第 29 卷第 1 期　　四川理工学院学报(自然科学版)　　Vol. 29　No. 1
2016 年 2 月　　Journal of Sichuan University of Science & Engineering(Natural Science Edition)　　Feb. 2016
文章编号:1673-1549(2016)01-0039-05　　DOI:10.11863/j.suse.2016.01.09

基于改进 BP 神经网络的个性化推荐算法研究

罗频捷[a], 温　荷[b]

(成都东软学院 a. 实验管理中心;b. 计算机科学与技术系, 成都　610000)

摘　要:个性化推荐是根据用户的喜好向用户推荐个性化的产品、信息或服务等,在大数据时代能有效地解决信息过载的问题。提出基于改进 BP 神经网络的个性化推荐算法,研究了如何建立混合推荐模型,采用 BP 神经网络训练提高算法中用户相似度计算的准确性。在真实数据集上的实验结果表明,该算法能有效减小用户相识度计算偏差,提高用户个性化推荐的推荐质量。

关键词:个性化推荐算法;改进 BP 神经网络;混合推荐系统;用户相似度;动量因子

中图分类号:TP391　　**文献标志码**:A

引　言

随着移动互联网技术的快速发展与成熟,现代电子商务呈现出爆炸式增长,以淘宝双十一购物节为例,2012 年销售为 191 亿,2013 年销售增长到 350 亿,2014 年销售总额高达 673 亿[1]。如此繁荣的电子商务产业在给用户带来更多选择机会的同时,也在增加用户选择负担,普通用户很难从大量商品信息中找出适合的商品。从另一个侧面看,在电子商务平台中还有海量商品信息无法被普通用户获知[2]。因此个性化的推荐系统在未来电子商务发展中将发挥越来越重要的作用。

通常推荐系统被用于两个方面目的:一是激发用户做某些事情;二是解决用户信息过载。个性化的推荐系统主要通过信息检索和信息过滤,从庞大集合里选择出最能满足用户需求的物品。常见的推荐算法有协同过滤推荐算法、基于内容的推荐算法和基于知识的推荐算法等,它们在不同的应用领域表现出的效果各有不同。协调过滤推荐算法主要思想是利用已有用户群过去的行为预测当前用户的选择,即通过目标用户群体的特定信息产生推荐结果[3];基于内容的推荐算法则依赖被推荐产品的特征进行推荐[4];而基于知识的推荐算法则是根据已知的知识模型推演出推荐结果,这些推荐算法都各有利弊。一个良好的推荐算法是可以将用户模型、产品特征和知识模型高效地输出产品推荐序列表。因此,通过将多个推荐算法组合起来构成一个推荐单元的混合推荐系统,在个性化推荐中更能发挥作用。

1　个性化推荐系统模型

1.1　个性化混合推荐系统模型

本文将个性化推荐系统模型建立在混合推荐系统基础上,将几种推荐算法整合到一个实现算法中,通过对算法进行内部调整,让多个推荐算法都发挥作用[5-6],如图 1 所示。

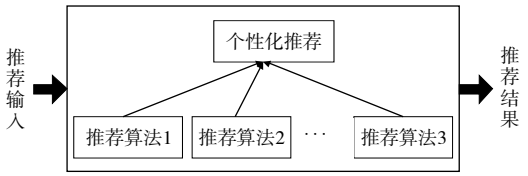

图 1　个性化混合推荐系统模型

个性化混合推荐系统模型将多种推荐算法进行内

收稿日期:2015-12-14
基金项目:四川省教育厅科研项目(14ZB0346);成都东软学院校级科研项目(NEU2015-1)
作者简介:罗频捷(1981-),男,四川泸州人,高级工程师,硕士,主要从事神经网络、机器视觉方面的研究,(E-mail)luopinjie@nsu.edu.cn;
温　荷(1982-),女,四川成都人,副教授,主要从事云计算、大数据方面的研究,(E-mail)wenhe@nsu.edu.cn

部调整后整合在一个推荐单元,通过对知识源进行预处理和组合多个知识源,利用不同类型的输入,扩大整个推荐系统的特征空间,实现推荐系统的高效性[7-8]。

1.2 个性化推荐系统算法

个性化推荐算法中用户 a 对产品 i 个性化推荐结果 $rec(a,i)$ 采用 Linkert 量值表示,影响 $rec(a,i)$ 值的因素有四个方面,分别是:用户之间相似度 $P_{a,u}$,用户评分个数 n_i,用户之间评分重叠个数 $n_{a,u}$,以及用户评分向量 $v_{u,i}$。$rec(a,i)$ 由协同推荐算法改进而来,表示为:

$$rec(a,i) = (sw_a c_{a,i} + \sum_{u=1u\neq a}^{n} hw_{a,u} P_{a,u} v_{u,i}) \\ (sw_a + \sum_{u=1u\neq a}^{n} hw_{a,u} P_{a,u}) \tag{1}$$

