CALIBRATING SENSOR NETWORKS FOR ENVIRONMENT SENSING APPLICATION: A SPECIFIC CASE STUDY ON CHEMICAL SENSING

ABHISHEK GROVER

BHARTI SCHOOL OF TELECOMMUNICATION TECHNOLOGY AND MANAGEMENT
INDIAN INSTITUTE OF TECHNOLOGY DELHI
SEPTEMBER 2021

Calibrating Sensor Networks for Environment Sensing Application: A Specific Case Study On Chemical Sensing

by

Abhishek Grover
Bharti School Of Telecommunication Technology and Management

submitted

in fulfilment of the requirements for the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

September 2021

THESIS CERTIFICATE

This is to certify that the thesis titled **Calibrating Sensor Networks for Environment Sensing Applications: A Specific Case Study On Chemical Sensing**, submitted by **Abhishek Grover**, to the Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Brejesh Lall
Department of Electrical Engineering
IIA-113, Bharti School of Telecom,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi-110016.

Place: New Delhi

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Brejesh Lall for his guidance and usefull insights. My supervisor introduced me to this application of environment sensing using sensor networks and motivated me to work on it. I want to thank him for his technical help, specially in the area of stochastic processes.

The work that I have done during my PhD has been published in various technical journals. I want to thank the editors and reviewers of these papers. Their technical comments and suggestions have helped me in improving my work.

I want to thank my family for allowing me to do what I liked and for their support in this endeavour. There were times when the way ahead was foggy but their motivation acted as the source of light. Thank You!

ABSTRACT

The process of measuring physical parameters at specific locations in the geographic region is known as environment sensing. The geographic region is monitored by deploying sensors at the requisite locations. The research problems can be classified as the problems at the device-level and the problems at the network-level. At the device-level, the variations in the sensor's response due to baseline drifts and sensitivity/selectivity variations are the major problems. Two device-level algorithms have been proposed to improve the accuracy of the sensor. A time series analysis-based approach is proposed for baseline drift removal. The response of a sensor is modeled as an autoregressive (AR) process associated with an observation zero-mean Gaussian noise. The parameters of the state-space model are learned using the calibration data. The variance of the Gaussian noise is estimated using an adaptive online method. The AR process equation and the observation equation are used to construct the adaptive Kalman filter. Each sensor response is passed through a separate Kalman filter, and then a regression technique is used to predict the sensor response. Numerical results show that the technique is effective in removing baseline drifts in sensors. Thus for a chemical sensor, the proposed factory calibration method can help in removing baseline drifts when the sensor is deployed on the field. Adapting the observation noise variance is the key to resolving the problem of baseline drifts in sensors. The second solution consists of modelling the fluctuations in the response of a chemical sensor. A stochastic differential equation model has been proposed. The model is developed by formulating the binding and unbinding reactions in the form of Bernoulli trials. The transition probability density function (pdf) of the number of binded analyte molecules X_t , at a given time instant t, has been evaluated using the Bernoulli trial representation. The obtained pdf is used to derive the Fokker Planck equation (FPE) for the stochastic process. The FPE has a direct correspondence with the desired SDE model. A maximum aposteriori (MAP) estimation method has been proposed for evaluating the parameters. This work presents a novel theoretical contribution to the chemical sensing literature as it models the operation of a surface based chemical sensor in form of a SDE. The proposed method is useful in compensating for the effects due to sensitivity/selectivity variations. At the network level, the major issues are finding optimal locations for sensor deployment and ensuring data reliability of the network. Two network-level algorithms have been proposed as a solution to these problems. A data-driven spatio-temporal model is proposed. The spatio-temporal representation of sensor values has been modeled as the sum of a systematic trend component and a residual process. The trend component is modelled as the sum of deterministic functions, and the residual component is modelled using support vector regression. The locations with maximum support vector count in the residual model are identified as optimal for the deployment of sensors. The method can be used for both static and dynamic deployments. The proposed method is data-driven and can be used to deploy sensors in all environment sensing applications. It only requires an initial data for the specific use case and the geographical covariates. An integro-difference equation (IDE) based spatio-temporal model is proposed for estimating missing data. The spatio-temporal field is modelled in the form of an adaptive linear state space model. The parameters of the model are identified using sensor values and are updated in a rolling window approach. Whenever missing data is encountered, the algorithm predicts the missing observations based on the learned state evolution equation. The proposed method can be used in all sensor network applications for estimating missing data as it makes no assumption about the application or on the underlying spatio-temporal field. In this thesis, the application of chemical sensing has been chosen to demonstrate the efficacy of the proposed methods. The drift correction and the deployment method has been tested on air pollution dataset. The utility of sensor modeling work is demonstrated in case of a pH sensor, molecular communication and in agent identification application. The missing data algorithm is applied on air pollution dataset and on an inertial sensor network used for human activity recognition.

