

基于建构主义学习理论的个性化知识推荐模型

谢振平 金晨 刘渊

(江南大学数字媒体学院 江苏无锡 214122)

(江苏省媒体设计与软件技术重点实验室(江南大学) 江苏无锡 214122)

(xiezhening@hotmail.com)

Personalized Knowledge Recommendation Model Based on Constructivist Learning Theory

Xie Zhenping, Jin Chen, and Liu Yuan

(School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122)

(Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, Jiangsu 214122)

Abstract Personalized recommendation is becoming a basic form of information network services in the era of Internet+ and big data. Its wide use in e-commerce and social media has produced huge commercial value, however, there are only limited research and applications in the field of personalized knowledge learning, which may have tremendous potential social value for public education and personalized information selection. This study proposes a novel personalized knowledge recommendation method—constructive recommendation model, based on constructivist learning theory. The new model uses knowledge networks to represent expected knowledge systems, uses the nearest neighbor priority strategy to select knowledge item candidates, and introduces top- K unstudied knowledge recommendation algorithm based on sorting knowledge candidate items by their learnable constructive degrees. The proposed constructive recommendation model can dig users' potential knowledge demands by comparing domain knowledge network structure and users' learnt knowledge network structure. Then it can orderly recommend most needful knowledge items to users for gaining the greatest constructive learning effect. We choose a very interesting healthy diet knowledge system as the experimental problem, in which 14 600 knowledge documents are grabbed from public Internet Websites in China with knowledge subjects 'health knowledge', 'dietary nutrition' and 'dietary misconceptions' etc. Some meaningful experimental analysis are executed in this paper, and corresponding results demonstrate that recommended knowledge sequences given by our model can gain stronger knowledge continuity and higher knowledge learning efficiency than the existing related methods.

Key words constructivist learning theory; personalized learning; knowledge network; constructive recommendation; healthy diet knowledge

摘要 个性化推荐正成为“互联网+”和“大数据”时代信息网络服务的基本形式,虽然其已在电子商务和社交媒体的广泛应用中产生了巨大的商业价值,但在具有巨大潜在社会价值的个性化知识学习领域,

收稿日期:2016-07-27;修回日期:2017-04-27

基金项目:江苏省自然科学基金项目(BK20130161);国家科技支撑计划基金项目(2015BAH54F01);国家自然科学基金项目(61572236)

This work was supported by the Natural Science Foundation of Jiangsu Province of China (BK20130161), the National Key Technology Research and Development Program of China (2015BAH54F01), and the National Natural Science Foundation of China (61572236).

相关研究与应用还较为稀少. 研究提出一种基于建构主义学习理论的个性化知识推荐方法——建构推荐模型. 新模型首先考虑将知识系统以知识网络的形式进行表达, 随后引入最近邻优先的候选知识选择策略, 以及基于最大可学习支撑度优先的 top-K 未学知识推荐算法. 建构推荐模型通过知识网络的知识关联结构挖掘用户知识需求, 并推荐给出最具建构学习价值的待学新知识. 以饮食健康知识系统学习为例的实验分析表明, 新模型在多种情况下推荐产生的个性化知识序列均具有较强的知识关联性和较高的知识体系覆盖率.

关键词 建构主义学习理论; 个性化学习; 知识网络; 建构推荐; 饮食健康知识

中图分类号 TP391

近年来, 随着无线网络和移动宽带互联网的快速发展, 以及成熟的便携式移动设备的大量涌现, 在线学习已越来越成为一种重要的知识学习方式^[1-2]. 在线学习具有学习资源广泛、资源更新及时、学习方式便捷等特点. 然而互联网上的海量学习资源在给用户带来便捷学习的同时也造成了“认知过载”、“知识碎片化”等困惑^[3]. 如何帮助不同学习者从海量的学习资源中找到个体所需的资源已成为一个巨大的挑战, 在其中引入个性化推荐技术是一个必然的选择^[3-10].

传统个性化推荐的主要目标是通过分析用户信息和项目信息, 获取用户兴趣偏好, 从而为用户推荐感兴趣的项目, 主要技术包括基于内容的推荐^[11-12]、协同过滤推荐^[13]、基于规则的推荐^[14-15]等. 这些技术在电子商务和社交媒体领域具有较高的适用性, 并已取得了大量的成功应用; 但在个性化知识推荐学习方面, 相关研究与应用还较少, 而直接使用现有推荐算法将面临 2 个问题:

1) 推荐产生的知识序列缺乏连续性. 分析人类知识学习过程可知, 较好的系统性和连续性是个性化知识学习序列的基本要求, 而现有的推荐方法在每次推荐新项目时通常考虑为独立的过程, 对于推荐结果序列的系统性和连续性较少关注.

2) 知识需求分析片面化. 个性化知识推荐的目的是帮助用户学习新知识, 弥补知识缺陷, 知识需求分析应主要依据用户当前知识水平和目标知识体系. 现有个性化推荐方法在分析用户推荐需求时, 缺少这方面的考虑, 难以充分挖掘用户的实际知识需求.

另一方面, 建构主义学习理论^[16-17]认为, 人类知识学习本质上是一个知识建构过程, 个体通过自身原有的知识经验建构生成新的知识; 如用户在学习了知识 A, B, C 的基础上能够根据三者的内在逻辑结构建构理解新知识 D. 为此, 本文提出一种基于建

构学习理论的个性化知识推荐模型——建构推荐模型. 建构推荐模型采用知识网络建模用户所需的知识系统, 基于知识网络的边信息存储知识间的建构关系, 据此结合用户已学知识内容, 引入支撑度最大优先的随机游走推荐算法, 连续地推荐产生最佳知识学习序列.

1 背景技术

1.1 个性化推荐

基于内容的推荐^[11-12] (content-based recommendation, CBR) 是最早被提出的一种推荐技术, 算法通常包括 3 个步骤: 1) 为资源库中每一个资源抽取特征项作为资源特征; 2) 分析用户过去喜好的资源, 从中学习出这些资源的特征作为用户喜好特征; 3) 通过计算用户喜好特征和候选资源特征的相似度, 为用户推荐相似度高的一组资源. CBR 算法的优点是准确度高, 不依赖大量的用户群, 新资源不存在冷启动问题; 缺点是特征抽取困难, 局限于文本资源的推荐, 并且很难挖掘用户潜在兴趣.

协同过滤推荐^[13] (collaborative filtering recommendation, CFR) 通常分为 2 类: 基于用户的协同过滤^[18-20] (user-based collaborative filtering, UBCF) 和基于项目的协同过滤^[21-22] (item-based collaborative filtering, IBCF). 基于用户的协同过滤首先寻找与目标用户有相似兴趣的用户, 然后根据相似用户对项目的评分来预测目标用户对未知项目的评分, 将评分高的项目推荐给目标用户. 基于项目的协同过滤则是根据用户对项目的评分数据来分析项目之间的相似性, 并将那些与用户之前喜欢的项目相似度高度的项目推荐给用户. CFR 算法的优点是适用性广, 主要依据用户对资源的评分, 与资源的形式无关, 因此理论上 CFR 算法适用于任何资源的推荐. 协同过滤推荐的缺点也很明显, 由于过分依赖用户

评分数据,导致新资源和新用户存在冷启动问题,并且通常用户评分项目较少,导致用户评分矩阵稀疏,造成相似性计算误差大,影响最终推荐的准确度.

关联规则的推荐^[14-15] (rule-based recommendation, RBR)关键在于挖掘不同资源项之间可能存在的价值关联项,然后基于已有的关联规则为用户推荐可能感兴趣的资源.关联规则的挖掘一般分为2个步骤:1)通过分析用户与资源项之间的历史数据生成所有的频繁项集;2)通过计算支持度、置信度、提升度来提取强关联规则. RBR 算法的主要不足之处在于:规则制定费时费力,规则一旦生成不能自动更新.

1.2 建构主义学习理论

建构主义^[16-17]源自瑞士著名心理学家皮亚杰(Jean Piaget)创立的关于儿童认知发展理论,他坚持从内因和外因相互作用的观点来研究儿童认知发展,他认为儿童是通过与周围环境的不断相互作用来逐步建构对外部环境的认识.他提出,知识并非单纯地来自主体或者客体,而是在双方相互作用的过程中生成的.主体获得新经验需要自身原有的经验基础,而新经验的获得又能使原有经验得到更新和完善.从这一角度看,人类学习的本质是一个主动建构知识的过程,而不是被动地接受信息的过程.

