

# 三元概念的启发式构建及其在社会化推荐中的应用



刘忠慧<sup>1</sup> 赵琦<sup>1</sup> 邹璐<sup>1</sup> 闵帆<sup>1,2</sup>

1 西南石油大学计算机科学学院 成都 610500

2 西南石油大学人工智能研究院 成都 610500

(Lz\_hui@126.com)

**摘要** 形式概念分析作为知识发现的方法,在理论分析和实际应用中已经取得很多成果。随着三维数据的涌现,许多学者开始了对三元形式概念分析的研究。但是,目前该领域的研究和应用较少,尤其还没有被应用到推荐系统。文中介绍了三元概念的构建及其社会化推荐应用。首先设计启发式信息,构造覆盖所有用户的三元概念集合,启发式信息旨在生成外延和内涵均有一定规模的强概念;然后根据拟推荐项目的属性来筛选用户合适的社会关系,并结合项目在概念中的流行度实现推荐预测。文中分别在真实数据集和抽样数据集中进行了3个实验。实验1对比了启发式方法和 $V_{\alpha}$ 运算构造的三元概念数量及其运行时间,其中 $V_{\alpha}$ 运算构造的概念数量少、耗时长且对推荐的提升效果不明显;实验2对比了推荐效果的精确度、召回率和F1值,揭示了增加条件可以有效提升推荐效果;实验3的结果表明,基于三元概念的推荐算法的推荐效果优于KNN及GRHC。

**关键词:**三元形式概念分析;启发式算法;项目条件;流行度;社会化推荐

中图法分类号 TP181

## Heuristic Construction of Triadic Concept and Its Application in Social Recommendation

LIU Zhong-hui<sup>1</sup>, ZHAO Qi<sup>1</sup>, ZOU Lu<sup>1</sup> and MIN Fan<sup>1,2</sup>

1 School of Computer Science, Southwest Petroleum University, Chengdu 610500, China

2 Institute for Artificial Intelligence, Southwest Petroleum University, Chengdu 610500, China

**Abstract** Formal concept analysis is a knowledge discovery method that has great achievements in theory and application. Recently, with the emergence of three-dimensional data, triadic formal concept analysis has been developed. However, there are few researches and applications in this field, especially it has not been applied to recommendation systems. This paper proposes an efficient triadic concept set construction method and applies it to social recommendation. Firstly, the heuristic information is designed to generate a set of triadic concepts covering all users. Heuristic information aims to construct strong concepts with a certain scale of extension and intension. Then, appropriate social relations are screened through the attributes of the proposed items, and the recommendation prediction is realized by combining the popularity of the items in the concept. Three experiments are carried out in real data set and sampled data set respectively. In the first experiment, the number of triadic concepts and running time constructed by the heuristic method and  $V_{\alpha}$  operation are compared respectively. The concepts constructed by the  $V_{\alpha}$  operation do not significantly improve the recommendation effect. The second experiment compares the accuracy, recall rate and F1 of the recommendation effect. It reveals that increasing the number of conditions can effectively improve the recommendation effect. The results of the last experiment show that the recommendation effect of the new algorithm is better than that of KNN and GRHC.

**Keywords** Triadic formal concept analysis, Heuristic algorithm, Item conditions, Popularity, Social recommendation

## 1 引言

形式概念分析(Formal Concept Analysis, FCA)<sup>[1]</sup>主要探讨二维数据中对象、属性的二元关系以及概念的层次结构,可以很好地对概念进行聚类、排序和分类,因此在机器学习、数据挖掘和知识发现等<sup>[2-4]</sup>领域已有广泛研究和应用。

FCA的主要研究方向包括形式概念的构造<sup>[5-7]</sup>和约简<sup>[8]</sup>、形式概念和粗糙集<sup>[9]</sup>的结合、形式概念与粒计算<sup>[10]</sup>的结合、基于形式概念的推荐应用<sup>[11-13]</sup>等。例如,基于树结构的概念格增量构造算法<sup>[7]</sup>可以快速找到父子节点范围,缩短搜索时间,降低概念格构造的时间复杂度;文献<sup>[10]</sup>将形式概念分析与粒计算相结合,证明了多粒度标记形式背景和多粒度

到稿日期:2020-05-27 返修日期:2020-08-17 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(41604114);四川省青年科技创新团队(2019JDTD0017)

This work was supported by the National Natural Science Foundation of China(41604114) and Sichuan Province Youth Science and Technology Innovation Team(2019JDTD0017).

