

doi:10.3969/j.issn.1000-7695.2019.07.026

基于模糊概念格的领先用户个性化知识推荐研究

张喜征, 蔡月月, 罗 文

(湖南大学工商管理学院, 湖南长沙 410082)

摘要: 针对创新社区日益增长的海量信息阻碍用户对知识进行有效获取和创造的现状, 将模糊形式概念分析 (FFCA) 理论应用于创新社区领先用户的个性化知识推荐研究。首先识别出创新社区领先用户并对其发帖内容进行文本挖掘, 得到用户-知识模糊形式背景, 然后构建带有相似度的模糊概念格对用户偏好进行建模, 最后基于模糊概念格和协同过滤的推荐算法为领先用户提供个性化知识推荐有序列表。以手机用户创新社区为例, 验证基于 FFCA 的领先用户个性化知识推荐方法有助于满足用户个性化知识需求, 促进用户更好地参与社区知识创新。

关键词: 模糊形式概念分析; 知识推荐; 领先用户; 协同过滤; 创新社区

中图分类号: G203; F224; G301

文献标志码: A

文章编号: 1000-7695 (2019) 07-0183-07

Research on Personalized Knowledge Recommendation for Leading Users Based on Fuzzy Concept Lattice in Innovation Community

Zhang Xizheng, Cai Yueyue, Luo Wen

(Business school of Hunan University, Changsha 410082, China)

Abstract: The increasing mass of information in innovative communities is hindering users effective acquisition and creation of community knowledge, so the fuzzy formal concept analysis (FFCA) theory is applied to the research of personalized knowledge recommendation for leading users in an innovative community in this paper. First of all, the user-knowledge fuzzy formal context is obtained by identifying the leading users of the innovative community and text mining the content of their posts. Then, a fuzzy concept lattice with similarities is constructed to model user preference. Finally, a recommendation algorithm based on fuzzy concept lattice and collaborative filtering is proposed to provide personalized knowledge recommendation for leading users. Taking the innovative community of mobile phone users as an example, the paper uses the algorithm proposed in this paper to get an ordered list of personalized knowledge recommendation for leading users, so as to improve their enthusiasm to participate in community knowledge innovation.

Key words: fuzzy formal concept analysis; knowledge recommendation; leading users; collaborative filtering; innovation community

1 研究背景

互联网的飞速发展使人类进入知识爆炸时代, 由此引发了“知识泛滥”“知识迷航”等问题。正如创新社区中渐增的海量信息正在加重用户的认知负担, 大量无效内容阻碍了用户对知识的有效获取和创造, 用户往往需要花费大量的时间和精力去寻找和筛选其需要的信息和知识, 在一定程度上阻碍了用户参与社区知识创新。个性化知识推荐因为能够依据用户的个性化需求主动为其提供合适的知识从而有效缓解上述问题^[1], 提高企业知识服务能力, 受到学者们的广泛关注。

目前常见的知识推荐方法主要有以下几类: (1) 基于社会网络分析 (social network analysis, SNA) 的知识推荐。例如, 赵英等^[2]基于 SNA 构建某科研团队的复杂网络, 进而设计知识共享推荐系统, 研究了组织内部网络结构对知识共享和推荐的影响; Fan 等^[3]在 SNA 理论的基础上, 提出一种面向知识管理系统用户的基于中心性分析和凝聚子群分析的知识推荐算法, 以提高用户的满意度。(2) 基于语义扩展的知识推荐。如董兵等^[4]、张鼎等^[5]均采用基于语义扩展的个性化推荐方法, 利用扩展激活模型建立用户兴趣偏好档案, 为用户提供个性化的

收稿日期: 2018-06-01, 修回日期: 2018-09-05

基金项目: 国家自然科学基金面上项目“跨界融合下企业知识网络的知识配置、整合与创新研究”(71571066)

文献推荐服务。(3) 基于情境感知的知识推荐。曾子明等^[6]在传统协同过滤算法基础之上,使用贝叶斯方法学习用户在不同情境下对各资源类别的偏好,提出符合用户当前情境的个性化信息推荐方法。不同于曾子明等^[6]的研究,Song等^[7]强调情景是动态变化的,基于时间序列的变化描述情景感知的过程,并使用甘特图描述知识推荐过程中的时序关系,给出一种面向企业生产员工的知识推荐方法。

