

# 基于知识地图的知识推送方法研究

渠国庆,熊 峰,牛 倩,吴祖伟,吕北轩

(上海大学 机电工程与自动化学院 上海市智能制造及机器人重点实验室,上海 200072)

**摘 要:**对于知识密集型的高新企业来说,有效利用丰富且复杂知识的方法与途径,一直是企业发展的一个重要环节。随着推荐系统的出现与发展,催生了诸多经典推荐算法,并在电子商务和社交网络中得到了成功应用,但现有经典推荐算法并不适用于更加专业化的企业知识推送。为了提高企业知识与企业员工的匹配程度,使企业员工能更高效地利用企业知识,提出了一种改进的协同过滤推荐算法。该算法基于知识地图的知识推送方法,由知识地图的关联度得到更具体的关联关系,并综合考虑路径关联权值对推荐算法的影响,改进了相似度构成与计算方法,以相似度衡量用户之间的相似性,加强了用户之间的相似度的评估。理论分析表明,所提出的算法提高了知识推送的匹配程度和可行性,为企业知识推送提供了新的思路和途径。

**关键词:**企业知识管理;知识地图;知识推送;协同过滤

**中图分类号:**TP301

**文献标识码:**A

**文章编号:**1673-629X(2017)09-0082-03

doi:10.3969/j.issn.1673-629X.2017.09.018

## Research on Knowledge Push Method with Knowledge Map

QU Guo-qing, XIONG Feng, NIU Qian, WU Zu-wei, LYU Bei-xuan

(Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China)

**Abstract:** For knowledge-intensive and high-tech enterprises, the effective use of method with rich and complex knowledge has become an important part in development of enterprises. With the emergence and development of recommended system, a lot of classical recommendation algorithms have been produced and applied in e-commerce and social networks successfully, but all of the existed algorithms are not practically suited to more professional knowledge push for business enterprises. In order to increase matching degree between business knowledge and business employees and be efficient use of enterprise knowledge for employees, an improved collaborative filtering algorithm has been proposed. Based on an knowledge push method of knowledge map, it obtains more specific relationship according to the relativity of knowledge map and improves the similarity construction and computation considering the impact of path associated weights on the recommendation algorithm. The similarity is utilized to measure the similarity between users, enhancement of similarity evaluation between users. Theoretical analysis shows that the matching degree and the feasibility of knowledge push has been increased by modified algorithm, which has provided a new way for future business knowledge push system in enterprises.

**Key words:** enterprise knowledge management; knowledge map; knowledge push; collaborative filtering

## 0 引 言

近年来,随着互联网技术的不断发展和数据信息的广泛传播,进入了“信息爆炸”时代,虽然人们可以方便地获取更多的数据信息,但是也被大量无关信息所淹没。信息呈现的多元化发展以及企业信息化的不断发展,使得企业需要及拥有的知识不断增加,而网络的快速发展导致信息与知识加速流通。企业所面临的竞争,已从原始资本的竞争、速度的竞争,转变为知识

的竞争。知识已成为在取得持久性竞争优势的过程中必不可少的重要元素。另一方面,在知识推送方面,大多知识管理系统还是采取基于类似于 RSS 的被动推送方式。但是 RSS 的订阅粒度太大,不能有效地进行跨领域订阅。更重要的是,被动推送方式需要用户对自己的喜好和所需信息与知识及其所属领域有一个清晰的了解。这些在很大程度上给用户带来了诸多不便。近年来,由于推荐系统中信息过滤的特殊属性,推

收稿日期:2016-05-19

修回日期:2016-08-25

网络出版时间:2017-07-05

基金项目:上海市科技计划项目(12DZ1505600)

作者简介:渠国庆(1990-),男,硕士研究生,研究方向为企业知识管理、推送;熊 伟,副研究员,研究方向为企业信息化。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20170705.1649.012.html>

