

基于并行决策树的微博互动数预测

黄林昊¹, 郭昆²

(1.福建广播电视大学 电子信息与计算机系, 福建 福州 350012;
2.福州大学 数学与计算机科学学院, 福建 福州 350116)

摘要: 社交网络的快速发展,微博成为主要的社交媒体平台,针对如何预测微博文本的未来互动数,对微博进行有效的分发控制的问题,提出一种基于并行决策树的微博互动数所属级数预测的方法。首先,对用户以往发表的微博进行用户特征和微博文本特征的处理;然后,使用并行决策树分类算法对训练数据进行分类模型的构建;最后使用得到的分类模型对新微博文本的互动数所属级数进行分类预测。通过对比算法的实验,验证了所提方法具有较高的分类精度和较好的可扩展性,能够对微博所属级数进行有效的分类预测。

关键词: 微博; 互动数; 并行; 决策树; 预测

中图分类号: TP 311.5 文献标志码: A 文章编号: 1672-4348(2017)03-0294-07

Interaction number prediction of micro-blog based on parallel decision tree

Huang Linhao¹, Guo Kun²

(1. Electronic Information and Computer Department, Fujian Radio and TV University, Fuzhou 350012, China;
2. College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China)

Abstract: To predict the future interaction number of micro-blog texts to implement effective distribution control of micro-blogs, a method of forecasting the series number of micro-blog interaction numbers based on parallel decision tree was proposed. Firstly, the user characteristics and micro-blog text features of the user's previous micro-blog were processed. Then, a classification model of the training data was constructed via a parallel decision tree classification algorithm. Finally, the series number of the interaction number of new micro-blog texts was classified via the classification model. The experimental results show that the proposed method has high classification accuracy and good scalability and can effectively forecast micro-blog series.

Keywords: micro-blog; interaction number; parallel decision tree; forecast

近年来,随着互联网技术、移动端技术的快速发展,特别是移动网络为代表的移动互联网技术的迅速发展,根据第 36 次《中国互联网络发展状况统计报告》的报告,截止 2015 年 6 月,中国网民的规模已经达到了 6.68 亿,其中互联网的普及率达到 48.8%,这个发展速度还在不断增加,其中手机网民规模达到了 5.94 亿,是互联网网民的主力军。这也带动了社交网络的迅速发展,而目前对于社交网络的研究主要集中在个性化内容推荐^[1]、社群挖掘^[2]、热点话题检测^[3]等方面。

微博作为社交网络中一个主要的社交媒体平台,可帮助用户发布的公开内容进行快速传播互动,它以较短的文字消息,在较短的时间内通过用户的传播达到信息的快速传播、共享,用以提高用

收稿日期: 2017-03-22
基金项目: 国家自然科学基金资助项目(61300104);福建省教育厅资助项目(JA14349)
通讯作者: 黄林昊(1979-),男,福建福州人,讲师,硕士,研究方向:移动应用、信息安全与数据挖掘。

户和内容的影响力,受到很多人的热爱。以国内主要的微博平台——新浪微博为例,目前已经超过5亿个注册用户,截止2015年9月,其微博的月活跃用户人数已经达到2.22亿,与2014年9月同比增长33%,而日活跃用户达到了1亿,比2014年同期增长30%。微博平台的发展完善,使其使用率不断提高,用户量持续增长,每天产生数以亿计的微博文本数量。如何对这些微博进行快速的分析挖掘^[4],找到受众认可度高的微博文本,对这些文本进行有效的分发控制,以提高受众认可度高的微博文本的曝光量和内容传播的互动量,具有重要的研究意义。

微博用一种短文本的形式表达用户的状态或心情,这些微博文本会被其他用户进行转发分享提高其传播量,同时用户也可以对微博进行评论、点赞等行为操作。一条微博若被用户大量的转发或评论或点赞等操作,可见其是一条比较有意义、有价值、受众认可度高的微博文本。若能提前发现这些互动数即微博转发数、微博评论数和微博点赞数高的微博文本,进行有效的分发控制,这对提高这些微博文本的曝光量具有重要的意义。

