

基于 ID3 决策树改进算法的稿刊推荐研究

贾笛笛, 陈智勇

(桂林电子科技大学 计算机与信息安全学院, 广西 桂林 541004)

摘要:在科研工作者的科研活动中,发表论文是其中非常重要的部分。论文承载着科研工作者的研究成果,只有发表在出版物上,才能得到世人的认可。现有稿刊推荐研究的推荐结果少,而且推荐结果的准确度不高,投稿人很难从推荐结果中发现适合投稿的期刊。鉴于此,提出利用改进的 ID3 决策树算法对期刊与稿件相关度进行分类建模,利用得到的分类规则为投稿人推荐合适期刊的方法。实验结果表明,基于 ID3 决策树改进算法的稿刊推荐方法推荐准确率较高。

关键词:稿刊推荐;数据挖掘;ID3 算法;多值逻辑;K-Means 聚类

DOI:10.11907/rjdk.171695

中图分类号:TP312

文献标识码:A

文章编号:1672-7800(2017)010-0042-05

Research on Manuscript Publication Recommendation Based on Using Improved ID3 Decision Tree Algorithm

JIA Di-di, CHEN Zhi-yong

(School of Computer Science and information security, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract:In the research activities of scientific research workers, to publish papers is one of the most important parts. The papers carry the research results of scientific research workers. Nowadays, the results of the research manuscript-publication recommendation are less and the accuracy is low. It is difficult for the contributors to find the journals suitable for submission from the recommendation results. Aiming at this phenomenon, the ID3 decision tree algorithm is proposed to classify the relevance of journals and manuscripts, and use the obtained classification rules to recommend the related journals to manuscripts. Firstly, the shortcomings of the traditional ID3 decision tree algorithm are summarized, and then the process of improving ID3 decision tree algorithm by multi-value logic is expounded. Finally, the classification rule extracted by the generated decision tree model is used to the manuscript-publication recommendation. The experimental results show that the proposed method based on the improved algorithm of ID3 decision tree is not only recommend the related journals, but also has high accuracy.

Key Words:manuscript-publication recommendation; data mining; ID3 algorithm; multi-value logic; K-means clustering

0 引言

稿刊推荐是向有投稿需求的科研工作者推荐适合其投稿的期刊。国外关于出版物与稿件的研究主要有两个方面:推荐与科研工作者研究领域相符的文献^[1-2];推荐与科研工作者稿件研究领域相符的出版物^[3-8]。Luong Hiep、Klamma R、Luong H 等^[3-4]通过社交网络的方法,探索网络中与稿件作者在同—个研究领域内的其他稿件作者所出版过的刊物;2015 年,Medvet Eric、Bartoli Alberto^[5]建立期刊查询系统,通过稿件标题和摘要进行期刊推荐;

2015 年,Anas Alzoghbi^[6]提出将研究者简介作为多元线性回归问题的模型,利用元数据实现有效的出版物推荐;2016 年,Tran Hung Nghiep、Huynh Tin^[7]使用研究人员引用的出版物构建数据样本,进行出版物推荐。国内关于稿刊推荐的研究方法,主要是建立稿件与期刊研究方向的关系。2014 年,徐镇^[9]提出基于垂直搜索引擎的论文投稿推荐系统研究;2014 年,秦臻等^[10]提出基于合著作者网络的期刊推荐方法;2015 年,王亮等^[11]提出基于主题模型文本挖掘的期刊推荐方法,大大提升了期刊推荐的准确度。

上述出版物推荐方法为有投稿需求的用户提供了方

收稿日期:2017-05-05

基金项目:广西可信软件重点实验室项目(KX201413);广西高校云计算与复杂系统重点实验室项目(14106)

作者简介:贾笛笛(1990—),女,山东菏泽人,桂林电子科技大学计算机与信息安全学院硕士研究生,研究方向为数据挖掘;陈智勇(1971—),男,湖北潜江人,硕士,桂林电子科技大学计算机与信息安全学院副教授、硕士生导师,研究方向为计算机体系结构。

便,但也存在不足之处,总结如下:

(1)现有出版物推荐方法数据样本有限。已有的推荐方法是建立与稿件作者具有合著关系的稿件作者关系网,这样导致推荐结果的数量少,有可能科研工作者真正需要的出版物并不包含在其中。

(2)现有出版物推荐方法没有考虑出版物与稿件的关联度。现有推荐方法没有量化推荐的出版物与稿件的研究方向是否一致,没有考虑出版物与稿件的质量是否相符。如果推荐的出版物质量与稿件质量不符合,稿件很难成功出版。

(3)现有出版物推荐方法不适合没有投稿经验的科研工作者。基于社交网络的出版物推荐方法是建立与稿件作者具有合著关系的作者关系网。没有投稿经历的科研工作者无法得到合著关系网,也就无法得到推荐结果。

1 材料与方法

分析整理 JCR(Journal Citation Reports)中所有的 SCI(Science Citation Index)期刊数据,建立期刊数据库。利用多值逻辑理论改进传统 ID3 算法,并利用改进后的 ID3 决策树算法,对期刊数据样本分类,提取分类规则,然后按照提取的分类规则进行稿刊推荐。

1.1 ID3 算法分析

1966 年, Hunt E B 在概念学习系统 CLS(Care Life of Science)^[12] 中首次应用了决策树方法,是决策树发展的基础;1975 年, J. Ross Quinlan 提出了一个决策树算法,称作 ID3 算法^[13],后来的决策树算法都是在 ID3 算法基础上得到的。ID3 算法利用信息增益^[14]的度量标准选择结点属性,算法学习能力较强,适合处理大规模数据。ID3 算法原理简单易懂,可以生成易于理解分类规则,而且算法的分类速度快^[15]。但 ID3 算法也有自身缺陷:对连续型数据无法处理,对属性值缺失的数据也难以解决;在选择信息增益时,一般选择取值较多的属性。

针对 ID3 算法的不足及本研究需要,现从以下两个方面加以改进:

(1)连续属性离散化。因为 ID3 算法无法处理连续型变量,但期刊与论文相关度的数据样本中,数据属性是连续型数据,如期刊刊登范围与论文研究方向相关度、5 年影响因子和载文量。

(2)确定属性取值。一般决策树算法处理数据集的类别属性是布尔类型的数据,只有是与否的对立情况,但在现实生活中两个对立的取值并不能解决问题。由于原有的 ID3 算法中决策属性的取值个数不一,造成分类后的决策树分类不均,会产生异常点或孤立点。

1.2 多值逻辑对 ID3 算法的改进

多值逻辑是一种非经典的逻辑系统。在经典逻辑中,每个命题取值为真或假。在客观世界中,有些事物无法由二值逻辑描述,于是产生了多值逻辑^[16]。多值逻辑指一切逻辑值的取值数大于 2 的逻辑,它由二值逻辑扩展而

来。

ID3 算法依据多组决策属性的值判断一个实例属于某个类别,若此时属性的取值用集合表示,记成 $V = \{v_1, v_2, v_3, \dots, v_m\}$ 。假设 $m=2$,属性取值为 2,属于二值逻辑,只有一个为真,另一个为假。在稿刊推荐中,若用“1 或 0”两个值表征属性的取值,分类结果将生成只有两个分枝的决策树,造成树的深度过大、分枝过少。若 $m=N$, N 为一个足够大的值,导致生成的决策树分枝过多。由生成的决策树得到大量的分类规则使得 ID3 算法的使用变得复杂。

属性取值个数决定着 ID3 决策树算法生成的决策树模型分枝,因此要控制属性取值的个数。本文利用多值逻辑理论改进 ID3 算法,以限定属性取值的个数。

以期刊与稿件关联度分类问题为例,利用多值逻辑改进 ID3 算法的步骤如下:

(1)计算类别属性的信息熵。若期刊信息样本集合 S 中有 n 个样本。利用多值逻辑理论将期刊与稿件的相关度信息的类别属性分为 0,1 之间的 m 个取值: $v_1, v_2, v_3, \dots, v_m$ 。此时类别属性 C 有 m 个不同的取值即 $C_i (i=1, 2, \dots, m)$ 。设 s_i 为 C_i 中的样本数,可得样本集 S 对应类别属性 C 的信息熵为:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