以上各种知识推荐方法都有其各自的适用条件和优缺点,而随着形式概念分析(formal concept analysis, FCA)理论(又名概念格理论)在数据挖掘、本体构建以及知识发现等领域的广泛应用,有不少学者发现基于概念格理论的个性化知识推荐有其独特的优势^[8-10]。Boucherryan等^[11]首次在协同过滤推荐系统中使用形式概念分析描述用户和项目间的二元关系,并将概念格作为用户评分矩阵的索引,实验结果表明在保证准确性的前提下,该方法能够大幅度减少寻找最近邻居的工作量。李宏涛等^[12]则以概念格的形式表达知识,构造了用户属性概念格和用户社交概念格,提出以概念格知识为指导的随机游走推荐算法,对社交网络用户进行更为精准的top-K朋友推荐。由于传统的形式概念分析只能反映对象与属性之间的精确关系,而客观世界中用户与知识间的关系很多时候是不能用“是”或“非”的概念进行确定性描述的,因此,一些学者开始将引入了模糊集理论的模糊形式概念分析(fuzzy formal concept analysis, FFCA)应用于个性化推荐中^[13]。Fang等^[14]就提出一种基于模糊形式概念分析的电影个性化推荐算法,并通过实验证明了该算法在准确性和时间效率方面都具有不错的表现。但目前有关利用FFCA进行个性化知识推荐研究的文献还较少,有待学者们进行更深入的研究。

研究表明,不同类型的用户群体对创新社区贡献的创新价值大小不同^[15]。作为创新社区内外部知识传递的中介,领先用户帮助社区不断获取外部新知识,并将其内化成新的内部知识,实现知识的转移和创新^[16],在企业创新社区的知识创新中起主导作用。然而创新社区的信息剧增给用户带来了知识困扰。针对目前国内外的相关研究大多集中在领先用户的特征与作用、创新过程及识别方法等^[17-19],鲜有学者对其进行个性化知识推荐研究的现状,为了更好地促进创新社区领先用户在企业开放式创新中的知识获取与创造,发挥其“意见领袖”的领导作用,本研究综合运用模糊形式概念分析和协同过滤推荐技术,探讨创新社区领先用户个性化知识推荐机制,并给出相应的个性化知识推荐算法,

通过模糊概念格构建用户兴趣模型,利用模糊概念的相似度及协同过滤的思想,为领先用户推荐符合其兴趣偏好的、有先后顺序的个性化知识,而不是任何用户都需要的知识。

2 模糊形式概念分析(FFCA)

形式概念分析理论最早由德国数学家 Wille^[20]于1982年提出,历经30多年的发展,已成为概念数据分析和知识处理的有效工具之一。传统的概念格由二值化形式背景得到,反映对象与属性间的精确关系^[21]。但在客观世界中,很多信息是模糊不确定的,因此,一些学者将模糊集理论引入形式概念分析中,形成模糊形式概念分析^[22-24]。依据文献^[24]给出相关概念如下:

定义1: 一个模糊形式背景是一个三元组 $F = (U, A, I = \varphi(U \times A))$, 其中 U 为所有对象的集合, A 为所有属性的集合, I 是一个在域 $U \times A$ 上定义的模糊集, 每个关系中的元素 (x, a) 均有一个隶属度 $\mu(x, a)$, $0 \leq \mu(x, a) \leq 1$ 。

定义2: 给定模糊形式背景 $F = (U, A, I = \varphi(U \times A))$ 及阈值 $T \in [0, 1]$, 对于对象集合的子集 $X \subseteq U$, 定义 $X^\uparrow = \{a \in A / \forall x(x \in X \wedge \mu(x, a) \geq T)\}$; 对于属性集合的子集 $M \subseteq A$, 定义 $M^\uparrow = \{x \in U / \forall a(a \in M \wedge \mu(x, a) \geq T)\}$ 。具有阈值 T 的模糊形式概念定义为 $C = (\varphi(X), M)$, 其中 $X^\uparrow = M$ 且 $M^\uparrow = X$, $\varphi(X)$ 是定义在 X 上的一个模糊集, 每个对象 $x \in X$ 均有一个隶属度 μ_x 定义为: $\mu_x = \min_{m \in M} \mu(x, a)$ 。 $\mu(x, a)$ 表示形式背景中对象 x 与属性 a 间的隶属度。如果 $m = \{\}$, 则对任意 x 均有 $\mu_x = 1$; 相应的, 称 X 为 C 的外延, M 为 C 的内涵。

定义3: 设 $C_1 = (\varphi(X_1), M_1)$ 和 $C_2 = (\varphi(X_2), M_2)$ 为 $F = (U, A, I = \varphi(U \times A))$ 上的两个模糊形式概念, 定义 $C_1 \leq C_2 \Leftrightarrow \varphi(X_1) \subseteq \varphi(X_2) (\Leftrightarrow M_2 \subseteq M_1)$, 此时, 称 C_1 为 C_2 的子概念, C_2 为 C_1 的超概念。所有模糊概念的集合 $CS(F)$ 及 $CS(F)$ 上的偏序关系“ \leq ”构成了阈值 T 下模糊形式背景 $F = (U, A, I = \varphi(U \times A))$ 的概念格。

定义4: 因为模糊形式概念的模糊度使用其对象的隶属度来表达, 因此, 一个模糊形式概念的基数定义为: $|C| = |\varphi(X)|$ 。

定义5: 模糊形式概念 $C_1 = (\varphi(X_1), M_1)$ 与其子概念 $C_2 = (\varphi(X_2), M_2)$ 之间的相似度定义为: $S(C_1 C_2) = \frac{|\varphi(X_1) \cap \varphi(X_2)|}{|\varphi(X_1) \cup \varphi(X_2)|}$ 。

