

Xishuang Dong

Assistant Professor, Center for Computational Systems Biology,
Department of Electrical and Computer Engineering,
Prairie View A&M University, Prairie View, Texas 77446
Office: +1(936)-261-9865, E-mail: xidong@pvamu.edu

Education

Harbin Institute of Technology

Ph.D., Harbin, Heilongjiang Province, China

Computer Engineering

04/2008 – 07/2013

Harbin Engineering University

M.S., Harbin, Heilongjiang Province, China

Computer Engineering

09/2005 – 03/2008

Harbin University of Science and Technology

B.S., Harbin, Heilongjiang Province, China

Computer Engineering

09/2001 – 07/2005

Professional Experience

Assistant Professor

10/2018 – Now

Prairie View A&M University

Prairie View, Texas, USA

- Presenting researches in conferences;
 - "Multiple Time-Series Data Analysis for Rumor Detection on Social Media", IEEE International Conference on Big Data, December 10-13, 2018, Seattle, WA, USA;
- Teaching Courses
 - Undergraduate
 - ◆ Machine Learning (Fall 2018);
 - ◆ Computer Application in Engineering (Fall 2018, Spring 2019, Fall 2019);
 - Graduate
 - ◆ Statistical Learning for Big Data (Fall 2019);

Postdoc

06/2016 – 09/2018

Prairie View A&M University

Prairie View, Texas, USA

- Guiding Students to Win the CyberC 2016 IEEE Big Data Analytics Competition organized by the IEEE Big Data Initiative held in Chengdu, China (10/13/2016–10/14/2016);
- Organizing an intensive training on "Deep Learning Tutorial and Hands-on Training" during The Second Workshop of Mission-Critical Big Data Analytics (MCBDA 2017);
- Supervising graduate thesis
 - Shanta Chowdhury, "Multi-task Bi-directional LSTM RNN for Named Entity Recognition on Electronic Medical Records", Department of Electrical and Computer Engineering, Prairie View A&M University, 2018.
 - Safat Mahmood, "CNN Based Forecasting of Hourly Rental Bike Demand in New York City", Department of Electrical and Computer Engineering, Prairie View A&M University, 2017.
- Presenting researches of deep learning in conferences;
 - "A Multitask Bi-directional RNN Model for Named Entity Recognition on Electronic Medical Records", International Conference on Intelligent Biology and Medicine (ICIBM 2018) June 10-12,

2018, Los Angeles, CA, USA.

- "A CNN Based Bagging Learning Approach to Short-Term Load Forecasting in Smart Grid", In the third IEEE International Conference on Cloud and Big Data Computing (CBDCom), 2017;
- "A multiclass classification method based on deep learning for named entity recognition in electronic medical records." In New York Scientific Data Summit (NYSDS), 2016;
- Teaching courses
 - Deep Learning (Spring 2017);
 - Machine Learning (Fall 2016);

Assistant Professor

Xinyang Normal University

07/2014 – 05/2016

Xinyang, Henan, China

- Teaching Courses
 - Data Structure;
 - Database Application Techniques;
 - Programming Languages: C++;
 - Software Testing;

Peer Reviewer

- Journal
 - Transactions on Asian and Low-Resource Language Information (TALLIP)
 - IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI)
 - Briefings in Bioinformatics
 - Frontiers
 - IEEE Access
- Conference
 - IEEE BIBM (2018, 2019)
 - MILCOM (2018, 2019)
 - CNB-MAC 2019

Awards

International Competitions

- CyberC 2016 IEEE Big Data Analytics Competition
 - Competition summary: to promote the awareness and analytic technology of Big Data, the IEEE Big Data Initiative (BDI) organized this Data Analytics Competition. It attracted 6 research groups from universities and companies all over the world.
 - Role: Supervisor;
 - Task: Mining knowledge from Bike Rental data for improving Bike sharing services in New York;
 - **Winner of Bike Rental Forecasting;**
- TREC-2010 Blog Track
 - Competition summary: this Blog track aimed to investigate the information seeking behavior in the blogosphere. It attracted 12 research groups from top universities such as Stanford NLP group.
 - Role: Team member;
 - Task: Faceted Blog Distillation Task and Top Stories Identification Task;
 - **Winner of the task of Faceted Blog Distillation;**

