

Jia Xu

CONTACT

Department of Computer Science
Graduate Center & Hunter College
City University of New York
695 Park Ave, NY, NY 10065, USA

Phone: +1-347-571-7013
e-mail: jia.xu@hunter.cuny.edu
<http://www.jiaxu.org>

RESEARCH INTERESTS

Artificial Intelligence, Machine Learning, Machine Translation

EMPLOYMENT

Apr. 2016 -

Assistant professor
Graduate Center & Hunter College, City University of New York

Feb. 2015 - Mar. 2016

Associate professor
Head of the Center for Machine Intelligence (CMI)
Key Laboratory of Intelligent Information Processing
Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS)

Dec. 2011 - Jan. 2015

Assistant professor
Head of the Center for Machine Intelligence (CMI)
IIIS, Tsinghua University

Nov. 2009 - Nov. 2011

Senior Researcher
German Institute for Artificial Intelligence (DFKI)

EDUCATION

Aug. 2003 - Sep. 2009

Department 6 of Computer Science (Informatik 6), RWTH-Aachen University
Ph.D. in Computer Science
Thesis: Sequence Segmentation for Statistical Machine Translation
Advisor: Hermann Ney

Oct. 1999 - Mar. 2003

Department of Computer Science, TU-Berlin
Diploma in Computer Science (MSc&BSc)
(5 year program completed in 3.5 yrs; 2nd student in the history of DCS)
Thesis: A Computational Model for Human Auditory System
Advisor: Werner Hemmert

PUBLICATIONS (RECEIVED ABOUT 300 CITATIONS)

- Zhixian Lei, Xuehan Ye, Yongcai Wang, Deying Li, Jia Xu. On the Efficient Online Model Adaptation by Incremental Simplex Tableau. accepted in Proceedings of AAAI, June 2017,
- P. Papakonstantinou, J. Xu, G. Yang (*authors alphabetically ordered – paper on the mathematics of Machine Learning*). On the Power and Limits of Distance-Based Learning. In Proceedings of ICML, NY, NY, June 2016.

- P. Papakonstantinou, J. Xu, C. Zhu (*first two authors alphabetically ordered*). Bagging by Design (On the Sub-optimality of Bagging). In Proceedings of AAAI, Quebec City, Quebec, Canada, July 2014.
- M. Dong, Y. Cheng, Y. Liu, J. Xu (*corresponding author*), M. Sun. Query Lattice for Translation Retrieval. In Proceedings of COLING, Dublin Ireland, August 2014,
- C. Gan, Z. Qin, J. Xu, T. Wan. Salient Object Detection in Image Sequences via Spatial-Temporal Cue. In Proceedings of Visual Communications and Image Processing (VCIP), Sarawak, Malaysia, November 2013.
- J. Xu, C. Kennington, C. Przywara and L. Wanzare. Comparable Corpora in Wikipedia Text for Machine Translation. In Proceedings of the 6th NIC Symposium 2012: 25 Years HLRZ/NIC (Book Section), ISBN: 9783893367580. Jülich, Germany. February 2012.
- J. Xu and W. Sun. Generating Virtual Parallel Corpus: A Compatibility Centric Method. In Proceedings of MT-Summit , Xianmen, China. September 2011.
- W. Sun and J. Xu. Enhancing Chinese word segmentation using unlabeled data. In Proceedings of the EMNLP 2011. Edinburgh, UK, July 2011.
- J. Xu, J. Gao, K. Toutanova and H. Ney: Synchronous Learning of Chinese Word Segmentation and Word Alignment. In the Handbook of Natural Language Processing and Machine Translation (Book Chapter), Springer, New York. 2011.
- A. Eisele and J. Xu: Improving Machine Translation Performance Using Comparable Corpora. In Proceedings of the LREC Workshop on Building and Using Comparable Corpora, Malta, May 2010.
- C. Federmann, A. Eisele, Y. Chen, S. Hunsicker, J. Xu and H. Uszkoreit: Further Experiments with Shallow Hybrid MT Systems. In Proceedings of the ACL Workshop on Statistical Machine Translation, pages 77-81, Uppsala, Sweden, July 2010.
- J. Xu, J. Gao and K. Toutanova: Character-based Chinese-English Statistical Machine Translation. Technical Report at RWTH-Aachen University. Aachen, Germany, October 2009.
- J. Xu, J. Gao, K. Toutanova and H. Ney: Bayesian Semi-Supervised Chinese Word Segmentation for Statistical Machine Translation. In Proceedings of the 22nd International Conference on Computational Linguistics , Manchester, UK, August 2008.
- Y. Deng, J. Xu and Y. Gao: Phrase Table Training for Precision and Recall: What Makes a Good Phrase and a Good Phrase Pair? In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies , Columbus, OH, June 2008.
- J. Xu, Y. Deng, Y. Gao and H. Ney: Domain Dependent Machine Translation. In Proceedings of the Machine Translation Summit XI , Copenhagen, Danmark, September 2007.
- J. Xu, R. Zens and H. Ney: Partitioning Parallel Documents Using Binary Segmentation. In Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL): Proceedings of the Workshop on Statistical Machine Translation , pp. 78-85, New York City, NY, June 2006.
- D. Vilar, J. Xu, L. F. D'Haro and H. Ney: Error Analysis of Statistical Machine Translation Output. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC) , pp. 697-702, Genova, Italy, May 2006.
- J. Xu, E. Matusov, R. Zens, and H. Ney: Integrated Chinese Word Segmentation in Statistical Machine Translation. In Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Pittsburgh, PA, October 2005.