表 1 展示了一个模糊形式背景，其中 $U = \{o_1, o_2, o_3, o_4\}$, $A = \{a_1, a_2, a_3, a_4\}$, I 描述了 U 中对象拥有 A 中属性的隶属度集合。图 1 是表 1 对应模糊概念格 (fuzzy concept lattice, FCL) 的 Hassae 图, 展示了形式概念的格结构, 格上的每个节点都是一个概念, 由内涵和外延两部分组成。概念格通过 Hassae 图生动简洁地体现了概念间的泛化和特化关系。

表 1 模糊形式背景示例

U/A	a_1	a_2	a_3	a_4
o_1	1.0	0	0.4	0.8
o_2	0.7	1.0	0	0
o_3	0	1.0	0.4	0.4
o_4	0	0.6	0.8	1.0

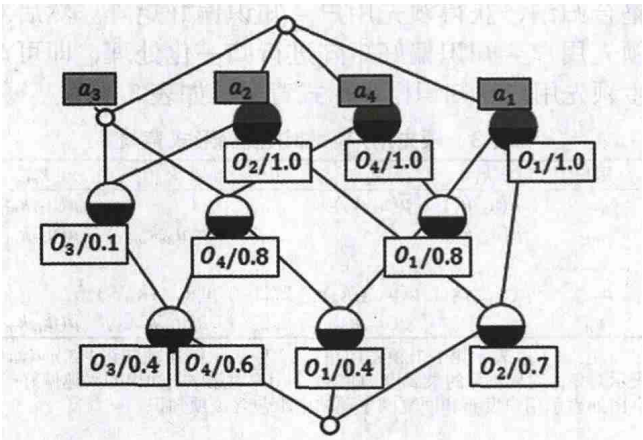


图 1 表 1 对应模糊概念格的 Hassae 图

3 基于 FFCA 的领先用户个性化知识推荐机制

在创新社区中, 原本孤立分散的用户摆脱了时间和地域的限制, 能够通过交流技术信息和使用经验、分享创新成果等自由交换和获取知识, 其他成员也都可以对创新成果提出意见和建议, 通过相互交流讨论完成显性知识和隐性知识的有效转化和互动, 形成知识创新^[25]。国外诸如戴尔公司的 Dell IdeaStorm、BMW 公司的 M Power community、Nike 等体育产品在线社区 Niketalk 等, 国内如小米社区、海尔 HOPE 创新平台和华为花粉俱乐部等, 都是用户创新社区的典型代表。通过对国内主流在线用户创新社区的观察发现, 用户所发表的内容分别处于不同的版块中, 这些版块主要依据产品类型进行区分; 而在每一个基于产品类型区分的版块中, 用户发表的内容按照主题功能的不同又被划分为产品功能讨论、产品使用建议、问题反馈和创意分享等几大模块, 这些内容反映了用户的产品知识、产品的使用经验、产品存在的问题、产品需求以及用户的创新动机和创新能力等, 同时也体现了用户发帖的内容信息特征。领先用户是企业社区的创新主体, 若对领先用户的意见及需求进行重点考虑并进行个性化知识推荐, 更易获得突破性创意, 有助于提高企业产品创新能力。因此, 针对不同版块下的领先用户, 挖掘其兴趣偏好并为其推荐可能感兴趣的个性化知识, 充分发挥用户创新的潜能, 以实现其创新价值的最大化, 对于社区的创新活动具有重要意义。

