MACHINE LEARNING FOR EFFECTIVE CONNECTIVITY IN HUMAN BRAIN USING NEUROIMAGING DATA

SHILPA DANG

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2018

MACHINE LEARNING FOR EFFECTIVE CONNECTIVITY IN HUMAN BRAIN USING NEUROIMAGING DATA

by

SHILPA DANG

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

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

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JULY 2018

To my family and parents

CERTIFICATE

This is to certify that the thesis entitled, "**Machine Learning for Effective Connectivity in Human Brain using Neuroimaging Data**", being submitted by **Mrs. Shilpa Dang** for the award of the degree of **Doctor of Philosophy** is a record of bonafide research work carried out by her in the department of Electrical Engineering of the Indian Institute of Technology Delhi.

Mrs. Shilpa Dang has worked under our guidance and supervision and has fulfilled the requirements for the submission of this thesis, which to our knowledge has reached the requisite standard. The results obtained here, have not been submitted to any other University or Institute for the award of any degree.

Dr. Santanu Chaudhury
Professor,
Dept. of Electrical Engineering,
Indian Institute of Technology Delhi,
New Delhi, India.
Director,
Central Electronics Engineering Research Institute,
Pilani, India.

Dr. Brejesh Lall
Associate Professor,
Dept. of Electrical Engineering,
Indian Institute of Technology Delhi,
New Delhi, India.

Dr. Prasun Kumar Roy
Professor,
Computational Neuroscience & Neuro-Imaging Laboratory,
School of Bio-Medical Engineering,
Indian Institute of Technology BHU,
Varanasi, India.

Place: New Delhi **Dated:**

Acknowledgements

I express my sincere thanks to my PhD supervisors Dr. Santanu Chaudhury, Dr. Brejesh Lall, and Dr. Prasun Kumar Roy for their able guidance, and also to my other research committee members, Dr. Sumantra Dutta Roy, Dr. Shiv Dutt Jhosi, and Dr. Sneh Anand for their valuable comments and feedback which helped to improve the work.

My heartfelt thanks to Dr. Brejesh Lall, my master's supervisor, for his most valuable discussion which culminated in the selection of my PhD topic.

I thank the reviewers of the manuscripts, which were communicated to peer-reviewed journals, for providing constructive comments that improved the quality of the thesis.

I thank my fellow lab mates Deepti Goyal, Meera Thapar, Nisha Pahal, and lab staff Manrika Lochan in the Multimedia Lab for their help during the study.

Most importantly, I thank the 1000 Functional Connectomes Project, Functional Imaging Biomedical Informatics Research Network, International Neuroimaging Data-Sharing Initiative, organizers of the International ADHD-200 competition, and the Neurobureau for publicly sharing their database without which this thesis would not have been possible.

I am grateful to my mother and father who have helped me by taking care of my daughter during my entire PhD so that I could spend time in lab and at home for my research work. Also, I am grateful to my husband for encouraging and being a constant support throughout in good and bad times and my daughter for being an understanding kid at such a young age.

Last but not the least, I thank the Almighty who gave me the much necessary endurance to tackle hardships that came upon during the study and life.

Place: New Delhi Shilpa Dang

Abstract

This dissertation shows that a machine learning approach for quantitative and qualitative learning of effective brain connectivity not only improves the exploration of the intricate space of brain connectivity but also provides clinically meaningful measures, quantifying normal and pathological variation in the brain. To this end, we introduce a set of computational and mathematical frameworks and algorithms to explore, understand, and characterize neuroimaging (functional and/or diffusion MRI) data derived effective brain connectivity.

We contribute towards learning brain connectivity under various categories, namely, complete fMRI data, incomplete fMRI data, and multimodal imaging data. The face validity of all proposed methods is established over synthetic data and then over experimental data. We also contribute towards assessing validity of assumptions made by the statistical modelling approach, namely, multivariate autoregressive framework for learning effective connectivity from fMRI data and thus paving a way for employment of machine learning techniques towards exploration of brain connectivity.

