

# 一种利用协同过滤预测和模糊相似性改进的基于内容的推荐方法\*

姜书浩<sup>1</sup> 薛福亮<sup>2</sup>

<sup>1</sup>(天津商业大学信息工程学院 天津 300134)

<sup>2</sup>(天津财经大学商学院 天津 300222)

**摘要:**【目的】基于模糊相似性的协同过滤预测和多样性选择算法对基于内容的推荐方法进行改进,提高推荐质量。【应用背景】基于内容过滤的推荐系统(CB-RS)目前已经有比较成功的应用。但是,推荐的多样性、项目特征的表示、用户偏好的建模仍然是基于内容推荐的关键。【方法】对用户特征提出采用一种新的模糊化表示方法,并根据此特征表示进行用户的相似性计算,在此基础上整合协同过滤的多样性预测,对基于内容的推荐方法进行改进。【结果】通过实验,该推荐方法在平均绝对误差、覆盖率以及多样性三方面明显优于当前流行的三种解决方案。【结论】推荐方案在一定程度上能够提高推荐质量,同时增强推荐的多样性。

**关键词:** 推荐系统 推荐多样性 模糊 CF-CBF 模糊相似性度量

**分类号:** TP301.6

## 1 引言

互联网的海量数据使得用户获取有价值的信息越来越困难,这一点在电子商务环境下尤为突出,为了应对上述挑战,智能推荐系统已成为为用户提供个性化推荐最重要的工具<sup>[1]</sup>,推荐系统(Recommender System, RS)基于用户的兴趣特征和购买行为,在复杂的网络环境中向用户提供其可能最感兴趣的项目。现在普遍使用的推荐系统中,主要包含两种最常用的方法:协同过滤(Collaborative Filtering, CF)和基于内容的过滤(Content-Based Filtering, CBF)。协同过滤通常是在用户群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,形成系统关于该指定用户对此信息喜好程度的预测,从而给予适当的推荐;而基于内容的过滤是通过比较包含在项目中的描述内容与用户感兴趣的内容,从而提供推荐,与其他相似的用户并没有任何关系<sup>[2]</sup>。当前,电子商务推荐系统在实际电子商务环境中已经有了比较好的应用,但仍然存

在许多问题。如协同过滤的推荐质量受到庞大而稀疏的用户评价数据的影响;基于内容的推荐则受推荐多样性限制,而严重影响推荐的效果。

基于内容的推荐系统通过分析项目描述来识别用户感兴趣的项目,但是项目的特征和用户的行为在本质上是高度主观的<sup>[3]</sup>,因此项目的代表性和用户行为建模是创建基于内容推荐系统的主要问题。一些学者提出了基于模糊集的个性化推荐方法,并介绍了该方法的依据和规则,同时在该领域也做了很多工作<sup>[4-6]</sup>。但是,基于内容的推荐系统在某些领域的应用是有缺陷的。例如:用户在购买单反相机时首先设定购买价格要低于 7 000 元、配备 SD 或 SDHC 卡、1500 万像素以上的全高清视频录制功能的单反相机。基于内容的推荐系统返回首选推荐的是一台包含上述功能的尼康 D7000。虽然这是一个不错的推荐,但如果第二、第三和第四的推荐只是在第一推荐的结果上稍作修改而做出的,那用户就不会太满意。如果用户决定选择其他

收稿日期: 2013-10-09

收修改稿日期: 2013-12-09

\*本文系教育部人文社会科学一般项目“电子商务环境下顾客购物偏好推荐及企业利润挖掘”(项目编号: 13YJC630195)的研究成果之一。

品牌,比如佳能,那基于内容的推荐系统就无法做出推荐。因此多样性问题是基于内容的推荐系统公认的缺陷。目前,解决这个问题的常见方法是在基于内容推荐的基础上结合协同过滤推荐,以消除基于内容推荐系统的多样性问题。一些学者提出了衡量推荐多样性指标。多样性在推荐系统中包含两个层次:用户间的多样性,衡量推荐系统对不同用户推荐不同商品的能力;用户内的多样性,衡量推荐系统对一个用户推荐商品的多样性<sup>[7]</sup>。