通信作者:闵帆(minfan@swpu.edu.cn)

标记信息系统在语义上等价;文献[11]以概念格为基础,在概念格中提取关联规则以实现推荐算法;文献[12]利用概念的父子关系寻找用户的邻居,通过相似度计算达到推荐效果;文献[13]针对概念格构造效率低、难以应对电子商务大规模数据的问题,通过启发式方法构造概念集合来实现组推荐。

随着大数据的发展,越来越多的三维数据被采集,如何有效利用这些数据进行知识发现变得日益重要。三元概念分析(Triadic Concept Analysis, TCA)<sup>[15]</sup>在FCA基础上增加了第三元“条件”,使对象与属性的联系更加具体和多元化,为数据处理和分析提供了新的思路。该方向的理论进展迅速,主要研究包括三元格的构造及简化<sup>[16-17]</sup>、三元概念的认知系统研究及三元决策形式背景下的信息融合<sup>[18-19]</sup>、三元概念的角色访问控制<sup>[20]</sup>、文本分类等。Wang等<sup>[16]</sup>提出了一种三元概念的构造方法,通过 $\vee_{oc}$ 运算在已有三元概念的基础上构造新的三元概念;Tang等<sup>[18]</sup>对三元形式概念分析下的认知系统进行了研究,提出了充分信息粒、必要信息粒及充分必要信息粒等概念,并分析了它们之间的转换方法;Wei等<sup>[21]</sup>对三元概念的理论、方法及应用进行了综合阐述;Li等<sup>[22]</sup>将三元概念和模糊集相结合,通过三元概念来实现文本分类算法。

然而,TCA在实际应用中面临时间及空间复杂度高的问题。传统形式概念分析的常用数据结构为概念格,而构造概念格的时间复杂度为指数型。三元形式背景比二元形式背景复杂,因此三元格构造的时间复杂度更高,增加了将三元概念应用到大数据分析和知识发现的难度。

本文采用启发式方法构造三元概念集合,降低了构造的时间复杂度,同时根据形式概念的特点实现了社会化推荐,主要贡献包括以下两个方面:

(1)利用启发式方法构造三元概念集合。通过内涵阈值和概念面积的约束构造覆盖用户集的三元概念集合,降低了时间复杂度。

(2)利用三元概念集合实现社会化推荐。将包含用户的三元概念视为其社会关系,通过待推荐项目的属性筛选可用的社会关系,再结合项目的流行度实现推荐。

实验结果表明:本文方法可以在大数据集上实现三元概念的构造;项目的条件个数增加能够提升推荐效果;基于三元概念的社会化推荐效果优于 $k$ 近邻( $k$  Nearest Neighbors,  $k$ NN)<sup>[23]</sup>算法和GRHC(Group Recommendation with the Concept of Heuristic Construction)<sup>[13]</sup>算法。

## 2 相关工作

本节主要介绍FCA和TCA的相关定义和知识,以及推荐系统的相关研究。本文仅针对推荐系统进行研究,因此使用“用户”和“项目”代替一般形式概念分析中的“对象”和“属性”。

### 2.1 相关定义

**定义 1<sup>[1]</sup>**(二元形式背景) 二元形式背景是一个三元组 $T=(U, A, R)$ ,其中 $U$ 为用户集, $A$ 为项目集, $R$ 为 $U$ 和 $A$ 之间的二元关系, $R \subseteq U \times A$ 。 $r(u, a) = 1$ 表示用户 $u$ 拥有项目 $a$ ,也可以表示为 $uRa$ 。

用户集 $X \subseteq U$ 和项目集 $B \subseteq A$ ,分别定义运算:

$$X^* = \{a \in A \mid \forall u \in X, uRa\} \quad (1)$$

$$B^\diamond = \{u \in U \mid \forall a \in B, uRa\} \quad (2)$$

**定义 2<sup>[1]</sup>**(二元形式概念) 在二元形式背景 $T=(U, A, R)$ 中,令 $E \subseteq U, I \subseteq A$ ,当二元组 $C=(E, I)$ 满足 $E^* = I, I^\diamond = E$ 时,则称 $C=(E, I)$ 为二元形式概念,也称二元概念。 $E(C) = E$ 表示二元概念的外延, $I(C) = I$ 表示二元概念的内涵。

**定义 3<sup>[17]</sup>**(三元形式背景) 三元形式背景是一个四元组 $T=(U, A, D, Y)$ ,其中 $U, A, D$ 为非空集合,分别称为用户集、项目集和条件集。 $Y$ 为 $U, A, D$ 之间的关系, $Y \subseteq U \times A \times D$ 。 $y(u, a, d) = 1$ 表示用户 $u$ 在条件 $d$ 下具有项目 $a$ ,也可以表示为 $(u, a, d) \in Y$ 。

表1列出了三元形式背景 $T=(U, A, D, Y)$ ,其中 $U = \{u_0, u_1, u_2, u_3, u_4\}$ , $A = \{a_0, a_1, a_2, a_3, a_4\}$ , $D = \{d_0, d_1, d_2\}$ 。

表1 三元形式背景 $T=(U, A, D, Y)$

Table 1 Triadic formal context  $T=(U, A, D, Y)$

	$d_0$					$d_1$					$d_2$				
	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$
$u_0$	1	0	1	0	0	0	0	1	1	0	0	1	0	1	0
$u_1$	0	0	1	0	0	0	0	1	1	0	0	1	0	1	0
$u_2$	0	0	1	0	1	0	0	1	1	0	0	0	0	1	1
$u_3$	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
$u_4$	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1

