underbrace{d\eta\, d\xi \cdots dx}_{D-1} \tag{5.2}$$

(5.1) can be simplified as the convolution of $u_D(t)$ and $\phi_D(t)$:

$$(u_D * \phi_D)(t) = \int_0^t u_D(\tau)\phi_D(t-\tau)\, d\tau = g(t) - g(0). \tag{5.3}$$

For a discrete-time system, the discrete convolution $(u_D \circledast \phi_D)[n]$ can be expressed compactly as $\boldsymbol{\Phi}_D \boldsymbol{u}_D$, where $\boldsymbol{\Phi}_D$ is constructed from the discretized $\phi_D[n]$ with zero-padding, as:

$$\boldsymbol{\Phi}_D = \begin{bmatrix} \phi_{D[0]} & 0 & \cdots & 0 \\ \phi_{D[1]} & \phi_{D[0]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{D[N-1]} & \phi_{D[N-2]} & \cdots & \phi_{D[0]} \end{bmatrix} \in \mathbb{R}^{N\times N}. \tag{5.4}$$

We take the 2^{nd}-order derivative as an example. Based on (5.2), $\phi_2(t) = t$. Then, $\boldsymbol{\Phi}_2$ can be expressed as:

$$\boldsymbol{\Phi}_2 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ N & N-1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times N}. \tag{5.5}$$

Accordingly, derivative estimation from noisy observation $\boldsymbol{y}$ can be regarded as a deconvolution problem. However, deconvolution is ill-posed. Inspired by recent developed kernel-based modelling [15, 81, 101], kernel-based Tikhonov regularization algorithm is proposed, whose cost function is given as:

$$J(\boldsymbol{u}_D) = \|\boldsymbol{y} - \boldsymbol{y}_0 - \boldsymbol{\Phi}_D \boldsymbol{u}_D\|_2^2 + \sigma^2 \boldsymbol{u}_D^\top \mathbf{K}^{-1} \boldsymbol{u}_D \tag{5.6}$$

where $\boldsymbol{y}_0 = [y_{[0]}, \ldots, y_{[0]}]^\top \in \mathbb{R}^N$ denotes the vector of initial value. Here, we apply $y_{[0]}$ instead of $g_{[0]}$. $\mathbf{K} = [\kappa\langle i,j\rangle] \in \mathbb{R}^{N\times N}$ is a positive-definite kernel matrix. Unlike the kernel-based learning algorithm (whose kernel function $\kappa\langle\cdot,\cdot\rangle$ operates on sample vectors, the $i, j = \{1, \ldots, N\}$ in $\mathbf{K}$ represents position indices.

Accordingly, the derivative $\boldsymbol{u}_D$ can be calculated by minimizing the cost function via regularized least squares as:

$$\begin{aligned} \boldsymbol{u}_D^* &= \arg\min_{\boldsymbol{u}_D} J(\boldsymbol{u}_D) \\ &= \mathbf{K}\boldsymbol{\Phi}_D^\top(\boldsymbol{\Phi}_D\mathbf{K}\boldsymbol{\Phi}_D^\top + \sigma^2\mathbf{I}_N)^{-1}(\boldsymbol{y} - \boldsymbol{y}_0) \end{aligned} \tag{5.7}$$

where $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ represents a identity matrix.

The kernel matrix $\mathbf{K}$ represents prior guessing regarding to the solution. Accordingly, we can construct $\mathbf{K}$ without considering observations. In (4.17), we list some widely used smoothing kernels. These kernel can also be used in this study.

The proposed method relies on the selection of kernel parameters. The kernel matrix can be written as a function of the hyper-parameter, i.e. $\mathbf{K}(\omega)$. Then, ω can be optimized by maximizing the marginal likelihood $\mathrm{P}(\boldsymbol{y} - \boldsymbol{y}_0|\omega)$, which is equivalent to minimize the following function:

$$\omega^* = \arg\min_{\omega}(\boldsymbol{y} - \boldsymbol{y}_0)^\top\boldsymbol{\Sigma}^{-1}(\omega)(\boldsymbol{y} - \boldsymbol{y}_0) + \log\det\boldsymbol{\Sigma}(\omega) \tag{5.8}$$

where $\boldsymbol{\Sigma}(\omega) = \boldsymbol{\Phi}_D\mathbf{K}(\omega)\boldsymbol{\Phi}_D^\top + \sigma^2\mathbf{I}_N$. Equation (5.8) is actually the problem of difference of convex programming (DCP). To solve this, we can use heuristic searching methods such as the DIRECT algorithm, or convert the problem to multiple kernel learning on a set of given $\mathscr{W} = \{\omega_m\}_{m\in M}$, i.e. $\mathbf{K}(\boldsymbol{\lambda}) = \sum_{\mathscr{W}} \lambda_m\mathbf{K}(\omega_m), \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M]^\top$, and then optimize $\boldsymbol{\lambda}$ using sequential convex optimization.

5.2.2 Boundary Distortion Reduction Using Polynomial Extension

The proposed method suffers from the influence of boundary distortion, which is derived from two aspects: i) the inaccurate initial value, i.e. $y(0) \neq g(0)$, caused by noise contamination; ii) a sharp increase in variance and bias in boundary region caused by boundary effect.

A simple improvement can be made by using extensions based on the D^{th}-degree polynomial regression, where $D \leqslant m < N/2$. The polynomial expansion helps to avoid artificial discontinuities of the extended signal and its high-order derivatives at the boundary. The underlying idea can be realized in Algorithm 5. An illustration of the extended signal is shown in Fig. 5.1.

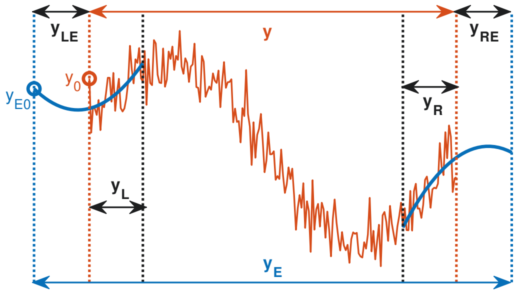


Figure 5.1: Illustration of polynomial extension. $\boldsymbol{y} \in \mathbb{R}^N$ is the input signal. $\boldsymbol{y}_L \in \mathbb{R}^m$ ($\boldsymbol{y}_R \in \mathbb{R}^m$) represents the first (last) m points of $\boldsymbol{y}$ used to training polynomial function $p_L(t)$ ($p_R(t)$). $\boldsymbol{y}_{LE} \in \mathbb{R}^m$ ($\boldsymbol{y}_{RE} \in \mathbb{R}^m$) represents the related extension. $\boldsymbol{y}_E$ denotes the polynomial extension. $y_{[0]}$ ($y_{E[0]}$) denotes the first point of $\boldsymbol{y}$ ($\boldsymbol{y}_E$).

Algorithm 5: Polynomial extension

Input: Signal $\boldsymbol{y} \in \mathbb{R}^N$, $D \in \mathbb{Z}^+$, $m \in \mathbb{Z}^+$.
Output: Extended signal $\boldsymbol{y}_E \in \mathbb{R}^{N+2m}$.

Initialization: $\boldsymbol{y}_L \leftarrow \boldsymbol{y}_{[0:m-1]}$; $\boldsymbol{y}_R \leftarrow \boldsymbol{y}_{[N-m:N-1]}$.
Train D^{th}-degree polynomial $p_L(t)$ via $\boldsymbol{y}_{[0:m-1]}$.
Train D^{th}-degree polynomial $p_R(t)$ via $\boldsymbol{y}_{[N-m:N-1]}$.
$\boldsymbol{y}_{LE} \leftarrow p_L([-m:-1])$.
$\boldsymbol{y}_{RE} \leftarrow p_R([N:N+m-1])$.
$\boldsymbol{y}_E = [\boldsymbol{y}_{LE}^\top, \boldsymbol{y}^\top, \boldsymbol{y}_{RE}^\top]^\top$.
Update the index: $\boldsymbol{y}_{E[0:N+2m-1]} \leftarrow \boldsymbol{y}_{E[-m:N+m-1]}$.
return $\boldsymbol{y}_E$.

Then, derivative estimation is performed on the extended signal $\boldsymbol{y}_E \in \mathbb{R}^{N+2m}$ and the estimated initial vector $\boldsymbol{y}_{E0} = [y_{E[0]}, \ldots, y_{E[0]}] \in \mathbb{R}^{N+2m}$. A simple implementation of the proposed derivative estimation algorithm with polynomial extension is shown in Algorithm 6.

5.2.3 Derivative Estimation for Arbitrary Differentiable Functions

For arbitrary differentiable functions, i.e. $\forall\, d \in D$, $\exists\, g^{(d)}(0) \neq 0$, (5.1) can be re-written as:

$$\overbrace{\int_0^t \int_0^x \cdots \int_0^\xi}^{D} u_D(\eta)\, \underbrace{d\eta\, d\xi \cdots dx}_{D} = g(t) - \sum_{d=0}^{D-1} \frac{1}{d!} g^{(d)}(0) t^d. \tag{5.9}$$

Accordingly, derivative estimation for such signals can be realized iteratively using the proposed kernel-based Tikhonov regularization method, as shown in Algorithm 7.

Algorithm 6: Derivative estimation (special case)

Input: Signal $\boldsymbol{y} \in \mathbb{R}^N$, $D \in \mathbb{Z}^+$, $m \in \mathbb{Z}^+$, noise level $\sigma^2 \in \mathbb{R}$, optimized kernel parameter $\omega \in \mathbb{R}$.
Output: D^{th}-order derivative $\boldsymbol{u}_D \in \mathbb{R}^N$.

Initialization: $\boldsymbol{a} \leftarrow [1,\ldots,N+2m]^\intercal$, $\mathbf{1}_{N+2m} \leftarrow [1,\ldots,1]^\intercal$, $\mathbf{P} \leftarrow \boldsymbol{a} \cdot \mathbf{1}_{N+2m}^\intercal$.
$\boldsymbol{y}_E \leftarrow D^{th}$-degree polynomial extension of $\boldsymbol{y}$ via Algorithm 5.
$\boldsymbol{y}_{E0} \leftarrow y_{E[0]} \cdot \mathbf{1}_{N+2m}$.
Construct $\boldsymbol{\phi}_{D[0:N+2m-1]}$ via (5.2).
Construct $\boldsymbol{\Phi}_D$ from $\boldsymbol{\phi}_{D[0:N+2m-1]}$ via (5.4).
Construct $\mathbf{K}$ from $\mathbf{P}$ with selected kernel via (4.17).
$\boldsymbol{u}_{DE} \leftarrow \mathbf{K}\boldsymbol{\Phi}_D^\intercal(\boldsymbol{\Phi}_D \mathbf{K} \boldsymbol{\Phi}_D^\intercal + \sigma^2 \mathbf{I}_{N+2m})^{-1}(\boldsymbol{y}_E - \boldsymbol{y}_{E0})$.
$\boldsymbol{u}_D \leftarrow \boldsymbol{u}_{DE[m+1:m+N]}$
return $\boldsymbol{u}_D$.

5.3 Numerical Experiment

5.3.1 Experiment Setup

Three signals adopted in other's work have been analyzed in this experiment [12, 153], as:

$$g_1(t) = sinc(\pi t) + t + 1, \quad t \in [-3,3],\ \Delta t = 0.012. \tag{5.10}$$

$$g_2(t) = e^{cos(2\pi t)}, \quad t \in [-2,2],\ \Delta t = 0.01. \tag{5.11}$$

$$g_3(t) = \frac{2.25}{4(1+100t)^{-0.4} + 0.45}, \quad t \in [1,30],\ \Delta t = 1. \tag{5.12}$$

Note that (5.12) is in the form of $V_{cc}R_{load}/(R_{sensor} + R_{load})$, where $R_{sensor}(t) = R_0(1 + kt)^{-\beta}$ is the Figaro TGS MOS gas sensor's theoretical function modelling the relation between resistance ($k\Omega$) and the concentration ($mol \cdot dm^{-3}$) [107].

Algorithm 7: Derivative estimation (arbitrary signal)

Input: Signal $\boldsymbol{y} \in \mathbb{R}^N$, $D \in \mathbb{Z}^+$, $m \in \mathbb{Z}^+$, noise level $\sigma^2 \in \mathbb{R}$, optimized kernel parameter $\omega \in \mathbb{R}$.

Output: D^{th}-order derivative $\boldsymbol{u}_D \in \mathbb{R}^N$.

Initialization: $\boldsymbol{a} \leftarrow [1,\ldots,N+2m]^\intercal$, $\mathbf{1}_{N+2m} \leftarrow [1,\ldots,1]^\intercal$, $\mathbf{P} \leftarrow \boldsymbol{a} \cdot \mathbf{1}_{N+2m}^\intercal$, $\hat{\boldsymbol{y}}_{E0} \leftarrow [0,\ldots,0]^\intercal$.

$\boldsymbol{y}_E^{(0)} \leftarrow D^{th}$-degree polynomial extension of $\boldsymbol{y}$ via Algorithm 5.

$\boldsymbol{y}_{E0}^{(0)} \leftarrow y_{E[0]} \cdot \mathbf{1}_{N+2m}$.

for $d = 1,\ldots,D$ **do**

  $\hat{\boldsymbol{y}}_{E0} \leftarrow \hat{\boldsymbol{y}}_{E0} + \frac{1}{(d-1)!}\boldsymbol{y}_{E0}^{(d-1)} \circ (\boldsymbol{a} - \mathbf{1}_{N+2m})^{\circ(d-1)}$.

  Construct $\boldsymbol{\phi}_{d[0:N+2m-1]}$ via (5.2).

  Construct $\mathbf{\Phi}_d$ from $\boldsymbol{\phi}_{d[0:N+2m-1]}$ via (2.11).

  Construct $\mathbf{K}$ from $\mathbf{P}$ with selected kernel via (4.17).

  $\boldsymbol{y}_E^{(d)} \leftarrow \mathbf{K}\mathbf{\Phi}_d^\intercal(\mathbf{\Phi}_d\mathbf{K}\mathbf{\Phi}_d^\intercal + \sigma^2\mathbf{I}_{N+2m})^{-1}(\boldsymbol{y}_E - \hat{\boldsymbol{y}}_{E0})$.

  $\boldsymbol{y}_{E0}^{(d)} \leftarrow y_{E[0]}^{(d)} \cdot \mathbf{1}_{N+2m}$.

end

$\boldsymbol{u}_D \leftarrow \boldsymbol{y}_{E[m+1:m+N]}^{(d)}$

return $\boldsymbol{u}_D$.

We also adopted the same environment setting, i.e. Gaussian white noise with a SNR of 15 dB was added to the ground truth signal $g(t)$ as a noise observation $y(t)$. Different derivative estimation algorithms are compared and evaluated by NRMSE (4.14).

5.3.2 Comparison of Different Numerical Differentiation Methods

The kernel parameters for the proposed method is optimized via (5.8) using the DIRECT algorithm. The parameters for other considered algorithms are selected by minimizing the NRMSE. To avoid boundary distortion, polynomial extensions are also applied to all compared algorithms. Each experiment is repeated for 1000 times. The statistical results are summarized in Table 5.1. All considered methods obtain more accurate result for $g_1(t)$ and $g_1(t)$ than $g_3(t)$ since derivative estimation is more difficult for a sparse set (e.g. $\Delta t = 1$) than for a dense set (e.g. $\Delta t = 0.01$). This limitation is also appeared in our proposed method. Despite this limitation, the proposed method still reaches the lowest errors in all considered functions in comparison with other methods. This result demonstrates that the proposed method is able to calculate more accurate derivatives for noise-contaminated signals.

Table 5.1: Normalized root mean square error in derivative estimation (SNR=15 dB)

Kernel	NRMSE					
	g_1		$g_2(t)$		$g_3(t)$	
	g'_1	g''_1	g'_2	g''_2	g'_3	g''_3
Proposed Method	$\mathbf{0.154 \pm 0.026}$	$\mathbf{0.159 \pm 0.032}$	$\mathbf{0.091 \pm 0.014}$	$\mathbf{0.179 \pm 0.031}$	$\mathbf{0.378 \pm 0.022}$	$\mathbf{0.466 \pm 0.027}$
Wavelet Filter [105]	0.278 ± 0.041	0.468 ± 0.068	0.259 ± 0.283	2.381 ± 7.176	0.851 ± 0.030	1.036 ± 0.039
Kernel FIR Filter [214]	0.198 ± 0.033	0.222 ± 0.041	0.104 ± 0.016	0.226 ± 0.040	0.554 ± 0.026	0.633 ± 0.029
Ensemble Regression [153]	0.432 ± 0.088	0.506 ± 0.091	0.328 ± 0.090	0.599 ± 0.150	0.751 ± 0.054	0.835 ± 0.033
Tikhonov (Category I) [147]	0.265 ± 0.028	0.506 ± 0.043	0.128 ± 0.014	0.255 ± 0.023	0.589 ± 0.058	0.737 ± 0.074
Tikhonov (Category II) [87]	0.193 ± 0.036	0.225 ± 0.045	0.200 ± 0.014	0.357 ± 0.018	0.618 ± 0.080	0.771 ± 0.101

5.4 Field Experiment

5.4.1 Experiment Setup

5.4.1.1 Material Preparation and Instrument Settings

The material preparation and instrument settings are the same as Section 4.4.1. The only difference is the material information, as shown in Table 5.2. In this experiment, we only considered four task whiskies. Another two whiskies are applied as disturbances.