近年来,国内外专家、学者也对社交网络中的微博文本的挖掘分析进行了广泛的研究。Boyd等人对Twitter即类似国内的新浪微博的一个社交平台进行研究,研究人们对Twitter上的Retweet操作即转发操作,研究其Retweet的动机,并对Retweet的文本内容进行主题倾向等方面的研究^[5-6]。Zan等人选取用户名、关注人数、Twitter包含的单词个数等特征,然后基于一种概率的协同过滤模型Matchbox^[7],对用户转发Twitter的行为进行预测^[8],该方法简单地将用户特征和微博特征抽取出来进行预测,没有考虑用户兴趣和微博内容之间的关系。杨子等对Twitter中用户转发行为提取了22个影响因素,使用因子图模型进行了转发行为的预测,获得了比较高的精度,但其对特征的量化处理过程比较简单,导致信息传播路径预测的精度比较低^[9]。Liben-Nowell等人对真实的社会网络中的传播特征和一些相关的问题进行了比较全面的研究,明确指出想要精确地预测信息的传播路径是比较困难的,用简单的模型进行预测得到的结果与真实的结果相差比较大^[10]。Fan等人通过对新浪微博的拓扑结构和信息扩散情况进行研究,指出新浪微博具有小世

界和无标度特性的拓扑结构,其中热门事件的扩散拓扑结构呈现星形或两级的结构^[11]。Webberley等人对微博中的传播扩散进行研究,指出微博的信息传播和扩散主要是依靠用户转发产生的,且一条微博的转发链具有一定的长度,随着微博的一次次转发,其被转发的概率随着微博链的长度的增加而减小^[12]。谢婧等人研究微博用户中的转发人群和未转发人群的微博内容、粉丝数、关注数等特征,基于贝叶斯预测模型提出一种新的预测用户转发行为的方法^[13]。匡冲等人根据贝叶斯个性化排序优化标准和分解机制,提出了一种对微博转发者进行预测的方法^[14]。

目前研究者对国外的Twitter研究比较多,而对国内的微博研究相对较少,且更多的是对微博文本被转发的行为进行预测研究,对于微博本身的互动数的研究相对较少。本研究以国内的新浪微博为例,利用微博发表用户的特征和微博文本自身的特征,提出一种基于决策树的微博互动数预测方法。同时为了适应海量微博文本数据的挖掘分析,利用Spark框架将方法进行并行化处理,以提高方法处理海量微博的能力。在真实的新浪微博数据上和对不同算法进行实验分析,验证所提出方法的有效性和可扩展性。

1 相关理论基础

1.1 新浪微博互动行为

新浪微博是国内主要的用户进行交流、分享的社交媒体平台,受到大众的喜爱。新浪微博文本以短信息的形式进行传播,其要求一条博文长度不能超过140个字符,用户可以对微博进行转发、评论、点赞的操作。新浪微博文本的价值信息可以通过其他用户对该微博文本的评价情况进行体现,而对微博文本的评价方面主要可以从微博的转发数、评论数和点赞数3方面即微博的互动数进行体现。一条原创的微博,通过其转发、评论、点赞等互动行为能够体现其他用户对该原创微博内容的兴趣程度。

微博转发,用户通过点击转发按钮即可实现对原创微博的转发。微博转发行为是微博能够快速传播的主要原因,转发时用户可以对微博加以评论,形成新的微博文本,如图1所示。同时转发会使该微博的转发数进行累加,这样微博被一个用户接一个用户地转发,形成一条转发微博链,微

