



Adaptive Sampling for Spatial Prediction in Wireless Sensor Networks

By

Van Linh NGUYEN

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)

December 2014

Certificate of Original Authorship

I, Van Linh Nguyen, certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgements

I would like to thank everybody who have helped me to accomplish this thesis. First and foremost, I would like to express my deepest gratitude to my principal supervisor, Associate Professor Sarath Kodagoda, who has always been a great mentor, giving me tremendously valuable suggestions, thoughtful guidance, and continuous support throughout past years. The accomplishment of this thesis would not have been possible without his invaluable advice.

I am sincerely grateful to my co-supervisor, Professor Gamini Dissanayake, for his research expertise and deep insight from face-to-face meetings at the beginning of my study having inspired me in selecting my research topic. I am thankful to Dr Ravindra Ranasinghe, who is also my co-supervisor, for informative comments and technical advice he has provided in group meetings and individual discussions.

I would like to express my deeply gratitude to my former supervisor at Vietnam National University of Agriculture, Dr. Duong Van Nguyen, now in Ministry of Education and Training, Vietnam for his support and advice on both research and a real life.

I would also like to thank the Vietnam International Education Development (VIED) for having offered me a PhD scholarship and having introduced me to the VIED-UTS joint scholarship. Without this scholarship, I would never have been able to create this thesis at this point.

I am indebted to the Vice-Chancellor's Postgraduate Research Student Conference Funding and the Faculty of Engineering and Information Technology (FEIT) Travel Funding committees for their generous financial support for my conference

Acknowledgements

trips. I would also send many thanks to the officers at the Research and Development Office, FEIT for their dedicated assistance.

I am very keen to appreciate the rest of the team at the Center for Autonomous Systems (CAS) for their help and support. Through many interesting discussions and seminars, I have learned so much about doing research and being part of scientific community.

I would like to extend further my thanks to the experts from the Sydney Water, including Heriberto Bustamante, Dammika Vitanage, Tung Nguyen, and Nur Nobi for giving me good advice from face-to-face and email meetings and providing me with the data sets for my simulated experiments.

Last but not least, I cannot thank enough my parents and my parents-in-law for always being there for me and for letting me choose my way. A very special thank goes to my beloved spouse, Huong, for her endless love, patience, and encouragement. In particular, my heartfelt gratitude is towards my daughter, Giang, for being part of my life.

Contents

Certificate of Original Authorship	i
Acknowledgements	ii
Contents	iv
List of Figures	viii
Acronyms and Abbreviations	xii
Abstract	xiv
1 Introduction	1
1.1 Background	2
1.2 Research Objectives	5
1.3 Main Contributions	7
1.4 Organization of the Thesis	12
1.5 Publications	13
1.6 Summary	15
2 Notations and Preliminaries	16
2.1 Introduction	16
2.2 Mathematical Notations	16
2.3 Random Field Models	17
2.3.1 Gaussian process	18
2.3.2 Gaussian Markov random field	19
2.4 Gaussian Process Regression	21
2.5 Voronoi Partition	23
2.6 Mobile Robotic Wireless Sensor Network	24

2.6.1	Network model	25
2.6.2	Physical collision avoidance	26
2.7	Submodularity Function	27
2.8	Summary	29
3	Literature Review	30
3.1	Introduction	30
3.2	Sensor Selection in SWSNs	30
3.2.1	Criteria for sensor selection	31
3.2.2	Algorithms for sensor selection	32
3.3	Sensor Placement in MRWSNs	34
3.3.1	Sensor placement in WSNs	35
3.3.2	Sensor placement for spatial prediction	36
3.4	GMRF for Spatial Prediction	39
3.5	Summary	41
4	Sensor Selection for Spatial Prediction	42
4.1	Introduction	42
4.2	Spatial Field Model	44
4.3	Sensor Selection in Unconstrained Environment	47
4.3.1	Metric for sensor selection	47
4.3.2	Simulated annealing based sensor selection	49
4.3.2.1	Simulated annealing technique	50
4.3.2.2	Simulated annealing based algorithm	51
4.4	Sensor Selection in Constrained Environment	53
4.4.1	Spatial modelling of Hydrogen sulphide	54
4.4.1.1	Hydrogen sulphide model as a multivariate normal distribution	55
4.4.1.2	Spatial distribution of hydrogen sulphide model using Gaussian process	56
4.4.2	Mutual information based sensor selection	58
4.4.2.1	Mutual information function	59
4.4.2.2	Mutual information based algorithm	60
4.5	Experiments and Results	62
4.5.1	Unconstrained Environment	63
4.5.1.1	Indoor environment	63
4.5.1.2	Outdoor environment	67
4.5.2	Constrained Environment	73
4.6	Summary	79
5	Sensor Placement for Spatial Prediction using Gaussian Processes	81
5.1	Introduction	81