其中, P_i 表示任意样本属于分类 C_i 的概率。

(2)计算决策属性的信息熵。选择样本集 S 中一个决策属性 A ,若 A 有 k 个不同的取值,决策属性 A 将样本集合 S 划分为 k 类别。由此,属性 A 划分样本集 S 的信息熵为:

$$E(A) = - \sum_{j=1}^k \frac{A_{1j} + A_{2j} + \dots + A_{mj}}{n} I(s_j) \quad (2)$$

其中, $A_{ij} (i=1, 2, \dots, m; j=1, 2, \dots, k)$ 表示子集 S_j 中类 C_i 的样本数量。因为决策属性的取值是连续变量,本文利用多值逻辑理论离散化决策属性的值。首先利用 K-Means 聚类算法将决策属性的值分组,按照多值逻辑的 m 个值: $v_1, v_2, v_3, \dots, v_m$, 对每个分组分别赋值。

(3)计算信息增益。这样可以根据以上信息计算属性 A 的信息增益,公式如下:

$$Gain(A) = I(S) - E(A) \quad (3)$$

最后选择信息增益最大的决策属性作为树的根结点,由此结点产生多个分枝,形成子树。每个子树根结点的选择也如此,以递归的方式最终生成一棵完整的树。

1.3 稿刊推荐数据准备

本文数据来源于 ISI(Institute for Scientific Information)2015 年出版的 JCR。JCR 对收录的全部 SCI 期刊数据进行统计、运算,并对每种期刊定义了影响因子、5 年影响因子、即引指数、载文量、引半衰期等指数。为了方便科研工作者的投稿需要,利用网络获取期刊网址和审稿周期。本文花费大量的时间对期刊引证报告中期刊的研究方向进行划分,由于时间和个人能力有限只完成了计算机方向相关工作。计算机领域的期刊划分是依据中国计算

机学会(CCF, China Computer Federation)中研究方向分类要求进行分类。分类后建立期刊研究领域与论文研究方向之间的关系,作为数据样本的一个数据属性(期刊与论文研究方向关联度属性, J_Category)。

期刊引证报告是世界权威性的综合数据库,涵盖了世界上大部分影响力较大的科技期刊,集合了各学科的重要

研究成果。因此,利用 JCR 建立期刊库,满足用户需求。5 年影响因子和载文量是期刊影响力和发展程度的重要指标^[17]。为了建立期刊与稿件之间的关系,选取期刊引证报告中期刊的 5 年影响因子和载文量作为样本数据的属性。在数据样本中,5 年影响因子的属性名为 J_IF,载文量的属性名为 J_Articles。整理好的期刊数据如表 1 所示。

表 1 期刊数据

ISSN	Journal_Title	category1	category2	J_Category	J_IF	J_Articles
0267-6192	COMPUT SYST SCI ENG	理论计算机科学应用研制		2.692		11
1383-469X	MOBILE NETW APPL	信息系统信息学理论	0			12
1370-4621	NEURAL PROCESS LETT	人工智能与模式识别机器学习	0.588			12
0378-7206	INFORM MANAGE-AMSTER	信息系统信息学理论	2.224			12
1566-2535	INFORM FUSION	体系结构、理论计算机科学计算机系统结构、算法			0	13
1936-2447	CRYPTOGR COMMUN	信息保密	0.476			13
0941-0643	NEURAL COMPUT APPL	人工智能与模式识别机器学习	0.929			13
0038-0644	SOFTWARE PRACT EXPER	软件工程软件开发过程	0.974			13
1386-7857	CLUSTER COMPUT	高性能计算,并行执行模型、分布式计算			0	14
0022-0000	J COMPUT SYST SCI	人工智能与模式识别机器学习、数据库理论	0.515			14

2 实验与结果

本文利用改进后的 ID3 算法进行稿件推荐实验。以《计算机学报》上刊载的论文《朴素贝叶斯分类中的隐私保护方法研究》为例,影响因子范围定为 0.6~2.6,论文关键词:数据挖掘、隐私保护、朴素贝叶斯分类、随机处理、特征重构。

