

基于大数据平台的企业画像研究综述

田 娟 朱定局 杨文翰
(华南师范大学计算机学院 广州 510631)

摘 要 随着国民经济的发展,企业的数量不断增加。基于企业运营时产生的海量数据,可以利用大数据技术结合企业画像的理论来对企业进行全面分析,为企业成长、行业发展、政府监管等各方面提供可靠的数据分析。文中首先对当前国内外企业画像构建及其技术进行总结和分析,构建了基于大数据平台的企业画像标签体系模型和建模框架;然后根据企业数据的特点,结合比较热门的用户画像技术,提出了几种处理企业的数据方法;最后提出使用大数据技术处理企业数据时值得讨论的几个问题。

关键词 企业画像,大数据技术,特征分析,标签提取

中图法分类号 TP301 **文献标识码** A

Research on Enterprise Portraits Based on Big Data Platforms

TIAN Juan ZHU Ding-ju YANG Wen-han

(College of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract With the development of economy, the number of enterprises is increasing. For the massive data generated by the enterprised, we can use big data technology processing methods and the theory of enterprise portrait to analyze business data, and provide reliable data analysis for enterprise development, industry development and government regulation. Firstly, this paper summarized and analyzed the construction and technology of corporate portraits at domestic and international. Then, according to the characteristics of enterprise data and combined with the persona technology, this paper put forward several methods to deal with enterprise data. At the last, this paper put forward several issues about handling enterprise data when using big data technology processing methods.

Keywords Enterprise portrait, Large data technology, Feature analysis, Label extraction

1 引言

随着国民经济的稳定发展,中国正处于市场经济改革的关键时期,尤其是随着经济全球化的发展,国家“引进来,走出去”以及“一带一路”政策的发展对于中国的企业来说是机遇也是挑战。2016 年末,国家统计局对全国中小型企业统计调查的数据显示,全国中小规模以上工业、企业达到了 37 万户,比 2015 年同期增加了 0.5 万户。其中,大型企业占总数的 2.5%;中型企业 5.4 万户,占中小企业户数的 14.6%;小型企业 31.6 万户,占中小企业户数的 85.4%。企业间关联合作越来越密切,但是企业贸易合作中商业风险无处不在,企业作为经济实体,在寻求合作投资时需要了解合作方的基本信息作为投资参考,企业自身更需要认清发展的利弊,提升企业品牌形象。随着企业数量以及交易信息数据的增长,其产生的数据量也越来越多,并且每年呈指数级增长。随着大数据信息时代的到来以及大数据处理技术的日益发展,我们可以从海量冗杂的企业活动数据中挖掘有价值的数据,帮助银行、券商、会计师、律师、投资方等挖掘和计算企业全方位的信息,包括发现和挖掘企业之间的关联关系,找寻未知关系以促进

企业合作,识别企业资本行为,构造企业风险评估模型等。本文以当下热门的“用户画像”理论为指导依据,提出了基于大数据技术的企业画像的构建模型以及构建方法。

2 企业画像概述

随着网络用户数量的快速增长,大数据技术的不断发展,我们可以对用户在不同网络环境中留下的大量个人痕迹等碎片化数据进行充分的挖掘,发现其中潜在的价值,由此产生了在大数据时代将“用户”“人”数据化的技术。Alan Cooper 最早提出了 persona 的概念。Persona 是真实用户的虚拟代表,是建立在一系列真实数据之上的目标用户模型。用户画像可以完善产品的运营,提升用户的体验。与以往的通过用户调研、市场调研的方式了解用户和市场需求的方式相比,用户画像模式的提出可以更加准确、快速地提升用户的体验,提高企业对外服务的效率,并且也可以根据产品的特点找到目标用户,在用户偏好的渠道上与其进行沟通,实现企业的精准营销。文献[1]阐述了在大数据背景下用户画像对企业实现精准营销提供了有力的支持;文献[2]根据用户画像模型对银行客户信息进行挖掘,分析得出用户的产品偏好、理财

