一种基于耦合对象相似度的项目推荐算法

余永红[1,2] 陈兴国[2] 高 阳[2]
(南京邮电大学通达学院 南京 210003)[1] (南京大学计算机软件新技术国家重点实验室 南京 210093)[2]

摘 要 推荐系统根据用户的偏好为用户推荐个性化的信息、产品和服务等,能够帮助用户有效解决信息过载问题。基于内容的协同过滤算法缺少合适的度量指标用来计算项目之间的相似度。提出一种基于耦合对象相似度的项目推荐算法,即通过耦合对象相似度捕获项目特征频率分布相似性和特征依赖聚合相似度。首先从项目文本中抽取项目的关键特征,然后利用耦合对象相似度构建项目相似度模型,最后使用协同过滤的方法为活动用户推荐用户可能感兴趣的项目。在真实数据集上的实验结果表明,基于耦合对象相似度的推荐算法可以有效解决基于内容推荐系统的项目相似度度量问题,在缺失大量项目特征数据的情况下改进传统基于内容推荐系统的推荐质量。

关键词 基于内容的推荐系统,耦合对象相似度,协同过滤
中图法分类号 TP311 **文献标识码** A

Coupled Object Similarity Based Item Recommendation Algorithm

YU Yong-hong[1,2] CHEN Xing-guo[2] GAO Yang[2]
(College of Tongda, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)[1]
(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)[2]

Abstract Recommender systems are very useful due to the huge volume of information available on the Web. It helps users alleviate the information overload problem by recommending users with the personalized information, products or services. For content-based recommendation algorithm, there are few suitable similarity measures for the content-based recommendation methods to compute the similarity between items. This paper proposed a coupled object similarity based item recommendation algorithm. Our method firstly extracts item features from items, and then constructs item similarity model by using coupled object similarity measure. The collaborative filtering technique is then used to produce the recommendations for active users. Experimental results show that our proposed recommendation algorithm effectively solves the problem of similarity measure between items for recommendation algorithm and improves the quality of traditional content-based recommendation when lacking most of the item features.

Keywords Content-based recommendation system, Collaborative filtering, Coupled object similarity

1 引言

互联网和电子商务技术的发展为用户获取信息提供了便利,但是另一方面也使用户面临信息过载问题,用户被"淹没"在互联网和电子商务系统呈现给用户的大量数据中,无法快速、准确获取目标数据。推荐系统可以为用户提供个人化的信息、产品和服务,满足用户的个性化需求,从而有效地解决信息过载问题。推荐系统在现实生活中的典型应用包括Amazon、Last.fm、Google News、Netflix等,它们分别利用推荐技术为用户提供个性化的商品、音乐、新闻和电影。而且,越来越多的电子商务系统通过部署推荐系统来改进用户体验,提高用户的忠诚度,增加企业的销售收入。

协同过滤算法[1,2]是应用最广的推荐技术,并在电子商务系统中取得了巨大的成功。协同过滤算法基于用户之间评分行为相似性为目标用户推荐个性化商品、服务等。与基于内容的推荐系统不同,协同过滤分析用户的行为数据,不考虑具体的推荐项目(例如:电影、新闻、产品等)的属性,在进行个性化推荐时完全忽略对项目内容的分析。传统的基于内容的推荐算法[2,3]通过分析用户和项目的内容来进行推荐。Balabanovic[3]和Melville[4]等用实验表明基于内容的推荐算法可以在推荐精度上较大地改进协同过滤算法。然而,基于内容的推荐算法难于抽取项目相关的合理特征。而且,基于内容的推荐算法缺少合适的度量方法来计算项目之间的相似性。为了解决以上问题,本文提出一种基于耦合对象相似度(COS, Coupled Object Similarity)[5,6]的推荐算法。基于耦合对象相似度的推荐算法首先利用COS分析项目特征之间的关系,在考虑特征值频率分布相似度和特征依赖聚合度相似性基础上计算项目之间的耦合对象相似度;进行推荐时,根据

到稿日期:2013-05-20 返修日期:2013-07-16 本文受国家自然科学基金项目(61035003,60875011,60721002),科技部国际科技合作计划项目(2010DFA11030),973计划项目(2010CB327903),江苏省自然科学基金项目(BK2010054)资助。
余永红(1978—),男,博士生,讲师,主要研究领域为推荐算法、数据挖掘,E-mail:yuyh.nju@gmail.com;陈兴国(1984—),男,博士生,主要研究领域为数据挖掘、强化学习;高 阳(1972—),男,博士,教授,主要研究领域为数据挖掘和机器学习,E-mail:gaoy@nju.edu.cn。