本文所提出的基于 FFCA 的领先用户个性化知识推荐过程如图 2 所示。

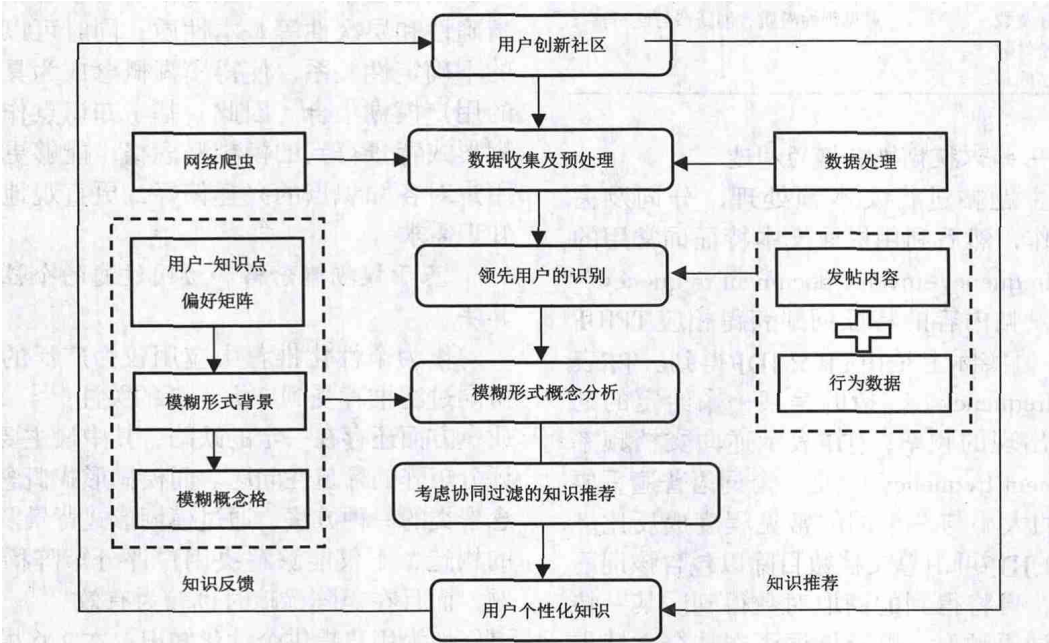


图 2 创新社区领先用户个性化知识推荐过程

3.1 领先用户的识别

以往的研究中,学者们主要基于用户发帖的内容信息来识别创新社区中的领先用户,比较有代表性的是网络志 (netnography) 方法^[26-27]。Belz 等^[27]认为创新社区中的领先用户应当具备以下基本特征:

(1) 具有领先于普通用户的需求; (2) 通常对现有的产品难以满足; (3) 具备丰富的产品知识; (4) 具备丰富的产品使用经验; (5) 拥有强烈的创新参与感; (6) 具有作为意见领袖的潜质。通过对创新社区中用户发布的主题内容进行综合评估,判断用户是否满足上述六大特征,最终确定该用户是否为领先用户。但是使用这种方法需要消耗大量的时间和人工成本,而且对用户的评判受到研究人员自身认知水平和情感因素的影响,因而存在一定的缺陷。本文通过观察发现,主流创新社区的用户资料中均存在着大量易于提取的行为数据信息,如积分、主题数、等级、贡献值、粉丝数和访问量等等,这些信息在某种程度上客观地反映了用户的领先特征且易于获取。为了更好地识别领先用户,本文将用户的行为数据作为领先用户识别的一部分,参考文献^[28],最终得到如表 2 所示的领先用户识别指标;基于表 2 的识别指标,结合网络志方法发现用户特征,完成对创新社区中领先用户的识别。

表 2 基于用户行为数据的领先用户识别指标

行为特征	识别指标	体现的领先用户特征
活跃度	主题数	创新参与性、产品知识、产品使用经验
	回复数	
	论坛积分	
社区影响力	等级	用户创新能力、意见领袖潜质
	贡献值	
	精华帖数量	
用户关系	好友数	意见领袖潜质、创新参与性
	粉丝数	
	访问量	

3.2 领先用户 - 知识模糊概念格的构造

首先,对主题帖进行文本预处理、分词及去除停用词等操作,然后利用提取文本特征词常用的 TF-IDF (term frequency-inverse document frequency) 算法提取用户发帖内容的特征词并获得相应 TFIDF 值^[29]。TFIDF 值实际上是由 $TF \times IDF$ 得到。TF 表示词频 (term frequency), 指的是某一个给定的词语在该文档中出现的频率; IDF 表示逆向文档频率 (inverse document frequency), 是一个词语普遍重要性的度量,它的大小与一个词的常见程度成反比。某一特定词语的 IDF 可由总文档数目除以包含该词语之文档的数目,再将得到的商取对数得到。某一特定文档内的高词语频率,以及该词语在整个文档集合中的低文档频率,可以产生出高权重的 TFIDF。

因此,TF-IDF 算法倾向于过滤掉常见的词语,保留重要的即能够代表文本特征的特征词。

用户 u_j 发帖内容的特征词及其 TFIDF 值的具体计算公式如下:

(1) $TF_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}$ 。其中: n_{ij} 是词语 t_i 在用户 u_j 的发帖文档中出现的次数; $\sum_k n_{kj}$ 则是在用户 u_j 的发帖文档中所有字词的出現次数之和。

(2) $IDF_i = \log \frac{|D|}{|\{j: t_i \in D_j\}|}$ 。其中: $|D|$ 表示全体用户的发帖文档数; $|\{j: t_i \in D_j\}|$ 表示全体用户发帖文档中包含词语 t_i 的文档数目。

(3) $TFIDF_{ij} = TF_{ij} \times IDF_i$ 。

按照上述公式计算得到所有用户的特征词及其 TFIDF 值,通过设置阈值提取排名靠前的特征词并整合归纳,获得领先用户 - 知识偏好矩阵。然后对领先用户 - 知识偏好矩阵进行归一化处理,即可产生领先用户 - 知识模糊形式背景,如表 3 所示。

表 3 领先用户 - 知识模糊形式背景

用户	k_1	k_2	k_{m-1}	k_m
μ_1	$\mu(u_1, k_1)$	$\mu(u_1, k_2)$		$\mu(u_1, k_m)$
μ_2	$\mu(u_2, k_1)$		$\mu(u_2, k_{m-1})$	$\mu(u_2, k_m)$
.....				
μ_{n-1}	$\mu(u_{n-1}, k_1)$	$\mu(u_{n-1}, k_2)$	$\mu(u_{n-1}, k_{m-1})$	
μ_n		$\mu(u_n, k_2)$	$\mu(u_n, k_{m-1})$	$\mu(u_n, k_m)$

