

基于字形编码与拼音编码的近似商标辨识算法研究

祁俊辉¹, 龙 华¹, 赖 华¹, 毕丹宏²

(1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650000;

2. 江苏省邮电规划设计院有限责任公司, 江苏 南京 210000)

摘 要: 为了实现高效、快速的商标自动评审工作, 提出将商标文字通过字形编码、拼音编码等方式映射为数字型字符串和字母型字符串的辨识算法。算法依照字形和拼音描述商标信息, 配合改进后的 Jaro-Winkler distance 算法计算出商标文字的字形相似度和拼音相似度, 并根据相似匹配算法得出商标文字的相似程度。实验结果表明, 该方法所得结果有更好的辨识度, 符合常规人体视觉和逻辑判断。

关键词: 商标辨识算法; 近似商标; 形近字; 近音字; 相似度

DOI: 10.11907/rjdk.172985

中图分类号: TP312

文献标识码: A

文章编号: 1672-7800(2018)006-0077-04

Research on Approximate Trademark Identification Algorithm Based on Glyph and Pinyin Coding

QI Jun-hui¹, LONG Hua¹, LAI Hua¹, BI Dan-hong²

(1. Faculty of Information Engineering and Automation,

Kunming University of Science and Technology, Kunming 650000;

2. Jiangsu Posts and Telecommunications Planning and Designing Institute CO., LTD, Nanjing 210000, China)

Abstract: In order to realize efficient and rapid automatic evaluation of trademark, an identification algorithm is proposed to map the trademark text into numeric and alphabetic string using ideographic and Pinyin encoding. The algorithm is used to describe the trademark according to the font and Pinyin, and with the improved Jaro-Winkler distance algorithm the similarity of trademarks font and Pinyin coordinated is calculated. The similarity of words in trademark is calculated by similarity-matching algorithm. The experimental results show that the method conforms to judgment of physical vision and logic and has high recognition rate.

Key Words: trademark identification algorithm; approximate trademark; homophonic characters; homomorphous characters; similarity

0 引言

商标作为商品生产者、经营者的重要特征标识, 是商品生产者、经营者的重要知识产权资源。随着全球经济迅速发展和经济一体化快速推进, 知识产权保护已经成为重要的商业活动。此外, 新型产品和产业迅速兴起, 为人们带来了更加丰富的产品, 商标产生速度也大大加快。2016年, 我国商标申请数量达 396.1 万件以上, 已连续 15 年位居世界第一, 且呈现逐年递增的趋势。在这样的形势和数据规模下, 依靠人工处理商标的评审工作已经非常困难, 急

需高效的自动评审方法实现快速、准确的商标评审工作。

中文商标作为国际化商标重要组成部分, 不法分子会针对某些火爆的中文商标申请仿冒伪劣商标, 进而制造出相似产品以迷惑消费者。事实上, 我国《商标法》对近似商标的定义有明确规定, 这也是人工处理商标评审工作应遵循的原则, 但是不免会因为工作人员疏忽与个人因素的影响, 造成评审工作不平衡。

近似商标具体指, 两个商标相比较, 商标文字的字形、读音、含义等相似。经研究发现, 文字字形相似的近似商标带给消费者的影响最大, 因为消费者在面对大多数产品时首先是通过视觉观察该产品, 由于中文汉字存在很多形

收稿日期: 2017-11-16

作者简介: 祁俊辉(1994—), 男, 昆明理工大学信息工程与自动化学院硕士研究生, 研究方向为中文信息处理; 龙华(1963—), 女, 博士, 昆明理工大学信息工程与自动化学院教授、硕士生导师, 研究方向为网络通信。本文通讯作者: 龙华。

近字,再加上快速阅读习惯,难免会在一定程度上造成视觉的错误判断;其次,文字读音相似的近似商标也会给消费者带来一定影响,因为消费者在面对大多数产品时会在潜意识中发出该产品的商标文本读音。