在二元形式背景中,若 $E \subseteq U, I \subseteq A, M \subseteq D$ ,定义诱导算子如下:

$$E^{(D)} = \{d \in D \mid \forall (u, a) \in E \times I, (u, a, d) \in Y\} \quad (3)$$

$$E^{(M)} = \{a \in A \mid \forall (u, d) \in E \times M, (u, a, d) \in Y\} \quad (4)$$

$$I^{(M)} = \{u \in U \mid \forall (a, d) \in I \times M, (u, a, d) \in Y\} \quad (5)$$

**定义 4<sup>[15]</sup>**(三元形式概念) 在二元背景 $T=(U, A, D, Y)$ 中,令 $E \subseteq U, I \subseteq A, M \subseteq D$ ,若 $E^{(D)} = M, E^{(M)} = I, I^{(M)} = E$ ,则称 $C=(E, I, M)$ 为三元形式概念,简称三元概念。 $E, I, M$ 分别为该三元概念的外延、内涵和方式。

**定义 5<sup>[15]</sup>**( $\vee_{oc}$ 运算) 在二元形式背景 $T=(U, A, D, Y)$ 中,设 $C_1=(E_1, I_1, M_1), C_2=(E_2, I_2, M_2)$ 为 $T$ 的两个三元概念,则 $(E, I, M) = (E_1, I_1, M_1) \vee_{oc} (E_2, I_2, M_2)$ 也为 $T$ 的三元概念,其中:

$$I = E_1^{(M_1)} \cap E_2^{(M_2)} \quad (6)$$

$$M = E_1^{(D)} \cap E_2^{(D)} \quad (7)$$

$$E = M^{(D)} \quad (8)$$

**定义 6<sup>[13]</sup>**(概念面积) 三元形式概念 $C=(E, I, M)$ 的面积:

$$S(C) = |E| \times |I| \quad (9)$$

其中, $|\cdot|$ 表示集合中的元素个数。

**定义 7<sup>[13]</sup>**(强概念) 令 $C=(E, I, M)$ 为一个用户 $u \in E$ 的三元概念且 $|I| \geq \beta$ ,若不存在 $C'=(E', I', M'), u \in E'$ ,使 $S(C) < S(C')$ ,则称 $C$ 为用户 $u$ 的强概念,其中 $\beta$ 为内涵阈值。

**定义 8<sup>[13]</sup>**(项目流行度) 三元形式背景 $T=(U, A, D, Y), C=(E, I, M)$ 为 $T$ 的一个三元概念。已知用户 $u \in E$ ,项目 $a \in A$ ,则 $a$ 在概念 $C$ 上的项目流行度为:

$$p(C, a) = \frac{|\{u \in E \mid \exists d \in D, (u, a, d) \in Y\}|}{|E|} \quad (10)$$

## 2.2 基于形式概念分析的推荐系统

推荐系统已经成功应用到电子商务、信息检索和社交网络等领域。传统的推荐系统主要包括基于内容的推荐、协同过滤和组合推荐<sup>[24]</sup>。基于内容的推荐通过待推荐项目和用户喜欢的项目之间的匹配程度来实现推荐；协同过滤主要包括基于模型的方法和启发式方法，前者利用概率统计模型或机器学习方法来构建用户特征模型，后者通过寻找用户的相似邻居用户来实现协同推荐；组合推荐是将基于内容的推荐和协同过滤相结合的推荐方法。

目前，社会化推荐系统逐渐成为研究热点<sup>[25]</sup>。其在传统推荐系统的基础上，把用户社会关系信息(如亲属关系、朋友关系等)作为重要影响因子引入到推荐系统中，更符合人类生活的社会化特征且具有可解释性。社会化推荐系统根据用户之间的社会化关系构建社会化关系网络，然后利用这种社会关系对用户进行推荐。与传统推荐系统一样，社会化推荐系统分为3种：基于内容的社会化推荐系统、基于协同过滤的社会化推荐系统和组合的社会化推荐系统。其中应用最普遍的是基于协同过滤的社会化推荐系统<sup>[26]</sup>。

传统的基于形式概念分析的推荐是一种协同过滤推荐方法。2005年，Boucher-ryan等<sup>[27]</sup>将形式概念分析与协同过滤算法相结合，提出了基于概念格的推荐算法。该算法基于概念格，利用概念之间的偏序关系，从父子概念和兄弟概念中获得推荐候选项目。Cheng等<sup>[11]</sup>通过构建概念格，利用概念的偏序关系获得与用户偏好相似的邻居用户，并从这些邻居用户中提取目标用户的推荐项。

