

基于数据挖掘的弱关系社交网络及弱关系强化的研究

潘树银 高建瓴

(贵州大学大数据与信息工程学院 贵阳 550025)

摘要 以新浪微博作为弱关系社交网络的代表。针对弱关系社交网络,在聚类和相关性方面做了详细分析,并对弱关系社交网络中弱关系的强化做了适当的条件假设。然后,随机选取了若干不同类型的话题微博,并获取了不同数量的样本点;对话题微博进行研究与分析,用实验证明了该假设的合理性,并找到了弱关系变强的具体条件。

关键词 社交网络,强关系,弱关系,凝聚层次聚类,相关系数

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.8.046

Research on Weak Relation Social Network and Weak Relation Strengthening Based on Data Mining

PAN Shu-yin GAO Jian-ling

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract Sina microblog is a representative of the social network which is weakly related. Aiming at the relevance among user behaviors of weak relational social network, this paper analyzed them in clustering and correlation in detail, and made certain assumptions at the same time. Then we randomly selected a number of topic microblog from different aspects, and got different number of sample points. Based on these topic microblog, this paper carried out research and analysis. The results prove the rationality of the hypothesis, and the specific conditions of weak relation getting stronger are found afterwards.

Keywords Social network, Strong relation, Weak relation, Hierarchical clustering, Correlation coefficient

1 引言

社交网络是虚拟的社会网络,由多个节点构成,可看作现实社会关系在网络中的映射与延伸。社交网络是由节点与边构成的图结构,社交网络的节点代表着社会关系中的个人或组织,连接两个节点的边体现社会关系,即社交网络用户可以将现实的社会关系虚拟化,以便于找到与自己兴趣爱好相似或相同的朋友圈^[1,2]。

随着在线社交的快速发展,有着不同行业背景的各类商家也纷纷开发出独具特色的社交网络产品。由此,实时了解社交网络用户的兴趣变化,明晰用户对社交网络产品的行为反馈,变得尤为重要。对诸如此类信息的洞悉不仅有助于商家制定营销策略,还能使消费者的个性化需求得到更好的满足^[3]。

正是由于社交网络研究具有重大的商业价值与应用价值,大量研究人员转向对网络关联关系的研究。在网络关联关系的研究中,社交网络用户的关注、转发、评论、提及等关联关系中隐藏着许多有价值的知识,挖掘这些数据中隐藏的知识已经成为近几年研究的热点。

2 社交网络研究的理论基础

2.1 关系型社交网络

根据社交网络用户间的联系紧密与否、用户之间产生互

动频率的高低,社交网络中的社交关系可以分为强关系、弱关系两种。强关系是指社交网络中的组织或个人同质性较强,弱关系则是指异质性较强。在社交网络中,维护感情、相关利益关系用强关系连边来表征,联系较少的弱情感关系用弱关系连边来表征,强关系的存在与调整可以推动社交网络的发展,弱关系的存在与调整则影响着社交网络信息的传播和网络结构的维护。社交网络的强关系表现为对可靠信息的接受度,而弱关系则表现为社交网络信息的传播广度与速度^[4],因此强关系与弱关系在社交网络的发展与维护方面都有着十分重要的作用。

强关系社交网络中的关系以增强联系、维系感情为主,通过关注好友的动态,快速回复,互动越多,感情越深。以QQ空间为代表,其特点是:1)圈子小,好友间为匹配关系;2)优质内容与闲聊内容并存;3)加入弱关系社交网络的用户更倾向于维系感情。

弱关系社交网络中的关系以展现自我、提高自我为主,普通用户从内容、形式等方面模仿优秀用户,以此提高发布内容的质量或热度。弱关系社交网络用户通常采用这种方式来提高自己的影响力,使自己能够成为弱关系社交网络中小群体或大群体中的名人。新浪微博、YY语音等是典型的弱关系社交网络,其特点是:1)圈子更大,可为单向,也可互粉;2)内容更优质;3)加入弱关系社交网络的用户更倾向于以此来提高自我和提高知名度。

