

面向科技服务的专家推荐模型研究

盖素丽^{1,2}, 檀改芳^{1,2}, 唐松^{1,2}, 董佳^{1,2}

(1. 河北省科学院应用数学研究所, 河北 石家庄 050081;

2. 河北省信息安全认证工程技术研究中心, 河北 石家庄 050081)

摘要:社会公众在日常生活中经常会遇到各种难题,需要专家提供专业性的咨询服务,由此催生出了专家推荐系统,使得专家及其专业知识得以广泛而深入地介入公众生产生活当中。但是,随着社会数字化发展,科技资源更加繁杂,给专家推荐工作带来一定的挑战。当前专家推荐系统并没有深入挖掘用户需求与专家研究兴趣之间潜在的知识关联,无法针对具体需求推荐出合适的服务专家。知识利用不当成为知识匮乏之后大数据环境下专家推荐的又一障碍。就如何为用户决策咨询需求推荐合适的服务专家进行系统研究,最终提出了一种科技咨询服务的新模式。通过构建需求、专家和案例的知识网络,挖掘需求与专家间潜在的知识关联,提出了一种面向科技领域用户决策咨询需求服务的专家推荐模型。

关键词:专家推荐;知识网络;隐性知识;关系抽取;科技服务

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2022)10-0021-05

doi:10.3969/j.issn.1673-629X.2022.10.004

Expert Recommendation Model for Technology Services

GE Su-li^{1,2}, TAN Gai-fang^{1,2}, TANG Song^{1,2}, DONG Jia^{1,2}

(1. Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang 050081, China;

2. Hebei Authentication Technology Engineering Research Center, Shijiazhuang 050081, China)

Abstract: Due to the strong demand for professional knowledge, public decision-making needs more and more experts to provide professional services. Thence experts and their expertise could be involved in the decision-making process widely and deeply. However, the current recommendation system could not deeply explore the tacit knowledge associations between the needs of users and the interests of experts, and could hardly recommend suitable experts for a specific need. The lack of the solution to access the proper knowledge instead of the lack of knowledge became another obstacle to expert recommendation, which could be solved by knowledge recommendation. Historical servicing cases were combined to conduct a systematic study on how to recommend proper experts to meet the needs of the public for decision-making consultation. Finally, a new recommendation method was proposed. The method proposed a solution based on the concept of "relationship", which was the knowledge web targeting the tacit knowledge connection. Based on the knowledge web, an expert recommendation model was proposed to meet the needs of technology services.

Key words: expert recommendation; semantic web; tacit knowledge; relation extraction; technology services

0 引言

关于专家推荐系统, Reichling T等^[1]曾提出过一个比较明确的定义,即专家推荐系统是为了满足用户在特定场景下对专家的咨询需要,帮助他们推荐合适专家来解决问题。专家推荐系统一般应用于专业性较强的领域,如科研项目申报^[2]、学术论文评审^[3]、企业需求对接^[4]等应用场景。这些领域大多拥有自己组织内部规范化、结构化的知识库,进行专家推荐时,除了通过人工筛选和关键词查询外,可以借助基本的计算

方法如模糊逻辑^[5]和向量空间模型^[6]来计算用户需求与专家领域的相关性,进而对专家进行排序并完成推荐。

但是,当专家推荐系统由专业机构转向社会公众,用户不再是规范的组织机构,那些推荐算法所依赖的高质量且层次分明的知识库也变成了散乱的知识碎片,它们涉及的学科更加丰富,彼此交叉、渗透,界限模糊。面对科技领域的海量数据资源,用户和工作人员难以快速精准找到满足需求的专家,便会陷入知识困

收稿日期:2021-12-07

修回日期:2022-04-08

基金项目:河北省中央引导地方科技发展资金项目(216Z0305G);石家庄市科学技术研究与发展计划项目(201130251A)

作者简介:盖素丽(1977-),女,硕士,副研究员,CCF会员(63420M),研究方向为计算机应用、大数据分析、深度学习;通讯作者:檀改芳(1994-),女,助理工程师,CCF会员(D9906M),研究方向为计算机应用、深度学习。

境。此时,任凭专家有多么强的才智能力、多么丰富的知识经验,如果不能满足用户需求,会导致极大的资源浪费。科技领域迫切需要采用一种行之有效的方法去处理日益繁重的专家推荐工作,实现良善的专家推荐目标。

在系统梳理河北省关于专家和用户决策咨询数据关系和程序研究的基础上,该文提出了契合社会公众需求的专家推荐模型,旨在挖掘符合专家研究兴趣的用户需求,并将用户需求有效推荐给可能解答的专家。模型推动专家与用户之间有效开展决策咨询工作,保证推荐专家的专业性以提升用户决策咨询质量,并做出如下工作:

- (1) 融合多个数据源构建科技服务知识库,对以往案例中需求与专家知识进行有效提取;
- (2) 提出知识融合模型,将需求与专家进行统一建模,并挖掘二者之间的潜在知识关联;
- (3) 提出一种基于 ALBERT 的专家推荐模型。

1 专家推荐算法

近年来,针对推荐算法^[7]的研究发展迅速,其中应用最为广泛的主要有以下几种:

(1) 基于协同过滤的推荐。

作为应用最早和最为成功的推荐技术之一,协同过滤算法面向用户^[8],通过计算用户偏好数据来预测新用户可能喜欢的主题或产品。Javier 等^[9]在此基础上,融合推荐和文本信息检索任务之间的联系,研究出了社交网络中有效联系人推荐算法的新视角。

(2) 基于内容的推荐。

作为信息过滤技术的延续,基于内容的推荐算法一般通过向量空间模型将文档等文本信息表示为多维空间中的向量。随着用户行为产生的标签、评论等文本信息剧增,使得基于内容的推荐显得尤为重要。Wang Donghui 等^[10]提出的基于卡方特征选择和 softmax 回归的有效混合模型能够根据手稿摘要内容的优先顺序推荐合适的投稿期刊或会议。

(3) 基于效用的推荐。

效用理论认为,决定物品价值的是该物品对人的有用程度,或是人对该物品的价值认可程度,亦即该物品的效用。基于效用的推荐方法建立在用户对物品的满意情况,并依靠对象自身属性和用户偏好来衡量其效用值。将效用函数应用在专家推荐系统^[11],可以把用户需求属性和专家兴趣同时进行量化,比依托单一的属性信息进行推荐的方法更加灵活。

(4) 其他推荐。

在某一具体问题中为了获得更好的推荐效果,常常需要将两种甚至以上推荐方法组合在一起,即为组

合推荐。此外,推荐算法与深度学习^[12]尤其是迁移学习的结合,能够轻松处理稀疏数据、捕捉数据间的隐藏关系。跨域推荐^[13]则利用迁移学习的方法,将知识从源域(相关领域)迁移到目标域(研究领域),获取用户在不同领域的有效信息,提高目标域的推荐性能。张甯等^[14]提出了一种基于语义扩展的知识推荐方法用于文献推荐。刘鲁等^[15]利用维基百科知识作为背景构建专家知识地图,能够直观度量专家知识组成和研究兴趣。

由此可以看出,现有学术资源推荐模型^[16]的研究目标是为专家个性化推荐感兴趣的文献或者类似专家。这些推荐模型或研究了文献之间的联系,或研究了专家之间的关系,不能充分捕捉需求、专家与方案三者之间的潜在知识关联(见图 1),导致专家推荐的准确性较低。为了解决这个问题,在综合考虑上述推荐方法的基础上,该文重点关注用户需求与专家之间的知识联系,基于预训练语言模型^[17-18]对需求描述、专家描述进行特征提取与知识建模,提出了一种深度融合辅助知识的推荐算法。该算法构建知识网络,对需求、专家和案例进行统一建模,并挖掘需求与专家之间的潜在知识关联,最后通过模型对需求进行专家推荐。

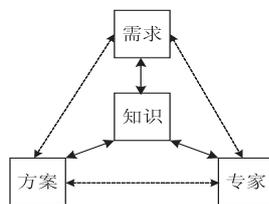


图 1 需求、专家与方案间潜在知识关联

2 专家推荐模型定义

专家推荐模型的形式化定义:假设 RS 是所有需求(requirements)的集合, ES 是所有专家(experts)的集合, $u(\cdot)$ 函数可以计算专家 $e(e \in ES)$ 对需求 $r(r \in RS)$ 的推荐度,还要考虑专家的可靠性等,即 $u: RS \times ES \rightarrow L$, L 是一定范围内的全序的非负实数,则:

$$\forall r \in RS, e^* = \arg \max_{e \in ES} u(r, e) \quad (1)$$

即专家推荐平台要研究的问题就是找到推荐度 L 最大的那些专家 e^* ,专家推荐的核心问题转化为计算 $u(\cdot)$,其计算值用来衡量专家对需求的有用程度,称为效用度。