事实上,对近似商标的辨识,就是对短字符串相似度的检测,但此处的相似度主要指字形、读音相似,那么传统的字符串相似度算法就显得捉襟见肘。相关研究工作中,文献[1]提出把汉字描述成部件为操作数、部件间位置关系为运算符的数学表达式形式;文献[2]也定义了一种语言采用笔画、部件分层次的方法对汉字进行描述;以后的研究工作中,基本都围绕这两种方法对汉字字形相似进行研究,文献[3]提出一种基于结构描述的汉字字形相似度计算方法,但由于汉字字形结构复杂,很多汉字无法用数学表达式进行精准描述,故还需进一步改进;文献[4]提出三元组递归表示的汉字字形相似度计算方法,但该方法需要对原子部件进行人工分类及相似度设置;文献[5]对字符串的相似度作了介绍,但对短字符串没有深入研究;文献[6]介绍了一种商标相似方法,但没有编程实现;文献[7]提供了一种对同音字和近音字的检索算法;文献[8]采用基于汉字、拼音和拼音改良的编辑距离对关键词进行匹配;文献[9]基于拼音输入法对中文字符串近似串匹配查询问题研究与改进;文献[10—12]也对短字符串的提取、识别做了介绍;文献[13]对手写字的识别做了相应研究,但只使用了笔画顺序技术;文献[14—15]则针对独体字的识别做了相应研究,利用相似的独体字也是商标识别中最常见的抄袭行为。

以上参考文献大多采用汉字结构(如上下结构、左右结构等)和笔画对汉字部件进行描述,进而通过编辑距离(edit distance)等算法计算其字形相似度。虽然这种方法在理论研究中有较高的价值,但由于汉字结构类型的复杂性,目前没有任何一个较为完整的汉字结构库可供参考,所以此方法在实际运用过程中可能比较困难;其次,将汉字描述为数学表达式后,用何种算法对其计算相似度影响重大。本文以实际运用为主,不考虑过多复杂的概念,根据相关编码规则对汉字进行字形编码和拼音编码,通过 Jaro-Winkler distance 分别计算两者的相似度,再根据相似匹配算法确定其相似程度。

1 商标文字形式化描述

1.1 文字字形编码

在汉字计算机编码标准中,编码方式为 Unicode 的中日韩统一表意文字基本字符集收录的汉字,共 20 902 个。因为每个中文汉字都可根据书写笔画顺序分为横、竖、撇、捺、折,故可按照表 1 所述编码规则对任意汉字生成其汉字字形的数字型字符串。例如,汉字“明”可根据所述编码规则,生成的数字型字符串为“25113511”;汉字“天”可根据所述编码规则,生成的数字型字符串为“1134”。将编码方式为 Unicode 的基本字符集中的 20 902 个汉字依照此

编码规则生成其数字型字符串并存入数据库,组建 Unicode 汉字笔画顺序数据库。

表 1 汉字字形编码规则

笔画	笔画编码
横(一)	1
竖(丨)	2
撇(丿)	3
捺(㇏)	4
折(乚、㇄、㇚)	5

对于商标文本来说,其往往由词语,即多个汉字所构成。遍历该商标文本中所有汉字,生成其汉字字形的数字型字符串,再按照商标文本中汉字出现顺序进行数字型字符串的合并,最终生成其文字字形的数字型字符串。

例如“脉动”一词,经查找数据库,“脉”字的数字型字符串为“351145534”,“动”字的数字型字符串为“115453”,则词语“脉动”的数字型字符串为“351145534/115453/”。

1.2 文字拼音编码

在处理中文汉字拼音时,由于汉字存在大量的多音字、同音字,当对两个汉字进行相似度匹配时,应考虑读音的相同或相近。但对于汉字的读音来讲,由于地方方言的影响,使人们不能发出标准的汉字读音,如将声母 l 和 n、c 和 ch、s 和 sh 等混淆,不易处理,故本文直接采用汉字拼音的相似度替代其读音的相似度。

对于多音字来讲,因为商标文本的内容大多是较为常用的字或音,所以取该汉字的常用拼音作为其拼音编码;而对于同音字来讲,如果仅仅考虑其拼音的相似程度,那么可以直接根据字符串相似算法计算其拼音的相似度。

