

# 基于云模型的协同过滤推荐算法<sup>①</sup>

万年红

(浙江东方职业技术学院 工程技术系, 温州 325011)

**摘要:** 传统的协同过滤推荐算法面临严峻的数据稀疏性和推荐实时性困境, 推荐质量明显不高. 为提高推荐效果, 首先对基于云模型的用户评分项和相似性度量方法展开研究. 然后定义基于云模型的推荐系统信任约束, 并改进主观信任云模型的约束函数、信任变化云模型的信任变化函数. 最后提出一种基于云模型的协同过滤推荐算法. 实验结果表明, 相比传统算法, 该算法在用户评分数据稀疏的状况下仍然可以取得良好的推荐效果, 具有较高的实用价值.

**关键词:** 云模型; 相似性度量; 约束函数; 信任变化函数; 协同过滤推荐

## Collaborative Filtering Recommendation Algorithm Based on Cloud Model

WAN Nian-Hong

(Department of Engineering Technology, Zhejiang Dongfang Vocational and Technical College, Wenzhou 325011, China)

**Abstract:** The traditional collaborative filtering recommendation algorithms face the dilemma of severe data sparsity and real time of recommendation, their recommendation quality is not obviously high. To improve recommendation efficiency, firstly, user rating items and similarity measurement method based on cloud model were researched. Then the definition of recommendation system trust constraint based on cloud model was given, and improved the constraint function of subjective trust cloud model and trust change function of trust change cloud model. Finally, a collaborative filtering recommendation algorithm based on cloud model was put forward. The experimental results show that the algorithm still obtains good recommendation efficiency on situation of user rating data sparsity compared to the traditional algorithms, it has high utility.

**Key words:** cloud model; similarity measure; constraint function; trust change function; collaborative filtering recommendation

随着互联网、智能计算和电子商务的发展, 协同过滤推荐系统已经成为电子商务 IT 技术中的一个重要研究内容. 实现准确、安全、高效的协同过滤推荐系统已成为电子商务研究中备受瞩目的热门课题. 目前, 对协同过滤推荐技术的研究主要从似然关系模型和用户等级<sup>[1]</sup>、用户和项目因素<sup>[2,3]</sup>、集成项目类别与语境信息<sup>[4]</sup>、领域最近邻<sup>[5]</sup>、不确定近邻<sup>[6]</sup>、用户聚类<sup>[7-9]</sup>等角度来讨论的, 其总体思想是基于相似用户行为的评分数据, 预测目标用户对未评分项目的评分, 从而将预测评分最高的若干项目推荐给目标用户<sup>[1]</sup>. 但是, 随着电子商务站点用户和商品项数量的不断增

加, 传统的协同过滤推荐系统也存在数据稀疏性和扩展性等局限, 面临严峻的用户评分数据稀疏性和推荐实时性挑战, 导致推荐质量下降<sup>[5]</sup>. 相比较而言, 作为电子商务的新一代支撑模式, 云模型在协同过滤推荐系统的构建中发挥着越来越重要的作用, 它通过估计用户评分的办法补充用户矩阵, 减小数据稀疏性对计算结果的负面影响, 取得理论上的推荐质量. 因此云模型非常适合在协同过滤推荐领域的研究与应用, 但是随着电子商务系统规模的扩大, 导致用于产生推荐的用户数据过于庞大, 从而影响推荐生成速度和质量<sup>[7]</sup>, 更系统更有效的有关云模型协同过滤推荐算法的研究

<sup>①</sup> 收稿时间:2014-09-18;收到修改稿时间:2014-10-13

并不多见。

本文针对已有研究的不足,利用云模型在知识表达中的重要作用,从基于云模型的用户评分项以及基于云模型的相似性度量角度,通过改进云模型的相似性度量方法、主观信任云模型的约束函数、信任变化云模型的信任变化函数,提出一种新的基于云模型的协同过滤推荐算法(Collaborative Filtering Recommendation Algorithm Based on Cloud Model, 简记为 CFRABCM), 以便实现更好的推荐效果。

## 1 CFRABCM总体架构

CFRABCM 主要包括基于云模型的用户评分项分析、基于云模型的相似性度量方法、基于云模型的推荐系统信任约束、基于云模型的协同过滤推荐算法这四个模块,其总体架构如图 1 所示。