5.2	Gaussian Predictive Inference	84
5.3	Centralized Sensor Placement	86
5.3.1	Learning model parameters	87
5.3.2	Locational optimization based sampling approach	90
5.3.2.1	Centroidal Voronoi tessellation	90
5.3.2.2	Locational optimization based algorithm	93
5.3.3	Entropy based sampling approach	96
5.3.3.1	Conditional entropy	97
5.3.3.2	Adaptive sampling strategy	98
5.3.3.3	Stopping Criterion Based on Lower Bound	104
5.4	Distributed Sensor Placement	108
5.4.1	Distributed computation tools	108
5.4.1.1	Jacobi over-relaxation approach	109
5.4.1.2	Discrete-time average consensus approach	110
5.4.2	Distributed Gaussian inference	110
5.4.3	Distributed estimators in MRWSNs over time	113
5.4.3.1	Distributed parameters learning	114
5.4.3.2	Predicted means and variances	116
5.4.4	Distributed sampling paths	117
5.5	Experimental Results	118
5.5.1	Centralized sampling scheme	118
5.5.2	Distributed sampling scheme	123
5.6	Summary	128
6	Spatial Estimation using Gaussian Markov Random Field	132
6.1	Introduction	132
6.2	Spatial Field Model via GMRF	135
6.3	Spatial Predictive Inference via GMRF	138
6.3.1	The SPDE approach	138
6.3.2	Centralized prediction schemes	140
6.3.3	Distributed prediction schemes	142
6.3.4	Parameters estimation	146
6.3.4.1	Centralized learning	146
6.3.4.2	Distributed learning	147
6.4	Adaptive Sampling Approach	148
6.4.1	One-step-ahead spatial inference	148
6.4.2	Centralized sampling strategy	150
6.4.3	Distributed sampling strategy	154
6.5	Experimental Results	155
6.5.1	Centralized navigation scheme	156
6.5.2	Distributed navigation scheme	162
6.6	Summary	168

7	Conclusions and Future Research	169
7.1	Thesis Conclusions	169
7.2	Future Research	172
A	Mathematical Background	173
A.1	Block-wise Inversion of Matrix	173
A.2	Matrix Determinant Lemma	174
A.3	Woodbury Matrix Identity	174
A.4	Cholesky Decomposition	174
B	Information Theory	176
B.1	Chain Rule for Entropy	176
	Bibliography	177

List of Figures

2.1	Predicted rainfall (1/10th of a mm) in Switzerland: (a) Posterior mean, (b) Posterior variance. Colour bars represent range of rainfall. X Coord and Y Coord are in km.	23
2.2	Allowable movement regions of sensors at time step t (red) versus Voronoi cells (blue).	28
4.1	Boundary constraints in a sewer system: Although 'A' and 'B' are proximally located, the H ₂ S contents are "isolated" from each other. .	57
4.2	A deployment of a sensor network with 54 nodes at the Intel Berkeley research lab.	63
4.3	Average root mean square errors of temperature distribution of the Intel Berkeley research lab.	64
4.4	Deploying sensors densely in the center: Square - the locations selected by using SA based algorithm; diamond - the locations selected by using direct greedy method; pentagon - the locations selected by using mutual information approach; and triangle - the locations selected by using entropy technique. X Coord and Y Coord are in m. .	65
4.5	Average root mean square errors of densely central deployment. . . .	66
4.6	Deploying sensors densely in the boundary: Square - the locations selected by using SA based algorithm; diamond - the locations selected by using direct greedy method; pentagon - the locations selected by using mutual information approach; and triangle - the locations selected by using entropy technique. X Coord and Y Coord are in m. .	67
4.7	Average root mean square errors of densely border deployment. . . .	68
4.8	Average root mean square errors of Swiss rainfall data.	68
4.9	Deploying gauges and prediction of rainfall (1/10th of a mm) using all gauges. (a) Entire 100 gauges deployment throughout Switzerland; (b) Chosen gauges deployment of 30 sensor locations selected by using SA based algorithm (square), using direct greedy method (diamond), using mutual information approach (pentagon), and using entropy technique (triangle); (c) Predicted rainfall means using all gauges; (d) Predicted rainfall variance using all gauges. Colour bar in (c) represents range of rainfall, but colour bar in (d) describes range of errors. X Coord and Y Coord are in km.	69

