

个性化推荐系统

关键技术

GEXINGHUA TUIJIAN XITONG
GUANJIAN JISHU

苏玉召 赵妍 著



郑州大学出版社

个性化推荐系统

关键技术

GEXINGHUA TUIJIAN XITONG
GUANJIAN JISHU

苏玉召 赵妍 著



郑州大学出版社
郑州

图书在版编目(CIP)数据

个性化推荐系统关键技术/苏玉召,赵妍著. —郑州:
郑州大学出版社,2013.9

ISBN 978-7-5645-1506-5

I. ①个… II. ①苏… ②赵… III. ①数据库-情报
检索 IV. ①G354.4

中国版本图书馆 CIP 数据核字 (2013) 第 142188 号

郑州大学出版社出版发行

郑州市大学路 40 号

出版人:王 锋

全国新华书店经销

郑州市诚丰印刷有限公司印制

开本:787 mm×1 092 mm 1/16

印张:10.25

字数:245 千字

版次:2013 年 9 月第 1 版

邮政编码:450052

发行电话:0371-66966070

印次:2013 年 9 月第 1 次印刷

书号:ISBN 978-7-5645-1506-5

定价:35.00 元

本书如有印装质量问题,请向本社调换



内容提要

本书重点阐述个性化推荐系统使用的数据及数据质量,提高数据质量的定制日志数据建模方法,数据挖掘算法和常用的数据挖掘工具。编写本书的目的是:一方面,帮助研究人员和系统开发人员系统地理解个性化推荐系统;另一方面,为解决推荐系统精度和效率难题提供借鉴思路。



前言

个性化推荐系统在研究和应用中不断发展,参与研究的科研人员不断地挖掘新方法,采用新技术,不但在数据挖掘算法上进行智能性改进,而且还应用新的信息技术提高服务的自动化程度。同时,随着大数据时代的到来,研究人员越来越多地关注数据质量问题,通过提高数据质量提高个性化推荐的精度,不但提高了数据中心的数据治理能力,而且提高了数据挖掘算法的效率和精度。个性化推荐系统应用领域越来越广,例如,电子商务、生物医学、数字图书馆、通信、金融、娱乐等。

可以分析,在许多领域和行业需要越来越多的智能化、自动化的个性化推荐系统研究人员和系统开发工程师,高校的研究生也是参与科研和开发的重要力量。因此,个性化推荐系统读者类型多样,包括在校大中专学生、研究生、科研人员和系统开发工程师等。

个性化推荐系统的关键技术主要包括:用户模型、推荐系统使用的推荐算法。其中,与用户模型相关的主要是数据,该部分涉及数据建模、用户建模通过对数据获取、预处理、传输和存储,形成用户模型。推荐系统使用的推荐算法主要包括:推荐系统使用的数据挖掘算法、数据挖掘工具、Web 挖掘等,其中,日志挖掘是推荐系统使用的关键技术。普通的通用日志格式存在缺失重要信息的问题,不利于深度挖掘,而定制日志格式存在格式不一致的问题,不利于数据集成和数据治理。本书提出一种定制日志的理论和方法,目的是通过提高数据质量的方法,提升用户建模的准确性和规格化,同时,提高数据集成和数据治理的效率,也便于深度挖掘潜在的价值。

在个性化推荐系统关键技术的大环境下,从提高推荐系统效率和精度的数据源着手,逐步展开推荐系统实现技术,将实践应用和理论分析从技术层面结合到一起。

本书重在提高推荐系统精度和效率的数据质量,并提出一套定制日志数据的理论和方法。帮助科研人员和系统开发人员从根源解决推荐系统面临的难题。

由于作者水平和知识有限,对于书中不足和错误之处,恳请读者批评指正!