KEYWORDS: Chemical Sensing, Baseline Drift Correction, Sensitivity/Selectivity variations, Sensor Deployment, Missing data

सारांश

भौगोलिक क्षेत्र में विशिष्ट स्थानों पर भौतिक मापदंडों को मापने की प्रक्रिया को पर्यावरण संवेदन के रूप में जाना जाता है। अपेक्षित स्थानों पर सेंसर लगाकर भौगोलिक क्षेत्र की निगरानी की जाती है। अनुसंधान समस्याओं को उपकरण-स्तर पर समस्याओं और नेटवर्क-स्तर पर समस्याओं के रूप में वर्गीकृत किया जा सकता है। डिवाइस-स्तर पर, बेसलाइन ड्रिफ्ट्स और संवेदनशीलता/चयनात्मकता भिन्नताओं के कारण सेंसर की प्रतिक्रिया में बदलाव प्रमुख समस्याएं हैं। सेंसर की सटीकता में सुधार के लिए दो डिवाइस-स्तरीय एल्गोरिदम प्रस्तावित किए गए हैं। आधारभूत बहाव को हटाने के लिए एक समय श्रृंखला विश्लेषण-आधारित दृष्टिकोण प्रस्तावित है। एक सेंसर की प्रतिक्रिया एक ऑटोरेग्रेसिव (एआर) प्रक्रिया के रूप में तैयार की जाती है जो एक अवलोकन शून्य-माध्य गाऊसी शोर से जुड़ी होती है। अंशांकन डेटा का उपयोग करके राज्य-अंतरिक्ष मॉडल के मापदंडों को सीखा जाता है। एक अनुकूली ऑनलाइन पद्धति का उपयोग करके गाऊसी शोर के विचरण का अनुमान लगाया जाता है। एआर प्रक्रिया समीकरण और अवलोकन समीकरण का उपयोग अनुकूली कलमन फिल्टर के निर्माण के लिए किया जाता है। प्रत्येक सेंसर प्रतिक्रिया को एक अलग कलमन फ़िल्टर के माध्यम से पारित किया जाता है, और फिर सेंसर प्रतिक्रिया की भविष्यवाणी करने के लिए एक प्रतिगमन तकनीक का उपयोग किया जाता है। संख्यात्मक परिणाम बताते हैं कि सेंसर में बेसलाइन ड्रिफ्ट को हटाने में तकनीक प्रभावी है। इस प्रकार एक रासायनिक सेंसर के लिए, प्रस्तावित फ़ैक्टरी कैलिब्रेशन विधि बेसलाइन ड्रिफ्ट को हटाने में मदद कर सकती है जब सेंसर को मैदान पर तैनात किया जाता है। सेंसर में बेसलाइन ड्रिफ्ट की समस्या को हल करने के लिए अवलोकन शोर विचरण को अपनाना महत्वपूर्ण है। दूसरे समाधान में रासायनिक सेंसर की प्रतिक्रिया में उतार-चढ़ाव की मॉडलिंग शामिल है। एक स्टोकेस्टिक अंतर समीकरण मॉडल प्रस्तावित किया गया है। मॉडल को बर्नौली परीक्षणों के रूप में बाध्यकारी और अनबाइंडिंग प्रतिक्रियाओं को तैयार करके विकसित किया गया है। बर्नौली परीक्षण प्रतिनिधित्व का उपयोग करके किसी निश्चित समय तत्काल t पर, बाध्य विश्लेषक अणुओं की संख्या X_t के संक्रमण संभाव्यता घनत्व फ़ंक्शन (पीडीएफ) का मूल्यांकन किया गया है। प्राप्त पीडीएफ का उपयोग स्टोकेस्टिक प्रक्रिया के लिए फोककर प्लैंक समीकरण (एफपीई) प्राप्त करने के लिए किया जाता है। एफपीई का वांछित एसडीई मॉडल के साथ सीधा पत्राचार है। मापदंडों के मूल्यांकन के लिए एक अधिकतम एपोस्टीरियरी (एमएपी) अनुमान पद्धति का प्रस्ताव किया गया है। यह काम रासायनिक संवेदन साहित्य में एक उपन्यास सैद्धांतिक योगदान प्रस्तुत करता है क्योंकि यह एक एसडीई के रूप में सतह आधारित रासायनिक सेंसर के संचालन को मॉडल करता है। प्रस्तावित विधि संवेदनशीलता/चयनात्मक भिन्नताओं के कारण होने वाले प्रभावों की क्षतिपूर्ति करने में उपयोगी है। नेटवर्क स्तर पर, प्रमुख मुद्दे सेंसर परिनियोजन के लिए इष्टतम स्थान ढूंढ रहे हैं और नेटवर्क की डेटा विश्वसनीयता सुनिश्चित कर रहे हैं। इन समस्याओं के समाधान के रूप में दो नेटवर्क-स्तरीय एल्गोरिदम प्रस्तावित किए गए हैं। एक डेटा-संचालित अनुपात-अस्थायी मॉडल प्रस्तावित है। सेंसर मूल्यों के अनुपात-अस्थायी प्रतिनिधित्व को एक व्यवस्थित प्रवृत्ति घटक और एक अवशिष्ट प्रक्रिया के योग के रूप में तैयार किया गया है। प्रवृत्ति घटक को नियतात्मक कार्यों के योग के रूप में तैयार किया गया है, और अवशिष्ट घटक को समर्थन वेक्टर प्रतिगमन का उपयोग करके तैयार किया गया है। अवशिष्ट मॉडल में अधिकतम सपोर्ट वेक्टर काउंट वाले स्थानों को सेंसर लगाने के लिए इष्टतम के रूप में पहचाना जाता है। विधि का उपयोग स्थिर और गतिशील दोनों परिनियोजन के लिए किया जा सकता है। प्रस्तावित विधि डेटा-चालित है और इसका उपयोग सभी पर्यावरण संवेदन अनुप्रयोगों में सेंसर लगाने के लिए किया जा सकता है। इसके लिए केवल विशिष्ट उपयोग के मामले और भौगोलिक सहसंयोजकों के लिए एक प्रारंभिक डेटा की