4.10	Comparisons of the Swiss rainfall (1/10th of a mm) predicted means using 30 best sensors obtained by (a) SA based algorithm, (b) direct greedy approximation method, (c) mutual information approach, and (d) entropy technique. Colour bars illustrate range of rainfall. X Coord and Y Coord are in km.	70
4.11	Comparisons of the Swiss rainfall (1/10th of a mm) predicted variances using 30 best sensors obtained by (a) SA based algorithm, (b) direct greedy approximation method, (c) mutual information approach, and (d) entropy technique. Colour bars demonstrate range of errors. X Coord and Y Coord are in km.	71
4.12	Average root mean square errors of solid calcium ingredient. Results of the greedy and MI based approaches mostly overlap.	72
4.13	Forced ventilation zone in the North Head system, Sydney Water. . .	74
4.14	(a) The gas phase H ₂ S field. The predicted field using (b) 50, (c) 40, (d) 30 gas sensors, respectively. Colour bars represent range of hydrogen sulfide in ppm.	75
4.15	The H ₂ S predictions at particular locations using (a) 50, (b) 40, (c) 30 gas sensors, respectively.	76
4.16	The results for H ₂ S prediction. Root mean square errors.	77
5.1	True field of spatial diffusion	88
5.2	Estimated model parameters: (a) σ^2 , (b) ϕ and (c) τ^2	89
5.3	True field of spatial temperature. Colour bar shows range of temperature.	119
5.4	Conditional entropies $H(\hat{y}_{t+1} y_{1:t})$ in the networks of 5, 10, 20 sensors, respectively.	120
5.5	The predicted field (a) at time $t = 19$ using $N = 5$ sensors, (c) at time $t = 12$ using $N = 10$ sensors, and (e) at time $t = 6$ using $N = 20$ sensors. The predicted error variances and the sampling paths (b) at time $t = 19$ using $N = 5$ sensors, (d) at time $t = 12$ using $N = 10$ sensors, and (f) at time $t = 6$ using $N = 20$ sensors. Colour bars in left-wise column represent range of temperature, and colour bars in right-wise column represent range of errors.	121
5.6	Comparison of the predicted error variances and the sampling paths between (a) locational optimization and (b) entropy based approaches. Range of errors is shown in colour bars.	122
5.7	The globally estimated: (a) mean parameter β_0 (b) marginal variance σ^2 , (c) bandwidth of the kernel ϕ , and (d) noise variance τ^2	124
5.8	The true field of the soil organic matter. Percentage of the soil organic matter is shown in color bar.	125

5.9	The column-wise pairs compare the predicted fields obtained by distributed and centralized approaches at times (a) and (c) $t = 1$, (b) and (d) $t = 5$. Percentage of the soil organic matter is shown in color bar.	126
5.9	The column-wise pairs compare the predicted fields obtained by distributed and centralized approaches at times (e) and (g) $t = 10$, (f) and (h) $t = 15$. Percentage of the soil organic matter is shown in color bar.	127
5.10	The globally predicted variances and the sampling paths at times (a) $t = 1$, (b) $t = 5$, (c) $t = 10$, and (d) $t = 15$. Range of errors is illustrated in colour bars.	129
5.11	Root mean square errors.	130
5.12	Convergences of the DAC algorithms of different sensors at time step 10: (a) estimated mean parameter β_0 of sensor 18, (b) estimated marginal variance σ^2 of sensor 7, (c) estimated bandwidth of the kernel ϕ of sensor 15, and (d) estimated noise variance of sensor 8.	131
6.1	The spatial field $\mathcal{Q} \subset \mathbb{R}^2$. Voronoi cells are partitioned by dash red lines and a mesh of a triangulation is created by solid blue lines.	136
6.2	The true field of the numerically generated data set. Range of the fields is shown in color bar.	156
6.3	The results for numerically generated field in Case 1: The predicted fields at time (a) $t = 1$, (b) $t = 10$, and (c) $t = 20$. Range of the fields is illustrated in color bars.	157
6.4	The results for numerically generated field in Case 1: The predicted variances in the field at time (a) $t = 1$, (b) $t = 10$, and (c) $t = 20$. Range of the errors is demonstrated in color bars. The sampling paths for MRWSs up to time steps 1, 10 and 20, respectively, are illustrated by white circles, where current mobile sensor locations are shown in white dots.	160
6.5	The sampling paths for MRWSs up to time step 20 are illustrated by white circles, where current mobile sensor locations are shown in white dots. (a) Case 2 and (b) Case 3.	161
6.6	The root mean square errors for numerically generated field at different starting conditions.	162
6.7	The results for agricultural field: The model parameters of GMRF are estimated by centralized and distributed approaches proposed. These parameters are (a) mean parameter β_0 , (b) $\log(\tau)$, and (c) $\log(\kappa)$	163
6.8	The true field of the agricultural data set. Percentage of the soil organic matter is shown in color bar.	164

