

Shutian Ma

Email: mashutian0608@hotmail.com

Phone: +86 15005156531

Brief Bio

- Doctoral Degree in Nanjing University of Science & Technology
- Conduct research using NLP, machine learning and deep learning models

Areas of Research

- Document clustering and classification
- Text embedding and node embedding
- Scientific software recommendation, Citation recommendation (Dissertation Topic)

Education

Ph.D. in Information Science

- Nanjing University of Science & Technology (2014-2020)
- Department of Information Science
- Supervisor: [Chengzhi Zhang](#)

B.M. in Information Management and System

- Nanjing University of Science & Technology(2010-2014)
- Information Management and Information System, Department of Information Science

Exchange student in Economic Management

- Ajou University, South Korea (Spring 2012)
- Information Management and Information System, Department of Information Science

Experience

Data Scientist Intern, Tencent, Shenzhen (07/2020-Now)

- Customer Research & User Experience Design Center
- Data preprocessing, Name entity extraction, Knowledge Graph

NLP Engineer Intern, Aegis, Nanjing (03/2020-04/2020)

- Data preprocessing, Name entity extraction

NLP Engineer Intern, Bytedance, Beijing (11/2018-01/2019)

- Department of Data and NLP
- Data preprocessing, modeling training and query classification service construction

Visiting scholar, Indiana university Bloomington, United States (10/2017-10/2018)

- Cyberinfrastructure for Network Science (CNS) Center, Department of Information and Library Science, School of Informatics, Computing, and Engineering
- Working on the projects of job recommendations.
- Supervisor: [Katy Boner](#)
- Supervisor: [Xiaozhong Liu](#)

Publications

1. **Ma, Shutian**, Chengzhi Zhang*, and Xiaozhong Liu. Chronological Citation Recommendation with Time Preference. *Scientometrics*. 2020. (Under review)
2. **Ma, Shutian**, Chengzhi Zhang*, and Xiaozhong Liu. A review of citation recommendation: from textual content to enriched context [J]. *Scientometrics*. 2020, 122(3): 1445-1472. (SSCI)
3. **Shutian Ma**, Jin Xu, Chengzhi Zhang*. Automatic identification of cited text spans: a multi-classifier approach over imbalanced dataset [J]. *Scientometrics*, 2018, 116(2): 1303-1330. (SSCI)
4. **Shutian Ma**, Yingyi Zhang, Chengzhi Zhang*. Using multiple Web resources and inference rules to classify Chinese word semantic relation [J]. *Information Discovery and Delivery*, 2018, 46(2): 120-126. (SSCI)
5. Börner, Katy, Olga Scrivner, Mike Gallant, **Shutian Ma**, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy [J]. *Proceedings of the National Academy of Sciences (PNAS)*, 2018, 115(50): 12630-12637. (SCI)
6. Zhang, Chengzhi, Zijing Yue, Qingqing Zhou, **Shutian Ma**, and Zi-Ke Zhang. Using social media to explore regional cuisine preferences in China [J]. *Online Information Review*. 2019, 43(7): 1098-1114. (SCI)
7. Qiangbing Wang, **Shutian Ma**, Chengzhi Zhang. Predicting Users' Demographic Characteristics in a Chinese Social Media Network [J]. *The Electronic Library*. 2017, 35(4): 758-769. (SSCI)
8. Börner, Katy, Olga Scrivner, Leonard E. Cross, Michael Gallant, **Shutian Ma**, Adam S. Martin, Elizabeth Record, Haici Yang, and Jonathan M. Dilger. Mapping the co-evolution of artificial intelligence, robotics, and the internet of things over 20 years (1998-2017). arXiv preprint arXiv:2006.02366 (2020).
9. **Ma, Shutian**, Heng Zhang, Tianxiang Xu, Jin Xu, Shaohu Hu, and Chengzhi Zhang. IR&TM-NJUST@ CLSciSumm-19 [C]. In: *Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2019)*, Paris, France, 2019:181-195. (EI)
10. **Shutian Ma**, Heng Zhang, Jin Xu, Chengzhi Zhang*. NJUST @ CLSciSumm-18 [C]. In: *Proceedings of the 3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2018)*, Ann Arbor, MI, USA, 2018: 114-129 (EI)
11. **Shutian Ma**, Jin Xu, Jie Wang and Chengzhi Zhang*. NJUST @ CLSciSumm-17 [C]. In: *Proceedings of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017)*, Tokyo, Japan, 2017: 16-25. (EI)
12. **Shutian Ma**, Chengzhi Zhang*. Document Representation and Clustering Models for Bilingual Documents Clustering [C]. In: *Proceedings of 2017 Annual Meeting of the Association for Information Science and Technology (ASIST'2017)*, Washington, DC, USA, 2017: 499-502.
13. **Shutian Ma**, Chengzhi Zhang*. Using Full-text to Evaluate Impact of Different Software Groups [C]. In: *Proceedings of the 16th International Conference on Scientometrics and Informetrics (ISSI 2017)*, Wuhan, China, 2017: 1666-1667. (EI)
14. **Shutian Ma**, Chengzhi Zhang*. Documents Representation for Comparable Corpora Clustering: A Preliminary Study [C]. In: *Proceedings of iConference2017*, Wuhan, China, 2017: 876-880.
15. **Shutian Ma**, Chengzhi Zhang*, Daqing He. Document Representation Methods for Clustering Bilingual Documents [C]. In: *Processing of the 2016 Annual Meeting of the Association for Information Science and Technology (ASIST'2016)*, Copenhagen, Denmark, 2016: 1-10
16. **Shutian Ma**, Chengzhi Zhang*. Automatic Collection of the Parallel Corpus with Little Prior Knowledge [C]. In: *Proceedings of the 13th China National Conference on Computational Linguistics (CCL2014)*, Wuhan, China, 2014: 95-106. (EI)
17. **Shutian Ma**, Xiaoyong Zhang, Chengzhi Zhang*. NLPCC 2016 Shared Task: Chinese Words Similarity Measure via Ensemble Learning based on Multiple Resources [C]. In: *Proceedings of the Fifth Conference on Natural Language Processing and Chinese Computing & The Twenty Fourth International Conference on Computer Processing of Oriental Languages (NLPCC-ICCPOL 2016)*. Kunming, China, 2016: 862-869. (EI)
18. Xu, Jin, Chengzhi Zhang*, and **Shutian Ma**. Ensemble System for Identification of Cited Text Spans: Based on Two Steps of Feature Selection [C]. In: *Proceedings of the 25th China Conference on Information Retrieval (CCIR2019)*. Fuzhou, China, 2019: 95-107. (EI)
19. Gao, Zheng, Vincent Malic, **Shutian Ma**, and Patrick Shih. How to Make a Successful Movie: Factor Analysis from both Financial and Critical Perspectives [C]. In: *Proceedings of the International Conference on Information, Springer, Cham, 2019: 669-678. (EI)*