明显地,个性化知识推荐服务目标与建构主义学习理论有着天然的本质联系,后者应能较好地指导个性化知识推荐方法的设计实现,使得用户能更轻松且高效地获得新知识.

2 模型框架

本文将建构主义学习理论引入到知识推荐系统中,提出一种新的个性化知识推荐方法——建构推荐模型,其中主要考虑2个问题:1)如何表示用户当前知识经验;2)如何基于已有的知识经验推出下一时刻最适合建构学习的新知识.本文考虑基于知识网络图覆盖的形式帮助用户逐步建构学习来完善自身的知识,并提出知识的可学习支撑度来评估用户对新知识的可学习建构优先级.

设计考虑的建构推荐模型如图1所示,主要由2部分构成:知识网络模块和建构推荐模块.知识网络^[23-24]是一个具有节点间互联关系的知识环境,节点代表知识单元,边表示知识单元之间的相关关系.知识网络的构建包括:知识术语抽取、词向量模型以及通过语义距离的约束生成知识网络.这样,用户知识背景可认为是知识网络的一个子结构,表示用户已学习的知识子系统.模型框架的另一个核心是建构推荐模块,主要任务是通过知识网络建立用户已有知识与需求知识间的逻辑关联,并分析用户需求,为用户推荐最具建构学习价值的新知识.

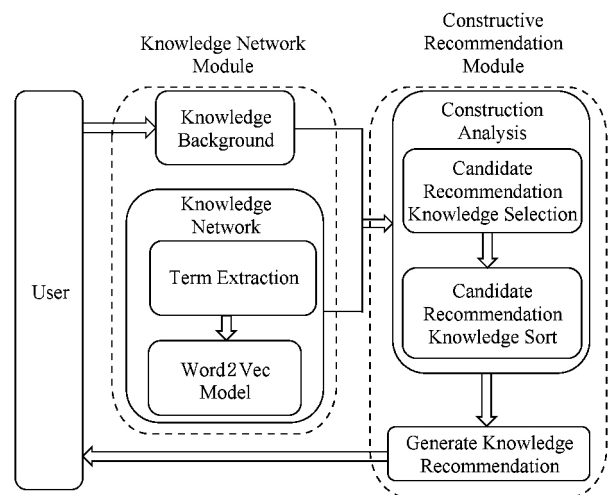


Fig. 1 Framework of the proposed constructive recommendation model

图1 建构推荐模型框架

2.1 知识网络模块

建构推荐模型的知识网络构建过程包含3个步骤:1)通过术语抽取技术自动识别领域知识术语;2)采用词向量工具为每个知识术语训练生成一个实数向量;3)通过获取的实数向量计算知识术语之间的语义关系,从而建立知识之间的结构关联.

2.1.1 知识术语抽取

知识术语抽取的主要任务是从特定领域的文本语料中自动识别完整独立的知识术语,每一个知识术语表示一个知识概念.知识术语抽取具体流程如图2所示:首先获取特定领域的一定量的文本语料,

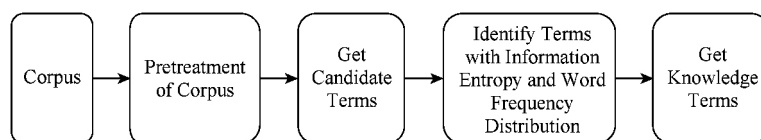


Fig. 2 Term extraction process

图2 术语抽取流程

通过分词、去除停用词进行语料预处理,将预处理的字串重新组合生成候选术语;然后通过信息熵和词频分布筛选候选术语^[25]. 其中,信息熵^[26]用于计算候选术语在语料中左右边界的稳定性,以判断相应术语的独立成词程度;词频分布通过计算候选术语的词频来区分普通词和领域术语,通常地,领域术语在相关领域出现的频率较高,在不相关领域出现的频率较低^[25]. 实际中,可结合信息熵和词频为每个候选术语生成一个综合权重,按权重大小排序,选取权重大的前 T 个候选术语作为最终知识术语.

2.1.2 词向量模型

词向量模型是神经网络在自然语言处理领域应用的产物,最早由 Hinton 于 1986 年提出. 词向量的核心思想是通过文本语料训练,将每个词映射成一个高维的实数向量,然后通过计算向量之间的距离可直观描述词与词之间存在的语义关系. 目前词向量已经广泛应用在文本情感分类^[27]、情感新词发现^[28]、词义消歧^[29]等自然语言处理领域.

本文采用 Word2Vec 工具^[30-32]训练词向量. Word2Vec 中包含 2 个重要模型:CBOW 模型和 Skip-gram 模型^[30-31]. 这 2 个模型的区别在于:前者是在已知上下文的基础上预测当前词;而后者是在已知当前词的基础上预测上下文. 每个模型均有 2 套框架,分别基于 Hierarchical Softmax 和 Negative Sampling 来设计实现. 与 Hierarchical Softmax 相比,Negative Sampling 不使用复杂的 Huffman 树,而是利用相对简单的随机负采样,能提高训练速度并改善词向量的质量. 综合地,本文采用基于 Negative Sampling 的 CBOW 模型来训练词向量. 如图 3 所示, CBOW 模型包括输入层、投影层和输

出层,工作原理是通过输入上下文 n 个词来预测当前词 W_t 出现的概率. 本文实验中设置 $n=4$,输入经过术语识别的分词语料,最终为每个术语训练生成一个 200 维的实数向量.

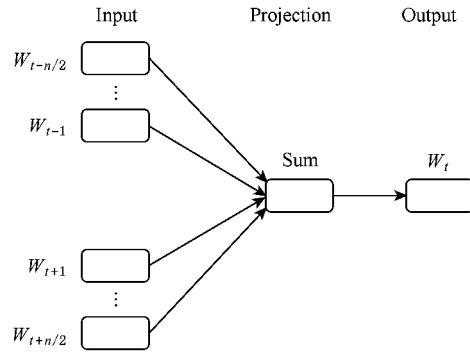


Fig. 3 CBOW term vector model

图 3 词向量模型示意图

2.1.3 知识网络构建

基于获得的每个知识术语所对应的词向量,可通过词向量间的数学运算来计算知识之间可能存在的知识关系,并将有效的知识关系进行存储,最终建立一个完整的知识网络.

建构推荐模型的知识网络构建流程如图 4 所示,知识网络构建遵循近邻优先原则,即知识库中 T 个知识分别寻找语义距离最近的 M 个知识为其近邻知识,并进行记忆存储构成知识关系,则总的边关系数量为 $T \times M$ 个. 通常, M 值太小会丢失必要的近邻知识关系,降低知识网络的整体连通性,无法有效表达目标知识系统;而 M 值过大只会增加知识网络冗余度,并且增加推荐计算效率,对知识网络的整体连通性提升价值有限. 结合小世界网络理论和实验分析,本文考虑 $M=10$.

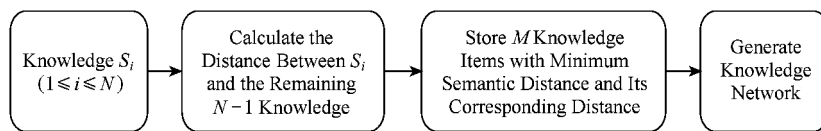


Fig. 4 Knowledge network building process of constructive recommendation model

图 4 建构推荐模型的知识网络构建流程

图 5 所示为一个建构推荐模型的知识网络结构实例,圆点表示知识,有向边表示近邻知识关系.

上述方法构建的知识网络具有 3 个特征:

- 1) 近邻关系是单向的,即若 b 是 a 的近邻知识, a 不一定是 b 的近邻知识.
- 2) a 的所有近邻知识总数为 a 的入度,近邻知识包含 a 的知识总数为 a 的出度,任何一个知识节

点的入度固定为 M ,出度则不固定.

3) 知识网络中任意 2 个知识节点 a, b 间的连接关系有 4 种情况:两者互为近邻知识关系;两者无直接近邻关系;一方是另一方的近邻知识,但反之则不是.

