DATA DRIVEN RESOURCE OPTIMIZATION SCHEMES FOR EDGE DEVICES IN SMART IoT COMMUNICATIONS

SHARDA TRIPATHI

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2019

DATA DRIVEN RESOURCE OPTIMIZATION SCHEMES FOR EDGE DEVICES IN SMART IoT COMMUNICATIONS

by

SHARDA TRIPATHI

Department of Electrical Engineering

Submitted
in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2019

Certificate

This is to certify that the dissertation entitled **Data-Driven Resource Optimization Schemes for Edge Devices in Smart IoT Communications**, submitted by **Ms. Sharda Tripathi**, a Research Scholar, in the *Department of Electrical Engineering*, *Indian Institute of Technology Delhi*, *New Delhi*, *India*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by her under my supervision and guidance. The dissertation fulfills all requirements as per the regulations of this Institute and in my opinion has reached the standard needed for submission. Neither this dissertation nor any part of it has been submitted for any degree or academic award elsewhere.

Prof. Swades De
(Supervisor)
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi 110016, India

Acknowledgements

Foremost, my revered regards to the Almighty for bestowing on me the fortune and ability of being here to accomplish this dissertation. I have always been blessed with more than what I deserve. This day I promise to do my best in whatever is planned for me.

I would then like to express my sincere gratitude to my supervisor Prof. Swades De, who gave me the opportunity of pursuing PhD at IIT Delhi. His academic excellence, enthusiasm for research and zeal for perfection have always inspired me to work harder. This work would not have been possible without his guidance and support ever since the commencement of this journey till date. I am thankful for his patience in bearing with my incapabilities and constraints, yet believing in me and helping me come stronger. I feel privileged to be working under his mentorship.

My earnest thanks to committee members, Prof. I.N. Kar, Prof. S. Dharmaraja, and Prof. Nilanjan Senroy for their valuable feedback, constructive criticism and appreciation during my research tenure which contributed to the successful completion of this work. A special mention of thanks to my seniors and fellow lab mates from Communication Networks Research Group at IIT Delhi. I gratefully acknowledge their co-operation and support. They have been compassionate friends and I will always cherish their warmth.

I owe thanks to a very special person, my better half, Ankur, for his understanding of my goals and aspirations. I deeply appreciate his efforts in supporting all my endeavours. Special thanks to my sibling Dr. Laxmi, who helped me keep things in perspective during difficult times. I greatly value her love, care and encouragement throughout this experience. Also, heart felt regards to my in laws for their affection and moral support.

Finally, my profound gratitude to the most significant people in my life, my parents, Smt Saroj Tripathi, and Shri Raj Narayan Tripathi. Their unconditional love and valuable prayers have sustained me so far. I will forever be indebted to the pains and sacrifices they have made to shape my life. They are my strength and I dedicate this dissertation to them.

Sharda Tripathi

Abstract

Internet of Things (IoT) has gained tremendous popularity with the recent fast-paced technological advances in embedded programmable electronic and electro-mechanical systems, miniaturization, and their networking ability. IoT is expected to change the way of human activities by extensively networked monitoring, automation, and control. However, widespread application of IoT is associated with numerous challenges on communication and storage requirements, energy sustainability, and security. Also, IoT service quality requirements are application-specific. In this dissertation, we have identified novel methodologies to exploit the data-driven IoT framework for optimization of resources and development of context-aware cognitive applications in a massive machine type communication context. Through practical case studies, IoT application specific unique approaches and optimization techniques are proposed to reduce the data handling footprint, leading to communication bandwidth, cloud storage, and energy saving, without compromising service quality, thereby making them viable for wide-ranging adoption.

In the first part of dissertation, we introduce a novel data-driven framework for data pruning in wide area monitoring and control in smart grid, which is an emerging IoT application. Due to stringent latency constraints, packet losses and end-to-end transmission delay exceeding the permitted threshold values may jeopardize the stability of power grid. However, transient occurrences in the grid are relatively sparse, and much of the data is routine monitoring data having high redundancy. The proposed framework exploits the temporal correlatedness in the consecutive samples to dynamically prevent redundant Phasor Measurement Unit (PMU) data from being transmitted without affecting the quality of power grid health monitoring. The missing samples are predicted at the receiving end using ϵ-support vector regression model. It is noted that though PMU data has high temporal correlation, it actually characterizes a non-stationary process. Consequently, hyper-parameters of the prediction model are recomputed as necessary to maintain the accuracy and robustness of prediction. Also, for low runtime