本文受国家社会科学基金重大项目(14ZDB101),国家自然自然基金(61105133)项目,广东省科技计划项目软科学研究项目(2014A070703045)资助。

田 娟(1994—),女,硕士生,主要研究方向为大数据、人工智能,E-mail:1280301862@qq.com;朱定局(1978—),男,博士,副教授,主要研究方向为大数据、人工智能,E-mail:zhudingju@m.senu.edu.cn;杨文翰(1990—),男,硕士生,主要研究方向为大数据,E-mail:850168675@qq.com。

偏好、风险偏好、消费层次以及财富等级等重要的营销信息，然后对不同等级的客户进行理财产品的推荐。文献[3]基于 4C 理论构建用户画像数据库，并且以三枪集团营销数据为数据依托，对用户身高这一因素进行分析，建立了消费者画像模型，为企业精准营销提供依据。文献[4]从 UCG 网站用户的自然属性、社会属性、兴趣属性等方面对用户进行了较为全面的用户模型构建，根据不同的场景分析用户行为。

用户画像是根据用户的社会属性、生活习惯和消费行为等信息而抽象出的一个标签化的用户模型，用户画像的提出可以让企业更有利地赢得用户^[5]。在用户画像的引导下，我们可以进行双向画像，不仅企业可以给用户画像，用户基于企业的数据也可以给企业进行画像。因此**对于企业画像来说，就是把企业信息标签化，在一系列真实数据的基础上为企业建立标签模型体系，将企业的具体行为属性进行归类，最终形成一个多元化的企业标签对象。**今品指数^[6]以 60% 公共媒体数据+20% 自媒体数据+10% 企业官网数据+10% 企业管理层数据为依托建立企业画像，为国家品牌建设发展和有关政府制定宏观调控政策提供重要的依据。大连国税^[7]利用大数据技术对税务系统、纳税人第三方的数据进行分析，为出口企业画像汇集了企业经营、诚信、风险和贡献等多个成要素，并且以图文的形式来可视化呈现。文献[8]利用企业画像概念设计并实现了企业公示数据查询系统，实现了海量企业数据的多维度查询。南京商业区引入了全球最大的“SAP 供应链交易平台”，旨在通过为国内企业画像，将国内企业引入国际“圈子”，使得国内企业能够快速融入全球供应链中^[9]。美国邓白氏公司是国际上最著名的企业征信的管理公司，它的数据整合流程采用多元化的数据收集方式，用精准的数据为企业画像助力企业征信系统，建立企业关联，预测企业征信系统的相关指数。

本文以企业画像和大数据分析为关键字对相关文献进行检索，并对其进行研究归纳和总结，对文献中有关企业画像的相关信息整理，对企业画像标签体系建立的技术进行相关研究，并且在用户画像模型的基础上提出改进，提出构造产品画像、企业画像等事物画像的新应用。

3 企业画像生成模型的研究

当前国内外对企业画像的相关研究成果比较少，在已有的企业画像应用中可以大致把企业画像的构建过程分为 4 个阶段。

(1)明确企业画像建立的战略意义以及企业画像建立的初级目标和效果期望值。企业画像是在真实数据的基础上，对企业数据进行分类整理，帮助企业自身、政府、银行、券商、会计师、律师、投资方等用户计算企业全方位的信息，包括发现和挖掘企业之间的关联关系，找寻未知关系以促进企业合作；在企业征信中对其规模、信誉、风险能力进行评估，识别企业资本行为，构造企业风险评估模型等；在企业品牌构建、传播以及营销时提供了重要的数据支持；企业画像也给消费者和投资者提供了有利的数据参考。