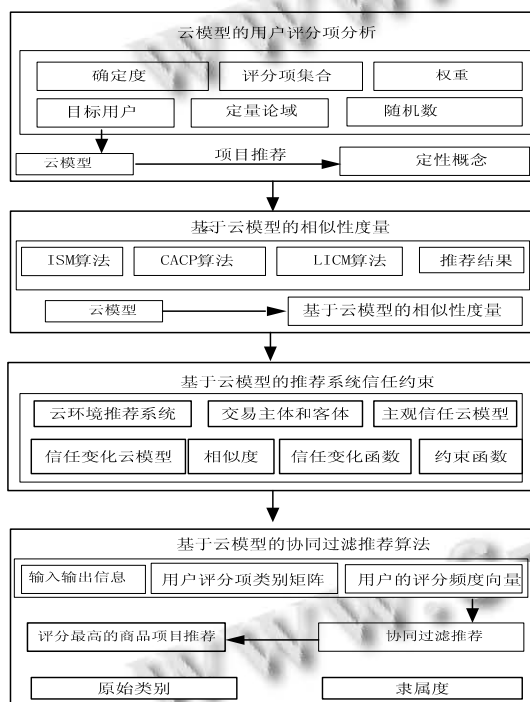


图 1 CFRABCM 总体架构

CFRABCM 架构主要流程如下:

首先需要得到基本用户与云模型的用户评分项之间的相似性矩阵,获得目标用户所属基于云模型的相似性度量计算向量,考虑客体信任度变化对信任决策的影响,有效解决用户评分项数据极端稀疏的问题,提高推荐质量。然后利用基于云模型改进的约束函数和信任变化函数建立起来的信任约束条件搜索类别所

属矩阵,确定目标用户的最近邻居,有效反映用户与推荐系统交互的程度以及提炼。最后结合基于云模型的用户评分项、基于云模型的相似性度量方法、基于云模型的推荐系统信任约束条件设计基于云模型的协同过滤推荐算法,对推荐系统中多个目标用户评分项的类属进行划分,准确寻找最近邻居目标用户,选取评分最高的商品项目推荐给目标客户,从而实现目标用户之间的推荐。

### 1.1 基于云模型的用户评分项分析

基于云模型的用户评分项是指在实际的电子商务网站中,所有商品项都通过用户评分划分到若干个项类中,以降低评分数据稀疏性,从而产生推荐列表。用户对项目的评分是为用户推荐的主要依据,如果用户对一些项目的评分比较相似,则他们对其他项目的评分也将会比较相似,电子商务商品项类实质上对应着用户的兴趣领域。因此,用户评分项关系着协同过滤推荐系统的质量和效率,它是计算用户相似性度量值的基础。对于用户评分项的分析,文献[5]采用目标用户  $u$  和  $v$  的评分项并集  $I_{uv}$  来计算用户相似性,有效解决了用户评分项数据极端稀疏的问题,提高了推荐质量。但是,  $I_{uv}$  中的用户  $v$  实际上可分为无推荐能力和有推荐能力两种类型,而文献[5]未能区别对待,导致在其算法中存在不必要的计算耗费。另一方面,云模型所表达的概念的整体特性可以用数字特征来反映,而基于云模型的用户评分项又是定性概念的度量,反映了在论域空间可被概念接受的取值范围。对此,首先给出云和云滴的定义,然后分析基于云模型的用户评分项。

定义 1. 云和云滴<sup>[10-12]</sup>

设  $U$  是一个用数值表示的定量论域,  $C$  是  $U$  上的定性概念,若定量值  $e \in U$  是定性概念  $C$  的一次随机实现,  $e$  对  $C$  的确定度  $\mu(e) \in [0,1]$  是有稳定倾向的随机数,且当  $\mu: U \rightarrow [0,1]$ , 有  $\forall e \in U, e \rightarrow \mu(e)$ , 则  $e$  在论域  $U$  上的分布称为云,记为  $C(e)$ 。每一个  $e$  称为一个云滴。