- R. Zens, O. Bender, S. Hasan, S. Khadivi, E. Matusov, J. Xu, Y. Zhang, and H. Ney: The RWTH Phrase-based Statistical Machine Translation System. In Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Pittsburgh, PA, October 2005.
- J. Xu, R. Zens, and H. Ney: Sentence Segmentation Using IBM Word Alignment Model 1. In Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT 2005), pp. 280-287, Budapest, Hungary, May 2005.
- J. Xu, R. Zens, and H. Ney: Do We Need Chinese Word Segmentation for Statistical Machine Translation? In Proceedings of the Third SIGHAN Workshop on Chinese Language Learning, pp. 122-128, Barcelona, Spain, July 2004.

MANUSCRIPTS

The following are available upon request:

- J. Xu: Sampling-added Machine Translation (archived in Oct./2013 Project Report at IIS and presented in Nov./2013 at MSRA)
- G. Chen and J. Xu: Language Model Smoothing Using Nested Chinese Restaurant Process (archived in Jun./2013 Bachelor Thesis of Beida)

COMPETITION RECORDS

I won leading ranks in the following machine translation evaluations:

- *NIST 2015*, representing ICT/CAS, Chinese-English, Rank 4th (among all) and Rank 1st (among academic affiliations), **leader**
- *WMT 2011*, representing DFKI, English-German, Rank 1st (in BLEU), **leader**
- *NIST 2008*, representing MSR Redmond, Chinese-English, Rank 1st
- *NIST 2006*, representing RWTH-Aachen, Chinese-English, Rank 4th, **leader**
- *NIST 2005*, representing RWTH-Aachen, Chinese-English, Rank 4th
- *NIST 2004*, representing RWTH-Aachen, Chinese-English, Rank 2nd
- *GALE 2008*, representing RWTH-Aachen, Chinese-English, Rank 2nd (as NightInGale)
- *GALE 2007*, representing RWTH-Aachen, Chinese-English, Rank 2nd (as NightInGale)
- *GALE 2006*, representing RWTH-Aachen, Chinese-English, Rank 2nd (as NightInGale)
- *TC-Star 2007*, representing RWTH-Aachen, Chinese-English, Rank 1st
- *TC-Star 2006*, representing RWTH-Aachen, Chinese-English, Rank 1st
- *TC-Star 2005*, representing RWTH-Aachen, Chinese-English, Rank 1st

US PATENT

- Jianfeng Gao, Kristina Nikolova Toutanova and Jia Xu: Unsupervised Chinese Word Segmentation for Statistical Machine Translation. Application No. 12/163.119, (ASM 8310) MS 323635.01 Redmond, Washington, June 2008.