Competitions in China

- Chinese Opinion Analysis Evaluation 2011
 - Competition summary: it was organized by Fuzhou University, Chinese Academy of Sciences, and Shanghai Jiao Tong University. The goal is to improve the performance of opinion analysis in different domains and contexts. It attracted 20 research groups from top universities and companies in China.
 - Role: Team member;
 - Task: 4 sentiment analysis tasks;
 - **Winner of the task of mining sentiment words in different domains;**
- Chinese Opinion Analysis Evaluation 2009
 - Competition summary: it was organized by Institute of Chinese Academy of Sciences, Shanghai Jiao Tong University, and Fudan University. Its motivation is to promote researches of recognizing Chinese sentiment words and sentences, and to improve the performance of opinion analysis. It attracted 17 research groups from top universities such as Beijing University in China.
 - Role: Team leader;
 - Task: 5 sentiment analysis tasks;
 - **Winner of two tasks: (1) recognizing and classifying sentiment words; (2) recognizing emotions of sentences;**

Research Experience

Research Fields

- **Deep Learning Theory and Application**
 - Distributed deep transfer learning;
 - Semi-supervised deep learning;
 - Ensemble deep learning;
 - Deep learning for streaming data analysis;
 - Time-series data analysis via deep learning
 - Complex Neural Networks;
- **Machine Learning for Biomedical Big Data Analytics**
 - Single-cell data Analysis via semi-supervised Deep Learning;
 - Feature Selection of biomedical data via recurrent neural networks;
 - Information extraction of electronic medical records
- **Deep Learning for Cyber Physical Systems**
 - Distributed deep learning on IoT;
 - Deep learning based seismic data analysis;
 - Deep learning based underwater communication data analysis;
 - Load forecasting via deep learning for smart grid;
- **Natural Language Processing via Machine Learning**
 - Sentiment analysis via deep learning;
 - Neural machine translation;
 - Word representation learning;
 - Dependency parsing;
 - Text classification and clustering;
- **Image Processing**
 - Image classification;
 - Image semantic segmentation;
 - Object detection and tracking;

Research Projects

Single-cell Data Analysis via Deep Learning

09/2019 – present

- Task
 - Feature Selection on Single-Cell Data;
 - Single-Cell Subtype Classification;
- Proposed Methods
 - Recurrent Convolutional Neural Networks Based Semi-supervised Learning;
- Related papers have been accepted in conferences such as BigData 2019;

Fake News Detection via Deep Learning

09/2018 – present

- Task
 - Classify News into Fake and True;
 - News Representation
- Proposed Methods
 - Ensemble Learning via Deep Learning
 - Convolutional Neural Networks Based Semi-supervised Learning;
- Related papers have been accepted in conferences such as MILCOM 2019;

Word-agent Based Autonomous Reinforcement Learning for Sentence Similarity Measurement

Sponsored by National Natural Science Foundation of China

01/2010 – 12/2012

- Task
 - Building the adaptive immune system simulator;
 - Constructing word-agent learning model;
- Proposed Methods
 - Adaptive immune simulation via plasma cell negative regulation;
 - Multi-word-agent Autonomous Learning;
- Related papers have been published in AI top conferences such as AAAI 2014 and Journals;

Sentiment Analysis on Web Texts for Recognizing Abnormal Events

Sponsored by National Natural Science Foundation of China

01/2010 – 12/2012

- Task
 - Sentiment analysis on words and sentences;
 - Sentiment analysis on Microblog;
- Proposed Methods
 - Set-Similarity Joins Based Semi-supervised Sentiment Analysis;
 - HIT_LTRC at TREC 2010 Blog Track: Faceted Blog Distillation;
- Related papers have been published in AI conferences such as ICONIP2012;