(2)为企业画像进行建模。在这一阶段中，我们需要结合实际的数据以及企业品牌的特性来规约数据实体并明确企业与企业、企业与市场之间的关联关系。

(3)企业画像的维度分解。根据我们调查获取的相关性原则关系，对企业的属性进行维度分解，在不同维度下把握企业画像的颗粒度。我们可以通过以下 5 个维度来刻画用户画像，并根据每一维度的特点来规定画像的颗粒度。1)企业属性：企业经营证件类型、经营范围、经营资质起始日期、经营资质截止日期、法人证件类型、登记机关、法人名称、企业员工数量等。根据对企业属性的描述，我们可以知道企业的基本信息特征，对企业有初步的了解。2)企业信用属性：企业公示信息、纳税信息、发票购销实际交易数据等。企业信用属性描述企业的信誉度，可以帮助消费者、投资者和企业利益相关者在购买股票、企业投资和企业合作中提供重要的数据参考。3)企业交易特征，我们可以根据企业交易的内部数据、产品销售地、产品用户人群等数据来描述企业交易特征。企业交易特征一方面有助于企业建立良好的营销策略；另一方面也有助于政府监控整个地区内行业的交易情况，实现对市场的宏观调控，及时对地区内的产业结构做出相应的调整升级和招商引资政策的变化。4)企业内外关联特征，根据企业间合作链信息、企业内部高管信息以及股东代表等信息建立企业外部联系图谱和企业内部关系图谱，这样有助于了解企业内部管理信息和外部合作发展趋势。5)企业评价信息，根据社交舆论信息、企业网站招聘评论数据、企业员工内部评价信息，提升企业管理和服务质量，根据数据分析结果趋利避害，构造良好的企业形象。我们还可以根据企业自身的特色从横向上扩展企业维度，从纵向上细化颗粒度属性。

(4)企业画像的应用。根据我们对用户画像维度分解的程度，用户画像可以应用在企业营销、政府税收系统管理、股票证券交易系统以及企业招聘系统等多个与企业相关的领域。在企业画像的具体应用阶段，针对不同人员的角色需求来设计企业画像平台的功能和具体的应用流程。在明确了企业画像构建的步骤后，我们可以对企业画像的标签体系进行建模，从企业基本属性、企业业务往来数据、企业税务数据以及企业行政处罚数据和企业招聘数据等分析企业活动，提取企业标签。在企业画像应用时要具体需求具体分析，一般的企业画像标签体系建模如图 1 所示。

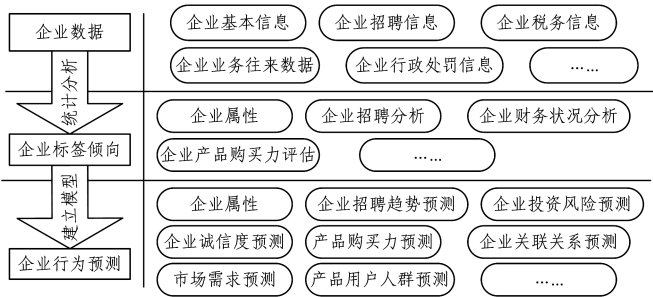


图 1 企业画像标签体系建模

4 基于大数据平台对企业画像设计方法的研究

利用大数据技术研究企业画像，可以从定性化和量化两个角度来分析。前者利用数据获取的方法对企业进行前期调研，从宏观上对企业进行定位，根据企业固有的属性、法人代表、投资人、财务负责人、企业选址、交易范围、企业规模等数据对企业进行定性化的画像。对于企业的量化分析，我

