

# 面向大数据的并行聚类算法在股票板块划分中的应用

海沫<sup>1</sup>, 牛怡晗<sup>2</sup>, 张悦今<sup>1</sup>

1. 中央财经大学信息学院 北京 100081; 2. 上海浦东发展银行昆明分行 昆明 650000

## 摘要

上市公司的经营业绩在一定程度上反映股票的投资价值, 因此以反映上市公司盈利能力、偿债能力、成长能力、资产管理质量及股东获利能力5个方面共15项财务指标作为股票投资价值的衡量指标, 首次尝试使用面向大数据的并行聚类算法Mahout中的K-means聚类算法和模糊K-means聚类算法对中国A股市场约2 600支股票依据其财务指标进行聚类, 以便进行股票板块的划分, 并比较两种算法在不同距离度量方式下的迭代次数、执行时间、聚类间密度和聚类内密度。实验结果表明, 谷本距离度量方式下的K-means聚类效果最好, 因此可将该实验结果作为最终股票板块划分结果进行分析, 从而为投资决策提供参考。

## 关键词

财务指标; 并行聚类算法; K-means; 模糊K-means; 股票板块划分

doi: 10.11959/j.issn.2096-0271.2015036

## *Application of Parallel Clustering Algorithms for Big Data in the Division of Stock*

Hai Mo<sup>1</sup>, Niu Yihan<sup>2</sup>, Zhang Yuejin<sup>3</sup>

1. School of Information, Central University of Finance and Economics, Beijing 100081, China;

2. Kunming Branch, Shanghai Pudong Development Bank, Kunming 650000, China

## *Abstract*

For the operating performance of listed corporations reflects the value of stock investment to a certain extent, financial index reflecting the operating performance of listed corporations was taken as the evaluation index of stock investment value, and for the first time the parallel clustering algorithms for big data both K-means and fuzzy K-means of Mahout were used to cluster nearly 2 600 stock of China's A shares market according to their financial index, afterwards the clustering results of these two algorithms under different distance metrics were compared. Experimental results show that the clustering quality of K-means algorithm adopting Tanimoto distance metric is the best. Therefore, this result can be used as the final result of the division of stock, which can provide a reference for the investment decision.

## *Key words*

financial index, parallel clustering algorithm, K-means, fuzzy K-means, division of stock

## 1 引言

股票市场作为连接上市公司和投资者的桥梁,在金融系统中有不可替代的重要作用。随着中国股票市场的快速发展,如何在股票市场上选择具有投资价值的股票进行投资对投资者来说非常重要,关系到投资者的利益问题,但影响股票市场的因素很多,例如政治、政策、经济等因素都会不同程度地影响股票市场。上市公司的经营业绩在一定程度上反映股票的投资价值,而上市公司的财务指标是企业总结和评价财务状况和经营成果的相对指标,因而选取了能反映上市公司经营业绩的财务指标,并对股票按照这些财务指标进行合理的板块划分。通过对股票进行合理的板块划分,能帮助投资者准确地了解和把握股票的总体特征,确定投资范围,并通过各类的总体价格水平预测股票价格的变动趋势,选择有利的投资时机<sup>[1]</sup>。

聚类技术可以将股票市场上的股票根据特定的特征进行合理划分,得到具有指导意义的股票板块分类,便于投资者根据需求从合适的分类中选择股票进行投资。聚类结果对于投资者的指导作用主要包括<sup>[1]</sup>:了解各股票板块的基本特征及总体状况,初步划分出业绩优良和业绩一般的版块;根据选取的财务指标,了解各板块的特征,如收益性、成长性等,帮助投资者判断股票的投资价值;得到同一板块的均衡价格,找出受市场因素影响而低于这一价格的股票,将其视为具有升值空间且投资风险相对较小的股票。

本文以近2 600个上市公司的财务指标为依据,第一次在Hadoop平台上使用面向大数据的并行聚类算法Mahout软件库中的K-means算法和模糊K-means算

法对财务指标数据进行聚类。通过实验比较K-means算法及模糊K-means算法在Hadoop环境下对财务指标聚类的效率和质量,以找到适合大规模财务指标的并行化聚类方法,最终得到以财务指标为标的的股票板块划分,帮助投资者准确了解和把握股票的总体特性,选出各板块及板块内的绩优股和潜力股,使投资者能做出最佳的投资决策。