在编码方式为 Unicode 的基本字符集中的 20 902 个汉字,其中有 31 个汉字无拼音,统一以“*”填充,对其它 20 871 个汉字按照常用拼音生成其字母型字符串并存入数据库,组建 Unicode 汉字拼音数据库。

与生成商标文字字形的数字型字符串生成规则相同,也按此规则生成商标文字拼音的字母型字符串。例如“脉动”一词,经查找数据库,“脉”字的字母型字符串为“mai”,“动”字的字母型字符串为“dong”,则词语“脉动”的字母型字符串为“mai/dong/”。

2 商标文字相似度算法

2.1 文字字形、拼音相似度算法

为量化文字之间的相似度,引入改进后的 Jaro-Winkler distance 算法分别对描述其字形的数字型字符串和描述拼音的字母型字符串进行相似度计算。

算法 1 商标文字字形/拼音相似度算法

输入:商标文本 X、商标文本 Y。

输出:商标文本 X、Y 之间的字形/拼音相似度 Sim。

Step1:从 Unicode 汉字笔画顺序数据库/Unicode 汉字拼音数据库中读取商标文本 X、Y 的数字型/字母型字符串 str_x 和 str_y ,并获取 str_x 和 str_y 的长度 len_x 和 len_y ,生

成检测矩阵 $I(X, Y)_{len_x \times len_y}$ 。

Step2: 根据公式(1)计算匹配窗口值 MW 。

$$MW = \frac{\max(|len_x|, |len_y|)}{2} - 1 \quad (1)$$

Step3: 检测矩阵 $I(X, Y)_{len_x \times len_y}$ 及匹配窗口值 MW , 根据相关规则, 计算匹配字符数 m 和匹配字符换位数 n , 并根据公式(2)计算数字型/字母型字符串 str_x 和 str_y 之间的 Jaro distance。

$$Dis_j = \begin{cases} 0 & m = 0 \\ \frac{1}{3} \left(\frac{m}{len_x} + \frac{m}{len_y} + \frac{m-n}{m} \right) & m > 0 \end{cases} \quad (2)$$

Step4: 获取数字型/字母型字符串 str_x 和 str_y 的最长公共子串 str_{xy} , 并得到其长度 len_{xy} 。定义 b_t 为是否需要进一步计算阈值, p 为缩放因子, 根据公式(3)进一步计算数字型/字母型字符串 str_x 和 str_y 之间的 Jaro-Winkler distance。

$$Dis_{jw} = \begin{cases} Dis_j & Dis_j < b_t \\ Dis_j + \left[len_{xy} \times \min\left(\frac{1}{\max(len_x, len_y)}, p\right) \times (1 - Dis_j) \right] & Dis_j \geq b_t \end{cases} \quad (3)$$

Step5: 如公式(4)所示, 商标文本 X, Y 之间的字形/拼音相似度 Sim 即为数字型/字母型字符串 str_x 和 str_y 的 Jaro-Winkler distance。

$$Sim = Dis_{jw} \quad (4)$$

对算法 1 作进一步补充说明:

(1) 数字型/字母型字符串 str_x 和 str_y 的长度 len_x 和 len_y 应满足 $len_x, len_y \in N_+$ 。

(2) 匹配字符数 m 和匹配字符换位数 n 的计算: 若数字型/字母型字符串 str_x 和 str_y 中相同字符相差距离小于匹配窗口值 MW , 则视为该字符匹配。但应注意, 在匹配过程中, 需排除被匹配过的字符, 若找到匹配字符, 则需跳出此次匹配, 进行下一字符的匹配, 所有匹配字符的个数即为 m 。然后根据数字型/字母型字符串 str_x 和 str_y 中对于匹配字符集的顺序是否一致计算匹配字符换位数 n , 若一致, 则 $n=0$; 若不一致, 则 n 为换位数目的一半。另外, 匹配字符数 m 和匹配字符换位数 n 应满足公式(5)的要求。

$$\begin{cases} 0 \leq m \leq \min(len_x, len_y) \\ 0 \leq n \leq \frac{m}{2} \end{cases} \quad (5)$$