式中,$hw_{a,u}$ 为混合权重因子,是由权重因子 $sg_{a,u}$ 和 $hm_{a,u}$ 组成,这两个权重因子取自于基于内容的推荐算法,均倾向于具有更多的共同评分的产品。$hw_{a,u}$、$sg_{a,u}$ 和 $hm_{a,u}$ 分别表示为式:

$$hw_{a,u} = sg_{a,u} + hm_{a,u} \tag{2}$$

$$sg_{a,u} = \begin{cases} \frac{n_{a,u}}{50} & n_{a,u} < 50 \\ 1 & n_{a,u} \geqslant 50 \end{cases} \tag{3}$$

$$hm_{a,u} = \frac{2m_a m_u}{m_a + m_u} \tag{4}$$

$$m_i = \begin{cases} \frac{n_i}{50} & n_i < 50 \\ 1 & n_i \geqslant 50 \end{cases} \tag{5}$$

由用户 i 的原始评分值的个数决定,反映了算法基于内容的预测可信度。sw_i 计算公式为:

$$sw_i = \begin{cases} \frac{n_i}{50} \times \max & n_i < 50 \\ \max & n_i \geqslant 50 \end{cases} \tag{6}$$

式中的 max 为常量,一般设置为 2。

式(1)中的用户评分向量 $v_{u,i}$ 值表示为:

$$v_{u,i} = \begin{cases} s_{u,i} & 若\ s_{u,i}\ 存在 \\ c_{u,i} & 若\ s_{u,i}\ 不存在 \end{cases} \tag{7}$$

式中,若用户 u 已给产品 i 评过分数,则将评分结果 $s_{u,i}$ 作为用户评分向量 $v_{u,i}$;若用户 u 还未曾给产品 i 评过分数,则采用基于内容的推荐算法计算结果 $c_{u,i}$ 作为用户评分向量 $v_{u,i}$。

2 BP 神经网络模型改进

在式(1)中,除了用户相似度 $P_{a,u}$ 外,其余各个参数均可由实际统计数据得出。在推荐系统中,$P_{a,u}$ 常常可以使用修正余弦相似性度量方法求出[9]。但是修正余弦相似性度量方法成立的前提条件是默认所有用户对产品的评判标准是一致的,这在实际应用中明显不合理,所以采用修正余弦相似性度量方法的计算结果与真实的 $P_{a,u}$ 有时偏差较大,继而影响最终推荐结果 $rec(a,i)$。因此,本文提出采用改进 BP 神经网络算法进行用户相似度 $P_{a,u}$ 测算,从而提高 $P_{a,u}$ 的准确性和可靠性。

2.1 BP 神经网络模型

BP 神经网络因其具备完善的理论体系和学习算法,是目前人工神经网络中使用最为广泛的神经网络模型[10]。BP 神经网络是一种前馈型神经网络,模仿人脑神经元对外部信号的反馈机理,建立多层次的感知模型,并采用信号的正向传播和误差的反向调节相结合的方式进行自主学习。BP 神经网络中神经元节点使用 S 型传递函数,能够输出从 0 到 1 之间的连续信息,从而可以实现从输入层到输出层之间的任意非线性映射。

BP 神经网络是一种三层神经网络,分别是网络输入层、网络隐含层和网络输出层,每层是由多个并行运算的神经元构成。BP 神经网络中同一层次的神经元之间不进行连接,而是将相邻两层的全部神经元相互连接,神经元之间连接权值决定了 BP 神经网络的信息处理能力。BP 神经网络拓扑结构如图 2 所示。

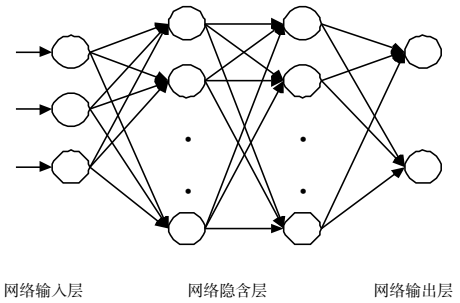

图 2 BP 神经网络拓扑

BP 神经网络算法主要分为信号正向传播和误差反向传播两个阶段。首先,输入信息从输入层输入,经过隐含层后再到输出层输出,在整个传输过程中神经元的连接权值固定不变。一般而言这样的输出结果与期望输出结果之间存在误差,此时,将误差由输出层开始逐层向前传输,在传输过程中根据误差情况动态调整各个神经元之间连接权值,直到得到实际输出与期望输出之间的最小均方差值时为止[11]。

