An Improved EEG Pattern Classification System Based on Dimensionality Reduction and Classifier Fusion

By

Akram Saleh Mousa AlSukker

Thesis submitted as a requirement for the degree of

Doctor of Philosophy

School of Electrical, Mechanical and Mechatronic Systems
Faculty of Engineering and Information Technology

University of Technology, Sydney (UTS)

July, 2012

Certificate

I, Akram Saleh Mousa AlSukker, declare that this thesis titled, An Improved EEG Pattern Classification System Based on Dimensionality Reduction and Classifier Fusion, and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has been previously submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Production Note: Signature removed prior to publication.

Date: 6/7/2012

Acknowledgements

First and foremost, all my praise be to Allah the Most Gracious, the Most Merciful whose guidance constantly inspires me and has given me courage and strength to complete my PhD. I especially express my deepest gratitude to my supervisor, **Dr. Ahmed Al-Ani**, for his continual guidance, positive comments and tremendous support throughout my PhD candidature. His perpetual energy, enthusiasm and devotion to biomedical signal processing have motivated me to continue my research in the area. Special thanks go to **Dr. Rami Kushaba** for his collaboration, creative ideas and continuous support. He was my companion throughout my candidature and was of great help in my thesis writing. My associate supervisor **A/Prof. Adel Al-Jumaily** also deserves a special thank for his guidance, advice and assistance during the course of my PhD candidature. I wish to acknowledge the Department of Medical Informatics at the University of Technology, Graz, in Austria and the BCI competition web site for making the EEG datasets available for the research. I would also like to thank Ms Rosalind Elliott from Royal North Shore Hospital, Sydney, Australia for providing sleep study datasets. I would like to express my appreciation to Prof. Hung Nguyen for his support and for facilitating the use of laboratory equipment. My thanks are also due to A/Prof. Fazlul Huq of Discipline of Biomedical Science, The University of Sydney for kindly going through the thesis draft and for making useful suggestions towards improvement of language, its flow and coherence. I wish to thank my colleagues at UTS, in particular Mr. Thamer Hanoun for their company, friendship and support. My greatest heartfelt thanks go to my loving wife Zaynab Al-Eisawi for her constant support, encouragement and sacrifice throughout my candidature. I am also grateful to Allah for blessing us with two beautiful boys, Bassam and Yousef who have been a source of constant joy, love and affection in our heart and mind.

Dedication

I would like to dedicate this thesis to my mother and father, Noura and Saleh AlSukker. Throughout my lifetime, they always attempted to instill in me the desire to acquire new knowledge and their words of wisdom are a source of inspiration and strength.

Abstract

Analysis of brain electrical activities (Electroencephalography, EEG) presents a rich source of information that helps in the advancement of affordable and effective biomedical applications such as psychotropic drug research, sleep studies, seizure detection and brain computer interface (BCI). Interpretation and understanding of EEG signal will provide clinicians and physicians with useful information for disease diagnosis and monitoring biological activities. It will also help in creating a new way of communication through brain waves.

This thesis aims to investigate new algorithms for improving pattern recognition systems in two main EEG-based applications. The first application represents a simple Brain Computer Interface (BCI) based on imagined motor tasks, whilst the second one represents an automatic sleep scoring system in intensive care unit. BCI system in general aims to create a non-muscular link between brain and external devices, thus providing a new control scheme that can most benefit the extremely immobilised persons. This link is created by utilizing pattern recognition approach to interpret EEG into device commands. The commands can then be used to control wheelchairs, computers or any other equipment. The second application relates to creating an automatic scoring system through interpreting certain properties of several biomedical signals. Traditionally, sleep specialists record and analyse brain signal using electroencephalogram (EEG), muscle tone (EMG), eye movement (EOG), and other biomedical signals to detect five sleep stages: Rapid Eye Movement (REM), stage 1,... to stage 4. Acquired signals are then scored based on 30 seconds intervals that require manually inspecting one segment at a time for certain properties to interpret sleep stages. The process is time consuming and demands competence. It is thought that an automatic scoring system mimicking sleep expert rules will speed up the process and reduce the cost.

Practicality of any EEG-based system depends upon accuracy and speed. The more accurate and faster classification systems are, the better will be the chance to integrate them in wider range of applications. Thus, the performance of the previous systems is further enhanced using improved feature selection, projection and classification algorithms.

As processing EEG signals requires dealing with multi-dimensional data, there is a need to minimize the dimensionality in order to achieve acceptable performance with less computational cost. The first possible candidate for dimensionality reduction is employed using channel/feature selection approach. Four novel feature selection methods are de-

veloped utilizing genetic algorithms, ant colony, particle swarm and differential evolution optimization. The methods provide fast and accurate implementation in selecting the most informative features/channels that best represent mental tasks. Thus, computational burden of the classifier is kept as light as possible by removing irrelevant and highly redundant features.

As an alternative to dimensionality reduction approach, a novel feature projection method is also introduced. The method maps the original feature set into a small informative subset of features that can best discriminate between the different class. Unlike most existing methods based on discriminant analysis, the proposed method considers fuzzy nature of input measurements in discovering the local manifold structure. It is able to find a projection that can maximize the margin between data points from different classes at each local area while considering the fuzzy nature.