(3) 进一步计算阈值 b_t , 本文取值为 0.85, 可根据实际检测结果作小幅度调整, 主要是为了提高检测准确性; 所述缩放因子 p , 通常取值为 0.1, 可根据实际检测结果做小幅度调整, 主要是为了避免最终计算结果大于 1 的情况发生, 但本方法新增编码字符串 str_x 和 str_y 中最长距离的倒数 $\frac{1}{\max(len_x, len_y)}$, 改

进此处的计算公式 $\min\left(\frac{1}{\max(len_x, len_y)}, p\right)$, 所以缩放因子

p 的取值对最终计算结果影响并不大。

2.2 相似匹配算法

由算法 1 计算所得的商标文本 X, Y 之间字形/拼音相似度 Sim 以一个 $[0, 1]$ 之间的数值反映了商标文本 X, Y 之间字形/拼音相似程度, 数值越大说明相似程度越高。设商标文本 X, Y 经算法 1 计算后, 字形相似度为 Sim_{zx} , 拼音相似度为 Sim_{py} , 两个相似度分别从不同方面反映了商标文本 X, Y 的相似程度, 但两者并不是相互独立的, 如对于由形近字引起的相似重复, 则 Sim_{zx} 较大而 Sim_{py} 较小; 对于由同音字或近音字引起的相似重复, 则 Sim_{zx} 较小而 Sim_{py} 较大; 如果文本之间差异很小, 那么 Sim_{zx} 和 Sim_{py} 都比较大; 相反, 如果文本没有重复, Sim_{zx} 和 Sim_{py} 才会都较小。基于此, 本文认为商标文本 X, Y 之间, 若 Sim_{zx} 和 Sim_{py} 都较大, 则反映文本之间具有最高的相似程度; 若 Sim_{zx} 和 Sim_{py} 之间有一个较大, 另一个较小, 则反映文本之间具有较高的相似程度; 若 Sim_{zx} 和 Sim_{py} 都较小, 则反映文本之间具有较低的相似程度。

单独使用字形相似或拼音相似去衡量商标文本是否近似不够严谨, 因为任何汉字都是由五笔构成, 所以即使是两个完全不同的字, 经算法 1 计算后可能也存在一定的相似度; 对于拼音相似来说, 因为经拼音编码后的字母字符串长度相对较短 (一般不超过 5 位), 但拼音编码由 26 个字母组成, 所以对于有重复字母发音的其相似度会很高, 没有重复字母发音的其相似度则会很低。故本文针对字形相似度和拼音相似度, 设置字形相似阈值和拼音相似阈值, 进而判断其相似程度。

算法 2 相似匹配算法

输入: 商标文本 X, Y 的字形相似度 Sim_{zx} 和拼音相似度 Sim_{py} 。

输出: 商标文本 X, Y 之间的相似程度 (非常相似、字形相似、拼音相似、不相似)。

Step1: 设字形相似度 Sim_{zx} 的相似阈值为 θ_{zx} , 拼音相似度 Sim_{py} 的相似阈值为 θ_{py} 。

Step2: 如果满足公式(6)的要求, 则商标文本 X, Y 之间的相似程度为非常相似。

$$\begin{cases} Sim_{zx} \geq \theta_{zx} \\ Sim_{py} \geq \theta_{py} \end{cases} \quad (6)$$

Step3: 如果满足公式(7)的要求, 则商标文本 X, Y 之间的相似程度为字形相似。

$$\begin{cases} Sim_{zx} \geq \theta_{zx} \\ Sim_{py} < \theta_{py} \end{cases} \quad (7)$$

Step4: 如果满足公式(8)的要求, 则商标文本 X, Y 之间的相似程度为拼音相似。

$$\begin{cases} Sim_{zx} < \theta_{zx} \\ Sim_{py} \geq \theta_{py} \end{cases} \quad (8)$$