## 2 相关工作

聚类分析在股票市场板块分析中的研究可分为两类。

### 2.1 聚类指标体系的建立

周焯华等<sup>[2]</sup>将聚类分析方法引入证券投资分析中,对股票的行业因素、公司因素、收益性、成长性等基本层面进行考察,建立了较为全面的综合评价指标体系,以衡量样本股票的相似程度;然后通过聚类分析模型确定投资范围和投资价值。实证研究表明,该方法对指导证券投资具有有效性和实用性。劳兰珺等<sup>[3]</sup>提出对行业指数收益率序列分阶段进行聚类分析的动态分析方法,以考察行业间的相互关系及其演化过程;并基于深交所的行业指数数据进行实证研究,分析了各行业间的相似程度,有助于加深投资者及监管部门对行业间相互关系的了解,对投资决策具有参考价值。李云飞等<sup>[4]</sup>以全部上市公司为样本,通过模糊聚类和指标筛选得到了一个包含5个方面15项指标的股票投资价值评价指标集,为数据挖掘技术在股票价值投资方面的应用提供了实证依据。孙磊平<sup>[5]</sup>通过数据挖掘中的Logistic回归模型、决策树及神经网络模型3种方法对上市公司的财务

比率指标和股票投资价值的内在联系进行分析研究,从公司盈利能力、偿债能力、发展能力、运营能力以及现金流5个方面找出对股票投资价值有较大影响的上市公司财务指标。

## 2.2 聚类方法的选择和实现

邓秀勤<sup>[6]</sup>通过选取高科技板块中31个上市公司5个反映综合盈利能力的指标,运用SAS软件中的聚类过程Cluster对31个样本进行聚类,最终得到4个与公司的实际财务状况和经营状况相吻合的类。杨富勇<sup>[7]</sup>使用Clementine软件中的K-means、Kohonen和TwoStep 3种聚类算法,以13个反映上市公司五大方面情况的财务指标作为聚类指标体系对中国A股市场800多支股票进行聚类分析,发现TwoStep聚类方法在股票聚类分析过程中具有较好的分析结果。张传琦<sup>[8]</sup>使用优化的蚁群聚类算法从财务指标和个股收益率波动两方面对中国A股市场上1 800多支股票进行聚类,结果表明该算法对大样本数据进行聚类分析具有可行性和良好的效果。

综上所述,国内关于聚类算法在股票板块分析中的应用研究大都是以少于10维的财务指标作为聚类指标体系,并选取50支股票以下的小样本量进行聚类分析。参考文献[8]对1 800多支股票按照优化的蚁群聚类算法进行了聚类,但并没有采用面向大数据的并行聚类算法进行聚类。本文首次将面向大数据的并行聚类算法应用于整个中国A股市场的股票板块划分,在Hadoop平台下应用Mahout算法库中的聚类算法,基于选取的多于10维的财务指标,对目前A股市场上所有股票进行并行化聚类,从而得到整个A股市场的板块划分,对于投资者进行合理的投资决策具有重要的实际意义。

## 3 财务指标

影响股票投资价值的上市公司财务数据指标有很多,本文以李云飞、李鹏雁<sup>[4]</sup>通过对全部上市公司的财务数据进行模糊聚类后筛选得到的评价指标集作为参考,选取了能反映上市公司经营业绩5个方面共15项指标作为聚类实验的财务指标。这15项指标基本涵盖了股票投资价值所包含的全部信息。然后将数据进行标准化处理,用15维的数组作为聚类算法的输入。

### (1) 盈利能力指标

- 总资产净利润率=净利润/总资产余额;
- 净资产收益率=净利润/股东权益余额;
- 营业净利润=净利润/营业收入。

### (2) 偿债能力指标

- 流动比例=流动资产/流动负责;
- 现金比率=现金及现金等价物期末余额/流动负责;
- 资产负债率=负债合计/资产总计。

### (3) 资产管理质量指标

- 总资产周转率=营业收入/资产总额期末余额;
- 每股现金净流量=现金及现金等价物净增加额本期值/实收资本本期末值;
- 存货周转率=营业成本/存货期末余额。

### (4) 成长能力指标

- 总资产增长率=(资产总计本期期末值-资产总计本期期初值)/资产总计本期期初值;
- 营业利润增长率=(营业利润本年本期单季度金额-营业利润上一个单季度金额)/营业利润上一个单季度金额;