们可以按照不同用户群体的关注点进行划分,具体的用户画像的建设框架如图 2 所示。

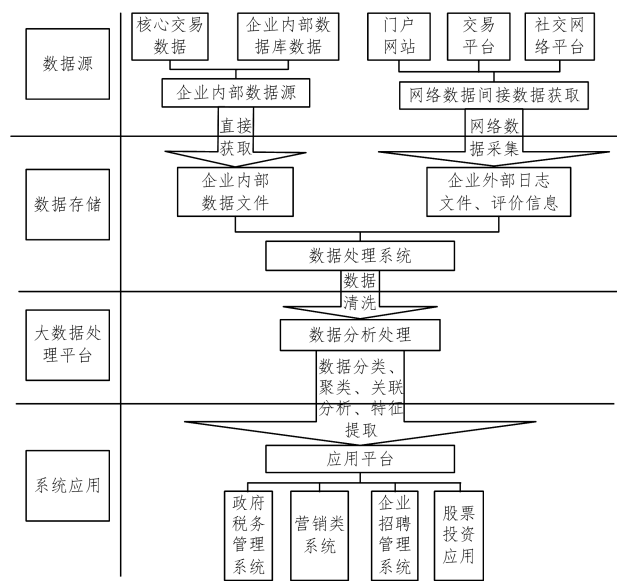


图 2 企业画像的研究框架

数据源获取:获取企业画像的数据通常有两种渠道:1)对于企业基本信息数据、财务数据、企业交易数据以及其他企业内部核心数据等数据信息,可以从企业中直接获取;2)对于企业不能直接获取到的数据,如企业社交平台中用户对企业的评价信息、招聘网站中求职应聘者对企业的评价等数据,我们可以在相关的门户网站以及一些社交平台进行间接获取。目前,我们常用的获取网络数据的方法有 Larbin,Nutch 和 Heritrix,spiderpy,它们都是开源的网络爬虫工具。文献[10]提出了一款并行的可迁移的网络爬虫工具,提高了网络效率和网络数据的下载速度。文献[11]通过把 Scrapy 和 Reids 相结合并加以改进,设计了效率更高、更稳定的分布式的 Scrapy-Reids 网络爬虫框架。文献[12]对网站信息组织以及网页 URL 进行研究分析,提出一种基于 URL 模式集的主题爬虫模型,它可以快速爬取相关网站的主题页面,提高了爬取数据的查全率和查准率。但是,随着网络信息资源爆炸式的增长,网络爬虫的爬取速度和准确性还需要不断地提高。

数据存储技术:对于企业的数据,文本形式数据占大多数,但是也有许多音频和视频等非结构化的数据。为了解决这种海量数据集合、多重数据种类带来的难题,可以把企业数据信息存入 NOSql 中,它最大的特点就是可以并发式地处理大量的、多类型的数据。在对这些海量数据进行任务调度处理时,可以借助 HDFS 这一分布式文件处理系统,它具有高容错性等优点,可以提供很高的聚合数据带宽,一个集群中支持数百个节点千万级别的文件。

大数据处理平台:在处理用户画像的问题上利用大数据处理技术可以更加实时、精确地得到用户标签,实现海量企业数据的批处理计算。一个完整的大数据平台应该提供离线计算、即时查询、实时计算这几个方面的功能。目前用得最多的大数据处理架构是 Hadoop,以及在它基础上提升的 spark 大数据处理框架。对于企业画像,国内外相关的研究较少,我们可以把用户画像的处理技术应用到企业画像中。利用 Hadoop 技术框架处理企业的海量数据,分析数据,进行并行化

处理。文献[13]利用 MPP+Hadoop 的大数据处理技术框架对手机用户进项画像,把用户画像应用到个人征信中,实验证明 MPP+Hadoop 的框架在数据处理的实时性、可扩展性以及数据复杂查询与分析的能力上有很大的提高。文献[14]利用手机 APP 用户的使用行为数据,通过支持向量机(SVM)的方法建立了用户流失预测分析的模型,并且利用 Hadoop+mahout 对用户的 20 多项行为指标进行聚类分析,对用户偏好和特征进行分析预测。在 Hadoop 基础上提出的 Spark 可以支持分布式数据集上的迭代作业,可以用来构建大型的、低延迟的数据分析应用程序。文献[15]基于 Spark 集群的数据管理架构,实现数据仓库 ETL 自动化的操作,使用 NOSql 数据库存储标签,构建可扩展的用户画像系统。文献[16]利用 Spark 分布式流处理框架,用最小二乘法改进传统的协同过滤的推荐算法,并且将 solrcloud 分布式技术引入对新闻文本的处理中,从而提高分布式系统的计算能力。

数据处理方法:创建企业画像的过程是基于企业的基本属性和企业行为活动数据来提取企业特征以及建立企业之间的关联关系并对企业活动预测指导的一个数据分析的过程。在构建企业画像过程中常用的数据处理方法有特征标签提取、关联分析、分类与预测等。