Step5: 如果满足公式(9)的要求, 则商标文本 X, Y 之间的相似程度为不相似。

$$\begin{cases} Sim_{zx} < \theta_{zx} \\ Sim_{py} < \theta_{py} \end{cases} \quad (9)$$

3 实验与结果

为了验证基于字形编码和拼音编码的近似商标辨识算法,本文设计一个实验。实验的主要目的在于:比较近似商标辨识算法与人工评审结果之间的差异,考察该方法是否能够真实、准确地反映商标文本的相似程度。

3.1 实验设计和评测方法

实验使用 10 对已知相似程度的样本商标文本,即样本商标文本已有人工评审结果,根据算法 1 计算出样本商标文字基于字形编码、拼音编码的相似度,再利用算法 2 对样本商标文字进行相似匹配,然后将其计算结果与人工评审结果进行比较。通过比较本文提出的算法和人工评审结果对同一样本的相似程度,考察本方法对商标文本近似检测反映的真实程度。

3.2 实验结果和分析

实验所用的 10 对样本商标文本分属不同的相似程度,如表 2 所示。

表 2 样本商标文本及人工评审结果

样本编号	相似程度	样本	样本编号	相似程度	样本
1	完全相似	脉动/脉动	6	拼音相似	脉动/迈动
2	比较相似	脉动/脉动	7	拼音相似	脉动/震动
3	比较相似	脉动/脉动	8	拼音相似	脉动/玫动
4	字形相似	脉动/脉劲	9	不相似	脉动/康师傅
5	字形相似	脉动/咏动	10	不相似	脉动/哇哈哈

根据算法 1 所述规则和步骤,计算出样本商标文字基于字形编码、拼音编码的相似度,定义字形相似阈值为 0.95,拼音相似阈值为 0.90,再利用算法 2 所述步骤对样本商标文字进行相似匹配,所得结果如表 3 所示。

表 3 样本商标文本的相似度及相似程度

样本编号	字形相似度	拼音相似度	相似程度
1	1	1	非常相似
2	0.983 6	1	非常相似
3	0.965 2	0.955 6	非常相似
4	0.974 3	0.805 6	字形相似
5	0.962	0.644 4	字形相似
6	0.916 3	1	拼音相似
7	0.775 7	1	拼音相似
8	0.756 5	0.977 8	拼音相似
9	0.628 1	0.611 1	不相似
10	0.660 8	0.555 6	不相似

为了更直观地分析表 3 所示数据,将计算所得的字形相似度和拼音相似度以折线图的方式画出,如图 1 所示。

从表 3 和图 1 看出,对于样本编号为 1、2、3 的样本商标文本,其字形相似度和拼音相似度都较大,即高于字形相似阈值和拼音相似阈值,故相似程度为非常相似;样本编号为 4、5 的样本商标文本,其字形相似度高于字形相似阈值,但拼音相似度低于拼音相似阈值,故相似程度为字形相似;样本编号为 6、7、8 的样本商标文本,其字形相似度低于字形相似阈值,但拼音相似度高于拼音相似阈值,故相似程度为拼音相似;样本编号为 9、10 的样本商标文

