

基于 FNN 的家纺产品个性化推荐系统的研究

华铨平

(浙江纺织服装学院 信息工程学院, 浙江 宁波 315211)

摘要:在传统推荐系统模型的基础上,结合预购买家纺产品的特性,建立了预购买家纺产品的推荐模型。研究了推荐系统预购买家纺产品的偏好度的数据提取算法。在此基础上,结合用户在购买家纺产品过程中具有确定性与不确定性的特性以及模糊神经网络的特性,提出了一种模糊神经网络推荐算法。通过对推荐系统的算法的量化研究、仿真研究,显示算法具有较小的 MAE 值,且随着预购买家纺产品采样人数的增加,MAE 依然保持较小的数值,证明了推荐系统的基准集与推荐集具有较好的一致性、稳定性,推荐系统具有较好的聚类特性,说明了基于模糊神经网络算法在家纺产品电子商务推荐系统是可行的。

关键词:推荐系统;模糊神经网络;聚类;MAE;仿真

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2011)09-0183-04

Research on Personalized Recommendation System of Textile Products Used by Family Based on FNN

HUA Quan-ping

(Department of Information Engineering, Zhejiang Textile & Fashion College, Ningbo 315211, China)

Abstract: Recommendation model to buy textile products used by family is established after the traditional recommend system model is analyzed, combining to buy the textile products characteristics. First, it analyzes data collection and data refining algorithm of recommender system to buy textile products used by family. Then a fuzzy neural network recommendation algorithm is built, which is based on fuzzy neural network features and on fuzzy to buy textile products. Finally, simulation experiment show that these values of MAE is quite small, and keeping these small values of MAE with sampling data increasing. All these results demonstrate that the recommendation algorithm performed well.

Key words: recommendation system; fuzzy neural network (FNN); clustering; MAE; simulation

0 引言

电子商务推荐系统能够提高电子商务网站的竞争力,满足客户个性化需求,是发展的必然趋势。目前国内外对推荐技术和系统的研究比较活跃。但具体到家纺产品上,并没有引起大的关注。国内外销售家纺产品的电子商务网站比较多,把当前先进的推荐技术与家纺产品的特殊性结合起来,开发具有实用价值的家纺产品个性化推荐系统,研究适合家纺行业的推荐模型与技术,具有十分重要的意义。

本课题研究的目的是在传统推荐系统模型的基础上,结合了预购买家纺产品的特性,建立了预购买家纺产品的推荐模型。在模型的进一步研究中,首先研究

了推荐系统预购买家纺产品偏好度的数据提取算法,结合用户在购买家纺产品过程中具有确定性与不确定性的特性以及模糊神经网络(Fuzzy Neural Network, FNN)的特性^[1],提出了一种 FNN 推荐算法,通过对算法的量化、仿真研究,证明了推荐系统的基准集与推荐集具有较好的一致性和稳定性。该推荐系统有较好的聚类特性,说明了基于 FNN 算法在家纺产品电子商务的推荐系统是可行的。

1 方案设计

1.1 问题描述

根据文献[2],做如下家纺产品个性化推荐系统的定义:设 C 为访问家纺产品的用户(User)的集合, $C = \{c_1, c_2, \dots, c_n\}$, S 为所有可以推荐给访问家纺产品的用户的对象(object)的集合, $S = \{s_1, s_2, \dots, s_m\}$, 设效用函数 u 为计算推荐对象 $s_i (1 \leq i \leq m)$ 与用户 $c_j (1 \leq j \leq n)$ 的推荐度和产品的信任度。关系式如(1)式:

收稿日期:2011-02-19;修回日期:2011-05-12

基金项目:浙江省教育厅科研项目(Y201017814);宁波市自然科学基金项目(201001A6001043)