企业画像的核心工作是提取企业标签,企业标签要求呈现出两个主要的特征:1)语义化,即能很快理解每个标签的含义;2)短文本,即每个标签通常只表示一种含义^[4]。对于企业的非结构化数据,处理标签提取时,分析难度较大,这些非结构化数据大多是短文本类型的数据,长度有限,特征的共现性不大,而且传统的文本数据处理方法会导致文本语义特征稀疏和语义敏感等问题^[17]。对企业进行特征提取时,我们可以根据企业数据的特性,如一些关键词“上市”“融资”等敏感性词,增加权重,对于与企业相关的名字,更新原有词典的知识库。现有的对于短文本预处理的方法有基于词性标注以及文献[18]提出的引入拼音序列表征进行文本处理输出。在计算单词权重方面,最为有效的实现方法是 TF-IDF 和 word2vec。它们用来评估一个词语对于一个文件集或者语料库中的一份文件的重要程度,字词的重要性与它在文件中出现的次数呈正比关系^[19]。其中,word2vec 在百万数量级的词典和上亿的数据集上进行高效训练^[20]。除了上述两种文本特征权重评估算法,常用的特征评估算法有度量两个对象之间相互性的计算语言模型常用的互信息法、反映文本主题概率的期望交叉熵算法以及信息增益算法等,这些算法各有优劣,我们要根据实际数据对它们进行取舍。在文本预处理后的提取特征值阶段,常用的提取特征的方法如表 1 所列。

表 1 特征提取方法的比较分析

特征提取方法	优点	缺点
k-means	可以在无标签的情况下进行聚类分析,实现相对比较简单	实验初始值和噪声点对结果影响大,容易陷入局部最优
LDA	适用于文档大并且多主题的文本	处理短文本数据在主题切分上的准确度比较低
NB	所需估计的参数很少,对缺失数据不大敏感,算法也比较简单	无法处理基于特征组合所产生的变化结果
CNN	提高了模型准确度,有优异的特征自抽取能力	训练耗时严重,需要大量的样本训练

近年来,随着计算机计算速度的提升以及深度学习技术

的不断发展,越来越多的学者开始把深度学习的技术运用到自然语言处理和特征提取上来,深度学习通过多层神经元进行计算,使得最终训练结果最优^[21]。卷积神经网络(CNN)在文本处理方面取得了很好的效果, Kim^[22]在 CNN 基础上提出 4 种改进的基准模型,即 CNN-rand,CNN-static,CNN-non-static 和 CNN-multichannel,并且通过实验证明,改进的 4 种基准模型在短文本处理的精确度上有所提高。文献[18]中,作者利用 CNN-static 和 CNN-non-static 模型,引入中文的拼音序列来对原始的文本进行语义拓展,建立了字符级和词级的双输入矩阵,分类实验效果提升明显。

对数据进行特征提取后,我们会进行后续的企业间的关联分析以及分类预测推荐等应用。首先,通过分析挖掘海量的企业数据信息,发现它们之间的关联关系,从中找寻规律以进行企业行为预测和企业推荐。根据对企业之间的交易数据以及全地区范围内企业管理层、负责人、董事会成员进行关联分析预测并构造关系图谱,建立企业之间的关系网络图。通过企业关系的关联预测可以方便政府部门对本地区市场经济进行监控,提高企业之间寻求合作的效率。数据挖掘中常用的关联分析算法有 APRIORI 算法、GRI 关联规则算法和 FP-growth 算法,如表 2 所列。

表 2 关联规则方法的比较分析		
特征提取方法	优点	缺点
APRIORI 算法	基于频繁集的关联分析,首先要找出数据项的所有频繁集,在满足最小支持度和可信度的前提下产生强关联规则	计算过程中可能会产生大量的候选数据集,在寻找频繁数据集时需要不断地重复扫描数据库,因此会增加算法的时间复杂度
GRI 关联规则算法	可以处理多重输出栏位,既能处理字符型输入字段,又能处理数值型输入字段	训练速度低于 APRIORI 算法