## 2 研究背景

基于内容的推荐与协同过滤推荐是目前推荐系统中的主流技术,协同过滤推荐通过对未评分项进行评分预测来实现。基于用户的协同过滤算法基于一个这样的假设:“跟你喜好相似的人所感兴趣的东西,也很有可能是你所感兴趣的”,基于用户的协同过滤就是找出某个用户其兴趣度最相似的邻居,从而根据其最相似邻居的兴趣偏好做出未知项的评分预测。协同过滤推荐相对于基于内容的推荐的最大优势在于推荐的新颖性。基于内容的推荐,主要是基于项目的特征和用户过去的历史偏好的匹配度实施推荐。基于内容的推荐系统难以区分待推荐项目的品质和风格,而且不能为用户发现新的感兴趣的项目,而协同过滤推荐系统存在可扩展性和稀疏性等问题<sup>[3]</sup>。因此,很多研究人员将协同过滤和基于内容的推荐技术结合起来,以克服它们各自的缺点。Fab 系统通过为每个用户生成一个特征模型<sup>[8]</sup>,并基于该特征模型计算用户的兴趣相似度,从而将协同过滤和基于内容的推荐进行整合。有些学者采用基于用户模糊聚类的协同过滤技术<sup>[9]</sup>,以及采用基于模糊聚类的协同过滤技术<sup>[10]</sup>都是对协同过滤技术进行的基于用户的模糊聚类的改进,目的也是解决协同过滤系统的稀疏性问题。

在应用系统中,即使是准确率比较高的推荐也不能保证用户对其推荐结果满意。原因是用户得到的推荐结果可能是非常流行的商品或用户熟悉的商品,对于这些商品或者商品,用户很可能早已从其他渠道得到,因此对用户来说这些推荐意义不大;而对其他相对陌生或新颖的产品,系统可能会因为评价较少等原因很少推荐或不予推荐。对于用户间的多样性,朱郁

筱等提出采用计算汉明距离来作为多样性的指标<sup>[7]</sup>。张富国等分析探讨推荐多样性的度量方法,提出了基于信任的提高推荐多样性的算法<sup>[11]</sup>。而对于用户内的多样性的研究相对并不多见,本文研究的多样性正是针对用户内产品多样性的改善。在此基础上,本文又在以下三方面做了改进:

(1) 基于本地和全局的模糊距离,并利用特征兴趣度测量设计了一个模糊的协同过滤模型。

(2) 通过引入一种“模糊的项目特征表示方法”和“相似性度量”提出了“模糊的基于内容推荐方案”。

(3) 利用模糊协同过滤推荐的项目多样性预测的优势,以改进模糊基于内容的推荐。提出了一种模糊的协同过滤与基于内容的混合推荐模型。

## 3 推荐框架

整个推荐框架如图 1 所示:

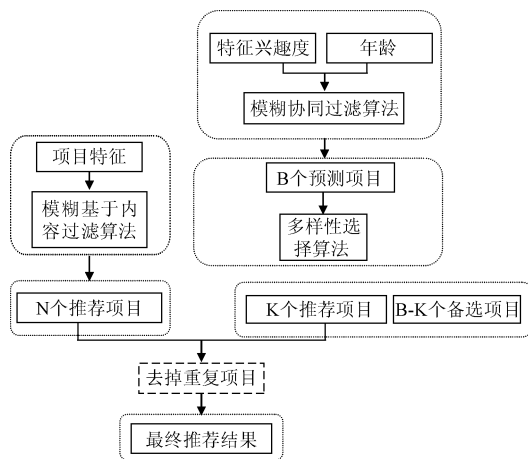


图 1 推荐流程

### 3.1 模糊协同过滤(模糊 CF)

本文总共采用了两次模糊算法,一次是在协同过滤中计算用户相似性时用到,另一次是在计算项目相似度时。无论是项目还是用户在进行分类时很可能存在某一用户(项目)既属于 A 类又属于 B 类,唯一不同的地方可能是隶属的程度不同,如果严格按照非 A 即 B 的方式,很可能造成运算结果的偏差,所以采用模糊的划分方式,将样本的隶属度值在 $[0,1]$ 划分到不同类别中。

