

《智能信息处理》课程考试

## 基于本体的推荐算法研究综述

于慧敏

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021 年 12 月 12 日

# 基于本体的推荐算法研究综述

于慧敏

(大连海事大学 信息科学技术学院, 辽宁省大连市 116026)

**摘 要** 基于本体的推荐技术在解决传统推荐系统中存在的数据稀疏性、语义缺乏、推荐精度低等问题中发挥了重要的作用。文章首先梳理了基于本体的推荐系统在实现过程中的关键技术,帮助读者了解基于本体的推荐系统的基本特征和技术实现上的重要问题,然后对基于本体的推荐系统的主流方法以详细介绍应用实例的方式进行了分析与总结,最后对基于本体的推荐系统有待深入的研究和发展趋势进行了展望。

**关键词** 推荐算法; 本体; 推荐系统; 语义

## Research on recommendation algorithm based on formal concept analysis

Yu Huimin

(Computer science and technology, Dalian maritime university, Liaoning Dalian, 116026)

**Abstract** Recommendation technology based on ontology plays an important role in solving the issues of data sparsity, lack of semantic, low recommendation accuracy of the traditional recommendation system. Firstly, this paper summarizes the key technology in the process of implementation of the recommendation system based on ontology, to help readers understand the basic characteristics and the key problems which are proposed to solve in the technical implementation of the recommendation system based on ontology. Then this paper systematically analyzes and summarizes the mainstream technology method of recommendation system based on ontology which is according to the general classification of recommendation system and detailed introduction of concrete application examples. Finally, the prospects for further research and development trend of the recommendation based on ontology are discussed.

**Key words** recommendation algorithm; ontology; recommendation system; semantic

## 0 引言

在大数据时代,人们已不再为资源匮乏的问题而感到担心,如今困扰人们的是资源过载这一问题。信息技术带来资讯极其丰富的好处的同时,也带来了用户的选择困难症。而最好的选择,就是不需要选择,这就是推荐系统所要解决的问题,推荐系统的目标就是从大集合中选取用户最感兴趣的对象。

历史上,推荐算法由于综合的应用了人工智能和信息过滤领域的方法而广受关注,采用这些方法可以推荐 Web 站点或对新闻进行排序、过滤。现在的推荐技术已在众多行业得到了应用和推广,为

不同领域提供着解决方案,比如金融产品、房地产、电影、书籍、购物等,但同时这些推荐系统也存在一些缺陷:

- (1) 冷启动和数据稀疏性问题;
- (2) 不能表达项目和用户需求之间的语义关系,导致的推荐精度低的问题;
- (3) 资源内容结构化程度低,在资源重用和共享方面存在限制等问题。

本体的引入为上述问题的解决提供了新的思路,随着国内外研究者的深入探索和实践,本体已在个性化推荐领域显露优势。本文系统地梳理了基于本体的推荐系统在实现过程中的关键技术,并对基于本体的推荐系统的主流方法以详细介绍应用实例的方式进行了综述。

# 1 基于本体推荐的关键技术

## 1.1 领域本体的构建

领域本体的构建是实现基于本体推荐的基础。本体是对领域概念及其相互关系的形式化表示, 是对领域知识共同的理解与描述, 是实现语义网技术的关键。Perez 等<sup>[1]</sup>认为按分类法组织的本体是最简单的本体, 类或概念、关系、函数、公理和实例是本体的 5 个基本要素。面向不同的知识领域、描述语言 and 对象, 本体构建的过程和方法会有所不同, 本体的构建方法有很多, 主要包括企业建模法、骨架法、IDEF5 法、循环获取法、知识工程方法和原型进化法等。这里简要介绍知识工程法和基于 OWL 的原型迭代法。

知识工程法<sup>[2]</sup>是斯坦福大学的 Natalya 等人提出的一种基于本体开发工具 Protégé 的开发方法, 包括以下不断重复的过程: (1) 确定领域和范围; (2) 重用已有本体; (3) 列举重要概念; (4) 定义类及类的层级结构; (5) 定义类的属性和分面; (6) 创建本体实例。

考虑标准化的描述语言 OWL 对本体构建的要求, 提出了使用原型迭代模型进行本体的构建, 该过程包括领域需求分析、现有本体复用、迭代分析、本体评估修正和本体存储等 5 个步骤。基于 OWL 的原型迭代本体构建模型见图 1.1。