5.4.1.2 Experiment Process

To tackle the 3D issue, the following framework is proposed, (shown in Fig. 5.2), as:

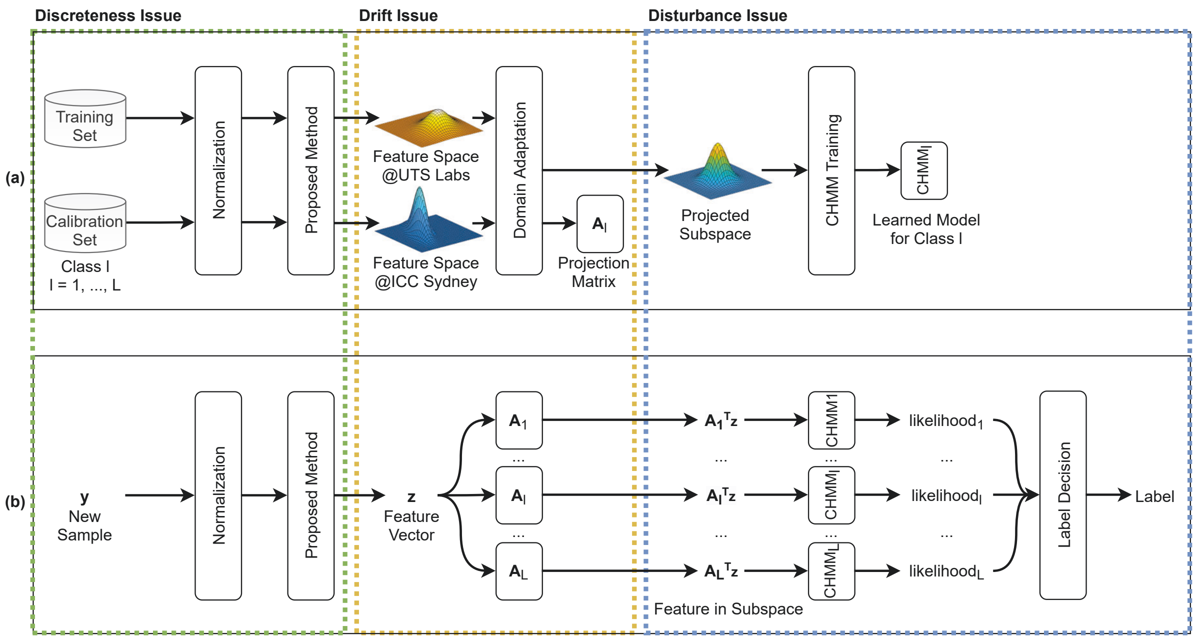


Figure 5.2: Framework of experiment process to tackle the 3D issue. (a) Training procedure. (b) Testing procedure.

- **Discreteness**: Normalization is applied to reduce the discreteness caused by odour concentration variation and sensor manufacturing discrepancy. The proposed method is applied to ensure a stable and accurate transient-state feature extraction in resistance to sensor noises. The extracted transient-state features are illustrated in Fig. 5.3.

Table 5.2: Whisky information (four targets and two disturbances)

Category	Lable	Product Information	Sample Size		
			Training Set	Calibration Set	Testing Set
Target	1	Ardbeg 10 Year Old (Single Malt, ABV 46%)	240	20	40
	2	Macallan 12 Year Old (Single Malt, ABV 43%)	240	20	57
	3	Johnnie Walker Red Label (Blended, ABV 40%)	240	20	62
	4	Chivas Regal 12 Year Old (Blended, ABV 40%)	240	20	41
Disturbance	X	Smokehead (Single Malt, ABV 43%)	-	-	20
		Johnnie Walker Black Label (Blended, ABV 40%)	-	-	22

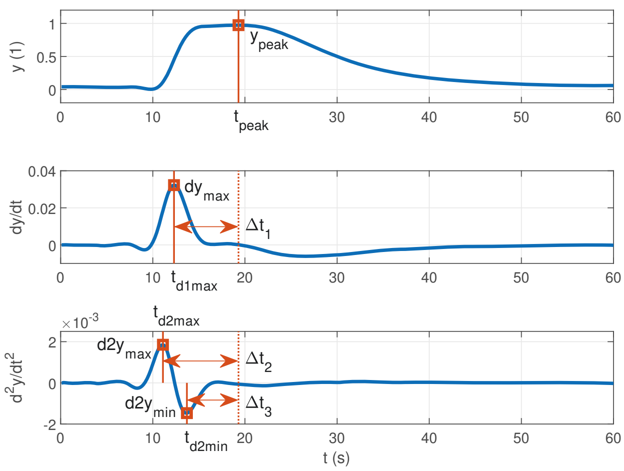


Figure 5.3: Representative example of feature extraction. Six transient-state features are extracted, including: $z_1 = dy_{max}$, $z_2 = d2y_{max}$, $z_3 = d2y_{min}$, $z_4 = \Delta t_1$, $z_5 = \Delta t_2$, and $z_6 = \Delta t_3$.

- **Drift**: Similar to Section 4.4.1, DRCA is adopted to reduce the effect of sensor drift. Considering the subsequent disturbance issue, in this experiment, the projection matrix is learned for every single class.

- **Disturbance**: The disturbance issue refers to the influence of unknown odours on the trained classifier. It is expensive and unrealistic to train models for all possible disturbance odours. In this experiment, we adopt the strategy of one-class classification to tackle this problem. Here, we apply CHMMs as classifiers for our task. We use the dataset obtained after domain adaptation to establish CHMM for each target whisky. The hyper-parameters of CHMM, i.e. the number of hidden states, the number of Gaussian mixtures, and the threshold (likelihood) for one-class classification, can be optimized by maximizing the classification accuracy of the five-fold cross validation (5-CV) on both training and calibration set. In the 5-CV, both the training set and the calibration set are divided into five subsets and composed as five pairs. Four pairs are used for training, and the rest one pair for validation. This procedure is then repeated for five times. When optimized parameters are achieved, all training and calibration samples are used to generate the final classifier for testing. In the testing procedure, each CHMM would assign a likelihood to the test sample. If the likelihood of a sample is not within the threshold of any model, the sample is considered as a disturbance sample. Otherwise, the label of the sample belongs to the model that generates the largest likelihood.

5.4.2 Comparison of Different Feature Extraction Methods

We compare the accuracy of the traditional transient-state features processed by our method and by the standard filtering process in identifying whiskies. Moreover, we also compare other noise-resistant feature extraction algorithms, including the polynomial curve fitting-based features and SAE features. The results are illustrated in Fig. 5.4. Under the proposed framework, we realize the identification of both target whiskies and unknown disturbances. The proposed method improves the quality of transient-state features by increasing the classification accuracy (including the identification of disturbances) from 73.14% to 86.36%. The processed transient-state features even perform better than the learning-based (SAE) method. The result of the field experiment shows the potential of applying the proposed method and framework in the area of e-noses.

5.5 Conclusion

In this chapter, we propose a novel numerical differentiation algorithm using kernel-based Tikhonov regularization. A numerical experiment is conducted to verify the effectiveness of the proposed method. Then, the proposed method is applied to the area of e-noses in estimating 1^{st}- and 2^{nd}-order derivatives from the noise-contaminated gas sensor's reading to ensure an accurate and stable transient-state features extraction. The extracted features are hence no longer sensitive to the discreteness issue. On the basis of accurate feature representation, the subsequent drift and disturbance issue is tackled by applying the strategy of domain adaptation and one-class classification. A field experiment is also conducted to verify this framework. The result outperforms the classification using other considered features by reaching an accuracy of 86.36% in the task of identifying four target and two disturbance whiskies, which shows its strong potential to be employed in the proposed method and the adopted framework for the purposes of odour identification.

In the area of e-noses, recent research points out the similarity between the derivative of the slower but inexpensive MOS-type gas sensors and the signal of the faster but expensive photoionization detector (PID) [152]. Accordingly, the proposed numerical differentiation method would also benefit the research in using the derivative of the

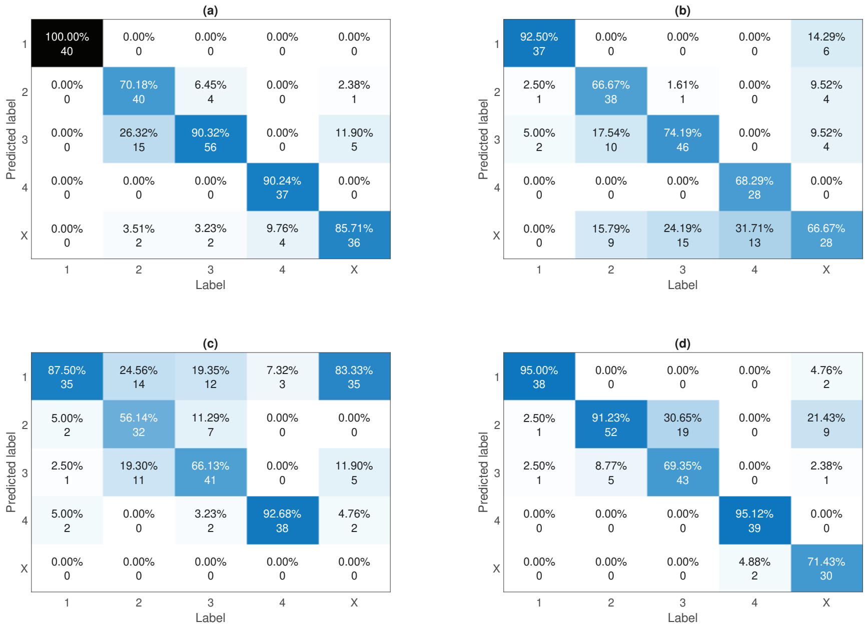


Figure 5.4: Confusion matrix for the field experiment (the testing procedure @ICC Sydney). (a) Traditional transient-state features (proposed method). Accuracy (Acc.): 86.36%; Averaged sensitivity (A. Sen.): 87.29%; Averaged specificity (A. Spe.): 96.50% (b) Traditional transient-state features (wavelet filter). Acc. 73.14%; A. Sen.: 73.66%; A.Spe.: 93.18%. (c) Polynomial curve fitting-based features. Acc.: 60.33%; A. Sen.: 60.49%; A.Spe.: 89.97%. (d) Sparse auto-encoder features. Acc.: 83.47%; A. Sen.: 84.43%; A. Spe.: 95.75%.

MOS sensor's response modulated by periodic flow rate change to achieve rapid olfactory recognition similar to the function of PID sensors. Future improvement of the proposed method would be automatic kernel selection and real time derivative estimation based on the segment-wise regularization framework.

CHAPTER 6

DATA-DRIVEN MEAT FRESHNESS MONITORING AND EVALUATION USING ELECTRONIC NOSE

6.1 Introduction

Meat constitutes a vital part of the human diet since it is a critical source of high-quality protein and other nutrients [37]. Nevertheless, meat undergoes a rapid freshness deterioration process, during which the proliferation of bacteria occurs [72]. Ingesting such contaminated meat, people with a weak immune system, such as the elderly, children, and pregnant women, are prone to foodborne diseases [197]. Besides, meat products are also commercially important in the food industry; freshness plays a crucial role in deciding the market values of this product due to its relatively short product shelf-life. Thus, the demand for a stable and rapid meat spoilage monitoring and evaluation method is becoming increasingly urgent.

Previous studies have demonstrated that the concentration of volatile organic compounds (VOCs), such as fatty acids profiles, aldehydes, and trimethylamine, gradually increases as meat spoilage progresses. [58, 108]. These VOCs take part in the perception of odour, which enables distinguishment of the spoilage process using olfaction-based approaches. Since e-nose was introduced in 1982 [130], much research has been devoted to investigating its efficacy and practicality in food quality control [108, 136, 143]. E-nose offers a non-destructive detection on meat products [58, 136]. Moreover, this approach is more time and cost efficient compared to other methods in the industry [58], including chemical test [86], microbiological analysis [21], spectroscopy [8] and computer vision [113].

A number of approaches have been proposed for e-nose-based food quality control. We classify the approaches in two categories, that is, food quality monitoring and food quality evaluation. Food quality monitoring is a dynamic process that uses monitoring indices to track the changes of the extracted features over time. A significant change in the monitoring indices is indicative of a freshness degradation. Unlike monitoring, food quality evaluation is static. Since the evaluation model is trained exactly once utilizing historical dataset and do not update frequently. Evaluation models directly associate the features with the given labels. These labels that divide the foods into different quality groups are often acquired from two aspects: non-data-driven approaches (i.e. according to the storage time [97], sensory analysis [94] and chemical or microbiological analysis [62, 115]); and data-driven approaches (i.e. according to clustering analysis [103]).

Food quality monitoring algorithms can be further subdivided into unsupervised learning and supervised learning. In terms of unsupervised learning, principal component analysis (PCA) has been widely applied in the monitoring of various food products. The mechanism of this method is to compress the whole life-span samples of monitored food from the high-dimensional feature space to a low-dimensional subspace (i.e., 1- to 3-dimension [4, 29, 38]) and to find the coordinates that maximize variances of these samples. Once new observations are added to the historical dataset, the previously calculated PCA projection matrix needs to be updated. After the whole life-span data has been fed to the model, a final PCA score plot is provided to visualize the monitoring result by tracking the position change of these samples. Alternative methods include kernel PCA (KPCA) [168], which adopts kernel functions to construct non-linear mapping, and independent component analysis (ICA) [9], which maximizes the higher order statistics. In terms of supervised learning, a similar analytic tool is linear discriminant analysis (LDA). It also illustrates the monitoring result by using a 2- or 3-dimensional [29, 119] score plot. Unlike PCA, LDA utilizes the label information. This method reduces the dimension of the original dataset by selecting coordinates that maximize the generalized Rayleigh quotient of the between-classes scatter matrix to the within-class scatter matrix. An alternative method is kernel LDA (KLDA) [168], which is regarded as a non-linear expansion of the standard LDA. Both PCA- and LDA-based methods apply the coordinates change of the tested sample in the transformed feature space as monitoring indices to reveal the quality change of food. However, these monitoring models are usually static models. The major shortcoming of static models is the lack of focus on time dependence [34], which restricts their applications in real-time monitoring.

Food quality evaluation algorithms can be further subdivided into classification and regression. Classification algorithms enable qualitative analysis (e.g., freshness level or spoilage degree evaluation) by assigning the new food sample observation to a set of mutually exclusive quality groups. According to the review papers published recent years [58, 78, 108, 154, 186], existing food quality evaluation methods cover nearly all classical classification algorithms, such as k-nearest neighbours (kNN), naive Bayes classifier (NB), discriminant analysis (DA), logistic regression (LR), decision tree (DT), support vector machines (SVM) and artificial neural network (ANN). Improvements are often made in the area of feature extraction, since discriminant features will facilitate classification accuracy. With the development of deep learning, some deep neural networks, such as auto-encoders and Boltzmann machines, have been applied to reduce the effect of long-term sensor drift. Regression algorithms are able to realize quantitative analysis (e.g., shelf life or storage time prediction) by modelling the relationships among input feature vectors and the object variables. Linear regression is a common method to build the predictive model. Improved algorithms include support vector regression (SVR), which overcomes the presence of non-linearity in the data, as well as principal component regression (PCR) and partial least squares regression (PLSR), which overcome the presence of a large number of highly correlated features in the data.

Although considerable research has been devoted to the advancement of the technology. There are gaps that prevent industry adoption of e-nose to food quality control. The purpose of this chapter is to: i) provide the ability to monitor the meat freshness change regardless the storage condition; ii) provide a data-driven method to grade the meat freshness level rather than biological or chemical tests when the specific storage condition is determined; and iii) improve the stability of the data-driven method. This study applies a powerful statistical tool, hidden Markov model (HMM), to establish the freshness monitoring and evaluation models.

The main contributions of this chapter are as follows:

- To enable monitoring regardless the storage condition, we propose a monitoring method based on a single HMM trained only by fresh samples. Monitoring is realized by measuring the distance (i.e. likelihood) between the fresh sample and the tested sample.

- When the whole life-span data of meat stored in a specific storage condition is available, a meat freshness evaluation method based on multiple HMMs is proposed.

The freshness level is uncovered by a decoding algorithm (i.e. the Viterbi algorithm). Then, HMM for each level is trained and applied in parallel as a classifier to evaluate the freshness.

- To ensure stable and reliable monitoring and evaluation results, we improve the traditional training algorithm of HMM (i.e. the Baum-Welch algorithm) by proposing the segmental rapid centroid estimation (RCE) algorithm.

The rest of the chapter is structured as follows: Section 6.2 is devoted to describing the whole methodology. Experiment setup is introduced in Section 6.3. Then, the experiment results and discussion will be presented in Section 6.4. Finally, the conclusion of this chapter will be outlined in Section 6.5.

6.2 Methodology

6.2.1 Preliminary of HMMs

The hidden Markov model (HMM) is a powerful statistical tool with a double-layered stochastic process: an unobserved Markov process $Q := q_{1:D}$, where the current state is only dependent on its previous state, associates with an observed stochastic process $O := o_{1:D}$, where the current observation only generates from the current hidden state. An HMM $\lambda := \{\pi, \mathbf{A}, \mathbf{B}\}$ can be expressed in joint distribution as (3.9).

There are three fundamental questions of interest with HMM. The solution for the questions are of great use in our proposed data-driven approaches: i) Learning: Given a set of observation sequences $\mathbf{O}$, learn the model λ. This question relates to the establishment of models for meat freshness monitoring or evaluation, which can be solved by the Baum-Welch algorithm; ii) Decoding: Given a trained model λ and one observation sequence O, find the most likely hidden state sequence Q that generated this observation sequence. This question relates to meat freshness level grading, which can be solved by the Viterbi algorithm; iii) Evaluation: Given a trained model λ, and one observation sequence O, calculate the likelihood $P(O|\lambda)$ that the model produces this observation sequence. This question relates to meat freshness evaluation, which can be solved by the Forward algorithm. Details of the algorithms mentioned above can be found in [175].

6.2.2 Segmental RCE-Based Training

Algorithm stability is required in food quality control. However, the Baum-Welch algorithm suffers from the high reliance on a good initial guess of HMM parameters since the convergence is only guaranteed to local optima. The sensitivity to initialization is mainly derived from the estimation to Gaussian mixture density, i.e. $\{\mathbf{C}, \mu, \Sigma\}$. Segmental rapid centroid estimation (RCE)-based training method is therefore proposed in this section to alleviate the influence of estimating Gaussian mixture parameters and to ensure a stable model.