实际应用中,过于清晰的特征描述并不能反映用户的实际决策需求<sup>[12]</sup>,例如,18 岁和 20 岁的两个用户

的年龄差为 2 年,而这两个用户属于同一年龄组,即青年。因此,对用户特征进行模糊化的描述具有很大的优势,反而能检索到最相似的用户群体<sup>[13]</sup>。使用不同模糊集能够用来模糊化各种不同特征的用户。首先,年龄被模糊化为三个模糊集:青年、中年和老年,如图 2 所示:

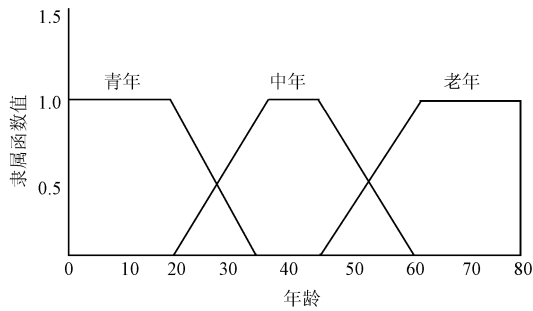


图 2 年龄隶属函数

此外,特征兴趣度模糊化分为 6 个模糊集合:很不感兴趣(V1)、不感兴趣(V2)、一般(V3)、感兴趣(V4)、很感兴趣(V5),以及非常感兴趣(V6)。特征兴趣度在不同的模糊集隶属函数值,如图 3 所示:

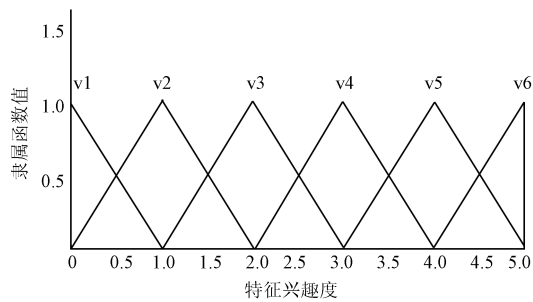


图 3 特征兴趣度隶属函数

用户之间的相似性度量是协同过滤推荐系统获取其最相似用户的关键。用户的特征被模糊化为不同的模糊集,为了计算两个用户之间的相似度,对于不同特征之间的局部模糊距离的计算,笔者提出以下公式:

$$LFD(a_i, b_i) = \frac{\sum_{j=1}^k |a_{ij} - b_{ij}|}{k} \quad (1)$$

其中,  $a_{ij}$  为用户  $a$  的第  $j$  个模糊集特征  $i$  的隶属度值,而  $k$  是模糊集的数量。为了获得用户之间的全局模糊距离(GFD),笔者提出通过以下公式得到:

$$GFD(a, b) = \frac{\sum_{i=1}^l LFD(a_i, b_i)}{l} \quad (2)$$

其中,  $l$  是  $a$  和  $b$  的用户特征的数目。对于用户之间的相似度,笔者使用全局模糊距离来进行计算,提出公式如下:

$$Sim(a, b) = 1 - GFD(a, b) \quad (3)$$

通过用户的相似性计算,协同过滤系统可为当前用户产生其最相似邻居集合  $N$ ,以下公式<sup>[14]</sup>可用来预测该用户在项目  $S$  上评分值  $P_{a,s}$ :

$$P_{a,s} = \bar{a} + \frac{\sum_{u \in N} Sim(a, u) \times (r_{u,s} - \bar{u})}{\sum_{u \in N} |Sim(a, u)|} \quad (4)$$