此外,一个好的知识网络在不考虑连接边的方向性时应该具有较强的全连通性.

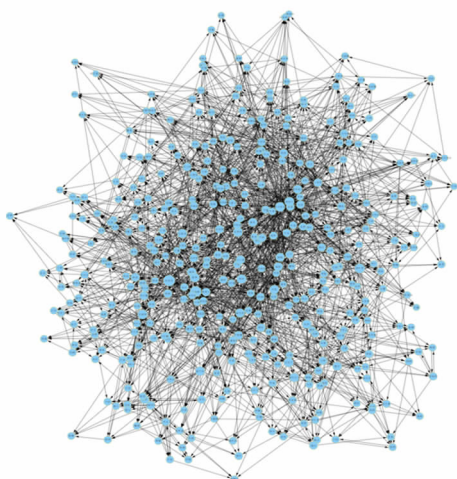


Fig. 5 An example of knowledge network structure of constructive recommendation model

图5 建构推荐模型的知识网络结构实例

2.2 建构推荐模块

建构推荐的主要任务是结合知识网络分析用户知识需求,从而为用户生成语义连续的推荐序列.建构推荐模块主要包括候选推荐知识提取和候选知识排序输出 2 个部分.候选推荐知识考虑从最近学习知识序列的关联知识中提取不超过 N 个的未学知识;候选知识排序输出推荐引入支撑度指标作为排序量.为此,首先引入关于支撑知识的定义.

定义 1. 知识支撑关系. 设知识 b 与 a 存在近邻关系,且设 b 是 a 的 top- M 近邻,记作 $a \leftarrow b$,则称 b 是 a 的支撑知识,也称 a 是 b 的外延知识.

进一步,可设知识 a 的所有支撑知识构成的集合为 a 的支撑知识集,记为 S_a . 如图 6 所示,知识 t 的支撑知识集为 S_t .

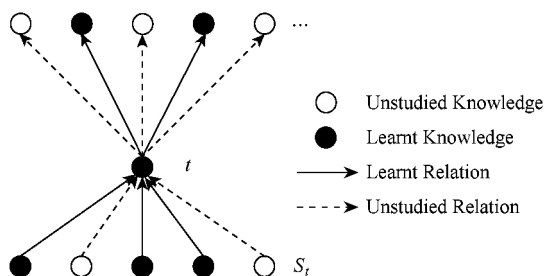


Fig. 6 An example of knowledge network substructure

图6 知识网络子结构示例

进一步引入支撑度概念,表示用户当前已学知识对新知识的直接可建构程度.例如“营养素”的支撑知识集为“蛋白质”、“脂肪”、“碳水化合物”、“维生素”、“纤维素”等,用户对支撑知识的掌握体现了建

构理解“营养素”的能力,本文引入支撑度概念量化描述这种能力.任一知识 t 相对于用户当前状态的支撑度记为 $W_{sp}(t)$,具体定义如下:

$$W_{sp}(t) = \frac{\sum_{s_i \in S_t} Sim(t, s_i)}{\sum_{s_j \in \bar{S}_t} Sim(t, s_j)}, \quad (1)$$

$$Sim(t, s) = \exp(-d_{ts}^2 / (2\sigma^2)), \quad (2)$$

其中, S_t 表示知识 t 的支撑知识集, \bar{S}_t 表示 S_t 中已学习的知识子集; $\exp()$ 表示指数函数, d_{ts} 为知识 t 与 s 间的距离, σ 为尺度因子. 如图 6 所示,实心节点表示已学习知识,空心节点表示未学知识,相应地,实线表示已学知识关系,虚线表示未学知识关系.

3 模型实现

3.1 建构主义推荐策略

建构主义学习理论的核心在于建构认知,学习者通过已掌握知识建构获得新知识.相应地,引入如图 7 所示的建构推荐策略.同样地,图 7 中实心圆点表示已学知识,线条圆点表示候选知识,空心圆点表示其他知识;实线表示已经建构的知识关系,虚线表示未建构学习的知识关系.

图 7(a)表示一次知识推荐的开始状态,图 7(b)(c)(d)分别表示候选推荐知识选择、候选知识推荐排序以及推荐知识学习后的更新结果.图 7(a)中有 2 个知识为已学知识,其他均为未学知识,它们之间基于知识网络关系构成一个整体.

如图 7(b)所示,候选推荐知识选择考虑从已学知识的直接相关但还未学的知识中选取,其中直接相关知识包括支撑知识和外延知识.候选知识选择按 2 个优先次序进行选取:最近学习知识的关联知识为第 1 优先选择,而同一已学知识的不同关联知识则按最短距离优先选择.同时考虑实际应用需要,引入最大候选知识容量参数 N ,即每次按上述方法选择的最大候选知识项数不超过 N .

如图 7(c)所示,候选知识推荐排序对所有已选择的候选知识项按未学知识的支撑度从大到小排序,并优先推荐给出支撑度较大的候选项作为推荐结果.不失一般性,可考虑推荐结果为排序结果的 top- K 输出,即输出排序靠前的 K 个候选知识项作为推荐选项,供用户选择.

最后,如图 7(d)所示,用户选择一个推荐项(图中为用户选择了排序第 1 的知识项)进行学习后形成了新的已学知识状态(同图 7(a)相似的状态结构).

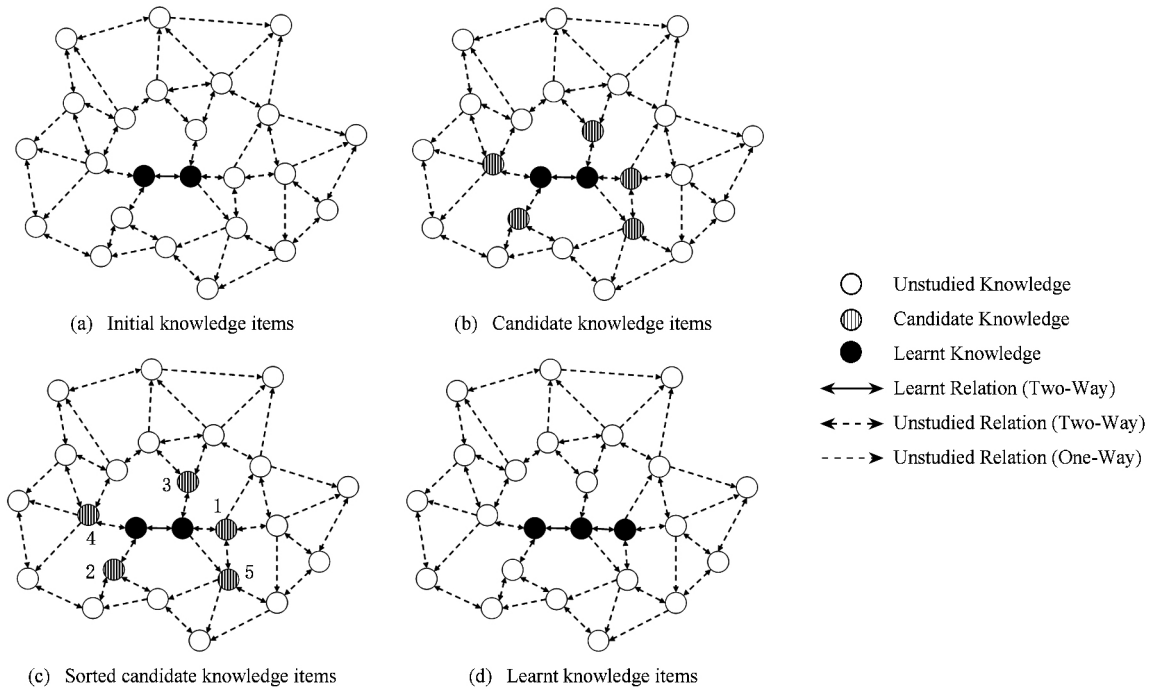


Fig. 7 The diagram of constructive recommendation strategy

图7 建构推荐策略过程示意图

3.2 建构推荐算法描述

基于上述建构主义推荐策略,进一步对建构推荐算法过程进行具体描述.算法输入为用户初始背景知识,然后连续地推荐给用户最具学习价值的top-K个学习知识项,用户选择其中一个知识学习后形成新的背景知识,如此不断迭代,用户知识得到渐进增长.