● 营业总收入增长度=(营业总收入本年本期金额-营业总收入上年同期金额)/营业总收入上年同期金额。

(5) 股东获利能力指标

● 每股营业收入=营业收入本期值/实收资本本期期末值;

● 每股营业收入=营业收入本期值/实收资本本期期末值;

● 市盈率=今收盘价当期值/(净利润上年年报值/实收资本本期期末值)。

## 4 实验结果及分析

本文实验数据为国泰安数据库上中国上市公司财务指标分析数据库中下载的所有上市公司的2014年6月的财务数据,有效数据共包含2 544个样本,每个样本包含15项财务指标,即有15个维度。对各项指标值进行标准化处理后,进行聚类实验。本实验环境为在一台阿里云服务器上搭建的伪分布式Hadoop平台,包括一个主节点和一个从节点。具体软硬件配置如下:CPU为双核;内存为2 GB;带宽为100 Mbit/s;操作系统为CentOS 6.5 64 bit;JDK为jdk-7u71-linux-x64;Hadoop版本为Hadoop 1.2.1;

Mahout版本为mahout-distribution-0.9。在该平台上使用Mahout中的K-means算法和模糊K-means算法对数据进行聚类。算法参数见表1。

对K-means和模糊K-means两种聚类算法在欧几里得距离、平方欧几里得距离、曼哈顿距离、余弦距离和谷本距离5种不同的距离测量方法下分别进行了财务指标聚类,并最终计算了每个实验的聚类间密度和聚类内密度<sup>[9]</sup>,将其作为聚类质量的评价指标。实验结果见表2和表3。

由表2和表3可以看出,对于该财务指标数据集,模糊K-means算法相比K-means算法在不同距离测度选择下都有更少的迭代次数和更短的执行时间,说明模糊K-means算法比K-means算法收敛得更快,有更高的执行效率。比较以上5种距离算法,模糊K-means和K-means算法在选择余弦距离作为距离计算参数时,都在1至2次迭代后就收敛,有很高的执行效率。但由于该距离算法不考虑两个向量的长度,只关注从原点到两个点的方向,因此分析两种算法选择余弦距离作为距离参数的聚类输出,发现50个分类中股票数量分布非常极端,有的类有几百只股票,而有的类仅有几只股票。这样的结果没有现实的

表1 聚类算法参数设置

算法	聚类个数	迭代次数	收敛阈值	模糊参数
K-means	20	50	0.5	/
模糊K-means	20	50	0.5	2

表2 不同距离算法选择下K-means的聚类结果

距离度量方法	迭代次数	收敛时间/min	聚类间密度	聚类内密度
欧几里得距离	20	6.57	0.093	0.553
平方欧几里得距离	15	4.951	0.091	0.554
曼哈顿距离	17.3	5.631	0.099	0.544
余弦距离	1.7	0.758	0.077	0.555
谷本距离	4.3	1.602	0.045	0.557

表3 不同距离算法选择下模糊K-means的聚类结果

距离度量方法	迭代次数	收敛时间/min	聚类间密度	聚类内密度
欧几里得距离	3.7	1.484	0.12	0.549
平方欧几里得距离	3.7	1.477	0.05	0.581
曼哈顿距离	5	1.929	0.124	0.544
余弦距离	1	0.585	0.115	0.573
谷本距离	7.7	2.85	0.063	0.489

指导意义,因此在对股票的财务指标聚类时不适合选择余弦距离作为距离参数。

在排除余弦距离算法后,比较K-means算法在不同距离测度选择下的聚类结果,当选择谷本距离作为距离算法参数时,K-means算法有最少的迭代次数和最短的执行时间,同时还有最小的聚类间密度和最大的聚类内密度,表明K-means算法在选择谷本距离对财务指标数据集进行聚类计算时有最高的执行效率以及最好的聚类质量。而具体分析该实验的聚类输出,发现50个分类中有较为平均的股票数量,对实际应用具有指导意义,可将其作为财务指标的股票板块划分结果。比较模糊K-means算法的聚类结果,当选择平方欧几里得距离作为距离算法参数时有最高的执行效率和最好的聚类质量,同样分析该实验的聚类输出,也有相对平均的聚类划分,但由于模糊K-means算法不是像K-means算法一样的硬性聚类,它会划分成有重叠的簇,因此其聚类输出的聚类划分较K-means算法更不平均。