2.1 数据预处理

通过计算期刊的刊登范围标签与稿件关键词标签的相似度,近似计算期刊与稿件之间的研究方向相关度 J_Category 的值。 $\{A_1, A_2, A_3, \dots, A_m\}$ 表示期刊 A 的标签集合, $\{B_1, B_2, B_3, \dots, B_n\}$ 表示稿件 B 的关键词标签集合, $\{a_1, a_2, a_3, \dots, a_m\}$ 表示期刊 A 向量, $\{b_1, b_2, b_3, \dots, b_n\}$ 表示稿件 B 向量, $k = \min(m, n)$ 。由于期刊的标签数量大,期刊标签与稿件关键字匹配完成后,按照大小降序排列,截取前 n 个值定为向量 B,然后计算它们之间的标签相似度。假如稿件 B 的关键词标签为{数据挖掘,隐私保护,朴素贝叶斯分类,随机处理,特征重构},期刊 A 的标签为{数据挖掘,分类,估计,熵,朴素贝叶斯分类,相关性分组,关联规则,聚类,特征重构,决策树}。A 与 B 匹配之后用向量形式表示为: $A(1,1,1,1,1), B(1,0,0,0,1,0,0,0,1,0)$ 。将 B 向量倒序排列,截取和 A 一样的长度,得 $B(1,1,1,0,0)$ 。利用式(4)计算 A 与 B 的相似度结果为 0.77,然后按照此方式获得 J_Category 的其它值。

$$Sim(A, B) = \sum_{k=0}^n \frac{a_k \times b_k}{\sqrt{\sum_{k=0}^n a_k^2} \sqrt{\sum_{k=0}^n b_k^2}} \quad (4)$$

利用 K-Means 算法对研究方向相关度(J_Category)、5 年影响因子(J_IF)、载文量(J_Articles)进行聚类,聚类结果如图 1、图 2、图 3 所示。其中 J_Category 聚为 5 类,每个类别按照多值逻辑的 5 个值即 1、0.75、0.5、0.25、0

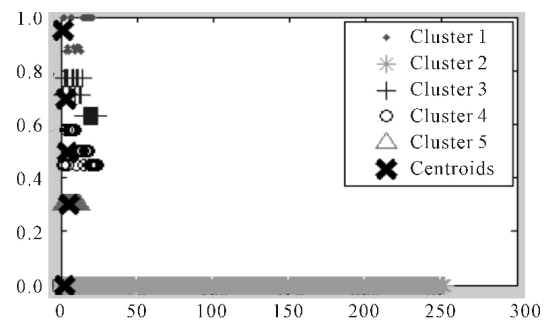


图 1 研究方向聚类结果(J_Category)

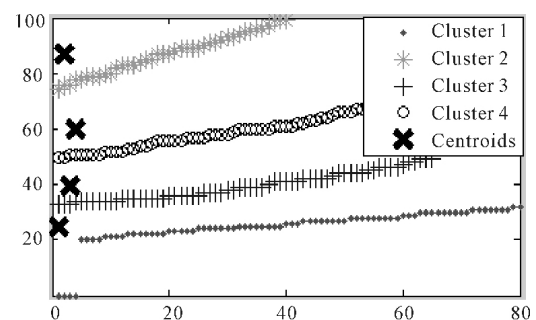


图 2 载文量聚类结果(J_Articles)

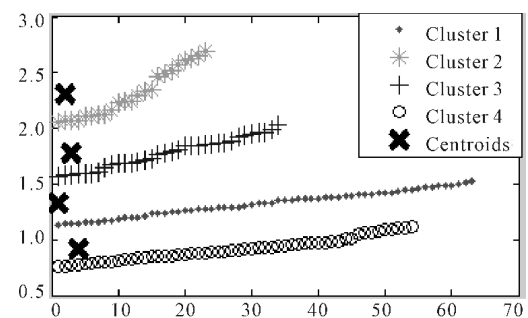


图 3 影响因子聚类结果(J_IF)

