

# 数据挖掘技术在互联网领域的应用研究

吴琛

(江西信息应用职业技术学院 江西南昌 330043)

**摘要:**互联网随着时代的发展,演变成信息传输的综合体。随着现代计算机信息技术的飞速发展,网络数据不断地增长,我们步入了大数据时代。多年来,人们使用数据挖掘技术分析、整合海量的数据,从大量数据中,发现规律和知识,挖掘出有价值的信息。本文从数据挖掘技术的研究入手,结合不同领域的特点,剖析了数据挖掘技术在不同网络平台中的应用。我们可以发挥数据挖掘的社会价值,改善人们的生活,完善平台的数据服务,最大化数据挖掘的积极作用。

**关键词:**互联网;数据挖掘;应用;网络平台

## Research on the Application of Data Mining Technology Based on Internet Majors

WU Chen

(Jiangxi Vocational & Technical College of Information Application 330043)

**Abstract:** Along with the fast pace of the times, Internet has evolved into a synthesis of information transmission. With the rapid increase of modern computer information technology and the continuous growth of network data, we have entered the era of big data. Over the years, data mining technology has been used to analyze, integrate massive data; thus, discover laws and find out valuable information from a large number of data. Based on the research of data mining technology and the characteristics of different fields, this paper analyzes the application of data mining technology in different network platforms. The social value of data mining can be properly explored to improve people's lives, improve data service of the platform, and maximize the positive role of data mining.

**Key Words:** Internet; Data Mining; Application; Network Platform

### 1 前言

互联网随着时代的发展,演变成信息传输的综合体。无论是个人,还是企业、平台等都频繁使用互联网,如个人通信、企业宣传、媒体平台传播等都依赖于它。随着现代计算机信息技术的飞速发展,网络数据不断地增长。社交网络、电商平台、金融平台、教育平台、音乐平台、医学平台等不同领域的平台,各种服务工具等,都提供和处理着大量网络数据,同时,也作为数据的来源。在互联网上,有着丰富的数据信息,如某电商平台数亿的用户每天产生的交易数据约 20TB 左右,某教育平台课程资源等数据也超过 100TB。数据通过互联网地传输,每时每刻都在更新,逐渐形成大量且复杂性较

高的信息。互联网为人们提供了许多应用及服务,人们在生活中几乎离不开互联网,换句话说,每个使用互联网的人都在为互联网提供着数据,并且这些数据是实时进行处理的。

### 2 大数据时代

跟随着信息化的脚步,我们已经正式步入了大数据时代。在信息日益渐增的情况下,一定程度上加速了人们生活的进步,也促进了不同领域的信息交流的时效性。在日常生活中,联络他人可以直接用网络通信,买东西可以足不出户,想了解天下事只需要上网,甚至人们的衣、食、住、行都可以通过网络方式获取信息。在数据时代,几乎所有的信息都可以通过网络来看到相应数据。这也意味着,网络数据在生活

中已经占据了至关重要的角色。

在大数据时代,人们的身边每天充斥着大量的信息。在这些信息当中,有准确或者不完善的信息,有对于不同需求的人来说,看似有用或者无价值的信息等。因此,不同领域的网络平台对于数据的应用,就需要将不同人的看法搜集起来,利用数据挖掘技术,进行有效整合,充分将大数据的价值利用起来。此外,通过数据挖掘技术,可以将繁杂的数据进行分类,提取有效信息,防止数据的重复性和枯竭,改善网络环境。

### 3 数据挖掘技术

信息的交互、网络数据的实时分享,在一定程度上加速了人们的生活发展,每个人都能享受到大数据所带来的便携。与此同时,随着互联网行业的不断发展,平台大规模的增加,大量网络数据递增。数据具有海量且复杂的特征,其中,相对价值较高的数据所占百分比比较小。数据中包含很多噪声数据,数据内容可能缺失或者分散,导致数据质量不一致。需要利用数据挖掘技术从大量的随机并且杂乱的数据集合中,通过采用一定的算法对信息进行提取,发现规律和有用的价值信息,重新整合数据信息库。

数据挖掘不仅是收集某类数据,也是通过从大量不同类型的数据中,挖掘出各种有价值的信息如趋势、模式及关系的过程,从而达到改善社会治理、提高效率、提高数据系统使用率的效果。数据挖掘技术是对大量数据进行分整合、统计与分析并通过机器学习方法、人工智能方法或数据挖掘方法深度分析,发现规律和知识,并运用于各个领域。