आवश्यकता होती है। लापता डेटा का अनुमान लगाने के लिए एक पूर्णांक-अंतर समीकरण (आईडीई) आधारित अनुपात-अस्थायी मॉडल प्रस्तावित है। अनुपात-अस्थायी क्षेत्र को एक अनुकूली रैखिक राज्य अंतरिक्ष मॉडल के रूप में तैयार किया गया है। मॉडल के मापदंडों को सेंसर मानों का उपयोग करके पहचाना जाता है और रोलिंग विंडो दृष्टिकोण में अद्यतन किया जाता है। जब भी लापता डेटा का सामना करना पड़ता है, तो एल्गोरिथ्म सीखे गए राज्य विकास समीकरण के आधार पर लापता टिप्पणियों की भविष्यवाणी करता है। प्रस्तावित विधि का उपयोग सभी सेंसर नेटवर्क अनुप्रयोगों में लापता डेटा का अनुमान लगाने के लिए किया जा सकता है क्योंकि यह आवेदन के बारे में या अंतर्निहित स्थानिक-अस्थायी क्षेत्र पर कोई धारणा नहीं बनाता है। इस थीसिस में प्रस्तावित विधियों की प्रभावकारिता को प्रदर्शित करने के लिए रासायनिक संवेदन के अनुप्रयोग को चुना गया है। वायु प्रदूषण डेटासेट पर बहाव सुधार और परिनियोजन विधि का परीक्षण किया गया है। एक पीएच सेंसर, आणविक संचार और एजेंट पहचान आवेदन के मामले में सेंसर मॉडलिंग कार्य की उपयोगिता का प्रदर्शन किया जाता है। लापता डेटा एल्गोरिदम वायु प्रदूषण डेटासेट और मानव गतिविधि पहचान के लिए उपयोग किए जाने वाले एक जड़त्वीय सेंसर नेटवर्क पर लागू होता है।

Contents

List of Figures

List of Tables

ABBREVIATIONS

WSN	Wireless Sensor Network
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression
KF	Kalman Filter
PCA	Principal Component Analysis
PLS	Partial Least Squares
RMSE	Root Mean Square Error
MAE	Mean of Absolute Error
R	R programming language
FFT	Fast Fourier Transform
DWT	Discrete Wavelet Transform
OSC	Orthogonal Signal Correction
SDE	Stochastic Differential Equation
PDF	Probability Density Function
FPE	Fokker Planck Equation
ODE	Ordinary Differential Equation
PSD	Power Spectral Density
MAP	Maximum Aposteriori
FES	Fluctuation Enhanced Sensing
SP	Signal Processing
TSA	Time Series Analysis
LRK	Ligand-Receptor Kinetics
ET	Estimation Theory
GT	Geometric Techniques
MBT	Model-Based Techniques
SC	Spatial Correlation
LR	Low-Rank based methods
DL	Deep Learning
SSM	State Space Models
RHS	Right Hand Side
LHS	Left Hand Side
pH	Potential of Hydrogen
ST	Spatio-Temporal

CFD	Computational Fluid Dynamics
RF	Radio Frequency
LUR	Land Use Regression
SVD	Singular Value Decomposition
SEOF	Smooth Empirical Orthogonal Functions
IDE	Integro-Difference Equation
KLT	Karhunen Loeve Theorem
EM	Expectation Maximization
MARE	Mean of Absolute Relative Error
ARE	Absolute Relative Error
KNN	K-Nearest Neighbour
GP	Gaussian Process
SVT	Singular Value Thresholding
LSTM	Long-Short Term Memory
KFST	Kalman Filter-based Spatio Temporal Regression