综上所述,对于该股票财务数据聚类实验,K-means算法应选择谷本距离算法作为距离参数,而模糊K-means算法应选择平方欧几里得距离算法作为距离参数。

根据以上分析,选择将谷本距离作为距离计算参数的K-means算法和将平方欧几里得距离作为距离计算参数的模糊K-means算法对股票财务指标进行聚类,并对其结果进行分析。具体处理过程包括

以下几个步骤。

(1) 使用Mahout中的Clusterdump类读取聚类结果,并将结果输出保存为文本文件。

(2) 根据财务数据找到每一条数据对应的股票代码。

(3) 统计具有相同聚类编号的股票数量,即每个板块内的股票数量。

(4) 考虑到聚类结果的实际指导意义,在K-means算法的聚类结果中筛选出每一类中股票数量大于30支且小于100支的分类(共26类),在模糊K-means算法的聚类结果中筛选出每一类中股票大于10支且小于100支的分类(共27类)。

(5) 计算每一类中能反映上市公司盈利能力、偿债能力、资产管理质量、成长能力和股东获利能力的各财务指标的标准差和平均标准差。具体计算结果见表4和表5。

根据股票财务指标板块划分的原理,同一类型的公司应当具有相似的经营状况,即同一类内上市公司财务指标的标准差应当小于总体的标准差1,且该值越小表示同一类内的股票财务数据相似度越高。根据各财务指标的平均值和标准差的计算结果可知,K-means算法和模糊K-means算法都只有一个分类的平均标准差大于1,其他分类的平均标准差均小于1,表明两种算法的划分结果都较合理。但K-means算法得到的划分结果与模糊K-means算法相比,只有5个类的各项指标平均标准差大于模糊K-means算法;并且K-means算法

表4 模糊K-means 聚类结果财务指标标准差

类别	股票数量/支	盈利能力	成长能力	股东获利能力	偿债能力	资产管理质量	平均标准差
1	20	0.226	0.053	0.053	0.063	0.191	0.117
2	13	0.113	0.111	0.111	0.038	0.235	0.122
3	12	0.168	0.047	0.047	0.179	0.150	0.118
4	17	0.256	0.386	0.386	0.311	0.374	0.343
5	11	0.168	0.104	0.104	0.068	0.262	0.141
6	16	0.218	0.115	0.115	0.206	0.253	0.182
7	16	0.161	0.228	0.228	0.751	0.267	0.327
8	10	0.139	1.026	1.026	0.083	0.257	0.506
9	20	0.186	0.146	0.146	0.480	0.139	0.219
10	14	0.180	0.055	0.055	0.032	0.237	0.112
11	17	0.431	0.123	0.123	0.081	0.250	0.201
12	20	10.133	7.617	7.617	10.451	7.635	8.690
13	33	0.153	0.433	0.433	0.141	0.253	0.283
14	29	0.126	0.095	0.095	0.404	0.295	0.203
15	12	0.173	0.048	0.048	0.388	0.236	0.179
16	34	0.209	0.215	0.215	0.266	0.793	0.340
17	16	0.162	0.053	0.053	0.534	0.132	0.187
18	19	0.210	0.053	0.053	0.147	0.151	0.123
19	22	0.175	0.694	0.694	0.421	0.356	0.468
20	38	0.296	1.631	1.631	0.117	0.244	0.784
21	21	0.455	0.055	0.055	0.046	0.243	0.171
22	12	0.130	0.238	0.238	0.055	0.285	0.189
23	17	0.781	0.058	0.058	0.456	0.379	0.346
24	22	0.188	0.078	0.078	0.044	0.235	0.125
25	11	0.108	0.068	0.068	0.116	0.323	0.137
26	16	0.400	0.275	0.275	0.157	0.359	0.293
27	23	0.237	0.117	0.117	0.320	0.335	0.225

表5 K-means 聚类结果财务指标标准差

类别	股票数量/支	盈利能力	成长能力	股东获利能力	偿债能力	资产管理质量	平均标准差
1	50	0.085	0.077	0.275	0.062	0.117	0.123
2	77	0.172	0.093	0.363	0.060	0.161	0.170
3	89	0.185	0.166	0.444	0.073	0.271	0.228
4	44	0.159	0.045	0.152	0.048	0.120	0.105
5	64	0.071	0.069	0.147	0.057	0.084	0.086
6	79	0.084	0.088	0.151	0.075	0.087	0.097
7	58	0.193	0.084	1.852	0.159	0.486	0.555
8	76	0.278	0.107	0.297	0.117	0.256	0.211
9	62	0.069	0.045	0.135	0.106	0.088	0.088
10	43	0.105	0.057	0.160	0.043	0.088	0.091
11	30	0.047	0.029	0.096	0.039	0.051	0.052