注: 1) μ_i 表示第 i 个领先用户; 2) k_j 表示第 j 项知识; 3) $\mu(u_i, k_j)$ 表示对象 u_i 对属性 k_j 的隶属度,即第 i 个用户对第 j 项知识的兴趣偏好; 4) 空白处表示用户发布和回复的主题帖中不包含该项知识

本文参考比较经典的 Fuzzy Lattice 算法^[30], 基于领先用户 - 知识模糊形式背景构建领先用户 - 知识模糊概念格。模糊概念格具有概念格的完备性、精确性和层次性等基本性质,同时可以描述概念间的不确定性关系,使得模糊概念成为具有特定兴趣的用户模糊集合。因此,基于知识点作为属性的模糊形式背景得到的模糊概念格,能够更准确地刻画用户对各知识点的兴趣偏好,更直观地反映用户的知识需求。

3.3 基于模糊概念格和协同过滤的个性化知识推荐算法

作为个性化推荐中应用较为广泛的推荐算法,协同过滤推荐受到广大学者的关注^[31],但其在推荐质量方面还存在一定的缺陷,其中最主要的就是用户评分矩阵的稀疏性问题。而模糊形式概念分析作为概念聚类的一种方法,通过模糊形式背景及模糊概念格的构造,不仅能够解决用户评分矩阵稀疏性高的问题,而且在矩阵稀疏时也较为有效^[14]。因此,为了更好地为用户提供个性化知识,本文将模糊形式概念分析理论与协同过滤思想结合起来,提出一种基于模

糊概念格和协同过滤的用户个性化知识推荐算法——FCLCF 算法。该算法的具体实现过程如下：

输入：用户 - 知识偏好矩阵。

输出：各用户的个性化推荐列表。

步骤 1：对用户 - 知识偏好矩阵设置阈值及归一化处理，获得模糊形式背景 F 。

步骤 2：依据模糊形式背景 F 及定义 8 中的相似度计算公式，使用上文中的模糊概念格构造算法构建带有相似度的模糊概念格 L 。

步骤 3：遍历模糊概念格 L ，寻找包含某一推荐对象 u_i 的概念节点 c 。

步骤 4：令 $O(c)$ 为 c 中包含的推荐对象集合；相反的， $\tilde{O}(c) = \{u_1, u_2, \dots, u_h\}$ 。则表示除推荐对象集合外 c 中包含的剩余对象集合。

步骤 5：如果 $\tilde{O}(c) \neq \emptyset$ ，那么对于 $\forall u_k \in \tilde{O}(c)$ ，计算 $\mu(u_k)$ 与 $\mu(u_i)$ 的差值： $d_{ki} = |\mu(u_k) - \mu(u_i)|$ ，并对 d_{ki} 由大到小排序并建立相应的索引。

步骤 6：按照步骤 5 中的索引依次考虑包含对象 u_k 且与节点 c 相似度更近的子节点 $c' : u_k \in O(c')$ ，用 $A(c') = \{a'_1, a'_2, \dots, a'_j\}$ 表示该节点的知识点集。若 $\forall a_k \in A(c') \wedge a_k \notin A(c)$ ，将知识点 a_k 推荐给用户 u_i 并按推荐的先后顺序进行排列，更新 $A(c) : A(c) = A(c) \cup a'$ 。当 $\tilde{O}(c)$ 中的元素全部被考虑到，输出对用户 u_i 的推荐列表。

步骤 7：如果 $\tilde{O}(c) = \emptyset$ ，考虑节点 c 的父节点： $P(c) = \{c_1, c_2, \dots, c_s\}$ （注意 $\forall i = 1, 2, \dots, s$ ，都有 $u_i \in G(c_i) \wedge u_i \in G(c)$ ，其中 $G(c_i)$ 代表节点 c_i 的对象集合）。对于 $\forall c_k \in P(c)$ ，转至步骤 4 继续操作。

步骤 8：重复步骤 3 至步骤 7，对每个用户都做同样的推荐，待所有用户的推荐列表都输出，算法终止。

4 实例研究及分析

领先用户不仅有助于企业把握市场需求趋势、协助企业进行产品开发，而且作为不同用户群体之间的桥梁，能够加速新产品的知识扩散^[32]，因此领先用户积极参与知识创造对于企业创新至关重要。然而，社区中大量的冗余信息阻碍了用户的知识吸收与创新，为了更好地发挥其领先作用，企业需要充分调动领先用户的参与积极性。鉴于此，本文以国内某手机用户创新社区为例，实证分析如何识别领先用户并为其提供有针对性的知识推荐，以更好地促进其参与社区知识创新。