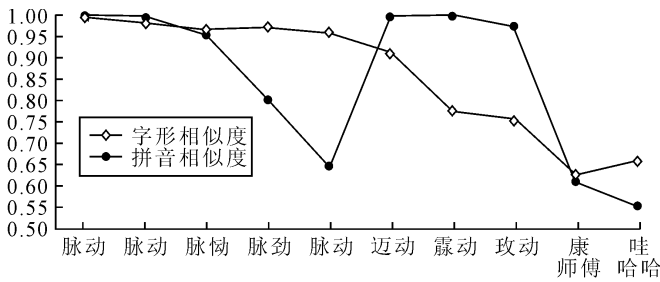


图 1 样本商标文本的相似度折线

本,其字形相似度和拼音相似度都低于字形相似阈值和拼音相似阈值,故相似程度为不相似。

实验结果表明,使用本文算法所得结果与人工评审结果一致,实验结果符合常规人体视觉和逻辑判断,能够真实、准确地反映商标文本的相似程度。

4 结语

实现高效、快速的商标自动评审工作,是中文信息处理领域的一项重要研究。事实上,可以将近似商标等价为短字符串在字形、读音上的相似。本文先将编码方式为 Unicode 基本字符集中的 20 902 个汉字通过字形编码和拼音编码映射为数字型字符串和字母型字符串,建立 Unicode 汉字笔画顺序数据库和 Unicode 汉字拼音数据库;将商标文字通过数据库调用出相应的编码字符串,无需进行字形样本和拼音样本实例的学习,只依照字形和拼音描述信息,配合使用改进后的 Jaro-Winkler distance 算法就可以计算出商标文字的字形相似度和拼音相似度;然后根据相似匹配算法得出商标文字的相似程度;最后通过实验证明该算法的有效性,不仅明显降低了计算的复杂性,还能够真实、准确地反映商标文本的相似程度。

参考文献:

[1] 孙星明,殷建平,陈火旺,吴泉源,景新海.汉字的数学表达式研究[J].计算机研究与发展,2002(6):707-711.

[2] 林民.基于结构描述的汉字字形相似度计算[C].第三届学生计算语言学研讨会论文集,2006:5.

[3] 林民.汉字字形形式化描述方法及应用研究[D].北京:北京工业大学,2009.

[4] 王东,熊世桓.一种新颖的汉字字形相似度计算方法[J/OL].计算机应用研究,2013,30(8):2395-2397.

[5] 刁兴春,谭明超,曹建军.一种融合多种编辑距离的字符串相似度计算方法[J].计算机应用研究,2010,27(12):4523-4525.

[6] 孔军民,谢军.一种商标文本相似性评审方法[P].北京:CN106095865A,2016-11-09.

[7] 阎红灿,张淑芬,谷建涛,阎少宏.基于音码相似度的拼音模糊查询算法[J].计算机与现代化,2008(8):18-20.

[8] 曹肇,郭晓钧,夏云庆,郑方.基于拼音索引的中文模糊匹配算法[J].清华大学学报:自然科学版,2009,49(S1):1328-1332.

[9] 刘兵.基于拼音输入法的中文字符串近似匹配技术研究[D].沈阳:东北大学,2010.

[10] 王耀华,李舟军,何跃鹰,巢文涵,周建设.基于文本语义离散度的自动作文评分关键技术研究[J].中文信息学报,2016,30(6):173-181.

(下转第 84 页)

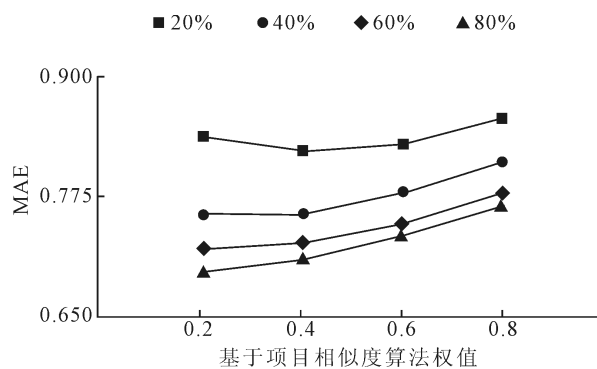


图 8 混合算法在不同权值下的 MAE 对比

表 1 不同稀疏度下的推荐算法 MAE 精度对比

	源数据中训练数据占比			
	20%	40%	60%	80%
item	0.8969	0.8578	0.8206	0.8103
als	0.8685	0.7737	0.7338	0.7044
hybrid	0.8223	0.7552	0.7201	0.6967