假设目标用户  $u$  的评分项集合  $I_u = \{u_i | i \in \text{自然数 } n\}$ , 目标用户  $v$  的评分项集合  $I_v = \{v_i | i \in \text{自然数 } n\}$ ,  $u_i$  和  $v_i$  的确定度  $\mu(u_i)$  和  $\mu(v_i) \in [0,1]$  是有稳定倾向的随机数,  $I_{uv} = I_u \cup I_v$  表示定量论域。若  $I_v < I_u$ , 即对于  $\forall i \in I_v$ , 都有  $i \in I_u$  成立, 定量值  $u_i, v_i \in I_{uv}$  是定性概念的一次随机实现, 则用户  $v$  的所有评分项都已被用户  $u$  评价过, 因此  $v$  不可能向  $u$  推荐项目; 若  $I_v > I_u$ , 即对于  $\forall i \in I_u$ , 都有  $i$

$\in I_v$  成立, 且当  $\mu: I_{uv} \rightarrow [0, 1]$ , 有  $\forall u_i \in I_{uv}, u_i \rightarrow \mu(u_i)$ , 则用户  $u$  的所有评分项都已被用户  $v$  评价过, 因此用户  $v$  必定向用户  $u$  推荐项目.

## 1.2 基于云模型的相似性度量方法

查询目标项目的若干邻居项目是通过度量不同项目之间的相似性来实现的. 相似性度量是指电子商务用户评价若干个项目时所具有的操作相近程度, 是项目查询的基础, 是协同过滤推荐算法研究中的热点. 传统的相似性度量方法主要包括余弦相似性度量、修正的余弦相似性度量和相关相似性度量这 3 种方法<sup>[2,4,7]</sup>, 传统的相似性度量方法考虑了不同用户的评分尺度问题, 减去了用户对项目的平均评分. 但传统的相似性度量方法在用户项目数据急剧增加的情况下不能准确地反映数据稀疏对相似性度量的影响, 算法质量较为低下. 对此, 文献[2]提出的改进的相似性度量方法(ISIM)能根据数据的稀疏状况自适应地调节相似度的大小, 更准确地反映数据稀疏情况对相似性计算的影响; 文献[4]基于项目类别的修正条件概率相似性(CACP)计算方法在对项目进行分类的基础上采用条件概率来计算项目之间的相似性; LICM方法<sup>[10-12]</sup>在用户投票整体层面上粗粒度考虑用户的相似性和对象类的整体信息, 避免了基于向量的相似度计算方式严格匹配对象属性的不足. 但是随着电子商务系统规模的扩大, 用户数目和项目数据急剧增加, 用户评分数据出现极端稀疏性, 使得得到的最近邻居集合不够准确, 导致算法的推荐质量降低<sup>[10-12]</sup>. 此外, 作为一种定性定量转换的模型, 云模型能够实现定性概念与其数值表示之间的不确定性转换<sup>[10-12]</sup>. 因此本节结合使用云模型、ISM算法、CACP算法、LICM算法的优点, 提出一种改进的基于云模型的相似性度量方法.

在第 1.1 节定义 1 和假设的基础上, 利用 LICM 算法, 则基于云模型的相似性度量公式如式(1)所示:

$$S(u_i, v_i) = \frac{|I_{uv}| \cdot \sum_{i=1}^n \mu(u_i)}{|I_u| \cdot |I_v| \cdot \sum_{i=1}^n \mu(v_i)} \quad (1)$$

式(1)的合理性和改进意义如下: 与传统的只基于用户或基于项目的相似性度量方法相比, 式(1)根据当前评分数据的稀疏情况, 动态调节相似度计算值, 自适应地调节目标用户和目标项目的最近邻对最终推荐结果的影响权重, 真实地反映彼此之间的相似性, 能显著地提高推荐系统的推荐质量. 改进算法所获得的推荐

效果有较大提高, 克服了传统基于向量的相似度比较方法严格匹配对象属性的不足. 因此相比改进前的 ISM、CACP、LICM 方法, 式(1)更好地解决了在数据稀疏的情况下用户相似性度量的自适应调节问题.

## 1.3 基于云模型的推荐系统信任约束

在面向大众用户的云环境推荐系统中, 当选择推荐对象时, 基于云模型的相似性度量计算成了主要的衡量指标. 但是, 在实际的用户操作过程中, 由于云模型的高度复杂性、随机性和模糊性, 且推荐系统存在大量的交易主体和客体, 相似性度量计算可能仅仅针对少数几个商品项打分, 在一些特殊的状况下, 甚至可能仅针对一个共同评分的商品项打分, 这样的相似性度量计算, 存在较大的偶然性. 另一方面, 推荐系统提供了对客体进行信任约束评价的评分机制, 并根据这种信任约束, 辅助主体完成客体选择. 因此研究基于云模型的推荐系统信任约束以保证云环境推荐系统的用户交易满意度已成为一个重要的研究课题.