到稿日期:2015-07-12 返修日期:2015-09-24 本文受贵州省科学技术基金项目:云环境中半监督聚类算法的并行化研究(黔科合J字[2015]2045)资助。

潘树银(1990-),女,硕士生,主要研究方向为数据挖掘,E-mail:psy0410chen@163.com;高建瓴(1969-),女,硕士生导师,主要研究方向为数据挖掘、云计算。

2.2 社交网络用户行为

在弱关系社交网络新浪微博中,用户可以采取的行为主要包括以下几种:发布微博、关注某人/取消对某人的关注、浏览所关注用户发布的微博、给某人发私信、对微博评论、转发微博、收藏微博。

Meeyoung 等人^[5]分别根据入度、转发与提及 3 个指数对 Twitter 消息传播性质进行了研究,实验结果表明粉丝数量多的用户不一定被很多人引用且其消息不一定被转发很多次。

Romero 等人^[6]对社交网络进行研究,实验表明用户影响力不仅取决于社交网络用户粉丝的数量,还与粉丝的消极性有关联关系,而粉丝的消极性不属于用户行为的范畴,而是属于社交网络中用户的情感。在社交网络消息的传播过程中,某用户看到其好友发消息时,转发用户行为伴随着转发概率 P 这一指标,而这个转发概率 P 与该用户以往转发同一好友的消息比例相关。Romero 等人^[6]的研究表明用户消息被浏览的次数与用户粉丝数量之间的关联比较弱。

3 相关的数据挖掘算法介绍

3.1 层次聚类

层次聚类是采用某种方法对数据集进行层次分解的一种数据挖掘算法,其分布为树结构,除叶结点以外,树中每个结点都是该结点的子女的并,树根是树结构中的一种特殊情况,是包含所有对象的簇^[7]。层次聚类按分类原理的不同,可以分为凝聚聚类和分裂聚类两种方法。凝聚的层次聚类是自底向上^[8],分裂的层次聚类采用的策略与凝聚的层次聚类策略相反,是自顶向下的策略^[9]。

3.2 相关分析

数据分析和数据挖掘的过程中,在了解个体间差异的大小后评价个体的相似性和类别,是最常用的一种方法。在此类方法中,最常见的是数据分析中的相关分析^[10]。

Pearson 简单相关系数用来度量定距型变量之间的线性相关关系。其数学定义如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Spearman 等级相关系数用来度量定序变量的线性相关关系,该系数的设计思想与 Pearson 简单相关系数完全相同。但是,在涉及相关计算时,由于数据是非定距型的,因此计算时是利用数据的秩。其数学定义如下:

$$r = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad (2)$$

4 社交网络弱关系分析

以新浪微博为弱关系社交网络的代表进行分析,新浪微博的多种用户行为都是只有用户本人或者新浪微博系统平台才能检测到。作为一个第三方的围观者,我们能够从公开的数据中得到的用户行为数据就是用户发布微博的数量、收藏微博的数量、关注其他用户的数量及粉丝的数量。

将发布微博的数量收藏微博的数量、关注其他用户的数量及粉丝的数量作为 4 个变量,进行以下几方面的分析。

4.1 弱关系社交网络的聚类分析

选用 3055 条新浪微博数据进行层次聚类分析,并将用户粉丝数(fans)、用户关注数(concern)、用户微博数(blog)、用户收藏数(collection)作为变量,用户 id 作为标记。个体距离采用平方欧氏距离,类间距离采用组间平均链锁距离,由于数据存在数量级上的差异,先对数据进行标准化处理,得到微博用户变量的相似度矩阵,如表 1 所列。

表 1 weibo_user 相似度矩阵

Case	Matrix File Input			
	Fans	Concern	Blog	Collection
Fans	0.000	111.415	48.627	60.648
Concern	111.415	0.000	83.117	103.376
Blog	48.627	83.117	0.000	22.892
Collection	60.648	103.376	22.892	0.000