Under complete data assumption for fMRI data, our contributions are twofold. First, we propose employing high-order dynamic Bayesian network (HO-DBN), as with the advent of fast sampling rates dependency beyond first-order is a possibility. However, a fundamental problem faced in the structure learning of HO-DBN is high computational-burden and low accuracy by the existing heuristic search techniques used for learning effective connectivity from fMRI. Second, we propose using dynamic programming (DP) principle along with integration of properties of scoring-function in a way to reduce search space for structure learning of HO-DBNs and finally, for learning effective connectivity from fMRI. The proposed exact search-&-score learning approach HO-DBN-DP is an extension of the technique which was originally devised for learning a BN's structure from static data (Singh and Moore, 2005). The algorithm reaches globally-optimal solution in appreciably reduced time-complexity than the static counterpart due to integration of properties. The proof of optimality is provided.

fMRI signal is an indirect measure of underlying neuronal activity. Moreover, there are lost samples and thus low temporal resolution due to down-sampling present in the acquisition process of fMRI signal. Under incomplete data assumption for fMRI data, again our contributions are twofold. First, we propose a novel autoregressive hidden Markov model with missing data (AR-HMM-md) framework which aims at addressing aforementioned issues while allowing accurate capturing of fMRI time series characteristics. The proposed work models unobserved neuronal activity over time as sequence of discrete hidden states and lost data over time as variables, estimates their values by joint optimization given fMRI observation sequence, and shows how exact inference can be obtained with missing fMRI data under the "Missing not at Random" (MNAR) mechanism. This mechanism requires explicit modelling of the missing data along with the observed data. The proposed model captured the fMRI characteristics efficiently and thus converged to better posterior probability resulting into higher classification accuracy over existing latent models. Second, we propose use of AR-HMM-md in dynamically multi-linked (DML) framework for learning effective connectivity using multiple fMRI time series. This framework overcomes the disadvantage of low-temporal resolution, improves cross-validation accuracy significantly, and leads to reliable estimates of effective connectivity due to presence of missing data variables and autoregressive process.

By definition effective connectivity is "the causal influence exerted by one neuronal group on another" which is constrained by anatomical connectivity (axonal connections). Anatomical connectivity is necessary for effective connectivity but does not fully determine it, because synaptic communication occurs dynamically in a context-dependent fashion. Although there is a vast emerging evidence of structure-function relationship using multimodal imaging studies, till date only a few studies have done joint modelling of the two modalities: fMRI and DTI. We aim to propose a unified probabilistic framework that combines information from both sources to learn effective connectivity using DBNs. Specifically, we propose a novel anatomically-informed (AI) score that evaluates fitness of a given connectivity structure to both DTI and fMRI data simultaneously. The AI score is employed

in structure learning of DBN given the multimodal imaging data. Multimodal analysis provide a more reliable basis for differentiating brain under abnormal/diseased conditions than single modality analysis.

The results show that the machine learning approaches, proposed in this dissertation, under different data categories are computationally more efficient, reliable and robust than existing approaches for learning effective connectivity estimates of human brain.

सार

इस शोध प्रबंध से पता चलता है कि प्रभावी मस्तिष्क कनेक्टिविटी की मात्रात्मक और गुणात्मक सीखने के लिए एक मशीन सीखने का दृष्टिकोण न केवल मस्तिष्क कनेक्टिविटी की जटिल जगह की खोज में सुधार करता है बल्कि मस्तिष्क में सामान्य और पैथोलॉजिकल भिन्नता को मापने, नैदानिक अर्थपूर्ण उपायों को भी प्रदान करता है। इस अंत में, हम प्रभावी मस्तिष्क कनेक्टिविटी व्युत्पन्न न्यूरोइमेजिंग (कार्यात्मक और / या प्रसार एमआरआई) डेटा का पता लगाने, समझने और विशेषता के लिए कम्प्यूटेशनल और गणितीय ढांचे और एल्गोरिदम का एक सेट पेश करते हैं।