博格式如://@username:微博内容//@username:
微博内容。

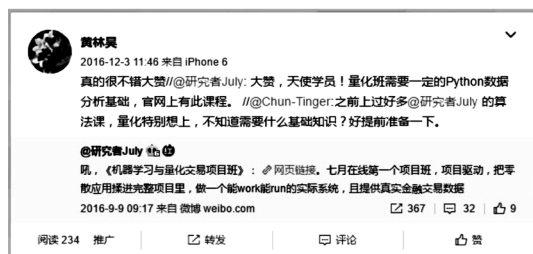


图 1 微博转发

Fig.1 Micro-blog forwarding

微博评论,用户可以直接在某条微博文本的下方进行评论,表达自己对该微博文本的认识。同时被评论微博的评论数会相应地累加。

微博点赞,用户可以直接点击微博的“赞”的按钮,即可对该微博进行点赞,以表达用户对该微博的认可度。同时被点赞的微博的点赞数会相应地累加。

1.2 Spark 分布式并行计算

Spark 是 Apache 的一个开源项目,是近年来发展较快的分布式并行数据处理框架,是伯克利大学在 2012 年提出的一种基于内存的分布式计算框架^[15],它允许重复地使用加载到内存中的数据,并且可以将计算的中间结果持久地保存在内存中^[16],从而减少磁盘 IO 操作,提高数据运算效率。Spark 采用了一种新的数据抽象模型即弹性分布式数据集(resilient distributed dataset, RDD),使其能够在多次迭代计算过程中重复利用内存数据,这也是 Spark 的核心,是一个不可变的带分区的记录集合。RDD 的基本操作包括 Transformation 和 Action^[17],其中 Transformation 是得到一个新的 RDD,可以从数据源或是 RDD 中生成,而 Action 是得到一个结果。Transformation 是采用懒策略,只有当 Action 提交时才执行相应计算。

Spark 广泛应用在计算量大、效率要求高的场景当中,通常在互联网广告、报表、推荐系统等业务中做应用分析、效果分析与优化。例如腾讯大数据精准推荐利用 Spark 快速迭代实现实时并行高维算法;淘宝技术团队将 Spark 应用于淘宝推荐相关算法,还利用 GraphX 解决生产问题。

1.3 决策树分类

决策树分类方法是一个比较经典的分类算

法,它通过使用树的结构来记录数据分类的规则,即每个树的叶节点代表某个条件下的一个数据记录集。根据数据属性字段的不同取值建立树的分支,然后在每个分支子集上重复建立下层的分支节点,最终生成一颗树。对生成的原始的决策树进行修剪,可以很快地得到具有商业价值的信息,以供决策者决策时参考。决策树分类一般分为两个步骤^[18]:(1)使用训练数据集进行学习,形成决策树分类模型的构建;(2)利用已经得到的分类模型对未知的数据进行分类。

决策树分类最重要的是选择属性进行树的分裂。其中引用率较高的决策树算法 ID3 算法使用信息增益来进行属性的划分。信息增益是基于信息熵进行属性选择的,一棵决策树对一个记录数据进行判断所需要的信息熵如式(1)所示:

$$\text{Info}(D) = - \sum_{i=1}^m p_i \log_2 p_i \quad (1)$$

其中 D 是用于存放数据记录的, p_i 是数据记录 D 中任意记录属于 C_i 的非零概率,用 $|C_i|/|D|$ 进行估计^[19]。而信息增益是原来的信息需求与新的信息需求(对属性 A 进行划分之后)之间的差,如式(2)所示:

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D) \quad (2)$$

信息增益的决策树使用信息增益最大的属性作为树节点的划分,即最小化 $\text{Info}_A(D)$ 。

2 互动数预测

2.1 数据描述与特征提取

数据选取天池大数据科研平台(<https://tianchi.shuju.aliyun.com>)提供的新浪微博文本数据,包含了 2015-02-01~2015-07-31 部分用户发表的微博文本数据,共计 1 626 750 条微博文本。其中微博数据包含的内容如表 1 所示。微博文本互动数预测是预测一条微博发表 1 周之后,被用户转发、评论、点赞的数量,同时对于数量的预测关注的是这个数量所属的一个数量级,高级别的互动数受到大众认可度高,具有较高的价值。因此将微博的互动数量分为 5 级,分别是第 1 级互动数最少,几乎没有互动数,即互动数为 0 到 5 的微博;第 2 级互动数较少,为 6 到 10 的微博;第 3 级互动数一般为 11 到 50 的微博;第 4 级的互动数较高,为 51 到 100;第 5 级的互动数最高,具有最高的大众认可度,即互动数大于 100 的微博。

表 1 微博数据格式

Tab.1 Micro-blog data format

| 字段 | 字段说明 |
|---------------|---------------|
| uid | 用户标记(加密) |
| mid | 博文标记(加密) |
| time | 发博时间(精确到天) |
| content | 博文内容 |
| comment_count | 博文发表 1 周后的评论数 |
| like_count | 博文发表 1 周后的赞数 |
| forward_count | 博文发表 1 周后的转发数 |