续表

类别	股票数量/支	盈利能力	成长能力	股东获利能力	偿债能力	资产管理质量	平均标准差
12	35	0.053	0.038	0.079	0.039	0.050	0.052
13	31	0.070	0.031	0.097	0.154	0.057	0.082
14	87	0.123	0.068	0.217	0.102	0.098	0.122
15	57	0.058	0.043	0.166	0.055	0.063	0.077
16	48	0.109	0.108	0.244	0.070	0.139	0.134
17	52	0.070	0.047	0.196	0.037	0.095	0.089
18	77	0.112	0.079	0.192	0.078	0.113	0.115
19	41	0.066	0.061	0.097	0.040	0.082	0.069
20	33	0.093	0.045	0.211	0.146	0.101	0.119
21	67	0.221	0.174	0.588	0.091	0.679	0.351
22	54	0.074	0.028	0.105	0.036	0.075	0.064
23	55	0.094	0.048	0.157	0.094	0.091	0.097
24	59	0.043	0.038	0.096	0.039	0.061	0.055
25	35	0.196	0.082	0.480	2.096	2.965	1.164
26	80	0.057	0.058	0.093	0.055	0.065	0.066

得到的有效划分中包含的股票总数量为1 483支,而模糊K-means算法仅有511支。由此可知,K-means算法对于股票财务数据的聚类效果优于模糊K-means算法。因此,本文选择使用平方欧几里得距离作为距离度量方法K-means算法的聚类结果,也作为最终对中国上市公司股票板块划分的结果,见表6。

每一大类中股票的分析结果如下。

#### (1) 第一大类

7类:该类别股票盈利能力较强、股东回报率较高、成长能力一般、投资安全性

较高,属于对风险厌恶的投资者短期投资的对象。

8、13类:该类别股票盈利能力较强、股东回报率一般、成长能力较强、投资安全性高,具有较好的发展前景,属于对风险厌恶的投资者长期投资的对象。

#### (2) 第二大类

3类:该类别股票盈利能力、股东回报率、成长能力均较强,投资安全性一般。对于风险中立的投资者来说,无论是长期还是短期都是极佳的投资标的。

4、9、11类:该类别股票盈利能力和成

表6 股票财务数据K-means算法聚类结果

类别	特征	股票类别
第一类	具有较高的投资价值、较高的投资安全性	(7、8、13)
第二类	具有较高的投资价值、一般的投资安全性	(3、4、9、11、16、21)
第三类	具有较高的投资价值、较低的投资安全性	(1、5、6、12、22、23)
第四类	具有一般的投资价值、较高的投资安全性	(25)
第五类	具有一般的投资价值、一般的投资安全性	(26、18、20、2)
第六类	具有一般的投资价值、较低的投资安全性	(14、15、19)
第七类	具有较低的投资价值、一般的投资安全性	(17)
第八类	具有较低的投资价值、较低的投资安全性	(10、24)