作者

2013 年 4 月



目 录

1	个性化推荐系统概述 /	1
1.1	概念界定 /	2
1.2	相关研究 /	3
1.3	个性化推荐系统实现方法 /	5
2	个性化推荐系统数据模型 /	11
2.1	数据与数据应用 /	12
2.2	数据模型类型 /	16
2.3	数据建模规则 /	16
2.4	数据建模过程 /	19
3	个性化推荐系统用户模型 /	45
4	数据挖掘与个性化推荐系统 /	51
4.1	数据挖掘 /	52
4.2	Web 挖掘 /	83
4.3	个性化推荐系统中的数据挖掘技术 /	83
4.4	一种智能混合过滤个性化推荐模型 /	86
5	个性化推荐系统中的 Web 日志挖掘 /	89
5.1	日志数据 /	90
5.2	日志数据预处理 /	92
5.3	定制 Web 日志数据 /	103
5.4	Web 日志挖掘算法和技术 /	114

5.5 日志数据应用 /	129
6 个性化推荐系统应用 /	145
6.1 推荐系统组成 /	146
6.2 推荐技术的发展和应用 /	147
6.3 发展趋势 /	152
参考文献 /	153

1

个性化推荐系统概述

1.1 概念界定

关于个性化概念,不同的学者和研究人员,在不同的时期对其理解及其所持观点也不尽相同。有些从事 IT 行业的专家认为,个性化是一种能力。例如,前剑桥福雷斯特研究公司的分析师 Paul Hagen 在 1999 年的一次电子商务会议上说:“个性化是一种能力,根据每个用户的爱好和行为知识,对其提供简洁的内容和服务。”长期从事 IT 行业实践和研究的专家 Jill Dyche 女士,在 2002 年由艾迪生-韦斯利出版公司出版的“CRM 指南”认为:“个性化是一种定制用户交流的能力,这种能力的实现是借助于用户进行交流时的爱好和行为知识的获得。”而有些计算机技术专家认为,个性化是一种服务。例如,前 IBM 沃森研究中心的研究人员 Doug Riecken 在 2000 年美国计算机协会(ACM)通信年会上提出:“个性化是关于通过建立一种有意义的一对一关系,从而建立用户的忠诚度。同时,通过理解每一个用户特殊需要的内容,为其提供高效、有价值的服务。”一些从事个性化技术研究的专家学者认为,个性化是一种技术。例如,在 2003 年的 Web 个性化智能技术国际研讨会上,各国的学者和研究人员就个性化概念达成共识,认为“个性化是一种技术,这种技术根据用户的信息定制商业和每个用户电子商务的交互。用户的信息可以是以前获得的,也可以是实时方式获得的。根据可用的用户信息,商业不同部门之间进行变更以适应用户的需求。”还有一些从事信息技术研究的专家和学者认为,个性化是一种过程。例如,Kwon 分别在 2007 年和 2009 年美国信息技术协会(ITAA)年会上分别提出:“个性化是一个过程,在电子服务应用方面,是一个应用用户信息为其发送目标方案的过程。”这些关于个性化的概念重点是从个性化是什么和个性化能够做什么方面定义,但是,这些观点没有系统化、全面性定义个性化。

美国芝加哥德保罗大学的 Bamshad Mobasher 教授从 20 世纪 90 年代就开始个性化的研究,他对个性化的特点、历史、现状和未来的发展趋势具有全面和独到的见解。他对个性化领域研究的兴趣包括:①Web 个性化推荐系统;②个性化推荐系统安全;③个性化 Web 语义挖掘;④Web 使用挖掘;⑤Web 个性化使用挖掘的数据准备;⑥Web 内容挖掘和文本挖掘。Mobasher 对个性化的定义具有重要的参考价值,认为“在 Web 上下文中,个性化意味着动态内容的发送,例如文本元素、链接、广告和产品推荐等,这些内容专门为特定用户或者一部分用户的需要或者兴趣定制”。他把 Web 个性化过程看作是一个包含数据挖掘循环所有阶段的典型应用。这些阶段包括数据收集、预处理模式发现、性能评价和在用户和 Web 网站之间应用实时发现的知识。

1.2 相关研究

Web 个性化智能技术国际研讨会从 2001 年开始连续举办 5 届,并从 2008 年第 6 届开始,该会议改名为 Web 个性化智能技术和推荐系统国际研讨会。美国芝加哥德保罗大学的 Bamshad Mobasher 教授和英国华威大学计算机科学系的 Sarabjot Singh Anand 教授一直担任该会议的联合主席。这些会议的研究内容及他们对个性化的历史、现状和未来发展趋势的观点更具有权威性。