由于数据中只有微博文本的发表时间等信息,其所具有的特征信息比较稀少,难以直接进行有效的分析,需要提取用户发表的微博文本背后的一些特征信息,主要分为用户特征和微博特征。用户特征指的是用户发表微博所得到的互动数的特点,而微博特征指的是微博文本本身的特点使其互动数发生变化的特性。

用户发表的微博特性,主要关注于用户自身是否是一个比较受欢迎,被大量用户关注的用户,即其具有的粉丝数量等,可以从用户以往发表的微博的互动数情况进行侧面反映。本研究提取了用户的 11 个特性如表 2 所示。

表 2 用户特征

Tab.2 User characteristics

| 特征 | 特征说明 |
|---------------|-----------------------|
| weiboSum | 发表的微博总数 |
| tapOneNum | 发表微博属于第 1 级数占发表微博总数比例 |
| tapTwoNum | 发表微博属于第 2 级数占发表微博总数比例 |
| tapThreeNum | 发表微博属于第 3 级数占发表微博总数比例 |
| tapFourNum | 发表微博属于第 4 级数占发表微博总数比例 |
| tapFiveNum | 发表微博属于第 5 级数占发表微博总数比例 |
| avgInteract | 发表的微博的平均互动数 |
| avgInteactTap | 发表的微博的平均互动数所属的级别 |
| interactSum | 发表的微博总的互动数 |
| avgDayNum | 平均每天发表的微博数量 |
| publishRate | 每天发表微博的频率 |

微博文本的特征,主要是通过识别微博文本本身的特性,判断其是否是一条受大众认可喜欢的微博文本,对以往的微博文本进行挖掘提取,判断微博是否是原创微博,微博中“@”符号的个数,微博中是否有网页链接等特点。本研究提取微博文本 7 个主要特征,如表 3 所示。

表 3 微博文本特征

Tab.3 Micro-blog text features

| 特征 | 特征说明 |
|----------------|----------|
| publish_hour | 微博发表的时间段 |
| exist_aite | 是否存在@ 符号 |
| aite_num | @ 符号的个数 |
| exist_http | 是否有链接 |
| http_num | 链接个数 |
| original_weibo | 是否为原创微博 |
| content_length | 微博文本长度 |

2.2 流程设计

基于 Spark 框架对所设计的微博互动数预测流程如下:

- (1)对原始数据进行采集划分,得到训练数据和测试数据两个数据集;
- (2)对数据进行特征的提取转化等预处理操作;
- (3)使用基于 Spark 的分类算法对训练数据集进行训练,得到分类模型;
- (4)使用得到的分类模型对测试集数据进行分类以预测其未来的互动数所属级数;
- (5)对分类得到的结果进行验证,得到分类模型的准确度,具体流程如图 2 所示。

2.3 评估指标

对分类模型的准确性与有效性指标的判定可以通过其混淆矩阵进行计算得到^[20]。通过混淆矩阵(表 4)可以计算分类模型的正确率如公式 3 所示,正确率越高代表模型分类结果越好。

表 4 分类结果混淆矩阵

Tab.4 Classification results confusion matrix

| 实际类别 | 预测类别 | |
|------|------|----|
| | Yes | No |
| Yes | TP | FN |
| No | FP | TN |

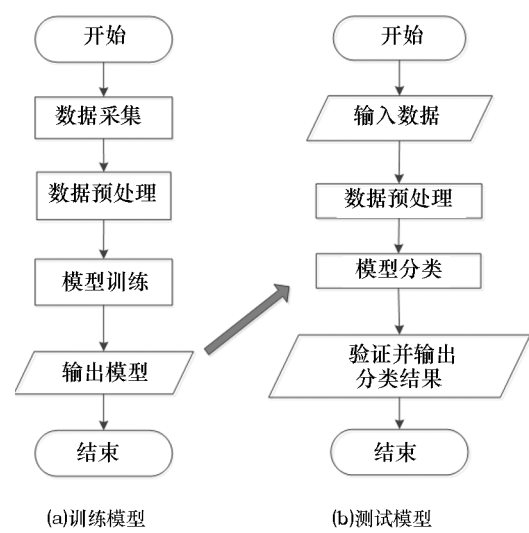


图 2 决策树训练测试流程

Fig.2 Parallel decision tree training testing process

$$Acc = \frac{(TP + TN)}{(TN + TP + FP + FN)} \quad (3)$$

由于对微博文本的互动数的预测是预测其所属的级数,通过不同的级数可看出该微博的一个受认可度情况,对不同的微博文本预测结果更看重互动数级数高的微博能否被分类正确。对微博互动数预测结果根据不同的级数赋予不同的权重值如表 5 所示,最后计算所有微博的分类结果的加权分数,分数越高代表分类结果越好,如公式 4 所示。