算法1. 个性化知识学习的建构推荐算法.

输入:用户初始背景知识 $U = \{t_1, t_2, \dots, t_L\}$.

过程:

- ① $C_n = \{\langle c, W_{sp} \rangle \mid C_n = \emptyset\}$, $i = 1$;
- ② while $|C_n| < N$ and $i \leq T$ do
- ③ $C = \text{getCandNeib}(t_i)$;
- ④ if $\text{isempty}(C) == \text{TRUE}$ then
- ⑤ $i = i + 1$;
- ⑥ else
- ⑦ 根据式(1)计算 $W_{sp}(c) \mid c \in C$, $\text{addToCn}(\langle c, W_{sp}(c) \rangle)$, $i = i + 1$;
- ⑧ end if
- ⑨ end while
- ⑩ 根据支撑度 W_{sp} 对 C_n 降序排序;
- ⑪ 输出排序列表 C_n 中前 top-K 推荐项;
- ⑫ 用户选择推荐项 c_u , 结束本次学习;
- ⑬ $L = L + 1$, $t_L = c_u$, $U = U \cup \{t_L\}$;

⑭ goto Step②;

算法1中, $C_n = \{\langle c, W_{sp} \rangle\}$ 表示当前候选知识项集合, $\text{getCandNeib}()$ 表示获取某个知识节点的未学习且不在 C_n 中的最近直接相关知识(包括支撑知识和外延知识), $\text{addToCn}()$ 表示将一个新的候选知识和相应的支撑度值构成的元素项加入当前候选知识项集合 C_n , 步骤⑩中的排序操作表示对 C_n 中的项按每个项的支撑度值从大到小排序.

3.3 分析讨论

在建构推荐模型中,若将模型中的候选项排序指标“支撑度最大优先”替换为“距离最小优先”,则相应的模型将等价于某种基于内容的个性化推荐模型.由此,知识网络上的基于内容相似的个性化推荐模型同样可直接用于个性化学习知识推荐,本文也将此方法作为参照,在实验研究中进行比较分析.另一方面,相比于协同过滤推荐算法,建构推荐模型由于直接基于知识网络进行用户知识需求分析,避免了冷启动问题和用户历史数据稀疏的问题.

进一步分析本文中模型的复杂度.假设在建构推荐算法中,已学习知识数为 T , 候选知识项集合容量为 N , 则算法主要时间复杂度体现在步骤③和步骤⑦, 每执行1次语句认为检索1次数据库, 步骤③表示获取某个知识节点未学习的最近直接相关知识, 最大循环次数为 T , 步骤⑦表示计算候选知识

支撑度,最大循环次数为 N ,则算法的时间复杂度可表示为 $O(T+N)$.相对地,最大相似度优先排序推荐算法与建构推荐算法相比无需执行步骤⑦,则时间复杂度应为 $O(T)$.虽然如此,通常在检索产生 N 个候选样本时,所需的实际检索次数要远小于 T ,本文实验部分显示的结果为大部分情况下均小于 5.如此,检索计算复杂度分别约为 $O(5+N)$ 和 $O(5)$,新算法的计算复杂度同样是实用可接受的.

此外,虽然与电子商务和社交网络中的传统个性化推荐目标具有较大差别,但个性化知识推荐的核心需求仍是发现并给出用户最需要的信息内容.本文受建构主义学习理论启发,提出的建构推荐模型给出了一种思想上直观、算法上可行的新型个性化推荐方法,是对现有个性化推荐方法的有益补充,理论上也能拓展应用于传统的电子商务和社交网络领域.

4 实验研究

4.1 实验方案

考虑个性化知识推荐的特点及当前研究现状,本文考虑以饮食健康知识学习为实验对象,对建构推荐模型进行性能分析.

1) 我们使用八爪鱼采集器^①从中国食品科技网^②和 39 健康网^③抓取“健康知识”、“膳食营养”、“饮食误区”等主题的 14 600 篇饮食相关科普文章,共计约 2 000 万字,作为语料素材.经过语料预处理和术语抽取技术识别获得 1 000 条饮食知识术语,部分术语抽取结果如表 1 所示,其中综合权重为 2.1.1 节中所述的基于信息熵和词频生成的术语权重.

Table 1 Part of Healthy Diets Knowledge Terms

表 1 获取的部分健康饮食术语

Diet Term	Weight	Diet Term	Weight
牛奶	0.928 2	榴莲	0.703 2
鸡蛋	0.894 8	豆腐	0.695 6
水果	0.828 1	西红柿	0.658 7
木瓜	0.813 5	苹果	0.658 6
青枣	0.739 1	大豆	0.655 5
柠檬	0.723 5	樱桃	0.649 0
蜂蜜	0.714 2	蔬菜	0.631 0

① <http://www.bazhuayu.com/>

② <http://www.tech-food.com/>

③ <http://food.39.net/>

2) 利用抽取的知识术语筛选文本语料,去除不包含知识术语的句子,用经过筛选的文本语料训练得到知识术语词向量,并通过计算知识术语间的欧氏距离作为不同术语词间的语义距离.表 2 给出了部分健康饮食知识术语间的语义距离值.直观分析可以发现,“牛奶”与“豆奶”、“奶制品”的语义距离明显小于与“水果”、“蔬菜”的语义距离,这一结果与术语的实际语义相一致,也表明本文所训练生成的健康饮食知识术语词向量具有较高的合理性.

Table 2 Semantic Distance of Some Healthy Diet Knowledge Terms

表 2 部分健康饮食知识术语间的语义距离

Diet Term a	Diet Term b	Semantic Distance
牛奶	牛奶	0
牛奶	脱脂牛奶	1.571 8
牛奶	豆奶	1.586 6
牛奶	奶制品	1.728 1
牛奶	水果	2.557 3
牛奶	蔬菜	2.851 2

3) 基于建构推荐模型的知识网络构建方法生成含有 1 000 个健康饮食知识术语的知识网络,图 8 给出了相应的知识网络中语义距离的分布情况,距离值主要落在区间 $[0, 0.0312, 3.1215]$ 内,分布形状呈现一定的规则性.

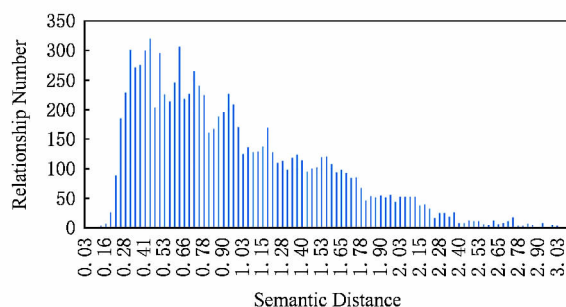


Fig. 8 Semantic distance distribution among network nodes of our knowledge network

图 8 知识网络节点间语义距离分布

为了对比研究建构推荐模型的性能,我们引入基于最大相似度优先排序的推荐 Similarity_Rec 方法和随机推荐 Random_Rec 方法作为参照,类似地记建构推荐方法为 Support_Rec. 基于最大相似度优先排序的推荐指候选推荐知识按相似度最大优先

排序,选择相似度最大的 top-K 个知识作为推荐输出. 随机推荐指每次从候选推荐知识中等概率随机选择 K 个知识作为推荐输出.

4.2 评价指标

从个性化知识推荐的目标需求出发,我们引入学习效率和知识序列关联度 2 个性能指标. 学习效率表示用户在连续选择学习一段时间后,学习获得的总知识量与学习的知识数间的相对比值. 在学习相同个数的知识前提下,学习获得的知识量越多则学习效率越高. 学习知识序列关联度表示用户在推荐给出的学习知识中连续选择的知识间的相关程度,基于学习的一般认知,学习知识序列间的相关程度越高则越有利于新知识的快速掌握.

为此,首先引入知识网络上的知识量定义.