荐系统已应用于任何规模的电子商务网站中,为这些网站提供了海量的推荐结果。推荐系统可以帮助企业决定向哪个客户出价,实现一对一的营销战略目的。推荐系统可以说是当前大型商业网站的必备模块,如亚马逊、淘宝网和京东商城等,都建设了性能优良的推荐系统,通过成功地应用推荐系统,亚马逊成功将其销售额提高了三成<sup>[1]</sup>。但是,由于企业知识推荐的特殊性,原有的推荐方法不能完全适用于企业内部,造成了企业知识孤岛问题严重。而且,现有知识推荐系统中,主要采用两种方式。一种是被动的知识推送方法,通过用户对一类知识的订阅来获取知识,但是这种方法的主要问题是需要用户对自己需要的知识是什么非常清楚。同时这种方法很难达到细粒度的知识获取,往往使用户得到许多不必要的知识。所以这种方法的使用越来越少。另外一种推荐系统是,当今推荐系统的主流算法是协同过滤推荐算法,通过分析知识与用户、知识与知识、用户与用户的相互关联而得到一个合理的推荐结果。然而在实际情况中,往往由于知识与知识、用户与用户之间的关系不便获得与鉴定,使得现在的主要算法多是基于知识与用户间的关联来得到,从而大大降低了算法的有效性。

为此,借鉴电子商务中推荐系统的发展思路,提出了基于知识地图的知识推送方法,加强了用户与用户间的相似性度量,并改进了现有的协同过滤算法,利用知识地图展现的知识节点间的权重关系,通过知识源和知识点以及知识源和知识源间的图论路径权重优化传统的协同过滤算法。

1 相关综述

1.1 知识地图发展

知识地图作为一种导航系统,可以显示不同知识存储之间的静态结构和动态联系。它是知识管理过程中的输出层,将整合后的知识内容输出,实现知识关联和知识的汇聚。构建知识地图的过程需要依赖一定的信息和准则,很多研究者根据不同的信息构建了知识地图。Chung W 等<sup>[1]</sup>利用信息间的链接构建了知识地图;Yoon B 等<sup>[2]</sup>利用信息间的引用关系构建了知识地图;Gordon J L 等<sup>[3]</sup>利用信息间的依赖关系(学习依赖)构建了知识地图。

然而以上方法并不能完全适用于企业知识地图构建,一方面企业知识间存在链接关系、引用关系和依赖关系,不同的部门、员工间所需要的知识分类层次区别很大。另一方面,这些方法并没有考虑在企业知识过程中的知识分层问题,在不同层次和不同节点的知识推送权重会有所不同。综上所述,现有知识地图虽然取得了很大发展,但针对企业的知识地图并不

能有效表达知识节点间的权重关系。

1.2 推荐算法发展

推荐算法是个性化推荐算法的核心。现有的推荐算法包括基于内容的推荐算法<sup>[4]</sup>、基于项目的协同过滤<sup>[5]</sup>、基于用户的协同推荐<sup>[6]</sup>、基于模型的协同过滤<sup>[7]</sup>以及最近兴起的基于网络结构的推荐算法。王有远等<sup>[8]</sup>提出了基于集对分析的知识多维度筛选模型;蒋翠清等<sup>[9]</sup>根据设计过程中的知识需求模型,建立了一种面向产品设计知识的协同推送模型。但是上述研究大都限于产品设计本身,很少涉及企业内部知识协同的需求,同时,在对设计人员进行知识推送时,并没有考虑用户对某一知识的匹配程度权重。综上,现有协同过滤算法在互联网电子商务和社交等大数据量应用中效果比较明显,但在专业性更强的企业知识推送过程当中准确度不高。

2 知识地图模型的构建

2.1 分层立体知识地图

依托金字塔式知识管理体制,将庞大业务和技术状态知识分别纵向梳理到专业层、项目层、子项目层、知识节点层,另外在分层系统中关联用户和知识,并为用户贴标签。在分层立体知识地图的框架<sup>[10]</sup>基础上,根据各层知识的构成特点,构建横向的单层知识地图,并在每一层的知识节点上定位用户,以进一步优化知识互联的关联关系,特指用户即为知识源。