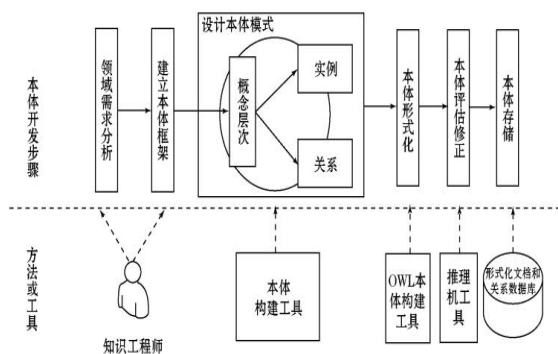


图 1.1 原型迭代本体构建模型

利用本体技术进行知识的表示和描述, 当用户刚加入 Web 时只根据用户配置文件分析其兴趣爱好, 为用户推荐其感兴趣的内容, 当用户在 Web 上发布的信息不断增加时, 可以根据这些内容进一步全面地分析用户兴趣, 实现更加精准的推荐。同时, 在领域本体构建完成之后, 其更新和完善的工作量是很小的, 相对于数据挖掘等技术来说, 基于本体的推荐节约了大量的计算时间。此外, 由于本体中

每个概念中不仅含有概念的名称, 同时还含有概念的属性、关系和实例等, 这些都可以作为最后的推荐结果向用户推送, 因此, 推荐结果的精确性和可扩展性都获得了很大的提升。

## 1.2 语义标注

本体定义了资源描述所使用的词汇集, 但对于语义网信息组织来说, 在成功完成本体构建后, 在资源描述上仍存在两大问题: 一是在创建网页、发布网络信息时, 如何方便、有效地利用本体中定义的词汇描述网络信息资源; 二是对于现有的万维网而言, 如何利用本体词汇揭示其中潜在的语义信息。这两个问题都属于语义标注的范畴。

语义标注(Semantic Annotation)又称本体标注(Ontology Annotation), 是指通过本体中定义的词汇显示地揭示和表达网络中的语义信息。可以将本体看作是由概念分类构成的一个分类词表, 语义标注的任务就是依据本体中定义的分类体系, 将应用实例或目标文档进行归类, 也就是分类标引的过程, 这个过程需要借助本体自动标注工具来实现。如果将语义标注过程看作一个过程系统, 本体和目标文档则为该系统的输入, 本体实例或个体的属性值则为系统的输出, 系统的处理过程就是依据给定的本体, 通过各类算法, 对目标文档进行自动分类。图 1.2 是一种基于特征向量进行概念归类的本体标注模型。

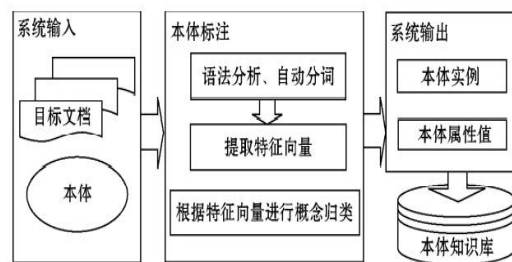


图 1.2 本体自动标注模型

在基于领域本体的推荐系统中, 语义标注根据 Web 上的语义关系实现智能推理, 推进语义网技术走向实际应用。通过语义标注, 领域本体中的类与项目之间建立了丰富的语义关系, 提高了个性化推荐的精确性。

## 1.3 语义相似度的计算

领域本体使用层次化的树状结构描述类之间的语义关系, 利用领域本体中项目的类型及属性计算项目之间的语义相似度。国内外关于语义相似度的计算方法可以归结为: 依据特征模型的、依据语

义距离的、依据信息内容的以及混合方法,其中基于语义距离的计算方法在推荐系统中的应用较为广泛。通常情况下,推荐系统中项目的语义距离与项目的相似度之间存在着反比关系,即两个项目的语义距离越大,其相似度越低;两个项目的语义距离越小,其相似度越大。定义两个项目  $a$  和  $b$  之间的相似度函数为  $\text{Sim}(a,b)$ , 两个项目之间的语义距离和相似度之间存在如下的对应关系: (1)  $\text{Sim}(a,b) \in [0, 1]$ ; (2)  $\text{Sim}(a,b) = 1$ , 则  $a, b$  两个项目等价; (3)  $\text{Sim}(a,b) = 0$ , 则  $a, b$  无共同特征; (4)  $\text{Sim}(a,b) = \text{Sim}(b,a)$ , 则  $a, b$  两个项目对称。利用领域本体中类的层次关系表示和度量各个节点间的语义关系,可以从语义角度衡量节点间的相似性程度,从而提高推荐的精准度。