The proposed method is an improvement of the traditional k-means initialization [13]. The mainly difference is the replacement of particle swarm optimization-based clustering (PSC) algorithm to the traditional k-means algorithm. PSC is more stable and can obtain a more globally solution. However, PSC will become computationally demanding when the dimension of data is high. RCE is a simplified version of PSC, which introduced a new update strategy. Our previous works have demonstrated that RCE can greatly reduce the computational complexity [203, 204]. The proposed training algorithm separates the observation sequences into several segments, and then, estimates Gaussian mixture parameters within each segment. The algorithm is assumed convergence when the distance score between two re-estimated models reaches the threshold. The final model can be regarded as the learning result. A simple implementation of segmental RCE-based training is presented in Algorithm 8. The pseudo-code of RCE is shown in Algorithm 9.

6.2.3 HMM-Based Meat Freshness Monitoring

As mentioned in the previous section, conventional statistical models such as PCA or LDA are not suitable in dealing with real-time monitoring. Therefore, an HMM-based model is proposed in this section to monitor the dynamic process.

However, in the real situation, a large number of factors may cause the meat to spoil, such as bacteria, mould and oxidative rancidity. Different factors will generate different spoilage patterns. It is impossible for a pattern library to include every interested spoilage pattern, and it is also difficult to obtain enough samples to train those patterns. To avoid the problem over complicated, we can simplify it by only identifying the fresh meat.

The monitoring model is trained by segmental RCE algorithm only using the feature vectors of fresh samples, $\mathbf{O}_0 = [o_0^{(1)}, o_0^{(2)}, ..., o_0^{(K)}]^\intercal$. We use subscripts 0 to denote the fresh meat sample. $o_0^{(K)} \in \mathbb{R}^{1\times dim}$ represents the k^{th} fresh meat sample. Then, the incoming

Algorithm 8: Segmental RCE-based training

Input: Observations $\mathbf{O} \in \mathbb{R}^{K \times dim}$, state number $N \in \mathbb{Z}^+$, and mixture number $M \in \mathbb{Z}^+$.

Output: Trained model $\lambda = \{\pi, \mathbf{A}, \mathbf{C}, \mu, \Sigma\}$.

Initialization: Initialise λ randomly or according to requirements, iter $\leftarrow 0$.
repeat
    Partition $\mathbf{O}$ into N segments as $\{\mathbf{O}_1, \mathbf{O}_2, ..., \mathbf{O}_N\}$ according to the optimal path of λ calculated via the Viterbi algorithm.
    for $j = 1, 2, \ldots, N$ **do**
        if $|\mathbf{O}_j| < M$, *where* $\mathbf{O}_j = \{o_{j,1}, ..., o_{j,T_j}\}$. **then**
            $\forall m, \hat{c}_{jm} \leftarrow \frac{1}{M}$.
            $\forall m, \hat{\mu}_{jm} \leftarrow \frac{\sum_{o_{j,t} \in O_j} o_{j,t}}{|\mathbf{O}_j|}$.
            $\forall m, \hat{\Sigma}_{jm} \leftarrow$ Identity matrix.
        else
            Cluster $\mathbf{O}_j$ into M clusters via Algorithm 9.
            $\forall m, \hat{c}_{jm} \leftarrow \frac{\text{The number of } o_{j,t} \text{ grouped in cluster } m}{|\mathbf{O}_j|}$.
            $\forall m, \hat{\mu}_{jm} \leftarrow$ the m^{th} centroid of $\mathbf{O}_j$.
            $\forall m, \hat{\Sigma}_{jm} \leftarrow cov(\text{All } o_{j,t} \text{ grouped in cluster } m)$.
        end
    end
    $\hat{\lambda} \leftarrow \{\pi, \mathbf{A}, \hat{\mathbf{C}}, \hat{\mu}, \hat{\Sigma}\}$.
    $\lambda \leftarrow$ re-estimate $\hat{\lambda}$ via the Baum-Welch Algorithm.
    iter $\leftarrow$ iter + 1.
until iter reaches maximum iteration;
return Trained model $\lambda = \{\pi, \mathbf{A}, \mathbf{C}, \mu, \Sigma\}$.

sample, $o_t, t > 0$, is evaluated by the model via the forward algorithm. The monitoring index (i.e. the output of the forward algorithm) is log-likelihood, which denotes the probability of the observation generated by the given model. The monitoring index will maintain at a similar value if the incoming sample is still fresh but will drop to a very low value when the sample has spoiled. The monitoring is achieved by tracking the change of likelihood over time. The flowchart of the proposed monitoring method is illustrated in Fig. 6.1.

6.2.4 HMM-Based Meat Freshness Grading and Evaluation

When the whole life-span data obtained under a specific storage condition is determined, an interesting question is how to grading its freshness level. In addition to expert

Algorithm 9: RCE algorithm

Input: Observations $\mathbf{O} \in \mathbb{R}^{K \times dim}$, cluster number $M \in \mathbb{Z}^+$.
Output: Locally optimum centroids $\mu_{\text{Best}} \in \mathbb{R}^{M \times dim}$.

Initialization: Initialise swarm: M particles, where $\mu \leftarrow \text{rand}(\Omega), \mathbf{c} \leftarrow \mathbf{0}, \mathbf{s} \leftarrow \mathbf{0}, \mu^0_{\text{Best}} \leftarrow \mu, \triangle\mu^0 \leftarrow \mathbf{0}$. ▷ Ω: searching space
iter $\leftarrow 1$; stag $\leftarrow 0$. ▷ stag: count for stagnation
repeat
  Calculate pairwise Euclidean distance between each O_k and each μ_m.
  $\forall m$, $c_m \leftarrow$ The closest position, μ_m, of the particle m in relation to the observation O_k.
  $\forall k$, $s_k \leftarrow$ The position a particle that has been closest to the observation O_k.
  $\mu^{\text{iter}}_{\text{Best}} \leftarrow \underset{\{\mu^{\text{iter}-1}_{\text{Best}}, \mu\}}{\text{argmin}}\{f(\mu^{\text{iter}-1}_{\text{Best}}), f(\mu)\}$. ▷ f: cost function
  $\text{stag} = \begin{cases} \text{stag}+1 & f(\mu^{\text{iter}}_{\text{Best}}) - f(\mu^{\text{iter}-1}_{\text{Best}}) > -\epsilon \\ 0 & \text{others} \end{cases}$.
  $\mu_{\text{win}} \leftarrow$ Position of a particle which constituted cluster contains the most observations.
  Assign each observation to its corresponding cluster $\mathbf{O}_m$.
  for $m = 1, 2, \ldots, M$ **do**
    if $|\mathbf{O}_m|>0$. **then**
      $C_m \leftarrow c_m - \mu_m$. ▷ Cognitive term
      $S_m \leftarrow \frac{\sum_{\forall O_k \in \mathbf{O}_m}(s_k - \mu_m)}{|\mathbf{O}_m|}$. ▷ Social term
      $R_m \leftarrow \frac{\sum_{\forall O_k \in \mathbf{O}_m}(O_k - \mu_m)}{|\mathbf{O}_m|}$. ▷ Self-organising term
      $\triangle\mu^{\text{iter}}_m \leftarrow \omega \cdot \triangle\mu^{\text{iter}-1}_m + \phi_1 \circ C_m + \phi_2 \circ S_m + \phi_3 \circ R_m$.
      $\mu_m \leftarrow \mu_m + \triangle\mu^{\text{iter}}_m$.
    else
      $\mu_m \leftarrow \mu_m + \phi_4 \circ (\mu_{\text{win}} - \mu_m)$. ▷ $\phi_{1,\ldots,4}$: uniform random vector
    end
  end
  $\omega \leftarrow 0.95 \times \omega$. ▷ ω: inertia weight
  iter $\leftarrow$ iter $+ 1$.
until iter reaches maximum iteration *or* stag reaches threshold;
return Locally optimum centroids μ_{Best}.

experience and chemical or biological analysis, clustering algorithms can be applied to solve this problem. However, conventional clustering algorithms, such as k-means, PSC and GMM, lack the focus on the dependence of sample on time. Therefore, a meat freshness grading method is proposed based on the Viterbi algorithm in this section.

Meat spoilage is an irreversible process, where the freshness level degrades progressively with storage time and cannot change back. The freshness level can be regarded as

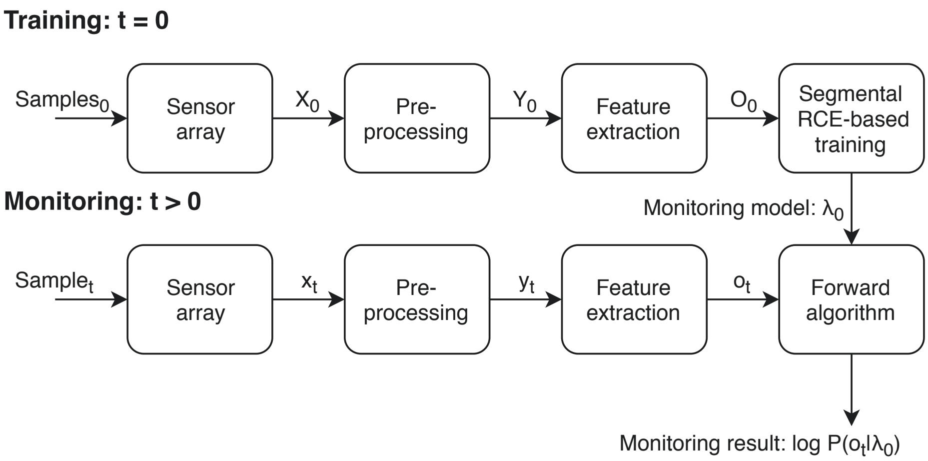


Figure 6.1: Flowchart of HMM-based meat freshness monitoring. Sensor array's responses of meat samples are first pre-processed before extracting features. In the training procedure, only the feature vectors extracted from fresh meat samples (i.e. $t = 0$) can be sent to segmental RCE algorithm to learn the monitoring model, λ_0. In the monitoring procedure, the monitoring result can be achieved by evaluating the feature vector of an incoming sample (i.e. $t > 0$) via the forward algorithm.

a hidden state since it usually hides beneath the observable feature vectors extracted from the sensor array's responses. Therefore, we can use the Viterbi algorithm to uncover the exact freshness level from the observations. In order to model this dynamic process, we select the no-jump left-right HMMs, whose structure is illustrated in Fig. 6.2.

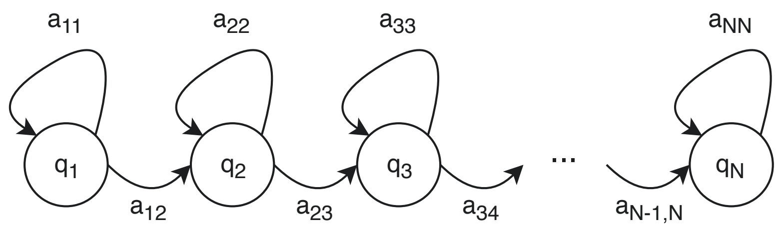


Figure 6.2: Representative example of a N-state no-jump left-right HMM. In such HMMs, the current state is restricted to transit only to itself and the following state. Jumps of more than one state are not allowed.

To build the freshness grading model, the whole life-span data should be arranged in chronological order as $\mathbf{O} = [o_1^\intercal, o_2^\intercal, ..., o_T^\intercal]$, where $o_t \in \mathbb{R}^{1\times dim}$ is the feature vector of the sample obtained at time t. The initial state vector of the no-jump left-right HMM is fixed as $\pi = [1, 0, ..., 0]^\intercal \in \mathbb{R}^N$ during the training procedure since the spoilage process usually starts at the first freshness level. The initial guess of state transition matrix, A, is as:

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 & \cdots & 0 \\ 0 & 0.5 & 0.5 & \cdots & 0 \\ 0 & 0 & 0.5 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{N\times N} \tag{6.1}$$

After training, the Viterbi algorithm is applied to finding a hidden state sequence that is most likely to generate $\mathbf{O}$. The optimal hidden sequence, Q, is regarded as the grading result. The hidden states number, N, is representative of the number of freshness level, which can be determined by the one that maximize the BIC. The flowchart of the proposed monitoring method is illustrated in Fig. 6.3.

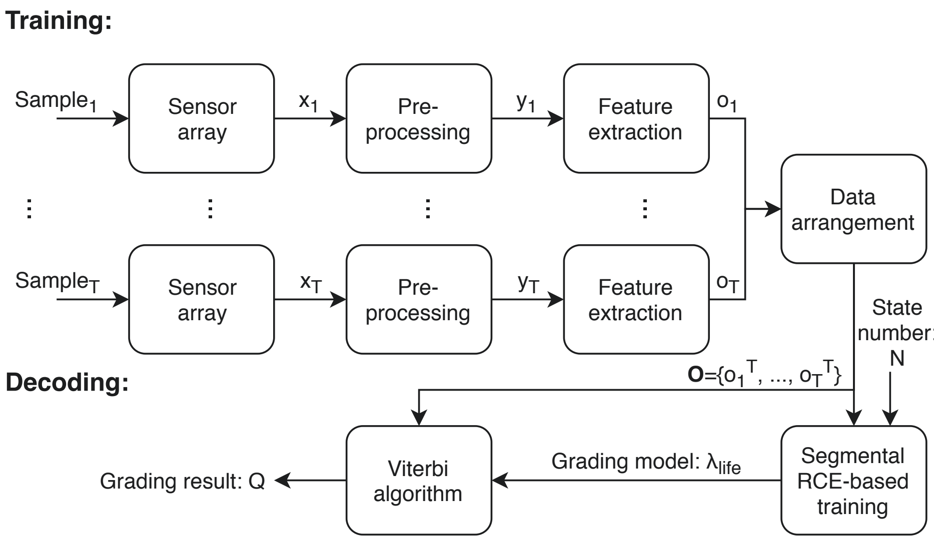


Figure 6.3: Flowchart of HMM-based meat freshness grading. Sensor array's responses of the whole life-span meat samples are first pre-processed before extracting features. Feature vectors are arranged from the first test to the last test. In the training procedure, the arranged dataset is sent to segmental RCE algorithm to learn a stable grading model, λ_{life}. In the decoding procedure, the most probable state path is calculated according to λ_{life} via the Viterbi algorithm to illustrate the grading result.

Then, the pattern of each freshness level can be trained using the data from its corresponding cluster. We can still apply HMMs as the classification tool. In freshness evaluation, the type of HMMs is not limited to no-jump left-right HMM. The valuation procedure can be achieved by assigning a meat sample to a specific freshness level whose model obtains the highest log-likelihood via the Forward algorithms. The flowchart of evaluation procedure is illustrated in Fig. 6.4.

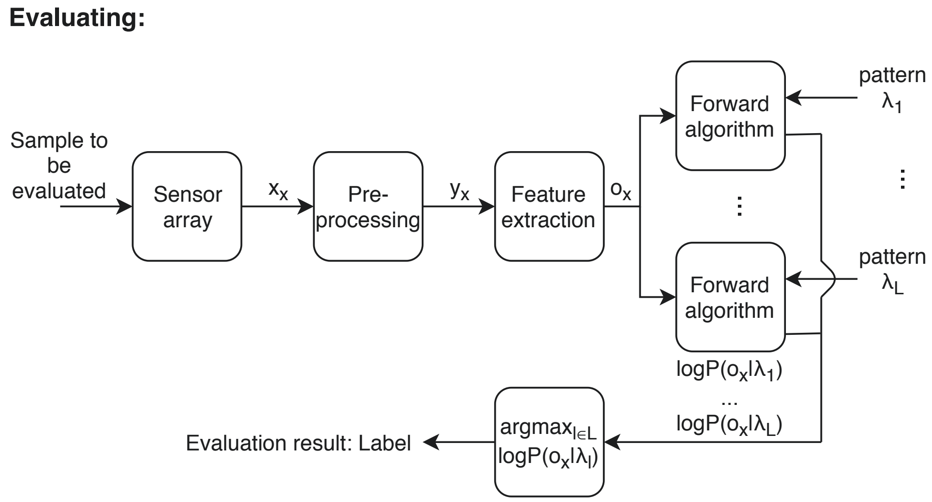


Figure 6.4: Flowchart of HMM-based meat freshness evaluation. The pattern of each freshness level is trained based on its corresponding samples. In evaluating procedure, log-likelihood of the sample to be evaluated is evaluated by all trained models. Its specific freshness level is determined by the model that achieves the highest log-likelihood.

6.3 Field Experiment Setup

6.3.1 Material Preparation

Three different materials were measured in this study: i) three salmon fillets (skinned and boned); ii) three packs of beef mince and iii) three packs of chicken mince (hereinafter referred to as fish, beef and chicken). All materials were purchased from a local supermarket in Sydney. Salmon fillets were sold in a polystyrene box covered with ice. The salmon was processed at a commercial seafood processing plant and delivered to the supermarket within one day after slaughter. We purchased the three salmon fillets on the same day. Beef mince and chicken mince were sold in modified atmosphere packages. The three packs of each material were purchased on the same day from the same brand produced by the same supplier with the same labelled expiration of four days.

In the preparation of fish samples, we cut each salmon fillet into forty small cubes (weighing approx. 4g each), known as group A, B and C. Then, each small cube was placed in a 10-ml-volume headspace crimped top vial. After placing the sample, the headspace crimped top vial was sealed with its cap immediately and stored at 4°C (refrigerator temperature). In the preparation of beef samples, the 184g of each pack of beef mince was divided equally into forty-six parts (weighing approx. 4g each). Each part was then placed in a 10-ml-volume headspace crimped top vial and sealed with the cap. Forty samples of each pack were stored at 4°C, known as group A, B and C, and the rest six of each pack were stored at 25°C (room temperature), known as group A′, B′ and C′. The sample preparation procedure for chicken mince was the same as the preparation for beef samples.