2001 年美国西雅图举办了第 1 届 Web 个性化智能技术(ITWP2001)研讨会,该会议的主要研究内容包括个性化数据建模、系统结构、集成技术和管理技术。其中,个性化数据建模采用的数据主要是用户访问 Web 的行为记录,主要的集成技术采用信息抽取和数据挖掘技术,系统结构采用的技术是在线的 Web 智能化系统。这些个性化研究的早期阶段采用的技术特点是结合了人工智能和 Web 技术,具有现代个性化推荐系统的雏形。例如,Damiani 等人的用户浏览历史建模研究,就是根据用户的历史数据进行个性化数据建模;Mobasher 等人的根据匿名 Web 使用数据提高协作过滤效率的研究,就是根据数据挖掘和 Web 技术实现个性化过滤系统。

2003 年在墨西哥阿卡普尔科举办第 2 届 Web 个性化智能技术(ITWP2003)研讨会,与第 1 届会议相比较而言,本次会议研究的内容除了对个性化的实现技术、系统结构进行深入研究外,突出的特点是 Web 个性化在更多领域进行了应用尝试,并对个性化的推荐系统进行了更深入和复杂的研究。例如,Kritikopoulos 等人研究的 Web 交流的个性化搜索引擎,就是 Web 个性化在搜索引擎领域的应用;Mobasher 等人研究的基于关联规则和序列模式的推荐模型对网站特点的影响,和基于内容结构和用户行为的 Web 个性化集成平台,就是个性化的推荐系统更深入和复杂的研究。

2005 年在苏格兰爱丁堡举办第 3 届 Web 个性化智能技术(ITWP2005)研讨会,本次会议除了深入数据挖掘技术在个性化推荐系统的研究,有一些新的特点值得关注,例如出现了电子学习的个性化、个性化安全方面的研究和个性化评价技术等。其中 Burke 等人的协作过滤系统中欺骗攻击研究,就是个性化安全方面的研究;Kevin 等人的用户对推荐系统的动态评价研究,就是关于个性化评价技术的研究。

2006 年在美国马萨诸塞州波士顿举办第 4 届 Web 个性化智能技术(ITWP2006)研讨会,本次会议上除了继续个性化协作过滤技术更多的应用研究外,开始出现了语义和本体技术在个性化协作过滤方面的研究。例如,Park 等人的协作过滤技术在电影搜索和浏览的应用研究,就是个性化协作过滤技术新的应用研究;Toivonen 等人的本体角色在个性化内容过滤应用需求的研究,Symeonidis 等人的基于潜在语义标引的可扩展协作过滤研究,就是语义和本体技术在个性化协作过滤系统的研究。

2007 年在加拿大温哥华举办第 5 届 Web 个性化智能技术(ITWP2007)研讨会,在以前研究的基础上,本次会议的特点是出现了关于个性化用户建模的数据收集方面的相关

个性化推荐系统关键技术

研究,个性化推荐系统使用收集到的这些数据挖掘用户的偏好;个性化推荐系统用户评价研究的出现,为个性化应用集成技术理论奠定了基础。例如,Evgeny 等人的以呈现为中心偏好的个性化推荐系统评价,Patricia 等人的协作推荐系统的语义使用评价,就是个性化的评价系统研究;Silvana 等人的推荐系统信息源的选择方法学研究,Nicolas 的 Web 用户概要关系聚类算法学习,就是个性化用户建模研究。

2008 年在美国芝加哥举办第 6 届 Web 个性化智能技术和推荐系统(ITWP2008)研讨会,本次研讨会的重点是个性化用户建模理论、推荐系统算法和策略、个性化安全和个性化社会网络应用等研究。其突出特点是个性化研究开始采用社会标签的社会网络等新技术。例如,Gemmis 等人的基于用户概要的个性化搜索信息抽取模型,就是个性化用户建模理论研究;Hung 等人的基于标签用户建模的社会媒体推荐系统,既是个性化用户建模理论研究,又是个性化社会网络应用研究;Sandvig 等人的分析社会标签系统攻击的框架研究,就是个性化推荐系统在安全方面的进一步研究;Kirmemis 等人的基于内容的用户模型生成和电影推荐系统优化方法,既是个性化建模技术理论研究,又是个性化推荐系统应用研究。

