形式概念分析在推荐系统中的应用

1 引言

随着互联网的迅速发展，其用户数量也在迅速增长，但伴随而来的是信息过载问题的出现。海量的信息使用户在获取资源时无从下手。为了解决这样的问题，推荐系统应运而生。通过分析用户以及与用户相关的数据来推测出用户的兴趣所在和行为趋势，以此为基础向用户推荐他们所需要的信息和服务。推荐系统的产生有效的解决了信息过载问题，使数据增长不再成为用户按需获取信息的阻碍。目前推荐系统被广泛的应用在电子商务、视频网站、音乐媒体和社交网络领域。

推荐系统领域的核心部分主要由推荐算法构成，协同过滤(Collaborative Filtering)是目前应用最广泛的方法之一。其核心思想是，通过分析用户记录，在用户群体中找到指定用户的相似用户，综合这些相似用户对某一信息的评价，形成系统对指定用户对此信息的喜好程度预测。协同过滤的方法具体可以分为基于用户和基于物品两种，在一些企业的实际应用中更多的选择基于物品的协同过滤方法，因为基于物品的协同过滤算法可以相对于用户而言物品之间的关系更加的稳定，从而训练过程更加可控。

形式概念分析(Formal Concept Analysis，FCA)，是德国数学家Wille基于序理论提出的理论体系，其核心数据结构概念格已经广泛地应用于数据挖掘、信息检索、数据抽取、软件工程等领域，是一种强有力的数据分析与规则挖掘工具。由于概念格存在特殊的结构及性质，使得它在多个领域取得了较好的应用效果。但在个性化推荐领域，形式概念分析及概念格理论的应用仍处于探索阶段，所以具有一定的研究意义和价值。

2 形式概念分析

概念在哲学中被理解为外延与内涵所组成的思想单元，德国数学家Wille在1982年首次提出了形式概念分析，用于概念发现、排序与显示。概念格作为形式概念分析的核心数据结构，是根据形式北京中对象与属性之间的二元关系建立的一种概念层次结构。概念格能够通过Hasse图清晰地体现概念间的泛化和特化关系，因此被认为是进行数据分析有力的工具。

定义1. (形式背景)一个形式背景是由两个集合和和以及与间的关系组成。的元素称为对象，的元素称为属性。或表示对象 具有属性。

定义2. （伽罗瓦联系）设A是对象集合G的一个子集，定义@@@（A中对象共同属性的集合）。相应地设B是属性集合M的一个子集，@@@（具有B中所有属性对象的集合）。若@@@@，则称集合 满足伽罗瓦联系。

定义3. （形式概念）设@和@是格中的两个概念，且@@，称@是@的子概念，@是@的超概念

表1 形式背景示例

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 考古遗址 | 沙滩 | 欧元 | 溪流 | 滑雪区 |
| 雅典 | 1 | 1 | 1 | 0 | 0 |
| 因斯布鲁克 | 0 | 0 | 1 | 1 | 1 |
| 巴黎 | 0 | 0 | 1 | 1 | 0 |
| 罗马 | 1 | 1 | 1 | 1 | 0 |

表3 形式背景

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | I1 | I2 | I3 | I4 |
| U1 | 0 | 1 | 0 | 1 |
| U2 | 0 | 0 | 1 | 1 |
| U3 | 0 | 0 | 0 | 1 |
| U4 | 1 | 0 | 0 | 0 |
| U5 | 1 | 1 | 1 | 0 |
| U6 | 0 | 0 | 1 | 0 |
| U7 | 1 | 1 | 0 | 0 |