For the groups that stored at 4°C, four samples of that group were measured as fresh samples on the same day of purchase. And for the remaining thirty-six samples, we measured one sample every 8 hours. For the groups that stored at 25°C, we only measured one sample every 8 hours. Each measurement was repeated for three times.

6.3.2 Instrument Settings

The emitted odour profiles of the tested meat samples were measured by a self-designed e-nose system, known as UTS NOS.E. This system consists of an AC to DC power transformer, an air-intake system controlled by a control module, a sealed sensor chamber to place gas sensors, an A/D converter to convert sensor's reading to digital signal and a Wi-Fi module to communicate with a Pad-based panel. The measurement circuit of the e-nose can be found in our previous work [106]. The schematic diagram of experiment setup and the control logic for the NOS.E system is illustrated in Fig. 6.5(a) and 6.5(b), respectively. During the measurement, the ambient temperature and humidity were well-regulated at 25°C and 50% RH, respectively. Seven different Figaro metal oxide-based sensors were installed in a sealed sensors chamber. The sensor array was selected according to some widely applied biomarkers used in meat and fish freshness evaluation, including nitrogen and sulphur volatile compounds, 1-butanal, trimethyl amine, ethanol, *etc*. [108]. The sensitive characteristics of the gas sensors are shown in Table 6.1.

6.3.3 Data Pre-Processing and Feature Extraction

Traditional features were adopted to represent the response curve in this study. Before extracting features, we first applied the fractional conductance (2.2) to align the baseline.

Then, the proposed numerical differentiation algorithm is applied to ensure a stable and accurate hand-crafted feature extraction. The extracted features include (see Ta-

Table 6.1: Gas sensors used for meat freshness monitoring and evaluation

No.	Sensor Name	Sensitivity Characteristics*
1	TGS2600	Hydrogen, Ethanol, Hydrocarbons, ect.
2	TGS2602	Hydrogen sulfide, Ethanol, Ammonia, ect.
3	TGS2603	Trimethyl amine, Methyl mercaptan, Hydrogen sulfide, ect.
4	TGS2610_D	Hydrocarbons, Hydrogen, Ethanol, ect.
5	TGS2611_E	Methane, Hydrogen, Iso-butane, ect.
6	TGS2612	Propane, Iso-butane, Methane, ect.
7	TGS2620	Ethanol, Hydrogen, CO, Methane, ect.

* The response of sensors is non-specific. They are also sensitive to other gases, which are not list in this table.

ble 2.1): f_1, f_2, f_3, f_4, f_6, f_7, f_9, f_{10}, and the integral from $t_{\text{gas-on}}$ to t_{peak}. The dimension of the feature vector to represent the sensor array's response is 63.

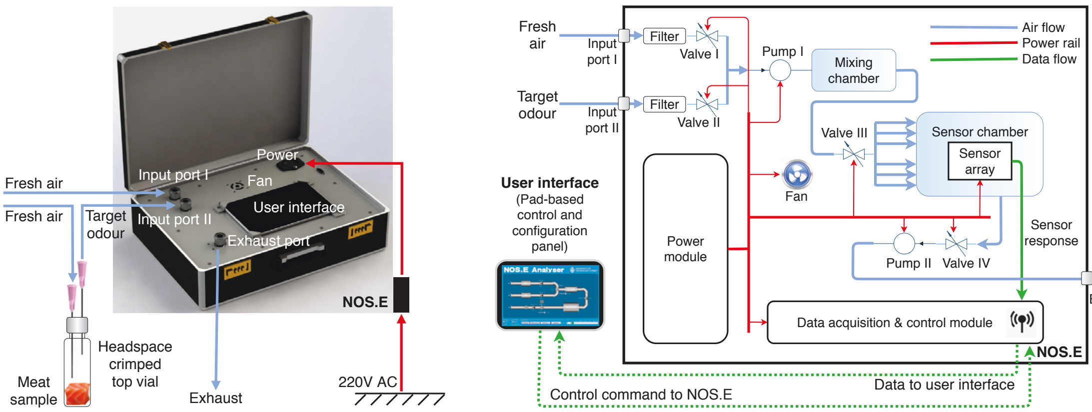
6.4 Results and Discussion

6.4.1 Meat Freshness Monitoring

6.4.1.1 Results of HMM-Based Monitoring

The performance of single HMM-based monitoring algorithm is discussed in this section. For each type of meat, we used only fresh samples from group A, B, and C to train the monitoring models separately, and then applied each model to track all three groups of freshness changes. Specifically, for beef and chicken, we also tested the performance of the trained model for monitoring samples under different storage conditions. The log-likelihood changes of fish, beef and chicken with storage time are illustrated in boxplots, Fig. 6.6(a) to 6.6(c), respectively.

For the meat stored at 4°C, the boxplots above show a downward trend in the log-likelihood change, indicating the degradation process of meat freshness. We roughly divide the monitoring curve into three phases. In the first phase, the change in the monitoring curve is usually not noticeable. The monitoring index is also close to zero, indicating that the meat stored in the refrigerator remained fresh during the first few days. As the meat spoilage process develops, the concentration of emitted VOCs is increasing. The response of the Figaro TGS sensor to the concentration increase is basically consistent with an exponential model (4.15). The resistance of the sensor decreases at a



(a) Schematic diagram of experiment setup

(b) Control logic for the NOS.E system

Figure 6.5: (a) Experiment setup. The odour of the tested meat sample was pumped to the sensor chamber through the input port II by sampling needle. After one measurement, fresh air was pumped to clean odour molecules from the sensor array. Meanwhile, the sensors' responses were sent to and stored in a tablet PC (Pad). (b) Control logic of the NOS.E system. Connect the meat sample to input port II. At the baseline setting stage, open Valve I, III, IV and Pumps I, II and keep the flow rate at 1.1L/min for 30 seconds letting the fresh air to wash the sensor chamber. The baseline of each sensor can be obtained. At the testing stage, keep Valve III, IV and the Pump I, II open, and control the flow rate at 1.1L/min. Then, close Valve I while opening Valve II. This state would last for 90 seconds, during which the sensor's response would increase and reach the steady value. At the recovery stage, keeping the state of Valve III, IV and the Pump I, II unchanged, close Valve II while opening Valve I. This state would last for 300 seconds, during which the sensor's response would return to its baseline. After the measurement, remove the meat sample from input port II. Open all valves and pumps, and set the flow rate at 2.2L/min letting the fresh air clean the residual odour molecules in the gas circuit for 80 seconds. Meanwhile, the sensor's response were converted into digital signal at a 1-Hz sampling rate and sent to the Pad via the Wi-Fi module.

high rate at low ppm concentrations. Thus, in the second phase, the monitoring index falls obviously with storage time, which makes the monitoring algorithm capable of detecting the early spoilage. According to the exponential model, the resistance of the sensor will reduce the rate of decrease at high ppm concentrations. This explains why the monitoring index fluctuated around a low value for a long time in the final phase. We believe the fluctuation is caused by the vary of humidity inside the top vials since other external influences had been well-regulated during the whole experiment. Tissue exudation increased during the spoilage process and had been kept by the sealed top vial, leading to the internal humidity change. Despite the large fluctuation in the final phase of spoilage, the proposed method is still applicable in detecting the early freshness

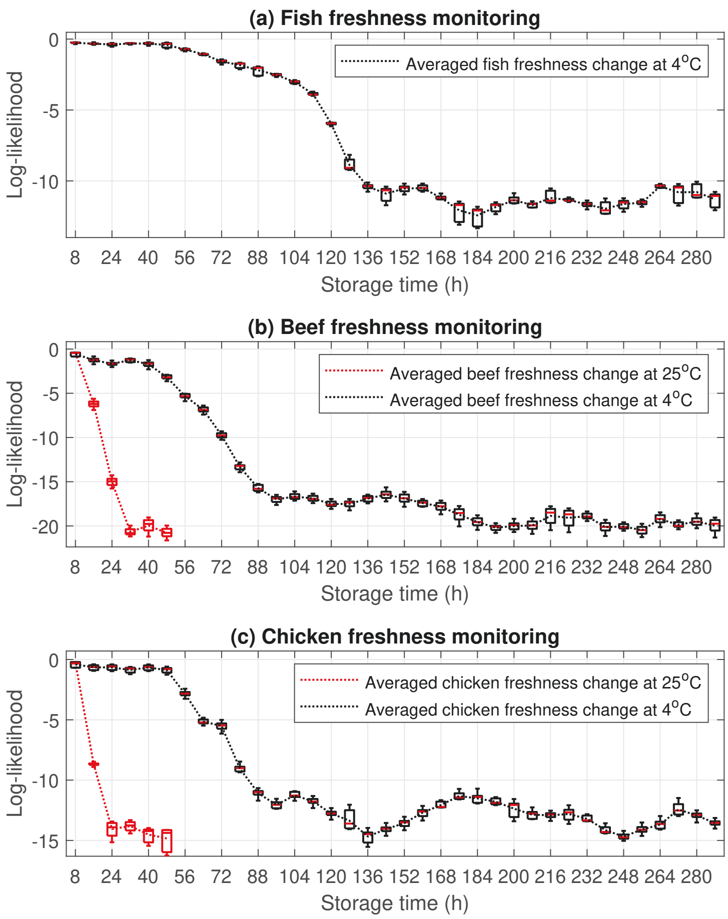


Figure 6.6: Boxplots of HMM-based meat freshness monitoring. (a) Fish monitoring result. During the first 48 hours of storage at 4°C, log-likelihood values remained at a very similar level around -0.52. The monitoring index began to decrease gradually after 56 hours, and then dropped sharply after 112 hours. After 136 hours of storage, the log-likelihood values began to fluctuate from -13.45 to -10.12 until the end of the experiment. (b) Beef monitoring result. For samples stored at 4°C, the log-likelihood value decreased slightly from 0 to -2.31 during the first 40 hours of storage. Then, it dropped obviously to an average of -17.28 after 96 hours, and eventually fluctuated from -18.01 to -21.64. For samples stored at 25°C, the log-likelihood values dropped directly to the final fluctuation range of the counterparts stored in the refrigerator, within 36 hours of storage. (c) Chicken monitoring result. For samples stored at 4°C, the log-likelihood values remained at around -0.70 during the first 48 hours of storage. After 88 hours of storage, it falls rapidly to an average of -11.30. And then, the log-likelihood value fluctuated from -15.71 to -11.62 until the end of the experiment. For samples stored at 25°C, the log-likelihood value dropped directly to the final fluctuation range of the counterpart stored in the refrigerator, within 24 hours of storage.

change of meat.

6.4.1.2 Comparison with PCA-based monitoring

The proposed algorithm enables real-time monitoring. To verify this, we illustrate this process and compare it with PCA-based monitoring algorithm in Fig 6.7. We take the fish freshness monitoring as an example.

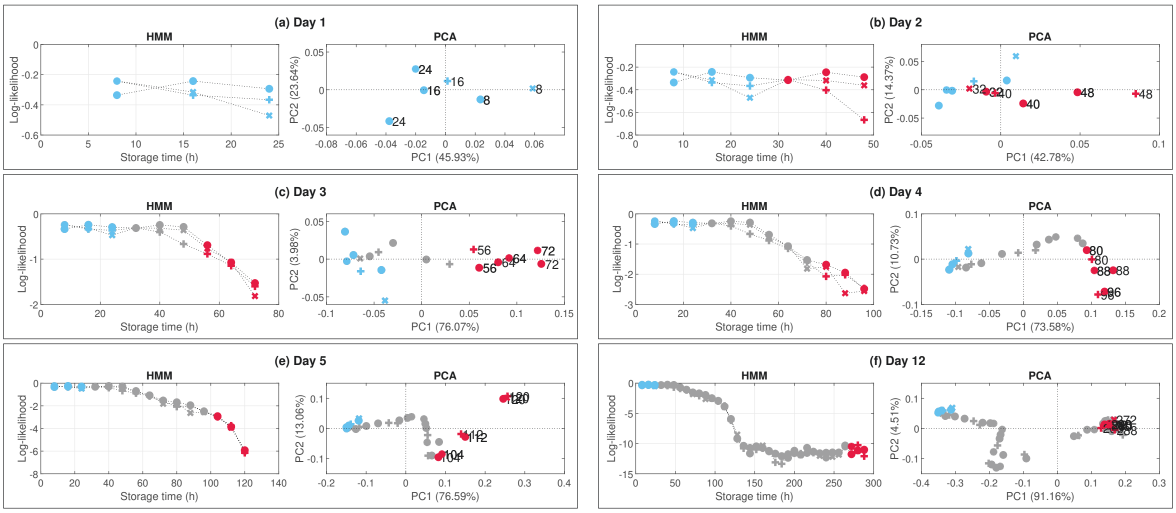


Figure 6.7: Representative example of real-time monitoring. This figure compares HMM- and PCA-based monitoring algorithms in tracking the real-time freshness change of the salmon stored at 4°C. (a) to (e) are the results of monitoring obtained for five consecutive days. Subplot (f) shows the result of the last day. Blue symbols denote the samples obtained from the first day. Red represents the newly obtained data, and gray the historical data. Symbol "•", "×" and "+" represent data in group A, B and C, respectively. The number (i.e.: $8, 16, 24, \ldots, 288$) in PCA score plots represents the storage time.

The spoilage trend also appears obviously in the PCA score plot: moving from left to right along the PC1 axis. However, for the PCA-based approach, the previously calculated projection matrix should be updated as new samples are added to the historical dataset. The coordinates of the PCA score plot will also vary with the model, which is not conducive to establishing a warning line for separating edible and spoiled meat. For the proposed algorithm, the monitoring model only needs to be trained once. The assessment of the incoming food sample does not change the original model and the coordinates. Because the evaluation result is a single metric, we can adopt a threshold to instead the warning line. Thus, the proposed method is more applicable in tracking the real-time freshness change of meat.

6.4.2 Meat Freshness Grading

6.4.2.1 Results of HMM-Based Decoding

Meat spoiled rapidly at 25°C, resulting in insufficient data for clustering analysis. Therefore, in this section, we focused on the samples stored at 4°C. As mentioned in the previous section, the BIC index can be used to determine the optimal number of hidden states, N (a.k.a. the number of freshness level). The BIC formula used in this study is according to Pelleg and Moore's work [128]:

$$\mathrm{BIC}(N) = \log l(\mathbf{O}|\lambda;N) - \frac{K_N}{2}\log(|\mathbf{O}|) \tag{6.2}$$

where $l(\mathbf{O}|\lambda;N)$ denotes the likelihood of the observation set, $\mathbf{O}$, given the model, λ, parameterized by N, and $|\mathbf{O}|$ the number of samples. K_N represents the number of free parameters.

The optimal N is the one that obtains the highest value of the BIC. The selection result is shown in Fig. 6.8. Then for each type of meat, the grading model, λ_{life}, is trained based on the whole life-span data according to the optimal number of hidden states. The hidden path, Q, is decoded via the Viterbi algorithm and illustrates in Fig. 6.9.

The no-jump and left to right assumptions restricted the state transited only to itself and the next state. Recall that the monitoring model was trained only by the fresh samples, while the grading model used the whole life-span samples for training. But, compared with Fig. 6.6, we can find that the decoding results (Fig. 6.9) is roughly consistent with the monitoring results. Samples with a similar rate of change are decoded in the same hidden state. The decoding results further illustrate the effectiveness of the monitoring method.

6.4.2.2 Stability Analysis

In this section, the stability of the proposed meat freshness grading algorithm is discussed and compared with other considered methods.

We mainly compared two types of methods: decoding and clustering. As to the former, the instability is mainly derived from the training of a HMM. Initial guess of a Gaussian mixture density will lead to different local optima. Thus, we compared the proposed

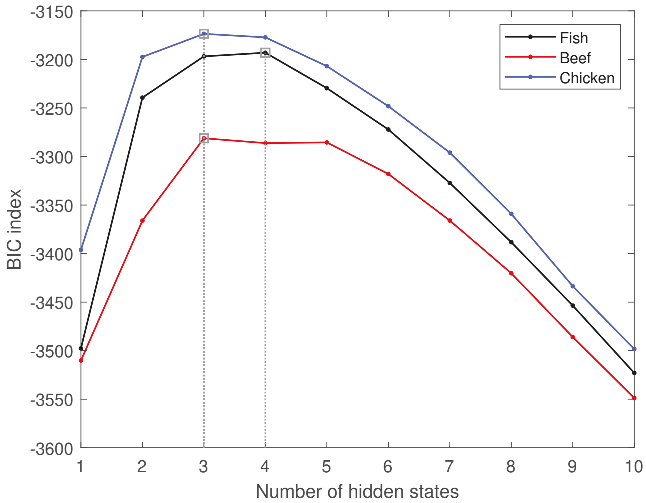


Figure 6.8: BIC index for selection of the number of hidden states. The selection results are: Fish, $N = 4$; Beef, $N = 3$; Chicken, $N = 3$.

training algorithm with other HMM training methods, including the classical Baum-Welch algorithm and the training with k-means^{++} initialization. As to the latter, the instability is derived from the guess of the initial centroids and from the high dimension of the input data. Thus, we also compared k-means clustering, k-means^{++} clustering, RCE clustering as well as their combination with PCA-based feature reduction method.

To test the stability, we can run the considered algorithm with random initialization for D times and compare the results. The difference between the two decoding/clustering solutions can be measured by calculating their normalized Hamming distance, $\frac{1}{|L|}||L_i - L_j||_{\mathrm{H}}$, between the two label vectors, L_i and L_j. For a set of solutions, we use the mean and standard deviation of pairwise normalized Hamming distance, $\{\frac{1}{|L|}||L_i - L_j||_{\mathrm{H}} \mid \forall i \neq j \text{ and } i, j \leq D\}$, to measure the stability of the algorithm. This measure reflects the diversity of the solutions, hence, the smaller the mean and standard deviation values of the pairwise normalized Hamming distance, the more stable the algorithm is.