2.2 改进 BP 神经网络模型

由于 BP 神经网络使用神经元之间的连接权值来存储信息,所以 BP 神经网络具有较高的容错性和鲁棒性。并且,BP 神经网络可以通过反馈学习不断加强神经元之间的连接强度,进而使得 BP 神经网络对信息具有较

高的灵敏度。BP 神经网络虽然优点众多,但它在实际应用中仍存在不足。BP 神经网络非线性优化存在局部极小点问题,从而对网络性能带来消极影响[12]。同时,BP 神经网络初始值的设置对算法收敛速度影响较大,BP 神经网络修正权系数 $w(k)$ 通常不考虑 k 时刻之前的梯度方向,只从 k 时刻误差的梯度下降方向进行权值调整,导致 BP 神经网络训练过程中发生震荡,致使算法收敛速度缓慢[13-14]。

为解决 BP 神经网络上述问题,引入动量因子改进 BP 神经网络算法。BP 神经网络增加动量因子后,可以使得权值的调整朝着误差曲面底部平方的方向变化,并将最后一次权值变化的影响通过动量因子进行传递。引入动量因子的权值调节表达式为:

$$\Delta\omega(k+1) = \Delta\omega(k) + \eta(1-\alpha)E(k) + \alpha E(k-1) \quad (8)$$

式中:k 为训练次数,α 为动量因子,η 为学习速率,$E(k)$ 为关于权值向量 $\omega(k)$ 的梯度。当 BP 神经网络权值进入误差曲面底部的平坦区域时,$E(k)$ 将变得很小,即有:

$$\Delta\omega(k+1) \approx \Delta\omega(k) \quad (9)$$

由式(9)可知,引入动量因子后,可以防止 $\omega(k) = 0$ 的情况出现,这将使得 BP 神经网络可从误差曲面局部极小值中跳出。

2.3　改进 BP 神经网络模型训练

在个性化推荐过程中,具有相同特征的用户一般拥有相似的行为习惯,用户之间的行为习惯越相似,用户相似度越高,将 a 用户与其他用户之间相似度记为 $P_{a,u}$。在计算 $P_{a,u}$ 时,通常进行归一化处理,把 $P_{a,u}$ 取值范围标准化到[0,1]集合上。本文采用改进 BP 神经网络进行用户相似度 $P_{a,u}$ 测算。

为了便于 BP 神经网络模型的初始权值设定,本文采用 3 层神经网络:第 1 层为输入层,第 2 层为隐含层,第 3 层为输出层。输入层主要输入商品属性信息;隐含层则为用户特征值;输出层为用户相似度 $P_{a,u}$。

改进 BP 神经网络的用户相似度模型算法的训练步骤如下:

(1)设定输入层和输出层节点的数量,并初始化各层权值和阈值。

(2)将商品主要属性信息输入到 BP 网络中。

(3)按照 BP 模型算法进行网络学习。

(4)计算权值梯度和训练误差。

(5)根据式 8 修正各层权值与阈值。

(6)判断样本集的均方误差是否满足要求。满足,则结束学习,否则继续学习,直至满足为止。

(7)使用以上训练好的网络对 $P_{a,u}$ 进行测算。

3　实验与分析

为了验证基于 BP 神经网络的个性化算法的准确性和有效性,本文在真实数据集上开展实验仿真,并进行分析。

3.1　实验数据集

本文实验数据采用美国明尼苏达大学 GroupLeans 项目提供开源的 Movieleans 数据集进行仿真实验[15-16]。Movieleans 数据集创建于 1997 年,是一个基于 Web 的推荐算法研究数据集。目前,Movieleans 数据集包含 72 000名用户对 1 万部电影的 1000 万条评分数据,电影的每条评分数据取值[1,5]之间的整数值,评分数值越高代表用户对该电影的认可度越高,即表示用户越喜欢该电影。本文从 Movieleans 数据集中随机的选择了 10 000条评分数据作为实验仿真数据集,该数据集共包含了 196 个用户对 1029 部电影的评分。实验仿真数据集被划分为训练数据集和测试数据集两部分,训练数据集作为输入用于算法训练,而测试数据集则用于算法的性能测试。