在前述定义了专家科研背景和需求基本特征的基础上,通过计算向量空间余弦夹角来衡量专家研究兴趣与需求之间的匹配程度,效用函数 $u(\cdot)$ 定义为:

$$u(r, e) = \text{sim}(D_r, D_e) = \cos \theta = \frac{\sum_{k=1}^m d_{kr} \cdot d_{ke}}{\sqrt{\sum_{k=1}^m d_{kr}^2 \sum_{k=1}^m d_{ke}^2}} \quad (2)$$

式中, D_r 、 D_e 是两个待计算相似度的目标对象, 即为需求与专家, d_{kr} 、 d_{ke} 是对象 D_r 、 D_e 的特征向量权值。通过计算, 得到其夹角余弦, 值越大, 二者相似度越大, 反之越小, $0 \leq \text{sim}(D_r, D_e) \leq 1$ 。

假设 r_x 为需求集 RS 中的一个需求样本, e_x 为专家集 ES 中的一个专家样本。给定需求-专家对 (r_x, e_x) , 为它设置一个状态标签 $y_x \in \{0, 1, 2, 3\}$, 其中 0 代表未接受需求邀请, 3 代表接受需求邀请并回答。 (r_x, e_x, y_x) 是一个训练样本。根据上述定义, n 个训练样本就组成了训练数据集, 如公式(3)所示:

$$D_{\text{train}} = ((r_1, e_1, y_1), (r_2, e_2, y_2), \dots, (r_n, e_n, y_n)) \quad (3)$$

通过训练数据集 D_{train} 构建推荐模型, 并定义损失函数对模型进行优化, 判断当前专家 e_x 是否会接受某个新需求 r_x 邀请的标签 y_x 。对得到的标签 y_x 对所有专家进行排序或者加权等操作, 找到最适合需求的专家。

$$y_x = u(r_x, e_x) \quad (4)$$

3 专家推荐模型实现

3.1 特征提取与知识融合

结合专家动态研究兴趣特征将研究重点放在挖掘其学术论文、发明专利、工程案例中的相关知识, 获取需求与专家描述的有效信息, 构建科技服务知识库。通过文本多特征提取与知识融合可以对需求和专家文本数据进行统一建模, 模型结构如图 2 所示。

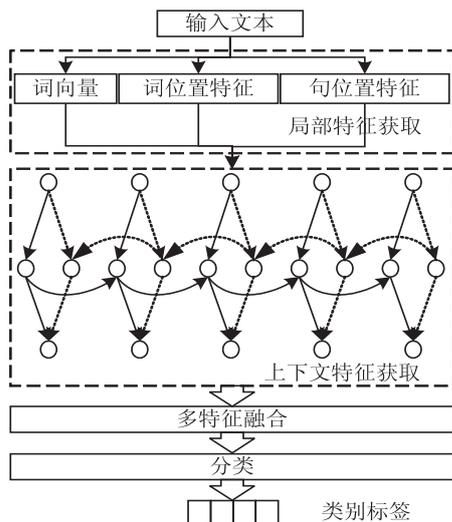


图 2 知识融合模型结构

知识融合模型主要由局部特征获取、上下文特征获取、多特征融合和分类四部分组成。首先, 获取三种不同输入特征, 包括词向量表示、词位置特征、句子位置特征; 然后, 获取上下文特征, 增强模型的特征学习能力; 最后, 将以上特征进行融合, 并输入到分类层进行分类, 预测并得到类别标签。

该文选取 ALBERT 模型融合多种特征, 并对输入文本进行语义理解, 提高文本的表征能力及文本分类的准确率。同时, 文本输入时的处理上要充分考虑任务需求, 比如输入是句子对, 输出是对该句子每个字的包含上下文信息的编码结果。然后, 对需求及专家分别进行个性化表征后应用于推荐系统。

3.2 专家推荐模型实现

实现专家推荐模型需要深入分析需求服务案例中的需求与专家, 挖掘二者之间的潜在知识关联, 然后进行专家推荐。在上述知识融合模型基础上提出一种面向公众决策需求咨询服务的专家推荐模型, 结构如图 3 所示, 其具体实现流程如下:

步骤 1: 对过往服务案例中需求与专家知识进行有效提取, 建立科技服务知识库, 筛选训练语料。

步骤 2: 以标题及内容的知识表征需求, 以专业领域及研究成果的知识表征专家, 将二者联合信息作为模型输入, 获取联合需求属性与专家偏好的深度特征。

步骤 3: 专家推荐模型的编码层采用在通用领域预训练好的 ALBERT 模型, 并在事先准备好的数据集上进行训练, 获取需求与专家的隐性知识关联, 使模型由通用领域迁移到科技服务领域。