#### 3.1 数据挖掘方法和步骤

数据挖掘方法有许多,例如有基于遗传算法,粗集方法,决策树方法和神经网络方法,通常可以选择一种来建模或者多种来对比校验。通常数据挖掘的步骤为:分析问题、对数据的选择和预处理、创建、调试数据模型、挖掘数据和维护数据挖掘模型和评估结果。下面以教育平台数据挖掘为例,解析数据挖掘的一般步骤:

##### 3.1.1 分析问题

明确数据库和挖掘目的。如为了推广优质课程,可分析教育平台中评分高的课程,可根据课程内容分类、课件点击率、用户关注度、评分、评价等。

##### 3.1.2 对数据的选择和预处理

获取、清洗数据,将不完整的数据去除或者修正,把数据按照一定格式来进行整合。如收集课程平台中的评分、关注度、课件点击率等数值,课程内容分类、评价等文本信息。针对不同格式的信息,设置自定义规则,如评价为空则不收集,评分高于某数值则收集对应的评价信息。

##### 3.1.3 创建、调试数据模型

选择所需的数据挖掘算法,将数据代入算法中来创建模型。再多次比对、校验、调整中,获得满足需求的模型。如可以在分析软件中对比数据,判断模型准确性。

##### 3.1.4 挖掘数据和维护数据挖掘模型

随着数据量的增加,需要对模型进行调整和维护,从验算中找到数据间的关系,并且不断完善模型,保持模型的活力。如分析课程评分高的影响因素。

##### 3.1.5 评估结果

根据模型和数据分析的结果,发现规律和有用的价值信息,重新整合数据信息库。如:点击率多、评分高且评价好的课程大多数为某一学科,可预测该学科为近期关注度较高课程。平台可以将推荐规则设为该学科,优先给用户推荐。

#### 3.2 数据挖掘的功能

对于不同领域来说,可以充分发挥数据挖掘的不同功能。数据挖掘有分类、估计、预测、关联等相关功能。

##### 3.2.1 数据分类

可以根据需求中分析对象的属性,按照某些属性来进行分类,并且建立不同类别组。如将教育平台中使用者的教育程度进行分类,小学、初中、高中、大学等,挖掘出某些信息后,则可向不同教育程度的用户,推荐对应教学层次的课程。

##### 3.2.2 数据估计

即根据已有连续性数值的相关属性数据,以获取某一属性未知的值。如在金融平台中按照使用者的教育程度、行为来推估信用卡消费额。

##### 3.2.3 数据预测

根据对象属性中之前的一系列信息,来推测出该属性值未来的变化。如:在音乐平台中某用户关注了某歌曲,来预测其未来可能会关注同类型的歌曲或歌手。

##### 3.2.4 数据关联分组

即判断哪些相关对象应该放在一起,设计出吸引人的产品群组,且购买的概率将会大幅提升。如:在购物平台中,一个用户搜索了牙膏,那么在物品推荐页面中,可以推荐牙膏和牙刷的组合。

##### 3.2.5 数据群集

从某一特殊特征分隔出相同属性群组,定义为某一数据群集。观察其中数据奥秘,通过观察数据为何被群集在一起的,可以了解数据间的关系,以及这些关系将会如何影响预测的结果。

如:使用相同品牌电脑的人群,热爱出国旅行的人群,喜好自助餐的人群等,可以从不同人群中,分析他们的消费水平,教育程度,家庭成员人数等,找到人群数据的特征加以分析。

##### 3.2.6 时序数据序列模式挖掘

以时间序列为主要属性,找出该数据与时间相关的行为模式,从中分析序列因时间而发生的改变。由此进行数据挖掘,来预测未来的效果,如:预测未来的股市走向、股价的波动。

### 4 数据挖掘的应用领域

不同行业中已经有许多平台运用了数据挖掘技术,将数

据挖掘结果有效地提升平台服务品质。在互联网中,常见的电子商务、金融、教育、音乐、医疗保健等平台通过数据挖掘技术,把数据信息进行挖掘并且整合,再将有益数据结果融合在不同应用中。数据挖掘的结果不仅来源于每一个人,也关系到个人、企业、社会未来的发展。需要发挥数据挖掘的社会价值,改善人们的生活,完善平台的数据服务,最大化数据挖掘的积极作用。

每一个用户都有和他人相同或不同的属性,这些属性都意味着数据的价值。数据产生的同时,也伴随着计算机数据的产生,例如网络中的文字信息、图片信息、音视频信息等。可以采用数据挖掘技术对网络数据进行分析,通过细分用户或用户行为,挖掘出不同用户的需求。不仅可以提供给用户个性化的服务,还可以研发适合不同用户的产品,来满足当下大数据时代人们对于网络智能化的要求。