4.1 数据获取及预处理

本研究实例选取某手机用户创新社区中的最新款 X 型号手机版块。用户通过发帖或回帖的形式在该版块发布有关对 X 型号手机新功能的看法、使用该款手机时的一些小技巧、自己的使用经验以及想要咨询的问题或解答其他用户提出的问题等内容。利用 Python 编程语言抓取社区论坛中 X 型号手机版块中最近一年来用户发布和回复的主题帖及主题数、回复数、论坛积分、等级、贡献值、精华帖数量、好友数、粉丝数和访问量等数据。此次共抓取 1 512 条记录，经过筛选，剔除无效数据 53 条，最终原始数据为 1 459 条。

4.2 创新社区领先用户

依照上文领先用户识别方法，基于原始数据在 X 型号手机版块中共筛选出 6 位满足领先用户基本特征的用户，其行为数据如表 4 所示。在识别过程中发现，相对于普通用户，领先用户普遍积极参与社区交流与讨论，发表的主题帖大多都是精华帖，积分和等级相对较高，粉丝数和访问量也都较大，同时较高的贡献值意味着其对创新社区的贡献较多。

表 4 某手机用户创新社区领先用户的行为数据

用户	主题 / 个	回复 / 个	积分	等级	贡献值	精华帖 / 个	好友 / 个	粉丝 / 个	访问量 / 次
用户 1	77	438	79 528	LV.5	4 850	45	78	5 416	44 562
用户 2	25	262	12 851	LV.3	2 820	20	56	3 434	34 562
用户 3	17	486	12 299	LV.2	1 550	12	35	1 537	12 567
用户 4	24	136	12 662	LV.3	2 800	18	58	3 501	34 370
用户 5	13	492	7 890	LV.2	1 270	11	26	2 600	10 900
用户 6	18	149	13 424	LV.3	2 580	15	40	3 124	25 508

4.3 创新社区领先用户 - 知识点模糊概念格

对以上识别出来的 6 位领先用户主题帖进行结巴分词和去除“手机”“用户”及其他无意义的停用词等操作，使用 TF-IDF 算法提取各领先用户 TFIDF 值排名前 10 的特征词并整合归纳，获得用户对知识的偏好矩阵，如表 5 所示。其中： k_1 表示系

统配置方面的创新知识，主要包括与内存、GPU、高配、处理器等有关的知识； k_2 表示新兴技术方面的创新知识，主要包括与 VoLTE、NFC、指纹识别、防水等有关的知识； k_3 表示基本功能方面的创新知识，主要包括屏幕、双摄、耳机、电池等有关的知识； k_4 表示系统操作方面的创新知识，主

要包括与设置、识别、Dock、重启等有关的知识节点； k_5 表示智能生态方面的创新知识，主要包括与创新、生态链、平台、场景式体验等有关的知识节点。由此可以看出，领先用户们普遍对手机的系统配置知识较为关注，这是用户创新的热点，对有关系统操作、基本功能等知识的密切关注展示了用户创新的核心知识所在；而像新兴技术、生态模式等有关知识也开始引起用户的兴趣，反映了该款手机主要的创新需求点及未来发展趋势。

表 5 某手机用户创新社区的领先用户 - 知识偏好矩阵

用户	k_1	k_2	k_3	k_4	k_5
μ_1	0.488 8	0.662 3		0.331 1	
μ_2	0.733 2		0.559 8	0.788 4	0.275 9
μ_3	0.788 4	0.417 9	0.165 6	0.591 3	
μ_4	0.551 9		0.756 9		
μ_5	0.512 5	0.496 7	0.788 4	0.788 4	0.551 9
μ_6	0.528 2		0.630 7		0.654 4

对领先用户 - 知识偏好矩阵进行归一化处理，过滤掉评分低于0.6的指标数据，即可获得领先用户 - 知识模糊形式背景，如表 6 所示。

表 6 某手机用户创新社区的领先用户 - 知识模糊形式背景

用户	系统配置	新兴技术	基本功能	系统操作	智能生态
U_1	0.62	0.84		0.42	
U_2	0.93		0.71	1.00	0.35
U_3	1.00	0.53	0.21	0.75	
U_4	0.70		0.96		
U_5	0.65	0.63	1.00	1.00	0.70
U_6	0.67		0.80		0.83

基于领先用户 - 知识模糊形式背景构造带有相似度的模糊概念格，如图 3 所示。由图 3 可以看到，基于领先用户 - 知识模糊形式背景生成的模糊概念格共有 7 个概念节点，分别为 c_1 、 c_2 、 c_3 、 c_4 、 c_5 、 c_6 、 c_7 。其中：最顶端上黑下白的节点 c_1 表示该节点拥有属性 K_1 并且包含所有的对象，即所有领先用户都对系统配置相关知识感兴趣；最底端上白下黑的节点 c_7 拥有所有的属性，表示用户 U_5 对 K_1 至 K_5 这五大主题知识都感兴趣；剩下的节点则是拥有部分对象和属性的普通节点，表示对其中某些知识感兴趣的集合。依据定义 8 中概念相似度公式计算得到各相邻节点间的相似性。相似度的值越大，表明两个概念节点越相似。如节点 c_1 与子概念节点 c_2 、 c_3 、 c_5 间的相似度的值分别为 0.51、0.60 和 0.27，相较于 c_2 与 c_5 、 c_1 与 c_3 各自代表的用户集合的兴趣偏好更相近。