依次赋值。J_Articles 属性一部分取值聚为 4 类,每个类别按照多值逻辑的 4 个值 0.75、0.5、0.25、0 依次赋值,J_

Articles 属性剩余的数据赋值为 1。选取 J_IF 在 0.6~2.6 范围内的所有值聚为 4 类,范围外的所有值赋值为 0,聚类后的每个类别按照多值逻辑的值即 1、0.75、0.5、0.25 进行赋值。利用多值逻辑处理后的数据样本如表 2 所示。

表 2 数据样本			
J_Category	J_IF	J_Articles	J_Correlation
0	0.75	0.25	0
0	0	0	0
0	0.25	0.75	0
0.75	0	0.5	0
1	0.75	1	1
0.75	0.75	0.25	0.75
0.25	0.5	0.25	0.25
0.75	0.25	1	0.75
0.5	1	0.5	0.5
0.75	0	0.5	0
0.25	0.25	0.5	0.25
0	0	0.25	0

2.2 数据建模

利用 Java 语言在 Eclipse 平台实现改进后的 ID3 算法,并生成决策树模型,如图 4 所示。本文分类结果的准确度由 Weka^[18] 工具分析,如图 5 所示。由图 4 可得决策模型分类规则,按照分类规则获取数据集中类别属性值大于 0.25 的所有期刊数据,输出结果如图 6 所示。投稿人可以将图 6 的输出结果作为参考选择适合自己的期刊进行投稿。

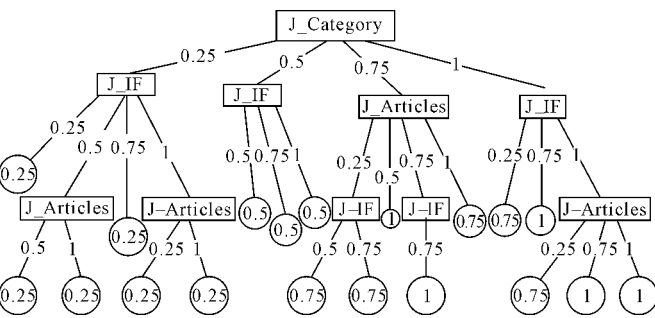


图 4 改进 ID3 算法生成的决策树模型

Correctly Classified Instances	321	96.3964 %
Incorrectly Classified Instances	12	3.6036 %
Kappa statistic	0.806	
Mean absolute error	0.0137	
Root mean squared error	0.0998	
Relative absolute error	16.6389 %	
Root relative squared error	50.339 %	
Total Number of Instances	333	

图 5 由 Weka 平台输出的分类准确度

3 结果分析

本文运用多值逻辑理论改进了 ID3 算法,使原本杂乱无序的属性值变得有规律可循,使得生成的决策树模型层

ISSN	Abbreviated_Journal_Title	AVG_IF	Articles
0168-7433	J AUTOM REASONING	2.332	44
0955-792X	J LOGIC COMPUT	2.111	36
0738-4602	AI MAG	2.102	689
0218-2130	INT J ARTIF INTELL T	1.962	43
0004-3702	ARTIF INTELL	1.929	101
1063-6560	EVOL COMPUT	1.879	38
1076-9757	J ARTIF INTELL RES	1.871	261
0167-9236	DECIS SUPPORT SYST	1.861	41
0954-0091	CONNECT SCI	1.846	27
0266-4720	EXPERT SYST	1.846	222

图 6 期刊推荐结果

数减少,而且分类更加均匀合理。本文从 3 个方面对改进 ID3 算法的分类结果进行分析:

- (1)改进前后 ID3 算法对比。改进前的 ID3 算法对连续型数据的数据样本无法处理。从图 4 可以看出,利用多值逻辑改进 ID3 算法生成的决策树,分布均匀,树的分枝减少,从众多的分枝减少到有限的几个分枝,属性取值也从无序变为有序。
- (2)期刊与稿件相关度。改进 ID3 算法生成的决策树以研究方向相关度(J_Category)作为树的根节点,将该属性作为更重要的属性,符合稿刊推荐首先要求期刊与稿件研究方向一致的条件。
- (3)准确度。如图 5 所示,由 Weka 平台输出结果可知,对 333 个实例进行分析,有 321 个正确的分类结果,5 个错误结果,分类准确率为 96.39%。分类准确率较高,因此改进后的 ID3 算法适合稿刊推荐的分类工作。