表 5 权重系数
Tab.5 Weight coefficient

| 级数 | 权重值 | 级数 | 权重值 |
|----|-----|----|-----|
| 1 | 1 | 4 | 100 |
| 2 | 10 | 5 | 200 |
| 3 | 50 | | |

$$Score = \frac{\sum_{j=1}^n weight_j \times g(a, b)}{\sum_{j=1}^n weight_j} \quad (4)$$
$$g(a, b) = \begin{cases} 1, & \text{if } (a = b) \\ 0, & \text{else.} \end{cases}$$

3 实验结果分析

3.1 精度分析

由于微博文本的时效性特征,对微博文本的互动数预测,应从距离微博文本较近的时间段内的数据进行用户特征的提取,所以选取 2015 年 4 月到 6 月共 3 个月的微博文本数据作为训练集,用以构建分类模型,用 2015 年 7 月份的数据作为测试数据,以验证分类模型的准确度。

通过与基于 Spark 的决策树 (decision trees, DT)、基于 Spark 的朴素贝叶斯 (naive Bayes, NB) 和基于 Spark 的逻辑回归 (logistic regression, LG) 在训练数据集上进行构建分类模型,在测试集上验证分类模型得到的结果如表 6 所示。

表 6 实验结果

Tab.6 Experimental results

| 算法 | 转发 | | 评论 | | 点赞 | |
|----|-------|-------|-------|-------|-------|-------|
| | Acc | Score | Acc | Score | Acc | Score |
| DT | 0.973 | 0.691 | 0.981 | 0.762 | 0.979 | 0.760 |
| NB | 0.889 | 0.430 | 0.863 | 0.531 | 0.894 | 0.543 |
| LG | 0.948 | 0.382 | 0.959 | 0.495 | 0.956 | 0.525 |

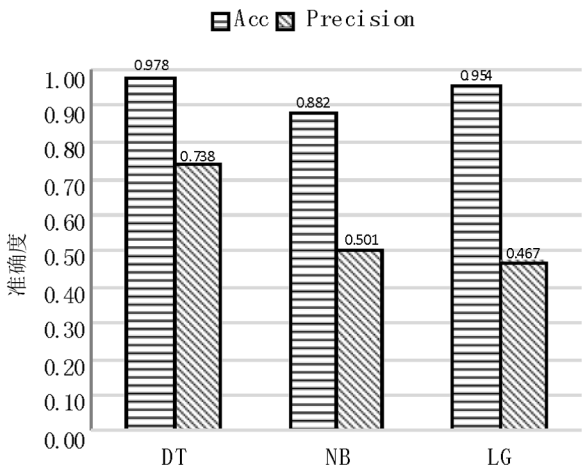


图 3 互动数平均实验结果

Fig.3 The average of interaction experimental result

从实验结果可以看出,从微博文本中提出的用户特征和微博文本特征,能够使分类算法有效地对新的微博文本进行预测分类,3 个算法均有较高的正确率,但本研究所提出的基于决策树的分类结果具有最高的正确率。同时通过对不同的

微博文本级数的分类, 本研究所提出的决策树方法分类的结果的 Score 得分最高, 能够对微博互动数级数高的文本进行正确的分类, 而另外两个分类算法虽然有较高的正确率, 但在级数高的微博文本中未能有效地识别, 导致其 Score 得分不高。