complexity, some of the hyper-parameters are precomputed using empirical optimization of offline data characteristics. Appropriate performance indices are defined to quantify the performance of the proposed algorithm, and its computational latency is estimated via online execution using Simulink model. It is found that the proposed dynamic prediction algorithm selectively transmits the PMU data, thereby achieving up to 90% reduction in channel bandwidth requirement without affecting the quality of stability monitoring of the system. Further, comparison of the proposed algorithm with the closest competitive scheme demonstrates 73% and 60% better performance, respectively, in terms of power system health monitoring and bandwidth saving.

Subsequently, in the next part intelligent data pruning is investigated for automated electric metering in smart cities, which, similar to PMU data, is expected to increase the volume of network traffic exponentially. However, it does not possess the same nature of dynamics as the PMU data, and is more relaxed in terms of delay tolerance. It may be noted that as the granularity of sampling average power consumption in smart meter increases, compressibility of the data reduces, owing to irregular load profile. Besides, the data appears incoherent in time domain, though it can actually be represented by a sparsifying basis. Thus, adaptively choosing the sparsity over optimum batch size before data transmission can be utilized for substantial reduction in data volume. To this end, considering high resolution data at the smart meter, the problem of smart meter data characterization and reduction is addressed to achieve higher compression gains and reduced bandwidth requirement for data transmission from smart meter to the data collector. A novel Gaussian mixture based model is proposed for the characterization of high frequency smart meter data, which is used in evaluating the quality of data reduction at the smart meter. Further, an adaptive data reduction scheme using compressive sampling is devised to operate at the smart meter which achieves about 40% bandwidth saving in data transmission to the nearest collection center without any appreciable loss of information. Performance comparison of the proposed data reduction scheme with an existing competitive approach demonstrates noise robustness during data transmission. Additionally, to achieve the same order of Root Mean Square Error (RMSE), bandwidth saving with the proposed scheme is 12.8% and 7.4% higher, respectively, for data sampled at 1 second and 30 seconds. Real-time implementation of the proposed system level design is tested on smart meters deployed at IIT Delhi campus.

Finally, in the last part, channel-adaptive transmission strategies based on simple yet efficient channel prediction frameworks using stochastic modeling and data-driven learning of channel variability are proposed for sporadic but time-critical PMU data. The proposed channel prediction frameworks are accompanied with adaptive channel coding that assigns redundant symbols to the packet in accordance with the current channel state. A probing-based transmission scheme is also proposed which is considered as the benchmark for comparing the stochastic model-based and learning-based approaches. Through large-scale simulations, the prediction and packet loss performance is analyzed at varying Signal-to-Noise Ratio (SNR) and fading conditions. The results demonstrate that, for a given channel fading condition, packet loss probability of the proposed learning-based transmission closely matches with the benchmark scheme, while with the stochastic model-based prediction the loss probability is found to be 12.3% higher. However, the respective signalling overhead requirements are 38% and 98% lower with respect to the benchmark.

सार

इंटरनेट ऑफ थिंग्स (IoT) ने हाल ही में तेजी से उभरती तकनीकी के साथ जबरदस्त लोकप्रियता हासिल की है। बड़े पैमाने पर नेटवर्क की निगरानी, स्वचालन और नियंत्रण द्वारा एम्बेडेड प्रोग्राम इलेक्ट्रॉनिक और इलेक्ट्रो-मैकेनिकल सिस्टम में प्रगति, लघुकरण, और उनकी नेटवर्किंग क्षमता। IoT से मानवीय गतिविधियों के तरीके को बदलने की उम्मीद की जाती है । हालांकि, व्यापक IoT का अनुप्रयोग संचार और भंडारण पर कई चुनौतियों से जुड़ा है आवश्यकताओं, ऊर्जा स्थिरता और सुरक्षा। इसके अलावा, IoT सेवा की गुणवत्ता की आवश्यकताएं हैं आवेदन विशेष। इस शोध प्रबंध में, हमने शोषण करने के लिए उपन्यास पद्धति की पहचान की है संसाधनों के अनुकूलन और संदर्भ-जागरूक के विकास के लिए डेटा-संचालित IoT ढांचा एक बड़े पैमाने पर मशीन प्रकार के संचार संदर्भ में संज्ञानात्मक अनुप्रयोग। व्यावहारिक के माध्यम से मामले के अध्ययन, IoT आवेदन विशिष्ट अद्वितीय दृष्टिकोण और अनुकूलन तकनीक हैं संचार बैंडविड्थ, बादल के लिए अग्रणी डेटा हैंडलिंग पदचिह्न को कम करने का प्रस्ताव भंडारण, और ऊर्जा की बचत, सेवा की गुणवत्ता से समझौता किए बिना, जिससे उन्हें व्यवहार्य बनाया जा सके व्यापक अपनाने के लिए।

