

基于弱概念相似度的组推荐方法

范敏^{1,2}, 张洁^{1,2}, 李金海^{1,2}

(1. 昆明理工大学数据科学研究中心, 昆明 650500; 2. 昆明理工大学理学院, 昆明 650500)

摘要: 网络数据下的概念认知与知识发现是网络背景下机器学习和人工智能的重要研究方向, 已被引入到推荐系统研究中。现有的基于概念格的推荐方法忽视了节点之间的网络结构关系, 同时构造概念格的效率低且构建概念集合的约束条件较严, 在大规模的社交网络中难以实现。为解决这些问题, 本文在网络形式背景的框架下, 综合复杂网络的拓扑结构和弱概念相似度, 提出了基于弱概念相似度的组推荐算法。首先, 定义属性度、属性密度来描述属性的重要性, 通过改进的节点影响力来确定专家节点; 其次, 利用专家节点划分社区, 在划分的社区中通过属性弱概念下限相似度进行组推荐研究, 进而获取推荐规则并对相应社区进行组推荐; 最后, 利用 MovieLens 数据集和 Filmtrust 数据集分析了各参数对本文所提算法的影响, 并确定了参数的合理取值。将本文所提算法与其他推荐算法进行比较测试, 实验验证了本文算法的有效性。

关键词: 形式概念分析; 网络形式背景; 节点影响力; 弱概念相似度; 组推荐

中图分类号: TP18 **文献标志码:** A

Group Recommendation Method Based on Weaken-Concept Similarity

FAN Min^{1,2}, ZHANG Jie^{1,2}, LI Jinhai^{1,2}

(1. Data Science Research Center, Kunming University of Science and Technology, Kunming 650500, China; 2. Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Concept cognition and knowledge discovery from network data are important research directions of machine learning and artificial intelligence under the network background, and have been introduced into the study of recommendation system. The existing recommendation methods based on concept lattice ignore the network structure relationship between nodes. At the same time, the efficiency of constructing concept lattice is low and the constraints of constructing concept set are strict, which is difficult to realize in large-scale social networks. In order to solve these problems, this paper integrates the topology of complex networks and weaken-concept similarity under the framework of network formal context, and proposes a group recommendation algorithm based on weaken-concept similarity. Firstly, the importance of attributes is described by defining attribute degree and attribute density, and then the expert nodes are determined by using the improved node influence. Secondly, the community is divided by expert nodes, the group recommendation research is carried out by using the lower limit similarity of attribute weaken-concept in the divided community, and then the recommendation rules are obtained and the group recommendation is applied to the corresponding communities. Finally, the influence of various parameters on the algorithm is

analyzed on MovieLens and Filmtrust datasets, and reasonable values of the parameters are determined. After that, the proposed algorithm is compared with other recommended algorithms, and the experiments show that the proposed algorithm is effective.

Key words: formal concept analysis; network formal context; node influence; weaken-concept similarity; group recommendation

引言

形式概念分析(Formal concept analysis, FCA)是 Wille 教授^[1]于 1982 年提出的一种在形式背景上进行数据分析和规则提取的有效工具。而概念格作为形式概念分析的核心数据结构,许多学者对其构造、模型推广以及应用等进行了深入研究^[2-6],比如,文献[7]从算子和布尔矩阵出发,提出了 3 种类型的对象(属性)概念的概念特征以及属性约简方法;文献[8]提出了一种面向属性的语言值概念格,在此基础上设计了面向属性的语言值概念格的提取方法,并通过基于神经网络的多维推理模型来判断语言值决策信息;文献[9]提出了一种新的构建三支概念格的方法;文献[10]研究了基于图的粒度变换和不可约元判断理论。随着 FCA 的快速发展,它与各领域的交叉融合研究亦日渐深入,已被广泛应用于数据挖掘、机器学习和推荐系统等领域^[11-14]。目前,FCA 在推荐系统上的应用主要是基于概念格,并取得了一些研究成果^[15-16],如:文献[17]将形式概念分析和协同过滤推荐相结合,通过将概念作为评分矩阵的索引,提出了基于概念格的推荐算法;文献[18]考虑到外延用户之间的相似性,利用模拟退火法构造概念集,进而在概念集上实现个性化推荐;文献[19]在构造概念格后,通过概念之间的偏序关系寻找近邻概念,进而通过概念相似度实现推荐。这些研究都是通过寻找概念格,进而利用概念用户之间的关系对网络进行推荐。事实上,在大规模的社交网络数据中,构造概念格效率低,且很难找到这样的经典概念,并对论域形成划分。为解决以上问题,本文结合网络形式背景,通过节点在网络中的影响力选取专家节点,进而根据专家节点对网络进行社区划分,在划分的社区中结合属性弱概念下限相似度获取推荐规则,从而实现组推荐。

1 相关工作

1.1 网络形式背景

网络形式背景是将复杂网络中的拓扑结构以邻接矩阵的方式与形式背景相结合,将它们统一在一个数据框架下进行分析与研究。

定义 1^[1] 三元组 (U, A, I) 称为形式背景,其中 $U = \{x_1, x_2, \dots, x_n\}$ 是非空有限对象集, $A = \{a_1, a_2, \dots, a_m\}$ 是非空有限属性集, I 是笛卡儿积 $U \times A$ 上的二元关系。约定 $(x, a) \in I$ 表示对象 x 拥有属性 a , $(x, a) \notin I$ 表示对象 x 不拥有属性 a 。为了从形式背景中诱导出概念,给出如下算子的定义:对于任意 $X \subseteq U, B \subseteq A$,有

$$\begin{cases} X^* = \{a \in A \mid \forall x \in X, (x, a) \in I\} \\ B^* = \{x \in U \mid \forall a \in B, (x, a) \in I\} \end{cases} \quad (1)$$

式中: X^* 表示 X 中所有对象共同拥有的属性; B^* 表示拥有 B 中所有属性的对象。

定义 2^[1] 设 (U, A, I) 为形式背景。对于 $X \subseteq U, B \subseteq A$,若二元组 (X, B) 满足 $X^* = B$ 且 $B^* = X$,则称 (X, B) 为形式概念,简称概念。 X 表示概念 (X, B) 的外延, B 表示概念 (X, B) 的内涵。更多关于形式概念研究的介绍参见文献[20]。