其中,  $\bar{a}$  为对用户  $a$  的平均评分,  $N$  表示一个邻居集合。

### 3.2 多样性选择算法

上述算法与大多数推荐算法存在一个相同的问题,推荐时只考虑了单个项目的准确度,而忽略了整个推荐项目集对用户满意度的影响。通过以上的模糊协同过滤计算,可以获得包含  $B$  个最优待推荐项目的集合,但该集合中有些待推荐项目可能存在很高的特征相似性,这会严重影响到用户对整个项目集的满意度,所以本文在此基础上提出多样性选择算法以消除上述集合中相似性过高的项目。由于推荐项目集还需进行二次推荐和精简,所以其中  $B$  的取值可以相对较大一些,这样既增加了推荐集中项目的多样性,同时在使用多样性算法后去掉相似的项目也不会增加推荐项目集的复杂度。

多样性选择算法将会从  $B$  个预测项目中选择  $K$  个最不相似的项目,  $K$  个项目的选择主要目的是用于提高基于内容推荐的多样性,其算法思想如下:

输入:前  $B$  个项目与它们的评分值。

输出:  $K$  个不同项目的集合  $R$ 。

步骤:

①从  $B$  个项目选择第一个项目,作为多样性集合  $R$  的当前项目;

②用公式(6)计算当前项目与剩余  $B-1$  个项目的相似度;

③将剩余  $B-1$  个项目中最不相似的项目加入到  $R$ ,并把它作为当前项目;

④对  $R$  中的每一个当前项目重复步骤①—步骤③,直到  $K$  个不同的项目被找到为止。

多样性选择算法的关键技术是项目间的相似度计算,推荐项目集在经过多样性选择算法处理后,会有  $B-K$  个项目从推荐集中滤出,滤出的项目对推荐

系统并非毫无意义,其被滤出的原因是与推荐集中某些项目具有较高的相似性,可能会给用户造成重复推荐,后续可根据与 K 个推荐项的相似性距离进行分组,在用户需要时作为推荐的备选项。

### 3.3 模糊基于内容的过滤(模糊 CBF)

本节提出一种模糊基于内容推荐的框架,其中模糊建模将会被用来描述项目特征。

#### (1) 项目特征描述

从特征的角度来描述项目。在 MovieLens 电影数据集里,电影的特征通过其所属风格分类来描述,如:Movie1(枪战、恐怖、爱情……),而大部分的项目都不只具有一个分类特征,很多项目之间都有 1-2 个分类特征是相同的。假设  $S = \{S_1, S_2, \dots, S_N\}$  作为项目的集合,而  $f$  为项目的特征,  $A = \{F_1, F_2, \dots, F_N\}$  为项目的特征集合,  $f$  可从集合  $A$  中取多个值,项目  $S_i$  隶属于特征  $f_j$  的高斯隶属函数取值可用以下公式<sup>[14]</sup>描述:

$$\mu_{f_j}(s_i) = \frac{r_k}{2\sqrt{\rho \times \gamma(r_k - 1)}} \quad (5)$$

其中,  $\gamma$  为与项目  $s_i$  与集合  $A$  相关联的特征的数目,  $r_k$  是  $f_j$  在集合  $A$  中的位置等级,  $\rho$  取值大于 1。

#### (2) 项目相似度计算

很多学者做过项目的相似度计算的研究,例如基于项目分类的相似性计算<sup>[15]</sup>。本文的计算方式为项目  $S_i$  和  $S_p$  可以分别定义为  $\{(f_j, \mu_{f_j}(s_i)) | j=1, 2, \dots, m\}$  和  $\{(f_j, \mu_{f_j}(s_p)) | j=1, 2, \dots, m\}$ , 项目之间的相似度计算可用以下公式<sup>[14]</sup>获得:

$$\text{Sim}(s_i, s_p) = \left( 1 - \frac{\sum_{j=1}^m |\mu_{f_j}(s_i) - \mu_{f_j}(s_p)|}{m} \right) \quad (6)$$

