

分类号 TP39
UDC 004

学校代码 10590
密 级 公开

硕士学位论文

XXXXXXXXXXXXX

YYYYYYYYYYYYYYYYYY

学位申请人姓名 XXX

学位申请人学号 XXXXXXXXXX

专 业 名 称 计算机科学与技术

学 科 门 类 工学

学院 (部, 研究院) 计算机与软件学院

导 师 姓 名 XXXXXXXXXXXXX

二〇二四年五月

学位论文原创性声明

日期: 年 月 日

学位论文使用授权说明

日期： 年 月 日

摘 要

中文摘要

关键词：短文本，主题模型，数据增强，变分自编码器，数据挖掘

ABSTRACT

ABSTRACT

Key word: short text, topic model, data augmentation, variational autoencoder, data mining

目 录

摘要	I
ABSTRACT.....	II
第一章 绪论.....	1
1.1 研究背景.....	1
附录	2
参考文献.....	2
致谢	5
攻读硕士学位期间的研究成果.....	6

第一章 绪论

1.1 研究背景

随着信息技术和互联网媒体的崛起，如博客、维基百科、社交媒体平台等，文本数据已经成为当代社会信息传播的重要载体。其中，短文本作为信息传播的一种高效形式，其数量在互联网时代经历了爆炸性的增长。短文本通常指的是字数较少、内容简洁的文本数据。它们的主要特点是信息量密集，但表达形式极为简洁。比如在社交平台中，不论是用户发表的微博和小红书，还是标题、弹幕以及评论等，绝大多数都以短文本的形式存在。由于短文本在个人日常交流、商业广告、新闻报道等领域扮演着重要的角色，对短文本进行分析研究不仅对于理解和挖掘网络社会的信息动态具有重要意义，也对于商业智能和公共管理等领域的决策支持具有实际价值。

附录 A IETM 的吉布斯采样公式推导

附录 B SpareNTM 的损失函数推导细节

致 谢

攻读硕士学位期间的研究成果

- [1] Chen, J., Wang, R., He, J., Li, M. J. (2023, September). Encouraging Sparsity in Neural Topic Modeling with Non-Mean-Field Inference. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, ECML-PKDD (pp. 142-158). Cham: Springer Nature Switzerland. (CCF B 类会议)
- [2] Li, M. J., Chen, J., Li, J., Wang, R., Zhang, Q.. Transferring Knowledge from Large Language Models for Short Text Topic Modeling. International Conference on Data Engineering, ICDE. (CCF A 类会议, 在投)
- [3] He, J., Chen, J., Li, M. J. (2022, November). Multi-knowledge Embeddings Enhanced Topic Modeling for Short Texts. In International Conference on Neural Information Processing (pp. 521-532). Cham: Springer International Publishing. (CCF C 类会议)
- [4] Li, M. J., Wang, R., Li, J., Bao, X., He, J., Chen, J., He, L. (2023, November). Topic Modeling for Short Texts via Adaptive Pólya Urn Dirichlet Multinomial Mixture. In International Conference on Neural Information Processing (pp. 364-376). Singapore: Springer Nature Singapore. (CCF C 类会议)