定义 3^[21] 四元组 (U, M, A, I) 称为网络形式背景, 其中 $U = \{x_1, x_2, \dots, x_n\}$ 是非空有限节点集, $A = \{a_1, a_2, \dots, a_m\}$ 是非空有限属性集, $M = \{M_1, M_2, \dots, M_k\}$ 是网络邻接矩阵集, $M_k = (m_{ij}^k)_{n \times n}$ 为网络的 k 阶邻接矩阵; $I = \{I_s | s = 1, 2, \dots, k+1\}$, $a_p \in A$ 。当 $s = 1, 2, \dots, k$ 时, I_1, I_2, \dots, I_k 是笛卡尔积 $U \times U$ 上的二元关系, $(x_i, x_j) \in I_s$ 表示节点 x_i 和节点 x_j s 阶邻接。当 $s = k+1$ 时, I_{k+1} 是笛卡尔积 $U \times A$ 上的二元关系, $(x_i, a_p) \in I_{k+1}$ 表示节点 x_i 拥有属性 a_p 。

表 1 给出的评分矩阵 Q 描述了用户 u_i 对项目 d_j 的评分情况。其中, $p(u_1, d_1) = 2$ 表示用户 u_1 对项目 d_1 的评分为 2; $p(u_1, d_2) = 0$ 表示用户 u_1 没有对项目 d_2 作出评分。表 2 为评分矩阵 Q 结合用户矩阵转化后的网络形式背景, 描述了用户之间的连接关系和用户对项目的拥有情况。在表 1 中, 若 $p(u_i, d_j) > 0$, 则在表 2 中 $I_A(u_i, d_j) = 1$ 表示用户 u_i 拥有项目 d_j ; 同理, 在表 1 中, 若 $p(u_i, d_j) = 0$, 则在表 2 中 $I_A(u_i, d_j) = 0$ 表示用户 u_i 不拥有项目 d_j 。 $m_{12}^1 = 1$ 表示用户 u_1 和用户 u_2 之间有一条边相连, $m_{2n}^1 = 0$ 表示用户 u_2 和用户 u_n 之间没有边相连, $m_{21}^k = 1$ 表示用户 u_2 和用户 u_1 之间通过 k 条边相连, $m_{2n}^k = 0$ 表示用户 u_2 和用户 u_n 之间不能通过 k 条边相连。本文将以一阶邻接矩阵为例进行分析与讨论。

表 1 评分矩阵 Q
Table 1 Scoring matrix Q

U	A			
	d_1	d_2	...	d_m
u_1	2	0	...	4
u_2	1	5	...	0
⋮	⋮	⋮	⋱	⋮
u_n	0	4	...	5

表 2 网络形式背景 (U, M, A, I)
Table 2 Network formal context (U, M, A, I)

U	M_1				...	M_k				A			
	u_1	u_2	...	u_n		u_1	u_2	...	u_n	d_1	d_2	...	d_m
u_1	0	1	...	1	...	0	1	...	0	1	0	...	1
u_2	1	0	...	0	...	1	0	...	1	1	1	...	0
⋮	⋮	⋮	⋱	⋮	⋱	⋮	⋮	⋱	⋮	⋮	⋮	⋱	⋮
u_n	1	0	0	0	...	0	1	...	0	0	1	...	1

1.2 推荐系统相关研究

基于群组的推荐系统和基于邻域的推荐系统是常见的两类推荐系统。群组推荐^[22]是指通过对用户的个人推荐聚合为对整个群组的推荐, 主要由 3 个部分组成: (1) 根据用户偏好的相似性划分用户群组; (2) 通过偏好融合策略缩小用户群组成员之间的偏好差异; (3) 对项目评分进行推荐预测。基于邻域的推荐^[23]是协同过滤推荐系统中最常见的推荐方法, 主要分为基于用户的协同过滤推荐^[24]和基于项目的协同过滤推荐^[25]。前者的主要思想是利用用户之间的相似性对目标用户进行预测评分, 后者的主要思想是利用项目之间的相似性对目标用户进行预测评分。

目前, FCA 在推荐系统的应用主要是基于概念格实施推荐, 具体是通过概念之间的偏序关系寻找近邻概念, 进而基于概念邻域完成推荐, 但它存在以下问题: 传统的利用概念集划分社区的方法是从对象集 X 出发, 找到它们共同拥有的属性 X^* , 当 $|X^*|$ 数量足够大时, X 被认为共同具有一些特点, 因而可以视作一个社区, 进一步找出所有的社区并对论域 U 形成划分。然而, 这种做法在实际数据中很难找到足够多的经典概念, 对论域 U 形成划分。于是本文提出用弱概念来对 U 进行划分。

综上, 本文将结合 FCA、群组推荐和基于项目推荐的思想, 提出基于弱概念相似度的组推荐算法。

具体步骤为:根据专家节点划分邻域(社区),接着在划分的社区中利用弱概念相似度 $\underline{\gamma}_p(q)$ 对社区成员进行群组推荐。

2 推荐系统相关问题描述与解决

本文将重点关注下列问题:(1)通过网络对象弱概念对论域进行划分,从而构建社区;(2)利用网络属性弱概念相似度实现组推荐。

2.1 网络弱概念

文献[15]在构建概念集进行组推荐时,需满足约束条件:(1)挖掘的概念集需覆盖形式背景的所有对象,即保证每个对象都至少属于1个概念;(2)挖掘的概念内涵规模不小于预设阈值。但是,在规模较大的复杂网络中,由于这种方式构建概念集的要求太严格,将会导致无法找到足够多的经典概念满足上述约束条件。为了避免出现上述情况,本文将通过构建网络弱概念来进行组推荐。在描述问题之前先给出网络弱概念的定义。