预测当前用户  $a$  未评价项目的评分值,公式<sup>[14]</sup>如下:

$$P_{a,s} = \frac{\sum_{t \in L} \text{Sim}(s,t) \times r_{a,t}}{\sum_{t \in L} |\text{Sim}(s,t)|} \quad (7)$$

其中,  $L$  是用户  $a$  喜欢的项目的集合。

### 3.4 混合方法(模糊 CF-CBF)

本文提出的模糊的协同过滤与基于内容的混合推荐方法的主要步骤如下:

(1) 基于特征兴趣度和年龄的混合特征,采用模糊协同过滤算法,获得前  $B$  个待推荐项目和它们的预测评分值,在表 2 中可以看到采用模糊协同过滤算

法(模糊 CF)获得的项目预测分值平均绝对误差比未采用模糊预测的简单协同过滤(CF)要低很多,这也充分说明了模糊方法的意义与优势。

(2) 通过多样性选择算法在预测的  $B$  项目中选择前  $K$  个最不相似的项目,实验中选择  $B$  的值为 20,对这 20 个项目选择第一个项目(项目 1)与剩余 19 个项目采用公式(6)进行相似性计算,选取相似性最低的项目作为当前项目(项目 4),将当前项目与剩余 18 个项目继续采用公式(6)进行相似性计算,以此类推。实验数据如表 1 所示:

表 1 项目间的相似性

	项目 1	项目 2	项目 3	项目 4	项目 5	项目 6	……
项目 1	1.0000	0.4131	0.7310	0.2344	0.6212	0.8760	
项目 2	0.4130	1.0000	0.6547	0.3580	0.5614	0.5432	
项目 3	0.7310	0.6547	1.0000	0.4622	0.8324	0.8824	
项目 4	0.2344	0.7467	0.4622	1.0000	0.5430	0.3421	
项目 5	0.6212	0.5617	0.8324	0.5430	1.0000	0.6547	
项目 6	0.8760	0.5432	0.8824	0.3421	0.6547	1.0000	
……							

从表 1 中可以看出,项目 1 与项目 2、项目 3、项目 5 以及项目 6 相似度是相对较高的,尤其是与项目 3 和项目 6,分别高达 0.7310 和 0.8760,如此高的相似度,在同一推荐结果中出现显然是严重影响了推荐结果的多样性,而通过多样性选择算法,最终选取的项目 1 与项目 4,其项目相似度仅为 0.2344,可以说这两个项目相似度极低,依据相似性最后可得项目 1、项目 4、项目 11 等 5 个项目序列为  $B$  中最不相似的项目集合,消除其多样性。

(3) 通过模糊协同过滤可以获得与当前用户具有相似偏好的用户最喜欢的且当前用户未评分的项目集合,但模糊协同过滤忽略掉了项目之间相似性,即忽略掉了与当前用户最喜欢的项目具有最大特征相似性的其他项目。因此将模糊协同过滤与模糊基于内容过滤进行融合,模糊基于内容的过滤可以获取 top  $N$  个最优推荐项目集合,本文  $N$  取值为 5,去掉模糊协同过滤 top  $K$  与模糊基于内容 top  $N$  中的重复项目,最后剩余的项目集合作为最终推荐项目集合。

## 4 实验框架和实验结果

为了证明本文提出方法的有效性,笔者在公开 MovieLens(ML)数据集上进行了实验。

### 4.1 实验设计

MovieLens 站点(<http://MovieLens.umn.edu/>)是一个基于 Web 的研究型推荐系统,用于接收用户对电影的评分并提供相应的电影推荐列表。MovieLens 电影数据集包含电影的属性、用户评分和用户人口学特征。该数据集包括 943 个用户为 1 628 部电影提供的评分,评分范围在 1-5 之间,分数值越大说明用户的喜爱程度越高。每个用户至少对 20 部电影进行了评分。