对 3055 条新浪微博数据进行层次聚类分析,得到微博用户的聚类图,如图 1 所示。

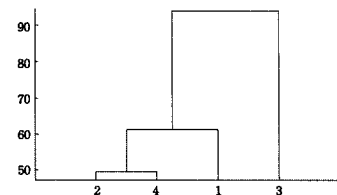


图 1 weibo_user 层次聚类图

分析以上结果,得出当聚成 3 类时,用户收藏数与用户关注数为一类,其他指标各成一类;当聚成 2 类时,用户粉丝数、用户关注数、用户收藏数为一类,用户微博数自成一类。用户收藏数与用户关注数的距离最近,其次与用户粉丝数距离相近,与用户微博数距离最远。

4.2 弱关系社交网络相关性分析

由于变量均为定距型变量,因此用皮尔森相关系数来度量用户粉丝数、用户关注数、用户微博数、用户收藏数这 4 个变量两两之间的关系,得出相关系数,如表 2 所列。

表 2 weibo_user 相关系数表

Variable	Index	Fans	Blog	Collection	Concern
Fans	Pearson	1	0.347**	0.040*	0.218**
	Sig.		0.000	0.026	0.000
	N	3055	3055	3055	3055
Blog	Pearson	0.347**	1	0.180**	0.344**
	Sig.	0.000		0.000	0.000
	N	3055	3055	3055	3055
Collection	Pearson	0.040*	0.180**	1	0.180**
	Sig.	0.026		0.000	0.000
	N	3055	3055	3055	3055
Concern	Pearson	0.218**	0.344**	0.180**	1
	Sig.	0.000	0.000	0.000	
	N	3055	3055	3055	3055

上述 4 个变量之间都存在显著相关关系,但是相关性较弱。变量间的 6 组相关关系按由强到弱排序为:微博-粉丝、微博-关注、关注-粉丝、微博-收藏、收藏-关注、收藏-粉丝。由此得出,经常发微博的用户可能更易去关注其他用户,也更容易引起其他用户的关注,与其他用户相比,粉丝数易更多。接下来,结合散点图,分析相关系数较低是否是由异常点造成的。分别对微博-关注、关注-粉丝、微博-粉丝作散点图,如图 2—图 4 所示。

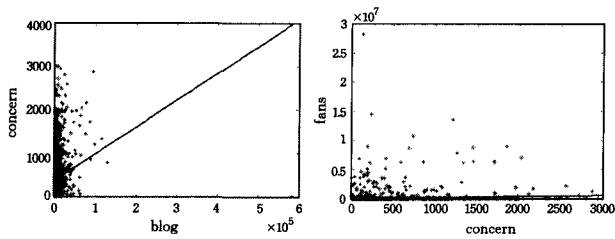


图2 微博-关注散点图

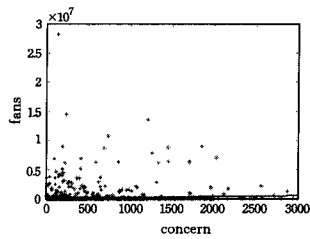


图3 关注-粉丝散点图

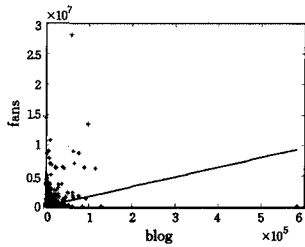


图4 微博-粉丝散点图

结合散点图分析出,异常点较少,线性拟合程度不高;但样本点较多,因此少量的异常点不足以影响变量间的相关性,变量相关性较弱并非异常点导致。同时,选取了不同数量的新浪微博数据进行了实验,发现样本点的数量对变量的相关性影响不大。

在上述研究结论的启发下,做出以下思考:弱关系能否变为强关系,而这种关系的转变又需要什么条件?这个问题值得做出理论、实践上的探究。下面对该问题进行研究。

5 社交网络弱关系的强化

通过分析强关系社交网络与弱关系社交网络,做出以下假设:弱关系社交网络中,通过找到与自己兴趣爱好相似或相同的人,弱关系可以变强。该假设也可表述为:弱关系社交网络中,因共同爱好产生的良好关系可以改善弱关系用户不匹配产生的用户行为相关性弱的问题。