3.2 扩展性实验

为进一步验证算法的可扩展性能力, 通过使用不同的集群规模对所提出的方法进行扩展性实验, 计算算法运行时的加速比, 公式如 5 所示:

$$S = T_s / T_p \tag{5}$$

其中 T_s 表示单机版算法运行所消耗的时间, T_p 表示并行版算法运行所消耗的时间, p 表示并行的节点个数。算法的加速比结果如图 4 所示。

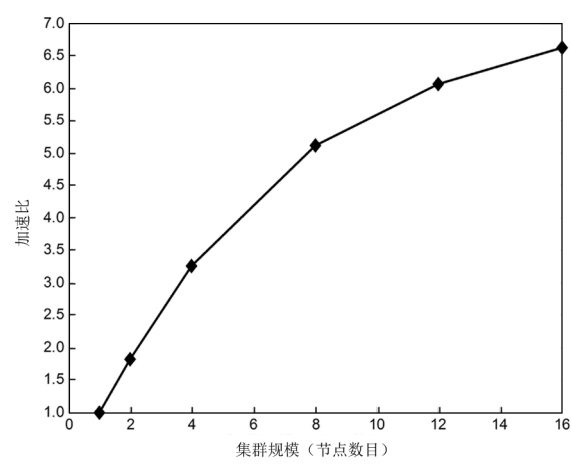


图 4 不同集群规模加速比

Fig.4 Different clusters scale acceleration ratios

从图 4 不同集群规模加速比的实验结果可以发现, 随着集群规模的增加, 算法运行的速度加快, 加速比增加。当集群规模从 1 增加到 4 时, 加速比增长迅速, 因为算法将数据分散到各个节点进行运行, 进而减少了算法处理所需要的时间, 大大提高了整体的运行速度。但随着集群规模的不断增加, 加速比的增长速度变慢, 这是因为随着集群规模的增加, 算法需要耗费更多的时间在数据传输和调度上, 从而导致了加速比增长缓慢。可见所提出的基于 Spark 的并行决策树方法具有较好的可扩展性能力。

4 结语

新浪微博作为国内主要的社交媒体平台, 如何对一条微博文本的互动数进行有效的预测, 进而根据互动数级数对微博文本进行有效的分发控制管理具有非常重要的意义。本研究首先通过对用户发表的微博进行有效的用户特征和微博文本自身特征的提取。然后基于 Spark 分布式框架使用决策树分类算法对数据进行分类模型的构建。最后在新的微博文本上使用分类模型进行分类以验证分类模型的有效性。通过与并行的朴素贝叶斯和逻辑回归分类算法的对比实验, 验证所提出的基于决策树分类算法的微博互动数预测的有效性与可扩展性能力, 能够对微博文本未来的互动数级数进行正确的分类。接下来, 将对微博的文本内容进行内容挖掘分析研究, 提取更多有价值的特征, 以进一步提高互动数级数高的微博文本的分类正确率。

参考文献:

[1] 王洁, 汤小春. 基于社区网络内容的个性化推荐算法研究[J]. 计算机应用研究, 2011, 28(4): 1248-1250.

[2] Yang B, Cheung W, Liu J. Community mining from signed social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(10): 1333-1348.

[3] Phuvipadawat S, Murata T. Breaking news detection and tracking in Twitter[C]//2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Aug. 31 - Sept. 3, 2010, Toronto, Ontario, Canada. Washington: IEEE Computer Society, 2010, 3: 120-123.

[4] 曹玖新, 吴江林, 石伟, 等. 新浪微博网信息传播分析与预测[J]. 计算机学报, 2014, 37(4): 779-790.

[5] Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter[C]//43rd Hawaii International Conference on System Sciences (HICSS), Koloa, Kauai, Hawaii, Jan 5-8, 2010. Washington: IEEE, 2010: 1-10.

[6] Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media? [C]//Proceedings of the 19th International Conference on World Wide Web. Apr 26-30, 2010, Raleigh, North Carolina, USA. New York: ACM, 2010: 591-600.

[7] Stern D H, Herbrich R, Graepel T. Matchbox: large scale online Bayesian recommendations[C]//Proceedings of the 18th International Conference on World Wide Web. Apr 20-24, 2009, Madrid, Spain. New York: ACM, 2009: 111-120.

[8] Zaman T R, Herbrich R, Van Gael J, et al. Predicting information spreading in twitter[J]. Computational Social Science

and the Wisdom of Crowds. Citeseer, 2010, 104(45):17599–17601.