定义 2. 一个知识网络或其一个子网络的知识量由 2 部分构成:知识节点的自身知识量以及知识节点间的关系知识量,其中记知识节点 a 的知识量为 $KI(a)$,知识节点 a 与 b 间的关联知识量为 $KI(a,b)$,具体定义如下:

1) $KI(a)$ 根据知识内涵先验给出(本文研究中默认取 $KI(a)=1$);

2) 若 a 与 b 间存在直接邻接关系,即其中至少有一个知识节点为另一个知识节点的支撑知识,则 $KI(a,b)$ 由式(3)定义,否则 $KI(a,b)=0$.

$$KI(a,b)=1-\exp\left(\frac{-d_{ab}^2}{2\sigma^2}\right), \quad (3)$$

其中, d_{ab} 表示 a 与 b 间语义距离;尺度参数 σ 与式(2)中定义相同,且可取相同值,本文研究中经实验比较均取 $\sigma=1$. 上述定义在考虑知识系统的总知识量时,不仅关注知识本体的知识量,也关注知识间的语义关系知识价值. 其中,2 个语义距离较近的直接相关联知识节点间的关系知识量较小,即若两者的信息差异度小,则建构形成的关系知识量也小.

基于定义 2,可对知识网络中的任意已学知识子网给出学习效率定义:

$$Ef(U_{cur}) = \left(\sum_{t \in U_{cur}} KI(t) + \sum_{t,s \in U_{cur}} KI(t,s) \right) \times \frac{|U_{all}|}{KI_{all} \times |U_{cur}|}, \quad (4)$$

其中, $U_{cur} = \{t_1, t_2, \dots, t_L\}$ 表示已学知识子网的知识节点集, U_{all} 为整体知识网络所含的知识节点集, KI_{all} 为整个知识子网的总知识量,有:

$$KI_{all} = \sum_{t \in U_{all}} KI(t) + \sum_{t,s \in U_{all}} KI(t,s). \quad (5)$$

根据上述学习效率定义,个体在学习某个知识

系统的过程中所选择的不同知识学习序列将对应不同的学习效率. 而在现实学习中,我们通常希望高效快速地掌握一个知识系统的信息内涵,即通过学习有限个节点知识,掌握知识系统中较多的知识信息. 例如给定 2 个知识序列,序列 1:“苹果、香蕉、西瓜、桃子、柑橘、猕猴桃、葡萄、草莓”;序列 2:“苹果、香蕉、西瓜、西红柿、红薯、黄瓜、青椒、胡萝卜”. 序列 1 中所有知识都围绕水果主题;序列 2 中“苹果、香蕉、西瓜”属于水果,“西红柿、红薯、黄瓜”是介于水果和蔬菜之间的食物,“青椒、胡萝卜”属于蔬菜. 虽然 2 个学习序列推荐了相同数量的知识,序列 2 的知识量显然更多. 根据式(4)的学习效率公式定义,较高的学习效率值也表示个体在学习等量的知识数后获得了更高的知识量,两者具有良好的一致性. 这也表明,我们引入的学习效率性能指标具有较好的合理性.

进一步,对最新学习的知识 t ,引入学习知识序列关联度定义:

$$Rele(t) = \sum_{s \in V_t} Ic(t,s) \times Sim(t,s), \quad (6)$$

其中, V_t 表示最近历史学习的知识集,即学习 t 之前最近学习的一定数量的知识,本文取最近的前 5 条历史知识; $Ic(t,s)$ 为 t 与 s 间是否存在直接关联(存在 $t \leftarrow s$ 或 $s \leftarrow t$)的指示函数,是则 $Ic(t,s)=1$,否则 $Ic(t,s)=0$;函数 $Sim()$ 定义同式(2).

综合地,上面引入的学习效率和知识序列关联度计算直接取决于已完成的知识学习历史以及目标学习知识系统. 此外,由于推荐算法不同而产生的任一不同知识学习过程点均可对应求得相应的学习效率和平均学习知识序列关联度,据此可定量地评价不同个性化知识推荐方法的性能.

4.3 实验结果与分析

建构推荐模型有 N 和 K 两个模型参数需要初始设定,分别表示每次推荐时选取的候选知识容量上限和推荐输出时给出的用户可选的知识项数. 首先考虑取 $K=1$,即每次只推荐输出排序最高的 1 个候选知识项时, N 在不同取值下 3 种对比算法的推荐性能情况. 表 3,4 给出了相应的一组实验结果,分别显示了学习效率和知识序列关联度. 实验中, N 分别取 10, 20, 30 和不限(Inf),每次随机选取 1 个初始知识,然后进行连续推荐模拟. 表 3 中实验结果为 20 次模拟的平均结果及相应的标准差. 为了更直观地表示,表 3 中不同已学知识比例时的学习效率值直接用已学知识量的比例间接表示.

Table 3 Learning Efficiency Values Obtained by Three Recommendation Algorithms on Different N with K=1
表 3 K=1 时不同 N 下的 3 种推荐算法的学习效率性能比较

N	Learnt Knowledge	Learning Efficiency/%		
		Support_Rec	Similarity_Rec	Random_Rec
10	10	6.03±2.5	2.99±0.02	3.35±0.2
10	30	22.41±4.4	12.34±0.1	15.73±0.5
10	50	43.93±3.7	28.04±0.2	35.97±0.9
10	70	67.62±4.0	52.43±0.5	62.09±1.2
10	90	90.79±4.1	87.05±0.6	88.19±0.6
20	10	7.32±0.9	2.99±0.02	3.40±0.3
20	30	26.49±1.1	12.43±0.2	16.32±1.1
20	50	51.35±1.9	27.86±0.4	36.29±1.2
20	70	75.96±1.2	52.08±0.3	61.30±1.3
20	90	94.46±0.3	87.00±0.5	88.30±0.9
30	10	7.35±0.7	3.00±0.03	3.75±0.4
30	30	26.24±1.5	12.36±0.09	16.08±1.4
30	50	49.78±2.3	27.85±0.3	35.83±1.2
30	70	73.18±3.8	52.37±0.3	60.63±1.2
30	90	94.36±1.0	86.64±0.5	88.20±1.0
Inf	10	6.68±0.6	2.89±0.02	3.72±0.5
Inf	30	23.58±1.6	11.21±0.02	14.52±1.3
Inf	50	47.53±1.9	24.62±0.02	31.24±1.7
Inf	70	74.30±0.9	45.75±0.03	53.60±1.5
Inf	90	93.63±0.2	78.53±0.002	82.88±1.0

分析表 3 中的学习效率值可知,在相同 N 取值下,建构推荐算法在学习完不同知识比例情况下均取得了更好的学习效率.相比较而言,随机推荐方法的学习效率要略好于基于最大相似度优先排序的推荐方法 Similarity_Rec.虽然如此,结合表 4 结果可以看出,随机推荐 Random_Rec 的平均学习知识序列关联度要低于 Similarity_Rec.对比不同 N 取值

下的学习效率结果可知,对于本文所构建的知识网络,建构推荐算法在 N=20 时表现出了最佳的学习效率,而另外 2 种算法在 N=20 时也表现出了较佳水平的学习效率.而对比表 4 结果也可以发现,虽然 N 取值较小时,平均的学习知识序列关联度会增加,但 N=20 时已与 N=10 时的最好结果相接近,但要明显好于更大的 N 取值.为此,进一步的实验中将考虑取 N=20 时,分析不同 K 取值下的推荐算法性能.

Table 4 Knowledge Sequence Correlation Values Obtained by Three Recommendation Algorithms on Different N
表 4 不同 N 下的 3 种推荐算法的学习知识序列关联度比较

N	Support_Rec	Similarity_Rec	Random_Rec
10	0.83	0.70	0.66
20	0.82	0.69	0.59
30	0.76	0.66	0.51
Inf	0.61	0.10	0.08

综合地,3 种对比推荐算法中,建构推荐方法不仅具有更优的学习效率,且学习知识序列间的平均关联度也较高,更符合用户学习新知识的需要.图 9 给出了一个典型情况下不同推荐算法随着推荐知识增加的学习效率相对值变化曲线;图 10 给出了相应的学习知识序列关联度变化曲线,为了清晰展示不同曲线之间的差异,我们将原始知识关联度数据进行约简,每隔 10 个数据取 1 个平均值,图 10 每条曲线实际展示了 100 个数据点.

