文章编号: 1009 - 2552(2015)03 - 0070 - 04 DOI: 10.13274/j.cnki.hdzj.2015.03.018

基于项目属性和 BP 神经网络的协同过滤推荐

陈丹儿，应玉龙

(浙江纺织服装学院，浙江 宁波 315211)

摘　要: 针对协同过滤推荐中存在的数据稀疏性问题，文中提出了一种基于项目特征属性和 BP 神经网络相结合的协同过滤推荐算法，并通过 Movielens 数据集验证了该模型的有效性。此方法首先将用户评分数据映射为用户对项目特征属性的偏好，然后使用 BP 神经网络训练得到目标用户的特征属性偏好模型并对新项目的评分进行预测，从而降低用户项目评分矩阵的数据稀疏性，最后使用协同过滤推荐算法，形成最近邻并生成推荐建议。

关键词: 协同过滤; 稀疏性; 特征属性; BP 神经网络; 用户偏好模型

中图分类号: TP183　　**文献标识码**: A

Collaborative filtering method based on item's characteristics and BP neural network

CHEN Dan-er, YING Yu-long

(Zhejiang Fashion Institute of Technology, Ningbo 315211, Zhejiang Province, China)

Abstract: For the solution of the problem that data sparseness exists in the collaborative filtering recommendation, this paper proposed a collaborative filtering recommendation algorithm based on item's characteristics and BP neural network, and verified the validity of the model by Movielens datasets. This method maps the user's rating data into the user's features property preference, and then establishes the target user's preference model of characteristics used by the BP neural network training, which can predict the program's rating and reduce the data sparsity of user's item rating matrix. In conclusion, it provides some suggestion that compute nearest-neighbors and produces the recommendations.

Key words: collaborative filtering; sparsity; characteristics; BP neural network; user's preference model

0　引言

随着电子商务的发展和大数据时代的到来，如何有效利用用户的购买记录和评分数据等相关信息来向用户推荐其可能感兴趣的商品，是目前解决信息过载严重、提高电商个性化服务水平的有效途径。协同过滤推荐技术是最成功的推荐技术之一[1-3]，著名的电子商务网站，如亚马逊和 CD - NOW 等，在向客户推荐产品时都应用了协同过滤技术，它改善了服务的质量和效率。

协同过滤是基于这样一个假设，找到一个特定用户感兴趣的内容是在找到与他志趣相投的人的结果之上。该方法通过分析用户的历史数据来计算用户之间的相似度，生成与目标用户行为兴趣最相近的最近邻集合，根据最近邻对产品的评分预测目标用户对这些产品的评分，将评分最高的前 N 项产品推荐给目标用户[4-5]。从协同过滤推荐的流程来看，用户评分数据对相似性计算起决定性作用，但在大型电子商务系统中，用户评分过的项目一般不会超过系统项目总数的 1%[6]，造成用户 - 项目评分矩阵的稀疏性问题，从而影响了推荐系统的推荐质量和性能。

很多时候单纯通过用户对项目的评分来判断用

收稿日期: 2014 - 04 - 17
基金项目: 宁波市自然基金资助项目(2012A610070)
作者简介: 陈丹儿(1978 -)，女，副教授，硕士研究生，研究方向为数据挖掘、信息推荐。

户间是否存在共同兴趣具有一定的片面性[7]。例如用户 A、B 虽然对项目 I 的评分相同，但两者评价的侧重点不同，A 可能偏重项目 I 的功能属性，而 B 可能偏重项目 I 的外观属性。基于此，本文提出了一种基于项目特征属性和 BP 神经网络相结合的协同过滤推荐算法，首先将用户 - 项目评分映射为用户对项目的特征属性偏好，然后采用 BP 神经网络模型来预测用户对项目的评分，从而减少用户 - 项目矩阵的数据稀疏性，提高推荐的质量。

1 用户偏好模型

用户偏好模型反映了用户对项目特征属性的偏好程度，可以从项目特征属性矩阵和用户偏好两个方面来构建用户偏好模型。

1.1 项目特征属性矩阵

每个项目均有各自的属性特征，如服装有款式、风格、领子、袖子等多个属性，用户对项目的喜好往往通过项目特征属性表现出来，如用户 A 偏好运动型、百搭、圆领的短袖。所有项目的特征属性矩阵可表示如下:

$$\boldsymbol{P}_{m\times n}=\begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & p_{ik} & \vdots \\ p_{m1} & \cdots & p_{mt} \end{bmatrix}$$

p_{ik} 表示项目 i 的特征属性，其中:

$$p_{ik}=\begin{cases} 0, \text{项目 } i \text{ 不具有特征 } h_k \\ 1, \text{项目 } i \text{ 具有特征 } h_k \end{cases}$$

n 表示项目特征属性的个数，一般比较固定，m 表示项目总数。当新增项目时，可以向 $P_{m\times n}$ 中添加一条记录。

1.2 用户特征属性偏好模型

用户 u 对项目特征属性的偏好可以用矩阵 $\boldsymbol{H}_u=[h_1,h_2,\cdots,h_t]$ 表示，h_i 的取值范围为 $[0,1]$，1 为最喜欢，0 为最不喜欢。用户 u 对第 i 个特征属性的偏好 h_{ui} 定义为:

$$h_{ui}=\frac{\sum_{j=1}^{n} I_j}{\sum_{j=1}^{m} I_j}$$

其中，m 表示用户评价过的所有项目数，n 表示具有第 i 个特征向量的用户评价过的项目数($n\leqslant m$)，I_j 表示用户 u 对项目的评分。因此用户 u 对所有项目的特征属性偏好模型 y_u 可定义为:

$$y_u=P\times H_u^{T}$$

上式公式表明项目所具有的特征属性决定了用户对其的喜好程度。上述函数所描述的用户特征属性偏好模型，其函数关系是很复杂的，随着项目和用户评分数据的不断增加，如何不断地修正用户偏好模型，以便更准确地得到用户对项目的评分，本文采用 BP 神经网络工具来模拟用户偏好模型。

2 BP 神经网络模型

2.1 BP 神经网络

BP 神经网络是一种多层前馈型神经网络，包括输入层、隐含层和输出层，其中隐含层可以有多个，一般采用三层网络结构。BP 神经网络的主要思想是通过对输入学习样本的正向传播，得到期望输出与实际输出的误差，然后使用反向传播算法根据误差的大小对网络的权值和阈值进行调整，经过反复的调整和训练使得实际输出越来越逼近期望值，当误差达到可容忍范围内时，停止训练，并保存此时的网络权值和阈值[8-9]。

BP 算法是一种梯度下降算法，其中的网络权值沿均方误差函数的负梯度转移[10-11]。输入向量和相应的目标向量用于训练网络，直至达到满意的误差。经过训练的 BP 神经网络有能力推广，也就是说，一旦经过训练处理，该系统能够处理以前看不见的样本数据，并产生一个可以接受的结果。

假设 $yl(t)$ 是神经网络的预期输出，用 MSE 代表实际输出与期望输出的误差，可以通过如下公式进行计算:

$$E=\frac{1}{2}\sum_{t=1}^{k}\sum_{l=1}^{n}\left(y_l(t)-y_l^{*}(t)\right)$$

2.2 用户特征属性偏好模型训练

本文使用三层 BP 神经网络来训练用户特征属性偏好模型，如图 1 所示。

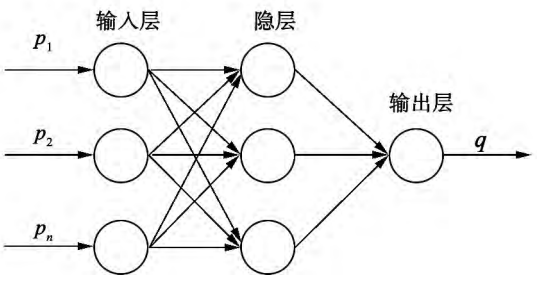

图 1 用户特征属性偏好模型 BP 神经网络结构

图 1 中，BP 神经网络的输入因子为产品的特征向量，输出因子为用户对产品的评分。对于待推荐产品，在提取其产品特征后，将产品特征向量作为输入，经过训练好的用户偏好模型，可以得到用户对该产品的评分。

由于 BP 神经网络的输入变量为产品的特征向量，隐层神经元表征为用户对特征属性的偏好程度，

因此输入层神经元和隐层神经元的相同，都等于产品特征向量维数。

3 产生推荐

通过使用 BP 神经网络技术可以填补用户项目评分矩阵中空缺的数值，从而降低矩阵的稀疏度，然后应用基于用户的协同过滤推荐算法，产生一个目标用户的预测。

3.1 用户评分相似性度量

目前，已经有很多个相似性测量方法应用在协同过滤推荐系统中，如 Pearson 相关相似性度量、余弦相似性度量、调整的余弦相似性度量等[12-13]，本文使用余弦相似性度量方法来计算用户之间的相似性。