4 结语

针对现有期刊推荐方法推荐结果准确率不高的问题,将多值逻辑理论改进的 ID3 决策树算法应用到稿刊推荐研究中,节约了投稿人大量的科研时间。本文利用改进的 ID3 决策树算法处理海量数据,符合投稿人的推荐结果更多,推荐结果准确率也高。但由于时间有限,本文仅开展了计算机领域相关工作。在后续工作中,将在期刊数据库中增加更多核心期刊,扩大数据样本,为投稿人提供更多选择;同时增加每个期刊对应的标签库,提高推荐的准确率。

参考文献:

[1] PERA MARIA SOLEDAD,NG YIU KAI. A personalized recommendation system on scholarly publications[C]. Proceedings of the 2011 ACM International Conference on Information and Knowledge Management, CIKM11,Glasgow,2011:2133-2136.

[2] LE ANH VU,HOANG HAI VO,TRAN HUNG NGHIEP,et al. SciRecSys: a recommendation system for scientific publication by discovering keyword relationships[J]. Lecture Notes in Computer Science,2014, 8733:72-82.

[3] KLAMMA R,CUONG P M,CAO Y. You never walk alone: recommending academic events based on social network analysis [M].

- Complex Sciences. Springer Berlin Heidelberg, 2009: 657-670.
- [4] LUONG HIEP, HUYNH TIN, GAUCH SUSAN, et al. Exploiting social networks for publication venue recommendations[C]. Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, KDIR, Barcelona, 2012: 239-245.
- [5] MEDVET ERIC, BARTOLI ALBERTO, PICCININ GIULIO. Publication venue recommendation based on paper abstract[C]. Proceedings of International Conference on Tools with Artificial Intelligence, ICTAI, Limassol Cyprus, 2014: 1004-1010.
- [6] ANAS ALZOGHBI, VICTOR ANTHONY ARRASCUE AYA-LA, PETER MFISCHER, et al. PubRec: recommending publications based on publicly available meta-data[C]. Proceedings of the LWA 2015 Workshops: KDML, FGWM, IR, and FGDB, Trier, Germany, 2015: 11-18.
- [7] TRAN HUNG NGHIEP, HUYNH TIN, HOANG KIEM. A potential approach to overcome in scientific publication recommendation[C]. Proceedings of 2015 IEEE International Conference on Knowledge and Systems Engineering, Ho Chi Minh City, 2015: 310-313.
- [8] HUYNH TIN, NGUYENTRAC-THUC, TRAN HUNG-NGHIEP. Exploiting social relations to recommend scientific publications[J]. Lecture Notes in Computer Science, 2016, 9795: 182-192.
- [9] 徐镇. 基于垂直搜索引擎的论文投稿推荐系统研究[D]. 兰州: 兰州大学, 2010.
- [10] 秦臻. 学术社会网络建模和学术资源推荐方法研究[D]. 北京: 北京邮电大学, 2015.
- [11] 王亮, 张绍武. 基于主题模型的文本挖掘的研究[D]. 大连: 大连理工大学, 2015.
- [12] PORTER BW, BARESS E R, HOLTE R. Concept learning and heuristic classification in weak theory domains[J]. Artificial Intelligence, 1989, 45: 229-263.
- [13] QUINLAN J R. Induction of decision trees[J]. Machine learning, 1986(1): 81-106.
- [14] 王小巍, 蒋玉明. 决策树 ID3 算法的分析与改进[J]. 计算机工程与设计, 2011, 32(9): 3069-3076.
- [15] 黄爱辉, 陈湘涛. 决策树 ID3 算法的改进[J]. 计算机工程与科学, 2009, 31(6): 109-111.
- [16] 刘任任, 欧阳建权. 多值逻辑函数结构理论研究[M]. 北京: 科学出版社, 2010: 2-20.
- [17] 陶维丽. 科技期刊的综合评价比较研究[D]. 武汉: 华中师范大学, 2012: 1-30.
- [18] 孟晓明, 陈慧萍, 张涛. 基于 WEKA 平台的 Web 事务聚类算法的研究[J]. 计算机工程与设计, 2009, 30(6): 1332-1334.

