

基于模糊 C 均值聚类有效性的协同过滤算法

葛林涛,徐桂琼

(上海大学 管理学院,上海 200444)

摘要:针对电子商务系统中传统协同过滤算法普遍存在的稀疏性和扩展性问题,文中提出了基于模糊 C 均值聚类有效性的协同过滤算法。首先依据四种不同的聚类有效性函数确定合理的聚类数区间,并在合理聚类数区间中根据 Xie-Beni 方法搜寻得到最佳的聚类数,然后使用最佳聚类数对项目进行模糊 C 均值聚类,将用户对单个项目的偏好转化为对相似群组的偏好,将稀疏的用户-项目偏好信息构造成密集的用户-模糊类的偏好信息,最后在项目所属类别中寻找目标用户的最近邻并产生推荐。在数据集 MovieLens 上与传统推荐算法相比的实验结果表明,新算法在平均绝对偏差、召回率、准确覆盖率等方面都有了较大改善,提高了推荐质量。

关键词:协同过滤;模糊 C 均值聚类算法;聚类有效性函数;最佳聚类簇数

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2016)01-0022-05

doi:10.3969/j.issn.1673-629X.2016.01.005

A Collaborative Filtering Algorithm Based on Fuzzy C-means Clustering Validity

GE Lin-tao, XU Gui-qiong

(School of Management, Shanghai University, Shanghai 200444, China)

Abstract: Considering the sparsity and the scalability of traditional collaborative filtering recommendation algorithms in electronic commerce system, a new collaborative filtering algorithm is presented based on fuzzy C-means clustering validity. Firstly, a reasonable cluster number range is presetted, and then an optimal cluster number is determined based on some representative fuzzy clustering validity functions and Xie-Beni method. Secondly, using the optimal number of cluster, this algorithm transforms the users' preferences of single item to similar groups with fuzzy C-means clustering, and sparse user-item preferences is established to dense user-fuzzy preferences. Finally, according to the item's cluster it finds the nearest neighbors of the object user and generates recommendations. The experimental results in MovieLens show that the new algorithm improves recommendation quality in MAE, recall and coverage.

Key words: collaborative filtering; fuzzy C-means clustering; clustering validity; optimal number of clustering

1 概述

互联网已成为当下人们获取信息的一个重要途径。然而,随着互联网信息的不断增加,用户很难从中挑选出自己真正需要的信息。针对互联网中的“信息过载”问题,推荐系统作为一种有效的解决手段被广泛应用在电子商务领域,它能主动为用户推荐需要但无法轻易获取的信息,在提供更具针对性的个性化服务的同时也提高了电子商务网站的销售量。因此,推荐系统已成为当下电子商务应用领域中的研究热点。其中,协同过滤推荐算法是目前最流行和最成熟的推

荐技术^[1]。

然而在实际应用中,推荐系统存在两个问题:一是稀疏性,由于用户评价过的物品只占有所有物品的一小部分,因此,用户-项目评分矩阵非常稀疏,从而影响推荐质量;二是可扩展性,随着系统中用户和资源的增多,性能越来越低。

针对以上问题,许多学者进行了深入研究并取得了一系列的研究成果^[2-14]。其中,将聚类算法应用到推荐系统中有助于缓解数据的稀疏性问题。如文献[5-7]使用硬聚类算法 K -means 方法对用户进行聚类,减少了寻找最近邻的计算量,然而硬聚类算法很难

收稿日期:2015-04-30

修回日期:2015-08-06

网络出版时间:2016-01-04

基金项目:国家自然科学基金资助项目(11201290,61104042)

作者简介:葛林涛(1990-),男,硕士研究生,研究方向为数据挖掘、个性化推荐;徐桂琼,博士,教授,博士生导师,研究方向为数据挖掘与人工智能、复杂系统建模。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160104.1505.028.html>