### 定义 2. 基于云模型的推荐系统信任约束

基于云模型的推荐系统信任约束是指在云计算环境下, 在用户主体信任决策过程中, 通过基于云模型的用户评分项和相似性度量方法, 综合主观信任云模型和信任变化云模型<sup>[10-12]</sup>对用户评分进行条件约束, 定性或定量刻画客体信任度差异性.

该定义的主要思想是基于云模型的用户评分项和相似性度量方法, 通过合理表达信任度变化率的随机性和模糊性, 综合主观信任云模型的约束函数和信任变化云模型的信任变化函数建立用户评分矩阵并划分项目类属, 对信任度的稳定性进行量化评价, 使得  $I_u$  与  $I_v$  之间的相似度达到最小值  $S(u_i, v_j)_{min}$ , 而  $\mu(u_i)$  和  $\mu(v_i)$  之间的相似度达到最大值  $S(u_i, v_j)_{max}$ . 其中基于云模型的相似性度量方法、主观信任云模型的约束函数和信任变化云模型的信任变化函数起核心作用, 它们可以把用户对所有项目的评分作为若干个向量进行信任度的计算, 再根据基于云模型的用户评分项, 把整体信任约束特征表示为定量数值, 得到信任约束的输入云滴, 实现概念空间与信任空间的相互转换. 但已有的主观信任云模型的约束函数和信任变化云模型的信任变化函数不足之处在于没有考虑信任本身的模糊性以及用户项目的分类信息, 其推理建立在纯粹的概率模型上, 从而影响了推荐质量. 因此需要对传统的主观信任云模型的约束函数和信任变化云模型的信任变化

函数的定义进行改进.

### 1.3.1 约束函数的改进

在 1.1 节定义 1 及相关假设的基础上, 假设定量论域  $I_{uv}=I_u \cup I_v$  是云的信任度空间, 作为基本用户的集合; 推荐系统中的数据源  $D=\langle I_{uv}, I_u, I_v \rangle$ ;  $I=\{I_1, I_2, \dots, I_n\}$  是项目集合;  $u_i$  和  $v_i \in I_{uv}$  是定性概念  $u_i$  和  $v_i$  的一次定量信任约束;  $m \times n$  阶矩阵  $R$  是基本用户对各项目的评分矩阵;  $u_i$  和  $v_i$  的确定度  $\mu(u_i)$  和  $\mu(v_i) \in [0, 1]$  是有稳定倾向的随机数; 给定目标用户  $a_i$  及其项目评分向量  $A(1, n)$ ;  $\mu: I_{uv} \rightarrow [0, n]$ ,  $\forall u_i \in I_{uv}, u_i \rightarrow \mu(u_i)$ ; 对于  $\forall i \in I_{uv}$ , 将  $S(u_i, v_i)$  最大的  $n$  个基本用户组成集合. 则改进的主观信任云模型的约束函数用数学公式表示为如式(2)所示:

$$f(a_i) = \frac{\left( \prod_{i=1}^n u_i \cdot \mu(u_i) \right) \cdot |D| \cdot |I| \cdot |A|}{\left( \sum_{i=1}^n v_i \cdot \mu(v_i) \cdot S(u_i, v_i) \right) \cdot |R|} \quad (2)$$

其中  $f(a_i)$  表示主观信任云的约束函数, 它的值决定了主观信任云模型的具体实现. 与改进前的主观信任云模型的约束函数只关注随机变量  $u_i$  不同的是, 式(2)还具有通过  $a_i$ 、 $\mu(u_i)$ 、 $S(u_i, v_i)$  因子提供离散值或连续值信任约束机制的功能, 考虑信任本身的模糊性以及用户项目的分类信息, 并且不强制要求对信任空间内定性概念进行排序及规定信任等级, 能够体现云模型的定性定量转换的特点.