目前，基于形式概念的推荐可被视为一种基于协同过滤的社会化推荐方法。根据评分信息和社交信息构造概念，其外延互为相关性用户，推荐时将相关性用户感兴趣的项目推荐给目标用户。Liu等<sup>[13]</sup>通过构建概念集合，利用包含了社交信息概念实现推荐。

## 3 问题描述

本文旨在通过构建三元概念来实现用户的社会化推荐，因此需要解决构造三元概念和通过三元概念实现推荐的问题。本节将对这两个问题逐一进行分析。

### 3.1 三元概念构造

第一个子问题讨论三元概念应该具有怎样的特性以及如何构造三元概念集合。借鉴文献<sup>[13]</sup>的思路，我们希望构造的三元概念在外延和内涵上具有一定规模，以保证社会关系的稳定性，并且概念集合需要覆盖用户集，确保每个用户都有能用于推荐的社会关系。基于这两点，三元概念应该具有强概念的特点。

问题1 三元概念集合的构建

输入：三元形式背景  $T=(U, A, D, Y)$ ，内涵阈值  $\beta$

输出：三元概念集合  $\mathcal{C}$

约束条件1:  $\bigcup_{(E, I, M) \in \mathcal{C}} E = U$ ;

约束条件2:  $\forall (E, I, M) \in \mathcal{C}, |I| \geq \beta$ ;

优化目标:  $\min |\mathcal{C}|$ 。

其中，三元形式背景表示在所有条件下用户对项目的行为数据。内涵阈值  $\beta$  控制三元概念中内涵的项目个数。第一个约

束条件表示三元概念集合覆盖形式背景的所有用户，确保每个用户至少与一个概念建立联系。第二个约束条件表示所挖掘的三元概念的内涵规模不小于阈值  $\beta$ ，从而控制概念的规模和三元概念集合中概念的个数。优化目标为最小化概念集合，其目的是获得更简洁的表示，提高模型的泛化性。

### 3.2 基于三元概念的推荐

第二个子问题是如何利用三元概念实现推荐。每个用户至少被一个三元概念包含，因此需要选择与用户和待推荐项目都有紧密联系的概念。

问题2 基于三元概念集合的社会化推荐

输入：三元形式背景  $T=(U, A, D, Y)$ ，三元概念集合  $\mathcal{C}$ ，内涵阈值  $\beta$

输出：预测矩阵  $L$

约束条件:  $\forall C=(E, I, M) \in \mathcal{C}, \forall u \in E, a \in A - \bigcup_{d \in D} \{u\}^{(d)}, M \cap (\bigcup_{u \in U} \{u\}^{(a)}) \neq \emptyset, p(C, a) \geq \gamma$ 。

其中， $\mathcal{C}$  为问题1构造的三元概念集合。输出结果  $L$  是一个二元矩阵。约束条件根据待推荐项目  $a$  的条件，在三元概念集合中选择方式  $M$  和与其有交集的概念  $C$ ，然后在概念  $C$  中计算项目  $a$  的项目流行度。 $\gamma(0 < \gamma < 1)$  是用户自定义参数，当项目流行度大于等于  $\gamma$  时推荐该项目。

## 4 算法实现

本节设计了基于三元概念集合的推荐算法(Recommendation Based on Triadic Concept Set, RTCS)，用于解决问题1和问题2，并描述了一个RTCS的运行实例。

### 4.1 三元概念构造算法

三元概念构造分为两个部分：算法1采用启发式方法构造三元概念集合；算法2利用算法1生成的三元概念，采用  $\bigvee_{\alpha}$  运算构造新的三元概念。所有三元概念在算法3中用于推荐。

算法1 启发式方法构造三元概念集合

输入：三元形式背景  $T=(U, A, D, Y)$ ，内涵阈值  $\beta$

输出：三元概念集合  $\mathcal{C}$

1.  $\mathcal{C} = \emptyset$ ;
2. for(each  $d \in D$ ) do
3.  $P = U$ ;
4. while( $P \neq \emptyset$ ) do
5.  $s = 0, E = \emptyset, I = \emptyset, M = \emptyset$ ;
6.  $u = \operatorname{argmax}_{p \in P} |\{p\}^{(d)}|$ ;
7.  $a = \operatorname{argmax}_{i \in \{u\}^{(d)}} |\{i\}^{(d)}|$ ;
8.  $I = I \cup \{a\}$ ;
9. while(true) do
10.  $a = \operatorname{argmax}_{i \in (\{u\}^{(d)} - I)} |\{i\}^{(d)}|$ ;
11.  $s' = S((I \cup \{a\})^{(d)}, I \cup \{a\})$ ;
12. if( $(s' > s) \vee (|I| + 1 < \beta)$ ) then
13.  $I = I \cup \{m\}$ ;
14.  $s = s'$ ;
15. else
16. break;
17. end if

```

18. end while
19.  $E = I^{(d)}$ ,  $M = E^{(1)}$ ;
20.  $C = C \cup \{(E, I, M)\}$ ;
21.  $P = P - E$ ;
22. end while
23. end for
24. return  $C$ .