文献[23]利用 APRIORI 算法对图书销售数据进行分析,发现用户购买的书籍之间的关联关系,进而调整销售策略,为用户推荐购买书籍。文献[24]利用 APRIORI 算法来进行基于标签的关联规则挖掘研究,并通过实验证明标签预测法结合 APRIORI 算法能达到很高的准确性。Han 等提出了基于频繁模式树发现频繁模式的 FP-growth 算法。FP-growth 算法只进行两次数据库扫描,它不使用候选集,直接压缩数据库成一个频繁模式树,最后通过频繁模式树生成关联规则^[25]。文献[26]对挖掘出的关联规则进行聚类分析,然后对聚类后的关联规则对照已有的知识库进行新颖度分析,并筛选出规则中高价值的规则进行输出。对关联规则的重复挖掘方法大大提高了数据关联分析的效率和精确性。

处理企业信息分类预测分为两步:1)提取已经分类数据的数据特征,根据属性特征建立分类函数或分类模型,然后计算总结出数据特征,对企业进行分类;2)进行预测分析^[27],我们可以根据每一类对象的属性、观测值和相关的资料进行预测分析。通常对数据分类的目的是得到一个分类器,通过所得到的分类器将经过预处理的数据对象归入一个给定的类别;预测的目的是从历史数据记录中自动推导出对给定数据的推广描述,从而能够对事先未知的数据进行预测。分类与预测中常用的方法有 K 近邻算法、决策树算法、SVM 支持向量机算法、人工神经网络算法等,其优缺点如表 3 所列。

表 3 预测推荐方法模型的比较分析		
预测推荐方法	优点	缺点
KNN	不需要预先对文本进行参数预测,适合对稀有事件进行分类	理解性较差,计算量大,需要对所有样本数据进行距离的计算,然后再进行分类
决策树算法	典型的分类与回归算法,模型的可读性强,分类速度快	创建树时可能会出现过拟合的缺点,其结果也不太稳定
SVM	用于解决小样本情况下的机器学习问题,可以提高泛化性能,解决高维问题,避免神经网络结构选择和局部极小点问题	对缺失数据比较敏感,也不适用于大文本数据

企业画像的应用:企业画像是对企业数据进行分析处理,得到有价值的信息并以可视化的方式呈现出来,它的最终目的是应用于现实中。根据对企业数据分析的结果,我们可以把企业画像应用到政府的税务管理系统中、企业品牌的营销决策、企业招聘管理以及证券行业、股票交易等与企业相关的事务处理中。

结束语 通过文中的分析可知,企业画像从不同的维度来分析企业的行为属性,从而为企业品牌的营销、政府税务管理以及其他有关企业的应用提供更加准确的数据,进而提升企业的竞争力和政府的市场监管力度。近几年,国内外学者研究得最多的还是利用用户数据为用户画像,对企业画像的研究还是比较少,我们可以利用企业画像和用户画像的理论,结合大数据平台的技术支持,根据社会需求,对“事件”“产品”“疾病”等实体对象进行画像。例如,对于微博的热搜榜的话题事件,我们可以根据其点击率、微博用户评论等数据进行多维度的分析,监控事件的发展趋势,引导舆论的走向。