符合一个项目可能同时属于几个类的实际情况,同时,为了确定最佳的聚类数目,要计算不同聚类数目下的 MAE 值,增加了计算复杂度。文献[8]综合考虑用户对项目的关注和用户评价的影响,提出了改进的用户聚类的协同过滤推荐方法,使用了用户对项目关注相似度来替代传统的相似度。文献[9]提出了基于用户谱聚类的协同过滤推荐算法,并使用组合相似度^[10]在用户所属类中寻找最近邻并产生推荐,提高了推荐质量。文献[11-12]提出了基于模糊聚类的协同过滤算法,通过传递闭包法计算模糊相似关系,然后对项目进行聚类,缓解了由于用户项目评价数据稀疏而导致传统协同过滤中寻找最近邻居用户不准确的问题,然而在求解项目相似矩阵的过程中需要进行矩阵幂运算,当矩阵的阶数 n 较大时,工作量较大。文献[13-14]中将模糊 C 均值聚类应用到协同过滤算法中,有效缓解了数据稀疏性问题。文献[13]通过比较不同聚类数下的 RMSE 值,确定最佳聚类数,计算复杂度有待提高。文献[14]将用户数量的一半作为最佳聚类数,降低计算量,准确度有待提高。文献[15]对 Xie-Beni 方法^[16]进行了证明与改进,提出了一种优化的最佳聚类数选择策略,在提高了算法的准确性的同时也减少了计算量。

基于以上研究工作,文中提出了一种基于模糊 C 均值聚类有效性的协同过滤算法 (FCMC CF)。在使用模糊 C 均值聚类算法解决数据稀疏性问题的同时,引入聚类有效性判别函数,降低了寻找最佳聚类数的计算复杂度。在 MovieLens 公开数据集上的实验结果表明,提出的改进算法与传统协同过滤算法相比,较好地改善了平均绝对偏差、覆盖率、召回率指标,提高了推荐质量。

2 概念定义

2.1 协同过滤算法概述

协同过滤算法是推荐系统中应用最广、效果最好的算法之一。典型的协同过滤算法是基于用户的,其基本思想是通过用户-项目评分矩阵计算出用户之间的相似度,从中选出与用户最相似的前 k 个用户,根据这 k 个用户对当前用户的未评分项目的评分,预测当前用户对其未评分项目的评分,选出前 n 个推荐。协同过滤推荐算法的实现过程主要分为三步:

(1) 建立推荐模型。

协同过滤算法的输入是一个 $m \times n$ 的用户评价矩阵 \mathbf{R} 的数据, R_{ui} 表示第 u 个用户对第 i 项的评价值。一般把评分分成 5 个等级,评分越高说明这个用户对物品的评价越高。

(2) 查找最近邻居。

通过计算目标用户与其他用户之间的相似度,找出与目标用户最相似的“最近邻居”集,即对目标用户 u , $N_u = \{N_1, N_2, \dots, N_i\}$ 。首先计算用户之间的相似度,可采用 Pearson 相关系数、Cosine 相似性等度量方法。然后,对目标用户 u 产生一个以相似度 $\text{sim}(u, v)$ 递减排列的“邻居集合”。由于修正余弦相似度的准确性较高,因此文中采用的是修正的余弦相似性,见式(1)。

$$\text{sim}(u, v) = \frac{\sum_{i \in I_u} (R_{ui} - \bar{R}_u)(R_{vi} - \bar{R}_v)}{\sqrt{\sum_{i \in I_u} (R_{ui} - \bar{R}_u)^2} \sqrt{\sum_{i \in I_v} (R_{vi} - \bar{R}_v)^2}} \quad (1)$$

其中, \bar{R}_u 表示用户 u 的平均评分; \bar{R}_v 表示用户 v 的平均评分。

(3) 进行推荐。

最后,根据目标用户的 m 个最近邻居对项目的评分,可以进行该用户对其未评分项目的评分预测。计算方法如下:

$$P_{ui} = \bar{R}_u + \frac{\sum_{v \in S(u)} \text{sim}(u, v) \cdot (R_{vi} - \bar{R}_v)}{\sum_{v \in S(u)} |\text{sim}(u, v)|} \quad (2)$$