余弦相似性度量方法，把两个评分向量看作夹角来进行相似性测量，公式如下:

$$sim(i,j) = \frac{\sum_{k=1}^{n} R_{ik}R_{jk}}{\sqrt{\sum_{k=1}^{n} R_{ik}^{2} \sum_{k=1}^{n} R_{jk}^{2}}}$$

其中，$R_{i,k}$ 是用户 i 对项目 k 的评分数值; n 是用户之间共同评分的项目数量。

3.2 选取最近邻居

产生推荐选择给定目标用户的邻居将作为推荐人。本文采用 Top N 技术，设置阈值来选取最近邻居的数目。

在获得目标用户的最近邻居集合之后，可以计算最近邻居集合中用户的加权平均来对目标用户进行未评分项目预测。

目标用户 u 对目标项目 t 用如下算式进行预测:

$$P_{ut} = A_u + \frac{\sum_{i=1}^{c} (R_{it} - A_i) \times sim(u,i)}{\sum_{i=1}^{c} sim(u,i)}$$

其中，A_u 是目标用户 u 对所有评分项目的平均分; R_{it} 是用户 i 对项目 t 的评分数值; A_m 是邻居用户 m 对所有评分项目的平均分; $sim(u,i)$ 是目标用户 u 和邻居用户 i 之间的相似性数值; c 是目标用户最近邻居的个数。

基于 BP 神经网络的协同过滤算法描述如下:

(1) 选择将要评分的空项 I。

(2) 设置项 I 的特征属性值。

(3) 设定输入层、输出层和隐层，并初始化。

(4) 使用 BP 神经网络预测空项值。

(5) 判定学习精度是否收敛到最小值: 是，转步骤(7)，否，转步骤(6)。

(6) 判定迭代步数是否超过规定的步数: 是，转步骤(7); 否，转步骤(4)。

(7) 按照余弦相似性度量方法求 Ui 的 Top - N 个最近邻居集。

4 实验结果及分析

4.1 数据集

数据集合使用 MovieLens，该数据集是由 GroupLens 研究项目组在明尼苏达大学开发出来[9-10]。历史数据包括 943 个用户对 1682 部电影的 100,000 评分数目。其中，每个用户至少对 20 部电影进行评分。MovieLens 数据集的评分等级是五级制。其中，数值 1 和 2 表示用户对电影是负评价，数值 4 和 5 表示用户对电影是推荐评价，数值 3 表示中性评价。

Movielens 数据库中电影的特征属性有 18 个，因此神经网络输入层神经元个数设定为 18 个，设定一个隐层并且隐层神经元个数设定为 18 个，输出层神经元设为 1 个。

4.2 度量方法

协同过滤推荐系统评价标准可以分为三大类: 预测精确度度量、分类精确度度量和等级精确度度量[14-15]。本文使用预测精确度度量中的平均绝对偏差 MAE(mean absolute error) 和分类精确度度量中的接收器操作特性 ROC(receiver operation characteristic) 敏感度对算法进行评价。

平均绝对误差 MAE(mean absolute error) 是目前使用最广泛的评价指标，MAE 通过计算预测的用户评分与实际用户的评分偏差评价预测的准确性。MAE 越小，推荐质量越高。

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |p_i - q_i|}{n}$$

所谓 ROC 敏感度是指随机选择的“好”的项在推荐系统列表中的比例。ROC 敏感度值范围是 0 ~ 1，值越大表明推荐系统性能越好。

$$\mathrm{ROC} = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} v_i}$$

$$u_i = \begin{cases} 1, & p_i \geqslant 4 \text{ and } q_i \geqslant 4 \\ 0, & \text{otherwise} \end{cases}$$

$$v_i = \begin{cases} 1, & p_i \geqslant 4 \\ 0, & \text{otherwise} \end{cases}$$

以上公式中，p_i 表示预测的用户评分，q_i 表示对应的用户实际评分。

4.3 实验结果

把 MovieLens 按照 5/1 的比例划分为训练集/测试集，使用 MAE 和 ROC 两个度量标准，将本文所提出的 BP – Based CF 算法与传统的基于余弦法、基于修正余弦法寻找最近邻的协同过滤算法作比较，当邻居个数从 5 个增至 40 个时，分别得到图 2 – 3 所示的比较图。