2009 年美国帕萨迪纳举办的第 7 届 Web 个性化智能技术和推荐系统(ITWP2009)研讨会的主席,由来自英国华威大学计算机科学系的教授 Sarabjot Singh Anand、美国芝加哥德保罗大学的 Bamshad Mobasher 教授、美国加州大学欧文分校信息与计算机科学学院的 Alfred Kobsa 教授和德国多特蒙德工业大学计算机科学系的 Dietmar Jannach 教授组成。从研讨会的规模和参与人员来看,随着参与 Web 个性化研究的专家和学者人数的增加,Web 个性化研究也越来越受到行业的认可和重视。Web 个性化研究也越来越成熟,从个性化建模策略研究、用户建模和推荐系统算法等集成实现技术,到个性化在多种行业和领域的应用,都有很快的发展。例如,Rosenthal 等人的提高推荐系统在线预测精度,Bohnert 等人的应用高斯空间过程用户建模并预测其对博物馆展览的兴趣,这些就是个性化推荐系统建模预测用户兴趣和偏好的应用研究;Heglich 等人的移动网络中不同推荐系统算法效率问题,Gemmell 等人的标签推荐系统中 K -近邻算法,就是推荐系统不同的算法研究。

Web 个性化研究以 Mobasher 教授为代表的专家和学者,他们的研究成果和研究方向,尤其是 ITWP 国际会议具有权威性和代表性。Mobasher 教授一直作为 ITWP 组委会主席,其研究成果对于未来 Web 个性化研究具有借鉴意义。他本人及其合作者的许多研究成果和文献可以通过他的个人网站获得,他的研究领域与个性化相关的方向包括:① Web 个性化推荐系统;②推荐系统安全;③Web 挖掘和个性化中的语义技术;④Web 使用挖掘;⑤Web 使用挖掘的数据准备;⑥Web 内容挖掘和文本挖掘。

1.3 个性化推荐系统实现方法

1.3.1 个性化过程

个性化组成过程大致可以分为 5 个阶段：用户信息收集-数据预处理-数据挖掘-个性化推荐-用户评价。在个性化推荐系统开发的具体阶段，可以进行功能的细化。同时，个性化在应用过程中，根据用户的评价，调整个性化策略，从而改进个性化质量。根据个性化方案的不同设计思想，可以有多种形式的个性化推荐过程。本书重点讨论基于数据驱动和目标驱动的个性化推荐过程，如图 1.1 所示。