```

算法1根据强概念的定义设计启发式方法,生成的每一个概念满足内涵阈值的约束。第4—22行对三元形式背景构造强概念集合。选择一个包含项目最多的用户作为概念构造的起点(第6行),第6—20行对这个用户构造强概念。概念的初始化内涵为用户的项目中覆盖用户最多的项目(第7行),然后依次添加覆盖较多用户的项目(第10行),每次添加后都更新概念面积(第11行)。当内涵达到阈值并且概念面积最大时,确定概念的内涵(第12—17行)。第19行通过式(5)和式(3)来确定概念的方式和外延,将新概念添加到三元概念集合中。将新概念的外延从用户候选集 $P$ 中删除(第21行),以避免重复构造相同的概念。

#### 算法2 $V_\alpha$ 运算构建三元概念集合

输入:三元形式背景  $T=(U, A, D, Y)$ , 强概念集合  $C$

输出:三元概念集合  $\mathcal{C}$

```

1.  $\mathcal{C} = C$ ;
2. for(each  $C_1, C_2 \in \mathcal{C}$ ) do
3.    $C' = C_1 V_\alpha C_2$ ;
4.   if( $C' \notin \mathcal{C}$ ) then
5.      $\mathcal{C} = \mathcal{C} \cup C'$ ;
6.   end if
7. end for
8. return  $\mathcal{C}$ .

```

算法2在算法1的基础上,通过 $V_\alpha$ 运算构造新的三元概念。在集合 $C$ 中,任意两个不相等的概念进行 $V_\alpha$ 运算(第3行),若构造的概念为集合 $C$ 中不存在的新概念,则将其添加到三元概念集合中(第4—6行)。

#### 4.2 社会化推荐算法

在社会化推荐中,将包含了用户的三元概念视为其社会化关系,而推荐过程是对特定条件下社会关系的筛选和应用。

#### 算法3 基于三元概念集合的社会化推荐算法

输入:三元形式背景  $T=(U, A, D, Y)$ , 三元概念集合  $\mathcal{C}$ , 推荐阈值  $\gamma$

输出:推荐预测矩阵  $L$

```

1.  $L_{|O| \times |A|} \leftarrow 0$ ;
2. for(each  $u \in U, a \in (A - \bigcup_{a \in D} \{u\}^{(d)})$ ) do
3.   for(each  $u \in C(E, I, M) \in \mathcal{C}$ ) do
4.     if( $(\bigcup_{u \in U} u^{(a)}) \cap M \neq \emptyset$ ) then
5.       if( $p(C, a) \geq \gamma$ ) then
6.          $l_{(u, a)} = 1$ ;
7.       end if
8.     end if
9.   end for
10. end for
11. return  $L$ .

```

算法3是利用算法2构造的三元概念集合进行推荐,其中待推荐项目为用户在任何条件下都不包含的项目(第2行)。在包含目标用户的三元概念集合中,对待推荐项目的条

件与概念的方式求交,若结果不为空,则该概念可用于预测(第3—4行);若待推荐项目在概念中的流行度超过推荐阈值,则将其推荐给目标用户(第5—7行)。

#### 4.3 算法时间复杂度分析

假设三元形式背景有 $n$ 个用户、 $m$ 个项目和 $q$ 个条件,则算法1的时间复杂度为:

$$O((n+m)nmq) \quad (11)$$

假设算法1构造了 $k$ 个三元概念,且 $k \leq nq$ ,则算法2的时间复杂度为:

$$O(n^3 q^3 m) \quad (12)$$

三元概念集合构造的时间复杂度为式(11)与式(12)之和。实际情况下,条件个数较少,与用户和项目的规模相比可忽略不计,因此构造三元概念集合的时间复杂度可以简化为:

$$O((n^2 + n + m)nm) \quad (13)$$

算法3实现推荐的时间复杂度为:

$$O(n^2 m) \quad (14)$$

#### 4.4 运行实例

为了通俗地表示RTCS算法的执行流程,下面描述该算法对表1所列的三元形式背景进行三元概念构造和推荐的运行实例。设置内涵阈值为2,推荐阈值为0.6。

##### (1)构造三元概念集合

构造的三元概念集合如表2所列,其中0—8号概念通过启发式方法构造,9号概念由5号、6号概念通过 $V_\alpha$ 运算生成。

表2 三元概念集合

Table 2 Triadic concept set

编号	三元概念
0	$\{u_0\}, \{a_0, a_2\}, \{d_0\}$
1	$\{u_2\}, \{a_2, a_4\}, \{d_0\}$
2	$\{u_3\}, \{a_0, a_4\}, \{d_0\}$
3	$\{u_0, u_1, u_2\}, \{a_2\}, \{d_0, d_1\}$
4	$\{u_2, u_3, u_4\}, \{a_4\}, \{d_0, d_2\}$
5	$\{u_0, u_1, u_2\}, \{a_2, a_3\}, \{d_1\}$
6	$\{u_0, u_1\}, \{a_1, a_3\}, \{d_2\}$
7	$\{u_2\}, \{a_3, a_4\}, \{d_2\}$
8	$\{u_4\}, \{a_1, a_4\}, \{d_2\}$
9	$\{u_0, u_1, u_2\}, \{a_3\}, \{d_1, d_2\}$