Note that for the clustering algorithm, even if the centroids calculated from the two tests are the same, their label vectors are not always the same. The clustering algorithm lacks the focus on the dependence of the samples over time. Therefore, we sorted the clustering result based on the average storage time of the samples in the same cluster and then re-labelled the cluster. After this operation, clustering result with the same centroids will generate the same label vector. We ran each considered algorithm for 100 times for each of three test groups and calculate the overall stability measure. We

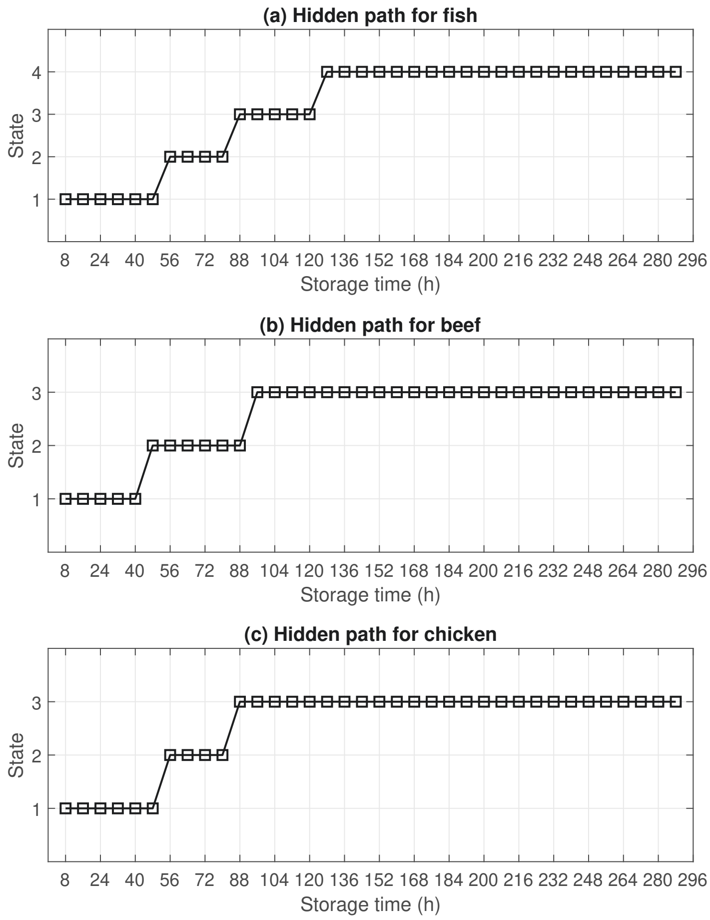


Figure 6.9: Representative example of HMM-based freshness grading. (a) Hidden path for fish. Fish samples stored at 4°C from 8 to 48 hours were generated by State I, from 56 to 80 hours by State II and from 88 to 120 hours by State III. The rest samples were generated by State IV. (b) Hidden path for beef. Beef samples stored at 4°C from 8 to 40 hours were generated by State I, from 48 to 88 hours by State II. The rest samples were generated by State III. (c) Hidden path for chicken. Chicken samples stored at 4°C from 8 to 48 hours were generated by State I, from 56 to 80 hours by State II. The rest samples were generated by State III.

eventually got 300 solutions for one algorithm. The size of the set of pairwise normalized Hamming distance is $\binom{300}{2}$. The comparison result is shown in Table 6.2.

According to the statistical results in Table 6.2, segmental RCE greatly increases the stability of decoding results to $1.2\% \pm 1.4\%$, $1.2\% \pm 1.4\%$ and $0\% \pm 0\%$ for fish, beef and chicken experiment respectively, which outperforms other considered methods. Especially for chicken freshness grading, the decoding results is the same for all three groups. For the fish and the beef freshness grading, the diversity of solutions is attributed to the between group differences. As to the fish experiment, the difference was derived from group C: the sample stored for 48 hours was decoded as State II while the one with same storage time was decoded as State I in group A and B. As to the beef experiment, the difference was also from group C: the sample stored for 88 hours was decoded as State III, while its counterparts were still in State II. For k-means initialization-based algorithm, the diversity of solutions is attributed to both between group differences and within group differences. Therefore, the proposed training algorithm ensures to decode a stable hidden path. In terms of clustering methods, RCE performs best in comparing with the traditional k-means and k-means^{++}. The stability can be further improved by using a PCA-based feature reduction method.

Moreover, to verify whether the decoding performance is consistent with the clustering, we selected one of the 100 repeated tests whose clustering result minimized the cost function and illustrated it in Fig. 6.10. Compared with Fig. 6.9, clustering results for fish and beef were the same as the decoding results. For chicken, the sample stored for 56 hours was classified in Cluster I, while it was graded as State II in decoding. But the decoding result achieved a smaller cost function than clustering, which means that the decoding method can obtain a more global solution. Therefore, the proposed method is applicable in grading the freshness level of meat.

6.4.3 Meat Freshness Evaluation

In this section, the performance of the proposed meat freshness evaluation method is validated and compared with other algorithms, including different HMMs, and some other classifiers mentioned previously, such as kNNs, softmax regression and SVMs.

Data from two of the three experimental groups was used to train the freshness evaluation model, which was then used to test whether the samples with the same freshness levels in the rest one group could be correctly identified. The validation procedure was repeated for three times, which ensured each experimental group was used for training only once. This operation performs like three-fold cross validation (3CV).

Table 6.2: Comparison result of algorithm stability

Category	Method	Normalized Hamming distance (mean% ± std%)		
		Fish	Beef	Chicken
Decoding	Baum-Welch algorithm	$64\% \pm 33\%$	$51\% \pm 35\%$	$56\% \pm 37\%$
	k-means^{++} initialization + Baum-Welch algorithm	$6.0\% \pm 4.5\%$	$1.3\% \pm 1.4\%$	$3.4\% \pm 3.8\%$
	Proposed method	$\mathbf{1.2}\% \pm \mathbf{1.4}\%$	$\mathbf{1.2}\% \pm \mathbf{1.4}\%$	$\mathbf{0}\% \pm \mathbf{0}\%$
Clustering	K-means clustering	$43\% \pm 23\%$	$28\% \pm 23\%$	$32\% \pm 25\%$
	K-means^{++} clustering	$31\% \pm 25\%$	$23\% \pm 21\%$	$27\% \pm 22\%$
	RCE clustering	$17\% \pm 15\%$	$16\% \pm 15\%$	$18\% \pm 20\%$
	PCA* + K-means^{++} clustering	$11\% \pm 8.5\%$	$20\% \pm 18\%$	$24\% \pm 21\%$
	PCA + RCE clustering	$6.3\% \pm 11\%$	$1.3\% \pm 3.4\%$	$0.70\% \pm 2.5\%$

* The number of principle components for fish, beef and chicken samples are 2, 3, and 3, respectively.

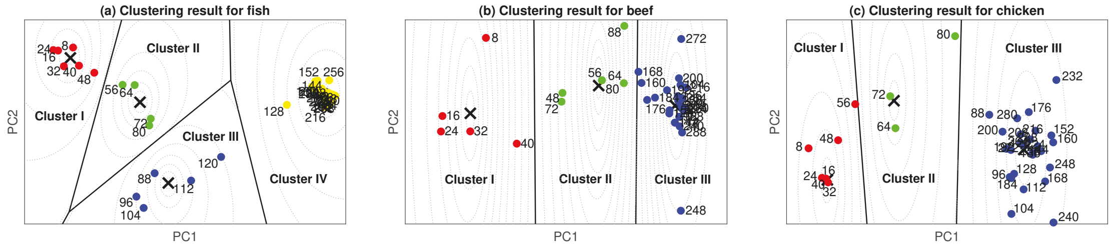


Figure 6.10: Voronoi diagram of meat freshness clustering based on PCA + RCE. (a) Clustering result for fish. Fish samples stored at 4°C from 8 to 48 hours were classified in Cluster I, from 56 to 80 hours in Cluster II and from 88 to 120 hours in Cluster III. The rest samples were classified in Cluster IV. (b) Clustering result for beef. Beef samples stored at 4°C from 8 to 40 hours were classified in Cluster I, from 48 to 88 hours in Cluster II. The rest samples were classified in Cluster III. (c) Clustering result for chicken. Chicken samples stored at 4°C from 8 to 56 hours were classified in Cluster I, from 64 to 80 hours in Cluster II. The rest samples were clustered in Cluster III.

We ran the considered algorithm for 100 times and then calculated the sensitivity (sen.) (i.e. TP/P, where P and TP denote the number of positive samples and true positive samples, respectively) and specificity (spe.) (i.e. TN/N, where N and TN denote the number of negative samples and true negative samples, respectively) for each freshness level. The sample size of the last freshness level is usually larger than that of the previous freshness levels. To measure the general performance of a considered classifier, we also adopted the weighted averaged sensitivity (WA sen.) and the weighted averaged specificity (WA spe.).

Hyper-parameters of each considered algorithm was optimized using the genetic algorithm (GA) by maximizing the accuracy (i.e. $(TP+TN)/(P+N)$). Recall that the type of HMM is not limited to no-jump left-right HMM in meat freshness evaluation. Hence, the hyper-parameters for HMM-based algorithms include the type of state transition matrix, hidden state number N and Gaussian mixture number M. We only select K closest training samples for KNN. softmax regression is a non-parametric algorithm. Hyper-parameters for SVMs include the regularization parameter C, kernel type and kernel parameters.

The comparison results among classifiers for fish, beef and chicken freshness evaluation are shown in Table 6.3 to 6.5, respectively. Standard variances for algorithms as kNNs, Softmax regression and SVMs are caused by between group differences, while the standard variances for HMM-based algorithms come from both the local optima and the between group differences. Meat freshness classification is a non-linear separable task.

Table 6.3: Comparison result of fish freshness evaluation (mean% ± std%)

Method	Freshness level I		Freshness level II		Freshness level III		Freshness level IV		General performance	
	sen.	spe.	sen.	spe.	sen.	spe.	sen.	spe.	WA sen.	WA spe.
Propose method	**94.44**%	**98.89**%	**93.33**%	**98.92**%	**100**%	**100**%	96.83%	95.56%	**96.32**%	**99.07**%
	±**8.13**%	±**1.63**%	±**9.76**%	±**1.57**%	±**0**%	±**0**%	±2.32%	±3.25%	±**2.83**%	±**0.69**%
k-means^{++} initialization + HMMs	**94.44**%	**98.89**%	**93.33**%	**98.92**%	**100**%	**100**%	87.30%	82.22%	95.63%	98.11%
	±**8.13**%	±**1.63**%	±**9.76**%	±**1.57**%	±**0**%	±**0**%	±2.32%	±3.25%	±2.85%	±0.75%
HMMs	88.89%	97.78%	92.00%	98.71%	**100**%	**100**%	88.25%	83.55%	93.89%	97.86%
	±8.13%	±1.63%	±12.65%	±2.04%	±**0**%	±**0**%	±3.54%	±4.95%	±3.07%	±0.83%
KNNs	**94.44**%	**98.89**%	61.66%	94.72%	66.67%	94.62%	93.65%	91.11%	74.48%	95.60%
	±**8.13**%	±**1.63**%	±10.63%	±1.58%	±9.76%	±1.57%	±2.32%	±3.25%	±3.36%	±0.74%
Softmax regression	66.67%	93.33%	38.33%	91.57%	33.33%	89.25%	95.24%	93.33%	49.16%	91.62%
	±**28.17**%	±5.63%	±10.63%	±1.60%	±9.76%	±1.57%	±4.02%	±5.63%	±9.46%	±1.74%
SVMs	88.89%	97.78%	**93.33**%	**98.92**%	93.33%	98.92%	**98.41**%	**97.78**%	92.98%	98.62%
	±8.13%	±1.63%	±**9.76**%	±**1.57**%	±9.76%	±1.57%	±**2.32**%	±**3.25**%	±3.91%	±0.49%

Table 6.4: Comparison result of beef freshness evaluation (mean% ± std%)

Method	Freshness level I		Freshness level II		Freshness level III		General performance	
	sen.	spe.	sen.	spe.	sen.	spe.	WA sen.	WA spe.
Proposed method	**100**%	**100**%	**97.78**%	**99.56**%	**100**%	**100**%	**99.09**%	**99.82**%
	±**0**%	±**0**%	±**5.86**%	±**1.17**%	±**0**%	±**0**%	±**2.40**%	±**0.48**%
k-means^{++} initial	**100**%	**100**%	96.30%	99.26%	92.10%	81.21%	97.73%	97.92%
-ization + HMMs	±**0**%	±**0**%	±7.35%	±1.47%	±2.01%	±4.64%	±3.04%	±0.76%
HMMs	**100**%	**100**%	94.44%	98.89%	90.77%	78.18%	96.84%	97.47%
	±**0**%	±**0**%	±8.13%	±1.63%	±2.03%	±4.07%	±3.55%	±1.11%
KNNs	**100**%	**100**%	83.33%	96.67%	98.72%	96.67%	93.06%	98.34%
	±**0**%	±**0**%	±14.43%	±2.89%	±1.92%	±5.00%	±5.77%	±0.83%
Softmax regression	86.67%	97.85%	57.78%	92.33%	97.44%	93.33%	74.99%	95.06%
	±10.00%	±1.61%	±13.33%	±1.51%	±3.85%	±10.00%	±3.27%	±0.92%
SVMs	**100**%	**100**%	94.44%	98.89%	**100**%	**100**%	97.72%	99.54%
	±**0**%	±**0**%	±8.33%	±1.67%	±**0**%	±**0**%	±3.42%	±0.68%

Table 6.5: Comparison result of chicken freshness evaluation (mean% ± std%)

Method	Freshness level I		Freshness level II		Freshness level III		General performance	
	sen.	spe.	sen.	spe.	sen.	spe.	WA sen.	WA spe.
Proposed method	**100**% ±**0**%	**100**% ±**0**%	**100**% ±**0**%	**100**% ±**0**%	92.31% ±3.25%	80.00% ±8.45%	**99.35**% ±**0.27**%	**98.31**% ±**0.71**%
k-means^{++} initial -ization + HMMs	97.78% ±5.86%	99.56% ±1.17%	**100**% ±**0**%	**100**% ±**0**%	92.05% ±6.25%	79.33% ±16.24%	98.51% ±2.20%	98.09% ±1.43%
HMMs	77.78% ±11.52%	95.56% ±4.30%	**100**% ±**0**%	**100**% ±**0**%	91.28% ±7.48%	77.33% ±19.44%	91.12% ±7.34%	96.46% ±0.84%
KNNs	88.89% ±16.67%	97.78% ±3.33%	66.67% ±12.50%	95.83% ±1.56%	**100**% ±**0**%	**100**% ±**0**%	77.62% ±6.52%	96.90% ±1.09%
Softmax regression	94.44% ±8.33%	98.89% ±1.67%	58.33% ±12.50%	94.79% ±1.56%	94.87% ±5.09%	86.67% ±13.23%	74.65% ±9.03%	95.61% ±2.02%
SVMs	94.44% ±8.13%	98.89% ±1.63%	**100**% ±**0**%	**100**% ±**0**%	97.44% ±3.75%	93.33% ±9.76%	97.75% ±2.83%	99.03% ±0.74%

It can be revealed by the worst performance of the softmax regression, a linear classifier. Regarding the HMM-based algorithms, the proposed method shows superior to the k-means^{++} initialization-based HMM. It can be reflected that the segmental estimation procedure can improve the overall stability of the HMM optimization algorithm, and can help to achieve a more "global" solutions. Compared with the proposed algorithm, SVM shows superior in correctly identifying the samples from the last freshness level. This may caused by the unbalanced dataset. The number of training samples from the last freshness level is larger than that from the previous freshness levels, and SVM tends to correctly classify those samples to reduce the error of the cost function. Regarding the general performance, the proposed method achieves highest WA sensitivity and WA specificity among all considered algorithms. Accordingly, the proposed method can be applied in e-nose to make an accurate freshness evaluation.

6.5 Conclusion

This work introduced a powerful statistical tool, hidden Markov models, to improve the performance of e-nose in the area of food quality control. Three different uses of HMM were discussed: i) When the spoilage pattern of one meat was unknown, a single HMM method was proposed to track the freshness change of the tested meat. Since the single model was trained only using the fresh samples, it could be used to monitor the meat spoilage process under different storage temperatures. ii) When a specific spoilage pattern was known, a decoding method was proposed to automatically grade the freshness levels that the tested meat had undergone through its whole life-span. With the help of the proposed segmental RCE-based training algorithm, the stability of the decoding result was improved, which was demonstrated by the comparison experiments. iii) When the freshness levels were obtained, meat freshness evaluation could be achieved by classifying the test sample into a set of HMMs trained for each freshness level. The experiment result showed that the sensitivity and specificity for fish, beef and chicken samples reached 96.32%±2.83% and 99.07%±0.69%, 99.09%±2.40% and 99.82%±0.48%, as well as, 99.35%±0.27% and 98.31%±0.71%, respectively.

A preliminary attempt to apply HMMs achieved an acceptable result. Future research along this work will be towards exploring the following aspects: i) Warning line establishment. To make the HMM-based monitoring method more applicable for users, a warning line (threshold) is required to separate the edible and spoiled meat. How to establish this threshold is a further study of the proposed method. ii) Storage time prediction.

The proposed freshness grading and evaluation algorithm only provides a qualitative evaluation (i.e. the classification result). A quantitative evaluation is also required in some situation. A time-variant HMM, which enables the storage time prediction, can be established according to the decoding results. However, its hidden state duration obeys the geometric distribution. Whether this distribution is suitable to model the meat spoilage process needs further study. iii) More application scenarios. The proposed framework can be used not only for food freshness monitoring but also for other scenarios of food quality control, such as food fermentation process monitoring and fruit ripeness monitoring. For example, in fruit ripeness monitoring, we can train a HMM for ripe fruit and then track the similarity change of the tested fruit.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

This thesis aims to improve the performance of MOS-type gas sensor array-based electronic noses by using stable and discriminant feature representations. In order to achieve the goal, we conducted our research from two aspects: i) Design novel feature extraction algorithm to provide stable feature representations; and ii) Design novel signal processing algorithm to ensure a stable feature extraction using existed algorithms.