作者简介:华铨平(1968-),男,副教授,研究方向为智能信息系统,知识管理,数据挖掘,信息安全处理。

$$u: C \times S \rightarrow u \quad (u > 0, u \in R_n) \quad (1)$$

现在的问题为:如何找到推荐度 u 最大的那个家纺产品的对象 s^* 。用公式(2)表示为:

$$s^* = \operatorname{argmax}_u(c_i, s_j) \quad (2)$$

$$i \in [1, \dots, n], j \in [1, \dots, m], s^* \in S$$

1.2 模糊神经网络

模糊神经网络系统实现了模糊逻辑控制思想与神经网络学习能力的结合。通过对样本数据的学习,得到模糊控制规则和隶属函数,克服了人为选择模糊控制规则主观性较大的缺陷。此外,模糊神经网络系统容易融入领域专家知识,从而提高了神经网络的控制能力^[3-7]。

由于用户在购买家纺产品的过程中,购买行为除了具有确定性外,还有一个很重要的特征是购买行为具有不确定性,即具有模糊特征,因此,如何合理地把确定性购买行为与模糊性购买行为结合在一起,便是模糊神经网络的一个很大的优势。

1.3 基于模糊神经网络推荐模型的建立

个性化家纺电子商务推荐系统包含 2 个核心部分,即通过建立对象特征提取模型,对用户浏览家纺产品信息进行采样,再由基于模糊神经网络的推荐算法处理采样信息,形成聚类输出,完成向用户进行个性化家纺产品推荐的整个过程^[8,9]。模型如图 1 所示。

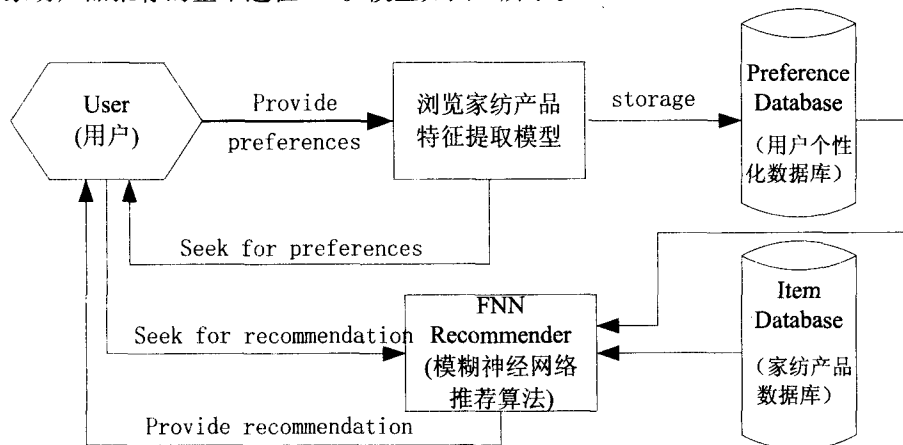


图 1 模糊神经网络推荐模型

模型说明:浏览家纺产品特征提取模型的功能为完成对注册用户浏览家纺产品特征的提取计算;用户个性化数据库的功能为完成对生成家纺产品为特征的提取计算结果的存贮;家纺产品数据库的功能为存贮家纺产品方面的产品信息;模糊神经网络推荐算法的功能为通过输入注册用户浏览家纺产品特征值,结合家纺产品数据库,推荐给用户满意的家纺产品。

1.4 浏览家纺产品特征提取与处理

对 Web 日志文件及 Web 数据库中的数据进行数据预处理。Web 服务器保存了对 Web 页面每一次访

问的日志项,它包含了所请求的 URL、发出请求的 IP 地址和时间戳^[10-12]。文中选用显式浏览输入与用户购买历史输入。显示浏览输入包括的内容有:用户浏览了家纺产品(home textile product)的 URL、浏览时间(browse time),用户购买历史是指用户对家纺产品的购买史。数据预处理模块主要包括以下操作流程:(1)数据清理;(2)路径完善;(3)相关性分析;(4)数据变换。对某一家纺产品的 Web 日志文件与 Web 数据库数据处理以后,得到表 1。

表 1 用户浏览家纺产品与购买家纺产品数据表

域	描述
PageDate	请求家纺产品页面的日期、时间
Vtime	完成浏览的时间
hit	访问某家纺产品
Record	购买家纺产品的记录

从表 1 中,可以挖掘出某一用户对某一家纺产品的购买爱好,也就是偏好值。具体的数据处理方法如下:

$$P_{ik}^j = a * (\text{PageDate}_{ik}^j - \text{Vtime}_{ik}^j) + b * \text{hit}_{ik}^j + c * \text{Record}_{ik}^j \quad (3)$$