(责任编辑: 孙 娟)

(上接第 34 页)

表 4 综合算法实验结果比较

	1	2	3	4	5	6	7	...	准确率(%)
SVM 算法	涨	涨	跌	跌	涨	涨	跌	...	67.5
时间序列	涨	跌	跌	涨	涨	跌	涨	...	71.4
综合算法	涨	跌	跌	涨	涨	跌	涨	...	77.6
正确集	涨	涨	跌	涨	跌	跌	跌	...	100

从实验结果可以看出, 单独的 SVM 算法和时间序列算法得到的准确率并不是很高, 普遍只有 70% 左右, 而综合算法的准确率接近 80%。因此, 相比于单一算法, 综合算法的效果更好。因为综合算法结合了时间序列和 SVM 算法的优点, 使 SVM 算法能够考虑到时间序列上的时间特性比如季节性等, 又结合了 SVM 算法的优点, 即其自身的错误率较低, 而且计算开销小, 适合运行大批量数据, 得到的结果也更加直观具体。

4 结语

本文结合 K-近邻算法、支持向量机算法和时间序列算法的优点, 提出了一种综合预测算法, 并将其应用到沪深 300 指数的涨跌预测中, 取得了较好效果。然而, 综合算法虽然相比于单一算法, 准确率有改善, 但是未来提升的空间还有很大, 需要通过发掘探索, 不断组合一些更高效的算法, 以得到更高的准确率。

参考文献:

- [1] 何永沛. ARMA 模型参数估计算法改进及在股票预测中的应用[J]. 重庆工学院学报: 自然科学版, 2009(2): 6-8.
- [2] 于志军, 杨善林. 基于误差校正的 GARCH 股票价格预测模型[J].

中国管理科学, 2013(S1): 28-32.

- [3] 费时龙, 任洪光. 多重马氏链模型在股市预测中的应用[J]. 德州学院学报, 2016(8): 28-35.
- [4] 王领, 胡扬. 基于 C4.5 决策树的股票数据挖掘[J]. 计算机与现代化, 2015(10): 38-51.
- [5] 张鹏. 基于 SVR 的股市预测与择时研究[J]. 重庆文理学院学报, 2016(3): 148-155.
- [6] 孙海波, 王丽敏. 引入趋势因子的 BP 模型在股市预测中应用[J]. 统计与决策, 2015(19): 87-89.
- [7] 李雁. 基于 ARIMA 模型和神经网络模型的股票价格预测[J]. 金融商务, 2014(4): 77-80.
- [8] 刘海翔, 白艳萍. 时间序列模型和神经网络模型在股票预测中的分析[J]. 管理科学, 2011(4): 22-31.
- [9] 金得宝. 基于支持向量机的股市预测研究[D]. 杭州: 浙江大学出版社, 2010: 2-10.
- [10] PETER HARRINGTON. 机器学习实战[M]. 北京: 人民邮电出版社, 2013: 15-20.
- [11] 王波, 程福云. KNN 算法在股票预测中的应用[J]. 武汉: 科技创业月刊, 2015(16): 10-15.
- [12] 张晨希, 张燕平. 基于支持向量机的股票预测[J]. 计算机技术与发展, 2006, 16(6): 34-35.
- [13] PANG NING TAN, MICHAEL STEINBACH. 数据挖掘导论[M]. 北京: 人民邮电出版社, 2011: 157-170.
- [14] BURGERS B C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [15] 张文彤, 董伟. SPSS 统计分析高级教程[M]. 北京: 高等教育出版社, 2013: 393-395.
- [16] 林利敏. 基于支持向量机的股价短期预测研究[D]. 杭州: 杭州电子科技大学, 2010: 45-50.

(责任编辑: 黄 健)