#### 4.1 电子商务领域的应用

互联网的普及颠覆了许多传统的实体行业,网络购物也成为了一种时尚,中国的营销市场取得了突破性的进展。计算机数据挖掘技术在电子商务领域的应用多数体现在用户的喜好上。运用挖掘技术,根据个人在网络浏览商品的记录,从购物习惯推算出用户对于商品的喜好。大多数平台则通过这项技术,采集大量的数据进行挖掘,分析用户群体所喜好的商品。为企业提供智能化的数据分析,优化电子商务平台的运营投放,改善营销手段,提升营销服务。

此外,除了购物网络化,电子商务平台正为用户打造一个完整的生态服务体系,提供用户日常生活服务等,如出行选择方式,订餐外卖,节假日的出行旅游和美食口味偏好等。根据每个用户所提供的数据来源,加以数据挖掘,为用户提供最契合自身特性和符合私人需求的定制化服务体验。

#### 4.2 金融领域的应用

银行业已经深刻意识到大数据战略对其经营管理、客户营销和产品优化等方面的重要作用,以及对银行未来发展方向的深刻影响。深度挖掘大数据蕴涵的巨大价值,从而推动银行产品、服务和管理的创新。在银行产品方面,可根据银行产品特点,结合客户消费习惯,计算出客户贡献率。例如可以查询当前客户使用网上银行的频率,将这一数据进行排名,对排名较高的客户进行适当的广告投放或者电话营销,推荐相关产品。在客户管理方面,依据客户的资产、购买产品等情况,计算他们的内部产品贡献率。这部分数据则可挖掘出优质客户群体和有潜力的客户群体。银行则可提供针对不同客户的差异化服务,对优质客户交叉销售提供支持,提高客户的满意度和忠诚度。

在金融投资方面,无论是股市的投资或是期货等金融类的投资,往往获得巨大收益的同时一定也伴随着巨大的风险。投资者们希望通过数据分析来降低风险,实现利益最大化。数据挖掘技术恰好能对数据进行整合分析,并且进行科学的预测,某种程度上可以帮助用户降低投资风险。

#### 4.3 教育领域的应用

大数据将传统的课堂教学模式改变,从“小教室”到“世界大讲堂”,数以万计的课程教学内容呈现在互联网上。网络在线教育,点击量达到亿万次。在网络平台中的知识资源里,大数据使得教育发展有着更广阔的空间。大数据将掀起新的教育革命,每一条数据记录了每个学习者独有的信息。而这些数据信息蕴含着学习者的学习,反馈着课程的教学质量,影响着教师的教学,完善着教育政策制定的方式。

以往学习者的学习特征都消散于教室、自习室等地方,很多数据难以收集。现在利用大数据技术,将这些数据存放在网络上或者下载到本地,反复研究。例如在网上学习者点击的某门课程频率,停留在某部分知识内容上的时间,知识测验的分数,学习方式的喜好等都以信息的方式呈现。通过研究群体学习的特征,看似杂乱无章的大量信息,则可利用数据挖掘技术分析其中的规律。对于教学者而言,数据挖掘技术可以通过学生的网络行为,了解学生对知识掌握的程度,还可以规划教学内容的重、难点,提升教案品质。除此之外,针对点击率高且评价好的课程、教案等,挖掘出其中受欢迎的原因。在网络上的大量学习者、教学者、学校的信息,都值得深度挖掘。一方面,通过数据挖掘,我们还能发现一些重要信息,利用学生反馈的信息,来改善学生的成绩,为他们提供个性化的服务。另一方面,通过分析课程资料数据、教案数据、学生行为数据、课程反馈等,挖掘数据的结果有助于教学质量,教学者们不仅可以互相分享,还能够教学沟通,优选出精品课程,提升教师整体素质。

#### 4.4 音乐领域的应用

近年来网络上的音频、视频剧增,海量的音乐数据依托于网络平台。大众获取音乐信息的模式逐渐网络化,通常在网络平台收听音乐、查询音乐动态等。音乐网络平台不仅迅速整合了多文化、多形式的音乐资源,还将受众用户的喜好和想法进行信息反馈。对于各大音乐平台来说,音乐资源量和用户体验度是平台优势的重要指标。通过对音乐数据的挖掘,一方面可以分析出当下流行的音乐资源,将其收录到平台,提升资源量。另一方面,结合用户的听歌数据等反馈信息,研发出人性化的推荐功能,如某平台的推荐相似歌曲、相似艺人等功能,大幅度提升用户体验。音乐数据挖掘有不同的方式,可以根据用户的基本信息、收听记录中的歌曲、歌曲曲风信息、艺人信息来寻找潜在联系,进行多维度关联挖掘。例如根据用户收听过某些歌曲,分析这类歌曲的曲风,将属于同类曲风的其他相似歌曲推荐给用户。也可以根据用户所关注的艺人信息,推送相关专辑、演唱会等。