如图 9 直观显示,3 条变化曲线中,支撑度最大优先的建构推荐方法在学习等量的知识节点后,所获得的知识量比例要高于另 2 种方法;而图 10 中对应的知识序列关联度变化曲线则表明,支撑度最大优先的建模推荐方法在学习经过一段时间后,学习知识序列间的关联度开始增强,并明显高于另 2 种

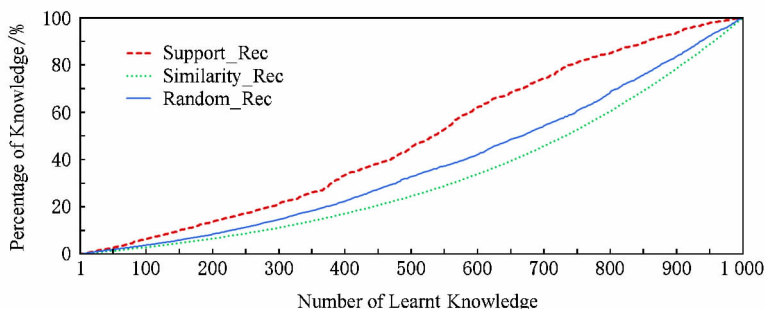


Fig. 9 Typical learning efficiency curves of compared recommendation algorithms

图 9 不同推荐算法的典型学习效率变化曲线

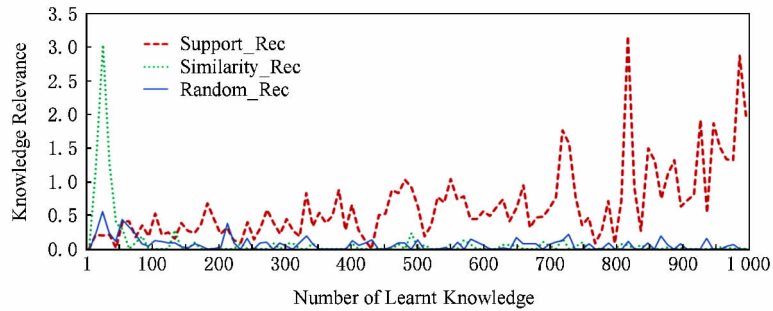


Fig. 10 Typical value tendency curves of knowledge sequence correlation obtained by compared recommendation algorithms

图 10 不同推荐算法的典型学习知识序列关联度变化趋势曲线

方法. 相对地, 随机推荐方法的学习知识序列关联度从开始到结束无明显变化, 这一结果与经验逻辑结果相一致, 也间接表明了本文所定义的相关评价指标的合理性. 对于相似度最近优先的推荐方法, 其生成的学习知识序列关联度开始较高, 但随后则表现不佳. 这一结果表明: 相似度最近优先的推荐方法在学习知识序列产生过程中, 不能很好地兼顾知识系统的整体需求, 每次过于简单地选择最相似的候选知识并非是一个有效的学习方式. 这一结论可能也将有助于进一步发展传统电子商务和社交网络的个性化推荐技术, 特别是基于内容相似性优先的推荐策略存在一定的局限性. 而基于支撑度最大优先的建构推荐策略则可以提供一些非常有价值的思想借鉴.

进一步, 我们考虑 top- K 推荐输出策略中不同 K 值对算法性能的影响, 根据前面实验结果, 固定平均最佳候选知识容量 $N=20$. 与上述实验过程稍微不同, 考虑每次为用户输出 K 个推荐知识项时, 模拟用户均匀随机地选择一个知识进行学习, 这样相同的初始背景知识也会产生非常多不同可能的学习知识序列. 为此, 实验中对于同一初始知识模拟 10 次不同结果, 同时随机选择了 20 个健康饮食知识术语作为初始背景知识. 表 5 给出了相应的实验结果, 其中性能值为 200 次模拟运行的统计结果, 随机选择的 20 个初始知识为: 低热量饮食、槟榔、柑橘、紫菜、羊肉串、辣味食物、山芋、牛奶、菊花茶、菌类、全谷物、酒酿、红枣、高盐食品、山银花、西瓜、菜子油、抗癌食品、牛排、苹果.

表 5 给出了 Support_Rec 和 Similarity_Rec 这 2 种算法的学习效率度量值. 显然地, 对于 Random_Rec, 在上述模拟实验中, 不同 K 值在理论上是等价的, 所以表 5 中没有给出重复对比结果. 同时 $K=1$ 时, 相应的结果即如表 3 结果所示. 此外, 实验中考

虑实际应用中, 不给用户造成过多的选择障碍, K 取值考虑在相对不大的个位数范围内.

Table 5 Learning Efficiency Values Obtained by Three Recommendation Algorithms on Different K with $N=20$

表 5 $N=20$ 时不同 K 值下的推荐算法的学习效率性能比较

K	Learning Efficiency/%		
	Learnt Knowledge	Support_Rec	Similarity_Rec
3	10	7.67±1.4	3.06±0.06
3	30	27.98±1.2	12.75±0.3
3	50	50.84±2.9	28.66±0.4
3	70	73.87±3.5	52.64±0.7
3	90	94.73±0.2	86.70±0.8
5	10	6.40±1.6	3.10±0.07
5	30	23.69±5.2	13.20±0.04
5	50	45.80±6.3	29.26±0.6
5	70	70.09±6.8	53.35±0.8
5	90	91.76±3.0	86.99±0.6
7	10	6.68±1.3	3.13±0.06
7	30	24.13±4.2	13.40±0.03
7	50	44.88±5.5	29.75±0.09
7	70	70.32±3.8	53.62±1.0
7	90	92.36±1.4	86.64±1.0

分析表 5 结果可知, 对于 Support_Rec, 其随着 K 值增大, 性能会有所下降, 而 $K=3$ 与 $K=1$ 时的性能结果相近. 而对于 Similarity_Rec, K 值增大时, 性能结果仅有微小的变化, 据此, 我们可以考虑在实际应用中取 $K=3$. 这样, 既给用户提供了一定的个性化选择空间, 又保证了用户学习知识序列的高效性. 同样地, 横向比较看, 基于支撑最大优先的推荐方法在不同 K 值条件下仍具有显著的学习效率优势. 进一步显示了本文提出的新模型的有益价值.

4.4 模拟推荐实例分析

进一步,我们对不同算法推荐产生的知识术语实例序列进行研究分析,表6、表7给出了2组典型的模拟实验结果,其中 $N=20, K=3$,初始知识分别选取了较为常见的知识概念“苹果”和“牛奶”。

Table 6 Knowledge Sequence Example 1 Generated by Compared Recommendation Algorithms

表6 不同推荐算法产生的知识序列实例1

Support_Rec	Similarity_Rec	Random_Rec
苹果	苹果	苹果
杨梅	香蕉	柑橘
樱桃	橙子	新鲜蔬菜
菠萝	葡萄柚	沙丁鱼
芒果	杨桃	香肠
木瓜	茯苓	燕麦
橘子	山银花	花生
柿子	冷泡茶	干果
葡萄	润喉片	糙米
西瓜	婴儿食品	奇异果
香蕉	凉拌酱油	米粉
草莓	烹饪酱油	橘子
柠檬	椰果	杨桃
柑橘	无籽水果	黄豆
新鲜蔬菜	冷鲜肉	粟米
绿叶蔬菜	纤体食品	酸牛奶
⋮	⋮	⋮

Table 7 Knowledge Sequence Example 2 Generated by Compared Recommendation Algorithms

表7 不同推荐算法产生的知识序列实例2

Support_Rec	Similarity_Rec	Random_Rec
牛奶	牛奶	牛奶
脱脂牛奶	豆奶	鱼肝油
奶酪	酸牛奶	蜂王浆
谷物	香草	银华
奶油	清茶	魔芋
白砂糖	山银花	豆苗
蜂蜜	冷泡茶	果仁
葡萄糖	婴儿食品	甜饮料
黑糖	凉拌酱油	荷兰豆
蔗糖	烹饪酱油	鲜笋
淀粉	椰果	砂仁
香精	无籽水果	苏打饼干
糖精	冷鲜肉	松饼
味精	纤体食品	黑面包
食盐	猪肝粥	乳清蛋白
沙拉酱	酸橙	苹果醋
⋮	⋮	⋮

从人工语义理解的角度分析表6、表7中术语知识可知,支撑度最大优先的建构推荐方法所产生的推荐知识序列的整体层次性和连续性要显著优先其他2种方法.如表6中结果所示,Support_Rec产生的推荐知识序列首先从“苹果”到“柑橘”一直围绕“水果”主题,然后从“新鲜蔬菜”和“绿叶蔬菜”开始推荐“蔬菜类”知识.如表7中结果所示,Support_Rec产生的推荐知识序列首先“牛奶”到“奶油”都与“奶类”密切相关,然后从“白砂糖”到“糖精”属于“糖类”相关知识,最后从“味精”到“沙拉酱”都属于“调味品”.由此可见,Support_Rec产生的推荐知识序列在局部确实具有很强的关联性,而当用户对某一小类知识掌握到一定程度后又会有效地转移推荐其他大类相关的节点知识.例如从“水果”到“蔬菜”、从“奶类”到“糖类”再到“调味品”.上述实验结果确实是令人惊奇的,也进一步表明了本文所述建构推荐模型的有效性.