### 1.3.2 信任变化函数的改进

信任变化是指相邻时间窗口间客体信任度的变化率<sup>[10-12]</sup>. 在 1.1 节定义 1 及相关假设的基础上, 假设把信任度变化空间  $E=(-\infty, \infty)$  作为云的定量论域  $U$ ; 目标用户的预测评分集合  $P=\{p_1, p_2, \dots, p_n\}$ ;  $e \in E$  是信任变化空间上的定性概念; 实际评分集合  $Q=\{q_1, q_2, \dots, q_n\}$ ;  $x \in E$  是定性概念  $e$  的一次定量评价; 目标用户的预测评分与实际评分之间的偏差  $MAE$ ;  $x$  对  $e$  的确定度  $\mu(x) \in [0, 1]$  是有稳定倾向的随机数, 即  $x$  是随机变量  $e$  的某一具体取值;  $\mu: E \rightarrow (-\infty, \infty), \forall x \in E, x \rightarrow \mu(x)$ ,  $\mu$  主要用于基于式(1)云模型的相似性度量公式建立信任决策词条. 则信任变化云模型的信任变化函数可以按式(3)进行改进:

$$TC(u_x, v_x) = \frac{|P| \cdot |U| \sum_{x=1}^n x \cdot MAE \cdot S(u_x, v_x)}{|Q| \cdot |e| \cdot \prod_{x=1}^n \mu(x)} \quad (3)$$

其中  $TC(u_x, v_x)$  表示信任变化云模型中用户  $u_x$  变到用户  $v_x$  的趋向性. 信任变化函数  $TC(u_x, v_x)$  表示在定量论域  $U$  范围里, 客体信任度随时间而发生变化. 相比改进前的信任变化云模型的信任变化函数, 式(3)避免了相似性度量计算中侧重利用相关性而非相似性的弱点, 可以有效平衡基于云模型的用户评分项及推荐结果所带来的不稳定影响, 使得那些缺少共同评分项目但有整体共同偏好的用户变得可实现, 有效地缓解了用户评分数据稀疏的情况所带来的问题.

## 2 基于云模型的协同过滤推荐算法

本节结合第 1.1 节基于云模型的用户评分项分析、第 1.2 节基于云模型的相似性度量方法、第 1.3 节基于云模型的推荐系统信任约束、改进的约束函数、改进的信任变化函数, 设计基于云模型的协同过滤推荐算法, 对推荐系统中多个目标用户评分项的类属进行划分. 首先需要得到基本用户与云模型的用户评分项之间的相似性矩阵, 获得目标用户所属基于云模型的相似性度量计算向量, 然后利用基于云模型的推荐系统信任约束搜索类别所属矩阵, 确定目标用户的最近邻居.

算法的主要设计步骤如下<sup>[5,7]</sup>:

输入: 目标用户  $u$  的评分项集合  $I_u$ ; 目标用户  $v$  的评分项集合  $I_v$ ; 定量论域  $I_{uv}=I_u \cup I_v$ ;  $u_b, v_b, \mu(u_i), \mu(v_i)$ ; 项目评分矩阵.

输出: 基于云模型的相似性度量  $S(u_i, v_j)$ ; 数据源  $D=\langle I_{uv}, I_u, I_v \rangle$ ; 项目集合  $I=\{I_1, I_2, \dots, I_n\}$ ; 目标用户的预测评分集合  $P=\{p_1, p_2, \dots, p_n\}$ ; 定性概念  $e$ ; 实际评分集合  $Q=\{q_1, q_2, \dots, q_n\}$ .

第 1 步: 经过预处理计算产生  $n$  个原始类别, 根据用户评分项列表中各项目所属类别, 建立用户评分项所属类别矩阵  $S$ . 再根据  $S$  建立修正的用户评分项类别矩阵. 行代表用户, 用户数为  $m$ , 列表示项目, 项目数为  $n$ .

第 2 步: 按照定义 1, 根据用户评分项矩阵, 统计各个用户的评分频度向量. 通过式(1)计算  $u_i$  和  $v_i$  之间的相似度  $S(u_i, v_i)$ , 并抽取出最大值  $S(u_i, v_i)_{max}$  和最小值  $S(u_i, v_i)_{min}$ .