但该假设也具有一定的局限性,即只有社交网络限定在一定的圈子内,因共同爱好产生的良好关系才可以改善弱关系用户不匹配产生的用户行为相关性弱的问题。若社交网络涉及的范围广泛,则该假设不再适用。

为了验证弱关系社交网络中弱关系强化的这一猜想,做如下实验进行分析。

为了排除话题本身对分析的影响,此次分析选取的话题微博覆盖了生活、体育、企业、通讯、电影5个方面。从这5类微博中分别随机选取了转基因、恒大、火箭、魅族、同桌的你5个话题的微博,对每个话题微博分别从转发、评论、态度3种用户行为进行分析。

另外,为了排除数据量对分析结果的影响,对每一种话题微博都选用了不同数量的样本点进行多次实验,因图表较多,不在文中一一列出。对每种话题微博,只随机选择一种实验结果作为代表。对转基因话题微博随机选取了491条数据,对恒大话题微博随机选取了1051条数据,对火箭话题微博随机选取了1015条数据,对魅族话题微博随机选取了1143条数据,对同桌的你话题微博随机选取了718条数据。

选用转基因话题微博数据,将转发(repost)、评论(comment)、态度(attitude)3个行为作变量,用皮尔森相关系数分析得到表3所列结果。

表3 转基因相关系数

Variable	Index	Repost	Comment	Attitude
Repost	Pearson	1	0.844**	0.834**
	Sig.		0.000	0.000
	N	491	491	491
Comment	Pearson	0.844**	1	0.768**
	Sig.	0.000		0.000
	N	491	491	491
Attitude	Pearson	0.834**	0.768**	1
	Sig.	0.000	0.000	
	N	491	491	491

转基因微博分析中,3个变量间都存在显著的相关性,且相关性较强。当样本点增多时,3个变量两两之间都呈现了极强的相关性。

选用恒大话题微博数据,将转发、评论、态度3个行为作变量,用皮尔森相关系数分析得到表4所列结果。

表4 恒大相关系数

Variable	Index	Repost	Comment	Attitude
Repost	Pearson	1	0.777**	0.697**
	Sig.		0.000	0.000
	N	1051	1051	1051
Comment	Pearson	0.777**	1	0.754**
	Sig.	0.000		0.000
	N	1051	1051	1051
Attitude	Pearson	0.697**	0.754**	1
	Sig.	0.000	0.000	
	N	1051	1051	1051

恒大微博分析中,4个变量间都存在显著的相关性,相关性比非话题微博分析时的剪强,但与同话题微博样本点选取较多的实验结果相比,变量间的相关性较弱。结合散点图进行分析,分别对评论-态度、转发-态度、评论-转发作散点图,如图5—图7所示。

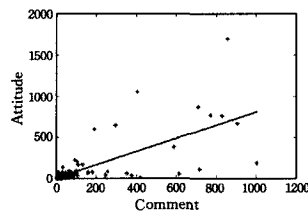


图5 恒大评论-态度散点图

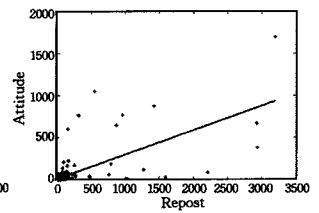


图6 恒大转发-态度散点图

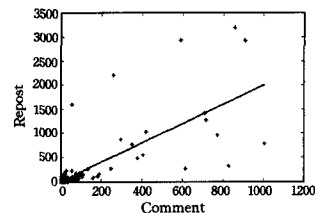


图7 恒大评论-转发散点图

结合散点图分析出,异常点是存在的,且在某种程度上影响了变量间的相关性。

选用火箭话题微博数据,将转发、评论、态度3个行为作变量,用皮尔森相关系数分析得到表5所列结果。

表5 火箭相关系数

Variable	Index	Repost	Comment	Attitude
Repost	Pearson	1	0.707**	0.873**
	Sig.		0.000	0.000
	N	1015	1015	1015
Comment	Pearson	0.707**	1	0.837**
	Sig.	0.000		0.000
	N	1015	1015	1015
Attitude	Pearson	0.873**	0.837**	1
	Sig.	0.000	0.000	
	N	1015	1015	1015