शोध प्रबंध के पहले भाग में, हम डेटा प्रूनिंग के लिए एक उपन्यास डेटा-संचालित रूपरेखा पेश करते हैं स्मार्ट ग्रिड में व्यापक क्षेत्र की निगरानी और नियंत्रण, जो एक उभरता हुआ IoT अनुप्रयोग है। कड़े विलंबता बाधाओं के कारण पैकेट के नुकसान और एंड-टू-एंड ट्रांसमिशन देरी से अधिक है अनुमत थ्रेशोल्ड मान पावर ग्रिड की स्थिरता को खतरे में डाल सकते हैं। हालाँकि, क्षणिक ग्रिड में होने वाली घटनाएं अपेक्षाकृत विरल हैं, और अधिकांश डेटा नियमित निगरानी है उच्च अतिरेक वाले डेटा। प्रस्तावित रूपरेखा अस्थायी सहसंबंध का शोषण करती है लगातार नमूनों में गतिशील पीएमयू डेटा को प्रसारित होने से रोकने के लिए गतिशील रूप से पावर ग्रिड स्वास्थ्य निगरानी की गुणवत्ता को प्रभावित किए बिना। लापता नमूने प्राप्त-अंत वेक्टर-प्रतिगमन प्रतिगमन मॉडल का उपयोग करके भविष्यवाणी की जाती है। यह उल्लेखनीय है कि हालांकि पीएमयू डेटा में उच्च अस्थायी सहसंबंध है, यह वास्तव में एक गैर-स्थिर है प्रक्रिया। नतीजतन, भविष्यवाणी मॉडल के हाइपर-मापदंडों को आवश्यक रूप से पुन: प्रतिष्ठित किया जाता है भविष्यवाणी की सटीकता और मजबूती को बनाए रखने के लिए। इसके अलावा, कम रनटाइम जटिलता के लिए, कुछ हाइपर-पैरामीटर ऑफ़लाइन डेटा के अनुभवजन्य अनुकूलन का उपयोग करके पूर्व-निर्मित हैं विशेषताएँ। उपयुक्त प्रदर्शन सूचकांकों के प्रदर्शन को निर्धारित करने के लिए परिभाषित किया गया है प्रस्तावित एल्गोरिथ्म, और इसकी कम्प्यूटेशनल विलंबता का उपयोग ऑनलाइन निष्पादन के माध्यम से किया जाता है सिमुलिंक मॉडल। यह पाया गया है कि प्रस्तावित गतिशील भविष्यवाणी एल्गोरिथ्म चुनिंदा प्रसारित करता है पीएमयू डेटा, जिससे चैनल बैंडविड्थ की आवश्यकता में 90% तक की कमी होती है सिस्टम की स्थिरता निगरानी की गुणवत्ता को प्रभावित किए बिना। इसके अलावा, की तुलना निकटतम प्रतिस्पर्धी योजना के साथ प्रस्तावित एल्गोरिथ्म 73% और 60% बेहतर प्रदर्शित करता है प्रदर्शन, क्रमशः, बिजली व्यवस्था स्वास्थ्य निगरानी और बैंडविड्थ बचत के संदर्भ में।