第 3 步: 抽取  $I_{uv}$  中所有属于  $I_u$  或  $I_v$  的任意项目组成项集, 对所有项目的评分组成评分矩阵. 通过式(2)改进的  $f(a_i)$  函数提供离散值或连续值信任约束机制,

产生  $m$  个云滴。

第4步: 按照式(3), 计算信任变化函数  $TC(u_x, v_x)$  对偏差 MAE 的隶属度  $L(u_x, v_x)$ , 公式如式(4)所示:

$$L(u_x, v_x) = TC(u_x, v_x) \cdot \sum_{x=1}^m \int_{|v|}^{|u|} x \quad (4)$$

第5步: 按照定义2, 计算未评分项和已未评分项的隶属度, 直到  $TC(u_x, v_x)$  不再发生变化, 使最近邻居用户数目稳定在  $\mu(x)$  附近。

第6步: 重复第2-5步, 计算所有用户对项目的平均评分, 根据待推荐用户、用户相似矩阵和用户项目矩阵, 在整个用户空间中查找用户, 并寻找最近邻居以便产生推荐。

第7步: 计算目标用户与最近邻居之间的相似性, 并设置阈值区间  $[M_i, M_j]$ , 如果  $S(u_x, v_x)_{max} > M_i \wedge S(u_x, v_x)_{min} < M_j$ , 将所有满足约束条件的用户评分项作为同一信任变化值划入同一个类别中, 从而选取评分最高的商品项目推荐给客户, 即实现协同过滤推荐。算法结束。

### 3 应用实验

#### 3.1 实验数据的整理

本文实验采用电影信息及用户对电影项目的评分信息建立模型, 并使用基于 Web 的推荐系统 MovieLens 数据作为数据集<sup>[11,12]</sup>, 其中一个数据集包含 95 位目标用户对 1876 部电影的 20000 条用户评分数据, 并从中选择 2101 条作为实验数据, 把数据集分为训练集和测试集。实验采用整个实验数据的 70% 作为训练集, 剩余的 30% 作为测试集<sup>[11,12]</sup>。  $|U_u|=4578$ ,  $|U_v|=5674$ ,  $|I|=987$ ,  $|I_{uv}|=8765$ ,  $\mu(u_i)=3458$ ,  $\mu(v_i)=6597$ ,  $|P|=1257$ ,  $|Q|=1346$ ,  $L(u_x, v_x)=14000$ 。目标用户包括老年、中年、青年、少年、儿童 5 种类型, 即目标用户评分个数为  $|U|=5$ 。每位目标用户分别对 10(预定阈值)部电影进行评分, 所有电影包含 12 种类别, 即  $M_i \geq 10$ , 其中包含了 569 个用户对 530 部电影的评分<sup>[11,12]</sup>。  $S(u_i, v_i)$  的取值范围限定为  $[0,1]$ ,  $TC(u_x, v_x)$  的评分值范围为 1,2,3,4,5。约束函数  $f(a_i)$  取值范围的集合 = {不满意, 一般满意, 较满意, 满意, 非常满意}, 相应的信任变化函数取值范围的集合 = {1,2,3,4,5}。

#### 3.2 实验云环境的搭建

实验云环境的硬件平台是配置 Intel Pentium IV,

4GHz CPU 和 1G RAM 的 PC, 操作系统运行 Windows XP, 算法采用 Java 编程实现, 数据库管理系统采用 SQL Server 2005。假设单位窗口长度内的评价数据数量相同, 则以当前决策时刻为时间原点, 共设计若干个云窗口, 并通过主窗口的窗口长度来计算其他云窗口的时效性权值。为简化实验, 本文随机地选取原始评价数据, 重新填充到每个云窗口。重复以上步骤, 可以看出, 本文实验环境体现出云环境的特点, 通过加权逆向云生成算法生成客体主观信任云数字特征值。

#### 3.3 实验结果分析

##### 3.3.1 评价标准

推荐效果的评价标准<sup>[1,7,10]</sup>主要有交互程度、推荐时间(参考值范围为 10-15 秒)、降维效果、自适应度(参考值范围为 16-21)、精确度、丢失率(信息丢失率, 不高于 31%)等。在统计精度度量方法中, 可以采用平均绝对误差(MAE)来直观地度量推荐质量<sup>[2]</sup>。