3个变量之间都存在显著相关关系,其中两组变量的相关性很强,一组变量的相关性一般。3组相关关系按由强到弱排序为:转发-态度、评论-态度、评论-转发。对相关性一般的评论-转发变量组作散点图,如图8所示。

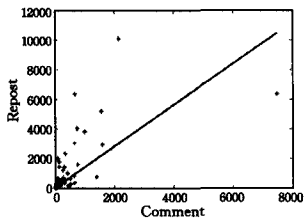


图8 火箭评论-转发散点图

结合散点图分析出,异常点是存在的,并在某种程度上影响了变量间的相关性。

选用魅族话题微博数据,将转发、评论、态度3个行为作变量,用皮尔森相关系数分析得到表6所列结果。

表6 魅族相关系数

Variable	Index	Repost	Comment	Attitude
Repost	Pearson	1	0.978**	0.955**
	Sig.		0.000	0.000
	N	1143	1143	1143
Comment	Pearson	0.978**	1	0.939**
	Sig.	0.000		0.000
	N	1143	1143	1143
Attitude	Pearson	0.955**	0.939**	1
	Sig.	0.000	0.000	
	N	1143	1143	1143

魅族话题微博分析中,3个变量间都存在显著的相关性,且相关性很强,接近完全正相关。样本点是5个话题中数量最多的,相关性非常强。当样本点增多时,转发、评论、态度3个变量两两之间的相关性十分接近完全正相关。

选用同桌的你话题微博数据,将转发、评论、态度3个行为作变量,用皮尔森相关系数分析得到表7所列结果。

表7 同桌的你相关系数

Variable	Index	Repost	Comment	Attitude
Repost	Pearson	1	0.565**	0.578**
	Sig.		0.000	0.000
	N	718	718	718
Comment	Pearson	0.565**	1	0.916**
	Sig.	0.000		0.000
	N	718	718	718
Attitude	Pearson	0.581**	0.916**	1
	Sig.	0.000	0.000	
	N	718	718	718

3个变量之间其中一组变量的相关性很强,两组变量的

相关性一般,但变量间的相关性比非话题微博分析时的剪性强。3组相关关系按由强到弱排序为:评论-态度、转发-态度、转发-评论。对相关性一般的转发-评论变量组、转发-态度变量组分别作散点图,如图9、图10所示。

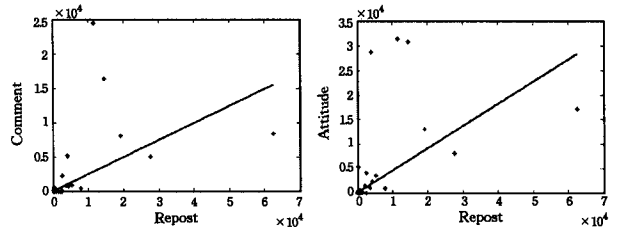


图9 同桌的你转发-评论散点图 图10 同桌的你转发-态度散点图

结合散点图分析,异常点是存在的,且数据中的有价值数据点少,这在某种程度上影响了变量间的相关性。

通过话题微博分析可以看出,其变量之间的剪剪性明显强于非话题微博时变量间的相关性,且在话题微博研究中,样本点的数量对变量的剪剪性影响较大。在一定范围内时,样本点越多,变量的剪剪性越强。实验验证了本文的猜想,在弱关系社交网络限定在一定的范围内时,用户关系要从弱变强,需要找到共同兴趣爱好的人。因共同爱好产生的良好关系可以改善弱关系用户不匹配产生的用户行为剪剪性弱的问题,拥有共同点越多,产生的话题也会越多,弱关系也越容易增强。