据稀疏程度下都可较好地提高预测评分的准确性,并且对混合算法权值进行动态调整可以有效解决数据的冷启动问题。

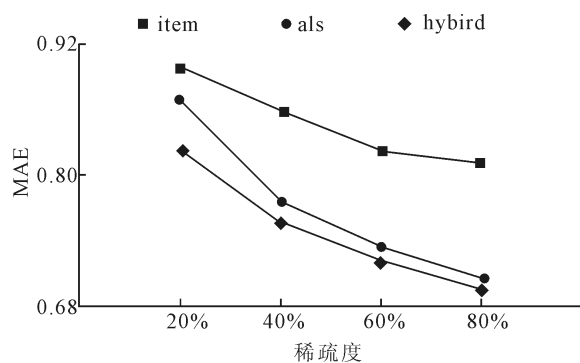


图 9 3 种推荐算法在不同稀疏度下推荐精度对比

4 结语

本文首先对 Spark 平台和建立在 Spark 平台上的推荐框架进行了介绍,指出其优点和使用的广泛性;然后对当前流行的基于项目相似度的协同过滤算法和交替最小二乘算法进行分析,提出使用修正的余弦相似度算法计算项目相似度,并提出利用权值将两种算法进行融合的计算方法;最后,在 Spark 集群环境下,利用 MovieLens 数据集对各算法进行实验,通过对比分析证实了该算法的优越性。

参考文献:

[1] EPPLER M J, MENGIS J. The concept of information overload: a

review of literature from organization science, accounting, marketing, MIS, and related disciplines[J]. IEEE Engineering Management Review, 2010,38(1):3.

[2] BOBADILLA J, ORTEGA F, HERNANDO A, et al. Recommender systems survey[J]. Knowledge-Based Systems, 2013,46(1): 109-132.

[3] YANG X, GUO Y, LIU Y, et al. A survey of collaborative filtering based social recommender systems[J]. Computer Communications, 2014,41(5):1-10.

[4] 文俊浩,舒珊.一种改进相似性度量的协同过滤推荐算法[J]. 计算机科学,2014,41(5):68-71.

[5] ZHAO Z D, SHANG M. User-based collaborative-filtering recommendation algorithms on Hadoop[C]. Third International Conference on Knowledge Discovery and Data Mining, IEEE Computer Society, 2010:478-481.

[6] HU W, YANG F, FENG Z. Item-based collaborative filtering recommendation algorithm based on MapReduce[M]. Multimedia, Communication and Computing Application,2015.

[7] 郭艳红,邓贵仕.协同过滤系统项目冷启动的混合推荐算法[J]. 计算机工程,2008,34(23):11-13.

[8] 郁雪,李敏强.一种有效缓解数据稀疏性的混合协同过滤算法[J]. 计算机应用,2009,29(6):1590-1593.

[9] GEUENS S. Factorization machines for hybrid recommendation systems based on behavioral, product, and customer data[C]. ACM Conference on Recommender Systems,ACM, 2015:379-382.

[10] 张新猛,蒋盛益,李霞,等.基于网络和标签的混合推荐算法[J]. 计算机工程与应用,2015,51(1):119-124.

[11] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.

[12] ZAHARIA M, CHOWDHURY M, DAS T, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing[C]. Usenix Conference on Networked Systems Design and Implementation. USENIX Association, 2012:2.

[13] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[M]. ACM, 2008.

[14] WANG P. A personalized collaborative recommendation approach based on clustering of customers[J]. Physics Procedia, 2012,24(24):812-816.

[15] BATES D, KLIEGL R, VASISHTH S, et al. Parsimonious mixed models[J]. Statistics, 2015.

[16] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]. ACM Conference on Recommender Systems, ACM, 2010: 135-142.

(责任编辑:黄健)

(上接第 80 页)

[11] 索红光,刘玉树,曹淑英.一种基于词汇链的关键词抽取方法[J]. 中文信息学报,2006(6):25-30.

[12] 龙翀,庄丽,朱小燕,黄开竹,孙俊,堀田悦伸,直井聪.手写中文地址识别后处理方法的研究[J]. 中文信息学报,2006(6):69-74.

[13] 郭军,蔺志青,张洪刚.一个新的脱机手写汉字数据库模型及其应用[J]. 电子学报,2000(5):115-116.

[14] 沈模卫,朱祖祥.独体汉字的字形相似性研究[J]. 心理科学,1997(5):401-405+478-479.

[15] 朱祖祥.独体汉字字形的多维相似性研究[C]. 北京:全国第七届心理学学术会议,1993.

(责任编辑:江艳)