企业画像构建的难点总结:利用大数据技术处理企业画像目前还处于初级阶段,虽然大数据技术日趋成熟,但是在企业画像这一应用领域应用得还是比较少,对于企业海量数据的获取、存储、处理以及分析还未形成一个统一的模式。首先从数据获取方面考虑,很多公司内部数据的保密程度相对较高,不同行业之间的信息流通相对困难,会形成信息孤岛的问题,对于企业的一些网络数据的真实性也是我们获取过程中需要考虑在内的;其次在企业数据的存储处理过程中,对于海量的数据来说就要求存储系统有较强的扩展能力,在涉及企业网上交易或者金融类相关的数据时,如企业股票行情的趋势、广告投放等,这就要求存储系统在必须能够支持上述特性的同时保持较高的响应速度。除了上述的大数据存储方面在扩展能力和实时性方面的问题我们还需要考虑数据存储的安全性、成本以及灵活性等问题。从企业数据的处理以及分析角度来说,虽然目前 Hadoop 和 spark 等大数据处理框架以及一些数据分析算法已经日趋完善,但是这些处理方式在企业数据上运用得还非常少,还需要我们不断地实验研究,提高计算速度以及准确性。针对企业数据获取困难和处理速度效率低下的问题,我们可以利用云存储和云计算的高效性特点对数据进行处理。在对企业数据处理时考虑企业数据的特殊性,对占大多数的非结构文本数据处理时可以建立企业敏感词库^[28];对短文本数据处理时根据短文本数据小共现性和高维稀疏性的特点,选择合适的文本处理方式,结合字符、拼音以及词性等特征,利用深度学习的方法以及多层输入、权值共享的特点提高特征提取的准确度^[29]。

参 考 文 献

[1] 郝胜宇,陈静仁. 大数据时代用户画像助力企业实现精准化营

[J]. 中国集体经济,2016(4):61-62.

[2] 孙晔,杨照东,陈德华,等. 大数据用户画像技术在商业银行的应用[J]. 数字通信世界,2016(9):86-88.

[3] 刘海,卢慧,阮金花,等. 基于“用户画像”挖掘的精准营销细分模型研究[J]. 丝绸,2015,52(12):37-42,47.

[4] 陈志明,胡震云. UGC 网站用户画像研究[J]. 计算机系统应用,2017,26(1):24-30.

[5] 吕辉,许道强,仲春林,等. 基于电力大数据的标签画像技术与应用研究[J]. 电力信息与通信技术,2017,15(2):43-48.

[6] 陈一平. 今品指数:用大数据为企业画像[J]. 华夏时报,2015-07-06(021).

[7] 王磊,郑国勇. 大连国税:大数据为企业精准画像[N]. 中国税务报,2016-10-17(A04).

[8] 王雪. 基于企业画像的公示数据查询系统设计与实现[D]. 大连:大连海事大学,2017.

[9] ZHONG C J. “Picture” for domestic enterprises to go to the western “circle”[N]. NanJing Times,2016-04-25(A02).

[10] 李代伟,谢丽艳,钱慎一,等. 基于 Scrapy 的分布式爬虫系统的设计与实现[J]. 湖北民族学院学报(自然科学版),2017,35(3):317-322.

[11] 胡萍瑞,李石君. 基于 URL 模式集的主题爬虫[J]. 计算机应用研究,2018(3):1-2.

[12] GOUNARIS A,TORRES J. A Methodology for Spark Parameter Tuning[J]. Big Data Research,2018,11:22-32.

[13] 李映坤. 大数据背景下用户画像的统计方法实践研究[D]. 北京:首都经济贸易大学,2016.

[14] DING W,WANG T,LIU X H,et al. Research on mobile phone user’s portrait and credit based on large data technology[J]. DesigTechn Post Telecommun,2016(3):64-69.

[15] 刘鹏. 基于 Spark 的数据管理平台的设计与实现[D]. 杭州:浙江大学,2016.

[16] 徐凯. 基于主题抽取演化模型的实时新闻推荐系统[D]. 广州:广东工业大学,2016.

[17] SEVERYN A,MOSCHITTI A. Learning to rank short text pairs with convolutional deep neural networks[C]// Proceedings of the 38th International ACM SIGIR Conference. 2015:373-382.

[18] 余本功,张连彬. 基于 CP-CNN 的中文短文本分类研究[J]. 计算机应用研究,2018,35(4).

[19] 李春梅. 基于 TF-IDF 的网页新闻分类的研究与应用[J]. 贵州师范大学学报(自然科学版),2015,33(6):106-109.

[20] 谢慧志. 基于用户选购及在线点评行为的门店画像研究[D]. 广州:华南理工大学,2016.