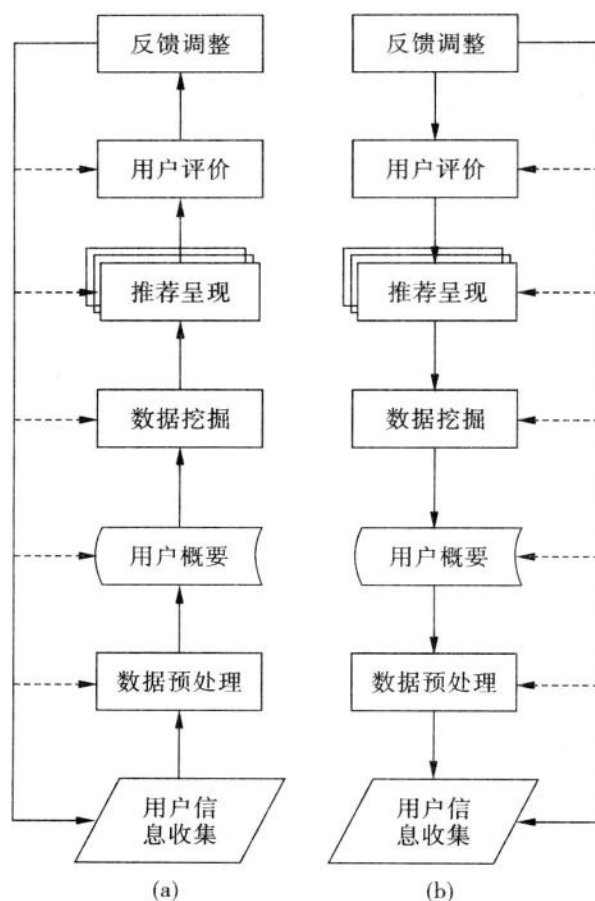


图 1.1 个性化过程

(a) 数据驱动个性化过程；(b) 目标驱动个性化过程

1.3.2 数据驱动的个性化过程

数据驱动的个性化采用“自底向上”的设计思想,如图 1.1(a)所示。

第一步,个性化过程从不同的交互渠道收集用户信息开始,比如,通过 Web、电话、直接发送邮件等。也可以通过各种各样的数据源获取每个用户的尽可能全面的信息。

第二步,一旦用户信息收集完毕,就要对收集的信息进行清理,也就是把一些冗余的、与用户个性化不太相关的数据清除,数据格式规范化。

第三步,在开发用户个性化应用中一个关键问题是,怎样构建一个精确的、容易理解的用户模型。经过数据预处理的数据,识别用户的兴趣爱好生成用户模型,并按照一定的格式存储起来。

第四步,用户模型构建完成后,进行数据挖掘为个性化的每个用户匹配合适的内容和服务。进行数据挖掘的技术包括:推荐系统、基于统计学的预测方法、基于规则的系统、专家指导发送内容和服务的管理规则。

第五步,完成数据挖掘之后,个性化信息发送并呈现给用户。可以采用几种方法发送个性化信息:可视化、相关性有序列表和可选择无序列表等。

第六步,个性化呈现的效果评价可以采用的方法,包括:精度、用户生命期、用户忠诚度和购买的商品。个性化评估结果的好坏,依赖于个性化过程前 5 个阶段采用技术的复杂程度。

最后,个性化系统对用户的评价进行反馈,可以用于调整并改进前 6 个阶段的功能。反馈决定是否需要收集更多的用户信息、在数据预处理时是增加还是减少冗余数据的清理、如何建立更好的用户模型、选择或者开发更好的数据挖掘算法、推荐呈现给用户的信息进行优化,或者采用更多和更复杂的评价方法。如果反馈阶段能够恰当地集成到个性化过程,与每个用户交互的质量逐步得到提高,个性化将会进入“良性循环”。

所谓的“良性循环”,是指在某种环境链中的一个元素的改进,导致了另外一个元素的改进,进而使得原始的元素进一步改进,如此循环,形成环境链中元素不断改进的一种情况。

1.3.3 目标驱动的个性化过程

目标驱动的个性化过程采用“自顶向下”的设计方法,之所以称为“目标驱动”,这是因为预先设定目标,然后再设计个性化的每个阶段。而“数据驱动”是先收集数据,然后再建立用户模型。这和商业管理的格言一致:我们无法管理不能预测的事情。换言之,我们设计个性化首先要确定目标,然后开始下一阶段工作的实施。目标驱动与数据驱动个性化过程的阶段类似,但是其反过程。

第一步,设定个性化评价指标,明确哪些因素会对个性化服务质量起到关键性作用。

个性化系统可以采用一种或者多种评价指标,例如,用户满意度、用户生命期、用户忠诚度和购买的商品。从这些评价指标中可以分析用户的兴趣爱好。

第二步,设定了个性化评价指标后,围绕这些指标选择哪些内容、以何种方式推荐呈现给用户。可以采用几种方法呈现个性化信息,包括:可视化、相关性有序列表和选择性无序列表等。用户兴趣爱好也体现在推荐呈现内容和方式上。

第三步,根据推荐呈现内容,选择合适的数据挖掘算法。与数据驱动个性化过程类似,可以采用的数据挖掘技术包括:推荐系统、基于统计学的预测方法、基于规则的系统、专家指导发送内容和服务的管理规则等。

第四步,由用户模型提供进行数据挖掘需要的数据。

第五步,数据预处理的结果用于生成用户模型。

第六步,用户信息收集。这一步与数据驱动个性化过程的最大不同,就是根据预先设定的目标,逐步向下推移,直到最后发现需要收集哪些数据。这样做的优点是避免做无用功,做到“有的放矢”。