##### 3.3.2 相似性度量方法的实验结果比较

在实验中, 利用基于云模型的用户评分项分析将每个用户的评分信息分成若干等份, 采用交叉验证的方法, 每次取一份为测试集, 其余若干份组合在一起为训练集, 进行5次实验。通过实验发现,  $S(u_i, v_j)$  交互程度为“高”, 推荐时间为10.2, 降维效果为“较好”, 自适应度为23.4, 精确度为“高”, 丢失率为17.6%, MAE为0.49, 所有用户对项目的平均评分为“高”, 目标用户  $u$  和  $v$  的评分项集合  $I_u$  和  $I_v$  均限制在可以接受的范围内,  $S(u_i, v_j)$  各项指标均优于传统的相似性度量方法。因此相比改进前的相似性度量方法, 在多维推荐空间中,  $S(u_i, v_j)$  能根据当前评分数据的稀疏现象, 将语境段作为约束条件, 动态调节相似性度量的计算值, 自适应地选择预测目标以及调节目标项目的最近邻, 使得评分模型只包含用户对项目的评分信息的条件表达式, 真实地反映彼此之间的相似性, 能显著提高推荐系统的推荐质量。

##### 3.3.3 基于云模型的推荐系统信任约束的实验结果比较

采用传统的基于云模型的推荐系统信任约束算法, 从 10 个最近邻开始逐步递增, 对本文提出的改进后的主观信任云模型的约束函数  $f(a_i)$  和信任变化云模型的信任变化函数  $TC(u_x, v_x)$  方法进行了比较。此时,  $f(a_i)$ 、 $TC(u_x, v_x)$  取值范围均为  $[0,1]$ , 隶属度  $L(u_x, v_x)$  取值范围为  $[5,10]$ , 预定阈值  $M_i \geq 10$ , 均符合约束条件; 基于云

模型的推荐系统信任约束的交互程度为“高”，推荐时间为 10 左右，降维效果为“较好”，自适应度为 23 左右，精确度为“高”，丢失率为 17% 左右，MAE 为 0.5 左右。 $f(a_i)$ 、 $TC(u_x, v_x)$  各项指标均优于其他算法。因此相比改进前的主观信任云模型和信任变化云模，改进后的主观信任云模型的约束函数  $f(a_i)$  和信任变化云模型的信任变化函数  $TC(u_x, v_x)$  可以有效平衡目标用户以及产品群推荐结果所带来不确定性影响，有效缓解用户评分数据极端稀疏情况使用传统性度量方法带来的问题，显著提高了推荐系统的推荐质量。

### 3.3.4 CFRABCM 与其他协同过滤推荐算法的实验结果比较

为了便于验证基于云模型的协同过滤推荐算法的优势，首先分别对结合似然关系模型和用户等级的协同过滤推荐算法(collaborative filtering recommendation algorithm combining probabilistic relational models and user grade, CFRACPRMUG)<sup>[1]</sup>、综合用户和项目因素的协同过滤推荐算法(Collaborative filtering recommend-

dation algorithm based on both user and item, CFRABOBUI)<sup>[2]</sup>、集成项目类别与语境信息的协同过滤推荐算法(Collaborative filtering recommender algorithm for integrating item category and contextual information, CFRAIICCI)<sup>[4]</sup>、基于用户聚类的协同过滤推荐算法(collaborative filtering recommendation algorithm based on clustering basal users, CFRACBU)<sup>[7]</sup>等传统算法进行验证，然后对本文 CFRABCM 进行验证。

具体的实验步骤如下：对数据的云模型类别文件进行分析，依据第 1.3 节改进的算法，得到云模型类别矩阵  $S$ ，从而使得评分模型只包含用户对项目的评分信息的条件表达式，进而得到修正的类别矩阵  $P$ 。最后在训练集的基础上对未评分项目的评分进行预测。实验的结果是：候选近邻数目减少，对推荐质量的影响较小，CFRABCM 算法的效率明显提高；当  $TC(u_x, v_x)=3$  时，CFRABCM 比 CFRACBU 算法的 MAE 减少 0.1。各种协同过滤推荐算法综合性能对比结果如表 1 所示。

表 1 各种协同过滤推荐算法综合性能对比

算法实例	交互程度	推荐时间	降维效果	自适应度	精确度	丢失率(%)	MAE
CFRACPRMUG	未测	15.7	较好	19.8	未测	30.5	0.53
CFRABOBUI	未测	17.8	未测	18.9	较高	30.7	0.56
CFRAIICCI	未测	18.5	较好	17.8	未测	29.9	0.60
CFRACBU	一般	19.8	较好	19.1	较高	28.8	0.50
CFRABCM	高	10.2	较好	23.4	高	17.6	0.40