其中, P_{ui} 代表目标用户 u 对任意项目 i 的预测评分; $S(u)$ 表示邻居用户集合。

在预测出目标用户对未评价项目的评分后,使用 TOP- N 方法挑选出预测评分最高的 N 项,并对用户进行推荐。

2.2 模糊 C 均值聚类算法简介

聚类算法往往分为硬聚类算法和软聚类算法。将每个数据对象归到一个类被称为硬聚类算法,然而数据对象经常可以被归到几个类中,属于每个类的程度也不相同。因此, Ruspini 和 Bezdek 于 1981 年提出了 C 均值的改进算法—模糊 C 均值聚类算法 (FCMC)^[17]。模糊 C 均值聚类算法相比于其他算法,具有简单、高效等特点,因此应用在很多领域。模糊 C 均值聚类算法属于一种有效的软聚类算法,根据每个样本隶属于某个类的程度(即隶属度)来对样本进行聚类,所以模糊 C 均值聚类算法在复杂的样本上也可以得到比较好的聚类效果^[18-19]。

对于论域中的有限个对象集合 $\{X_1, X_2, \dots, X_n\}$, 对于每个数据样本属于某个类别的程度,在模糊 C 聚类算法中,使用隶属度来定义。例如对于 n 个向量 $X_i (i = 1, 2, \dots, n)$ 来说,FCMC 首先将它们分成 c 个模糊组,然后求出每个类别的聚类中心,同时,要最优化价值函数(见式(3)),也就是将非相似性指标值达到最小。

$$J(U, H_1, \dots, H_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (3)$$

其中, U 为隶属度矩阵; $u_{ij} \in [0, 1]$; H_i 为模糊聚类 i 的聚类中心; $d_{ij} = \|H_i - X_j\|$ 为第 i 个聚类中心与第 j 个数据点之间的欧氏距离; $m \in [1, \infty)$ 是一个加权指数。

使式(3)达到最小的必要条件为:

$$c_i = \sum_{j=1}^n u_{ij}^m x_j / \sum_{j=1}^n u_{ij}^m \quad (4)$$

$$u_{ij} = 1 / \sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)} \quad (5)$$

FCMC 算法包括以下四个步骤:

步骤 1: 随机生成 0 到 1 之间的数, 并用来随机化隶属度矩阵 U 。

步骤 2: 用表示价值函数最优化条件(见式(4))计算出 c 个聚类中心 $c_i, i = 1, 2, \dots, c$ 。

步骤 3: 用式(3)计算价值函数。如果它相对上次价值函数值的改变量小于某个阈值, 或者大于某个阈值, 那么算法停止。

步骤 4: 用隶属度(见式(5))重新计算隶属度矩阵 U , 然后重复步骤 2。

尽管 FCMC 算法能快捷地判定数据集中样本的类别, 但是该算法属于无监督学习, 必须先确定一个聚类簇数。如何确定最佳的聚类簇数以达到最好的聚类效果便是聚类有效性问题。

3 基于模糊 C 均值聚类有效性的协同过滤算法

在协同过滤推荐算法中, 影响推荐精度的关键问题在于用户评分数据的极度稀疏性。事实上, 由于协同过滤是在有相似爱好的用户间进行的, 用户更可能选择有相似特征的项目。因此, 可以用 FCMC 算法将属性特征上相似的项目聚成一类, 然后在一类项目中寻找用户的最近邻居, 以提高协同过滤算法的精度。

基于 FCMC 的协同过滤算法的关键在于确定最佳聚类数 c_{\max} 。因此, 文中首先计算不同聚类的有效性函数值, 然后采用改进的 Xie-Beni 方法^[15]选择 c_{\max} 。

3.1 四类有效性函数

模糊 C 均值聚类算法属于无监督学习, 为了确定 FCMC 算法的最佳聚类数, 国内外学者提出了一些有效性判别函数, 主要分为两类:

一类是基于模糊划分的方法, 认为好的聚类对应于数据集是较分明较清晰的, 代表函数如 Bezdek 提出的分割系数 V_{PC} 和分割熵 V_{PE} ^[20];

另一类是基于几何结构的方法, 认为每个子类应当是紧致的子类与子类相互间尽可能分离, 代表函数

有 Xie 和 Beni 的 V_{XB} ^[16]、Fukuyama 和 Sugeno 的 V_{FC} ^[21]、Kwon 的 V_K ^[22] 等。

$$V_{PE} = -\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n \mu_{ij} \ln(\mu_{ij}) \quad (6)$$

$$V_{XB} = \frac{\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^2 \|x_j - v_i\|^2}{n \min_{i \neq j} \|v_j - v_i\|^2} \quad (7)$$

$$V_{FS} = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \|x_j - v_i\|^2 - \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \|v_i - \bar{v}\|^2 \quad (8)$$

$$V_K = \frac{\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^2 \|x_j - v_i\|^2 + \frac{1}{c} \sum_{i=1}^c \|v_j - \bar{v}\|^2}{\min_{i \neq k} \|v_i - v_k\|^2} \quad (9)$$

事实上在评价聚类结果时这四个有效性函数不一定同时达到最优, FCMC 算法中最优的聚类数的确定至今没有一个最优准则, 只能结合多个有效性判别函数对最优聚类数进行判断。然而, 通常这样认为: 在多个有效性函数取得较优值对应的聚类数为最佳聚类数。

3.2 算法描述

基于模糊 C 均值聚类有效性的协同过滤算法(FCMC CF)的具体步骤如下:

步骤 1: 最佳聚类数的确定。

记项目总数为 n , 分别计算出聚类数从 2 到 \sqrt{n} 对应的聚类有效性函数(见式(6)~(9)) V_{PE} 、 V_{XB} 、 V_{FC} 、 V_K 值, 并根据改进的 Xie-Beni^[15]方法找出聚类有效性函数值出现单调性改变的第一个聚类数 c 作为最佳聚类数 c_{\max} 。

步骤 2: 构造用户-模糊簇评价偏好矩阵。

用 c_{\max} 对项目进行 FCMC 聚类, 产生相应的模糊簇, 并用式(10)计算每个项目对应某个模糊簇中的隶属度。

$$\mu_j(x_i) = \frac{\|x_i - m_j\|}{\sum_{k=1}^c (\|x_i - m_k\|)} \quad i = 1, 2, \dots, n; j = 1, 2, \dots, c \quad (10)$$

其中, x_i 为每个项目的属性特征向量; m_j 代表每个模糊簇中心的属性特征向量; $\|x_i - m_j\|$ 为项目 i 与模糊簇 j 的相近程度(即欧几里德距离); c 代表模糊簇的数目。

通过原始稀疏用户-项目评价矩阵中的评分值和隶属度矩阵中项目属于某个模糊簇的隶属度, 可以通过式(11)来计算得到用户对某个模糊簇的偏好值, 并构造出如表 1 所示的用户-模糊簇的偏好矩阵 UC, 相对来说该矩阵比原始矩阵的稀疏性较低。

表 1 用户-模糊簇评价矩阵

	C_1	C_2	...	C_c
U_1	PC_{11}	PC_{12}	...	PC_{1c}
U_2	PC_{21}	PC_{22}	...	PC_{2c}
...
U_n	PC_{n1}	PC_{n2}	...	PC_{nc}

其中,偏好值的计算如式(11)所示:

$$PC_{u,j} = \frac{\sum_{i \in I_u} R_{u,i} * \mu_j(x_i)}{\sum_{i \in I_u} \mu_j(x_i)}, j = 1, 2, \cdots, c$$

(11)

其中, $PC_{u,j}$ 为用户 u 对模糊簇 j 的评价值; $R_{u,i}$ 为用户 u 对模糊簇 j 中的项目 i 的评价值; I_u 为所有用户 u 已经评价的项目集合。