知识节点和知识源节点以及它们间的关联关系构成了知识地图架构(见图1)。知识节点是知识的具体存储,知识源指企业不同的用户。在知识地图中,权重体现了节点间的关联关系,权重越大关联越紧密,反之说明两者的相关性不大,这种关联具体为知识节点间的关联和用户节点间的关联。从企业内部特征来说,知识节点与知识节点间的关联组成了企业知识网络,而用户间的关联则组成了企业内部的人际关系网络,体现的是用户间合作的紧密程度,知识网络和人际关系网络以及两者之间的关联构成了知识地图。

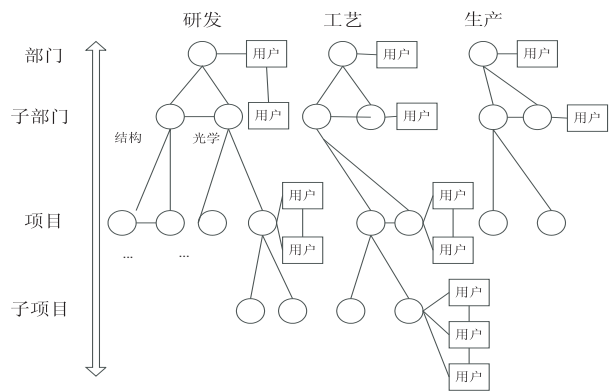


图1 企业知识地图模型

## 2.2 路径关联权重

根据知识地图模型很容易量化知识点和知识源之间的路径关联权重<sup>[11]</sup>, 节点语义相似度由聚类(Cluster)相似度和差异性(Specificity)决定, 聚类相似度基于知识地图建模中的知识点和知识源以及知识源之间的关联路径。由于推荐知识的匹配度和知识源所在的领域和范围相关, 知识地图把不同的部门, 不同的项目和不同的岗位按节点划分, 由此来区分知识源及用户。根据知识地图中的网络模型以及节点间的权重, 可以计算任意两个节点间的关联权重, 这种关联权重能够体现知识节点和用户间的关联程度, 可以得出知识点  $P$  和知识源  $I$  关联权重计算公式:

$$W(P, I) = \sum_{p, i} 1 \quad (1)$$

节点间并不都是两两相连的。绝大多数不相连的节点间都由一条或多条路径连接。在这种情况下, 对不相连的两个节点间的关联权重就由它们间的最短路径决定, 即关联路径。式(1)知识地图定义两相邻节点的关联路径为 1, 则  $W(P, I)$  是最短路径的相邻点权值的和, 得出知识源和知识点之间的关联权重:

$$w(P, I) = \frac{1}{\sqrt{W(P, I)}} \quad (2)$$

同理可得知识源和知识源间的关联权重, 其作用是进一步增加用户间的相似度, 计算方法是根据分层知识地图模型中用户个体所在的节点直接计算该路径的长度, 同式(2):

$$w(I, J) = \frac{1}{\sqrt{W(I, J)}} \quad (3)$$

## 3 基于知识地图的知识推送方法

### 3.1 协同过滤算法

协同过滤算法是目前较常用的推荐算法之一, 是通过采集其他用户对某个内容或知识的评价来获得相似用户对同知识的评价<sup>[12]</sup>。假设  $I_i$  是某用户  $i$  评价过的所有内容或者知识的集合, 可以得出该用户对内容或者知识的平均评价计算方法:

$$r_i = \frac{1}{|I_i|} \sum_{j \in I_i} r_{ij} \quad (4)$$

根据知识地图中节点间的关联权重, 改进协同过滤算法, 重新计算该用户的评价。计算两个用户间的相似度的方法有很多, 最常用的是根据两个用户评价过的内容或者知识之间的相似度来计算, 可以利用 Pearson 算法来衡量这种相似度, 公式如下:

$$\text{Sim}(i, j) = \frac{\sum_{c \in C} (r_{i,c} - \bar{r}_i)(r_{j,c} - \bar{r}_j)}{\sqrt{\sum_{c \in C} (r_{i,c} - \bar{r}_i)^2} \sqrt{\sum_{c \in C} (r_{j,c} - \bar{r}_j)^2}} \quad (5)$$