20. Zhang, Heng, **Shutian Ma**, and Chengzhi Zhang. Using Full-text of Academic Articles to Find Software Clusters [C]. In: Proceedings of the 17th International Conference on Scientometrics and Informetrics (ISSI 2019), Rome, Italy, 2019: 2776-2777. (EI)
21. Jie Wang, **Shutian Ma**, Chengzhi Zhang. CitationAS: A Summary Generation Tool Based on Clustering of Retrieved Citation Content [C]. In: Proceedings of Second Workshop on Mining Scientific Papers: Computational Linguistics and Bibliometrics (CLBib-2017), Wuhan, China, 2017: 1-9. (EI)
22. Yingyi Zhang, Guo Chen, Chengzhi Zhang*, **Shutian Ma**. Analyzing scientific user tagging behavior on academic blogs according to tag's content characteristics - a preliminary study [C]. In: Proceedings of iConference2017, Wuhan, China, 2017: 881-886.
23. Lu, Chao, Chengzhi Zhang, and **Shutian Ma**. How does citing behavior for a scientific article change over time? A preliminary study [C]. In: Proceedings of the 2015 Annual Meeting of the Association for Information Science and Technology (ASIST'2015), St. Louis, Missouri, USA, 2015: 1-4.
24. 章成志,徐津,马舒天.学术文本被引片段的自动识别研究[J].情报理论与实践,2019,42(09):139-145. (CSSCI)
25. 章成志,马舒天,揭春雨,姚旭晨.基于双语 URL 匹配模式可信度的平行网页识别研究[J].中文信息学报,2018,32(03):91-100. (CSCD)

NLP Competition Experience

Scientific Document Summarization, CL-SciSumm, 2017-2019

- Collaborated with two colleague (Leader)
- Placed 1st among 9 systems in Task 1A (2017) and placed 2nd among 9 systems in Task 1B (2018)
- Shared Task presented at BIRNDL 2017/2018/2019, co-located after the ACM SIGIR Conference

The Evaluation of Text Sourcing Technology, SMP-ETST, 2018

- Collaborated with two colleague (Leader), placed the 3rd among 13 systems
- Shared Tasks in Conference on Social Media Processing in 2018 (SMP CUP 2018)

Chinese Word Semantic Relation Classification, 2017

- Collaborated with one colleague (Leader), placed the 8th among 17 systems
- Shared Tasks in the 6th Conference on Natural Language Processing and Chinese Computing (NLPC 2017)

Chinese Word Similarity Measurement, 2016

- Collaborated with one colleague (Leader), placed the 11th among 24 systems
- Shared Tasks in Conference on Natural Language Processing and Chinese Computing & International Conference on Computer Processing of Oriental Languages (NLPC-ICCPOL 2016)

Skills and Abilities

- Software and Tools: Gensim, Scikit-Learn, Tensorflow, Pytorch, Gephi, Neo4j, Navicat
- Programming Languages: Java, Python, Linux command Line, SQL
- Languages: Chinese and English