定义4 给定1个网络形式背景 (U, M, A, I) , 对于 $X \subseteq U, B \subseteq A$, (1)若 $X^{**} \supset X$, 记 $\mu = \frac{|X|}{|X^{**}|}$, 则称 (X, X^*) 为对象弱概念;(2)若 $B^{**} \supset B$, 记 $\eta = \frac{|B|}{|B^{**}|}$, 则称 (B^*, B) 为属性弱概念。其中, $|X^{**}|$ 表示对象集 X 共同拥有的属性 X^* 对应的对象个数, $|B^{**}|$ 表示共同拥有属性 B 的对象集 B^* 共同拥有的属性个数。

问题1 通过网络对象弱概念对论域 U 进行划分

输入:网络形式背景 (U, M, A, I) 。

输出:划分集(弱概念集 C_w)。

优化目标: $\min(|C_w|)$ 。

约束条件:(1) $\bigcup_{(N_i, N_i^*) \in C_w} N_i = U$; (2) $\forall (N_i, N_i^*) \in C_w, |N_i^*| \geq \gamma$ 。其中, $C_w = \{(N_i, N_i^*), N_i \subseteq U\}$ 。

优化目标表示最小化社区数量,以提高模型的拟合能力。约束条件(1)表示弱概念集覆盖了论域 U 中所有对象,即每个对象至少划分在一个社区中;约束条件(2)表示挖掘出的弱概念的内涵规模不小于 γ , 即只有当用户共同拥有的项目足够多时,形成的社区才有意义。下面给出利用网络弱概念划分社区的具体过程。

在复杂网络中,节点的度可以描述节点的影响力。在节点具有内涵属性的复杂网络中,属性越多且重要性越大的节点在网络中的影响力也会越大。因此,综合考虑节点结构影响力和属性影响力能更好地刻画节点的综合影响力。

结合前面群组推荐的思路,可以根据节点综合影响力寻找网络中比较有影响力的节点作为专家,从而构成以专家为中心的1阶邻接节点作为群组。假设选取节点影响力排名前 h 的节点作为专家,则专家集可表示为 $\text{Expert} = \{e_1, e_2, \dots, e_h\}$, 记选取的专家比例为 $\lambda = \frac{|\text{Expert}|}{|U|}$ 。

为了度量各节点的综合影响力,下面先介绍节点结构影响力和属性影响力的相关概念。

定义5^[21] 节点 x_i 的结构影响力可表示为

$$w^S(x_i) = \frac{\overline{\omega}_i^S}{\max\{\overline{\omega}_i^S\}} \quad (2)$$

$$\overline{\omega}_i^S = \sum_{k=1}^r \left(|x_i^k| + \sum_{j \in I_i^k} \frac{1}{|c_D^k(j)|} \right) \tag{3}$$

式中: k 表示邻接矩阵的阶数; $|x_i^k|$ 表示与节点 x_i 有 k 阶邻接关系的所有节点 x_j 的个数; $c_D^k(j)$ 表示节点 x_j 的 k 阶邻接的边数。

定义6 给定1个网络形式背景 (U, M, A, I) ,若 $a_k \in A$,则 $D(a_k) = |a_k^*|$ 称为属性度, $D(a_k)$ 表示拥有属性 a_k 的对象个数。 $D(a_k)$ 越大,说明拥有属性 a_k 的对象越多,即属性 a_k 在网络中越普遍。

定义7 属性 a_k 的密度可表示为

$$\rho(a_k) = \frac{D(a_k)}{N} = \frac{|a_k^*|}{N} \tag{4}$$

式中 N 表示网络节点个数。特别地,若 $\rho(a_k) \rightarrow 1$,说明网络中几乎所有的节点都具有属性 a_k ;若 $\rho(a_k) \rightarrow 0$,说明网络中几乎所有的节点都不具有属性 a_k 。

定义8 属性 a_k 的影响力可表示为

$$\text{sig}(a_k) = \frac{\rho(a_k)}{\sum_{a_k \in A} \rho(a_k)} \tag{5}$$

事实上,推荐系统中有的用户评价的项目多,且项目的属性密度比较大,那么这类用户就相对重要一些。而有的用户评价的项目比较少,且项目的属性密度也比较小,那么这类用户就相对不太重要。由此,可以定义节点的属性影响力。

定义9 节点 x_i 的属性影响力可表示为

$$w^A(x_i) = \frac{\overline{\omega}_i^A}{\max\{\overline{\omega}_i^A\}} \tag{6}$$

式中: $\overline{\omega}_i^A = \sum_{a_k \in I_A(x_i)} \text{sig}(a_k)$, $I_A(x_i)$ 表示节点 x_i 拥有的属性集。

由定义5和定义9可得节点综合影响力。

定义10 节点 x_i 的综合影响力可表示为

$$w(x_i) = \alpha w^S(x_i) + (1 - \alpha) w^A(x_i) \tag{7}$$

式中: $0 \leq \alpha \leq 1$, α 为拓扑结构影响力系数; $1 - \alpha$ 为属性影响力系数。

考虑到专家节点对其邻接领域中节点的影响,可以依据专家节点将网络划分为不同的社区。将各专家节点及其邻接节点划分到一个社区,进而可以依据专家节点的信息对邻接节点进行推荐。用 $N = \{N_1, N_2, \dots, N_h\}$ 表示该网络划分的社区集合,其中 N_i 是由专家节点 e_i 及其邻接节点 x_j 组成的集合。

通过专家节点划分社区的思想反映了网络中综合影响力较大的节点更有可能将其拥有的属性推荐给相邻的节点。例如,在电影推荐系统中,影响力较大的用户,就比较容易将自己评价过的电影推荐给与其相邻(关系密切)的用户。

2.2 基于弱概念相似度的组推荐

在描述问题之前首先给出属性弱概念相似度、推荐列表函数和推荐规则的概念。

定义11 给定一个网络形式背景 (U, M, A, I) ,设 $a_p, a_q \in A$,有

(1)属性弱概念上限相似度定义为

$$\overline{\gamma}_p(q) = \frac{|a_p^* \cap a_q^*|}{\min\{|a_p^*|, |a_q^*|\}} \quad (8)$$