最后,在个性化应用完成后,根据用户评价结果,把反馈调整策略集成到个性化的前6个阶段。与数据驱动个性化过程不同的是,目标驱动是根据需要预先设定各个阶段需要的技术,然后再实施个性化的每个阶段,最后根据用户评价结果,把反馈调整策略集成到个性化各个阶段。数据驱动采用的是“推演技术”,从收集用户信息开始,一直到用户评价,进而反馈调整个性化的每个阶段的策略。

所谓的推演,即推论演绎、推移演变,见汉·陆贾《新语·明诫》:“观天之化,推演万事之类”;《三国志·蜀志·诸葛亮传》:“推演兵法,作八阵图,咸得其要云”。

1.3.4 二者比较

目标驱动的个性化过程从指定用于决定个性化影响的用户评价开始,选择的评价指标应该决定将要推荐呈现给用户的个性化内容属于哪种类型。推荐呈现给用户的个性化内容,又决定了采用的数据挖掘技术如何挖掘来自用户模型的数据。用户模型的生成,决定了用户信息的预处理采取的技术,哪些数据是冗余的、哪些数据是必需的。最后,建立全面的用户模型决定了需要收集哪些用户信息,如何收集需要的信息。很显然,在这个过程中,有一些阶段采用的技术针对性很强,需要额外处理的步骤会减少很多。例如,在数据预处理阶段,由于采用的数据已经限定范围,只要把生成用户模型的数据格式化就基本完成,所以,预处理步骤会简单得多。

数据驱动的个性化过程从用户信息收集开始个性化过程,由于无法确定哪些信息是必需的,所以,尽可能多的收集。如果收集太多的信息,会造成两点麻烦:一是加重服务器负担,例如收集信息时间过长影响系统性能;二是收集太多的冗余数据会给数据预处理阶段带来处理困难。用户模型生成后,经过数据挖掘技术处理,推荐呈现给用户的内容可能不是用户感兴趣和爱好的项目,这会使得个性化系统进行较大的调整,可能需要重新从策略上调整个性化过程每个阶段的技术,如此重复,不断调整。因此,数据驱动的

个性化过程比目标驱动的个性化过程复杂许多。

由于“目标驱动”方法,从一开始就预定目标,所以与“数据驱动”方法相比较而言,其优点是能够让个性化实现“良性循环”,从而为投资者带来更多的收益。然而,目标驱动方法在个性化文献研究中从没有出现过,我们的推测有待在未来的研究中证实。此外,其他影响个性化过程设计的因素还包括可信度和个性化过程中潜在的干扰等。这些情况也是有待于个性化过程中进行研究。

目标驱动的个性化过程能够更好地实现“良性循环”,其优点为:一方面,如果商业网站个性化实现了良性循环,将会为商家带来利润的不断增长,例如卓越亚马逊购物网的个性化;如果公共服务的网站个性化也实现了良性循环,将会为公众提供更好、更及时、快捷的便民服务,例如英国国家卫生健康网站的个性化;如果搜索引擎网站个性化实现了良性循环,将会为用户提供快捷和功能多样化的专业搜索服务。另一方面,良性循环对于个性化系统适应不断变化的环境、提高个性化服务至关重要。例如,用户的兴趣爱好发生改变时,个性化对其供应的商品和方式也要进行调整。

良性循环的对立面是去个性化过程。从开始的时候,用户就对供应的结果不满意便会发生去个性化情况。如果这种情况一直发展下去,或者系统没能及时做出策略调整,用户就不情愿继续使用个性化系统,直至不再使用它。去个性化很大程度上影响着工程是否有前途及其成败。因此,个性化开发者面临的一个巨大挑战就是如何实现良性循环,而不陷入去个性化漩涡。

1.3.5 目标驱动的个性化集成

目前为止,大部分个性化只是实现了前面5个阶段的功能,复杂一些的系统微弱地实现了个性化用户评价,但是还没有很好地实现反馈循环。因此,当前个性化亟待解决的问题是如何研究好的评价方法、个性化策略的调整方法和合适的反馈循环。实现个性化良性循环依赖于两个关键步骤:①个性化过程中6个阶段的每一个阶段都选择最好的技术;②个性化过程的每一个不同阶段集成时要有充分的依据原则。本节我们讨论通过目标驱动设计原理,分析个性化过程的几个关键技术及其特点。