6.9	The results for agricultural field: (a), (c) and (e) are the predicted field at times $t = 1$, $t = 5$, and $t = 10$, respectively. The predicted variances in the field at times (b) $t = 1$, (d) $t = 5$, and (f) $t = 10$. The sampling paths for MRWSs up to time steps 1, 5 and 10, respectively, are shown by white circles, where current mobile sensor locations are demonstrated in white dots.	165
6.10	The root mean square errors for agricultural field.	166

Acronyms and Abbreviations

WSN	Wireless Sensor Network
WSNs	Wireless Sensor Networks
GP	Gaussian Process
GPs	Gaussian Processes
SWSN	Stationary Wireless Sensor Network
SWSNs	Stationary Wireless Sensor Networks
MRWSN	Mobile Robotic Wireless Sensor Network
MRWSNs	Mobile Robotic Wireless Sensor Networks
MRWS	Mobile Robotic Wireless Sensor
MRWSs	Mobile Robotic Wireless Sensors
GMRF	Gaussian Markov Random Field
RMSE	Root Mean Square Error
RMSEs	Root Mean Square Errors
ARMSE	Average Root Mean Square Error
ARMSEs	Average Root Mean Square Errors
JOR	Jacobi Over-Relaxation
DAC	Discrete-time Average Consensus
H ₂ S	Hydrogen Sulphide
SPDE	Stochastic Partial Differential Equation
MI	Mutual Information
SA	Simulated Annealing
OF	Objective Function
MAC	Medium Access Control
VIED	Vietnam International Education Development

Acronyms and Abbreviations

CAS	Center for Autonomous Systems
FEIT	Faculty of Engineering and Information Technology
UTS	University of Technology, Sydney

UNIVERSITY OF TECHNOLOGY, SYDNEY

Abstract

Faculty of Engineering and Information Technology

Centre for Autonomous Systems

Doctor of Philosophy

Adaptive Sampling for Spatial Prediction in Wireless Sensor Networks

by Van Linh NGUYEN

Networks of wireless sensors are increasingly exploited in crucial applications of monitoring spatially correlated environmental phenomena such as temperature, rainfall, soil ingredients, and air pollution. Such networks enable efficient monitoring and measurements can be included in developing models of the environmental fields even at unobserved locations. This requires determining the number of sensors and their sampling locations which minimize the uncertainty of predictions. Therefore, the aim of this thesis is to present novel, efficient and practically feasible approaches to sample the environments, so that the uncertainties at unobserved locations are minimized. Gaussian process (GP) is utilized to statistically model the spatial field. This thesis includes both stationary wireless sensor networks (SWSNs) and mobile robotic wireless sensor networks (MRWSNs), and thus the issues are correspondingly formulated into *sensor selection* and *sensor placement* problems, respectively. In the first part of the thesis, a novel performance metric for the sensor selection in the SWSNs, named average root mean square error, which reflects the average uncertainty of each predicted location, is proposed. In order to minimize this NP-hard

and combinatorial optimization problem, a simulated annealing based algorithm is proposed; and the sensor selection problem is effectively addressed. Particularly, when considering the sensor selection in constrained environments, *e.g.* gas phase hydrogen sulphide in a sewage system, a modified GP with an improved covariance function is developed. An efficient mutual information maximization criterion suitable for this particular scenario is also presented to select the most informative gaseous sensor locations along the sewer system. The second part of this thesis introduces centralized and distributed methods for spatial prediction over time in the MRWSNs. For the purpose of finding the optimal sampling paths of the mobile wireless sensors to take the most informative observations at each time iteration, a sampling strategy is proposed based on minimizing the uncertainty at all unobserved locations. A novel and very efficient optimality criterion for the adaptive sampling problem is then presented so that the minimization can be addressed by a greedy algorithm in polynomial time. The solution is proven to be bounded; and computational time of the proposed algorithm is illustrated to be practically feasible for the resource-constrained MRWSNs. In order to enhance the issue of computational complexity, Gaussian Markov random field (GMRF) is utilized to model the spatial field exploiting sparsity of the precision matrix. A new GMRF optimality criterion for the adaptive navigation problem is also proposed such that computational complexity of a greedy algorithm to solve the resulting optimization is deterministic even with increasing number of measurements. Based on the realistic simulations conducted using the pre-published data sets, it has shown that the proposed algorithms are superior with appealing results.

*To the memory of my dearest grandfathers and
grandmother!*