#### 1.4 基于本体的推理

虽然本体清晰、准确地定义了概念的含义,并支持一定的推理,但是由于还没有定义具体的推理规则,因而不能判断命题的真伪。对规则的语义描述可使推荐系统获得更强大的推理能力。在面向语义网和本体的具体应用中,一般的规则描述都是在 OWL(Ontology Web Language)的基础上附加谓词逻辑,从而使 OWL 具有一定的规则表达能力。对 OWL 进行规则扩展比较热门的提案主要是 SWRL(Semantic Web Rules Language)。SWRL 仍用模型论进行语义解释,它是 OWL 语义的直接扩展。SWRL 可被视为规则和本体的结合,通过 SWRL,本体描述的关系和条件可以在编写规则时直接使用。在 SWRL 中仍然保留着用实例来表达推理假设的概念及关系和作为推理结果的这种模式。图 1.3 是一个用 SWRL 描述规则的实例。

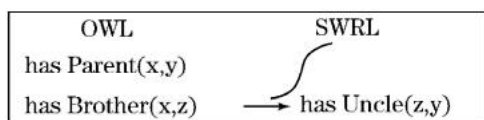


图 1.3 SWRL 的一个实例

## 2 基于本体的推荐算法

推荐算法是实现个性化推荐的核心,对推荐系统的性能优劣起着决定性作用。目前,无论是传统的推荐系统还是基于本体技术的推荐系统,根据推荐算法的不同,主流的推荐方法都可以分为:协同过滤推荐、基于内容的推荐、基于知识的推荐和混合推荐。基于本体技术的推荐系统仍然采用了传统

推荐系统中算法的核心思想,但是在具体实现细节上,结合本体技术的优势,对传统的推荐算法进行改进,从本质上解决传统推荐系统中存在的问题。

#### 2.1 协同过滤推荐扩展法

协同过滤推荐的主要思想是通过已有用户群的既往行为来预测当前用户的喜好。在传统协同过滤推荐系统中,如何确定相似用户集,推荐系统中通用的方法是 Pearson 相关系数和余弦相似度度量。随着用户与项目量级的急速膨胀,使曾在业界取得过一定成功的传统协同过滤推荐变得力不从心,该类推荐算法也暴露了一些缺陷,比如在冷启动和用户-项目评分矩阵稀疏的情况下,难以准确找出用户的最近邻居集合,从而不能有效计算用户间的相似度;没有考虑项目之间的语义关联,无法发现项目之间存在的隐含关系,从而导致推荐准确率低等问题。

针对传统协同推荐方法中暴露的问题,提出一种基于领域本体的协同过滤推荐算法。利用领域本体中项目的层次结构计算项目之间的语义相似度,基于最近邻的思想,依据用户历史项目评分记录来预测新项目的评分,用以填充稀疏的用户-项目评分矩阵,在此基础上实现推荐,以提高推荐的准确率。根据领域本体构建用户兴趣模型,通过对用户兴趣模型的相似度比较,预测 Top-N 个性化推荐列表,打破以往协同过滤推荐比较用户对不同项目评分的限制。同时,随着用户兴趣的动态变化,根据激活扩展推断理论,用户兴趣模型将被自动更新。该推荐系统框架如图 2.1 所示。

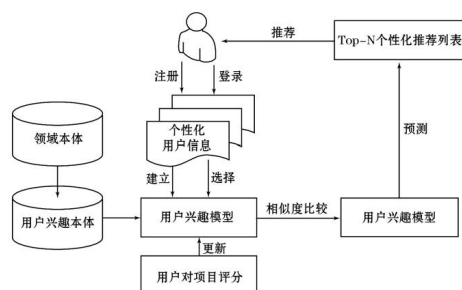


图 2.1 基于用户兴趣本体的协同过滤推荐系统框架

#### 2.2 基于内容的推荐扩展法

基于内容推荐系统的主要思想是得到项目的描述和特征的重要记录,将项目特征与用户兴趣偏好匹配起来,推荐其他类似特征的项目。用户兴趣偏好的获取方法是显式或隐式地分析用户的行为和反馈,或者直接询问用户的兴趣和偏好。实际评



研究。

## 参 考 文 献

- [1] Asuncion Gomez Perez., V. Richard Benjamins Overview of Knowledge Sharing and Reuse Components Ontologies and Problem-solving Methods[C]. Proceeding of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August2, 1999: 1-15.
- [2] Natalya F. Noy, Deborah L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology.[2006-11-27]. [http://protege.stanford.edu/publications/ontology\\_development/ontology101-noymcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noymcguinness.html).
- [3] Pretschner A, Gauch S. Ontology Based Personalized Search[C]. Proceedings of th11 IEEE International Conference on Tools with Artificial Intelligence, 1999: 391-398.
- [4] S. E. Middleton, N. R. Shadbolt, D. C. DeRoure. Ontological User Profiling in Recommender Systems, ACM Transactions on Information Systems22, 2005(2):179-197.