##### (2)通过三元概念实现推荐

###### 1)用户 $u_0$ 对待推荐项目 $a_4$ 的预测情况

待推荐项目 $a_4$ 的条件为 $d_0, d_1$ ,因此筛选与这些条件有交集的概念0,3,9为可用的社会关系。通过项目流行度与推荐阈值的比较结果可知,所有社会关系的预测为不喜欢,因此不向用户 $u_0$ 推荐项目 $a_4$ ,如表3所列。

表3  $u_0$ 的预测

Table 3 Prediction of  $u_0$

概念编号	项目流行度	预测	结果
0	不判断	—	
3	1/2=0.5	小于0.6 不喜欢	不推荐
9	1/2=0.5	小于0.6 不喜欢	

###### 2)用户 $u_2$ 对待推荐项目 $a_1$ 的预测情况

待推荐项目 $a_1$ 的条件为 $d_2$ ,因此筛选与该条件有交集的概念4,9为可用的社会关系,通过项目流行度与推荐阈值的比较,9号概念的社会关系预测结果为喜欢,因此向用户 $u_2$ 推荐项目 $a_1$ ,如表4所列。

表 4  $u_2$  的预测Table 4 Prediction of  $u_2$ 

概念编号	项目流行度	预测	结果
4	1/2=0.5	小于 0.6 不喜欢	推荐
9	2/2=1	大于 0.6 喜欢	

## 5 实验与结果

### 5.1 数据集

实验数据集包括两个原始数据集 MovieLens-100K 和 MovieLens-1M, 以及 3 个从豆瓣电影数据抽样得到的数据集 Douban1, Douban2, Douban3。这 5 个数据集均为用户对电影的评分数据集, 数据集中的用户集合、电影集合和电影类型集合分别对应形式背景中用户集、项目集和条件集。数据集的具体信息如表 5 所列。

表 5 数据集信息

Table 5 Dataset information

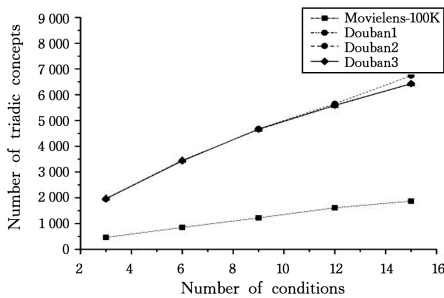
数据集	用户个数	项目个数	条件个数
MovieLens-100K	943	1682	19
Douban1	2965	1897	24
Douban2	2965	1897	24
Douban3	2965	1897	24
MovieLens-1M	6040	3952	18

### 5.2 评价指标

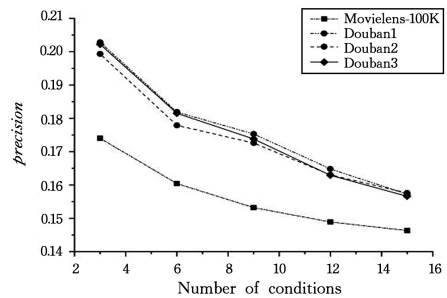
本文使用 3 种评价指标对算法进行评估。令  $L(U)$  为算法对所有用户的推荐矩阵,  $T(U)$  为所有用户在测试集中的行为矩阵。精确度 (precision)、召回率 (recall) 和 F1 这 3 个评价指标的计算式如下:

$$precision = \frac{|L(U) \cap T(U)|}{|L(U)|} \quad (15)$$

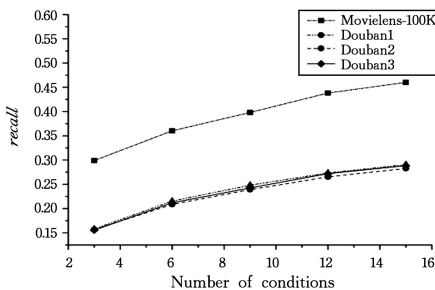
$$recall = \frac{|L(U) \cap T(U)|}{|T(U)|} \quad (16)$$