步骤 3:计算用户的最近邻居集合。

使用上一步计算出来的稀疏性较低的用户-模糊簇偏好矩阵来计算并形成目标用户的最近邻集合。用户 u 和 v 之间的相似度按式(1)计算。

步骤 4:为目标用户形成推荐。

从用户最近邻集合中选出目标用户的 K 个近邻,并根据式(2)来预测目标用户对没有评分过的项目的评分。然后,对预测评分进行排序,使用 TOP- N 策略进行推荐,即筛选出最高的 N 项推荐给目标用户。

4 实验与结果分析

4.1 数据集和度量标准

文中在实验中使用的数据是美国 Minnesota 大学 GroupLens 项目组提供的 Movielens 数据集 ml-100k 中的 u2 数据。该数据集包含了 943 名用户对 1 682 部电影的评价(评分值为数字 1 到 5,如果数值越高则用户喜爱该电影的程度越高),并含有电影项目的分类特征。该数据集仅包含评价过 20 部以上电影的用户评价数据,没有评分的电影数据占有所有数据的比重(稀疏度)为 93.7%。

文中使用平均绝对偏差(MAE)、召回率(Recall)和覆盖率(Coverage)来度量算法的推荐质量。其中,MAE 表示推荐算法对目标用户的预测评分与目标用户的实际评分的偏差值,这个指标可以用来评价推荐算法做出预测的误差,如果推荐质量越高,那么 MAE 的值越小。MAE 定义为:

$$MAE = \frac{\sum_{u,i \in T} |r_{ui} - p_{ui}|}{|T|}$$

(12)

一般来说,Recall 和 Coverage 可以较好地评价某个网站对目标用户做出推荐的预测准确率。

推荐结果的召回率定义为:

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$

(13)

推荐结果的覆盖率定义为:

$$Coverage = \frac{|\bigcup_{u \in U} R(u)|}{|I|}$$

(14)

4.2 实验结果

通过选取电影所属的类型作为项目的属性特征,文中做了以下几个实验来验证提出推荐算法的有效性。

(1) 确定最佳聚类数。

首先,通过比较不同聚类数相应的聚类有效性函数值来选出最佳聚类数 c_{max} 。实验结果如图 1 所示,具体的函数值如表 2 所示。

(a) V_{PE} 函数值

(b) V_{XB} 函数值

(c) V_{FC} 函数值

(d) V_K 函数值

图 1 不同聚类数对应的聚类有效性函数值

表 2 不同聚类数对应的有效性函数值

c	V_{PE}	V_{XB}	V_{FC}	V_K
2	0.807 7	0.000 0	797.17	0.000 0
3	1.126 3	0.000 0	403.52	0.000 1
4	1.356 4	0.000 1	221.33	0.000 1
5	1.540 6	0.000 1	114.06	0.000 2
6	1.697 2	0.000 2	43.678	0.000 5
7	1.833 6	0.000 3	-5.811 3	0.000 7
8	1.954 1	0.000 5	-42.414	0.000 9
9	2.061 8	0.000 5	-70.552	0.001 0
10	2.159 0	0.000 7	-92.845	0.001 4
11	2.247 5	0.018 9	-110.93	0.049 3
12	2.328 7	0.000 1	-125.91	0.000 1
...
40	3.486 7	0.565 1	-407.83	0.230 6

由上述结果可知,在 2 到 \sqrt{n} 的最佳搜索范围中,不同的聚类数 c 得到的 V_{PE} 值与 V_{FC} 值呈单调趋势,而

V_{XB} 与 V_K 函数值均在 $c = 12$ 时单调性发生改变。因此,根据 Xie-Beni 方法(见表 3),文中选取最佳聚类数 c_{max} 为 12。

表 3 Xie-Beni 方法确定的 c_{max}

V_{PE}	V_{XB}	V_{FC}	V_K
最佳聚类数 $c_{max} \geq 40$	$c_{max} = 12$	$c_{max} \geq 40$	$c_{max} = 12$