(2)属性弱概念下限相似度定义为

$$\underline{\gamma}_p(q) = \frac{|a_p^* \cap a_q^*|}{\max\{|a_p^*|, |a_q^*|\}} \quad (9)$$

比如,令 $a_1^* = \{x_1, x_2, x_3\}$, $a_2^* = \{x_2, x_3\}$, $a_3^* = \{x_1, x_3, x_5\}$, $a_4^* = \{x_3, x_2\}$, $a_5^* = \{x_1\}$, 则属性弱概念上限相似度分别为 $\overline{\gamma}_2(1) = 1$, $\overline{\gamma}_2(3) = 0.5$, $\overline{\gamma}_2(4) = 1$, $\overline{\gamma}_2(5) = 0$; 属性弱概念下限相似度分别为 $\underline{\gamma}_2(1) = 0.67$, $\underline{\gamma}_2(3) = 0.33$, $\underline{\gamma}_2(4) = 1$, $\underline{\gamma}_2(5) = 0$ 。

定理 1 若 $a_p, a_q \in A$, 则有以下结论成立: (1) $\overline{\gamma}_p(q) = 1 \Leftrightarrow a_p^* \subseteq a_q^*$ 或 $a_q^* \subseteq a_p^*$; (2) $\overline{\gamma}_p(q) = \underline{\gamma}_p(q) = 1 \Leftrightarrow a_p^* = a_q^*$; (3) $\overline{\gamma}_p(q) > 0, \underline{\gamma}_p(q) > 0 \Leftrightarrow a_p^* \cap a_q^* \neq \emptyset$; (4) $\overline{\gamma}_p(q) = 0, \underline{\gamma}_p(q) = 0 \Leftrightarrow a_p^* \cap a_q^* = \emptyset$ 。

证明 (1) $\overline{\gamma}_p(q) = 1 \Leftrightarrow \frac{|a_p^* \cap a_q^*|}{\min\{|a_p^*|, |a_q^*|\}} = 1 \Leftrightarrow a_p^* \subseteq a_q^*$ 或 $a_q^* \subseteq a_p^*$ 。

(2) $\overline{\gamma}_p(q) = \underline{\gamma}_p(q) = 1 \Leftrightarrow \frac{|a_p^* \cap a_q^*|}{\min\{|a_p^*|, |a_q^*|\}} = \frac{|a_p^* \cap a_q^*|}{\max\{|a_p^*|, |a_q^*|\}} = 1 \Leftrightarrow a_p^* = a_q^*$ 。

(3) $\overline{\gamma}_p(q) > 0, \underline{\gamma}_p(q) > 0 \Leftrightarrow \frac{|a_p^* \cap a_q^*|}{\min\{|a_p^*|, |a_q^*|\}} > 0, \frac{|a_p^* \cap a_q^*|}{\max\{|a_p^*|, |a_q^*|\}} > 0 \Leftrightarrow |a_p^* \cap a_q^*| > 0 \Leftrightarrow a_p^* \cap a_q^* \neq \emptyset$ 。

(4) $\overline{\gamma}_p(q) = 0, \underline{\gamma}_p(q) = 0 \Leftrightarrow |a_p^* \cap a_q^*| = 0 \Leftrightarrow a_p^* \cap a_q^* = \emptyset$ 。

由定理 1 知, 属性弱概念上、下限相似度可以推出属性弱概念之间的关系。

定义 12 $L(x_i, a_k) = 1$ 表示将属性 a_k 推荐给对象 x_i , $L(x_i, a_k) = 0$ 表示不将属性 a_k 推荐给对象 x_i 。

定义 13 推荐规则 $r(a_i, a_j)$ 表示向拥有属性 a_i 的对象推荐属性 a_j , 即 $r(a_i, a_j): a_j \rightarrow a_i$ 。在问题 1 输出划分集 $C_w = \{(N_i, N_i^*), N_i \subseteq U\}$ 的基础上, 研究以下问题。

问题 2 基于弱概念集合的推荐

输入: $N_i, \alpha, \beta, \lambda$ 。

输出: 推荐规则 $r(a_l, a_k)$ 。

优化目标: $\max(F_1)$ 。

约束条件: $\forall a_l \in e_i^*, a_k \in x_j^*, \underline{\gamma}_l(k) \geq \beta, a_l \notin x_j^*$, 则向 x_j 推荐 a_l 。其中, $e_i \in N_i, e_i^*$ 表示专家 e_i 拥有的属性。优化目标是最大化综合评价指标 F_1 , 提高模型的质量。约束条件表示在社区 N_i 中将专家节点 e_i 的属性推荐给社区中的其他节点, 需满足属性弱概念下限相似度达到阈值 β 的约束条件。

在训练集中经过专家节点划分的社区中(即专家节点的邻域), 考虑属性弱概念相似度, 若专家节点拥有的属性 a_p 与其邻接节点中的属性 a_q 的属性弱概念相似度达到阈值, 则将属性 a_p 推荐给属性 a_q , 进而获取推荐规则。然后在测试集中对满足推荐规则前件的节点进行预测推荐。

3 算法设计

本节给出基于弱概念相似度的组推荐算法(Group recommendation algorithm based on weaken-concept similarity, RAWs)。图1为基于弱概念相似度的组推荐算法的流程图。

算法1 基于弱概念相似度的组推荐算法(GRAWS)

输入:网络形式背景 (U, M, A, I) ,结构影响力系数 α ,属性弱概念下限相似度阈值 β, λ, γ

输出: $r(a_l, a_k)$

步骤1 计算网络中所有节点的影响力: $w(x_i) = \alpha w^S(x_i) + (1 - \alpha)w^A(x_i), x_i \in U$,将所有的 $w(x_i)$ 按从大到小的顺序排列,并按 λ 选取专家集,记为 $\text{Expert} = \{e_1, e_2, \dots, e_h\}, h = \lambda|U|$ 。

步骤2 依据选取的节点专家,将网络划分成社区集合 $N = \{N_1, N_2, \dots, N_h\}, (N_i, N_i^*) \in C_w$ 且 $|N_i^*| \geq \gamma, i = 1, 2, \dots, h$ 。