हम विभिन्न श्रेणियों, अर्थात्, पूर्ण एफएमआरआई डेटा, अधूरा एफएमआरआई डेटा, और बहुआयामी इमेजिंग डेटा के तहत मस्तिष्क कनेक्टिविटी सीखने में योगदान करते हैं। सभी प्रस्तावित तरीकों का चेहरा वैधता सिंथेटिक डेटा और उसके बाद प्रयोगात्मक डेटा पर स्थापित किया गया है। हम सांख्यिकीय मॉडलिंग दृष्टिकोण, अर्थात्, एफएमआरआई डेटा से प्रभावी कनेक्टिविटी सीखने के लिए बहुआयामी स्वैच्छिक रूपरेखा और इस प्रकार मस्तिष्क कनेक्टिविटी की खोज के लिए मशीन लर्निंग तकनीकों के रोजगार के लिए एक रास्ता तय करने के लिए प्रतिबद्धता की वैधता का आकलन करने में योगदान देते हैं।

एफएमआरआई डेटा के लिए पूर्ण डेटा धारणा के तहत, हमारे योगदान दो गुना हैं। सबसे पहले, हम हाई-ऑर्डर गतिशील बेयसियन नेटवर्क (एचओ-डीबीएन) को नियोजित करने का प्रस्ताव देते हैं, क्योंकि तेजी से नमूना दरों के आगमन के साथ पहले क्रम से परे निर्भरता एक संभावना है। हालांकि, एचओ-डीबीएन की संरचना सीखने में सामना की जाने वाली एक मौलिक समस्या एफएमआरआई से प्रभावी कनेक्टिविटी सीखने के लिए उपयोग की जाने वाली मौजूदा हेरिस्टिक खोज तकनीकों द्वारा उच्च कम्प्यूटेशनल-बोझ और कम सटीकता है। दूसरा, हम एचओ-डीबीएन की संरचना सीखने के लिए खोज स्थान को कम करने और अंततः एफएमआरआई से प्रभावी कनेक्टिविटी सीखने के लिए स्कोरिंग-फ़ंक्शन के गुणों के एकीकरण के साथ गतिशील प्रोग्रामिंग (डीपी) सिद्धांत का उपयोग करने का प्रस्ताव देते हैं। प्रस्तावित सटीक खोज - और - स्कोर सीखने का दृष्टिकोण एचओ-डीबीएन-डीपी तकनीक का एक विस्तार है जिसे मूल रूप से स्थिर डेटा (सिंह और मूर, 2005) से बीएन की संरचना सीखने के लिए तैयार किया गया था। एल्गोरिदम गुणों के एकीकरण के कारण स्थिर समकक्ष की तुलना में समय-जटिलता को कम करने में वैश्विक स्तर पर इष्टतम समाधान तक पहुंचता है। इष्टतमता का सबूत प्रदान किया जाता है।

एफएमआरआई सिग्नल अंतर्निहित न्यूरोनल गतिविधि का एक अप्रत्यक्ष उपाय है। इसके अलावा, एफएमआरआई सिग्नल की अधिग्रहण प्रक्रिया में डाउन-नमूनाकरण के कारण खोए गए नमूने और इस प्रकार कम अस्थायी संकल्प हैं। एफएमआरआई डेटा के लिए अपूर्ण डेटा धारणा के तहत, फिर से हमारे योगदान दो गुना हैं। सबसे पहले, हम लापता डेटा (एआर-एचएमएम-एमडी) ढांचे के साथ एक उपन्यास स्वैच्छिक छिपे हुए मार्कोव मॉडल का प्रस्ताव देते हैं जिसका लक्ष्य उपरोक्त मुद्दों को संबोधित करना है जबकि एफएमआरआई समय श्रृंखला विशेषताओं के सटीक कैप्चरिंग की अनुमति है। प्रस्तावित कार्य मॉडल अलग-अलग छिपे हुए राज्यों के अनुक्रम के रूप में समय के साथ अप्रसन्न न्यूरोनल गतिविधि और चर के रूप में समय के साथ खोए गए डेटा, एफएमआरआई अवलोकन अनुक्रम दिए गए संयुक्त अनुकूलन द्वारा उनके मूल्यों का अनुमान लगाते हैं, और दिखाते हैं कि "गुमशुदा एफएमआरआई डेटा" के साथ गायब एफएमआरआई डेटा के साथ कितना सटीक अनुमान प्राप्त किया जा सकता है रैंडम "(एमएनएआर) तंत्र पर नहीं। इस तंत्र के लिए मनाए गए डेटा के साथ लापता डेटा के स्पष्ट मॉडलिंग की आवश्यकता है। प्रस्तावित मॉडल ने एफएमआरआई विशेषताओं को कुशलता से कब्जा कर लिया और इस प्रकार मौजूदा पूर्ववर्ती संभावनाओं के लिए अभिसरण किया जिसके परिणामस्वरूप मौजूदा लेटेस्ट मॉडलों पर उच्च वर्गीकरण सटीकता हुई। दूसरा, हम कई एफएमआरआई समय श्रृंखला का उपयोग कर प्रभावी कनेक्टिविटी सीखने के लिए गतिशील रूप से बहु-लिंक (डीएमएल) ढांचे में एआर-एचएमएम-एमडी का उपयोग करने का प्रस्ताव देते हैं। यह ढांचा कम-अस्थायी संकल्प के