步骤 4: 使用验证集评估模型训练效果。在模型通过验证集之后, 使用测试集评估训练模型效果, 获取最佳模型即为专家推荐模型。

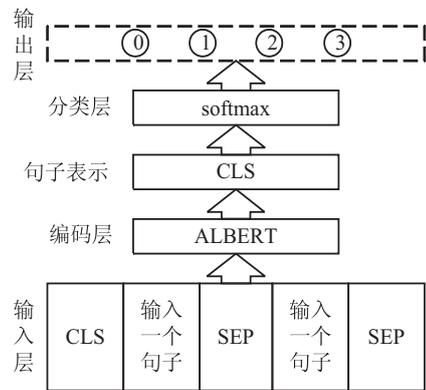


图 3 专家推荐模型结构

通过该模型可以对需求、专家、案例进行多特征知识融合, 并通过理解需求语义信息和学习专家特征之间的关系获取需求与专家的隐藏知识关联, 最终获得准确的专家推荐结果。

4 实验及结果分析

4.1 实验数据集

在系统梳理河北省关于专家和用户决策咨询数据关系的基础上, 该文融合学术论文、发明专利、工程案例等多个数据源中的有效信息(包括需求信息、专家信息、历史需求服务案例等内容), 组成科技服务知识

库。然后,经过团队人工标注最终筛选出 3 000 条专家信息,25 000 条历史服务需求专家匹配信息组成实验数据集,按比例 8 : 1 : 1 划分为训练集、测试集和验证集三个子集,数据已作脱敏处理。

4.2 实验环境

实验硬件、软件环境如表 1 所示。

4.3 实验对比与分析

该文总共选取了两种不同的预训练语言模型即 BERT 与 ALBERT 作为专家推荐模型的编码层,并分别采用 softmax 与 TextCNN 作为分类层进行训练模型推荐效果的对比实验,并对相应的推荐结果进行判定。

专家推荐模型在实验环节采用固定的参数配置(见表 2),此次对比实验采用谷歌官方发布的中文版 BERT 及 tiny 版的 ALBERT 进行文本特征的提取,并

通过数据处理阶段整理出的需求与专家数据集微调模型。

表 1 实验环境

名称	配置/版本
操作系统	Red HatEnterprise Linux 7.4
GPU	TeslaV100
内存	64 G
硬盘	30 T
python	3.6.5
tensorflow-gpu	1.14
keras	2.2.5
keras-bert	0.80.0
bert4keras	0.10.6

表 2 模型参数设置

模型名称	模型版本	dropout	epoch	learning rate	batch size
ALBERT(softmax)	albert_tiny_google_zh_489k	0.1	20	1e-5	8
BERT(softmax)	chinese_L-12_H-768_A-12	0.1	20	1e-5	8
BERT(TextCNN)	chinese_L-12_H-768_A-12	0.1	20	1e-5	8

专家推荐模型的准确度是评价模型的最基本指标,其衡量推荐模型能够准确预测用户需求对服务专家及专家对推荐需求的喜欢程度。本次实验主要考察训练及测试数据的一些统计特性,故采用分类准确率分数(accuracy_score)和多类别交叉熵损失函数(categorical_crossentropy)评价专家推荐模型效果。

由表 3 可以看出,基于 BERT 的专家推荐模型能够获得很好的准确率,但是由于 BERT 模型参数量巨大,模型训练耗时太久。以 BERT-TextCNN 为基础的专家推荐模型训练耗时有所增加而且推荐效果更差。基于 ALBERT 的专家推荐模型训练耗时更少,计算成

本更低,准确率上能够取得明显的提升。

表 3 不同语言模型训练效果对比

模型	验证集准确率/%	训练耗时/h
BERT(softmax)	84.13	15
BERT(TextCNN)	82.85	16
ALBERT(softmax)	88.88	0.5

