

《智能信息处理》课程作业

基于形式概念分析的推荐系统

靳少宁

作业	分数[20]
得分	

2020 年 11 月 10 日

基于形式概念分析的推荐系统

靳少宁¹⁾

¹⁾ (大连海事大学 信息科学技术学院, 大连 116026)

摘 要 形式概念分析是数据分析与规则提取的一种有效工具, 已经在机器学习、数据挖掘、软件工程等领域得到了广泛的应用。但是传统的形式概念分析理论在处理模糊关系的时候具有局限性, 尤其是现在海量数据的出现, 使得形式概念分析理论在模糊关系的处理应用上成为重要的研究方向。最近几年, 个性化推荐系统越来越火, 随着网购的兴起, 如何有效的提高协同过滤算法在处理模糊数据以及不完备数据时的推荐准确度, 将对提升系统的性能有很大的帮助。本文研究了一种形式概念分析 (FCA) 网络图在推荐系统中的应用, 本文使用 FCA 方法创建一个网络图并应用这个图来提高推荐系统的准确性。根据拥有相似特征信息的用户的基本知识, 他们可能喜欢相同的物品, 这个想法被用来创建一个 FCA 网络图, 在提出的过程中, 根据每个社区中新用户与现有用户的特征信息相似度, 为新用户找到合适的社区, 并推荐新用户更可能喜欢的东西。

关键词 形式概念; 形式背景; 数据挖掘; 推荐系统

Recommendation System based on formal concept analysis

Shaoning Jin¹⁾

¹⁾ (School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract Formal concept analysis is an effective tool for data analysis and rule extraction. It has been widely used in machine learning, data mining, software engineering and other fields. However, the traditional formal concept analysis theory has its limitations in dealing with the issues, especially with the appearance of mass data, so the theory of formal concept analysis has become an important research direction in the application of the mold and the relationship. In recent years, personalized recommendation systems have become increasingly popular, with the rise of online shopping, how to effectively improve the co-filtering Algorithm in the processing of modular and incomplete data recommendation accuracy, will be of great help to improve the performance of the system. In this paper, we study the application of a formal concept analysis (FCA) network graph in recommendation system. We use FCA method to create a network graph and apply this graph to improve the accuracy of recommendation system. Based on the basic knowledge of users with similar characteristic information, they may like the same items, the idea was used to create an FCA network diagram during the presentation process, find the right community for new users based on the similarity of their profile information to existing users in each community, and recommend things new users are more likely to like.

Key words formal concept; formal context; data mining; recommendation system

0 引言

随着互联网技术的快速发展，以及数据的海量增长，人们对于这种海量数据及海量数据之间的关系的处理凸显出了越来越大的局限性。而推荐系统是互联网或电子商务领域的主要功能。人们在购物或者进行网上冲浪的时候，如果网站或者软件可以根据用户的喜好分别进行推荐，这将极大地增强用户的使用体验，并且提高产品的销售量，反之，如果我们对用户的推荐总是出错，不能准确的判断出用户所希望看到的产品，而是推荐一些随机产品，长此以往，这会使我们失去用户，而形式概念分析（FCA），给我们提供了一个高精度的分析结果。

1 形式概念分析简介

形式概念分析是一种从形式背景出发，进行数据分析和规则提取的强有力工具，形式概念分析建立在数学基础之上，对组成本体的概念、属性以及关系等用形式化的语境表述出来，然后根据语境，构造出概念格，从而清楚地表达出本体的结构。形式概念分析强调以人的认知为中心，提供一种与传统的方法，成为了人工智能学科的重要研究对象，在机器学习、数据挖掘、信息检索等领域得到了广泛的应用。

1.1 概念定义

形式概念 (Formal Concept Analysis, FCA)由 R.Wille 于 1982 年提出是应用数学的一个分支、是信息处理的一种理论、是知识处理的一种理论。形式概念通俗的语言描述是：形式对象集和形式属性集的总和，定义如下：

定义^[1] 设 A 是对象集合 G 的一个子集，定义 $f(A) := \{m \in M | \forall g \in A, gIm\}$ (A 中对象共同属性的集合)。gIm 表示 $g \in G$ 与 $m \in M$ 之间存在关系 I，读作：

对象 g 具有属性 M。

相应地，设 B 是属性集合 M 的一个子集，我们定义 $g(A) := \{g \in G | \forall m \in B, gIm\}$ (具有 B 中所有属性的对象的集合)。

定义^{2[1]} 形式背景(G, M, I)上的一个形式概念 (Formal Concept) 是二元组(A,B)，其中 $A \subseteq G, B \subseteq M$ ，且满足 $f(A) = B$ 和 $g(B) = A$ 。我们称 A 是形式概念(A,B)的外延，B 是形式概念(A,B)的内涵。ξ (G,M, I) 表示形式背景(G,M, I)的所有形式概念的集合。

相应的概念三种表示方法，如图 1 所示。

形式概念的作用就是构建自然概念的层次连通结构。为了更好的解释此作用，根据上述描述的内容，在自然概念的基础之上可以建立一个新的形式概念内容，从而可以获得一些新的结论。如图 2 所示。

①表达式 形式概念 = (形式对象集，形式属性集)

②二维表法

	形式属性1	...	形式属性2
形式对象1			
...			
形式对象n			

(形式对象集，形式属性集)