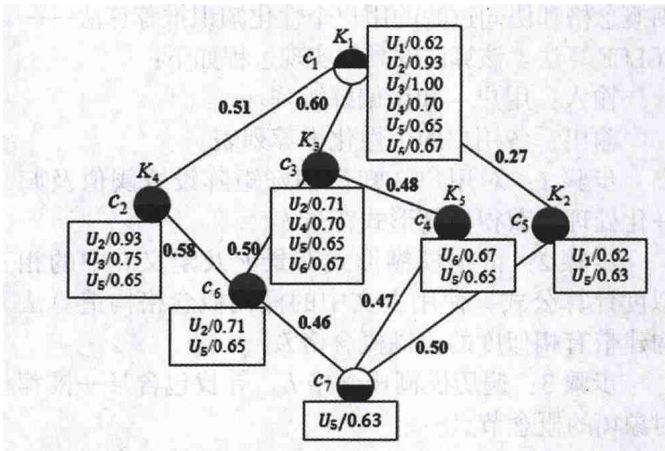


图 3 某手机用户创新社区的领先用户 - 知识模糊概念格

4.4 基于 FCLCF 算法的领先用户个性化知识推荐

使用 FCLCF 算法对图 3 中的领先用户进行知识推荐。这里以用户 U_4 为例，个性化知识推荐顺序为： $K_1 \rightarrow K_5 \rightarrow K_2$ 。即优先对领先用户 U_4 推荐有关系统操作方面的创新知识，如主题的设置、模式的识别等；然后推荐智能生态方面的创新知识，如智能硬件、生态链等；最后推荐新兴技术相关创新知识，如近距离无线通信技术 NFC、指纹识别技术等。同样的，基于 FCLCF 算法对其他各用户进行推荐，最终对各领先用户进行个性化知识推荐的结果如表 7 所示。从表 7 可以看出，使用本文提出的推荐算法，不仅能够对具有不同兴趣偏好的领先用户推荐有针对性的知识，而且为用户推荐的知识列表是有先后顺序的，这是因为领先用户对不同知识点的兴趣偏好是不同的，应优先为其推荐其更可能感兴趣的知識。

表 7 某手机用户创新社区领先用户的知识推荐列表

用户	知识推荐序列
U_1	$K_5 \rightarrow K_4 \rightarrow K_3$
U_2	$K_5 \rightarrow K_2$
U_3	$K_3 \rightarrow K_2 \rightarrow K_5$
U_4	$K_4 \rightarrow K_5 \rightarrow K_2$
U_6	$K_2 \rightarrow K_4$

为创新社区的领先用户进行个性化的知识推荐，一方面缓解了用户知识迷航问题，满足其个性化知识需求，有助于领先用户更好地全面了解外部市场需求从而进行更丰富的知识创造；另一方面有利于促进社区知识的传播与共享，实现企业创新社区与用户之间的良性互动，加快企业产品创新知识的产生。

5 结论

针对创新社区日益增长的海量信息阻碍了用户对知识进行有效获取和创造的现状，本文研究了面向创新社区领先用户的个性化知识推荐过程。（1）

通过文本挖掘及用户-知识模糊概念格的构造,将各用户对不同知识点的兴趣偏好进行显式呈现,并提出基于模糊概念格和协同过滤的推荐算法——FCLCF 算法,改善传统协同过滤推荐中由于数据稀疏性带来的问题,从而更好地为创新社区的领先用户进行个性化知识推荐。(2)同时本文首次将FFCA 应用于创新社区领先用户的个性化知识推荐中,进一步丰富了模糊形式概念分析理论的相关应用研究。

本文的不足之处在于,只针对社区内领先用户进行创新知识推荐研究,且仅对单个用户创新社区进行实证分析,数据来源单一。在后续研究中,将进一步扩大数据来源,充分发挥创新社区的作用与价值,更好地提升企业产品创新能力。

参考文献:

- [1] 易明,邓卫华,徐佳.社会化标签系统中基于组合策略的个性化知识推荐研究[J].情报科学,2011(7):1093-1097.
- [2] 赵英,袁莉.基于内部网络结构的知识推荐系统设计研究[J].情报杂志,2012,31(1):161-165.
- [3] FAN B, LIU L, LI M, et al. Knowledge recommendation based on social network theory [C] //IEEE.Proceedings of 2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE). New York: IEEE, 2009:1-3.
- [4] 董兵,吴秀玲.基于语义扩展的个性化知识推荐系统[J].图书馆学研究,2008(11):44-49.
- [5] 张翊,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.
- [6] 曾子明,李鑫.移动环境下基于情境感知的个性化信息推荐[J].情报杂志,2012,31(8):166-170.
- [7] SONG J, ZHAN H F, YU J H, et al. Enterprise knowledge recommendation approach based on context-aware of time-sequence relationship [J]. Procedia Computer Science, 2017,107:285-290.
- [8] ZOU C F, ZHANG D Q, WAN J F, et al. Using concept lattice for personalized recommendation system design [J]. IEEE Systems Journal, 2017,11(1):305-314.
- [9] 唐晓波,伍星蓉.基于形式概念分析的面向药物适应症的领域本体构建方法[J].情报杂志,2017(12):159-165.
- [10] GAO J F, QU J Y, XIN ZH. Towards on the MOOCs knowledge discovery based on concept lattice [J]. International Journal of Multimedia & Ubiquitous Engineering, 2015, 10(5):287-296.
- [11] BOUCHERRYAN P D, BRIDGE D. Collaborative recommending using formal concept analysis [J]. Knowledge-Based Systems, 2006,19(5):309-315.
- [12] 李宏涛,何克清,王健,等.基于概念格和随机游走的社交网朋友推荐算法[J].四川大学学报(工程科学版),2015,47(6):131-138.
- [13] BELOHLAVEK R. Concept lattices and order in fuzzy logic [J]. Annals of Pure & Applied Logic, 2004,128(1):277-298.
- [14] FANG P, ZHENG S. A research on fuzzy formal concept analysis based collaborative filtering recommendation system [C] //IEEE. International Symposium on Knowledge Acquisition and Modeling. New York: IEEE, 2009:352-355.
- [15] 唐洪婷,李志宏,秦睿.基于超网络的大众协同创新社区用户知识模型研究[J].管理学报,2017,14(6):859-867.
- [16] 丁志慧,刘伟.新产品开发中创新社区客户知识管理研究[J].科技进步与对策,2016,33(7):133-138.
- [17] VON HIPPEL E. Lead users: a source of novel product concepts [J]. Management Science, 1986,32(7):791-805.
- [18] LÜTHJE C. Characteristics of innovating users in a consumer goods field: an empirical study of sport-related product consumers [J]. Technovation, 2004, 24(9):683-695.
- [19] 杨波,刘伟.基于应用扩展和网络论坛的领先用户识别方法研究[J].管理学报,2011,8(9):1353-1358.
- [20] WILLE R. Restructuring lattice theory: an approach based on hierarchies of concepts [C] //RIVAL. NATO advanced study institutes series. Dordrecht: Springer, 1982,83:445-470.
- [21] GANTER B, WILLE R. Formal concept analysis, mathematical foundations [M]. Berlin: Springer, 1999.
- [22] QUAN T T, HUI S C, CAO T H. A fuzzy FCA-based approach for citation-based document retrieval [C] //IEEE. 2004 IEEE Conference on Cybernetics and Intelligent Systems. New York: IEEE, 2004:578-583.
- [23] QUAN T T, HUI S C, CAO T H. A fuzzy FCA-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data [C] //CLA. Proceedings of the CLA 2004 International Workshop on Concept Lattices and Their Applications. Ostrava: CLA, 2004:1-12.
- [24] MAJIDIAN A, MARTIN T, CINTRA M E. Fuzzy formal concept analysis and algorithm [C] //UKCI. Proceedings of the 11th Annual Workshop on Computational Intelligence. Manchester: UKCI, 2011:1-7.
- [25] VON HIPPEL E. Democratizing innovation [M]. Cambridge: The MIT Press, 2005:26.
- [26] BREM A, BILGRAM V. The search for innovative partners in co-creation: identifying lead users in social media through netnography and crowdsourcing [J]. Journal of Engineering and Technology Management, 2015,37:40-51.
- [27] BELZ F M, BAUMBACH W. Netnography as a method of lead user identification [J]. Creativity & Innovation Management, 2010,19(3):304-313.
- [28] 原欣伟,杨少华,王超超,等.基于用户特征抽取和随机森林分类的用户创新社区领先用户识别研究[J].数据分析与知识发现,2017,1(11):62-74.
- [29] 张瑾.基于改进 TF-IDF 算法的情报关键词提取方法[J].情报杂志,2014(4):153-155.
- [30] BELOHLAVEK R, BAETS B D, OUTRATA J, et al. Computing the lattice of all fixpoints of a fuzzy closure operator [J]. IEEE Transactions on Fuzzy Systems, 2010, 18(3):546-557.
- [31] 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.
- [32] KRATZER J, LETTL C, FRANKE N, et al. The social network position of lead users [J]. Journal of Product Innovation Management, 2016,33(2):1111-1117.

作者简介: 张喜征(1968—),男,湖南株洲人,教授,博士,主要研究方向为企业管管理、知识管理;蔡月月(1994—),通信作者,女,河南信阳人,硕士研究生,主要研究方向为数据挖掘与知识管理;罗文(1993—),女,湖北天门人,硕士研究生,主要研究方向为数据挖掘与知识管理。