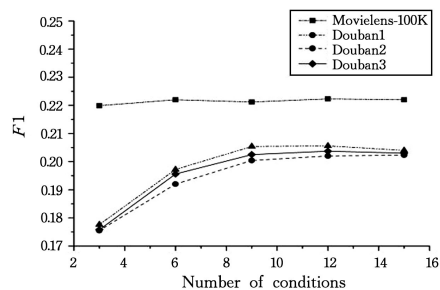
(a) 概念数量对比



(b) 精确度对比



(c) 召回率对比



(d) F1 对比

图 1 条件个数对推荐的影响

Fig. 1 Influence of number of conditions on recommendation

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (17)$$

### 5.3 $V_\alpha$ 运算对推荐的影响

$V_\alpha$  运算涉及到所有三元概念的两两计算, 每一对概念计算需要 3 次遍历三元形式背景, 时间复杂度较高。我们单独设计实验以观察  $V_\alpha$  运算对概念构造和推荐效果的影响。表 6 中,  $|C_{HCS}|$  表示启发式方法构造的三元概念个数,  $|C_{V_\alpha}|$  表示  $V_\alpha$  运算构造的三元概念个数,  $t_{V_\alpha}/t_{HCS}$  表示  $V_\alpha$  运算与启发式方法分别构造三元概念集合所用时间的比值。实验参数为  $\beta=2$ ,  $\gamma=0.5$ 。

表 6 概念数量与运行时间比较

Table 6 Number of concept vs running time

Dataset	$ C_{HCS} $	$ C_{V_\alpha} $	$t_{V_\alpha}/t_{HCS}$
MovieLens-100K	2076	87	45.15
Douban1	6940	112	69.15
Douban2	6993	123	66.66
Douban3	6978	105	57.53

表 6 所列的实验结果显示,  $V_\alpha$  运算得到的概念个数远小于启发式方法构造的三元概念个数, 且耗费的时间是启发式方法的几十倍。同时, 在 3 个评价指标下,  $V_\alpha$  运算构造的概念对推荐效果的提升有限, 未超过 0.5%。

### 5.4 条件个数对推荐的影响

项目通常有多个条件, 如某一部电影具有多种类型, 一个和多个条件对推荐是否有影响? 为了回答这个问题, 我们单独设计实验来比较条件个数对推荐效果的影响。图 1 给出了在 4 个数据集上, 概念生成、推荐性能与项目条件个数的关系。随着条件个数的增加, 三元概念数量增多, 推荐精确度降低, 召回率增大, F1 值保持恒定或有明显提升。因此, 可以通过增加条件个数来提高整体的推荐效果。

### 5.5 RTCS算法与两种推荐算法的对比

将 RTCS 算法与基于领域推荐的 KNN 算法<sup>[23]</sup>、基于群组推荐的 GRHC 算法<sup>[13]</sup>进行对比。从表 7 所列的实验结果可以看出,在精确度上,除了 Douban1 和 MovieLens-1M 数据集,RTCS 比 KNN 和 GRHC 算法的效果更好。在召回率上,3 种算法在实验数据集中的差别不大。在 F1 上,RTCS 算法优于 KNN 和 GRHC 算法。数据集 MovieLens-1M 的实验参数  $\beta=3, \gamma=0.7$ , 其他数据集的实验参数  $\beta=2, \gamma=0.6$ 。

表 7 RTCS 与两种算法的对比

Table 7 Comparison of RTCS and two algorithms

Dataset	Evaluation index	KNN	GRHC	RTCS
MovieLens-100K	precision	0.1623	0.2040	0.2110
	recall	0.3771	0.2114	0.3076
	F1	0.2269	0.2076	0.2503
Douban1	precision	0.1451	0.2016	0.2113
	recall	0.1580	0.1472	0.1501
	F1	0.1518	0.1702	0.1755
Douban2	precision	0.1518	0.2035	0.2176
	recall	0.1563	0.1460	0.1519
	F1	0.1543	0.1700	0.1789
Douban3	precision	0.1496	0.0952	0.2132
	recall	0.1599	0.2898	0.1524
	F1	0.1546	0.1433	0.1777
MovieLens-1M	precision	0.1375	0.3015	0.1893
	recall	0.3923	0.0631	0.2373
	F1	0.2036	0.1044	0.2106

**结束语** 本文研究了三元概念的启发式构造与社会化推荐应用,提出了 RTCS 算法,通过启发式方法和  $V_{\alpha}$  运算构造三元概念集合,根据待推荐项目筛选社会化关系来实现推荐。在实验中分析了通过  $V_{\alpha}$  运算生成的三元概念对推荐的贡献,讨论了项目条件个数对概念生成数量和推荐效果的影响,最后将 RTCS 算法与 KNN 和 GRHC 进行了比较,结果显示 RTCS 算法的推荐效果有较为稳定的提升。本文的研究对三元概念的第三元赋予了与项目相关的语义,拓展了三元概念的应用领域,今后还可以通过赋予三元概念不同的语义来解释更多的现实问题。利用启发式方法构造三元概念集合依然是一个值得拓展的方向,由于概念格构造的复杂度很高,因此还可以根据应用场景的需求采用不同的启发式方法。

### 参考文献

[1] WILLE R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts[J]. Springer, 1982, 83:314-339.

[2] GANTER B, STUMME G, WILLE R. Formal Concept Analysis: Theory and Applications[J]. J Ucs, 2004, 10(8):926.