与目标项目对象耦合相似度最近的 K 个项目的评分情况来计算当前活动用户对目标项目的评分。真实数据集上的实验结果表明，基于耦合相似度的推荐算法可以有效解决基于内容推荐算法中的项目之间相似度的度量问题，在一定程度上提高基于内容推荐系统的推荐质量。

本文第 2 节是推荐问题的形式化描述；第 3 节详细描述耦合对象相似度和基于耦合对象相似度的推荐算法；第 4 节在真实数据集上验证基于耦合对象相似度的推荐算法，并对实验结果进行分析；最后总结全文，并描述进一步的研究工作。

2 推荐问题描述

典型的场景下，一个推荐系统包含 n 个用户的集合 $U=\{u_1,u_2,\cdots,u_n\}$ 和 m 个项目的集合 $O=\{o_1,o_2,\cdots,o_m\}$。每个用户 $u\in U$ 对项目集合 O 中的部分项目进行评分。用户 u_i 评过分的项目集合表示为 O_{u_i} ($O_{u_i}\subseteq O$)。项目集合 O 中的每个项目 o_j 表示成一个特征向量 $o_j=\{a_{j1},a_{j2},\cdots,a_{jl}\}$，特征向量的每个值是从项目文本信息中抽取出的类别值。例如，如果项目集合 O 表示一组电影的集合，那么可以抽取“导演”、“演员”和“风格”等特征值来表示一个电影项目。

一般而言，推荐系统将用户对项目的评分数据转换为用户-项目矩阵 R_{n*m}。R_{n*m} 中的每项 r_{ij} 是用户 u_i 对项目 o_j 的评分，而且整数评分值 $r_{ij}\in[0,5]$，其中 0 值表示用户未对此项目进行评分。评分值越高意味着用户对当前项目越满意。

本质上，推荐系统的目标是利用各种数据挖掘和机器学习技术预测活动用户 u_a 对未评分过的项目 o_j 的评分值。

3 基于耦合对象相似度推荐算法

3.1 耦合对象相似度度量(COS)

在推荐系统中，从推荐项目内容中抽取的特征值都是类别型的值。例如，如果推荐项目是电影，可以抽取 director，actor 和 genre 特征值来描述一部电影，(“Hitchcock”，“Stewart”，“Thriller”)和(“Koster”，“Grant”，“Comedy”)特征向量值分别表示电影“Vertigo”和“Bishop's Wife”。如何计算电影“Vertigo”和“Bishop's Wife”之间的相似度是基于内容的推荐系统必须解决的核心问题。当项目的特征是由数值型数据描述时，可以使用类似于 Euclidean 和 Minkowski 距离的方法来度量项目之间的相似度。类别型数据之间的相似度计算没有数值型数据之间的相似度计算直接。由于类别型数据值之间的无序性，不能直接比较两个不同的类别值。对于类别型数据描述的项目，简单匹配相似度(SMS，Simple Matching Similarity)[7]仅仅使用 1 和 0 来区分相同类型值和不同类型值之间的相似度，从而不能有效获取类别型属性值之间的真正关系。耦合对象相似度同时兼顾同一特征内部耦合属性值相似性(IaAVS，Intra-coupled Attribute Value Similarity)和特征间耦合属性值相似度(IeAVS，Inter-coupled Attribute Value Similarity)，能够以较高的准确度和较低的算法复杂度获得特征值的频率分布情况和特征依赖聚合度[4]。

形式上，项目 X 和 Y 之间的耦合对象相似度定义如下：

$$COS(X,Y)=\sum_{j=1}^{l}\delta_j^{\mathcal{A}}(X_j,Y_j) \tag{1}$$

式中，X_j 和 Y_j 是项目 X 和 Y 在特征 j 上的属性值，$\delta_j^{\mathcal{A}}$ 为耦合属性值相似度(CAVS，Coupled Attribute Value Similarity)。

耦合属性值相似度(CAVS)由两部分组成：特征内耦合属性值相似度(IaAVS)和特征间耦合属性值相似度(IeAVS)。特征 j 上属性值 x 和 y 之间的耦合属性值相似度定义见式(2)：

$$\delta_j^{\mathcal{A}}(x,y)=\delta_j^{\mathcal{I}a}(x,y)*\delta_j^{\mathcal{I}e}(x,y) \tag{2}$$