另外,音乐的关注度意味着作品的热度。音乐人伴随着其高热度的作品,可获取相应互联网“流量”,即明星效应,能够给个人和经纪公司带来经济收益。其音乐的衍生品,例如演唱会门票、艺人专辑等。由此可见,用户所关注的信息非常重要,若将这些数据进行深度挖掘,例如把当下热门艺人、最

流行专辑等数据结果提供给音乐厂商,则可以辅助音乐衍生品的精准营销,开拓音乐的商务市场。

#### 4.5 医学领域的应用

随着医疗信息化的普及,医院管理逐步现代化。如今医院系统中将病人档案资料、医疗器具的管理、药品等方面的数据都录入系统,通过这些数据如诊患者结构、就诊时间等加以整合挖掘,数据报告则具有指导作用,如某科室的诊断时间过长,可以适当调整;患者服用的长期药品在某段时间有效等,能够帮助患者康复和使得医院服务质量提升。

在医疗工作中,工作者的诊断方式同样发生着改变。传统研究模式是利用自身经验诊断,而在大数据时代,人类疾病方面所蕴含的数据信息可以进行查阅。如果只是依靠自身经验诊断,说服力较弱。如今,工作者可以通过数据挖掘技术将患者的病史、出现的临床症状、相似患者的病史等多方面的数据,通过对数据库中信息进行综合研究。不仅可以挖掘出有意义的诊断规则,辅助疾病诊断,提高医生的诊疗效率,缓解医患矛盾,还可以切实对患者的疾病进行预警,推荐治疗措施。

## 5 结语

综上所述,跟随着信息化的脚步,已经正式步入了大数据时代。在信息日益渐增的情况下,通过数据挖掘技术来推进网络信息数据的处理是十分必要的。本文从数据挖掘技术的研究入手,结合不同领域的特点,剖析了数据挖掘技术在不同网络平台中的应用。人们使用数据挖掘技术,通过从海

量数据中挖掘出有利于不同领域发展的数据,达到改善社会治理、提高数据平台使用率的效果,并运用于各个领域。可以发挥数据挖掘的社会价值,改善人们的生活,完善平台的数据服务,最大化数据挖掘的积极作用。

#### 参考文献:

- [1] 闫伟伟. 基于数据挖掘的数字出版 CRM 应用研究[D].西安工业大学,2018.
- [2] 李小庆. 大数据挖掘在银行业务领域的应用[J].金融科技时代,2017(05):15-19.
- [3] 何光凝. 数据挖掘在计算机网络安全领域的应用研究[J].技术与市场,2016,23(08):13+15.
- [4] 王好姝. 基于大数据的数据挖掘技术与应用[J].数字技术与应用,2016(04):115.
- [5] 刘政宇. 大数据分析挖掘技术及其决策应用研究[J].科学技术创新,2019(23):84-85.
- [6] 于海波. 数据挖掘在银行客户关系管理中的应用研究[D].合肥工业大学,2010.
- [7] 阿里研究院. 互联网+ 从 IT 到 DT[M].机械工业出版社,2015.
- [8] 吕倩倩. 基于机器学习的音乐流行趋势预测 [D]. 兰州大学,2017.
- [9] 邢白夕. 情感驱动的音乐数据挖掘及检索[D].浙江大学,2013.
- [10] 任芳,刘硕. 数据挖掘技术在医学信息中的广泛应用[J].中国多媒体与网络教学学报(上旬刊),2019(06):9-10.

(上接第 42 页)

高学生的学习兴趣,激发学生的学习热情。采用 Flash 动画和视频演示把复杂问题简单化,化解了教学难点。实施本文所述教学方法后,学生的学习热情及学习成绩都有了显著的改变,取得了良好的教学效果。

## 4 结束语

与传统教学方法相比,采用信息化的教学方法,学生较好的掌握了知识的用法、拓展了编程的思路、增强了团队合作意识,养成了利用现代信息技术学习的习惯。随着时代的发展,教师要与时俱进,继续积极探索更好的教学模式,完善

教学方法。

#### 参考文献:

- [1] 闫磊. 信息化教学对提高职业院校课堂教学质量的作用探析[J].才智,2018(36):68.
- [2] 赵莹,王冠军,刘佰龙. 翻转课堂在卓越工程师班软件开发类课程中的应用与思考[J].现代计算机,2017.
- [3] 陆晓红,谢志昆. 基于 MOOC 的高校教师信息化教学能力提升的混合模式探究[J].中国教育信息化,2019(10).