इसके बाद, अगले भाग में स्वचालित इलेक्ट्रिक के लिए बुद्धिमान डेटा प्रूनिंग की जांच की जाती है स्मार्ट शहरों में पैमाइश, जो कि पीएमयू डेटा के समान है, की मात्रा में वृद्धि की उम्मीद है नेटवर्क ट्रैफ़िक का तेज़ी से होना। हालांकि, यह गतिशीलता की समान प्रकृति के अधिकारी नहीं है पीएमयू डेटा के रूप में, और देरी सहिष्णुता के मामले में अधिक आराम है। यह ध्यान दिया जा सकता है कि के रूप में स्मार्ट मीटर में औसत बिजली की खपत का नमूना लेने की क्षमता बढ़ जाती है, संपीड़ितता डेटा के कम होने से अनियमित लोड प्रोफ़ाइल के कारण। इसके अलावा, डेटा असंगत प्रतीत होता है समय डोमेन में, हालांकि इसे वास्तव में एक स्पार्सिफाइंग आधार द्वारा दर्शाया जा सकता है। इस प्रकार, अनुकूल रूप से डेटा ट्रांसमिशन के उपयोग से पहले इष्टतम बैच आकार पर स्पर्सिटी का चयन करना डेटा की मात्रा में पर्याप्त कमी के लिए। यह

अंत करने के लिए, उच्च संकल्प डेटा पर विचार कर रहा है स्मार्ट मीटर, स्मार्ट मीटर डेटा लक्षण वर्णन और कमी की समस्या को संबोधित किया जाता है उच्च संप्रेषण लाभ और डेटा ट्रांसमिशन के लिए बैंडविड्थ की आवश्यकता को कम करना स्मार्ट मीटर से डेटा कलेक्टर तक। एक उपन्यास गाऊसी मिश्रण आधारित मॉडल प्रस्तावित है उच्च आवृत्ति स्मार्ट मीटर डेटा के लक्षण वर्णन के लिए, जिसका उपयोग मूल्यांकन करने में किया जाता है स्मार्ट मीटर पर डेटा की कमी की गुणवत्ता। इसके अलावा, एक अनुकूली डेटा कमी योजना का उपयोग कर कंप्रेसिव सैंपलिंग को स्मार्ट मीटर में संचालित करने के लिए तैयार किया जाता है जो लगभग 40% प्राप्त करता है किसी भी प्रशंसनीय के बिना निकटतम संग्रह केंद्र के लिए डेटा ट्रांसमिशन में बैंडविड्थ की बचत जानकारी का नुकसान। एक के साथ प्रस्तावित डेटा कटौती योजना की प्रदर्शन तुलना मौजूदा प्रतिस्पर्धात्मक दृष्टिकोण डेटा ट्रांसमिशन के दौरान शोर की मजबूती को दर्शाता है। साथ ही, आरएमएसई के समान आदेश को प्राप्त करने के लिए, प्रस्तावित योजना के साथ बैंडविड्थ की बचत है 12.8% और 7.4% अधिक, क्रमशः, 1 सेकंड और 30 सेकंड में सैंपल लिए गए डेटा के लिए। रियल टाइम आईआईटी दिल्ली कैंपस में तैनात स्मार्ट मीटरों पर प्रस्तावित सिस्टम लेवल डिजाइन के कार्यान्वयन का परीक्षण किया जाता है।

अंत में, अंतिम भाग में, सरल अभी तक कुशल पर आधारित चैनल-अनुकूली पारेषण रणनीतियों स्टोकेस्टिक मॉडलिंग और डेटा-चालित शिक्षण का उपयोग करके चैनल की भविष्यवाणी की रूपरेखा चैनल परिवर्तनशीलता छिटपुट लेकिन समय-महत्वपूर्ण पीएमयू डेटा के लिए प्रस्तावित है। प्रस्तावित चैनल भविष्यवाणी चौखटे अनुकूली चैनल कोडिंग के साथ होती है जो निरर्थक बताती है वर्तमान चैनल स्थिति के अनुसार पैकेट को प्रतीक। एक जांच-आधारित ट्रांसमिशन स्कीम भी प्रस्तावित है जिसे तुलना करने के लिए बेंचमार्क माना जाता है स्टोकेस्टिक मॉडल-आधारित और सीखने-आधारित दृष्टिकोण। बड़े पैमाने पर सिमुलेशन के माध्यम से, भविष्यवाणी और पैकेट नुकसान के प्रदर्शन का विश्लेषण एसएनआर और लुप्त होती स्थितियों में किया जाता है। परिणाम प्रदर्शित करते हैं कि, किसी दिए गए चैनल लुप्त होने की स्थिति के लिए, पैकेट नुकसान की संभावना प्रस्तावित लर्निंग-आधारित ट्रांसमिशन बेंचमार्क योजना के साथ निकटता से मेल खाता है, जबकि स्टोकेस्टिक मॉडल-आधारित भविष्यवाणी के साथ नुकसान की संभावना 12.3% अधिक पाई जाती है। हालांकि, संबंधित सिग्नलिंग ओवरहेड की आवश्यकताएं सम्मान के साथ 38% और 98% कम हैं बेंचमार्क के लिए।