式中，$\delta_j^{\mathcal{I}a}$ 和 $\delta_j^{\mathcal{I}e}$ 分别表示特征内耦合属性值相似度(IaAVS)和特征间耦合属性值相似度(IeAVS)。

特征内耦合属性值相似度(IaAVS)度量属性值相似度时考虑了同一特征下属性值出现的频率，能够有效地从频率分布的角度刻画属性值之间的相似度。特征内耦合属性值相似度(IaAVS)的定义见式(3)：

$$\delta_j^{\mathcal{I}a}(x,y)=\frac{|g_j(x)|.|g_j(y)|}{|g_j(x)|+|g_j(y)|+|g_j(x)|\cdot|g_j(y)|} \tag{3}$$

式中，$g_j(x)$ 和 $g_j(y)$ 分别表示项目集合 O 中项目的特征 j 上的属性值等于 x 与 y 的项目集合。

在计算特征内耦合属性值相似度时，仅考虑同一特征 a_j 内属性值 x 与 y 之间的相互关系，并没有涉及到特征 a_j 与其他特征 $a_k(k\neq j)$ 之间的耦合关系。特征间耦合属性值相似度(*IeAVS*)考虑了特征 a_j 内属性值 x 与 y 之间的特征依赖聚合度。计算属性值 x 与 y 之间耦合相似度时，特征间耦合属性值相似度(*IeAVS*)综合考虑了在特征 a_j 内属性值为 x 与 y 的条件下的其他特征 $a_k(k\neq j)$ 属性值分布情况。特征间耦合属性值相似度(*IeAVS*)的定义见式(4)：

$$\delta_j^{\mathcal{I}e}(x,y)=\sum_{K=1,k\neq j}^{l}\alpha_k\delta_{j|k}(x,y) \tag{4}$$

式中，α_k 是特征 $a_k(k\neq j)$ 的权重参数，$\sum_{k=1}^{l}\alpha_k=1$，$\alpha_k\in[0,1]$。$\delta_{j|k}(x,y)$ 是属性值 x 与 y 在特征 $a_k(k\neq j)$ 下的特征间耦合属性值相似度。$\delta_{j|k}(x,y)$ 的定义见式(5)：

$$\delta_{j|k}=\sum_{w\in\cap}\min\{P_{k|j}(\{w\}|x),P_{k|j}(\{w\}|y)\} \tag{5}$$

式中，$\cap$ 表示特征 a_j 取属性值 x 条件下特征 a_k 的属性值的所有取值集合与特征 a_j 取属性值 y 条件下特征 a_k 的属性值的所有取值集合的交集。$P_{k|j}(\{w\}|x)$ 和 $P_{k|j}(\{w\}|y)$ 是信息条件概率，其定义见式(6)：

$$P_{k|j}(\{w\}|x)=\frac{|g_k(w)\cap g_j(x)|}{|g_j(x)|} \tag{6}$$

$P_{k|j}(\{w\}|x)$ 描述了特征 a_j 取属性值 x 条件下，特征 a_k 取值为 w 的属性值分布特征。

以表 1 描述的项目信息为例，每个项目由特征 a_1，a_2 和 a_3 表示。项目 o_2 和 o_3 的对象耦合相似度(COS)由 $\delta_1^{\mathcal{A}}('A_2','A_2')$，$\delta_2^{\mathcal{A}}('B_1','B_2')$ 和 $\delta_3^{\mathcal{A}}('C_1','C_2')$ 组成。$\delta_1^{\mathcal{A}}('A_2','A_2')$ 等于 $\delta_1^{\mathcal{I}a}('A_2','A_2')$ 与 $\delta_1^{\mathcal{I}e}('A_2','A_2')$ 之积。由于 o_2 和 o_3 在特征 a_1 下属性值相等，$'A_2'$ 在特征 a_1 下出现 2 次，根据 $g_j(x)$ 等计算公式，$\delta_1^{\mathcal{I}a}('A_2','A_2')=0.5$。$\delta_1^{\mathcal{I}e}('A_2','A_2')$ 为 $\delta_{1|2}|('A_2','A_2')$ 和 $\delta_{1|3}|('A_2','A_2')$ 的权重和。$\delta_{1|2}|('A_2','A_2')$ 和 $\delta_{1|3}|('A_2','A_2')$ 的值都为 1，取特征权重参数 $\alpha_2=\alpha_3=1/2$。$\delta_1^{\mathcal{I}e}('A_2','A_2')=1/2*1+1/2*1=1$。所以 $\delta_1^{\mathcal{A}}('A_2','A_2')=0.5*1=0.5$。类似可计算 $\delta_2^{\mathcal{A}}('B_1','B_2')=0.125$，$\delta_3^{\mathcal{A}}('C_1','C_2')$。因此项目 o_2 和 o_3 的对象耦合相似度 $\cos(o_2,o_3)=0.5+0.125+0.125=0.75$。