(2) MAE 比较。

使用上面得到的最优聚类数 c_{max} 对训练数据集合进行模糊 C 均值聚类,并比较了传统的协同过滤推荐方法(CF)和 $K - means$ 聚类协同过滤推荐方法(Kmeans CF)的结果,如图 2 所示,取最优聚类数 c_{max} 为 12 的实验结果。

图 2 不同算法 MAE 的比较

实验结果表明,FCMC 得到的 MAE 值要优于 CF 算法和 Kmeans CF。

(3) 召回率和覆盖率的比较。

在接下来的 TOP- N 实验中,选择 FCMC CF 算法与 CF 和 Kmeans CF 在召回率、覆盖率指标上进行比较。结果如图 3 和图 4 所示。

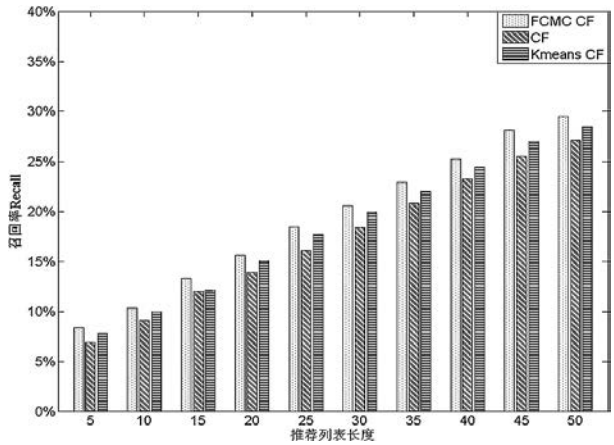


图 3 不同算法召回率的比较

4.3 实验结果分析

如图 2~4 所示,FCMC CF 算法在不同的最近邻水平下比 CF 算法以及 Kmeans CF 算法具有较小的 MAE 值,即新算法在推荐精度上有所改善;同时,FCMC CF 算法在不同的最近邻水平下比 CF 算法以及

Kmeans CF 算法具有更高的召回率和覆盖率,有效地改进了推荐精度。

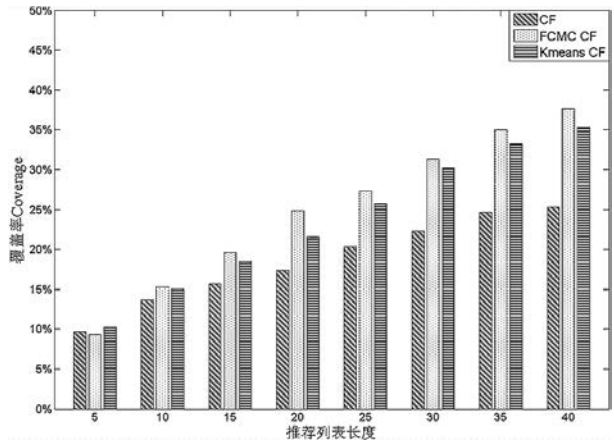


图 4 不同算法覆盖率的比较

5 结束语

当下,电子商务的流行导致了电商网站的规模越来越大,与此同时,网站上的用户和物品数据也随之增长,推荐系统面临着数据稀疏与可扩展性问题。

针对以上问题,文中提出基于模糊 C 均值聚类有效性的协同过滤算法。首先通过聚类有效性函数和改进的 Xie-Beni 方法^[15]确定了 FCMC 算法的最佳聚类数。其次根据属性特征对项目进行 FCMC 聚类,缓解了数据的稀疏问题,通过与 Kmeans CF 实验进行比较,更符合一个项目可能同属于几个类的实际情况。最后,对密集的用户-模糊簇的评价偏好矩阵进行协同过滤,由于簇的数目远远小于项目的数目,降低了推荐算法在时间上的计算复杂度。

该算法在 MovieLens 数据集上的实验结果表明,它可以较好地提高推荐系统的 MAE、召回率和覆盖率。然而,该算法在不同数据集上评价标准下的表现还需要进一步的研究才能证实。此外,社交媒体的流行为推荐系统提供了更多的用户社会信息,因此,如何将其应用到推荐系统中也值得进一步研究。

参考文献:

[1] Resnick P, Varian H R. Recommender systems[J]. Communications of the ACM, 1997, 40(3): 56-58.