万方数据

其中,  $\text{sim}(i, j)$  为用户  $i$  和  $j$  间的相似度;  $\bar{r}_i$  是用户  $i$  对所有知识的期望评价。

### 3.2 基于知识地图权重关系的算法改进

(1) 基于知识地图模型中的网络关系, 根据节点间的权重, 加强不同用户间的相似程度, 改进协同过滤算法。利用知识地图中知识源间的关系加强用户间的相似度。根据式(5)可得到两个用户之间的相似度, 是通过两个用户对其公共评价过的知识的评价相似度计算得到。但是通过知识地图的关联权重, 可以计算不同节点之间的关联程度。它是基于这样一个假设, 关联权重高的用户之间对同一个知识的评价总是相似的, 处在同一个部门的用户关联权重越高, 他们需要的知识内容重合度也就越高。通过这样的假设<sup>[13-14]</sup>, 对相似度计算公式进行改进: 将知识源相关信息从知识地图中独立出来; 按照知识地图的权值计算方法计算这个知识地图的连接权重; 根据式(5)得到新的相似度计算公式:

$$\text{Sim}(i, j) = \frac{\sum_{c \in C} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in C} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in C} (R_{j,c} - \bar{R}_j)^2}} * w(i, j) \quad (6)$$

其中,  $R_{i,c}$  和  $R_{j,c}$  分别是加权路径后的评分, 计算公式为:  $R_{i,p,c} = w(P, I) * r_{i,p,c}$ ;  $\bar{R}_i$ 、 $\bar{R}_j$  是路径加权后的平均评分;  $w(i, j) = w(I, J)$  是知识源之间的路径关联权重。

(2) 通过知识地图中的人际关系网络缩小预测的用户基数, 根据式(5), 与待预测用户相关程度不大的所有用户都被作为评价, 这些相关程度不大甚至不相关的用户很明显会降低预测的准确度。结合  $K$  近邻算法的思路, 基于建立企业知识地图网络及节点之间的权重关系, 采用如下方案:

步骤 1: 根据式(5)计算相似度并对相似度进行排序, 选取相似度最高的  $N * K$  个用户作为预测的集合。

步骤 2: 根据式(6)计算相似度并排序, 并根据步骤 1 计算的  $N * K$  个用户预测集合, 选取相似度最高的  $K$  个用户作为预测集合。

步骤 3: 求出相似度最高的  $K$  邻居, 基于此向目标用户推荐知识。

## 4 结束语

针对现有的推荐算法不能有效解决企业知识精确推送的问题, 在协同过滤推荐算法基础上, 提出了一种

348–354.

[3] Chekam T T,Zhai E,Li Z,et al. On the synchronization bottleneck of OpenStack swift-like cloud storage systems [C]//IEEE international conference on computer communications. San Francisco:IEEE,2016:135–144.

[4] Kamath A,Jaiswal A,Dive K. From idea to reality:Google file system[J]. International Journal of Computer Applications, 2014,103(9):8–10.

[5] Wu C H,Hsu P. Cost-effective and reliable cloud storage for big data [C]//ASE big data & social informatics 2015. New York:ACM,2015.

[6] Wei Y,Yong W F. A cost-effective and reliable cloud storage [C]//2014 IEEE 7th international conference on cloud computing. Anchorage:IEEE,2014:938–939.

[7] Faragardi H R,Shojaee R,Tabani H,et al. An analytical model to evaluate reliability of cloud computing systems in the presence of QoS requirements [C]//International conference on computer and information science. Niigata,Japan:[s. n.], 2013:315–321.

[8] Joolahluk J,Wen Y F. Reliable and available data replication planning for cloud storage [C]//IEEE international conference on advanced information networking and applications. Barcelona,Spain:IEEE,2013:772–779.