结束语 针对弱关系社交网络的用户行为间的相关性做了详细分析,同时对弱社交网络弱关系的强化做了假设,证明了该假设的正确性,找到了弱关系变强的具体条件。虽然文中对弱关系社交网络做了详尽的分析,并证明了弱关系可通过某些条件变强的猜想,但证明过程不尽完善,还需要做出进一步的改进。对比弱关系社交网络与强关系社交网络,探究弱关系与强关系之间的转化联系是笔者需要进一步研究的工作。

参考文献

[1] Chen Shao-qin. Research on the real time influence of social network users based on PageRank [D]. Shanghai: Shanghai Jiao Tong University, 2013 (in Chinese)
陈少钦. 基于PageRank的社交网络用户实时影响力研究[D]. 上海:上海交通大学, 2013

[2] Li Tao-tao, Zhou Bin, Wang Zhong-zhen. Research on the application of social networks based on graph data mining [J]. Computer Technology and Development, 2014, 24(10): 7-9 (in Chinese)
李桃陶, 周斌, 王忠振. 基于社交网络的图数据挖掘应用研究[J]. 计算机技术与发展, 2014, 24(10): 7-9

[3] Ding Zhen-guo, Song Wei, Li Jing. Analysis of social network user behavior based on sequential pattern mining [J]. Modern Information, 2013, 33(3): 56 (in Chinese)
丁振国, 宋薇, 李靖. 基于序列模式挖掘的社交网络用户行为分析[J]. 现代情报, 2013, 33(3): 56

表4 MinMax K 均值算法与改进算法的异常检测性能

数据集	MinMax K 均值算法异常检测			改进 K-mean 异常检测		
	检测率 (%)	误检率 (%)	平均用时 (ms)	检测率 (%)	误检率 (%)	平均用时 (ms)
Iris	70.3	20.8	892	85.6	6.3	916
Ecoli	68.1	25.6	946	79.3	8.9	1025
Yeast	55.2	36.5	1566	73.4	11.2	1332

实验表明,与原 K-mean 算法相比,无论是在检测率还是误报率方面,改进算法都优于原算法;而且在算法的平均运行时间方面,改进后的算法也更加高效。由于原算法选择聚类中心的随机性,导致算法可能选择异常点或其附近的点作为初始聚类中心,从而使聚类结果产生较大误差,因此改进算法在误检率方面明显优于原算法。

与 MinMax K 均值算法相比,改进算法用时偏多。但 MinMax K 均值算法运行得到的结果并非最优结果,所以在检测率和误检率方面,该算法性能低于改进算法。

结束语 本文通过均匀选择数据紧密区域及避免离群点区域,优化了 K 均值算法的初始聚类中心选取过程,提出了基于改进 K 均值聚类的异常检测算法。实验证明,改进之后的算法不仅在聚类性能方面更加高效,而且在异常检测方面也比其他算法更加具有优势。随着越来越多大数据的产生,研究面向更高维度、更大规模、适合数据流形式的数据挖掘算法将是本文下一步努力的方向。