表 1　项目信息表

特征项目	a_1	a_2	a_3
O_1	A_1	B_1	C_1
O_2	A_2	B_1	C_1
O_3	A_2	B_2	C_2
O_4	A_3	B_3	C_2
O_5	A_4	B_3	C_3
O_6	A_4	B_2	C_3

3.2　对象耦合相似度的项目推荐算法

本文提出的基于对象耦合相似度的项目推荐算法的框架如图 1 所示。基于对象耦合相似的项目推荐算法包含 3 个主要组成部分：特征抽取、项目耦合相似度模型构建和评分预测。

• 特征抽取：从面向特定推荐领域的源数据集中抽取用户的评分数据和项目特征信息，并将用户的评分数据转换为用户-项目矩阵 R_{n*m}，另外由项目特征信息构建项目特征向量集合 O，项目集合 O 中的每个项目由一个特征向量描述。例如，如果推荐的项目是电影，那么抽取电影的 *director*，*actor*，*genre* 等特征来描述每部电影。

• 项目耦合相似度模型构建：在项目特征向量集合 O 中，根据 COS 公式计算每对项目之间的对象耦合相似度。然后将项目对之间的相似度保存在哈希表中（如：$\langle o_1, \langle o_2, \cos(o_1, o_2)\rangle\rangle$），其中 o_1 为哈希记录的 key，$\langle o_2, \cos(o_1, o_2)\rangle$ 为对应的 value，$\cos(o_1, o_2)$ 是项目 o_1 和 o_2 之间的对象耦合相似度。通过哈希表模型，可以快速地查询 o_1 和 o_2 之间的对象耦合相似度。

• 评分预测：项目耦合相似度模型构建完成后，采用类似于 item-based 协同过滤[8] 的算法来计算用户 u 对项目 o_i 的预测评分。它将用户 u 对与 o_i 耦合对象相似度大的前 k 个项目的评分权重和作为用户 u 对目标项目 o_i 的预测评分。用户 u 对项目 o_i 的预测评分 P_{u,o_i} 由式(7)定义。

$$P_{u,o_i}=\frac{\sum\limits_{\forall N_j \in N}(\cos(o_i, N_j) * R_{u,N_j})}{\sum\limits_{\forall N_j \in N}|\cos(o_i, N_j)|} \tag{7}$$

式中，N 表示当前用户 u 评过分且与项目 o_i 耦合相似度大的前 k 个项目集合；$\cos(o_i, N_j)$ 表示项目 o_i 与项目 N_j 的对象耦合相似度；R_{u,N_j} 是用户 u 对相似项目 N_j 的评分。

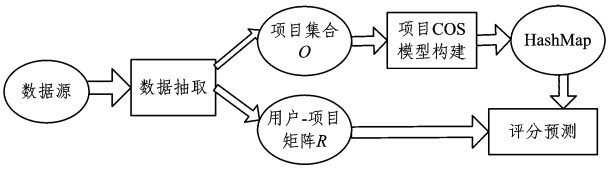

图 1　对象耦合相似度下的项目推荐算法框架

4　实验与分析

为了验证基于对象耦合相似度项目推荐算法的准确性，本文在真实数据集上进行了实验。

4.1　数据集与度量指标

本文使用 GroupLens 项目组提供的公开数据集 MovieLen100K[9] 验证对象耦合相似度下的推荐算法的性能，该数据集已被学术界广泛应用于推荐算法的评测。MovieLens100K 数据集是一个关于用户对电影评分的数据集，包含 943 位用户对 1682 部电影的 100000 条评分数据，每条评分值取 1 到 5 之间的整数值，评分值越高表示用户对电影的满意度越高。少于 20 条评分数据的用户已经从数据集中剔除。