- [9] Yang Z, Guo J, Cai K, et al. Understanding retweeting behaviors in social networks[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Oct 26–30, 2010, Toronto, ON, Canada. New York: ACM, 2010:1633–1636.
- [10] Liben-Nowell D, Kleinberg J. Tracing information flow on a global scale using Internet chain-letter data[J]. Proceedings of the National Academy of Sciences, 2008, 105(12):4633–4638.
- [11] Fan P, Li P, Jiang Z, et al. Measurement and analysis of topology and information propagation on Sina-Microblog[C]//2011 IEEE International Conference on Intelligence and Security Informatics (ISI), July 9–12, 2011, Beijing China. Washington: IEEE, 2011:396–401.
- [12] Webberley W, Allen S, Whitaker R. Retweeting: A study of message-forwarding in twitter[C]//2011 Workshop on Mobile and Online Social Networks (MOSN), Sept 8, 2011, Milan, Italy. Washington: IEEE, 2011:13–18.
- [13] 谢婧, 刘功申, 苏波, 等. 社交网络中的用户转发行为预测[J]. 上海交通大学学报, 2013, 47(4):585–588.
- [14] 匡冲, 刘知远, 孙茂松. 微博转发者的个性化排序[J]. 山东大学学报(理学版), 2014, 49(11):31–36.
- [15] 严玉良, 董一鸿, 何贤芒, 等. FSMBUS: 一种基于 Spark 的大规模频繁子图挖掘算法[J]. 计算机研究与发展, 2015, 52(8):1768–1783.
- [16] 丁圣勇, 闵世武, 樊勇兵. 基于 Spark 平台的 NetFlow 流量分析系统[J]. 电信科学, 2014, 30(10):48–51.
- [17] 牛海玲, 鲁慧民, 刘振杰. 基于 Spark 的 Apriori 算法的改进[J]. 东北师大学报(自然科学版), 2016, 48(1):84–89.
- [18] 徐鹏, 林森. 基于 C4.5 决策树的流量分类方法[J]. 软件学报, 2009, 20(10):2692–2704.
- [19] 韩家炜, 坎伯. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2012:213–222.
- [20] 陈羽中, 郭松荣, 陈宏, 等. 基于并行分类算法的电力客户欠费预警[J]. 计算机应用, 2016, 36(6):1757–1761.

(特约编辑: 黄家瑜)

(上接第 293 页)

通过对常用标志位 ZF、SF、CF 和 OF 生成逻辑的分析、实现及应用的详细阐述, 可以更好地了解运算器的工作原理及设计方法, 进一步理解运算器、指令系统以及控制器三者之间的联系, 提高汇编程序设计能力。

参考文献:

- [1] 王诚, 宋佳兴. 计算机组成与结构[M]. 2 版. 北京: 清华大学出版社, 2011.
- [2] 秦磊华, 王小兰, 管军. EDA 仿真在组成原理设计性实验中的应用[J]. 实验室研究与探索, 2009, 28(4):79–82.
- [3] 张磊, 郑榕, 张军峰. 计算机组成原理理论实验教学无缝结合的新方法[J]. 实验室研究与探索, 2013, 32(5):168–171.
- [4] 白中英. 计算机组成与系统结构: 第五版(立体化教材)[M]. 北京: 科学出版社, 2011.
- [5] 裘雪红, 李伯成. 计算机组成与体系结构[M]. 北京: 高等教育出版社, 2009.
- [6] 唐朔飞. 计算机组成原理[M]. 2 版. 北京: 高等教育出版社, 2010.
- [7] 王爱英. 计算机组成与结构[M]. 5 版. 北京: 清华大学出版社, 2013.
- [8] 袁波, 李树荣, 姚素英. 一种 8 位单片机中 ALU 的改进设计[J]. 微电子学与计算机, 2006, 23(4):71–74.
- [9] 朱一杰, 张曦, 俞军. 算术逻辑单元的优化设计[J]. 微电子学与计算机, 2004, 21(9):155–157.
- [10] 宋文强. 12 位 RISC 计算器设计[D]. 成都: 电子科技大学, 2011.
- [11] 齐广玉, 邹建伟. 计算机组成原理[M]. 北京: 兵器工业出版社, 1997.
- [12] 沈美明. IBM PC 汇编语程序设计[M]. 北京: 清华大学出版社, 1996.

(特约编辑: 黄家瑜)