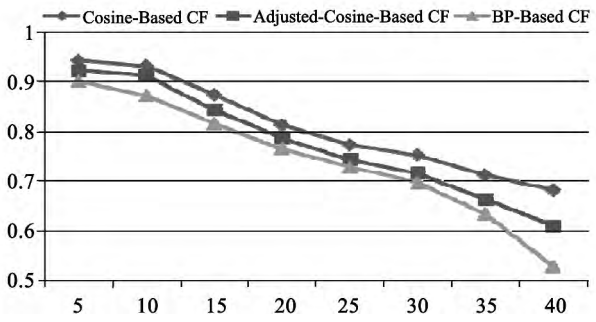

图 2 MAE 比较图

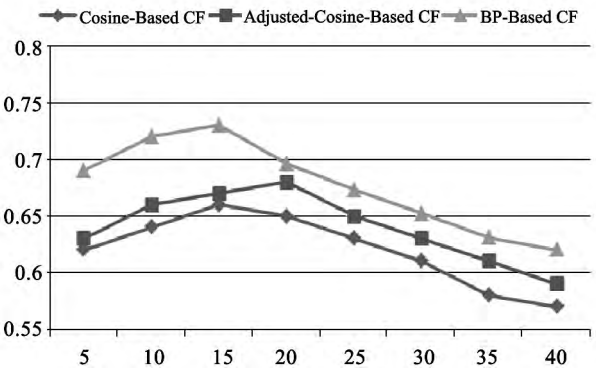

图 3 ROC 比较图

从图 2 – 3 可以看出，BP – Based CF 算法均有较好的性能。

5 结束语

协同过滤是推荐系统中最成功的技术之一，并广泛应用于许多个性化推荐领域，如电子商务，数字图书馆等。然而，大多数协同过滤算法存在数据集稀疏性问题，从而导致不准确的推荐。本文主要针对协同过滤算法中数据的极端稀疏性对推荐质量的影响，提出了一种基于项目特征属性和 BP 神经网络相结合的协同过滤推荐算法，该算法基于产品特征建立了用户偏好模型，采用神经网络训练得到具体的用户偏好模型并进行项目评分预测，从而降低用户评分矩阵的稀疏度。实验结果表明，本文提出的算法可以产生比传统方法更准确的推荐。

参 考 文 献:

[1] Hyung Jun Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem [J]. Information Sciences, 2008, 178: 37 – 51.

[2] Sarwar B, Karypis G, Konstan J, et al. Item-Based collaborative filtering recommendation algorithms [C] // Proceedings of the 10th International World Wide Web Conference, 2001: 285 – 295.

[3] Gong Song-jie, Shi Xiao-yan. A Collaborative Recommender Combining Item Rating Similarity and Item Attribute Similarity [M]. ISBIM2008, IEEE Computer Society Press.

[4] 辛菊琴，蒋艳，舒少龙. 综合用户偏好模型和 BP 神经网络的个性化推荐 [J]. 计算机工程与应用，2013，49(2)：57 – 60.

[5] Shih Y. -Y, Liu D. -R, Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands, Expert Systems with Applications, 2007.

[6] 余小鹏. 一种基于多层关联规则的推荐算法研究 [J]. 计算机应用，2007，27(6)：1392 – 1393.

[7] 李聪，梁昌勇. 基于属性值偏好矩阵的协同过滤推荐算法 [J]. 情报学报，2008，27(6)：884 – 890.

[8] 张月琴，刘翔，孙先洋. 一种改进的 BP 神经网络算法与应用 [J]. 计算机技术与发展，2012，22(8)：163 – 164.

[9] 罗斐. 基于 BP 算法的小波神经网络的研究 [J]. 信息技术，2012(11)：48 – 51.

[10] Gong Song-jie, YE Hong-wu. An Item Based Collaborative Filtering Using BP Neural Networks Prediction [M]. IIS 2009, IEEE CS Press.

[11] Zhang Feng, Chang Hui-you. Employing BP Neural Networks to Alleviate the Sparsity Issue in Collaborative Filtering Recommendation Algorithms [J]. Journal of Computer Research and Development, 2006, 43(4): 667 – 672.

[12] HE Fang-guo, Qi Huan. Back propagation neural network based on modified genetic algorithm and its application [J]. Journal of huazhong normal university (Nat. Sci.), Mar. 2007, 41(1): 51 – 54.

[13] 吴颜，等. 协同过滤推荐系统中数据稀疏性问题的解决 [J]. 计算机应用研究，2007，24(6)：94 – 97.

[14] 张锋，常会友，等. 使用 BP 神经网络缓解协同过滤推荐算法的稀疏性问题 [J]. 计算机研究与发展，2006，43(4)：667 – 672.

[15] Gong Song-jie, A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering, Journal of Software, Jul. 2010, 5(7): 745 – 752.

责任编辑：肖滨