[3] LI J H, MEI C L, LV Y J. Knowledge reduction in decision formal contexts[J]. Knowledge-based Systems, 2011, 24(5):709-715.

[4] ZHI H L, LI J H. Granule Description Based on Necessary Attribute Analysis [J]. Chinese Journal of Computers, 2018, 41(12):2702-2719.

[5] GODIN R, MISSAOUI R, ALAOUI H. Incremental concept formation algorithms based on Galois (concept) lattices[J]. Computational Intelligence, 1995, 11(2):246-267.

[6] LI Y, LIU Z T, CHEN L, et al. Attribute-based Incremental Formation Algorithm of Concept Lattice[J]. Journal of Chinese Computer Systems, 2004(10):1768-1771.

[7] XIE Z P, LIU Z T. A Fast Incremental Algorithm for Building Concept Lattice [J]. Chinese Journal of Computers, 2002(5):490-496.

[8] ZHANG W X, WEI L, QI J J. Theory and Method of Attribute Reduction in Concept Lattice[J]. Science in China Ser. E Information Sciences, 2005(6):628-639.

[9] YAO Y Y. Concept lattices in rough set theory[C]//IEEE Annual Meeting of the Fuzzy Information. 2004:796-801.

[10] LI J H, WU W Z, DENG S. Multi-scale theory in formal concept analysis[J]. Journal of Shandong University(Natural Science), 2019, 54(2):34-44.

[11] CHENG H W, WANG L M, ZHANG Z. Top-N Recommendation Algorithm Based on Conceptual Neighborhood[J]. Journal of Chinese Computer Systems, 2017, 38(11):2553-2559.

[12] LI H T, HE K Q, WANG J, et al. A Friends Recommendation Algorithm Based on Formal Concept Analysis and Random Walk in Social Network [J]. Advanced Engineering Sciences, 2015, 47(6):131-138.

[13] LIU Z H, ZOU L, YANG M, et al. Group recommendation with the concept of heuristic construction[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(4):703-711.

[14] IGNATOV D I, KUZNETSOV S O. Concept-based recommendations for internet advertisement[J]. Computer Science, 2009, 433:157-166.

[15] LEHMANN F, WILLE R. A Triadic Approach to Formal Concept Analysis [C]// International Conference on Conceptual Structures Applications. Springer, 1995:32-43.

[16] WANG X, JIANG S, LI J Y, et al. A Construction Method of Triadic Concepts[J]. Journal of Computer Research and Development, 2019, 56(4):844-853.

[17] WANG X, TAN S W, LI J Y, et al. Constructions and simplifications of concept lattices based on conditional attribute implications[J]. Journal of Nanjing University(Natural Science), 2019, 55(4):553-563.

[18] TANG Y Q, FAN M, LI J H. Cognitive system model and approach to transformation of information granules under triadic formal concept analysis [J]. Journal of Shandong University (Natural Science), 2014, 49(8):102-106.

[19] TANG Y Q, FAN M, LI J H. An information fusion technology for triadic decision contexts[J]. Int. J. Mach. Learn. & Cyber. 2016, 7:13-24.

[20] KUMAR C A, MOULISWARAN S C, LI J H, et al. Role based access control design using Triadic concept analysis[J]. Journal of Central South University, 2016, 23(12):3183-3191.

[21] WEI L, WAN Q, QIAN T, et al. An overview of triadic concept

- analysis[J]. Journal of Northwest University(Natural Science Edition), 2014, 44(5): 689-699.
- [22] LI Z, ZHANG Z, WANG L M. Research on Text Classification Algorithm Based on Triadic Concept Analysis[J]. Computer Science, 2017, 44(8): 207-215.
- [23] LUO X, OUYANG Y X, XIONG Z, et al. The effect of similarity support in K-Nearest-Neighborhood based collaborative filtering [J]. Chinese Journal of Computers, 2010, 33(8): 1437-1445.
- [24] HUANG L W, JIANG B T, LYU S Y, et al. Survey on Deep Learning Based Recommender Systems [J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647.
- [25] MENG X W, LIU S D, ZHANG Y J, et al. Research on social recommender systems[J]. Ruan Jian Xue Bao/Journal of Software, 2015, 26(6): 1356-1372.
- [26] ZHANG Q S, WENG L J. Review of Social Recommender Systems[J]. Computer Engineering and Applications, 2020, 56(1): 1-10.
- [27] BOUCHER-RYAN P D, BRIDGE D. Collaborative Recommending using Formal Concept Analysis[J]. Knowledge-Based Systems, 2006, 19(5): 309-315.



**LIU Zhong-hui**, born in 1980, postgraduate, associate professor, master supervisor, is a member of China Computer Federation. Her main research interests include machine learning, formal concept analysis and rough set.



**MIN Fan**, born in 1973, Ph.D, professor, Ph.D supervisor, is a member of CAAI granule computing and knowledge discovery committee and China Computer Federaton. His main research interests include granular computing, recommender systems and active learning.