

目 录

第一章 绪 论	1
1.1 研究工作的背景与意义	1
1.2 国内外研究现状.....	1
1.3 课题主要内容	1
1.4 本论文的结构安排	1
参考文献	2

第一章 绪论

1.1 研究工作的背景与意义

随着信息化进程的不断推进，我国互联网市场空前繁荣。从 WEB2.0 时代开始，到如今的“互联网+”时代，互联网已经融入了人们的生活中，也对各行各业产生了深远的影响。根据中国互联网络信息中心发布的《第 40 次中国互联网络发展状况统计报告^①》显示，截止到 2017 年 6 月，我国网民规模达到 7.51 亿，其中手机网民更是达到 7.24 亿，占总体网民的 96.3%，使用率排名前三的互联网应用分别是即时通信（92.1%）、网络新闻（83.1%）、搜索引擎（81.1%），使用率最高的三个 app 应用则是微信（84.3%）、QQ（65.8%）、微博（38.7%）。可以明显看出，人们使用互联网以及获取信息的方式，已经从以前的桌面台式电脑，转变为移动端的手机、IPAD 等掌上设备。

移动设备的盛行，使得人们发布和接收信息都更加方便，据统计，在 2012 年，微博用户已经增长到 3.68 亿，其中 69% 通过移动设备登陆，每天能产生 1.17 亿条微博。大量的手机用户增加，让互联网信息爆炸式增长，并且产生了大量碎片化的信息，如微博、QQ 说说、留言、商品评论等，如何理解并利用这些浩瀚如烟的短文本信息成了互联网应用最为迫切的需求之一。例如，根据一个用户每天在微博或朋友圈上发布的信息挖掘出该用户的兴趣倾向，从而让运营商精确投放用户感兴趣的广告，减少无关广告对用户体验的伤害。但是和传统文章相比，短文本过于短小（通常在 100 字以内，一般是一句话的长度），不能提供足够的词共现（word co-occurrence）或上下文，以至于很难从中提取出有效的文本特征^[1]。因此，常规机器学习技术与文本挖掘算法很难直接应用在短文本之上。

1.2 国内外研究现状

1.3 课题主要内容

1.4 本论文的结构安排

① <http://www.cnnic.net.cn/hlwfzyj/hlwxxzbq/hlwtjbg/201708/P020170807351923262153.pdf>

参考文献

- [1] G. Song, Y. Ye, X. Du, et al. Short text classification: A survey[J]. ISSN 1796-2048 Volume 9, Number 5, May 2014, 2014, 9(5): 635
- [2] X. F. Liu, B. Z. Wang, W. Shao, et al. A marching-on-in-order scheme for exact attenuation constant extraction of lossy transmission lines[C]. China-Japan Joint Microwave Conference Proceedings, Chengdu, 2006, 527-529
- [3] 竺可桢. 物理学 [M]. 北京: 科学出版社, 1973, 56-60
- [4] 陈念永. 毫米波细胞生物效应及抗肿瘤研究 [D]. 成都: 电子科技大学, 2001, 50-60
- [5] 顾春. 牢牢把握稳中求进的总基调 [N]. 人民日报, 2012 年 3 月 31 日
- [6] 冯西桥. 核反应堆压力容器的 LBB 分析 [R]. 北京: 清华大学核能技术设计研究院, 1997 年 6 月 25 日
- [7] 肖珍新. 一种新型排渣阀调节降温装置 [P]. 中国, 实用新型专利, ZL201120085830.0, 2012 年 4 月 25 日
- [8] 中华人民共和国国家技术监督局. GB3100-3102. 中华人民共和国国家标准—量与单位 [S]. 北京: 中国标准出版社, 1994 年 11 月 1 日
- [9] M. Clerc. Discrete particle swarm optimization: a fuzzy combinatorial box[EB/OL]. http://clere.maurice.free.fr/ps0/Fuzzy_Discrere_PSO/Fuzzy_DPSO.htm, July 16, 2010