在实验中,选择了数据的三个子集,分别包含 50、100 和 150 名用户,依次称作 ML50、ML100 和 ML150。这样选择的目的是为了证明在不同数量的用户参与下本方案的有效性。选择的每一个子集都被随机分为两部分:60%的训练数据和 40%的测试数据。为了测试方案的性能,采用平均绝对误差(MAE)和覆盖率作为测试指标来计算实验预测的准确性<sup>[6]</sup>,其中测试集中 25%的项目作为多样性选择算法中 K 的值。

### 4.2 实验结果

为了说明本设计方案(模糊 CF-CBF)具有提供更好的推荐精度的能力,实验比较了其于模糊基于内容推荐(模糊 CBF)、模糊协同过滤(模糊 CF)和混合协同过滤的基于内容推荐(CF-CBF)的平均绝对误差和覆盖率,结果如表 2 所示:

表 2 四种方式平均绝对误差和覆盖率的比较

数据集	模糊 CBF		模糊 CF		CF-CBF		模糊 CF-CBF	
	平均绝对误差	覆盖率 (%)	平均绝对误差	覆盖率 (%)	平均绝对误差	覆盖率 (%)	平均绝对误差	覆盖率 (%)
ML50	.896 8	94.50	.891 5	87.60	1.724 0	98.40	.879 1	98.40
ML100	.868 9	97.40	.862 0	86.40	1.760 5	97.40	.846 9	98.50
ML150	.883 0	97.60	.874 4	86.50	1.680 0	97.60	.863 0	98.20

其中,平均绝对误差和覆盖率的计算基于不同的数据集超过 20 次实验的平均值。推荐性能优越的方案应该有更低平均绝对误差和更高覆盖率。从表 2 可以清楚地看出,从预测准确性的角度来看,本方案的表现要好于其他几种技术方式。在 ML100 的案例中,20

次不同的实验运行的平均绝对误差如图 4 所示:

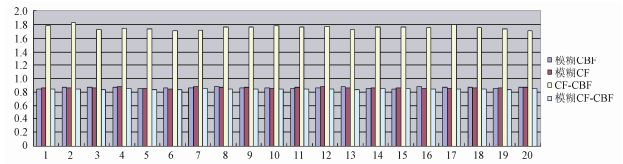


图 4 100 人运行 20 次的平均绝对误差

同时,由于本方案经过多样性选择算法的二次过滤,所以在第一次选择中适当增大了选择范围 K 的数值,其目的是通过扩大选择基数来提高项目推荐的多样性。实验分别选择采用多样性选择算法的最终推荐项目(项目 1、项目 4、项目 11、项目 15、项目 17)和未采用多样性选择算法的 top B 中前 5 个推荐结果的项目的相似性比较结果,如表 3 和表 4 所示,采用多样性选择算法的推荐结果相似性明显低于未采用的,其中未采用多样性选择算法的 top 5 推荐结果中项目 2、项目 3、项目 5 因具有较高的相似性,因此被多样性选择算法淘汰。从实验结果看出,多样性选择算法达到了增加项目多样性的目的,同时还增强了推荐结果的可信度。

表 3 采用多样性选择算法的推荐结果

	项目 1	项目 4	项目 11	项目 15	项目 17
项目 1	1.000 0	0.234 4	0.312 0	0.384 7	0.260 0
项目 4	0.234 4	1.000 0	0.295 4	0.358 0	0.362 3
项目 11	0.312 0	0.295 4	1.000 0	0.276 0	0.384 5
项目 15	0.384 7	0.358 0	0.276 0	1.000 0	0.339 0
项目 17	0.260 0	0.362 3	0.384 5	0.339 0	1.000 0

表 4 未采用多样性选择算法的推荐结果

	项目 1	项目 2	项目 3	项目 4	项目 5
项目 1	1.000 0	0.413 1	0.731 0	0.234 4	0.621 2
项目 2	0.413 0	1.000 0	0.654 7	0.358 0	0.561 4
项目 3	0.731 0	0.654 7	1.000 0	0.462 2	0.832 4
项目 4	0.234 4	0.746 7	0.462 2	1.000 0	0.543 0
项目 5	0.621 2	0.561 7	0.832 4	0.543 0	1.000 0