步骤3 依次选取 Expert 中的节点 e_i ,计算 e_i 对应的对象弱概念 (e_i, e_i^*) ,以及 x_j 对应的对象弱概念 (x_j, x_j^*) ,其中 $x_j \in N_i, x_j \neq e_i$ 。

步骤4 对于任意 $a_l \in e_i^*$,若存在 $a_k \in x_j^*$,使得 $\gamma_l(k) \geq \beta$ 且 x_j 不拥有属性 a_l ,则将 a_l 推荐给 x_j ,即 $L(x_j, a_l) = 1$ 。

步骤5 若 $L(x_i, a_l) = 1$,则输出推荐规则 $r(a_l, a_k)$ 。

步骤6 利用获取的推荐规则对网络中的节点进行推荐预测,结束算法。

算法1结合网络形式背景选取专家用户,进而根据专家用户划分群组,并在群组中通过属性弱概念下限相似度,以决定是否将专家用户的项目推荐给同一群组的其他用户。本文实验将通过精确度、召回率和 F_1 值这三个评价指标来验证算法1的有效性,由于 F_1 为精确度与召回率的调和平均值,因而以使得算法1的 F_1 值最大为例,设置算法中涉及的参数。

下面分析算法1的时间复杂度。假设网络形式背景中有 n 位用户和 m 个项目,则选取的专家个数为 λn ,进而通过专家节点划分社区的时间复杂度为 $O(\lambda n^2)$,因此在划分的社区中计算属性弱概念下限相似度的时间复杂度为 $O(\lambda n^2 m)$ 。由于在推荐过程中需要遍历每位用户的每个项目,所以算法1的时间复杂度为 $O(\lambda n^3 m^2)$ 。

4 实验与结果分析

下面通过实验讨论参数对算法1的影响,并将算法1与现有的5种算法进行对比分析,以说明本文算法的有效性。首先,对推荐系统常用的MovieLens数据集^①和Filmtrust数据集^②进行预处理。将处理后的数据随机划分为80%训练集和20%测试集。然后,在训练集中利用算法1发现推荐规则,并对测试集进行推荐列表预测,进而计算该算法在数据集上的精确度、召回率和 F_1 值。最后与现有的5种算法进行比较。

①<https://grouplens.org/datasets/movielens/>。

②<https://guoguibing.github.io/librec/datasets.html>。

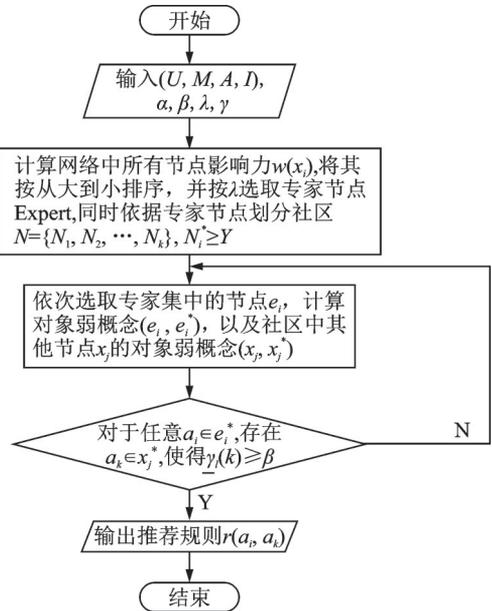


图1 算法流程框架图

Fig.1 The flowchart of the algorithm

4.1 数据集

本文实验数据集来源于 ML-1m、ML-100k-u1、ML-100k-u2、ML-100k-u3、ML-100k-u4、ML-100k-u5 和 Filmtrust。数据集的详细信息如表 3 所示。

4.2 评价指标

本文采取推荐系统的常用评价指标:精确度(Precision)、召回率(Recall)和 F_1 -measure。设 $R(u)$ 为根据训练集给出的推荐列表, $T(u)$ 为训练集中的行为列表(真实数据), 则精确度可定义为

$$\text{Precision} = \frac{|R(u) \cap T(u)|}{|R(u)|} \quad (10)$$

召回率可定义为

$$\text{Recall} = \frac{|R(u) \cap T(u)|}{|T(u)|} \quad (11)$$

F_1 -measure 为精确度与召回率的调和平均值, 是对推荐结果的综合评价, 可定义为

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

4.3 各参数对本文算法的影响

由于本文实验是通过评分的相似性来判断用户之间是否存在边相连, 进而通过专家用户划分的社区中各用户拥有的项目具有相似性, 即同一个社区中用户将会拥有足够多的项目。因此, 实验中不讨论 γ 对算法 1 的影响。下面分别给出参数 α 、 β 、 λ 对算法 1 的 F_1 值的影响。

4.3.1 参数 α 对算法 1 F_1 值的影响

图 2 为拓扑结构影响力系数 α 对算法 1 的 F_1 值的影响。由图 2 可知, 随着 α 的增大, 数据集 Filmtrust 的 F_1 值有先增加后减少的趋势, 其他数据集的 F_1 值呈现较为稳定的变化趋势, 但各数据集的 F_1 值均在 $\alpha = 0.3$ 时取得较大值。

4.3.2 参数 β 对算法 1 F_1 值的影响

图 3 为属性弱概念下限相似度阈值 β 对算法 1 的 F_1 值的影响。由图 3 可知, 随着 β 的增大, 各数据集的 F_1 值呈现先增加后减少的变化趋势。这是因为当属性弱概念下限相似度阈值达到一定值时, 若 β 再增大, 则将专家节点的属性推荐给群组中其他节点的可能性就越小, 从而获取的推荐规则就会越少, 甚至可能没有, 进而推荐结果的综合评价就会越小。当 $\beta > 0.5$ 时, F_1 值随着 β 的增大而减少, 直至为 0。

4.3.3 参数 λ 对算法 1 F_1 值的影响

图 4 为专家比例 λ 对算法 1 的 F_1 值的影响。由图 4 可知, 随着 λ 的增大, 各数据集的 F_1 值呈现上下波动的变化趋势, 且波动幅度较小, 但均在 $\lambda = 0.1$ 时取得较大值。这说明当专家比例达到 0.1 时, 算法在各数据上的 F_1 值均较大。