1.3.5.1 评价个性化

评价个性化最常用的方法是采用精度来度量,也就是说,精度和推荐的相关性能反映出喜欢或者不喜欢为其推荐商品的程度。但是基于精度的度量方法不能反映出更复杂、微妙的个性化,所以,一些研究者鼓吹采用一些更有效的、全面的个性化度量方法。例如,用户生命期、忠诚度、购买记录、消费记录和用户的反馈等。但是,如何在个性化应用中采用这些方法还有待于进行更深入的研究。

1.3.5.2 推荐呈现

个性化系统提供不同的呈现方式,由用户根据其各自的职业、习惯和年龄等特点进行选择。电子商务系统为用户推荐呈现个性化信息的方式通常有几种,包括描述、相关性列表、可选集合和可视化等类型。一种发送方法就是著名的“拉、推、送”方法。“拉”是通知用户有一些可用的个性化信息,但是没有显示出来,需要用户主动发送请求。“推”是把信息发送给没有使用个性化的用户。例如,发送电子邮件给访问电子商务网站的用户。“送”是把别的用户相似的行为给当前用户显示出个性化信息,例如,一个用户访问电子商务网站时会显示其相关的商品信息。

1.3.5.3 数据挖掘

数据挖掘技术是从数据源发现不明显的、潜在有用的和以前未知的信息,通过数据挖掘为用户生成个性化推荐。根据推荐方法的不同,分为基于内容的推荐、协作推荐和混合推荐等。

(1)基于内容的推荐 系统为用户推荐的是用户过去爱好的项目,例如内容、服务和商品。这种基于内容的推荐,分析用户对某些点击率高的项目的共性,只有与用户过去喜欢的项目类似时才被推荐给用户。

(2)协作推荐(或者称协作过滤) 系统推荐的项目与用户的兴趣和爱好相似的时候,才被推荐给用户。只有与每个用户的兴趣和爱好最相近时才被推荐。

(3)混合推荐 综合基于内容和协作的推荐可以形成几种方法。一种方法是分离基于内容和协作过滤的方法,单独建立系统,然后把两个系统产生的结果综合起来,形成最终的推荐结果。另外一种方法是,把基于内容和协作的技术综合到一个独立的推荐系统中产生推荐结果,而不是把两种技术建立的系统分离开来。

个性化数据挖掘方法大部分可以看作是协作过滤的扩展。根据网站上用户点击历史和过去浏览记录的信息作为数据挖掘的输入数据,根据用户兴趣爱好生成聚类用户模型。这种用户模型可以用来连接活动用户的资料,预测未来用户行为并生成推荐。

1.3.5.4 生成用户模型

要生产全面和有价值的用户模型,用户信息的收集可以通过多种渠道实现,例如,Web、电话和邮件等。可以通过调查的方式显示收集,也可以通过用户购物历史和搜索信息进行隐式收集,还可以通过人口统计学原理和心理行为学收集用户信息。这些信息收集完成后,可以进行处理、清洗,并存储到数据仓库中以便进一步分析使用。

用户信息用于建立用户模型。用户模型包含的信息应该包括:用户是谁,与个性化相关的用户行为。大部分个性化系统表示的用户模型包含收集到的事实数据,例如用户名、性别、出生日期和地址等人口统计学信息;也包括从过去交易中推导出来的事实数

个性化推荐系统关键技术

据,例如,在电子商务网站上购买喜爱的商品类别或者购买数量等。这些信息通常被定义为值,并存储在关系数据库中,一个用户对应一条记录。

以上分析的关键技术,在个性化集成过程中每一阶段所用技术的性能优劣,都会影响到系统整体性能。因此,采用目标驱动的个性化过程能够从系统整体结构分析,不同阶段采用最优技术能够减少系统优化时间,吸引用户使用个性化能够获得更好的效益。但是,如果系统在设计阶段没有分析好的话,在个性化推荐呈现时无法很好地满足用户需要。