从表 1 可以发现，CFRABOBUI 的精确度为“较高”，而 CFRABCM 可以达到“高”，证明精确度有所提高；推荐时间方面，CFRACPRMUG 为 15.7，而 CFRABCM 仅为 10.2，证明系统的推荐速度有所提高；自适应度方面，CFRACPRMUG 可以到达 19.8，而 CFRABCM 可以达到 23.4；交互程度方面，CFRACBU 为“一般”，而 CFRABCM 可以达到“高”；丢失率方面，CFRACBU 可以达到 28.8%，而 CFRABCM 仅为 17.6%，数据丢失程度不高；MAE 方面，CFRACPRMUG 为 0.53，而 CFRABCM 仅为 0.40。因此，改进后的基于云模型的相似性度量方法  $S(u_i, v_j)$  真实地反映了目标用户之间的相似性，显著地提高了推荐系统的质量。改进的主观信任云模型的约束函数  $f(a_i)$  能提供离散值或连续值信任约束机制的功能，能够体现云模型的定性定量转

换的特点。改进的信任变化云模型的信任变化函数  $TC(u_x, v_x)$  可以有效地缓解用户评分数据稀疏的情况所带来的问题。因此 CFRABCM 的综合推荐效果明显优于改进前的各种算法。

实验结果表明本文对基于云模型的相似性度量方法、主观信任云模型的约束函数以及信任变化云模型的信任变化函数所进行的改进是合理的。本文所提的基于云模型的协同过滤推荐算法具有较高的可行性，可以用于接收用户对电子商务商品项目评分并提供相应的推荐列表。

## 4 结语

基于云模型的用户评分项、基于云模型的相似性度量方法、基于云模型的推荐系统信任约束、主观信

任云模型的约束函数和信任变化云模型的信任变化函数在 CFRABCM 的研究中起着核心作用. CFRABCM 真实地反映了目标用户之间的相似性, 有效地解决了数据稀疏所带来的问题, 能显著地提高协同过滤推荐系统的推荐质量. 但是由于云模型和电子商务协同过滤推荐系统的高度复杂性, 基于本文改进算法所建立起来的 CFRABCM 并不能适用于所有的电子商务协同过滤推荐系统, 尤其在推荐时间、丢失率等方面仍然有待改善, 这也是笔者的后续研究任务.

### 参考文献

- 1 高滢, 齐红, 刘杰, 等. 结合似然关系模型和用户等级的协同过滤推荐算法. 计算机研究与发展, 2008, 45(9): 1463-1468.
- 2 黄裕洋, 金远平. 一种综合用户和项目因素的协同过滤推荐算法. 东南大学学报, 2010, 40(5): 917-921.
- 3 LI Q, KIM BM. An approach for combining content-based and collaborative filters. Proc. of the 9th IntWorkshop on Information Retrieval with Asian Language. Morristown, NJ, USA. Association for Computational Linguistics. 2008. 1224.
- 4 姚忠, 吴跃, 常娜. 集成项目类别与语境信息的协同过滤推荐算法. 计算机集成制造系统, 2008, 14(7): 1449-1455.
- 5 李聪, 梁昌勇, 马丽. 基于领域最近邻的协同过滤推荐算法. 计算机研究与发展, 2008, 45(9): 1532-1538.
- 6 黄创光, 印鉴, 汪静. 不确定近邻的协同过滤推荐算法. 计算机学报, 2010, 33(8): 1370-1376.
- 7 李涛, 王建东, 叶飞跃. 一种基于用户聚类的协同过滤推荐算法. 系统工程与电子技术, 2007, 29(7): 1178-1182.
- 8 Leung CW, Chan SC, Chung F. A collaborative filtering framework based on fuzzy association rules and multiple-level similarity. Knowledge Information Systems, 2008, 9(4): 492-511.
- 9 Sun XH, Kong FS, Ye S. A comparison of several algorithms for collaborative filtering in startup stage. Proc. of the 2006 IEEE International Conference on Networking, Sensing and Controlling. Los Alamitos, CA, USA. IEEE Computer Society Press. 2006. 25-28.
- 10 孟祥怡, 张光卫, 康建初, 等. 基于云模型的主观信任管理模型研究. 系统仿真学报, 2007, 19(14): 3310-3317.
- 11 周军锋, 汤显, 郭景峰. 一种优化的协同过滤算法. 计算机研究与发展, 2004, 41(10): 1842-1847.
- 12 李德毅, 刘常昱. 论正态云模型的普适性. 中国工程科学, 2004, 6(8): 28-34.