नुकसान पर विजय प्राप्त करता है, क्रॉस-सत्यापन सटीकता में काफी सुधार करता है, और गायब डेटा चर और स्वत: आक्रामक प्रक्रिया की उपस्थिति के कारण प्रभावी कनेक्टिविटी के विश्वसनीय अनुमानों की ओर जाता है।

परिभाषा के अनुसार प्रभावी कनेक्टिविटी "एक न्यूरोनल समूह द्वारा दूसरे पर लागू कारण प्रभाव" है जो रचनात्मक कनेक्टिविटी (अक्षीय कनेक्शन) से बाधित है। प्रभावी कनेक्टिविटी के लिए शारीरिक कनेक्टिविटी आवश्यक है लेकिन इसे पूरी तरह से निर्धारित नहीं करता है, क्योंकि संदर्भ-निर्भर फैशन में सिनैप्टिक संचार गतिशील रूप से होता है। यद्यपि मल्टीमोडाल इमेजिंग स्टडीज का उपयोग करके स्ट्रक्चर-फ़ंक्शन रिलेशनशिप के विशाल उभरते सबूत हैं, आज तक केवल कुछ अध्ययनों ने दो तरीकों के संयुक्त मॉडलिंग किए हैं: एफएमआरआई और डीटीआई। हमारा उद्देश्य एक एकीकृत संभाव्य ढांचा का प्रस्ताव देना है जो डीबीएन का उपयोग करके प्रभावी कनेक्टिविटी सीखने के लिए दोनों स्रोतों से जानकारी को जोड़ती है। विशेष रूप से, हम एक उपन्यास शारीरिक रूप से सूचित (एआई) स्कोर का प्रस्ताव देते हैं जो डीटीआई और एफएमआरआई डेटा दोनों को एक साथ कनेक्टिविटी संरचना की फिटनेस का मूल्यांकन करता है। एआई स्कोर नियोजित है।

डीबीएन की संरचना सीखने में बहुआयामी इमेजिंग डेटा दिया गया। मल्टीमोडाल विश्लेषण एकल मॉडेल विश्लेषण की तुलना में असामान्य / रोगग्रस्त स्थितियों के तहत मस्तिष्क को अलग करने के लिए एक अधिक विश्वसनीय आधार प्रदान करता है।

नतीजे बताते हैं कि विभिन्न शोध श्रेणियों के तहत, इस शोध प्रबंध में प्रस्तावित मशीन लर्निंग दृष्टिकोण मानव मस्तिष्क के प्रभावी कनेक्टिविटी अनुमानों को सीखने के लिए मौजूदा दृष्टिकोणों की तुलना में कम्प्यूटेशनल रूप से अधिक कुशल, भरोसेमंद और मजबूत हैं।

Table of contents

List of Figures

List of Tables