由于基于对象耦合相似度的推荐算法利用了推荐项目本身的属性特征，因此本文从 MovieLens100K 数据集中抽取每部电影的关键特征，将每部电影表示为一个特征向量 $o=\langle$mid，director，actor，country，genre$\rangle$。另外，由于 MovieLens100K 数据集中缺少 director，actor 和 country 等特征，本文仅在电影特征向量 o 中保留电影的 genre 信息。需要特殊注意的是一部电影往往拥有多种 genre，例如电影“Toy Story”既是一个动画篇，同时也是一部喜剧。为了利用电影的多种 genre 信息，我们对电影的特征向量进行扩展：如果数据集中所有电影一共包含 t 种 genres，那么在电影特征向量 o 中增加额外的 t 种 genre 特征。换句话说，电影项目的特征向量表示为 $o=\{g_1, g_2, \cdots, g_t\}$。若电影项目具有风格 g_i（$1 \leqslant i \leqslant t$），对应的特征 g_i 的属性值为 1，否则为 0。

目前，在推荐系统研究领域有很多不同的度量指标被用来度量推荐算法的质量。例如平均绝对值误差（MAE，Mean Absolute Error）、均方根误差（RMSE，Root Mean Squared Error）和正则化的平均绝对值误差（NMAE，Normalized Mean Absolute Error）等。

由于平均绝对值误差（MAE）计算简单，可以直观地解释，本文采用平均绝对值误差（MAE）来评价推荐算法的质量。平均绝对值误差（MAE）是最常用的衡量推荐质量的度量方法。该度量方法通过比较预测值与用户实际的评分值之间的偏差来度量推荐算法的准确性。平均绝对值误差（MAE）的定义如式(8)所示。

$$MAE=\frac{\sum\limits_{i=1}^{N}|p_i-q_i|}{N} \tag{8}$$

式中，p_i 和 q_i 分别表示实际的评分值和推荐系统预测的评分值，N 表示测试数据集中的记录条数。MAE 值越小，推荐算法的推荐质量越高。

4.2　实验过程与结果分析

在实验中，将数据集 MovieLens100K 随机分成训练数据和测试数据集，其中训练数据集占 80%，测试数据集占 20%，并进行 5 折交叉验证，取 5 个不同测试数据集上运行结果的平均值作为实验的 MAE。

为了验证基于对象耦合相似度项目推荐算法的有效性，本文在相同参数设置下，对比了基于对象耦合相似度项目推荐算法（COSRec）和简单匹配相似度（SMS）下推荐算法（SMSRec）的性能，其中每个特征的权重参数 a_k（$k \neq j$）设置为 $1/(l-1)$，l 为项目特征向量的维数。

由于项目近邻数量 K 在很大程度上影响推荐算法的性能，在实验中，本文以步长 5 将项目近邻数量 K 从 5 递增到 100，观察两种推荐算法的平均绝对误差值（MAE）。实验结果如图 2 所示。

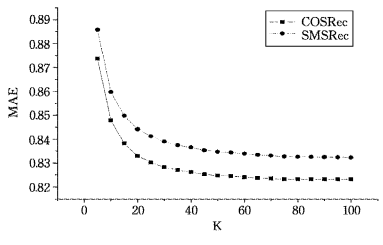

图 2　不同的项目近邻数量 K 对推荐算法性能的影响

（下转第 54 页）

[18] Chen Jie, Wang Rui-ping, Shan Shi-guang, et al. Isomap based on the image Euclidean distance[C]// IEEE International Conference on Pattern Recognition. Hong Kong, China, 2006: 1110-1113

[19] Li Jing, Lu Bao-liang. An adaptive image Euclidean distance[J]. Pattern Recognition, 2009, 42(3): 349-357

[20] 黄晓华，梁超，郑文明. 图像空间中的鉴别型局部线性嵌入方法[J]. 中国图象图形学报，2010，15(12)：1776-1782

[21] Sun Bing, Feng Ju-fu, Wang Li-wei. Learning IMED via Shift-Invariant transformation[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA, 2009: 1398-1405

[22] Gu Sui-cheng, Tan Ying, He Xin-gui. Laplacian smoothing transform for face recognition[J]. Science China Information Sciences, 2010, 53(12): 2415-2428

[23] Vapnik V. The nature of statistical learning theory [M]. New York: Springer Verlag, 1995: 1-50

[24] O'Sullivan Finbarr. Discretized Laplacian Smoothing by Fourier Method [J]. Journal of the American Statistical Association, 1991, 86(415): 634-642

[25] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]// Proceedings of Advances in MIT Press, Neural Information Processing Systems. MA, USA: 2001: 585-591

[26] He Xiao-fei, Yan Shui-cheng, Hu Yu-xiao, et al. Face recognition using laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340