[9] 赵 畅. 云存储系统可靠性分析[D]. 天津:南开大学, 2014.

[10] Song J,Deng J H. NOVA:a P2P-cloud VoD system for IPTV with collaborative pre-deployment module based on recommendation scheme [C]//International conference on materials science and information technology. Nanjing,China:[s. n.], 2013:1566–1570.

[11] Chakareski J. Cost and profit driven cloud-P2P interaction [J]. Peer-to-Peer Networking and Applications,2015,8(2): 244–259.

[12] Babalou O,Marzolla M,Trambrini M. Design and implementation of a P2P cloud system [C]//Proceedings of the ACM symposium on applied computing. Trento,Italy:ACM,2012: 412–417.

[13] 李小兵,许胤龙,林一施,等. X 再生码:一类适用于云存储的准确修复编码[J]. 计算机应用与软件,2014,31(8):241–244.

[14] 王 帅,常朝稳,王禹同,等. 面向多云盘的 N+i 模式的编码冗余存储机制[J]. 小型微型计算机系统,2016,37(2): 236–240.

[15] 吴礼乐. 基于双活容灾存储技术的云计算数据中心的设计及应用[J]. 电子设计工程,2015,23(6):190–192.

[16] 马玮骏,王 强,何晓晖,等. 一种基于策略的云存储系统故障管理方法[J]. 软件工程,2016,19(6):4–7.

(上接第 84 页)

结合路径权重,基于知识地图技术的企业知识推荐算法。即根据知识地图模型利用知识地图节点间的权重关系,得到相似度更高的用户预测集合,并以此来向目标用户推荐知识。理论分析表明,改进算法有效解决了企业知识推荐匹配程度低的问题,可以准确获取用户之间的关联度,提高推荐算法的质量和个性化程度。但该算法尚未考虑项目间的语义关系和用户间的语义关系,一定程度上影响了相似度计算和推荐效果,需要今后进一步研究解决。

参考文献:

[1] 朱文奇. 推荐系统用户相似度计算方法研究[D]. 重庆:重庆大学,2014.

[2] Chung W,Chen H,Nunamaker J F. A visual framework for knowledge discovery on the web:an empirical study of business intelligence exploration[J]. Journal of Management Information Systems,2005,21(4):57–84.

[3] Yoon B,Lee S,Lee G. Development and application of a keyword based knowledge map for effective R&D planning[J]. Scientometrics,2010,85(3):803–820.

[4] Gordon J L. Creating knowledge maps by exploiting dependent relationships[J]. Knowledge-Based Systems,2000,13(2):71–79.

[5] Selvi R T,Prakash-Raj G D. Information retrieval models:a survey[J]. International Journal of Research and Reviews in Information Sciences,2012,2(3):227–233.

[6] 刘 刚. 面向领域的软件需求一致性验证方法研究[D]. 哈尔滨:哈尔滨工程大学,2008.

[7] 程 飞,贾彩燕. 一种基于用户相似性的协同过滤推荐算法[J]. 计算机工程与科学,2013,35(5):161–165.

[8] 王有远,赵 璐. 面向产品设计的多维度知识推送研究[J]. 制造业自动化,2015,37(14):131–133.

[9] 蒋翠清,李斌生,高家飞,等. 面向协同的产品设计知识推送研究[J]. 中国机械工程,2012,23(16):1972–1977.

[10] 王晓堤,桑 婧. 基于云模型的时间修正协同过滤推荐算法[J]. 计算机工程与科学,2012,34(12):160–163.

[11] 张 艳,梁欣欣,张耐民,等. 基于知识地图的航天知识推送方法研究[J]. 航天工业管理,2015(5):35–37.

[12] 唐积益,黄树成. 优化相似度计算在推荐系统中的应用[J]. 电子设计工程,2015,23(23):46–48.

[13] 刘青文. 基于协同过滤的推荐算法研究[D]. 合肥:中国科学技术大学,2013.

[14] 熊 奇. 基于知识地图的知识检索与推荐方法研究[D]. 上海:上海交通大学,2009.