For the first aspect, in Chapter 3, we proposed a novel feature extraction algorithm using kernel-based non-parametric modelling. A stable spline kernel is applied in the regularization term, which improves the stability in model parameter estimation. Moreover, a hyper-parameter optimization algorithm is proposed to prevent the model from overfitting. The feature extracted is a series of impulse responses containing discriminant information of different odours. Considering the characteristic of the proposed features (i.e. impulse responses is actually a time series), we applied a time series classification tool, the Gaussian mixture density hidden Markov model (GMM-HMM). It should be noted that the stable feature representations obtained by the proposed method are calculated based on an accurate setting of the initial time. In order to find the initial time from the noise-contaminated signal, we also proposed a changing time searching algorithm. However, the searching of the initial time is time-consuming.

Aiming to the shortage of the modelling-based features, an improvement method is proposed in Chapter 4. The success of kernel methods highly relies on the capacity of kernel function. Here, a multiscale wavelet kernel derived from the orthogonal projection operator is applied as the regularization matrix in solving the inverse problem – estimating impulse response (i.e. the features for target odours) from noisy sensor's response. The use of the multiscale wavelet kernel equips the proposed method with ability in resistance to random noise. Moreover, the proposed method no more relies on the accurate setting of the initial time, due to the flexibility of the kernel function.

For the second aspect, in Chapter 5, we proposed a novel numerical differentiation algorithm using kernel-based Tikhonov regularization for the extraction of traditional transient-state features. The proposed method is able to directly estimate accurate high-order derivatives from the noise contaminated signal without applying any smoothing technique. The proposed method greatly improved the stability of the traditional hand-crafted feature extraction.

As to the 3D issue encountered in real applications, the field experiment in Chapter 3 only considered the discreteness issue, since all training and testing samples of six target perfumes were collected under the same environmental conditions. The proposed feature extraction algorithm is able to alleviate the influence of the discreteness issue. Under such conditions, the classification accuracy reached 93.56%.

The field experiment in Chapter 4 considered both discreetness issue and drift issue, since the training and testing samples of four target whiskies were collected under two different environments. This problem was solved under a "feature extraction – domain adaptation" framework by reaching an accuracy of 92.00%.

The field experiment in Chapter 5 considered the whole 3D issue (i.e. discreteness issue, drift issue, and disturbance issue). The training and testing samples of four target whiskies were collected under two different environments. Moreover, we added two new whiskies as disturbance samples. This problem was solved under a "feature extraction – domain adaptation – one-class classification" framework by reaching an accuracy of 86.36%.

The task in Chapter 6 is to monitor and evaluate the freshness level of meat products. In the field experiment, we only considered the discreteness issue and the disturbance issue, since all samples were collected under the same environmental conditions. We proposed a monitoring method and an evaluation method based on single and multiple HMMs, respectively. The proposed methods successfully evaluated the freshness level of three different meat samples, all obtaining a sensitivity and specificity higher than 95.00%.

7.2 Limitations

Thesis proposes a framework of "feature extraction – domain adaptation – one-class classification" to tackle the 3D issue. However, it should be noted that the 3D issue itself is only a simplification of the problems encountered in real applications. The real problem is more complex. For example, domain adaptation requires an assumption that temperature, humidity, and background smell in each environment are relatively constant. However, this assumption cannot be strictly guaranteed in all application scenarios (for example, in an open environment where wind influence must be considered).

7.3 Future Work

Interesting directions for future work are listed as follows.

In Chapter 3 and Chapter 4, we proposed feature extraction algorithms using kernel-based non-parametric modelling. The performance of these methods highly depends on the choice of kernel function as well as the optimization of kernel hyper-parameters. In these two chapters, we only give the method to optimize the kernel hyper-parameters. Therefore, the future improvement of the proposed methods will be automatic kernel selection.

In Chapter 5, we proposed numerical differentiation algorithm using kernel-based Tikhonov regularization. The weakness of this method is the computational cost in constructing the kernel matrix when the observation sequence is very long. This is why we introduce element-wise operations (matrix operations) in kernels. Moreover, since the kernel matrix is independent of the observation, we can calculate it off-line. In order to reduce the computation cost, our future work will focus on improvement through segmented derivative estimation and real-time derivative estimation.

In Chapter 6, we proposed a monitoring algorithm based on a single HMM. The proposed method has not considered the influence of the drift issue. To solve this problem, we can realize the domain adaptation using one reference odour. Then, all samples can be projected to the feature space calibrated by the reference odour. Finally, monitoring can be realized in the projected space.

Bibliography

[1] B. W. Ache and J. M. Young, *Olfaction: diverse species, conserved principles*, Neuron, 48 (2005), pp. 417–430.

[2] M. J. Anderson, J. G. Sullivan, T. K. Horiuchi, S. B. Fuller, and T. L. Daniel, *A bio-hybrid odor-guided autonomous palm-sized air vehicle*, Bioinspiration & Biomimetics, 16 (2020), p. 026002.

[3] M. J. Anderson, J. G. Sullivan, J. L. Talley, K. M. Brink, S. B. Fuller, and T. L. Daniel, *The "smellicopter," a bio-hybrid odor localizing nano air vehicle*, in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 6077–6082.

[4] I. Apetrei and C. Apetrei, *Application of voltammetric e-tongue for the detection of ammonia and putrescine in beef products*, Sensors and Actuators B: Chemical, 234 (2016), pp. 371–379.

[5] R. Arsat, M. Breedon, M. Shafiei, P. Spizziri, S. Gilje, R. Kaner, K. Kalantar-zadeh, and W. Wlodarski, *Graphene-like nano-sheets for surface acoustic wave gas sensor applications*, Chemical Physics Letters, 467 (2009), pp. 344–347.

[6] A. Asimakopoulos, D. Del Fabbro, R. Miano, M. Santonico, R. Capuano, G. Pennazza, A. D'amico, and E. Finazzi-Agro, *Prostate cancer diagnosis through electronic nose in the urine headspace setting: a pilot study*, Prostate cancer and prostatic diseases, 17 (2014), pp. 206–211.

[7] M. Awad and R. Khanna, *Support vector regression*, in Efficient learning machines, Springer, 2015, pp. 67–80.

[8] J. M. Balage, S. d. L. e Silva, C. A. Gomide, M. de Nadai Bonin, and A. C. Figueira, *Predicting pork quality using vis/nir spectroscopy*, Meat science, 108 (2015), pp. 37–43.

[9] S. Balasubramanian, S. Panigrahi, C. Logue, C. Doetkott, M. Marchello, and J. Sherwood, *Independent component analysis-processed electronic nose data for predicting salmonella typhimurium populations in contaminated beef*, Food Control, 19 (2008), pp. 236–246.

[10] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication*, Springer Science & Business Media, 2012.

[11] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains*, The Annals of Mathematical Statistics, 41 (1970), pp. 164–171.

[12] F. S. Bazán and L. Bedin, *Filtered spectral differentiation method for numerical differentiation of periodic functions with application to heat flux estimation*, Computational and Applied Mathematics, 38 (2019), p. 165.

[13] O. Ben-Harush, O. Ben-Harush, I. Lapidot, and H. Guterman, *Initialization of iterative-based speaker diarization systems for telephone conversations*, IEEE Transactions on Audio, Speech, and Language Processing, 20 (2012), pp. 414–425.

[14] J. L. Blanco, J. G. Monroy, A. Lilienthal, and J. Gonzalez-Jimenez, *A kalman filter based approach to probabilistic gas distribution mapping*, in Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013, pp. 217–222.

[15] L. Blanken and T. Oomen, *Kernel-based identification of non-causal systems with application to inverse model control*, Automatica, 114 (2020), p. 108830.

[16] N. Bos and P. d'Ettorre, *Recognition of social identity in ants*, Frontiers in psychology, 3 (2012), p. 83.

[17] J. Burgués, J. M. Jiménez-Soto, and S. Marco, *Estimation of the limit of detection in semiconductor gas sensors through linearized calibration models*, Analytica chimica acta, 1013 (2018), pp. 13–25.

[18] C. Bushdid, M. O. Magnasco, L. B. Vosshall, and A. Keller, *Humans can discriminate more than 1 trillion olfactory stimuli*, Science, 343 (2014), pp. 1370–1372.

[19] J. M. B. Calvo, S. B. i Badia, H. T. Simó, and P. F. Verschure, *The real-world localization and classification of multiple odours using a biologically based neurorobotics approach*, in The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, 2010, pp. 1–7.

[20] L. Capelli and S. Sironi, *Monitoring odour emissions from an oil & gas plant: Electronic nose performance testing in the field*, in 2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), IEEE, 2017, pp. 1–3.

[21] A. Casaburi, P. Piombino, G.-J. Nychas, F. Villani, and D. Ercolini, *Bacterial populations and the volatilome associated to meat spoilage*, Food microbiology, 45 (2015), pp. 83–102.

[22] T. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, *System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques*, IEEE Transactions on Automatic Control, 59 (2014), pp. 2933–2945.

[23] T. Chen and L. Ljung, *Implementation of algorithms for tuning parameters in regularized least squares problems in system identification*, Automatica, 49 (2013), pp. 2213 – 2220.

[24] F. Chraim, Y. B. Erol, and K. Pister, *Wireless gas leak detection and localization*, IEEE Transactions on Industrial Informatics, 12 (2015), pp. 768–779.

[25] M. Clapham, O. T. Nevin, A. D. Ramsey, and F. Rosell, *Scent-marking investment and motor patterns are affected by the age and sex of wild brown bears*, Animal Behaviour, 94 (2014), pp. 107–116.

[26] W. S. Cleveland and S. J. Devlin, *Locally weighted regression: an approach to regression analysis by local fitting*, Journal of the American statistical association, 83 (1988), pp. 596–610.

[27] R. De Luca and D. Botelho, *The unconscious perception of smells as a driver of consumer responses: A framework integrating the emotion-cognition approach to scent marketing*, AMS Review, (2019), pp. 1–17.

[28] R. De Vries, M. Muller, V. Van Der Noort, W. Theelen, R. Schouten, K. Hummelink, S. Muller, M. Wolf-Lansdorf, J. Dagelet,

K. MONKHORST, ET AL., *Prediction of response to anti-pd-1 therapy in patients with non-small-cell lung cancer by electronic nose analysis of exhaled breath*, Annals of Oncology, 30 (2019), pp. 1660–1666.

[29] F. DENG, W. CHEN, J. WANG, AND Z. WEI, *Fabrication of a sensor array based on quartz crystal microbalance and the application in egg shelf life evaluation*, Sensors and Actuators B: Chemical, 265 (2018), pp. 394–402.

[30] L. DESHMUKH, M. KASBE, T. MUJAWAR, S. MULE, AND A. SHALIGRAM, *A wireless electronic nose (wen) for the detection and classification of fruits: A case study*, in 2016 International Symposium on Electronics and Smart Devices (ISESD), IEEE, 2016, pp. 174–178.

[31] G. DESTEFANIS, A. BRUGIAPAGLIA, M. T. BARGE, AND E. DAL MOLIN, *Relationship between beef consumer tenderness perception and warner–bratzler shear force*, Meat science, 78 (2008), pp. 153–156.

[32] B. DING, J. KIM, Y. MIYAZAKI, AND S. SHIRATORI, *Electrospun nanofibrous membranes coated quartz crystal microbalance as gas sensor for nh3 detection*, Sensors and Actuators B: Chemical, 101 (2004), pp. 373–380.

[33] C. DISTANTE, M. LEO, P. SICILIANO, AND K. C. PERSAUD, *On the study of feature extraction methods for an electronic nose*, Sensors and Actuators B: Chemical, 87 (2002), pp. 274–288.

[34] Y. DONG AND S. J. QIN, *A novel dynamic pca algorithm for dynamic data modeling and process monitoring*, Journal of Process Control, 67 (2018), pp. 1–11.

[35] S. DRAGONIERI, J. T. ANNEMA, R. SCHOT, M. P. VAN DER SCHEE, A. SPANEVELLO, P. CARRATÚ, O. RESTA, K. F. RABE, AND P. J. STERK, *An electronic nose in the discrimination of patients with non-small cell lung cancer and copd*, Lung cancer, 64 (2009), pp. 166–170.

[36] T. EKLÖV, P. MÅRTENSSON, AND I. LUNDSTRÖM, *Enhanced selectivity of mosfet gas sensors by systematical analysis of transient parameters*, Analytica Chimica Acta, 353 (1997), pp. 291–300.

[37] C. EKMEKCIOGLU, P. WALLNER, M. KUNDI, U. WEISZ, W. HAAS, AND H.-P. HUTTER, *Red meat, diseases, and healthy alternatives: A critical review*, Critical reviews in food science and nutrition, 58 (2018), pp. 247–261.

[38] N. El Barbri, E. Llobet, N. El Bari, X. Correig, and B. Bouchikhi, *Electronic nose based on metal oxide semiconductor sensors as an alternative technique for the spoilage classification of red meat*, Sensors, 8 (2008), pp. 142–156.

[39] F. P. Elortondo, M. Ojeda, M. Albisu, J. Salmerón, I. Etayo, and M. Molina, *Food quality certification: An approach for the development of accredited sensory evaluation methods*, Food quality and Preference, 18 (2007), pp. 425–439.

[40] E. Esposito, S. De Vito, M. Salvato, V. Bright, R. L. Jones, and O. Popoola, *Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems*, Sensors and Actuators B: Chemical, 231 (2016), pp. 701–713.

[41] H. Fan, V. H. Bennetts, E. Schaffernicht, and A. J. Lilienthal, *A cluster analysis approach based on exploiting density peaks for gas discrimination with electronic noses in open environments*, Sensors and Actuators B: Chemical, 259 (2018), pp. 183–203.

[42] J. A. Farrell, S. Pang, and W. Li, *Plume mapping via hidden markov methods*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 33 (2003), pp. 850–863.

[43] R. Fend, A. H. Kolk, C. Bessant, P. Buijtels, P. R. Klatser, and A. C. Woodman, *Prospects for clinical application of electronic-nose technology to early detection of mycobacterium tuberculosis in culture and sputum*, Journal of Clinical Microbiology, 44 (2006), pp. 2039–2045.

[44] T. Feng, H. Zhuang, R. Ye, Z. Jin, X. Xu, and Z. Xie, *Analysis of volatile compounds of mesona blumes gum/rice extrudates via gc–ms and electronic nose*, Sensors and Actuators B: Chemical, 160 (2011), pp. 964–973.

[45] I. Ferain, C. A. Colinge, and J.-P. Colinge, *Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors*, Nature, 479 (2011), pp. 310–316.

[46] P. Finamore, C. Pedone, S. Scarlata, A. Di Paolo, S. Grasso, M. Santonico, G. Pennazza, and R. A. Incalzi, *Validation of exhaled volatile*

organic compounds analysis using electronic nose as index of copd severity, International journal of chronic obstructive pulmonary disease, 13 (2018), p. 1441.

[47] A. Fort, M. Gregorkiewitz, N. Machetti, S. Rocchi, B. Serrano, L. Tondi, N. Ulivieri, V. Vignoli, G. Faglia, and E. Comini, *Selectivity enhancement of sno2 sensors by means of operating temperature modulation*, Thin Solid Films, 418 (2002), pp. 2–8.

[48] M. Fulford, J. Walker, M. Martin, and P. Marsh, *Total viable counts, atp, and endotoxin levels as potential markers of microbial contamination of dental unit water systems*, British dental journal, 196 (2004), pp. 157–159.

[49] C. Gadenne, R. B. Barrozo, and S. Anton, *Plasticity in insect olfaction: to smell or not to smell?*, Annual review of entomology, 61 (2016), pp. 317–333.

[50] J. C. R. Gamboa, A. J. da Silva, L. L. de Andrade Lima, T. A. Ferreira, et al., *Wine quality rapid detection using a compact electronic nose system: Application focused on spoilage thresholds by acetic acid*, Lwt, 108 (2019), pp. 377–384.

[51] P. J. García-Nieto, E. Garcia-Gonzalo, F. S. Lasheras, J. A. Fernández, and C. D. Muñiz, *A hybrid de optimized wavelet kernel svr-based technique for algal atypical proliferation forecast in la barca reservoir: A case study*, Journal of Computational and Applied Mathematics, 366 (2020), p. 112417.

[52] J. W. Gardner and P. N. Bartlett, *A brief history of electronic noses*, Sensors and Actuators B: Chemical, 18 (1994), pp. 210–211.

[53] A. Gepperth and B. Hammer, *Incremental learning algorithms and applications*, in European symposium on artificial neural networks (ESANN), 2016.

[54] M. Ghasemi-Varnamkhasti, M. Tohidi, P. Mishra, and Z. Izadi, *Temperature modulation of electronic nose combined with multi-class support vector machine classification for identifying export caraway cultivars*, Postharvest Biology and Technology, 138 (2018), pp. 134–139.

[55] A. H. Gomez, J. Wang, G. Hu, and A. G. Pereira, *Discrimination of storage shelf-life for mandarin by electronic nose technique*, LWT-Food Science and Technology, 40 (2007), pp. 681–689.

[56] ——, *Monitoring storage shelf life of tomato using electronic nose technique*, Journal of Food Engineering, 85 (2008), pp. 625–631.

[57] L. A. Gonder-Frederick, J. H. Grabman, J. A. Shepard, A. V. Tripathi, D. M. Ducar, and Z. R. McElgunn, *Variability of diabetes alert dog accuracy in a real-world setting*, Journal of diabetes science and technology, 11 (2017), pp. 714–719.

[58] E. Górska-Horczyczak, D. Guzek, Z. Molęda, I. Wojtasik-Kalinowska, M. Brodowska, and A. Wierzbicka, *Applications of electronic noses in meat analysis*, Food Science and Technology, 36 (2016), pp. 389–395.