Building Electronic Medical Record Corpus

Second Affiliated Hospital of Harbin Medical University, China

01/2013 – 07/2013

- Collaborating with doctors to annotate electronic medical records;
- Modifying annotation guidelines in terms of feedbacks from the doctors;

Microblog Sentiment Analysis System

- Constructing sentiment lexicon and sentence corpus;
- Building sentiment analysis models via supervised machine learning to recognizing sentiment polarities of words, sentences and microblogs;

Software tools

- **Deep Learning Platform:** TensorFlow, Keras;
- **Programming Languages:** Python, C++, C, C#, Java;
- **Database:** Microsoft SQL, Oracle;
- **Big Data:** Spark, Slurm;
- **Programming IDE:** .Net, Eclipse;
- **Website Development:** ASP.Net;

Publications

Doctoral Dissertation

- **Title:** Immune Multi-word-agent Autonomy Learning Based Sentiment Analysis
- **Supervisors:** Prof. Yi Guan
- **Abstract:** This research involves three parts of developing novel methods to complete sentiment analysis of texts. First part is to present a semi-supervised model based on set-similarity joins to recognize sentiment polarities of words and sentences via Maximum Entropy model. In the second part, inspired by similarities between the human immune system and the human language system, an autonomy learning model via adaptive immune theories is constructed in three steps: (1) building an artificial immune system based on the plasma negative regulation mechanism by multi-agent system modeling; (2) presenting a word-agent model based on similarities between words and immune cells and molecules on the artificial immune system, where word attributes are simulated as receptors of B cell and Antigen; (3) proposing a multi-word-agent autonomy learning model on the system through simulating words in sentences as immune cells and molecules to optimize relations between words, which can learn continuously by interactions between immune word-agents in the on-line learning manner; The third part applies the autonomy learning model to complete sentiment analysis on collocations and dependency parsing, and experiments show that the performance could be improved significantly.

Journal

Published

1. **Xishuang Dong**, Chowdhury, Shanta, Lijun Qian, Xiangfang Li, Yi Guan, Jinfeng Yang, and Qiubin Yu. "Deep learning for named entity recognition on Chinese electronic medical records: combining deep transfer learning with multitask bi-directional lstm rnn. " PLOS ONE. 2019.
2. Chowdhury, Shanta, **Xishuang Dong**, Lijun Qian, Xiangfang Li, Yi Guan, Jinfeng Yang, and Qiubin Yu. "A multitask bi-directional RNN model for named entity recognition on Chinese electronic medical records." BMC bioinformatics 19, no. 17 (2018): 499.
3. Huang, Lei, **Xishuang Dong**, and T. Edward Clee, "A scalable deep learning platform for identifying geologic features from seismic attributes." The Leading Edge, Vol. 36 no. 3 pp. 249-256, 2017.
4. Yang, Jinfeng, Yi Guan, and **Xishuang Dong**. "A Multi-word-agent Autonomous Learning Model for Regulating Word Combination Strength." International Journal of Multimedia and Ubiquitous Engineering 10.4 (2015): 355-366.
5. **Xishuang Dong**, Xinbo Lv, Yi Guan, and Jinfeng Yang. Multi-word-Agent Autonomy Learning Based on

Adaptive Immune Theories. *JDCTA: International Journal of Digital Content Technology and its Applications*, 2013, 7(3): 723-745.