[21] GLOROT X,BORDES A,BENGION Y. Deep spares rectifier neural network[C]//Proceedings of International Conference on Artificial Intelligence and Statistice. 2011:315-323.

[22] KIM Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of Empirical Methods in Natural Language Processing. 2014:1746-1751.

[23] 张燕琴,潘利强. Apriori 算法在 WEB 的图书销售智能系统中的应用[J]. 网络安全技术与应用,2017(9):58-59.

[24] 刘志刚. 基于标签关联规则的挖掘与研究[J]. 科技创新与应用,2017(26):26.

[25] 左爱文. 数据挖掘技术在气象数据中的应用[D]. 西安:西安电子科技大学,2006.

[26] 张玲玲,周全亮,唐广文,等. 基于领域知识和聚类的关联规则深层知识发现研究[J]. 中国管理科学,2015,23(2):154-161.

[27] 夏光虎,贾宇波,范红丹. 分类与预测挖掘在信用风险评估中的应用研究[J]. 工业控制计算机,2012,25(7):71-72.

[28] SZEGED Y C,LIU W J,YANG Q. Going deeper With Convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston,USA,2015.

[29] JOHNSON R,ZHANG T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks[C]// Proceedings of North American Chapter of the Association for Computational Linguistics. 2014:103-112.

(上接第 42 页)

[50] 张玉洁,杜雨露,孟祥武. 组推荐系统及其应用研究[J]. 计算机学报,2016(4):745-764.

[51] ORTEGA F,BOBADILLA J,HERNANDO A,et al. Incorporating group recommendations to recommender systems: Alternatives and performance[J]. Information Processing & Management,2013,49(4):895-901.

[52] MASTHOFF J. Group Recommender Systems[C]//Combining Individual Models. 2010:677-702.

[53] KASSAK O,KOMPAN M,BIELIKOVA M. Personalized hybrid recommendation for group of users; Top-N multimedia recommender[J]. Information Processing & Management,2016,52(3):459-477.

[54] PERA M S,NG Y K. A group recommender for movies based on content similarity and popularity[J]. Information Processing & Management An International Journal,2013,49(3):673-687.

[55] BALTRUNAS L,MAKCINSKAS T,RICCI F. Group recommendations with rank aggregation and collaborative filtering[C]// ACM Conference on Recommender Systems(Recsys 2010). Barcelona,Spain,2010.

[56] GARCIA I,SEBASTIA L,ONAINDIA E. On the design of individual and group recommender systems for tourism[J]. Expert Systems with Applications,2011,38(6):7683-7692.

[57] NAAMANI-DERY L,KALECH M,ROKACH L,et al. Preference elicitation for narrowing the recommended list for groups[C]// ACM Conference on Recommender Systems. 2014.

[58] JAMESON A. More than the sum of its members; challenges for group recommender systems[C]// Working Conference on Advanced Visual Interfaces(AVI 2004). Gallipoli,Italy,2004.

[59] QUIJANO-SANCHEZ L,RECIO-GARCIA J A,DIAZ-AGUDO B,et al. Social factors in group recommender systems[J]. Acm Transactions on Intelligent Systems & Technology,2013,4(1):1199-1221.

[60] QUEIROZ S R M,DE CARVALHO F A T,RAMALHO G L,et al. Making recommendations for groups using collaborative filtering and fuzzy majority[C]//16th Brazilian Symposium on Artificial Intelligence(SBIA 2002). Bittencourt G,Ramalho G L,2002:248-58.

[61] GUO J,ZHU Y,LI A,et al. A Social Influence Approach for Group User Modeling in Group Recommendation Systems[J]. IEEE Intelligent Systems,2016,31(5):1.

[62] ARDISSONO L,GOY A,PETRONI G,et al. Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices[J]. Applied Artificial Intelligence,2003,17(8-9):37-41.

[63] GORLA J,LATHIA N,ROBERTSON S,et al. Probabilistic group recommendation via information matching[C]// International Conference on World Wide Web. 2013.