综上, 本文将算法 1 的参数设置为 $\alpha = 0.3$, $\beta = 0.5$, $\lambda = 0.1$ 。

表 3 数据集的信息

数据集	规模	评分范围
ML-1m	6 040 × 3 952	1~5
ML-100k-u1	943 × 1 682	1~5
ML-100k-u2	943 × 1 682	1~5
ML-100k-u3	943 × 1 682	1~5
ML-100k-u4	943 × 1 682	1~5
ML-100k-u5	943 × 1 682	1~5
Filmtrust	1 508 × 2 071	0.5~4

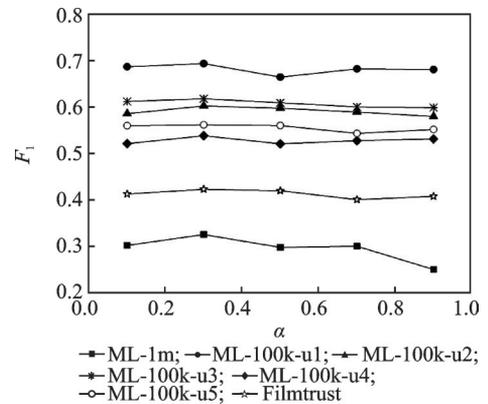


图 2 参数 α 对 F_1 的影响

Fig.2 Effect of α on F_1

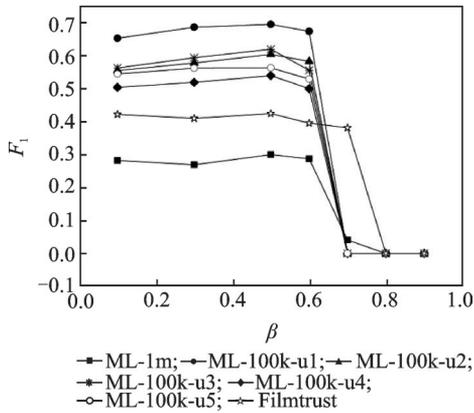


图3 参数 β 对 F_1 的影响

Fig.3 Effect of β on F_1

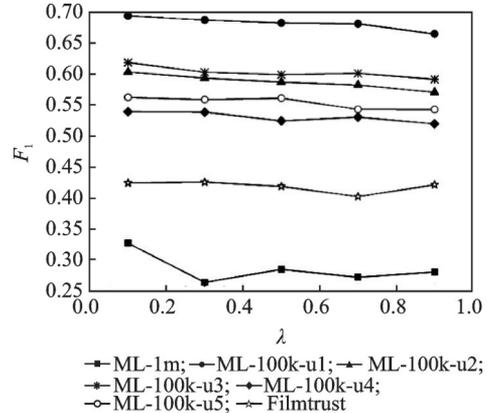


图4 参数 λ 对 F_1 的影响

Fig.4 Effect of λ on F_1

4.4 实验设置与结果对比分析

为了验证本文算法的可行性,将其与现有的5种推荐算法进行对比,5种算法如下:

(1)User-based collaborative filtering (UBCF^[24])。该算法是一种传统的协同过滤推荐算法,主要是通过用户之间的相似性对目标用户进行预测评分。

(2)Item-based collaborative filtering (IBCF^[25])。该算法是一种传统的协同过滤推荐算法,主要是利用项目之间的相似度,针对目标项目,寻找与其最相似的项目,进而通过评分信息预测目标项目。

(3)GraphRec(GR)^[26]。该算法利用图神经网络从社交和空间学习用户特征和项目特征,进而预测评分。

(4)ETBRec(ETBR)^[27]。该算法是将信任和专家用户相结合进行推荐,通过用户之间评分的相似性构造信任关系矩阵,并通过用户的信任度和积极态度计算各用户成为专家的可能,进而结合信任关系矩阵和专家进行预测评分。

(5)Term frequency*clarity (TFC)^[28]。该算法是将标签关联与用户社会关系相结合进行推荐,通过构建多标签相关矩阵和社交关系相似矩阵,进而获取用户标签矩阵,预测评分。

表4~6为GRAWS算法与其他算法在各数据集上的运行结果,表中加粗表示最优值。为了使对比更加直观,图5~7分别给出了6种算法在各数据集上的精确度、召回率和 F_1 值的对比情况。

表4 不同算法的精确度对比

Table 4 Precision comparison of different algorithms

数据集	GRAWS	UBCF	IBCF	GR	ETBR	TFC
ML-1m	0.436 4	0.412 3	0.405 5	0.424 7	0.428 9	0.416 3
ML-100k-u1	0.704 8	0.681 1	0.690 2	0.682 2	0.702 3	0.681 7
ML-100k-u2	0.619 2	0.602 3	0.594 3	0.614 3	0.620 1	0.604 3
ML-100k-u3	0.629 7	0.605 8	0.611 3	0.612 4	0.626 6	0.615 5
ML-100k-u4	0.549 5	0.523 4	0.526 6	0.536 6	0.541 3	0.536 6
ML-100k-u5	0.576 1	0.552 3	0.564 3	0.574 8	0.569 7	0.568 0
Filmtrust	0.423 9	0.417 7	0.420 1	0.416 9	0.431 8	0.427 8

表5 不同算法的召回率对比

Table 5 Recall rate comparison of different algorithms

数据集	GRAWS	UBCF	IBCF	GR	ETBR	TFC
ML-1m	0.218 4	0.196 2	0.204 3	0.206 2	0.213 4	0.203 8
ML-100k-u1	0.682 3	0.667 8	0.671 1	0.664 5	0.678 8	0.675 4
ML-100k-u2	0.586 8	0.573 4	0.563 7	0.571 1	0.581 4	0.580 1
ML-100k-u3	0.606 7	0.582 3	0.594 4	0.586 7	0.590 2	0.594 6
ML-100k-u4	0.527 7	0.496 8	0.530 6	0.511 9	0.523 1	0.528 9
ML-100k-u5	0.547 9	0.540 1	0.552 2	0.541 8	0.542 2	0.543 3
Filmtrust	0.422 6	0.421 6	0.413 3	0.409 7	0.418 8	0.420 1