[27] Ye Jie-ping, Janardan R, Qi Li. Two-dimensional linear discriminant analysis[C]// Proceedings of Advances in Neural Information Processing Systems. Vancouver, British, 2004: 1569-1576

[28] Chen Xiao-hong, Chen Song-can, Xue Hui, et al. A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data[J]. Pattern Recognition, 2012, 45(5): 2005-2018

(上接第 35 页)

从图 2 可以观察到项目近邻数量 K 对推荐算法平均绝对误差 MAE 有较大的影响。随着项目近邻数量 K 的增加，两种推荐算法的 MAE 值首先不断降低，推荐质量随 K 值增加而不断改进；$K>40$ 后，MAE 趋于平缓，推荐算法的推荐质量改进不大。在项目近邻数量 K 取任何值的情况下，基于耦合对象相似的推荐算法(COSRec)的 MAE 都低于基于简单匹配相似度的推荐算法(SMSRec)的 MAE 值，说明在相同条件下，基于耦合对象相似的推荐算法(COSRec)的推荐质量高于基于简单匹配相似度的推荐算法(SMSRec)。

然而，从两种推荐算法的 MAE 值变化曲线可以看出，虽然基于耦合对象相似度的推荐算法(COSRec)推荐质量在不同 K 值条件都高于基于简单匹配相似度的推荐算法(SMSRec)，但是两者的差距不大，保持在 0.01～0.02 之间。这主要是因为从 MovieLens100K 数据集中仅能抽取电影的 genre 特征，而且其它如 director、actor 等重要特征的缺失在很大程度上限制了基于耦合对象相似度推荐算法的性能。我们认为在能抽取更多项目特征的情形下，基于耦合对象相似度的推荐算法的推荐质量比传统基于内容的推荐算法有较大的改进。

结束语 推荐系统在为用户解决信息过载问题方面发挥着越来越重要的作用，它可以为其推荐符合其偏好的商品、新闻、电影和音乐等，甚至在社交网络系统中可以推荐与用户有共同爱好的好友。特别是在电子商务系统中，推荐系统根据用户的信息资料和推荐项目的文本信息，为用户推荐个性化的产品和服务信息，在提高用户体验的同时，也增加了电子商务企业的效益。针对传统基于内容的推荐算法的项目之间缺少合适度量手段计算项目之间的相似度的问题，本文提出了一种基于耦合对象相似度的项目推荐算法。该推荐算法由项目特征抽取、项目耦合相似度模型构建和评分预测等 3 个主要成分组成，其中的核心在于利用耦合对象相似度(COS)度量推荐项目之间的相似度，通过耦合对象相似度可以有效捕获特征值的频率分布情况和特征依赖聚合度。在现实数据集上的实验验证了在缺失项目特征的条件下，基于耦合对象相似度的项目推荐算法比传统的基于内容的推荐算法在推荐精度上有一定程度的改进。

由于目前公开的推荐算法评测数据集普遍包含用户的评分数据，且项目本身的特征信息较少，如 MovieLens100K 数据集中仅电影的风格特征可以用来计算电影的相似度，而且其他如导演、演员等特征信息缺失，本文下一步的研究将抽取更多的项目特征信息来改进基于耦合对象相似度的推荐算法。另外，冷启动问题是推荐系统的研究热点，基于内容的推荐算法在解决冷启动问题上较协同过滤的推荐算法有本质上的优势，本文下一步将研究基于耦合对象相似度推荐算法在解决推荐系统冷启动问题上的性能表现。

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] 赵亮，胡乃静，张守志. 个性化推荐算法设计[J]. 计算机研究与发展，2002，39(8)：986-991

[3] Balabanovic M, Shoham Y. Fab: content-based, collaborative recommendation[J]. Communications of the ACM, 1997, 40(3): 66-72

[4] Melville P, Mooney R, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C]// Proceedings of the National Conference on Artificial Intelligence. AAAI Press, MIT Press, 1999(2002): 187-192

[5] Wang C, Cao L, Wang M, et al. Coupled nominal similarity in unsupervised learning[C]// CIKM, ACM. 2011: 973-978

[6] Yu Y, Wang C, Gao Y, et al. A Coupled Clustering Approach for Items Recommendation[C]// The 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Gold Coast, Australia, 2013

[7] Boriah S, Chandola V, Kumar V. Similarity measures for categorical data: a comparative evaluation[C]// SDM 2008. 2008: 243-254

[8] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]// WWW. ACM, 2001: 285-295

[9] http://www.grouplens.org/node/73