3.2　度量标准

评价推荐系统的推荐质量的度量标准可分为三大类,分别是预测精度度量、分类精度度量和等级精度度量[17]。本文采用预测精度度量中常用的平均绝对偏差 MAE(Mean Absolute Error)作为度量标准。MAE 通过计算预测用户评分与实际用户评分之间的偏差度,可以非常直观地对推荐质量进行度量,便于对推荐系统结果的分析与评价[18-20]。MAE 评价度量标准公式为:

$$MAE = \frac{\sum_{i=1}^{n} S_i - T_i}{n} \quad (10)$$

式中,S 表示用户评分的集合 $\{s_1, s_2, \cdots, s_n\}$,$T$ 表示实际用户的评分集合 $\{t_1, t_2, \cdots, t_n\}$,$n$ 表示评测用户数量。计算结果的 MAE 值越小,表示推荐系统的推荐质量越高。

3.3　结果与分析

本文在仿真实验中使用 1 台惠普刀片服务器作为实验硬件环境,配置为:2 个 Intel 至强 E7 处理器(主频为 2.26 GHz),32 G 的 DDR3 内存,1 T 的 SATA 硬盘。仿真实验软件环境采用神经网络工具箱,主要使用了 newff()、train()和 sim()函数,分别用于 BP 神经网络生成、BP 神经网络训练和 BP 神经网络验证。由于 Movieleans 数据集中电影的特征由 18 个属性组成,所以在仿

真实验时设定输入神经元个数为 18 个,隐含层神经元也设置为 18 个。

改进后的 BP 神经网络经过训练可得到该网络模型,该模型的训练方差、测试方差和测试误差均值如图 3 所示。

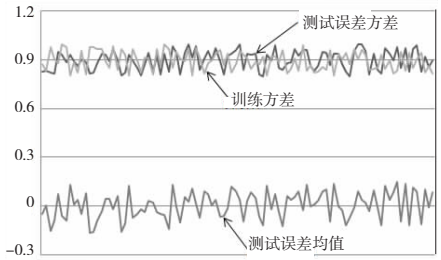

图 3 改进后 BP 神经网络有效性验证

由图 3 可见,训练后的网络模式的训练方差和测试误差方差大致在一条直线上,而且网络的测试误差均值也保持在 0 值附近,这说明改进后的 BP 神经网络模型具有较好的稳定性和有效性。

本次实验对改进 BP 神经网络相似性和修正余弦相似性的计算精度进行比较实验,计算其各自 MAE,最近邻居集合的个数 S 从 5 递增至 30,间隔为 5。实验结果如图 4 所示。

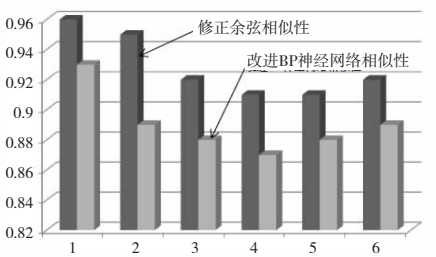

图 4 相似性度量方法的比较

由图 4 可见,改进 BP 神经网络相似性度量方法的 MAE 值均低于传统的修正余弦相似性度量值,因此改进 BP 神经网络相似性度量方法所计算的相似度 $P_{a,u}$ 较传统的修正余弦相似性度计算的相似度 $P_{a,u}$ 与真实 $P_{a,u}$ 较偏差要小。

为验证本文提出的算法的有效性,本文将基于改进 BP 神经网络的个性化推荐算法与混合推荐算法和常用的协调过滤推荐算法进行比较实验,实验采用 MAE 度量标准,实验结果如图 5 所示。从图 5 可以看出,本文提出的算法与其他算法相比均具有最小的 MAE 值,由此可知,本文提出的基于改进 BP 神经网络的个性化推荐算法可以显著地提高推荐系统的推荐质量。

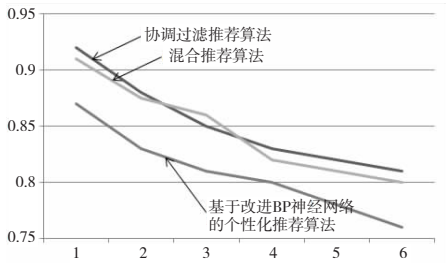

图 5 基于改进 BP 神经网络的个性化推荐算法效果

4 结束语

随着互联网技术的高速发展与应用,最近几年推荐系统几乎成为大型互联网平台的标准配置和功能。这反映了在当今信息过载时代,如何根据用户的品味和爱好自适应地发现满足用户个性化的需求是一个重要的研究领域。本文提出的基于 BP 神经网络的个性化推荐算法,将协同推荐算法、基于内容的推荐算法整合到一个混合推荐算法中,并采用改进 BP 神经网络算法解决了该算法中存在的用户相识度计算准确度偏低的问题。通过在真实数据集上的实验表明,本文提出的算法减小了用户相识度计算偏差,显著地提高了用户个性化推荐的推荐质量。

参 考 文 献:

[1] 许海玲,吴潇,李晓东,等.互联网推荐系统比较研究[J].软件学报,2009,20(2):350-362.