[2] Sarwar B, Karypis G, Konstan J, et al. Analysis of recommendation algorithms for e-commerce[C]//Proceedings of the 2nd ACM conference on electronic commerce. [s. l.]: ACM, 2000: 158-167.

[3] Konstan J A, Miller B N, Maltz D, et al. GroupLens: applying collaborative filtering to Usenet news[J]. Communications of the ACM, 1997, 40(3): 77-87.

[4] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推

系统会自动为其构建 Mashup 页面,以满足学习者个性化的需求。在学习者的学习过程中,系统会基于其订阅情况以及学习过程的记录,为其构建兴趣模型,将该兴趣模型作为智能推送的基础。然而由于聚合的资源来源于网络各处,该平台尚未考虑资源间的关联关系,以及学习者在移动学习的过程中如何引入情境化,可作为今后的研究方向。

参考文献:

- [1] 李晓丽,王晓军. 移动学习模式探讨及系统架构设计[J]. 北京邮电大学学报:社会科学版,2007,9(5):10-13.
- [2] 李 春,朱珍民,叶 剑,等. 个性化服务研究综述[J]. 计算机应用研究,2009,26(11):4001-4005.
- [3] 祁晓晖,郑 滔. Mashup 研究综述[J]. 计算机应用,2010,30(S1):203-208.
- [4] 纪姗姗,李 宇,周 强. Mashup 工具的分析与研究[J]. 现代图书情报技术,2010(1):41-45.
- [5] 路 跃,刘旭东,孙海龙,等. MM:面向最终用户的 Mashup 移动应用开发环境[J]. 计算机科学与探索,2012,6(3):237-247.
- [6] 汪名森,王 强. Mashup 系统构建研究[J]. 现代图书情报技术,2009(5):34-38.
- [7] Yu Jin, Benatallah B, Casati F, et al. Understanding Mashup development[J]. IEEE Internet Computing, 2008, 12(5):44-52.
- [8] 苏会杰. 基于 REST 和 SOAP 的 Mashup 平台 Web 服务研究与实现[D]. 北京:北京邮电大学,2012.
- [9] 李 峰,李春旺. Mashup 关键技术研究[J]. 现代图书情报技术,2009(1):44-49.
- [10] 廖铁宸. 基于移动网络的混合型信息推送系统的研究[J]. 计算机工程与设计,2012,33(8):3268-3272.
- [11] 魏 伟,陈正鸣,吕 嘉. 基于 Atom 的个性化数据重用[J]. 计算机应用与软件,2013,30(12):236-239.
- [12] 易 明,曹高辉,毛 进,等. 基于 Tag 的知识主题网络构建与 Web 知识推送研究[J]. 中国图书馆学报,2011,37(4):4-12.
- [13] 王 辉,高成英,刘 宁. 服务器端 Mashup 开发平台的设计与实现[J]. 计算机工程,2010,36(10):262-264.
- [14] 侯志鑫. 移动学习环境下学习资源建设模式的研究[D]. 北京:北京交通大学,2014.
- [15] 高 静,段会川. JSON 数据传输效率研究[J]. 计算机工程与设计,2011,32(7):2267-2270.
- [16] Acilar A M, Arslan A. A collaborative filtering method based on artificial immune network[J]. Expert Systems with Applications, 2009, 36(4):8324-8332.
- [17] Tsi C F, Hung C. Cluster ensembles in collaborative filtering recommendation[J]. Applied Soft Computing, 2012, 12(4):1417-1425.
- [18] 查文琴,梁昌勇,曹 镭. 基于用户聚类的协同过滤推荐方法[J]. 计算机技术与发展,2009,19(6):69-71.
- [19] 李振博,徐桂琼,查 九. 基于用户谱聚类的协同过滤推荐算法[J]. 计算机技术与发展,2014,24(9):59-62.
- [20] 查 九,李振博,徐桂琼. 基于组合相似度的优化协同过滤算法[J]. 计算机应用与软件,2014,31(12):323-328.
- [21] 张海燕,顾 峰,姜丽红. 基于模糊簇的个性化推荐方法[J]. 计算机工程,2006,32(12):65-67.
- [22] 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.
- [23] Birtolo C, Ronca D, Armenise R, et al. Personalized suggestions by means of collaborative filtering: a comparison of two different model-based techniques[C]//Proceedings of 2011 third world congress on nature and biologically inspired computing. [s. l.]:[s. n.], 2011:444-450.
- [24] 张付志,常俊风,周全强. 基于模糊 C 均值聚类的环境感知推荐算法[J]. 计算机研究与发展,2013,50(10):2185-2194.
- [25] 于 剑,程乾生. 模糊聚类方法中的最佳聚类数的搜索范围[J]. 中国科学:E 辑,2002,32(2):274-280.
- [26] Xie X L, Beni G. A validity measure for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8):841-847.
- [27] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[M]. New York:Plenum Press, 1981.
- [28] Berget L, Mevik B H, Nis T. New modifications and applications of fuzzy C-means methodology[J]. Computational Statistics & Data Analysis, 2008, 52(5):2403-2418.
- [29] 马 军,邵 陆. 模糊聚类计算的最佳算法[J]. 软件学报, 2001, 12(4):578-581.
- [30] Bedzek J C. Cluster validity with fuzzy sets[J]. Journal of Cybernetics, 1973, 3(3):58-72.
- [31] Fukuyama Y, Sugeno M. A new method of choosing the number of clusters for the fuzzy c-means method[C]//Proceedings of the fifth fuzzy systems symposium. [s. l.]:[s. n.], 1989:247-250.
- [32] Kwon S H. Cluster validity index for fuzzy clustering[J]. Electronics Letters, 1998, 34(22):2176-2177.