## 5 结 语

本文提出了一种模糊的项目特征描述方法,以及一种模糊的用户行为偏好建模,在此基础上着重探讨

了基于内容推荐的多样性的问题。本文将模糊协同预测与模糊基于内容推荐进行融合,减轻了基于内容推荐预测多样性的问题,从而提高了推荐质量。为了验证本文提出的混合模糊协同过滤与基于内容推荐方案的有效性,设计了一个实验方案与其他三种主流技术(模糊的基于内容推荐、模糊协同过滤和混合协同过滤与基于内容推荐)进行比较,实验结果表明,本设计方案优于其他三种方式。

目前的研究内容仅局限于特定的电影推荐系统,在今后的工作中,会将该方法拓展到其他特征项目来进一步完善,以期提高基于内容推荐的精确率。相信随着该方法的成熟,会很快推广到其他领域,例如音乐、书籍甚至电子产品等。但是在这些应用领域,计算项目特征时通常很少能获得用户对项目的评分,所以在用户评价较少甚至无评价的情况下生成用户对产品的评价是本算法需要继续探讨的问题<sup>[16,17]</sup>。

#### 参考文献:

- [1] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [2] 曾春, 邢春晓, 周立柱. 基于内容过滤的个性化搜索算法[J]. 软件学报, 2003, 14(5): 999-1004. (Zeng Chun, Xing Chunxiao, Zhou Lizhu. A Personalized Search Algorithm by Using Content-Based Filtering[J]. Journal of Software, 2003, 14(5): 999-1004.)
- [3] Zenebe A, Norcio A F. Representation, Similarity Measures and Aggregation Methods Using Fuzzy Sets for Content-Based Recommender Systems[J]. Fuzzy Sets and Systems, 2009, 160(1): 76-94.
- [4] 黄洪, 杨卓俊, 王奔. 模糊逻辑在电子商务商品推荐系统中的应用[J]. 计算机系统应用, 2012, 21(3): 171-175. (Huang Hong, Yang Zhuojun, Wang Ben. Application of Fuzzy Logic to E-commerce Recommendation System of Commodity [J]. Computer Systems & Applications, 2012, 21(3): 171-175.)
- [5] Shih Y Y, Liu D R. Product Recommendation Approaches: Collaborative Filtering via Customer Life Value and Customer Demands[J]. Expert Systems with Applications, 2008, 35(1-2): 350-360.
- [6] Kant V, Bharadwaj K K. Incorporating Fuzzy Trust in Collaborative Filtering Based Recommender Systems[C]. In: Proceedings of the 2nd International Conference of SEMCCO. Berlin, Heidelberg: Springer-Verlag, 2011: 433-440.
- [7] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报, 2012, 41(2): 163-175. (Zhu Yuxiao, Lv Linyuan. Evaluation Metrics for Recommender Systems [J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175.)
- [8] Balabanović M, Shoham Y. Fab: Content-based, Collaborative Recommendation[J]. Communications of the ACM, 1997, 40(3): 66-72.
- [9] 李华, 张宇, 孙俊华. 基于用户模糊聚类的协同过滤推荐研究[J]. 计算机科学, 2012, 39(12): 83-86. (Li Hua, Zhang Yu, Sun Junhua. Research on Collaborative Filtering Recommendation Based on User Fuzzy Clustering [J]. Computer Science, 2012, 39(12): 83-86.)
- [10] 王明佳, 韩景侗, 韩松乔. 基于模糊聚类的协同过滤算法[J]. 计算机工程, 2012, 38(24): 50-52. (Wang Mingjia, Han Jingti, Han Songqiao. Collaborative Filtering Algorithm Based on Fuzzy Clustering[J]. Computer Engineering, 2012, 38(24): 50-52.)
- [11] 张富国, 徐升华. 基于信任的电子商务推荐多样性研究[J]. 情报学报, 2010, 29(2): 350-356. (Zhang Fuguo, Xu Shenghua. Research on Recommendation Diversification in Trust Based E-commerce Recommendation Systems[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(2): 350-356.)
- [12] 牟向伟, 陈燕. 基于模糊描述逻辑的个性化推荐系统建模[J]. 计算机应用研究, 2011, 28(4): 1429-1433. (Mu Xiangwei, Chen Yan. Fuzzy Semantic Personalized Recommendation System Modeling[J]. Application Research of Computers, 2011, 28(4): 1429-1433.)
- [13] 严冬梅, 鲁城华. 基于用户兴趣和特征的优化协同过滤推荐[J]. 计算机应用研究, 2012, 29(2): 497-501. (Yan Dongmei, Lu Chenghua. Optimized Collaborative Filtering Recommendation Based on User's Interest Degree and Feature [J]. Application Research of Computers, 2012, 29(2): 497-501.)
- [14] 张慧颖, 薛福亮. 一种利用 Vague 集理论改进的协同过滤推荐算法[J]. 现代图书情报技术, 2012(3): 35-39. (Zhang Huiying, Xue Fuliang. A Collaborative Filtering Recommendation Algorithm Based on Vague Sets Rating Prediction [J]. New Technology of Library and Information Service, 2012(3): 35-39.)