Conference

Published

1. **Xishuang Dong**, Uboho Victor, Shanta Chowdhury, Lijun Qian. "Deep Two-path Semi-supervised Learning for Fake News Detection." arXiv preprint arXiv:1906.05659 (2019).
2. Kotteti, Chandra Mouli Madhav, **Xishuang Dong**, and Lijun Qian. "Multiple Time-Series Data Analysis for Rumor Detection on Social Media." In 2018 IEEE International Conference on Big Data (Big Data), pp. 4413-4419. IEEE, 2018.
3. **Dong, Xishuang**, Hsiang-Huang Wu, Yuzhong Yan, and Lijun Qian. "Hierarchical Transfer Convolutional Neural Networks for Image Classification." arXiv preprint arXiv:1804.00021 (2018).
4. Kotteti, Chandra Mouli Madhav, **Xishuang Dong**, Na Li, and Lijun Qian. "Fake News Detection Enhancement with Data Imputation." In 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), pp. 187-192. IEEE, 2018.
5. Shanta Chowdhury, **Xishuang Dong**, Lijun Qian, Xiangfang Li, Yi Guan, Jinfeng Yang, and Qiubin Yu. "A Multitask Bi-directional RNN Model for Named Entity Recognition on Electronic Medical Records", International Conference on Intelligent Biology and Medicine, 2018.
6. **Xishuang Dong**, Shanta Chowdhury, Lijun Qian, Yi Guan, Jinfeng Yang, Qiubin Yu. "Transfer Bi-directional LSTM RNN for Named Entity Recognition in Chinese Electronic Medical Records", 19th International Conference on E-health Networking, Application & Services, 2017.
7. **Xishuang Dong**, Lijun Qian, Lei Huang. "A CNN Based Bagging Learning Approach to Short-Term Load Forecasting in Smart Grid", IEEE CBDCOM Conference Proceedings, 2017.
8. **Dong, Xishuang**, Lijun Qian, and Lei Huang. "Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach." 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), 2017.
9. Dong, Xishuang, Lijun Qian, Yi Guan, Lei Huang, Qiubin Yu, and Jinfeng Yang. "A multiclass classification method based on deep learning for named entity recognition in electronic medical records." In 2016 New York Scientific Data Summit (NYSDDS), pp. 1-10.
10. Yang, Jinfeng, **Xishuang Dong**, and Yi Guan. "Words Are Analogous To Lymphocytes: A Multi-Word-Agent Autonomous Learning Model." Progress in Systems Engineering. Springer, Cham, 2015. 755-760.
11. Yang, Jinfeng, Yi Guan, **Xishuang Dong**, and Bin He. "Representing Words as Lymphocytes." In AAAI, pp. 3146-3147. 2014.
12. **Dong, Xishuang**, Yi Guan, and Jinfeng Yang. "Plasma Cell Negative Regulation Mechanism Based Artificial Immune System." Proceedings of the 2013 Fourth International Conference on System Science, Engineering Design and Manufacturing Informatization (icsem 2013). IEEE Computer Society, 2013.
13. **Xishuang Dong**, Qibo Zou, and Yi Guan. Set-Similarity Joins Based Semi-supervised Sentiment Analysis. In Proceedings of the 19th international conference on Neural Information Processing (ICONIP'12), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 2012, pp. 176-183.
14. Li, Yun, **Xishuang Dong**, and Yi Guan. "HIT_LTRC at TREC 2011 Microblog Track." In TREC. 2011.

15. Yang, Jinfeng, **Xishuang Dong**, Yi Guan, Chengzhen Huang, and Sheng Wang. "HIT_LTRC at TREC 2010 Blog Track: Faceted Blog Distillation." In TREC. 2010.
16. **Dong, Xishuang**, Xiaodong Chen, Yi Guan, Zhiming Yu, and Sheng Li. "An overview of learning to rank for information retrieval." In Computer Science and Information Engineering, 2009 WRI World Congress on, vol. 3, pp. 600-606. IEEE, 2009.

Accepted

6. Peng S., Rath S., Vuong C., Bollam S., Eschbacher J., **Dong X.**, Mehta S., Sanai N., Berens M., Kim S., and Dhruv H.. Probing glioblastoma and its microenvironment using single-nucleus and single-cell sequencing. BIBM'19 Single Cell-omics: Challenges and Opportunities. 2019.
7. Chowdhury, S., **Dong, X.**, and Li, X., Recurrent Neural Network Based Feature Selection for High Dimensional and Low Sample Size Micro-array Data. BigData'19 KDDBHI. 2019.