相比而言,其他2种对比方法所产生的推荐知识序列在人工语义角度不能找到明显的层次性和连续性,序列知识间的语义跳跃性很大.从这一角度看,本文提出的支撑度最大优先的建构推荐过程能较好地契合人类知识的有序渐进理解过程,也将能有效地支持用户的个性化知识学习,具有较高的实用价值.

4.5 真实用户模拟实验

为了进一步分析模型的实际推荐效果,我们邀请20位学生用户开展了模拟实验分析.实验中由用户自己设置初始知识,用户每次从系统推荐的3个知识中选择1个进行学习.结合前面实验结果,我们主要模拟分析了 $N=20$ 和 $K=3$ 时Support_Rec和Similarity_Rec这2种算法的推荐性能和算法复杂度,实验结果如表8和表9所示.对比表5和表8结果可知,真实用户模拟测试结果与随机模拟

Table 8 Learning Efficiency Values Obtained by Two Recommendation Algorithms

表8 推荐算法对真实用户模拟实验的学习效率性能比较

Learnt Knowledge	Support_Rec	Similarity_Rec/%
10	7.67±1.07	3.06±0.05
30	28.00±2.56	12.81±0.37
50	51.64±3.27	28.46±0.46
70	74.31±3.18	52.44±0.65
90	94.14±1.85	86.75±0.70

Table 9 Computational Efficiency Values Obtained by Two Recommendation Algorithms with $N=20$ and $K=3$

表 9 $N=20, K=3$ 时 2 种对比推荐算法的计算效率

Percentage of Learnt Knowledge/ %	Support_Rec Average Number of Searches	Similarity_Rec Average Number of Searches
10	22.72±0.21	2.01±0.31
30	22.99±0.31	2.70±0.28
50	23.16±0.18	3.05±0.15
70	23.32±0.28	4.06±0.11
90	23.78±0.27	8.30±0.49

结果高度一致,进一步显示了本文模型的有效性和实用性。

从表 9 结果可知,大部分情况下 Support_Rec 算法平均检索次数比 Similarity_Rec 高约 20 次,与 $N=20$ 的设定相关联。而具体的平均检索次数值表明,产生 $N=20$ 个候选样本所需的检索次数大部分情况下均小于 5,而 Support_Rec 算法相比 Similarity_Rec 需要增加约 N 次的支撑度计算所需数据检索,故检索计算复杂度平均要稍高一些,但仍维持在不大的常数水平,能够满足实用要求。此外需要指出的是,当已学知识比例达到 90% 时,由于知识网络中未学习知识数量较少,需检索较多的已学知识才能产生足量的候选相关知识。

5 结束语

为更好地适应网络资源知识的个性化学习需要,解决现有推荐技术没有重点关注推荐内容的整体性和序列关联性问题,研究提出了一种基于建构主义学习理论的个性化知识推荐新方法——建构推荐模型。新模型使用知识网络建模所期望学习的目标知识系统,并提出知识网络上的候选知识可学支撑度最大优先的推荐策略。本文理论与实验结果表明了建构推荐模型的合理性和有效性,并可作为当前基于内容的推荐和协同推荐方法的有益补充,其不仅仅可简单易行地应用于个性化知识推荐领域,也可为传统电子商务和社交网络中个性化推荐技术提供思想启发。虽然如此,针对不同实际应用场景,构建最为合理的知识网络系统是一个需要考虑的关键问题,而为了避免推荐算法相续产生高度相似的推荐项,考虑引入额外的知识概念语义分析也可进一步提升建构推荐算法的实用价值^[33]。

参 考 文 献

- [1] Salehi M. Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation [J]. *Data & Knowledge Engineering*, 2013, 87(9): 130-145
- [2] Manouselis N, Vuorikari R, Van Assche F. Collaborative recommendation of e-learning resources: An experimental investigation [J]. *Journal of Computer Assisted Learning*, 2010, 26(4): 227-242
- [3] Chen Yibo. Personalized knowledge service key technology based on linked data and user ontology [D]. Wuhan: Wuhan University, 2012 (in Chinese)
(陈毅波. 基于关联数据和用户本体的个性化知识服务关键技术研[D]. 武汉: 武汉大学, 2012)
- [4] Parveen R, Jaiswal A K, Kant V. E-learning recommendation systems—A survey [J]. *Int Journal of Engineering Research and Development*, 2012, 4(12): 10-12
- [5] Khribi M K, Jemni M, Nasraoui O. Automatic recommendations for e-learning personalization based on Web usage mining techniques and information retrieval [C] //Proc of the 8th Int Conf on Advanced Learning Technologies. Piscataway, NJ: IEEE, 2008: 241-245
- [6] Sikka R, Dhankhar A, Rana C. A survey paper on e-learning recommender system [J]. *International Journal of Computer Applications*, 2012, 47(9): 27-30
- [7] Klačnja-Milićević A, Vesin B, Ivanović M, et al. e-learning personalization based on hybrid recommendation strategy and learning style identification [J]. *Computers & Education*, 2011, 56(3): 885-899
- [8] Zhao Liang, Hu Najing, Zhang Shouzhi. Algorithm design for personalization recommendation systems [J]. *Journal of Computer Research and Development*, 2002, 39(8): 986-991 (in Chinese)
(赵亮, 胡乃静, 张守志. 个性化推荐算法设计[J]. *计算机研究与发展*, 2002, 39(8): 986-991)
- [9] Huang Zhenhua, Zhang Jiawen, Tian Chunqi, et al. Survey on learning-to-rank based recommendation algorithms [J]. *Journal of Software*, 2016, 27(3): 691-713 (in Chinese)
(黄震华, 张佳雯, 田春岐, 等. 基于排序学习的推荐算法研究综述[J]. *软件学报*, 2016, 27(3): 691-713)
- [10] Gao Ming, Jin Cheqing, Qian Weining, et al. Real time and personalized recommendation on microblogging systems [J]. *Chinese Journal of Computers*, 2014, 37(4): 963-975 (in Chinese)
(高明, 金澈清, 钱卫宁, 等. 面向微博系统的实时个性化推荐[J]. *计算机学报*, 2014, 37(4): 963-975)
- [11] Pazzani M J, Billsus D. Content-Based Recommendation Systems [M]. Berlin: Springer, 2007: 325-341