Contents

List of Figures

List of Tables

List of Symbols

Chapter 2

f_i	Actual value of powerline frequency
$\hat{f}_i$	Predicted value of powerline frequency
f_{A_i}	Attribute vector of powerline frequency
d	Optimum lag value
l	Training length of ϵ- support vector regression model
w_j	Parameters of ϵ- support vector regression model
$\phi(f_j)$	Set of basis function in ϵ- support vector regression model
b	Offset value
w_{A_i}	Array of latest weights with length equal to optimum lag
$\phi_{f_{A_i}}$	Array of latest basis functions with length equal to optimum lag
ξ, ξ^*	Real-valued slack variables
$\alpha, \alpha^*, \eta, \eta^*$	Lagrangian multipliers
$K(f_{A_i}, f_{A_j})$	Radial basis kernel function
C, γ	Hyperparameters of ϵ- support vector regression model
OTL	Optimum training length
E	Difference between actual frequency value and predicted frequency value
ϵ	Acceptable error tolerance in ϵ- support vector regression model
S	Status notification transmitted from PMU to PDC
l_{dist}	Actual number of frequency samples belonging to disturbed state

$\hat{l}_{dist}$	Estimated number of frequency samples belonging to disturbed state
k	Folds of cross-validation
Δ	Large time interval
x	Number of C values in the search space
y	Number of γ values in the search space
l'	Number of step-ahead predictions

Chapter 3

n	Length of samples in the data window
k	Number of components in Gaussian mixture model
x_i	Average power consumption samples in the data window
w_j	Mixing coefficient corresponding to each Gaussian component
μ_j	Mean of j^{th} mixing coefficient in Gaussian mixture model
σ_j	Variance of j^{th} mixing coefficient in Gaussian mixture model
H	Hellinger's distance
ψ	Sparse basis matrix
f	Column vector of coefficients corresponding to sparse basis
m	Size of compressed data window
s	Sparsity of samples in the data window
ϕ	Sensing matrix
y	Vector of transmitted samples in the compressed data window

Chapter 4

f_D	Doppler frequency
T_s	Slot duration
$R(t)$	Received complex signal
$R_I(t), R_Q(t)$	In-phase and quadrature component of complex received signal

$Z(t)$	Received signal envelope
$\dot{Z}(t)$	Rate of change of signal envelope with repect to time
$\dot{\sigma}$	Variance of distribution of $\dot{Z}(t)$
n	Length of blind interval in slots
$\Phi_1(\cdot)$	CDF of standard univariate normal distribution
L	Number of levels characterized by channel state boundaries
$CS(\kappa)$	Channel state during κ^{th} blind slot
$\psi_i(\kappa)$	Probability distribution of channel states in the i^{th} level during κ^{th} blind slot
$\delta Z(1)$	Temporal variation of $Z(t)$ in the next time slot
$\dot{\sigma_\kappa}$	Variance of PDF of $Z(t)$ in the κ^{th} slot
x_i	Observed channel gain value
$\hat{x}_i$	Predicted channel gain value
d	Optimum number of lagged samples
x_{F_n}	Feature vector channel gain of length d
$\mathcal{K}$	Covariance kernel function
$f(x_{F_n})$	Function that maps input x_{F_n} to the label x_n
ϵ_n	Normally distributed noise with mean 0 and variance σ^2
X_F	Feature matrix of Gaussian process regression model
X_α	Label vector of Gaussian process regression model
$\tilde{\mu}, \tilde{\sigma}^2$	Mean and variance of posterior function
$\hat{\mu}, \hat{\sigma}^2$	Mean and variance of predictive posterior function
a	Training length of Gaussian process regression model
b	Number of step-ahead predictions
k	Number of information symbols
c_i	Block length corresponding to i^{th} level in adaptive RS coding scheme
m	Length of each information symbol in bits

c_{max}	Maximum block length in adaptive RS coding scheme
c_{min}	Minimum block length in adaptive RS coding scheme
f_c	Carrier frequency
p_f	False prediction probability
p_{se}	Symbol error probability
p_l	Packet loss probability
BW_c	Bandwidth consumption
O_s	Signalling overhead