

学校代号 10532

学 号 S18100001

分 类 号 TP391

密 级 普通



湖南大学
HUNAN UNIVERSITY

硕士学位论文

基于机器学习技术的文本分类研究 与实现

学位申请人姓名 郭靖

培 养 单 位 信息科学与工程学院

导师姓名及职称 洪七公 教授

学 科 专 业 计算机科学与技术

研 究 方 向 并行分布式系统

论文提交日期 二〇二 x 年 x 月 xx 日

学校代号： 10532

学 号： S18100001

密 级： 普通

湖南大学硕士学位论文

基于机器学习技术的文本分类研究 与实现

学位申请人姓名： 郭靖

培 养 单 位： 信息科学与工程学院

导师姓名及职称： 洪七公 教授

专 业 名 称： 计算机科学与技术

论文提交日期： 二〇二 x 年 x 月 xx 日

论文答辩日期： 二〇二 X 年 x 月 xx 日

答辩委员会主席： 待定

Study on XX.

By

JING Guo

B.E. (Hunan University)201x

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of engineering

in

Computer Science and Technology

in the

Graduate School

of

Hunan University

Supervisor

Professor QIGONG Hong

June, 2020

湖 南 大 学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权湖南大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密 ☐，在 _____ 年解密后适用本授权书。

2、不保密 ☐。

（请在以上相应方框内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘 要

关键词：关键字 1；关键字 2；关键字 n

Abstract

Key Words: Keyword1; Keyword2; Keywordn

目 录

学位论文原创性声明和学位论文版权使用授权书	I
摘 要	II
Abstract	III
插图索引	VI
附表索引	VII
第 1 章 绪 论	1
1.1 多核系统概述	1
1.1.1 多核体系结构	1
1.2 选题背景与意义	1
1.3 本文主要工作	1
1.4 本文组织结构	1
第 2 章 第二章标题	2
2.1 2.1 节	2
2.1.1 2.1.1 子节	2
2.1.2 2.1.2 子节	2
2.2 2.2 节	2
第 3 章 基于多核系统的并发哈希表的评估与分析	3
3.1 实现并发哈希表的同步方法比较	3
3.2 典型的并发哈希表	3
3.2.1 缓存行哈希表	3
3.2.2 Cuckoo 哈希表	3
3.2.3 线程扩展性	3
第 4 章 章标题	4
4.1 节标题	4
第 5 章 章标题	5
5.1 节标题	5
5.1.1 子节标题	5
5.1.2 加锁与解锁	5
总结与展望	7
参考文献	8

附录 A 读学位期间所发表的学术论文·····	9
附录 B 读学位期间所参加的科研项目·····	10
致 谢·····	11

插图索引

附表索引

第1章 绪 论

1.1 多核系统概述

论文图、表、公式、伪代码的插入方式在 ch5 中的，格式仅供参考。

1.1.1 多核体系结构

多处理技术的重要性体现在以下几个方面：

-
-
-

1.2 选题背景与意义

1.3 本文主要工作

本文主要着眼主流多核处理器架构上的并发哈希表的优化、设计与应用研究。

1.4 本文组织结构

第 2 章 第二章标题

2.1 2.1 节

2.1.1 2.1.1 子节

2.1.2 2.1.2 子节

2.2 2.2 节

第 3 章 基于多核系统的并发哈希表的评估与分析

3.1 实现并发哈希表的同步方法比较

3.2 典型的并发哈希表

3.2.1 缓存行哈希表

3.2.2 Cuckoo 哈希表

3.2.3 线程扩展性

(1)

(2)

第 4 章 章标题

4.1 节标题

第 5 章 章标题

5.1 节标题

5.1.1 子节标题

定义 5.1 假阳性 (false positives) 也叫误判是指当前元素不在集合内，但由于哈希冲突的缘故存在其它元素被映射到部分相同 bit 位上，从而有一定的概率导致在判定该元素时认定其对应的所有位置都为 1，从而判定其在集合内，造成一次误判。这个概率本文称为误判率，误判率用 ϵ 表示。

5.1.2 加锁与解锁

接下来的内容介绍如何在论文中插入伪代码，仅供参考。

算法 5.1: 基于 Intel RTM 的 MCS 锁算法

```

1 struct {
2     mcs_lock_t lock;
3     uint32_t pad1[128/4 - sizeof(mcs_lock_t)];
4 } spec_mcs_lock_t;
5 struct {
6     mcs_lock_t mcs;
7     uint8_t mode;
8     char padding[];
9 } locklib_mutex_t;
10 int locklib_mutex_lock (locklib_mutex_t *mutex, uint8_t mode){
11     spec_mcs_lock_t *lock = (spec_mcs_lock_t *) mutex;
12     uint32_t reason = 0;
13 speculative_path:
14     XBEGIN(fallback_path, reason);
15     if (lock->lock) XABORT(1);
16     return 0;
17 fallback_path:
18     retries++;
19     while (lock->lock)
20         cpu_relax();
21     if (retries < MAX_RETRIES)
22         goto speculative_path;
23     /* 标准方式申请锁 */
24     my_node.locked = true;
25     qnode_t *prev = __sync_lock_TAS(&lock->lock, &my_node);
26     if (unlikely(prev != NULL)) {
27         prev->next = &my_node;
28         while (my_node.locked)
29             cpu_relax();
30     }
31     mutex->mode = mode;
32     return 0;
33 }

```

总结与展望

1. 本文工作总结

- 1.
- 2.
- 3.
- 4.

2. 下一步工作展望

参考文献

附录 A 读学位期间所发表的学术论文

- 1.
- 2.
- 3.

附录 B 读学位期间所参加的科研项目

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A

致 谢

TODO: 1 页