In classification phase, a number of improvements to traditional nearest neighbour classifier (kNN) are introduced. The improvements address kNN weighting scheme limitations. The traditional kNN does not take into account class distribution, importance of each feature, contribution of each neighbour, and the number of instances for each class. The proposed kNN variants are based on improved distance measure and weight optimization using differential evolution. Differential evolution optimizer is utilized to enhance kNN performance through optimizing the metric weights of features, neighbours and classes. Additionally, a Fuzzy kNN variant has also been developed to favour classification of certain classes. This variant may find use in medical examination. An alternative classifier fusion method is introduced that aims to create a set of diverse neural network ensemble. The diversity is enhanced by altering the target output of each network to create a certain amount of bias towards each class. This enables the construction of a set of neural network classifiers that complement each other.

All proposed feature reduction and classification algorithms have been tested and verified on different datasets including EEG (represented by BCIs and sleep scoring problem). The results obtained are very encouraging when comparing with their counterparts found in the literature.

Acronyms and Abbreviations

AAR: Adaptive Auto-Regressive
AASM: American Academy of Sleep Medicine
ACO: Ant Colony Optimization
ACOFUZZY: Fuzzy Ant Colony Optimization
ANOVA: Analysis of Variance
AR: Auto-Regressive
BCI: Brain computer Interface
BIC: Bayesian Information Criterion
BMI: Brain Machine Interface
BPSO: Binary Particle Swarm Optimization
BPSOMI: Mututal Information based Binary Particle Swarm Optimization
BSS: Blind Source Separation
CAT scan: Computerized Axial Tomography scan
*cb*NNE: Class biased Neural Network Ensemble
CSP: Common Spatial Pattern
CT scan: Computerized Tomography scan
CWkNN: Class Weighting k Nearest Neighbour
CWT: Continuous Wavelet Transform
DE: Differential Evolution
DEFS: Differential Evolution Feature Selection
DFT: Discrete Fourier Transform
DGA: Diverse Genetic Algorithm
DNF: Desired Number of Features
DSkNN: Dempster-Shafer k Nearest Neighbour
DWT: Discrete Wavelet Transform
EAs: Evolutionary Algorithms
ECG: electrocardiogram
ECoG: Electrocorticogram
EEG: Electroencephalogram
EMG: Electromyogram
EOG: electrooculogram
ERP: Event-Related Potential
ES: Exhaustive Search
EV: Evoked Potential
FCM: Fuzzy C-Means
FDR: Fisher Discrimination Rate
FFT: Fast Fourier Transform
FISkNN: Fuzzy Inference System k Nearest Neighbour

fkNN: Fuzzy k Nearest neighbour
FLDA: Fuzzy Linear Discriminant Analysis
fMRI: functional Magnetic Resonance Imaging
FT: Fourier Transform
GA: Genetic Algorithm
GARF: Relative Fitness scaling Genetic Algorithm
HGA: Hybrid Genetic Algorithm
IBPSO: Improved Binary Particle Swarm Optimization
ICA: Independent Component Analysis
ICU: Intensive Care Unit
kNN: k Nearest Neighbour
LDA: Linear Discriminant Analysis
LFDA: Local Fisher Discriminant Analysis
LLE: Locally Linear Embedding
LMS: Least-Mean-Squares
LMV: Local Mean Vector
LPP: Locality Preserving Projections
LSDA: Locality Sensitive Discriminant Analysis
LVQ: Linear Vector Quantization
MEG: Magnetoencephalography
MI: Mututal Information
MIFE: Mutual Information Feature Evaluation
MIFS: Mutual Information Feature selection
mkNN: Modified k Nearest Neighbour
ML: Machine Learning
MLP: Multi-Layer Perception
MRI: Magnetic Resonance Imaging
MSE: Mean Square Error
NCA: Neighbourhood Componants Analysis
ND: Negative Distribution Fator
NIRS: Near-Infrared Spectroscopy
NN: Neural Network
NNE: Neural Network Ensemble
NoS: Number of Evaluated Subsets
NPE: Neighbourhood Preserving Embedding
NREM: Non Rypid Eye Movment
NWkNN: Neighbourhood Weighting k Nearest Neighbour
OLSFDA: Orthogonal Locality Sensitive Fuzzy Discriminant Analysis
PCA: Principal Component Analysis
PD: Positive Distribution Factor

PET: Positron Emission Tomography

PkNN: Pseudo k Nearest Neighbour

PSD: Power Spectral Density

PSG: Polysomnogram

PSO: Particle Swarm Optimization

PTA: Plus Take-Away

R&K: Rechtschaffen and Kales Rules

RBF: Radial Basis Function

REM: Rypid Eye Movment

RLS: Recursive-Least-Squares

ROC: Receiver Operating Characteristics curve

SA: Simulated Annealing

SBS: Sequential Backward Search

SCI: Spinal Cord Injuriy

SCP: Slow Cortical Potentials

sEMG: surface Electromyogram

SFFS: Sequential Forward Floating Search

SFS: Sequential Forward Search

SGA: Simple Genetic Algorithm

SOFM: Self-Organizing Feature Maps

SSVEP: Steady-State Visual Evoked Potentials

STFT: Short-Time Fourier Transform

SVD: Singular Value Decomposition

SVM: Support Vector Machine

SWS: Slow-Wave Sleep

TMS: Transcranial Magnetic Stimulation

TS: Tabu Search

TSP: Travel Saleman Problem

UCI: University of California, Irvine

uLDA: uncorrelated Linear Discriminant Analysis

WGSSE: Within the Generalized Group Sum of Squared Error

wkNN: Weighted k Nearest Neighbour

wm$k$NN: Weighted Modified k Nearest Neighbour

WPD: Wavelet Packet Decomposition

WPT: Wavelet Packet Transform

WT: Wavelet Transform

WTL: Win-Tie-Loss

WVD: Wigner-Ville Distribution

Contents

List of Figures

List of Tables