参考文献

- [1] Yang Yu-zhou. Research and implementation of the clustering anomaly detection technology based on feature extraction[D]. Chengdu: University of Electronic Science and Technology of China, 2012(in Chinese)
杨宇舟. 基于特征提取的聚类异常检测技术的研究与实现[D]. 成都: 电子科技大学, 2012
- [2] Sun Na, Guo Yan-feng, Yao Yuan. Network data stream abnormal detection model based on SVM incremental learning method [J]. Computer Engineering and Applications, 2012, 48(29): 78-81(in Chinese)
孙娜, 郭延锋, 姚远. 增量式 SVM 的数据流异常检测模型[J]. 计算机工程与应用, 2012, 48(29): 78-81
- [3] Luo Yong-jian. Research on Data Flow Anomaly Detection Algorithm Cluster-based[D]. Harbin: Harbin Engineering University, 2010(in Chinese)
骆永健. 基于聚类的数据流异常检测算法的研究[D]. 哈尔滨: 哈尔滨工程大学, 2010
- [4] Fu Ying-ding, Lan Ju-long. Kernel-based adaptation for affinity propagation clustering algorithm [J]. Application Research of Computers, 2012, 29(5): 1644-1650(in Chinese)
付迎丁, 兰巨龙. 基于核自适应的近邻传播聚类算法[J]. 计算机应用研究, 2012, 29(5): 1644-1650
- [5] Jiang Min, Pi De-chang, Sun Lan. Research on Density Clustering Algorithm with a Multiple Constraints [J]. Computer Science, 2011, 38(10A): 143-164(in Chinese)
江敏, 皮德常, 孙兰. 一种多约束的密度聚类算法的研究[J]. 计算机科学, 2011, 38(10A): 143-164
- [6] Celeb M, Kingravi H, Vela P. A Comparative Study of Efficient Initialization Methods for the K-methods for the K-Means Clustering Algorithm [J]. Expert Systems with Applications, 2013, 40(1): 200-210
- [7] Tzortzis G, Likas A. The minmax k-means clustering algorithm [J]. Pattern Recognition, 2011, 44(4): 866-876
- [8] Jiang Da-yu. A fast and efficient parallel bisecting K-Means algorithm [D]. Harbin: Harbin Engineering University, 2013 (in Chinese)
蒋大宇. 快速有效的并行二分 K 均值算法[D]. 哈尔滨: 哈尔滨工程大学, 2013
- [9] Zhu Jian-yu. Research and Application of K-means algorithm [D]. Dalian: Dalian University of Technology, 2013(in Chinese)
朱建宇. K 均值算法研究及其应用[D]. 大连: 大连理工大学, 2013
- [10] Han Zui-jiao. An Adaptive K-means initialization method based on data density [J]. Computer Applications and Software, 2014, 31(2): 182-187(in Chinese)
韩最蛟. 基于数据密集性的自适应 K 均值初始化方法[J]. 计算机应用与软件, 2014, 31(2): 182-187
- [11] Macqueen J. Some methods for classification and analysis of multivariate observe [C] // Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967: 281-297
- [12] Asuncion A, Newman D. UCI Machine Learning Respository [EB/OL]. [2015-06-01]. <http://archive.ics.uci.edu/ml/datasets.html>
- [4] Zhu Y X, Zhang X G, Sun G Q, et al. Influence of Reciprocal Links in Social Networks [J]. PLoS ONE, 2014, 9(7): e103007
- [5] Cha M, Haddadi H, Benevenuto F, et al. Measuring User Influence in Twitter: The Million Follower Fallacy [C] // International AAAI Conference on Weblogs and Social Media (ICWSM). 2010
- [6] Romero D M, Galuba W, Asur S, et al. Influence and Passivity in Social Media [C] // ECML PKDD. 2011
- [7] Gong Shang-fu, Chen Wan-lu, Jia Peng-tao. Research on clustering algorithm of hierarchical clustering [J]. Computing Applications, 2013, 30(11): 3217-3218(in Chinese)
龚尚福, 陈婉璐, 贾澎湃. 层次聚类社区发现算法的研究[J]. 计算应用研究, 2013, 30(11): 3217-3218
- [8] Duan Ming-xiu. Research and application of hierarchical clustering algorithm [D]. Changsha: Central South University, 2009(in Chinese)
段明秀. 层次聚类算法的研究及应用[D]. 长沙: 中南大学, 2009
- [9] Zhang Li-hua. Application of data mining technology in special celestial discovery [D]. Jinan: Shandong University, 2009(in Chinese)
张丽华. 数据挖掘技术在特殊天体发现中的应用研究[D]. 济南: 山东大学, 2009
- [10] Hou Bin. Typical correlation analysis algorithm based on sparse representation [D]. Nanjing: Nanjing University of Science and Technology, 2013(in Chinese)
侯彬. 基于稀疏表示的典型相关分析算法研究[D]. 南京: 南京理工大学, 2013

(上接第 232 页)