[59] L. M. Gosling and S. C. Roberts, *Scent-marking by male mammals: cheat-proof signals to competitors and mates*, Advances in the Study of Behavior, 30 (2001), pp. 169–217.

[60] D. Grandjean, R. Sarkis, C. Lecoq-Julien, A. Benard, V. Roger, E. Levesque, E. Bernes-Luciani, B. Maestracci, P. Morvan, E. Gully, et al., *Can the detection dog alert on covid-19 positive persons by sniffing axillary sweat samples? a proof-of-concept study*, PloS one, 15 (2020), p. e0243122.

[61] C. Guest, M. Pinder, M. Doggett, C. Squires, M. Affara, B. Kandeh, S. Dewhirst, S. V. Morant, U. D'Alessandro, J. G. Logan, et al., *Trained dogs identify people with malaria parasites by their odour*, The Lancet Infectious Diseases, 19 (2019), pp. 578–580.

[62] T. Guo, M. Huang, Q. Zhu, Y. Guo, and J. Qin, *Hyperspectral image-based multi-feature integration for tvb-n measurement in pork*, Journal of Food Engineering, 218 (2018), pp. 61–68.

[63] J. Gutiérrez and M. C. Horrillo, *Advances in artificial olfaction: Sensors and applications*, Talanta, 124 (2014), pp. 95–105.

[64] A. Haehner, J. Draf, S. Dräger, T. Hummel, et al., *Predictive value of sudden olfactory loss in the diagnosis of covid-19*, Orl, 82 (2020), pp. 175–180.

[65] A. He and Z. Tang, *A novel gas identification method based on gabor spectrogram using self-adapted temperature modulated gas sensors*, in 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), IEEE, Aug. 2019, pp. 714–717.

[66] A. He, G. Wei, J. Yu, Z. Tang, Z. Lin, and P. Wang, *A novel dictionary learning method for gas identification with a gas sensor array*, IEEE Transactions on Industrial Electronics, 64 (2017), pp. 9709–9715.

[67] P. Herrero, L. R. Peterson, J. B. McGill, S. Matthew, D. Lesniak, C. Dence, and R. J. Gropler, *Increased myocardial fatty acid metabolism in patients with type 1 diabetes mellitus*, Journal of the American College of Cardiology, 47 (2006), pp. 598–604.

[68] A. E. Hoerl, R. W. Kannard, and K. F. Baldwin, *Ridge regression: some simulations*, Communications in Statistics-Theory and Methods, 4 (1975), pp. 105–123.

[69] O. Hotel, J.-P. Poli, C. Mer-Calfati, E. Scorsone, and S. Saada, *A review of algorithms for saw sensors e-nose based volatile compound identification*, Sensors and Actuators B: Chemical, 255 (2018), pp. 2472–2482.

[70] W. Hu, L. Wan, Y. Jian, C. Ren, K. Jin, X. Su, X. Bai, H. Haick, M. Yao, and W. Wu, *Electronic noses: from advanced materials to sensors aided with data processing*, Advanced Materials Technologies, 4 (2019), p. 1800488.

[71] R. Huber and M. Knaden, *Desert ants possess distinct memories for food and nest odors*, Proceedings of the National Academy of Sciences, 115 (2018), pp. 10470–10474.

[72] R. M. Humphries and A. J. Linscott, *Laboratory diagnosis of bacterial gastroenteritis*, Clinical microbiology reviews, 28 (2015), pp. 3–31.

[73] R. Hung, S. Lee, and J. W. Bennett, *Fungal volatile organic compounds and their role in ecosystems*, Applied Microbiology and Biotechnology, 99 (2015), pp. 3395–3405.

[74] P. Janbakhshi and M. B. Shamsollahi, *Ecg-derived respiration estimation from single-lead ecg using gaussian process and phase space reconstruction methods*, Biomedical Signal Processing and Control, 45 (2018), pp. 80–90.

[75] P. Jendrny, C. Schulz, F. Twele, S. Meller, M. von Köckritz-Blickwede, A. D. M. E. Osterhaus, J. Ebbers, V. Pilchová, I. Pink, T. Welte, et al., *Scent dog identification of samples from covid-19 patients–a pilot study*, BMC infectious diseases, 20 (2020), pp. 1–7.

[76] H. JI, W. ZENG, AND Y. LI, *Gas sensing mechanisms of metal oxide semiconductors: a focus review*, Nanoscale, 11 (2019), pp. 22664–22684.

[77] P. JIA, F. TIAN, Q. HE, S. FAN, J. LIU, AND S. X. YANG, *Feature extraction of wound infection data for electronic nose based on a novel weighted kpca*, Sensors and Actuators B: Chemical, 201 (2014), pp. 555–566.

[78] W. JIA, G. LIANG, Y. WANG, AND J. WANG, *Electronic noses as a powerful tool for assessing meat quality: a mini review*, Food Analytical Methods, 11 (2018), pp. 2916–2924.

[79] K. JIANG, F. YAN, AND H. ZHANG, *Data-driven modeling and ufir-based outlet no _{x} estimation for diesel-engine scr systems*, IEEE Transactions on Industrial Electronics, 67 (2019), pp. 5012–5021.

[80] B. JINADASA, *Determination of quality of marine fishes based on total volatile base nitrogen test (tvb-n)*, Nature and Science, 5 (2014), pp. 106–111.

[81] Y. JU, T. CHEN, B. MU, AND L. LJUNG, *On the influence of ill-conditioned regression matrix on hyper-parameter estimators for kernel-based regularization methods*, in 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, Dec. 2020, pp. 300–305.

[82] K. JUNEJA AND C. RANA, *Individual and mutual feature processed elm model for eeg signal based brain activity classification*, Wireless Personal Communications, 108 (2019), pp. 659–681.

[83] J. H. KAAS AND P. BALARAM, *Current research on the organization and function of the visual system in primates*, Eye and brain, 6 (2014), p. 1.

[84] G. KAISER, *A Friendly Guide to Wavelets*, Springer Science & Business Media, Nov. 2010.

[85] S. S. KEERTHI AND S. SHEVADE, *A fast tracking algorithm for generalized lars/lasso*, IEEE Transactions on Neural Networks, 18 (2007), pp. 1826–1830.

[86] D. KIM, S. LEE, K. LEE, S. BAEK, AND J. SEO, *Development of a ph indicator composed of high moisture-absorbing materials for real-time monitoring of chicken breast freshness*, Food science and biotechnology, 26 (2017), pp. 37–42.

[87] I. Knowles and R. J. Renka, *Methods for numerical differentiation of noisy data*, Electronic Journal of Differential Equations, 21 (2014), pp. 235–246.

[88] E. Krauss, J. Haberer, G. Barreto, M. Degen, W. Seeger, and A. Guenther, *Recognition of breathprints of lung cancer and chronic obstructive pulmonary disease using the aeonose® electronic nose*, Journal of breath research, 14 (2020), p. 046004.

[89] S. Labreche, S. Bazzo, S. Cade, and E. Chanie, *Shelf life determination by electronic nose: application to milk*, Sensors and Actuators B: Chemical, 106 (2005), pp. 199–206.

[90] M. Längkvist, L. Karlsson, and A. Loutfi, *A review of unsupervised feature learning and deep learning for time-series modeling*, Pattern Recognition Letters, 42 (2014), pp. 11–24.

[91] M. Längkvist and A. Loutfi, *Unsupervised feature learning for electronic nose data applied to bacteria identification in blood*, in NIPS 2011 workshop on deep learning and unsupervised feature learning, 2011.

[92] R. B. Lanjewar, S. Mathurkar, and N. Patel, *Implementation and comparison of speech emotion recognition system using gaussian mixture model (gmm) and k-nearest neighbor (k-nn) techniques*, Procedia Computer Science, 49 (2015), pp. 50–57.

[93] S. Lartviere and F. Messier, *Aposematic behaviour in the striped skunk, mephitis mephitis*, Ethology, 102 (1996), pp. 986–992.

[94] S. Ledauphin, D. Pommeret, and E. M. Qannari, *Application of hidden markov model to products shelf lives*, Food quality and preference, 19 (2008), pp. 156–161.

[95] S. Legg, A. Benavides-Serrano, J. D. Siirola, J.-P. Watson, S. Davis, A. Bratteteig, and C. D. Laird, *A stochastic programming approach for gas detector placement using cfd-based dispersion simulations*, Computers & Chemical Engineering, 47 (2012), pp. 194–201.

[96] S. Lekha and M. Suchetha, *Real-time non-invasive detection and classification of diabetes using modified convolution neural network*, IEEE journal of biomedical and health informatics, 22 (2017), pp. 1630–1636.

[97] J. Li, S. Zhu, S. Jiang, and J. Wang, *Prediction of egg storage time and yolk index based on electronic nose combined with chemometric methods*, LWT-Food Science and Technology, 82 (2017), pp. 369–376.

[98] W. Li, Y. Liu, X. Lu, Y. Huang, Y. Liu, S. Cheng, and Y. Duan, *A cross-sectional study of breath acetone based on diabetic metabolic disorders*, Journal of breath research, 9 (2015), p. 016005.

[99] X. Li, M. Parizeau, and R. Plamondon, *Training hidden markov models with multiple observations-a combinatorial method*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2000), pp. 371–377.

[100] S. Limbo, N. Sinelli, L. Torri, and M. Riva, *Freshness decay and shelf life predictive modelling of european sea bass (dicentrarchus labrax) applying chemical methods and electronic nose*, LWT-Food Science and Technology, 42 (2009), pp. 977–984.

[101] M. Lindfors and T. Chen, *Regularized lti system identification in the presence of outliers: A variational em approach*, Automatica, 121 (2020), p. 109152.

[102] B. Liu, X. Zeng, F. Tian, S. Zhang, and L. Zhao, *Domain transfer broad learning system for long-term drift compensation in electronic nose systems*, IEEE Access, 7 (2019), pp. 143947–143959.

[103] H. Liu, Y. Saito, D. F. Al Riza, N. Kondo, X. Yang, and D. Han, *Rapid evaluation of quality deterioration and freshness of beef during low temperature storage using three-dimensional fluorescence spectroscopy*, Food chemistry, 287 (2019), pp. 369–374.

[104] Q. Liu, X. Li, M. Ye, S. S. Ge, and X. Du, *Drift compensation for electronic nose by semi-supervised domain adaption*, IEEE Sensors Journal, 14 (2013), pp. 657–665.

[105] T. Liu, W. Zhang, P. McLean, M. Ueland, S. L. Forbes, and S. W. Su, *Electronic nose-based odor classification using genetic algorithms and fuzzy support vector machines*, International Journal of Fuzzy Systems, 20 (2018), pp. 1309–1320.

[106] T. Liu, W. Zhang, L. Ye, M. Ueland, S. L. Forbes, and S. W. Su, *A novel multi-odour identification by electronic nose using non-parametric modelling-based feature extraction and time-series classification*, Sensors and Actuators B: Chemical, 298 (2019), p. 126690.

[107] T. Liu, W. Zhang, M. Yuwono, M. Zhang, M. Ueland, S. L. Forbes, and S. W. Su, *A data-driven meat freshness monitoring and evaluation method using rapid centroid estimation and hidden markov models*, Sensors and Actuators B: Chemical, 311 (2020), p. 127868.

[108] A. Loutfi, S. Coradeschi, G. K. Mani, P. Shankar, and J. B. B. Rayappan, *Electronic noses for food quality: A review*, Journal of Food Engineering, 144 (2015), pp. 103–111.

[109] S. Lu, Z. Lu, J. Yang, M. Yang, and S. Wang, *A pathological brain detection system based on kernel based elm*, Multimedia Tools and Applications, 77 (2018), pp. 3715–3728.

[110] X. Lu, J. Wang, G. Lu, B. Lin, M. Chang, and W. He, *Quality level identification of west lake longjing green tea using electronic nose*, Sensors and Actuators B: Chemical, 301 (2019), p. 127056.

[111] Z. Lu, J. Sun, and K. Butts, *Multiscale support vector learning with projection operator wavelet kernel for nonlinear dynamical system identification*, IEEE Transactions on Neural Networks and Learning Systems, 28 (2017), pp. 231–243.

[112] H. Luo, P. Jia, S. Qiao, and S. Duan, *Enhancing electronic nose performance based on a novel qpso-rbm technique*, Sensors and Actuators B: Chemical, 259 (2018), pp. 241–249.

[113] J. Ma, D.-W. Sun, J.-H. Qu, D. Liu, H. Pu, W.-H. Gao, and X.-A. Zeng, *Applications of computer vision for assessing quality of agri-food products: a review of recent research advances*, Critical reviews in food science and nutrition, 56 (2016), pp. 113–127.

[114] A. Marjovi and L. Marques, *Swarm robotic plume tracking for intermittent and time-variant odor dispersion*, in 2013 European Conference on Mobile Robots, IEEE, 2013, pp. 379–384.

[115] A. McDermott, P. Whyte, N. Brunton, J. Lyng, J. Fagan, and D. Bolton, *The effect of organic acid and sodium chloride dips on the shelf-life of refrigerated irish brown crab (cancer pagurus) meat*, LWT, 98 (2018), pp. 141–147.

[116] A. McWilliams, P. Beigi, A. Srinidhi, S. Lam, and C. E. MacAulay, *Sex and smoking status effects on the early detection of early lung cancer in high-risk smokers using an electronic nose*, IEEE Transactions on Biomedical Engineering, 62 (2015), pp. 2044–2054.

[117] J. J. Miasik, A. Hooper, and B. C. Tofield, *Conducting polymer gas sensors*, Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases, 82 (1986), pp. 1117–1126.

[118] J. G. Monroy and J. Gonzalez-Jimenez, *Gas classification in motion: An experimental analysis*, Sensors and Actuators B: Chemical, 240 (2017), pp. 1205–1215.

[119] V. Y. Musatov, V. Sysoev, M. Sommer, and I. Kiselev, *Assessment of meat freshness with metal oxide sensor microarray electronic nose: A practical approach*, Sensors and Actuators B: Chemical, 144 (2010), pp. 99–103.

[120] A. Nayak, *A new regularization approach for numerical differentiation*, Inverse Problems in Science and Engineering, (2020), pp. 1–26.

[121] P. Nelli, G. Faglia, G. Sberveglieri, E. Cereda, G. Gabetta, A. Dieguez, A. Romano-Rodriguez, and J. Morante, *The aging effect on sno2–au thin film sensors: electrical and structural characterization*, Thin Solid Films, 371 (2000), pp. 249–253.

[122] Y. Niimura, *Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents*, Human genomics, 4 (2009), pp. 1–12.

[123] M. O,ÄôConnell, G. Valdora, G. Peltzer, and R. M. Negri, *A practical approach for fish freshness determinations using a portable electronic nose*, Sensors and Actuators B: chemical, 80 (2001), pp. 149–154.

[124] R. Paolesse, A. Alimelli, E. Martinelli, C. Di Natale, A. D'Amico, M. G. D'Egidio, G. Aureli, A. Ricelli, and C. Fanelli, *Detection of fungal con-*

tamination of cereal grain samples by an electronic nose, Sensors and Actuators B: Chemical, 119 (2006), pp. 425–430.

[125] A. PATLE AND D. S. CHOUHAN, *Svm kernel functions for classification*, in 2013 International Conference on Advances in Technology and Engineering (ICATE), IEEE, 2013, pp. 1–9.

[126] A. K. PAVLOU, N. MAGAN, C. MCNULTY, J. M. JONES, D. SHARP, J. BROWN, AND A. P. TURNER, *Use of an electronic nose system for diagnoses of urinary tract infections*, Biosensors and Bioelectronics, 17 (2002), pp. 893–899.

[127] T. C. PEARCE, S. S. SCHIFFMAN, H. T. NAGLE, AND J. W. GARDNER, *Handbook of Machine Olfaction: Electronic Nose Technology*, John Wiley & Sons, 2006.

[128] D. PELLEG, A. W. MOORE, ET AL., *X-means: extending k-means with efficient estimation of the number of clusters.*, in Icml, vol. 1, 2000, pp. 727–734.

[129] M. PENZA, G. CASSANO, F. TORTORELLA, AND G. ZACCARIA, *Classification of food, beverages and perfumes by wo3 thin-film sensors array and pattern recognition techniques*, Sensors and Actuators B: Chemical, 73 (2001), pp. 76–87.

[130] K. PERSAUD AND G. DODD, *Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose*, Nature, 299 (1982), pp. 352–355.

[131] G. PILLONETTO, F. DINUZZO, T. CHEN, G. D. NICOLAO, AND L. LJUNG, *Kernel methods in system identification, machine learning and function estimation: A survey*, Automatica, 50 (2014), pp. 657 – 682.

[132] G. PILLONETTO AND G. D. NICOLAO, *A new kernel-based approach for linear system identification*, Automatica, 46 (2010), pp. 81 – 93.

[133] L. D. PROCKOP AND R. I. CHICHKOVA, *Carbon monoxide intoxication: an updated review*, Journal of the neurological sciences, 262 (2007), pp. 122–130.

[134] P.-F. QI, M. ZENG, Z.-H. LI, B. SUN, AND Q.-H. MENG, *Design of a portable electronic nose for real-fake detection of liquors*, Review of Scientific Instruments, 88 (2017), p. 095001.

[135] J. Qian, Y. Luo, F. Tian, R. Liu, and T. Yang, *Design of multi-sensor electronic nose based on conformal sensor chamber*, IEEE Transactions on Industrial Electronics, (2020).

[136] S. Qiu and J. Wang, *The prediction of food additives in the fruit juice based on electronic nose with chemometrics*, Food chemistry, 230 (2017), pp. 208–214.

[137] L. R. Rabiner, *A tutorial on hidden markov models and selected applications in speech recognition*, Proceedings of the IEEE, 77 (1989), pp. 257–286.

[138] S. Rahayel, J. Frasnelli, and S. Joubert, *The effect of alzheimer's disease and parkinson's disease on olfaction: a meta-analysis*, Behavioural brain research, 231 (2012), pp. 60–74.