(上接第 26 页)

- [1] 荐算法[J]. 软件学报,2003,14(9):1621-1628.
- [2] 邓爱林,左子叶,朱扬勇. 基于项目聚类的协同过滤推荐算法[J]. 小型微型计算机系统,2004,25(9):1665-1770.
- [3] Tsi C F, Hung C. Cluster ensembles in collaborative filtering recommendation[J]. Applied Soft Computing, 2012, 12(4):1417-1425.
- [4] 查文琴,梁昌勇,曹 镭. 基于用户聚类的协同过滤推荐方法[J]. 计算机技术与发展,2009,19(6):69-71.
- [5] 李振博,徐桂琼,查 九. 基于用户谱聚类的协同过滤推荐算法[J]. 计算机技术与发展,2014,24(9):59-62.
- [6] 查 九,李振博,徐桂琼. 基于组合相似度的优化协同过滤算法[J]. 计算机应用与软件,2014,31(12):323-328.
- [7] 张海燕,顾 峰,姜丽红. 基于模糊簇的个性化推荐方法[J]. 计算机工程,2006,32(12):65-67.
- [8] 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.
- [9] Birtolo C, Ronca D, Armenise R, et al. Personalized suggestions by means of collaborative filtering: a comparison of two different model-based techniques[C]//Proceedings of 2011 third world congress on nature and biologically inspired com-

基于模糊C均值聚类有效性的协同过滤算法

作者：[葛林涛](#)，[徐桂琼](#)，[GE Lin-tao](#)，[XU Gui-qiong](#)
作者单位：[上海大学 管理学院, 上海, 200444](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：
年，卷(期)：2016(1)

引用本文格式：[葛林涛](#). [徐桂琼](#). [GE Lin-tao](#). [XU Gui-qiong](#) [基于模糊C均值聚类有效性的协同过滤算法](#)[期刊论文]-
[计算机技术与发展](#) 2016(1)