长能力较强、股东回报率一般、投资安全性一般。对于风险中立的投资者来说,由于其具有较好的扩张发展能力,因此是值得进行长期投资的标的。

16类:该类别股票盈利能力一般、成长能力较强、股东回报率较高、投资安全性一般。对于风险中立的投资者来说,亦是值得进行长期投资的标的。

21类:该类别股票盈利能力较强、成长能力一般、股东回报率较高、投资安全性一般。对于风险中立的投资者来说,是适合短期投资的标的。

### (3) 第三大类

1、5类:该类别股票盈利能力一般、成长能力较强、股东回报率较高、投资安全性较低。对于风险偏好的投资者来说,适合进行长期投资。

6类:该类别股票盈利能力、股东回报率、成长能力均较强,投资安全性较低。对于风险偏好的投资者来说,是长期、短期投资皆可的投资对象。

12、23类:该类别股票盈利能力和成长能力较强、股东回报率一般、投资安全性较低。对于风险偏好的投资者来说,是一个长期的投资对象。

22类:该类别股票盈利能力较强、成长能力一般、股东回报率较高、投资安全性较低。对于风险偏好的投资者来说,是短期的投资对象。

### (4) 其他大类

第四、五、六、七、八类投资价值意义不大。尤其第七、八类是明显的绩差股、垃圾股。

## 5 结束语

本文对Mahout中的K-means算法和模糊K-means算法使用不同的距离测量方

式,对中国股票市场上所有上市公司按照能反映公司盈利能力、偿债能力、成长能力、资产管理质量及股东获利能力5个方面共15项财务指标进行了聚类实验,通过对不同距离算法下两种聚类算法的迭代次数、执行时间、聚类间密度和聚类内密度进行比较分析,并结合实际聚类结果,得到适合于股票财务数据聚类的距离测量方法和聚类算法组合,同时以该组合的实验结果为依据,得到股票板块划分,为投资决策提供有益参考。

## 参考文献

- [1] 周鑫. 我国股票市场板块效应实证研究(硕士学位论文). 成都: 西南交通大学, 2012  
Zhou X. Empirical research on plate effect of Chinese stock market (master dissertation). Chengdu: Southwest Jiaotong University, 2012
- [2] 周焯华, 陈文南, 张宗益. 聚类分析在证券投资中的应用. 重庆大学学报(自然科学版), 2002, 25(7): 122~126  
Zhou Z H, Chen W N, Zhang Z Y. Application of cluster analysis in stock investment. Journal of Chongqing University(Natural Science Edition), 2002, 25(7): 122~126
- [3] 劳兰珺, 邵玉敏. 中国股票市场行业收益率序列动态聚类分析. 财经研究, 2004, 30(11): 75~82  
Lao L J, Shao Y M. Dynamic clustering analysis of return series of industrial indexes in Chinese stock market. Journal of Finance and Economics, 2004, 30(11): 75~82
- [4] 李云飞, 李鹏雁. 基于模糊聚类技术的股票投资价值评价指标选择. 燕山大学学报, 2008, 32(6): 551~556  
Li Y F, Li P Y. Evaluation indexes selection of stocks' investment value based on fuzzy clustering. Journal of Yanshan University, 2008, 32(6): 551~556
- [5] 孙磊平. 数据挖掘方法在股票分析中的应用与

- 研究(硕士学位论文). 成都: 西南财经大学, 2013
- Sun L P. The application and research of data mining in stock analysis (master dissertation). Chengdu: Southwestern University of Finance and Economics, 2013
- [6] 邓秀勤. 聚类分析在股票市场板块分析中的应用. 数理统计与管理, 1999, 18(5): 1~4
- Deng X Q. Application of cluster analysis in stock market board analysis. Journal of Applied of Statistics and Management, 1999, 18(5): 1~4
- [7] 杨富勇. 数据挖掘技术在股票投资中的应用

- (硕士学位论文). 长沙: 湖南大学, 2010
- Yang F Y. Application of data mining in stock investment (master dissertation). Changsha: Hunan University, 2010
- [8] 张传琦. 基于蚁群聚类算法的股票板块分类研究(硕士学位论文). 上海: 复旦大学, 2012
- Zhang C Q. Research on Stock plate classification based on ant colony clustering algorithm (master dissertation). Shanghai: Fudan University, 2012
- [9] Owen S, Anile R, Dunning T, *et al.* Mahout in Action. Greenwich: Manning Publications Co, 2012

## 作者简介



海沫, 女, 博士, 中央财经大学信息学院副教授, CCF高级会员, 主要研究领域为分布式系统、大数据处理和

分析。



牛怡晗, 女, 就职于上海浦东发展银行昆明分行, 主要研究领域为大数据分析。



张悦今, 女, 中央财经大学信息学院讲师, 主要研究领域为数据挖掘及其应用、知识管理、互联网金融。

收稿日期: 2015-09-30

基金项目: 北京高等学校青年英才计划资助项目(No.YETP0988), 2014年度中财121人才工程青年博士发展基金资助项目(No.QBJ1427)

**Foundation Items:** Beijing Higher Education Young Elite Teacher Project(No.YETP0988), 121 of CUFU Talent Project Young Doctor Development Fund in 2014(No.QBJ1427)

**论文引用格式:** 海沫, 牛怡晗, 张悦今. 面向大数据的并行聚类算法在股票板块划分中的应用. 大数据, 2015036

Hai M, Niu Y H, Zhang Y J. Application of parallel clustering algorithms for big data in the division of stock. Big Data Research, 2015036