ADMINISTRATION & SERVICE

Academic year 2012-2013, served full-time as:

- *Chair for Elite Tsinghua Computer Science Program*
- *Chair of student activities and leadership development of IIS*

**PHD
SUPERVISION**

- Yong Cheng (Sep. 2012 - Dec. 2016): Machine Translation
- Chuang Gan (Sep. 2013 - Dec. 2016): Image Recognition
- Mingxuan Wang (May 2015 - May 2016): Machine Translation
- Xiaojin Ye (Dec. 2016 -): Deep Learning
- Nur Dean (Dec. 2016 -): Semantics Analysis
- Pablo Gonzalez (May 2016 -): Machine Translation
Co-supervised with William Sakas

**COMPLETED
MSC THESES**

- Geliang Chen (Dec. 2016, Tsinghua Univeristy):
Phrase-based Language Model for Statistical Machine Translation
- Xiaojun Zhang (Nov. 2011, Univeristy of Saarland):
Two-level Parallel Text Extraction from Comparable Corpora
Co-supervised with Hans Uszkoreit

**COMPLETED
BACHELOR
THESES**

- Shun Zheng (May 2014, Beijing University of Posts and Telecommunications):
Improvements on Word Alignment Models in Statistical Machine Translation
- Zhengping Che (Jun. 2013, Elite Tsinghua Computer Science Program, Tsinghua University):
Dirichlet Process Model for Phrase-based Machine Translation
- Yulong Zeng (Jun. 2013, Elite Tsinghua Computer Science Program, Tsinghua University):
A Comparative Study of Generative Model and Discriminative Model
- Zhibo Zhang (Jun. 2013, Elite Tsinghua Computer Science Program, Tsinghua University):
Machine Learning-based Crime Prediction
- Suiqian Luo (Jun. 2013, Elite Tsinghua Computer Science Program, Tsinghua University):
Improving Training Models in Statistical Machine Translation
- Geliang Chen (Jun. 2013, Peking University):
A Novel Approach for Language Model
- Yong Cheng (Jun. 2012, Beijing Jiao Tong University):
Analysis of User Behaviors in Social Network

**CURRENT
GRANTS**

- 660,000 RMB (100,000 USD), NSFC (NSF-China) grant, 2017-2019, Co-PI, Key Problems for Tightly-coupled, Multi-signal Fusion based Simultaneously Locating and Mapping.
- 200,000 RMB (33,000 USD), KLIIP grant, 2015-2016, Principal Investigator, Novel machine learning methods, Grant No. 20156020
- 500,000 RMB (83,000 USD), ICT grant (Innovation subjects), 2015-2017, Principal Investigator, Ensemble learning in machine translation, Grant No. 20156020
- 660,000RMB (100,000 USD), NSFC grant, 2014-2017, *co-PI*, New Approaches to the Limits of Efficient Propositional Reasoning: Algorithms, Approximations and Foundations, Grant No. 20131351464

COMPLETED EU AND DARPA GRANTS

- ACCURAT, EU grant, 2010-2012, *project leader at DFKI and package leader*
- EuroMatrix-Plus, EU grant, 2010-2012, *member*
- Quaero, EU grant, 2009-2010, *member*
- GALE, DARPA grant, 2006-2008, *member*
- TC-Star, EU grant, 2004-2006, *member*

TEACHING

I designed and instructed the following courses:

- *Natural Language Processing C-SC 74040* (graduate course)
Graduate Center, CUNY, Fall 2016.
- *Language Technology LING 83600 32100* (graduate course)
Graduate Center, CUNY, Fall 2016.
- *Software Analysis and Design 1 CSCI 13500 - Xu Section* (undergraduate course)
Hunter College, CUNY, Fall 2016.
- *Speech Communication of Human and Machine* (undergraduate Yao-class course)
Tsinghua University, Fall 2013, co-taught with P.C. Ching, Tan Lee, Helen Meng and William S.-Y. Wang
- *Machine Learning and Machine Translation* (graduate course)
Tsinghua University, Fall 2012
- *Machine Learning and Pattern Recognition* (undergraduate Yao-class course)
Tsinghua University, Spring 2013
- *Language Technology* (graduate course)
University of Saarland, Spring 2011, co-taught with Martin Kay and Hans Uszkoreit

ACADEMIC EXCHANGE AND LONG-TERM VISITS

- Visiting Researcher, 2013/7-2014/1
MSRA Star-Track Visiting Young Faculty Program
- Ph.D. Internship, 2007/2-2007/5
IBM T. J. Watson Research Center, Yorktown Heights, USA
Research topic: Phrase extraction and domain adaptation
Practical work: large-scale Chinese-English statistical machine translation system
Manager: Yuqing Gao, Mentor: Yonggang Deng
One ACL Publication and one MT-Summit during this internship.
- Ph.D. Internship, 2007/10-2008/2
Microsoft Research NLP group, Redmond, USA
Research topic: Chinese word segmentation for statistical machine translation
Practical work: NIST MT evaluation campaign 2008
Manager: Bill Dolan, Mentor: Jianfeng Gao
One Coling Publication during this internship and one US patent.

INVITED TALKS

- Rutgers University, Newark, August, 2016
- Google Research, NYC, December, 2015
- University of Columbia, New York City, November, 2015
- University of Washington, Seattle, November, 2015
- USC, Los Angeles, November, 2015
- CWMT, Tutorial "Ensemble Learning for Machine Translation", Macau, November, 2014
- DFKI, Berlin, Germany, October, 2014
- RWTH-Aachen University, Germany, October, 2014
- Aarhus University, Denmark, October 2014
- Stanford University, USA, April 2014
- Facebook, USA, March 2014
- MSR-Asia, Star-Track Research Talk, Beijing, June 2013
- MSR-Asia, Natural Language Computing Group, Beijing, July 2013
- MSR-Asia, "MSRA-IIIS" Bilateral Seminar Series, Beijing, December 2012
- CUHK, Machine Learning Workshop, Hongkong, July 2012
- MSR, Natural Language Processing Group, Redmond (USA), September 2011
- SRI International, Menlo Park (USA), September 2011
- CAS, National Laboratory of Pattern Recognition Group, Beijing, November 2010
- MSR-Asia, Natural Language Computing Group, Beijing, November 2010
- Systran Software, Paris, December 2009
- Limsi, Spoken Language Processing Group, Orsay (France), 2009
- DFKI, Language Technology Group, Saarbruecken (Germany), October 2009
- IBM, Speech Technology Group, Watson (USA), February, 2008

REVIEW OF CONFERENCE PAPERS

- EMNLP 2016: Conference on Empirical Methods in Natural Language Processing
- NIPS 2016: The Thirtieth Annual Conference on Neural Information Processing Systems

REVIEW OF PH.D. THESIS

- Mr Aji Prasetya Wibawa, Advanced Javanese-to-Indonesian statistical machine translation, University of South Australia, August, 2013.
- Hongyu Liang, Efficient Approximation Algorithms for Graph Optimization Problems: Through a Combinatorial Lens, Tsinghua University, June, 2013.
- Jing He, Community Finding and Information Propagation in Social Networks, Tsinghua University, June, 2012.

WORKSHOP ORGANIZER LANGUAGES

- CUHK, Machine Learning Workshop, Hongkong, July 2012
- Fluent in English
and German
- Native Speaker in Chinese