
大规模文本分类的训练语 料去噪研究报告

中国科学技术信息研究所

2020. 12. 25

目 录

1 引言	3
1.1 研究背景和意义	3
1.2 相关研究进展	4
1.3 研究思路和总体方案	11
2 大规模文本分类的训练语料去噪研究	14
2.1 构建分类文件系统	14
2.2 分析噪声类型与多策略去噪	21
2.3 “语义错误”类型噪声的消除	22
2.4 实践结果与分析	24
3 结论	28
参 考 文 献	28
附录：三级类目嵌套字典	31