- [12] Ghauth K I, Abdullah N A. Learning materials recommendation using good learners' ratings and content-based filtering [J]. Educational Technology Research and Development, 2010, 58(6): 711-727
- [13] Leng Yajun, Lu Qing, Liang Changyong. Survey of recommendation based on collaborative filtering [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(8): 720-734 (in Chinese)
(冷亚军, 陆青, 梁昌勇. 协同过滤推荐技术综述[J]. 模式识别与人工智能, 2014, 27(8): 720-734)
- [14] Mobasher B, Dai Honghua, Luo Tao, et al. Effective personalization based on association rule discovery from Web usage data [C] //Proc of the 3rd Int Symp on Web Information and Data Management. New York: ACM, 2001: 9-15
- [15] Lee C H, Kim Y H, Rhee P K. Web personalization expert with combining collaborative filtering and association rule mining technique [J]. Expert Systems with Applications, 2001, 21(3): 131-137
- [16] Bodner G M. Constructivism: A theory of knowledge [J]. Journal of Chemical Education, 1985, 63(10): 873-878
- [17] Siemens G. Connectivism: A learning theory for the digital age [J]. International Journal of Instructional Technology & Distance Learning, 2004, 2(s101): 3-10
- [18] Zhao Zhidan, Shang Mingsheng. User-based collaborative-filtering recommendation algorithms on Hadoop [C] //Proc of the 3rd Int Conf on Knowledge Discovery and Data Mining, Pisacaway, NJ: IEEE, 2010: 478-481
- [19] Rong Huigui, Huo Shengxu, Hu Chunhua, et al. User similarity-based collaborative filtering recommendation algorithm [J]. Journal on Communications, 2014, 35(2): 16-24 (in Chinese)
(荣辉桂, 火生旭, 胡春华, 等. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014, 35(2): 16-24)
- [20] Wang Peng, Wang Jingjing, Yu Nenghai. A kernel and user-based collaborative filtering recommendation algorithm [J]. Journal of Computer Research and Development, 2013, 50(7): 1444-1451 (in Chinese)
(王鹏, 王晶晶, 俞能海. 基于核方法的 User-Based 协同过滤推荐算法[J]. 计算机研究与发展, 2013, 50(7): 1444-1451)
- [21] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C] //Proc of the 10th Int Conf on World Wide Web. New York: ACM, 2001: 285-295
- [22] Deng Ailin, Zhu Yangyong, Shi Baile. A collaborative filtering recommendation algorithm based on item rating prediction [J]. Journal of Software, 2003, 14(9): 1621-1628 (in Chinese)
(邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628)
- [23] Liu Xiang, Ma Feicheng, Wang Xiaoguang. Formation and process model of knowledge networks [J]. System Engineering Theory and Practice, 2013, 33(7): 1836-1844 (in Chinese)
(刘向, 马费成, 王晓光. 知识网络的结构及过程模型[J]. 系统工程理论与实践, 2013, 33(7): 1836-1844)
- [24] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review [J]. Journal of Computer Research and Development, 2016, 53(2): 247-261 (in Chinese)
(刘致远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 53(2): 247-261)
- [25] Li Lishuang. The research of term and relation acquisition methods for domain ontology learning [D]. Dalian: Dalian University of Technology, 2013 (in Chinese)
(李丽双. 领域本体学习中术语及关系抽取方法的研究[D]. 大连: 大连理工大学, 2013)
- [26] Ren He, Zeng Juanfang. A Chinese word extraction algorithm based on information entropy [J]. Journal of Chinese Information Processing, 2006, 20(5): 40-43 (in Chinese)
(任禾, 曾隽芳. 一种基于信息熵的中文高频词抽取算法[J]. 中文信息学报, 2006, 20(5): 40-43)
- [27] Zhang Dongwen, Xu Hua, Su Zengcai, et al. Chinese comments sentiment classification based on word2vec and SVM^{perf} [J]. Expert Systems with Applications, 2015, 42(4): 1857-1863
- [28] Yang Yang, Liu Longfei, Wei Xianhui, et al. New methods for extracting emotional words based on distributed representation of words [J]. Journal of Shandong University: Natural Science, 2014, 49(11): 51-58 (in Chinese)
(杨阳, 刘龙飞, 魏现辉, 等. 基于词向量的情感新词发现方法[J]. 山东大学学报: 理学版, 2014, 49(11): 51-58)
- [29] Tang Gongbo, Yu Dong, Xun Endong. An unsupervised word sense disambiguation method based on sememe vector in HowNet [J]. Journal of Chinese Information Processing, 2015, 29(6): 23-29 (in Chinese)
(唐共波, 于东, 荀恩东. 基于知网义原词向量表示的无监督词义消歧方法[J]. 中文信息学报, 2015, 29(6): 23-29)
- [30] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space [EB/OL]. Ithaca, NY: Cornell University Library, 2013 [2013-09-07]. <https://arxiv.org/abs/1301.3781>
- [31] Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations [C/OL] // Proc of the 2013 North American Chapter of ACL on Human Language Technologies. Atlanta, Georgia, USA: The Association for Computational Linguistics, 2013: 746-751. [2016-03-15]. <http://www.aclweb.org/anthology/N13-1090>
- [32] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [C/OL] //Proc of the 2013 Int Conf on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: NIPS Foundation Inc, 2013: 3111-3119. [2016-03-15]. <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>

- [33] Li Pohan, He Zhenying, Xiang Helin. A linkage clustering based query expansion algorithm [J]. Journal of Computer Research and Development, 2011, 48(Suppl 2): 197-204 (in Chinese)

(李珀瀚, 何震瀛, 向河林. 一种基于链接聚类的查询扩展算法[J]. 计算机研究与发展, 2011, 48(增刊 2): 197-204)



Xie Zhenping, born in 1979. PhD, associate professor. Member of CCF. His main research interests include evolutionary cognitive learning, knowledge network modeling and machine vision.



Jin Chen, born in 1991. Master. His main research interests include machine learning and knowledge network modeling (842808509@qq.com).



Liu Yuan, born in 1967. Professor. Senior member of CCF. His main research interests include network security and social network (lyuan1800@sina.com).

2018 年《计算机研究与发展》专题(正刊)征文通知

——物联网安全研究进展

目前,随着物联网应用的广泛普及以及支撑技术的不断发展,其安全问题也愈发严重。物联网带来新的安全问题不仅会给广大用户和企业带来严重的损失,还会威胁国家安全与社会稳定。由于物联网设备的异构性,通信的复杂性问题给广大科技工作者提出了大量的挑战性研究课题。现阶段物联网安全问题已经引起了学术界越来越多的重视,国内外相关学者开展了大量研究,也取得了许多不错的研究成果。

为了进一步推动我国物联网安全的科学研究和工程应用,及时报道物联网安全领域国内外科技工作者所取得的最新研究成果,同时展望物联网空间安全的研究方向,《计算机研究与发展》将于 2018 年 7 月出版物联网安全专辑,欢迎相关领域的专家学者和科研人员踊跃投稿。现将专题论文征集的有关事项通知如下。

征文范围(但不限于)

- 1) **物联网网络安全问题**:物联网安全通信协议;物联网网络态势感知与网络威胁评估;物联网入侵检测与防御技术。
- 2) **物联网系统安全问题**:物联网操作系统安全;物联网恶意代码防御技术;物联网系统与固件安全更新技术;物联网设备漏洞挖掘技术。
- 3) **物联网应用与服务安全问题**:物联网应用隐私保护方案;物联网匿名与认证方案;物联网 Web 与云服务安全保护技术;物联网授权管理与访问控制技术。
- 4) **其他**:物联网设备信任问题;物联网设备测试框架;物联网设备物理安全;物联网设备软硬件攻击方法;适用于轻量级物联网设备的密码学(密钥协商、加解密、认证等)算法;跨层次的物联网安全问题;物联网在具体应用场景(智能家居、工业与公共基础设施、医疗健康等)中的安全问题。

征文要求

- 1) 论文应属于作者的科研成果,数据真实可靠,具有重要的学术价值与推广应用价值,未在国内公开发行的刊物或会议上发表或宣读过,不存在一稿多投问题。作者在投稿时,需向编辑部提交签字版版权转让协议。
- 2) 论文一律用 Word 格式排版,论文格式体例参考近期出版的《计算机研究与发展》的要求(<http://crad.ict.ac.cn/>)。论文需附通信作者的联系地址、电话或手机及 E-mail 地址。
- 3) 论文请通过期刊网站(<http://crad.ict.ac.cn/>)进行投稿,作者留言中务必注明“物联网安全 2018 专题”(否则按自由来稿处理)。

重要日期

征文截止日期:2018 年 1 月 31 日

录用通知日期:2018 年 3 月 31 日

作者修改稿提交日期:2018 年 4 月 15 日

出版日期:2018 年 7 月

特邀编委

张玉清 教授 中国科学院大学、国家计算机网络入侵防范中心 zhangyq@ucas.ac.cn

Peng LIU 教授 美国宾州州立大学 pliu@ist.psu.edu

李 晖 教授 西安电子科技大学 lihui@mail.xidian.edu.cn

孙利民 研究员 中国科学院信息工程研究所 sunlimin@iie.ac.cn

联系方式

编辑部:crad@ict.ac.cn,010-62620696,010-62600350

通信地址:北京 2704 信箱《计算机研究与发展》编辑部 邮政编码:100190