[139] B. Raman, M. Stopfer, and S. Semancik, *Mimicking biological design and computing principles in artificial olfaction*, ACS chemical neuroscience, 2 (2011), pp. 487–499.

[140] H. L. Ramírez, A. Soriano, S. Gómez, J. U. Iranzo, and A. I. Briones, *Evaluation of the food sniffer electronic nose for assessing the shelf life of fresh pork meat compared to physicochemical measurements of meat quality*, European Food Research and Technology, 244 (2018), pp. 1047–1055.

[141] J. Ranstam and J. Cook, *Lasso regression*, Journal of British Surgery, 105 (2018), pp. 1348–1348.

[142] G. C. Rezende, S. Le Calvé, J. J. Brandner, and D. Newport, *Micro photoionization detectors*, Sensors and Actuators B: Chemical, 287 (2019), pp. 86–94.

[143] F. Röck, N. Barsan, and U. Weimar, *Electronic nose: current status and future trends*, Chemical reviews, 108 (2008), pp. 705–725.

[144] R. A. Russell, A. Bab-Hadiashar, R. L. Shepherd, and G. G. Wallace, *A comparison of reactive robot chemotaxis algorithms*, Robotics and Autonomous Systems, 45 (2003), pp. 83–97.

[145] R. A. Russell, D. Thiel, R. Deveza, and A. Mackay-Sim, *A robotic system to locate hazardous chemical leaks*, in Proceedings of 1995 IEEE International Conference on Robotics and Automation, vol. 1, IEEE, 1995, pp. 556–561.

[146] Y. Salinas, J. V. Ros-Lis, J.-L. Vivancos, R. Martínez-Máñez, M. D. Marcos, S. Aucejo, N. Herranz, I. Lorente, and E. Garcia, *A novel colorimetric sensor array for monitoring fresh pork sausages spoilage*, Food Control, 35 (2014), pp. 166–176.

[147] R. Sameni, *Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising*, Signal Processing, 133 (2017), pp. 52–63.

[148] J. P. Santos and J. Lozano, *Real time detection of beer defects with a hand held electronic nose*, in Electron Devices (CDE), 2015 10th Spanish Conference on, IEEE, 2015, pp. 1–4.

[149] S. S. Schiffman, B. G. Kermani, and H. T. Nagle, *Analysis of medication off-odors using an electronic nose*, Chemical senses, 22 (1997), pp. 119–128.

[150] W. Sears, K. Colbow, and F. Consadori, *Algorithms to improve the selectivity of thermally-cycled tin oxide gas sensors*, Sensors and Actuators, 19 (1989), pp. 333–349.

[151] L. Sela and N. Sobel, *Human olfaction: a constant state of change-blindness*, Experimental brain research, 205 (2010), pp. 13–29.

[152] P. Shakya, E. Kennedy, C. Rose, and J. K. Rosenstein, *High-dimensional time series feature extraction for low-cost machine olfaction*, IEEE Sensors Journal, (2020).

[153] X.-J. Shen, Y. Dong, J.-P. Gou, Y.-Z. Zhan, and J. Fan, *Least squares kernel ensemble regression in reproducing kernel hilbert space*, Neurocomputing, 311 (2018), pp. 235–244.

[154] H. Shi, M. Zhang, and B. Adhikari, *Advances of electronic nose and its application in fresh foods: a review*, Critical reviews in food science and nutrition, 58 (2018), pp. 2700–2710.

[155] M. Shirasu and K. Touhara, *The scent of disease: volatile organic compounds of the human body related to disease and disorder*, The Journal of Biochemistry, 150 (2011), pp. 257–266.

[156] B. Siegmund and W. Pfannhauser, *Changes of the volatile fraction of cooked chicken meat during chill storing: results obtained by the electronic nose in comparison to gc-ms and gc olfactometry*, Zeitschrift Für Lebensmitteluntersuchung Und-Forschung A, 208 (1999), pp. 336–341.

[157] S. Siyang, C. Wongchoosuk, and T. Kerdcharoen, *Diabetes diagnosis by direct measurement from urine odor using electronic nose*, in The 5th 2012 Biomedical Engineering International Conference, IEEE, 2012, pp. 1–4.

[158] B. C. Smith, *Human olfaction, crossmodal perception, and consciousness*, 2017.

[159] A. J. Smola and B. Schölkopf, *A tutorial on support vector regression*, Statistics and computing, 14 (2004), pp. 199–222.

[160] J. M. Soares, A. Marjovi, J. Giezendanner, A. Kodiyan, A. P. Aguiar, A. M. Pascoal, and A. Martinoli, *Towards 3-d distributed odor source localization: an extended graph-based formation control algorithm for plume tracking*, in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2016, pp. 1729–1736.

[161] U. Stockhorst and R. Pietrowsky, *Olfactory perception, communication, and the nose-to-brain pathway*, Physiology & behavior, 83 (2004), pp. 3–11.

[162] H. Sun, F. Tian, Z. Liang, T. Sun, B. Yu, S. X. Yang, Q. He, L. Zhang, and X. Liu, *Sensor array optimization of electronic nose for detection of bacteria in wound infection*, IEEE Transactions on Industrial Electronics, 64 (2017), pp. 7350–7358.

[163] R. C. Teixeira, M. Rodríguez, N. J. de Romero, M. Bruins, R. Gómez, J. B. Yntema, G. C. Abente, J. W. Gerritsen, W. Wiegerinck, D. P. Bejerano, et al., *The potential of a portable, point-of-care electronic nose to diagnose tuberculosis*, Journal of Infection, 75 (2017), pp. 441–447.

[164] F. Tian, S. X. Yang, and K. Dong, *Circuit and noise analysis of odorant gas sensors in an e-nose*, Sensors, 5 (2005), pp. 85–96.

[165] X. Tian, J. Wang, and S. Cui, *Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors*, Journal of Food Engineering, 119 (2013), pp. 744–749.

[166] J. Torres-Tello, A. V. Guaman, and S.-B. Ko, *Improving the detection of explosives in a mox chemical sensors array with lstm networks*, IEEE Sensors Journal, 20 (2020), pp. 14302–14309.

[167] C. Trimmer, A. Keller, N. R. Murphy, L. L. Snyder, J. R. Willer, M. Nagai, N. Katsanis, L. B. Vosshall, H. Matsunami, and J. D. Mainland, *Genetic variation across the human olfactory receptor repertoire alters odor perception*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 9475–9480.

[168] A. Tripathy, A. Mohanty, and M. N. Mohanty, *Electronic nose for black tea quality evaluation using kernel based clustering approach*, International Journal of Image Processing (IJIP), 6 (2012), p. 86.

[169] F. Tsai and W. Philpot, *Derivative analysis of hyperspectral data*, Remote Sensing of Environment, 66 (1998), pp. 41–51.
https://doi.org/10.1016/S0034-4257(98)00032-7.

[170] A. ur Rehman and A. Bermak, *Heuristic random forests (hrf) for drift compensation in electronic nose applications*, IEEE Sensors Journal, 19 (2018), pp. 1443–1453.

[171] P. M. van Oort, P. Povoa, R. Schnabel, P. Dark, A. Artigas, D. C. Bergmans, T. Felton, L. Coelho, M. J. Schultz, S. J. Fowler, et al., *The potential role of exhaled breath analysis in the diagnostic process of pneumonia‚Äîa systematic review*, Journal of breath research, 12 (2018), p. 024001.

[172] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, and R. Huerta, *Chemical gas sensor drift compensation using classifier ensembles*, Sensors and Actuators B: Chemical, 166 (2012), pp. 320–329.

[173] G. Viccione, T. Zarrab, S. Giulianib, V. Naddeob, and V. Belgiornob, *Performance study of e-nose measurement chamber for environmental odour monitoring*, Chemical Engineering, 30 (2012), pp. 109–114.

[174] B. L. Villarreal and J. Gordillo, *Bioinspired smell sensor: nostril model and design*, IEEE/ASME Transactions on Mechatronics, 21 (2015), pp. 912–921.

[175] S. Volant, C. Bérard, M.-L. Martin-Magniette, and S. Robin, *Hidden markov models with mixtures as emission distributions*, Statistics and Computing, 24 (2014), pp. 493–504.

[176] A. Voss, K. Witt, T. Kaschowitz, W. Poitz, A. Ebert, P. Roser, and K.-J. Bär, *Detecting cannabis use on the human skin surface via an electronic nose system*, Sensors, 14 (2014), pp. 13256–13272.

[177] G. G. Walter and J. Zhang, *Orthonormal wavelets with simple closed-form expressions*, IEEE Transactions on Signal Processing, 46 (1998), pp. 2248–2251.

[178] L. Wang, S. Pang, and J. Li, *Olfactory-based navigation via model-based reinforcement learning and fuzzy inference methods*, IEEE Transactions on Fuzzy Systems, (2020).

[179] W. Wang, S. Lu, B. Hofmann, and J. Cheng, *Tikhonov regularization with $\ell 0$-term complementing a˜ convex penalty: $\ell 1$-convergence under sparsity constraints*, Journal of Inverse and Ill-posed Problems, 27 (2019), pp. 575–590.

[180] Y. Wang, Y. Li, J. Yang, J. Ruan, and C. Sun, *Microbial volatile organic compounds and their application in microorganism identification in foodstuff*, TrAC Trends in Analytical Chemistry, 78 (2016), pp. 1–16.

[181] J. Webster, P. Shakya, E. Kennedy, M. Caplan, C. Rose, and J. K. Rosenstein, *Trufflebot: Low-cost multi-parametric machine olfaction*, in 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2018, pp. 1–4.

[182] D. R. Wijaya, R. Sarno, and E. Zulaika, *Noise filtering framework for electronic nose signals: An application for beef quality monitoring*, Computers and Electronics in Agriculture, 157 (2019), pp. 305–321.

[183] D. Wilches, J. Rovira, I. Jaime, C. Palacios, M. Lurueña-Martínez, A. Vivar-Quintana, and I. Revilla, *Evaluation of the effect of a maternal rearing system on the odour profile of meat from suckling lamb*, Meat science, 88 (2011), pp. 415–423.

[184] C. M. Willis, S. M. Church, C. M. Guest, W. A. Cook, N. McCarthy, A. J. Bransbury, M. R. Church, and J. C. Church, *Olfactory detection of human bladder cancer by dogs: proof of principle study*, Bmj, 329 (2004), p. 712.

[185] A. G. Wintjens, K. F. Hintzen, S. M. Engelen, T. Lubbers, P. H. Savelkoul, G. Wesseling, J. A. van der Palen, and N. D. Bouvy, *Applying the electronic nose for pre-operative sars-cov-2 screening*, Surgical endoscopy, (2020), pp. 1–8.

[186] W. Wojnowski, T. Majchrzak, T. Dymerski, J. Gębicki, and J. Namieśnik, *Poultry meat freshness evaluation using electronic nose technology and ultra-fast gas chromatography*, Monatshefte Für Chemie-Chemical Monthly, 148 (2017), pp. 1631–1637.

[187] E. J. Wolfrum, R. M. Meglen, D. Peterson, and J. Sluiter, *Calibration transfer among sensor arrays designed for monitoring volatile organic compounds in indoor air quality*, IEEE Sensors Journal, 6 (2006), pp. 1638–1643.

[188] A. Wu, B. Yu, and T. Komiyama, *Plasticity in olfactory bulb circuits*, Current opinion in neurobiology, 64 (2020), pp. 17–23.

[189] H. Wu and M. Siegel, *Odor-based incontinence sensor*, in Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference [Cat. No. 00CH37066], vol. 1, IEEE, 2000, pp. 63–68.

[190] Y. Xing, T. A. Vincent, H. Fan, E. Schaffernicht, V. H. Bennetts, A. J. Lilienthal, M. Cole, and J. W. Gardner, *Firenose on mobile robot in harsh environments*, IEEE Sensors Journal, 19 (2019), pp. 12418–12431.

[191] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, *Wavelet transform domain filters: a spatially selective noise filtration technique*, IEEE transactions on image processing, 3 (1994), pp. 747–758.

[192] D. Yan, H. Cao, P. Zhang, and S. Yang, *Fuzzy linguistic odor cognition for robotics olfaction*, IEEE Transactions on Cognitive and Developmental Systems, 11 (2018), pp. 234–243.

[193] J. Yan, X. Guo, S. Duan, P. Jia, L. Wang, C. Peng, and S. Zhang, *Electronic nose feature extraction methods: A review*, Sensors, 15 (2015), pp. 27804–27831.

[194] J. Yan, F. Tian, Q. He, Y. Shen, S. Xu, J. Feng, and K. Chaibou, *Feature extraction from sensor data for detection of wound pathogen based on electronic nose*, Sensors and Materials, 24 (2012), pp. 57–73.

[195] L. Ye, A. Argha, B. G. Celler, Y. Zhang, H. T. Nguyen, and S. W. Su, *Non-parametric modelling of vo 2 response to exercise*, in Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, IEEE, 2017, pp. 1525–1528.

[196] L. Ye, A. Argha, H. Yu, B. G. Celler, H. T. Nguyen, and S. Su, *Dynamic characteristics of oxygen consumption*, Biomedical engineering online, 17 (2018), pp. 1–18.

[197] F. Yeni, S. Yavaş, H. Alpas, and Y. Soyer, *Most common foodborne pathogens and mycotoxins on fresh produce: a review of recent outbreaks*, Critical reviews in food science and nutrition, 56 (2016), pp. 1532–1544.

[198] R. Yi, J. Yan, D. Shi, Y. Tian, F. Chen, Z. Wang, and S. Duan, *Improving the performance of drifted/shifted electronic nose systems by cross-domain transfer using common transfer samples*, Sensors and Actuators B: Chemical, (2020), p. 129162.

[199] Z. Yi, W. Shang, T. Xu, S. Guo, and X. Wu, *Local discriminant subspace learning for gas sensor drift problem*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, (2020).

[200] S. M. Yimenu, J. Kim, and B. Kim, *Prediction of egg freshness during storage using electronic nose*, Poultry science, 96 (2017), pp. 3733–3746.

[201] X. Yin, L. Zhang, F. Tian, and D. Zhang, *Temperature modulated gas sensing e-nose system for low-cost and fast detection*, IEEE Sensors Journal, 16 (2015), pp. 464–474.

[202] Y. J. Yoo, H.-I. Kim, and S.-I. Choi, *Robust classification of largely corrupted electronic nose data using deep neural networks*, IEEE Sensors Journal, (2020).

[203] M. Yuwono, S. W. Su, B. D. Moulton, Y. Guo, and H. T. Nguyen, *An algorithm for scalable clustering: Ensemble rapid centroid estimation*, in 2014 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2014, pp. 1250–1257.

[204] M. Yuwono, S. W. Su, B. D. Moulton, and H. T. Nguyen, *Optimization strategies for rapid centroid estimation*, in 2012 Annual International Confer-

ence of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 6212–6215.

[205] D. Zappa, *Low-power detection of food preservatives by a novel nanowire-based sensor array*, Foods, 8 (2019), p. 226.

[206] G. Zeng, C. Wu, Y. Chang, C. Zhou, B. Chen, M. Zhang, J. Li, X. Duan, Q. Yang, and W. Pang, *Detection and discrimination of volatile organic compounds using a single film bulk acoustic wave resonator with temperature modulation as a multiparameter virtual sensor array*, ACS sensors, 4 (2019), pp. 1524–1533.

[207] Z.-Q. Zeng, H.-B. Yu, H.-R. Xu, Y.-Q. Xie, and J. Gao, *Fast training support vector machines using parallel sequential minimal optimization*, in 2008 3rd international conference on intelligent system and knowledge engineering, vol. 1, IEEE, 2008, pp. 997–1001.

[208] L. Zhang and P. Deng, *Abnormal odor detection in electronic nose via self-expression inspired extreme learning machine*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49 (2017), pp. 1922–1932.

[209] L. Zhang, Y. Liu, Z. He, J. Liu, P. Deng, and X. Zhou, *Anti-drift in e-nose: A subspace projection approach with drift reduction*, Sensors and Actuators B: Chemical, 253 (2017), pp. 407–417.

[210] L. Zhang, F. Tian, and D. Zhang, *Electronic Nose: Algorithmic Challenges*, Springer, 2018.

[211] L. Zhang, F.-C. Tian, X.-W. Peng, and X. Yin, *A rapid discreteness correction scheme for reproducibility enhancement among a batch of mos gas sensors*, Sensors and Actuators A: Physical, 205 (2014), pp. 170–176.

[212] L. Zhang and D. Zhang, *Domain adaptation extreme learning machines for drift compensation in e-nose systems*, IEEE Transactions on instrumentation and measurement, 64 (2014), pp. 1790–1801.

[213] ——, *Efficient solutions for discreteness, drift, and disturbance (3d) in electronic olfaction*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48 (2018), pp. 242–254.

[214] W. Zhang, T. Liu, M. Ueland, S. L. Forbes, R. X. Wang, and S. W. Su, *Design of an efficient electronic nose system for odour analysis and assessment*, Measurement, 165 (2020), p. 108089.

[215] W. Zhang, T. Liu, L. Ye, M. Ueland, S. L. Forbes, and S. W. Su, *A novel data pre-processing method for odour detection and identification system*, Sensors and Actuators A: Physical, 287 (2019), pp. 113–120.

[216] W. Zhang, T. Liu, M. Zhang, Y. Zhang, H. Li, M. Ueland, S. L. Forbes, X. R. Wang, and S. W. Su, *Nos. e: a new fast response electronic nose health monitoring system*, in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 4977–4980.

[217] W. Zhao, Q.-H. Meng, M. Zeng, and P.-F. Qi, *Stacked sparse auto-encoders (ssae) based electronic nose for chinese liquors classification*, Sensors, 17 (2017), p. 2855.

[218] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, Journal of the royal statistical society: series B (statistical methodology), 67 (2005), pp. 301–320.