[2] 项亮.推荐系统实践[M].北京:人民邮电出版社,2013:5-9.

[3] 赵琴琴,鲁凯,王斌.SPCF:一种基于内存的传播式协同过滤推荐算法[J].计算机学报,2013,36(3):671-676.

[4] ADOMAVICIUS G,TNZHILIN A.Context-aware recommender systems[M]//FRANCESCO R, LIOR R, BRACHA S, et al. Recommender Systems Handbook, US:Springer,2011:217-253.

[5] 龙舜,蔡跳,林佳雄.一个基于演化关联规则挖掘的个性化推荐模型[J].暨南大学学报:自然科学与医学版,2012,33(3):264-267.

[6] GHAZANFAR M A,PRUGEL-BENNETT A.Building switching hybrid recommender system using machine learning classifiers and collaborative filtering[J].International Journal of Computer Science,2010,37(3):272-287.

[7] SHI Y,KARATZOGLOU A,BALTRUNAS L,et al. CLIMF:learning to maximize reciprocal rank with col-

laborative less-is-more filtering[C]//Proceedings of the sixth ACM conference on Recommender systems,Dublin,Ireland,September 9-13,2012:139-146.

[8] 谢娟英,高红超.基于统计相关性与 K-means 的区分基因子集选择算法[J].软件学报,2014,25(9):2050-2075.

[9] 范波,程久军.用户间多相似度协同过滤推荐算法[J].计算机科学,2012,39(1):23-26.

[10] 田增山,崔永全.基于 BP 神经网络的 GPS/SINS 姿态测量融合算法[J].重庆邮电大学学报:自然科学版,2014(8):478-506.

[11] 李小艳,唐智勇.基于 BP 神经网络的企业知识管理评价模型[J].计算机应用与软件,2014(2):93-95.

[12] 吴浩,罗毅,蔡亮.基于 RBF 神经网络的输电线路故障类型识别新方法[J].重庆邮电大学学报:自然科学版,2013(6):418-426.

[13] 李炯城,肖恒辉,李桂愉.神经网络中处理鞍点的 LMBP 改进算法[J].计算机工程,2012,38(23):173-176,180.

[14] 李翔,朱全银.Adaboost 算法改进 BP 神经网络预测研究[J].计算机工程与科学,2013,35(8):96-102.

[15] CHOI S M,KO S K,HAN Y S.A movie recommendation algorithm based on genre correlations[J]. Expert System with Application,2012,39(9):8079-8085.

[16] 冯本勇.基于 Hadoop 的协同过滤推荐算法[J].电子制作,2014(24):58-59.

[17] PAUL S,BOUTSIDIS C,MAGDON-ISMAIL M,et al. Random Projextions for Support Vector Machines[J]. ACM Transactions on Knowledge Discovery from Data,2014,8(4):498-506.

[18] BOBADILLA J,ORTEGA F,HERNANDO A.A collaborative filtering similarity measure based on singularities[J].Information Processing and Management,2012, 48(2):204-217.

[19] 文俊浩,舒珊.一种改进相似性度量的协同过滤推荐算法[J].计算机科学,2014,41(5):68-71.

[20] LOPS P,GEMMIS M D,SEMERARO G,et al.Content based and collaborative techniques for tag recommendation:an empirical evaluation[J]. Journal of Intelligent Information Systems, 2013,40(1):41-61.

Study of Personalized Recommendation Algorithm Based on BP Neural Network

LUO Pinjie[a], *WEN He*[b]

(a. Experimental Management Center; b. Department of Computer Science and Technology, Chengdu Neusoft University, Chengdu 610000, China)

Abstract: In the era of big data, personalized recommendation helps users alleviate information overload problem by recommending users with the personalized products, information or services. The personalized recommendation algorithm based on BP neural network is proposed, how to establish a hybrid recommendation model is studied, and the accuracy of user similarity algorithm is improved by using BP neural network training. Experimental results show that the algorithm can effectively reduce the user similarity calculation deviation and improve the quality of personalized recommendation.

Key words: personalized recommendation algorithm; improved BP neural network; the hybrid recommendation system; user similarity; momentum factor