由图 4 可以看出,基于 ALBERT 的专家推荐模型在上述各类模型中收敛速度更快、准确率更高,能够更高效地推荐出合适的服务专家。

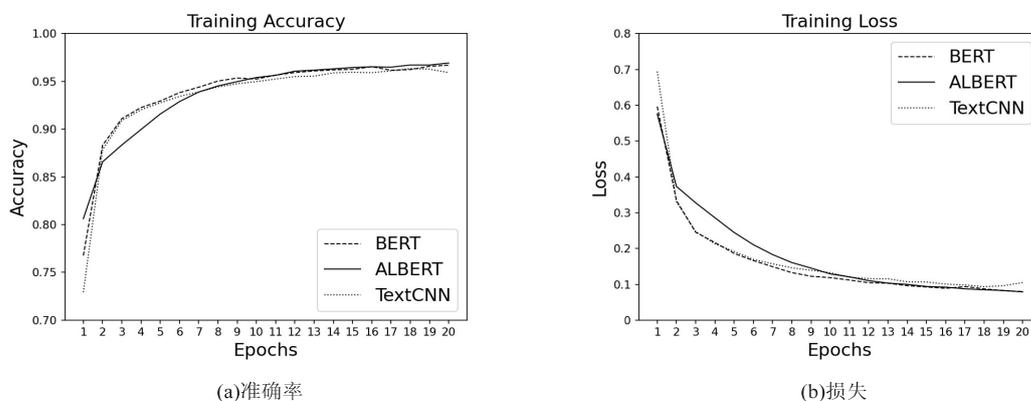


图 4 不同语言模型训练效果对比

5 结束语

该文研究了科技领域公众决策咨询专家推荐服务对知识整合和知识推荐的需求,为了解决专家推荐不

合理不准确的难题,在对知识整合、知识推荐理论及预处理语言模型进行全面分析的基础上,对推荐服务所需材料进行整合并统一建模,从而将科技领域现有的信息孤岛发展成一个数据库。最终提出了一种面向科

技领域用户决策咨询需求服务的专家推荐模型,模型能够深入分析需求服务案例中的需求与专家,并针对隐性知识的特点,挖掘二者之间的潜在知识关联,从而完成对所需服务专家的推荐。通过实验表明该专家推荐模型在使用过程能够取得很好的推荐效果。

参考文献:

- [1] REICHLING T, SCHUBERT K, WULF V. Matching human actors based on their texts: design and evaluation of an instance of the ExpertFinding framework[C]//Proceedings of the 2005 international ACM SIGGROUP conference on supporting group work. Sanibel Island: ACM, 2005.
- [2] 胡 斌. 科技项目评审专家推荐系统的研究与实现[D]. 杭州: 杭州电子科技大学, 2011.
- [3] 魏 捷. 论文审稿专家推荐系统的设计与实现[D]. 北京: 北京邮电大学, 2019.
- [4] 李 莹. 面向企业需求的专家推荐算法研究[D]. 北京: 北京交通大学, 2018.
- [5] 曹渝昆. 基于神经网络和模糊逻辑的智能推荐系统研究[D]. 重庆: 重庆大学, 2006.
- [6] 袁仁进, 陈 刚, 李 锋. 面向新闻推荐的用户兴趣模型构建与更新[J]. 计算机应用研究, 2019, 36(12): 3593-3596.
- [7] 秦 冲, 赵铁柱, 柳 毅. 个性化推荐算法的研究及发展综述[J]. 东莞理工学院学报, 2021, 28(3): 51-60.
- [8] 翁小兰, 王志坚. 协同过滤推荐算法研究进展[J]. 计算机工程与应用, 2018, 54(1): 25-31.
- [9] SANZ-CRUZADO J, CASTELLS P, MACDONALD C, et al. Effective contact recommendation in social networks by adaptation of information retrieval models[J]. Information Processing and Management, 2020, 57(5): 102285.
- [10] WANG Donghui, LIANG Yanchun, XU Dong, et al. A content-based recommender system for computer science publications[J]. Knowledge-Based Systems, 2018, 157: 1-9.
- [11] 尹 祎, 冯 丹, 施 展. 一种基于效用的个性化文章推荐方法[J]. 计算机学报, 2017, 40(12): 2797-2811.
- [12] HUANG L W, JIANG B T, LV S Y, et al. Survey on deep learning based recommender systems[J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647.
- [13] CREMONESI P, TRIPODI A, TURRIN R. Crossdomain recommender systems[C]//Proceedings of IEEE international conference on data mining workshops. Washington D. C.: IEEE, 2011: 496-503.
- [14] 张 鼎, 周年喜, 张 英. 基于语义扩展的个性化知识推荐技术研究[J]. 情报理论与实践, 2009, 32(8): 101-104.
- [15] 巩 军, 刘 鲁. 基于个人知识地图的专家推荐[J]. 管理学报, 2011, 8(9): 1365-1371.
- [16] 刘 伟, 刘柏嵩, 王洋洋. 海量学术资源个性化推荐综述[J]. 计算机工程与应用, 2018, 54(3): 30-39.
- [17] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics; human language technologies. Stroudsburg: Association for Computational Linguistics, 2019: 4171-4186.
- [18] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[C]//Proceedings of the 8th international conference on learning representations. La Jolla: ICLR, 2020: 1-17.