- [15] 熊忠阳, 刘芹, 张玉芳, 等. 基于项目分类的协同过滤改进算法[J]. 计算机应用研究, 2012, 29(2): 493-496. (Xiong Zhongyang, Liu Qin, Zhang Yufang, et al. Improved Algorithm of Collaborative Filtering Based on Item Classification[J]. Application Research of Computers, 2012, 29(2): 493-496.)
- [16] Albadvi A, Shahbazi M. Integrating Rating-based Collaborative Filtering with Customer Lifetime Value: New Product Recommendation Technique[J]. Intelligent Data Analysis, 2010, 14(4): 143-155.
- [17] Cho Y H, Kim J K. Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce[J]. Expert Systems with Applications, 2004,

26(2): 233-246.

### 作者贡献声明:

姜书浩: 提出研究思路, 设计研究方案;

薛福亮, 姜书浩: 进行整个实验过程的设计;

姜书浩, 薛福亮: 数据采集、分析和处理, 其中平均绝对误差、覆盖率以及多样性的计算主要是薛福亮完成; 用户特征、项目特征以及相似性计算主要是姜书浩完成;

姜书浩: 论文起草;

姜书浩, 薛福亮: 论文的修改以及最终版本修订。

(通讯作者: 姜书浩 E-mail: mr\_jiang1980@163.com)

## An Improved Content-based Recommendation Method Through Collaborative Predictions and Fuzzy Similarity Measures

Jiang Shuhao<sup>1</sup> Xue Fuliang<sup>2</sup>

<sup>1</sup>(Information Engineering College, Tianjin University of Commerce, Tianjin 300134, China)

<sup>2</sup>(Business School, Tianjin University of Finance & Economics, Tianjin 300222, China)

**Abstract:** [Objective] The authors improve content-based recommendation method through Fuzzy similarity-based collaborative filtering prediction and diversity selection algorithm to raise the recommendation quality. [Context] There are many successful applications of Content Based Recommender Systems (CB-RS). Recommendation diversity, representation of items as well as users' preference modeling are still critical parts in this field. [Methods] An effective collaborative Content-Based Filtering (CBF) is developed by introducing an item representation scheme, and measuring similarity based on the scheme, and fuzzy similarity measure and fuzzy-CF into the fuzzy-CBF with diversity, in order to improve content-based recommendation method. [Results] Experiment results show that the proposed hybrid scheme (fuzzy CF-CBF) is better than the other three popular schemes in Mean Absolute Error(MAE), coverage and diversity. [Conclusions] The proposed scheme improves the recommendation quality, while enhances the recommended diversity.

**Keywords:** Recommender system Recommendation diversity Fuzzy CF-CBF Fuzzy similarity measures