表6 不同算法的 F_1 值对比

Table 6 F_1 value comparison of different algorithms

数据集	GRAWS	UBCF	IBCF	GR	ETBR	TFC
ML-1m	0.298 8	0.267 5	0.272 7	0.280 1	0.272 9	0.280 0
ML-100k-u1	0.693 4	0.674 4	0.680 5	0.673 2	0.690 3	0.678 5
ML-100k-u2	0.602 5	0.587 5	0.578 5	0.591 9	0.600 1	0.592 0
ML-100k-u3	0.618 0	0.593 8	0.602 7	0.599 3	0.607 9	0.604 9
ML-100k-u4	0.538 4	0.509 7	0.528 6	0.524 0	0.532 0	0.532 7
ML-100k-u5	0.561 6	0.546 1	0.558 2	0.557 8	0.555 6	0.555 4
Filmtrust	0.423 3	0.419 6	0.416 7	0.413 3	0.425 2	0.423 9

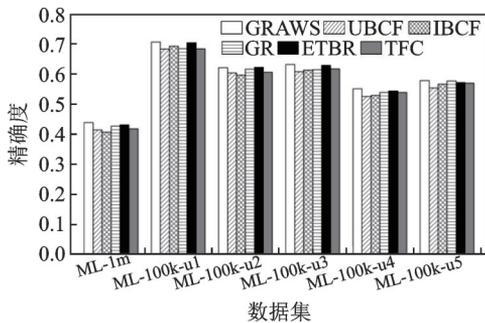


图5 不同算法精确度对比图

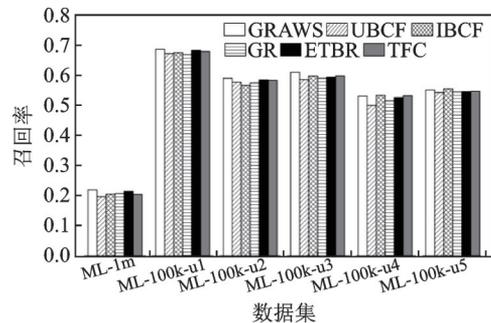


图6 不同算法召回率对比图

Fig.5 Precision comparison chart of different algorithms Fig.6 Recall rate comparison chart of different algorithms

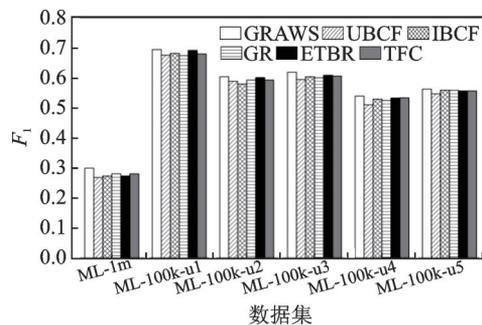


图7 不同算法 F_1 值对比图

Fig.7 F_1 value comparison chart of different algorithms

根据图5、6可知,在精确度和召回率方面,GRAWS算法在大部分数据集上都优于其他算法,这是由于GRAWS算法考虑的是在专家用户与其一阶邻接用户组成的社区中进行组推荐,通过评分的相似性来判断用户之间是否有边相连,因而在同一社区内专家节点对其他节点影响较大,各属性弱概念下限相似度就更大,进而在大部分的数据集中GRAWS算法的推荐效果就更好。但由于部分数据集中属性之间的相似性较小(即用户观看电影的相似性较小),进而可能导致统一社区内的属性弱概念下限相似度较小,GRAWS算法的推荐效果就略低于其他推荐算法,但数值相差不大。

从整体上看,在大部分数据集上,GRAWS算法在精确度、召回率和 F_1 值上均略优于其他算法。因此,在进行网络组推荐,可以采取GRAWS算法进行推荐,从而提高推荐效果。

5 结束语

本文结合网络形式背景,提出了基于弱概念相似度的组推荐算法。首先从属性度和属性密度出发,描述了属性的重要性,并通过节点影响力来选取专家节点;进而根据专家节点划分社区,为后续在划分社区中结合属性弱概念下限相似度进行组推荐奠定了基础;最后,将本文所提出的算法运用到MovieLens数据集和Filmtrust数据集中,通过对比实验发现,本文提出的推荐算法优于其他推荐算法。在本文的基础上,进一步还可以研究:(1)网络的中观层次即网络概念之间的属性依赖与合并组推荐算法,得到子网络间的推荐算法;(2)结合子网络间的相似度进行子网络合并推荐。

参考文献:

- [1] WILLE R. Restructuring lattice theory: An approach based on hierarchies of concepts[J]. *Ordered Sets*, 1982, 83: 445-470.
- [2] MA Jianmin, CAI Mingjie, ZOU Cunjun. Concept acquisition approach of object-oriented concept lattices[J]. *International Journal of Machine Learning and Cybernetics*, 2017, 8(1): 123-134.
- [3] WAN Qing, WEI Ling. Approximate concepts acquisition based on formal contexts[J]. *Knowledge-Based Systems*, 2015, 75: 78-86.
- [4] YAN Mengyu, LI Jinhai. Knowledge discovery and updating under the evolution of network formal contexts based on three-way decision[J]. *Information Sciences*, 2022, 601: 18-38.
- [5] 李金海, 魏玲, 张卓, 等. 概念格理论与方法及其研究展望[J]. *模式识别与人工智能*, 2020, 33(7): 619-642.
LI Jinhai, WEI Ling, ZHANG Zhuo, et al. Concept lattice theory and method and their research prospect[J]. *Pattern Recognition and Artificial Intelligence*, 2020, 33(7): 619-642.
- [6] 刘勇, 胡宇鹏, 李学庆. 一种基于约束格维护概念模型一致性的方法[J]. *控制与决策*, 2020, 35(2): 367-374.
LIU Yong, HU Yupeng, LI Xueqing. A method of maintaining consistency of conceptual model with constrained lattice[J]. *Control and Decision*, 2020, 35(2): 367-374.
- [7] 魏玲, 曹丽, 祁建军, 等. 形式概念分析中的概念约简与概念特征[J]. *中国科学: 信息科学*, 2020, 50(12): 1817-1833.
WEI Ling, CAO Li, QI Jianjun, et al. Concept reduction and concept features in formal concept analysis[J]. *Scientia Sinica (Informationis)*, 2020, 50(12): 1817-1833.
- [8] CUI Hui, YUE Guanli, ZOU Li, et al. Multiple multidimensional linguistic reasoning algorithm based on property-oriented linguistic concept lattice[J]. *International Journal of Approximate Reasoning*, 2021, 131: 80-92.
- [9] QIAN Ting, WEI Ling, QI Jianjun. Constructing three-way concept lattices based on apposition and subposition of formal contexts[J]. *Knowledge-Based Systems*, 2017, 116: 39-48.
- [10] WEI Ling, WAN Qing. Granular transformation and irreducible element judgment theory based on pictorial diagrams[J]. *IEEE Transactions on Cybernetics*, 2017, 46(2): 380-387.
- [11] KAYTOUE M, KUZNETSOV S O, NAPOLI A, et al. Mining gene expression data with pattern structures in formal concept analysis[J]. *Information Sciences*, 2011, 181(10): 1989-2001.
- [12] MI Yunlong, SHI Yong, LI Jinhai, et al. Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering[J]. *IEEE Transactions on Cybernetics*, 2022, 52(1): 582-593.

- [13] ZOU Caifeng, ZHANG Daqiang, WAN Jiafu, et al. Using concept lattice for personalized recommendation system design[J]. *IEEE Systems Journal*, 2017, 11(1): 305-314.
- [14] XU Weihua, LI Wentao. Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets [J]. *IEEE Transactions on Cybernetics*, 2016, 46(2): 366-379.
- [15] 刘忠慧, 邹璐, 杨梅, 等. 启发式概念构造的组推荐方法[J]. *计算机科学与探索*, 2020, 14(4): 703-711.
LIU Zhonghui, ZOU Lu, YANG Mei, et al. Group recommendation with concept of heuristic construction[J]. *Journal of Frontiers of Computer Science and Technology*, 2020, 14(4): 703-711.
- [16] MAIO C D, FENZA G, GAETA M, et al. RSS-based e-learning recommendations exploiting fuzzy FCA for knowledge modeling[J]. *Applied Soft Computing*, 2012, 12(1): 113-124.
- [17] BOUCHER-RYAN P D, BRIDGE D. Collaborative recommending using formal concept analysis[J]. *Knowledge-Based Systems*, 2006, 19(5): 309-315.
- [18] 刘忠慧, 陈建宇, 宋国杰. 基于模拟退火法的概念集构造算法[J]. *模式识别与人工智能*, 2021, 34(8): 723-732.
LIU Zhonghui, CHEN Jianyu, SONG Guojie, et al. Construction algorithm of concept set based on simulated annealing algorithm[J]. *Pattern Recognition and Artificial Intelligence*, 2021, 34(8): 723-732.
- [19] 陈昊文, 王黎明, 张卓. 基于概念邻域的 Top-N 推荐算法[J]. *小型微型计算机系统*, 2017, 38(11): 2553-2559.
CHEN Haowen, WANG Liming, ZHANG Zhuo. Top-N recommendation algorithm based on conceptual neighborhood[J]. *Journal of Chinese Computer Systems*, 2017, 38(11): 2553-2559.
- [20] 徐伟华, 李金海, 魏玲, 等. 形式概念分析理论与应用[M]. 北京: 科学出版社, 2016.
XU Weihua, LI Jinhai, WEI Ling, et al. *Theory and application of formal concept analysis*[M]. Beijing: Science Press, 2016.
- [21] 马娜, 范敏, 李金海. 复杂网络下的概念认知学习[J]. *南京大学学报(自然科学)*, 2019, 55(4): 609-623.
MA Na, FAN Min, LI Jinhai. Concept-cognitive learning under complex network[J]. *Journal of Nanjing University (Natural Science)*, 2019, 55(4): 609-623.
- [22] GARICA I, PAJARES S, SEBASTIA L, et al. Preference elicitation techniques for group recommender systems[J]. *Information Sciences*, 2012, 189: 155-175.
- [23] NING Xia, DESROSIERS C, KARYPIS G. A comprehensive survey of neighborhood-based recommendation methods[C]// *Proceedings of Recommender Systems Handbook*. Berlin, Heidelberg: Springer, 2015: 37-76.
- [24] LEKAKOS G, GIAGLIS G M. Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors[J]. *Interacting with Computers*, 2006, 18(3): 410-431.
- [25] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// *Proceeding of the 10th International Conference on World Wide Web*. New York, USA: ACM, 2001: 285-295.
- [26] FAN Wenqi, MA Yao, LI Qing, et al. Graph neural networks for social recommendation[C]// *Proceedings of the 2019 World Wide Web Conference*. New York, USA: ACM, 2019: 417-426.
- [27] DUAN Zhenchun, XU Weilong, CHEN Yuantao, et al. ETBRec: a novel recommendation algorithm combining the double influence of trust relationship and expert users[J]. *Applied Intelligence*, 2021, 52(3): 282-294.
- [28] MA Huifang, JIA Meihuizi, ZHANG Di, et al. Combining tag correlation and user social relation for microblog recommendation[J]. *Information Sciences*, 2017, 385/386: 325-337.

作者简介:



范敏(1975-),女,副教授,硕士生导师,研究方向:概念格、粗糙集、粒计算、社会网络分析等,E-mail: fmk-must@163.com。



张洁(1997-),女,硕士研究生,研究方向:粒计算、社会网络分析、概念格等,E-mail: zhangjie_104@163.com。



李金海(1984-),通信作者,男,教授,博士生导师,研究方向:大数据环境下的数据挖掘技术、概念认知学习、粒计算、智能系统分析与集成等,E-mail: jhlixjtu@163.com。