③图示法 

图 1 形式概念的三种表示

	屏幕	键盘	遥控器
电脑，电视			
电脑	✓	✓	
电视	✓		✓
∅	✓	✓	✓

图 2 形式概念作用距离

图 2 的表现形式，我们可以理解为：如果某个事物具有电脑和电视的两种性，也许就会有不一样的好处。例如，目前的电视主要使用有线信号进行节目播放。如果一个电

视机拥有两个系统，一个系统是 windows 操作系统；另一个系统就是普通的电视机系统，这样就可以实现电脑电视的结合。通过切换信号源就可以进行系统的改变，这样的产品既可以满足家中老年人对于简单的电视节目需求、也可以满足年轻人对智能设备的需求，一个简单的设备满足了不同的客户群体。事实证明，智能电视正在向这样的方向发展，一种是电视系统；一种是安卓系统。虽然目前还不是 windows 操作系统，这样的发展趋势已经展露。

1.2 形式背景

在这个客观世界中，单独的事物并不能够描述一个具体的系统。同样地，单独的形式概念并不足以描述一个形式概念集，于是我们引入形式背景来解释这个问题。

定义 3^[1] 一个形式背景 (Formal Context) 是一个三元组 $K = (G, M, I)$ ，其中 G 是对象的集合， M 是属性的集合， I 是 G 和 M 之间的一个二元关系，即， $I \subseteq G \times M$ 。

根据定义 3，一个形式背景能够用一个矩形表来表示，表的每一行是一个对象，每一列是一个属性。若 g 行 m 列的交叉处是 X ，则表示对象 g 具有属性 m ，如表 1 所示。

表 1 定义 3 的表格显示

	属性 1	属性 2	属性 n
对象 1	X			
对象 2	X	X		
.....				
对象 m				X

形式背景就是信息世界，相当于我们的客观世界。对于形式背景而言，把形式概念置于其中的二维表就是形式背景。

在一个复杂的形式概念集进行形式背景生成的时候会存在一些问题，其中包括形式背景生成过程中产生冗余信息。这样的冗余信息可能发生在不同的对象拥有相同的属性，我们把这样的冗余问题进行约简，此类问题称为“约简行”；有时候不同的事物虽然有不同的属性，但是属性和属性之间可能存在异构现象，我们把这样的约简问题称为“约简列”；有时候多个属性之间能相互

关联，某个属相可能由一个或者多个其他相关属性推出，这样的约简问题我们称为“关联规则抽取”通过行和列的约简我们就可以进行信息的合并。

2 材料和方法

2.1 材料

FCA 是一种数据分析理论，它描述了一组特定属性和一组特定对象之间的关系。它从输入数据生成两种类型的输出。第一个是概念格，它是数据中形式概念的集合。形式概念是根据子概念和超概念之间的关系来安排的。第二个输出是一组称为属性含义的属性。属性隐含描述对象属性之间的依赖关系，这些属性在连续的时间点上被观察到。

创建 FCA 网络图的过程是，首先根据形式上下文生成概念格。属性蕴涵函数或关联规则函数用于创建描述用户之间关系的关系表。最后，这个表用于创建网络图。

2.2 方法

本节详细描述了所提方法的场景。本文的主要思想是解决冷启动问题和在大型数据集上消耗大量时间的执行问题。冷启动问题涉及到对没有或很少历史记录的用户(新用户)的个性化推荐。因此，协同过滤模型的学习和预测能力有限，向历史记录较少的用户提供建议成为一个难题。按照一般的想法，具有相同特征信息或属性或个性化信息的用户可能喜欢相同的物品。在电影镜头数据集中使用用户的个性化信息，该数据集分为训练部分和测试部分。用于创建正式上下文表的训练数据中的信息。在正式上下文生成之后，然后使用 Lattice Miner 2.0 应用程序来创建概念格。概念格表示折线图，可通过收集概念节点到格的最小值的所有对象获得，其意图可通过访问从概念节点到格的最高值的所有属性获得。稍后，FCA 网络图将通过格点挖掘 2.0 中的关联规则函数创

建，该函数用于检查概念格中对象（用户）之间的关系。

3 结果

一旦概念格被生成,那么在格 Minor 2.0 应用程序中支持最好的属性暗示函数或关联规则将被用来在用户之间创建一个关系表。稍后，这个表用于创建网络图。FCA 网络图的结果如图 3 所示。此图还显示了从概念格导入后用户之间的关系。

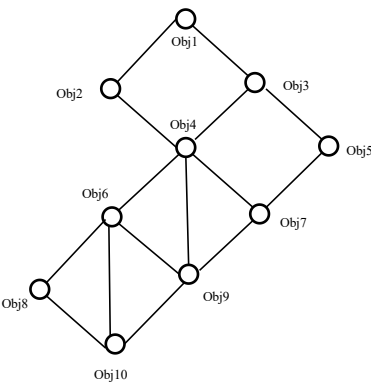


图 3 形式概念分析网络图

4 总结

由于推荐系统在互联网上的业务中是必不可少的，在线营销为企业所有者通过互联网或在线娱乐向消费者提供电影提供了很多好处。电影的数量已经增加，变得更加拥挤;通过现有的技术找到用户正在寻找的视频是非常困难的。因此，用户需要一个系统，可以推荐他们的电影需求和最好的技术，这是一个推荐系统。然而，协同过滤方法是推荐系统中用于预测用户需求的一种强有力的方法，因为这种方法有助于满足用户的预测。协同过滤机制内部仍然存在冷启动等问题，这与新用户登录时没有任何之前的互联网搜索信息或之前没有历史记录有关。

参 考 文 献

[1] 概念, <http://dwz.cn/4nJWbo> 2016,10,17

[2] Phonexay Vilakone, Doo-Soon Park, The Efficiency of a DoParallel Algorithm and an FCA Network Graph Applied to Recommendation System [J]. 2020,04

[3] 晏力, 刘鹏慧. 基于形式概念分析的属性约简[J]. 西华大学学报(自然科学版), 2012, 04: 37-41

[4] 张云中. 基于形式概念分析的领域本体构建方法研究[D]. 吉林大学, 吉林, 2009