

# 马舒天

---

地址: 南京理工大学经济管理学院

电子邮件: mashutian0608@hotmail.com

电话号码: +86 15005156531

## 个人简介与研究方向

- 管理科学与工程专业博士, 预计 2020 年毕业
- 利用自然语言处理, 机器学习与深度学习相关技术进行科研活动
- 研究方向涉及: 双语网页发现, 多语言文本聚类, 科研软件分析, 基于被引片段的自动摘要, 引文推荐系统

## 教育背景与工作经历

### 硕博连读, 南京理工大学经济管理学院, 2014-至今

- 管理科学与工程专业, GPA ( 3.37/4 ), 预计 2020 年毕业
- 导师章成志([https://www.researchgate.net/profile/Chengzhi\\_Zhang4](https://www.researchgate.net/profile/Chengzhi_Zhang4))

### 实习经历, 北京字节跳动网络技术有限公司, 自然语言处理工程师实习生, 2018.10-2019.1

- 基础数据预处理, 训练分类器, 完成一个广告推送线下服务

### 访问学者, 美国印第安纳大学布鲁明顿, 2017.10-2018.10

- 就职于 Cyberinfrastructure for Network Science (CNS) Center, School of Informatics, Computing, and Engineering
- 合作导师信息  
Katy Boner(<http://ella.slis.indiana.edu/~katy/>)  
Xiaozhong Liu([https://www.sice.indiana.edu/all-people/profile.html?profile\\_id=100](https://www.sice.indiana.edu/all-people/profile.html?profile_id=100))

### 本科学历, 南京理工大学经济管理学院, 2010 - 2014

- 信息管理与信息系统专业, GPA ( 3.29/4 )

## 已发表文章

1. **Ma, Shutian**, Heng Zhang, Tianxiang Xu, Jin Xu, Shaohu Hu, and Chengzhi Zhang. IR&TM-NJUST@ CLSciSumm-19. (BIRNDL2019), July 2019, Paris, France.
2. Jin Xu, Chengzhi Zhang and **Shutian Ma**. Ensemble System for Identification of Cited Text Spans: Based on Two Steps of Feature Selection, CCIR 2019 (Accept).
3. 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." Proceedings of the National Academy of Sciences 115, no. 50 (2018): 12630-12637. ( 美国国家科学院院刊, SCI 检索 )
4. **Shutian Ma**, Heng Zhang, Jin Xu, Chengzhi Zhang\*. NJUST @ CLSciSumm-18. In: Proceedings of the 3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2018), July 2018, Michgen, USA. ( EI 检索 )
5. **Shutian Ma**, Yingyi Zhang, Chengzhi Zhang\*. Using multiple Web resources and inference rules to classify Chinese word semantic relation. Information Discovery and Delivery 46.2 (2018): 120-126. ( SSCI 检索 )
6. **Ma, Shutian**, and Chengzhi Zhang. Using Full-text Academic Articles and Wikipedia to Find Alternative Free Bioinformatics Software. SIGMET 2018.
7. **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 检索 )
8. Qiangbing Wang, **Shutian Ma**, Chengzhi Zhang\*. Predicting Users' Demographic Characteristics in a Chinese Social Media Network. The Electronic Library. 2017, 35(4): 758-769. ( SSCI 检索 )

9. Jie Wang, **Shutian Ma**, Chengzhi Zhang\*. CitationAS: A Summary Generation Tool Based on Clustering of Retrieved Citation Content. In: Proceedings of Second Workshop on Mining Scientific Papers: Computational Linguistics and Bibliometrics (CLBib-2017), Wuhan, China, 2017. (EI检索)
10. **Shutian Ma**, Jin Xu, Jie Wang and Chengzhi Zhang\*. NJUST @ CLSciSumm-17. In: Proceedings of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017), Aug, 2017, Tokyo, Japan. (Winner of CL-SciSumm 2017) (EI检索)
11. **Shutian Ma**, Chengzhi Zhang\*. Document Representation and Clustering Models for Bilingual Documents Clustering. In: Proceedings of 2017 Annual Meeting of the Association for Information Science and Technology (ASIST'2017), Washington, DC, USA, 2017.
12. **Shutian Ma**, Chengzhi Zhang\*. Using Full-text to Evaluate Impact of Different Software Groups. In: Proceedings of the 16th International Conference on Scientometrics and Informetrics (ISSI 2017), Wuhan, China, 2017. (EI检索)
13. **Shutian Ma**, Chengzhi Zhang\*. Documents Representation for Comparable Corpora Clustering: A Preliminary Study. In: Proceedings of iConference2017, March 22-25, Wuhan, China, 2017.
14. 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. In: Proceedings of iConference2017, March 22-25, Wuhan, China, 2017.
15. **Shutian Ma**, Xiaoyong Zhang, Chengzhi Zhang\*. NLPCC 2016 Shared Task: Chinese Words Similarity Measure via Ensemble Learning based on Multiple Resources. 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检索)
16. **Shutian Ma**, Chengzhi Zhang\*, Daqing He. Document Representation Methods for Clustering Bilingual Documents. In: Processing of the 2016 Annual Meeting of the Association for Information Science and Technology (ASIST'2016), Copenhagen, Denmark, 2016.
17. **Shutian Ma**, Chengzhi Zhang\*. Automatic Collection of the Parallel Corpus with Little Prior Knowledge. In: Proceedings of the 13th China National Conference on Computational Linguistics (CCL2014), Wuhan, China, 2014: 95-106. (EI检索)
18. 章成志, 徐津, **马舒天**. 学术文本被引片段的自动识别研究 [J/OL]. 情报理论与实践 :1-11[2019-07-30].<http://kns.cnki.net/kcms/detail/11.1762.G3.20190610.1738.010.html>
19. 章成志, **马舒天**, 揭春雨, & 姚旭晨. (2018). 基于双语 URL 匹配模式可信度的平行网页识别研究. 中文信息学报, 32(3), 91-100.

## 比赛经历

1. 2019 年第四届数字图书馆文献计量增强信息检索和自然语言处理联合研讨会 (BIRNDL 2019), 第五届计算语言学领域自动文档摘要比赛, 引文域识别和自动摘要生成任务, 部分指标小组第一名
2. 2018 年全国社交媒体处理大会, 文本溯源技术评测, 小组第三名, 颁发二等奖
3. 2018 年第三届数字图书馆文献计量增强信息检索和自然语言处理联合研讨会 (BIRNDL 2018), 第四届计算语言学领域自动文档摘要比赛, 引文域类型分类任务, 小组第二名
4. 2017 年第二届数字图书馆文献计量增强信息检索和自然语言处理联合研讨会 (BIRNDL 2017), 第三届计算语言学领域自动文档摘要比赛, 引文域识别任务, 小组第一名
5. 2017 年自然语言处理与中文计算国际会议中文词语语义关系分类测评, 小组第八
6. 2016 年自然语言处理与中文计算国际会议中文词语相似度计算测评, 小组第十一

## 获奖经历

1. 2018 年研究生国家奖学金一等奖, 工信部创新创业奖学金一等奖
2. 2016 年第六届全国情报学博士生论坛二等奖
3. 2015 年全国情报学博士生学术论坛优秀论文一等奖
4. 2015 年第五届全国情报学博士生论坛暨 2015 中国信息资源管理论坛优秀论文三等奖

## 个人技能

- 软件库: Gensim, Scikit-Learn, Tensorflow, Gephi, Neo4j, Navicat
- 编程语言: Java, Python, Linux 命令, SQL
- 语言能力: 英文熟练