参数说明: P_{ik}^j 代表用户 i 对第 j 系列家纺产品的第 k 类别的偏好值, Record_{ik}^j 表示用户 i 对第 j 系列家纺产品的第 k 类别的购买历史记录, $(\text{PageDate}_{ik}^j - \text{Vtime}_{ik}^j)$ 表示用户 i 对第 j 系列家纺产品的第 k 类别的浏览时间, hit_{ik}^j 表示用户 i 对第 j 系列家纺产品的第 k 类别的点击次数, a, b, c 均为常数。

为了去掉采样信号中的杂质,使采样数据尽可能逼近准确,进行如下的数据处理:

$$\text{Time}_{ik}^j = \text{PageDate}_{ik}^j - \text{Vtime}_{ik}^j \quad (4)$$

$$\text{averageTime}_{ik}^j = \frac{1}{n_{ik}^j} \sum_{k=1}^n \text{Time}_{ik}^j \quad (5)$$

$$\text{averagehit}_{ik}^j = \frac{1}{n_{ik}^j} \sum_{k=1}^n \text{hit}_{ik}^j \quad (6)$$

$$E_T = \frac{|\text{Time}_{ik}^j - \text{averageTime}_{ik}^j|}{\text{averageTime}_{ik}^j} \quad (7)$$

$$E_H = \frac{|\text{hit}_{ik}^j - \text{averagehit}_{ik}^j|}{\text{averagehit}_{ik}^j} \quad (8)$$

$$\text{NewTime}_{ik}^j = \begin{cases} \text{Time}_{ik}^j, E_T \in [0, 1) \\ 0, \text{other} \end{cases} \quad (9)$$

$$\text{Newhit}_{ik}^j = \begin{cases} \text{hit}_{ik}^j, E_H \in [0,1] \\ 0, \text{other} \end{cases} \quad (10)$$

为了便于研究,对变量 NewTime_{ik}^j 、 Newhit_{ik}^j 进行线性归一化处理,这里采用线性归一化方法:

$$T_{ik}^j = \frac{(\text{NewTime}_{ik}^j - \text{NewTime}_{ik\min}^j)}{(\text{NewTime}_{ik\max}^j - \text{NewTime}_{ik\min}^j)} \quad (11)$$

$$H_{ik}^j = \frac{\text{Newhit}_{ik}^j - \text{Newhit}_{ik\min}^j}{\text{Newhit}_{ik\max}^j - \text{Newhit}_{ik\min}^j} \quad (12)$$

参数说明: T_{ik}^j 表示用户 i 对第 j 系列家纺对产品的第 k 类别的浏览时间的线性归一化时间,取值范围为 $[0,1]$; H_{ik}^j 表示用户 i 对第 j 系列家纺对产品的第 k 类别的点击次数的归一化点击次数,取值范围为 $[0,1]$; $\text{NewTime}_{ik\min}^j$ 表示用户 i 对第 j 系列家纺对产品的第 k 类别的最短浏览时间, $\text{NewTime}_{ik\max}^j$ 表示用户 i 对第 j 系列家纺对产品的第 k 类别的最长浏览时间; $\text{Newhit}_{ik\min}^j$ 表示用用户 i 对第 j 系列家纺对产品的第 k 类别的点击次数的最少点击数, $\text{Newhit}_{ik\max}^j$ 表示用户 i 对第 j 系列家纺对产品的第 k 类别的点击次数的最大点击数;故(3)式为:

$$P_{ik}^j = a * T_{ik}^j + b * H_{ik}^j + c * \text{Record}_{ik}^j \quad (13)$$

考虑用户 i 如果已购买过第 j 系列家纺对产品的第 k 类别,说明用户依然对 j 系列家纺产品有偏好,故继续推荐。具体处理方法为:如果购买过,则取 Record_{ik}^j 为 1,否则为 0;故(13)式为:

$$P_{ik}^j = \begin{cases} a * T_{ik}^j + b * H_{ik}^j + 1, \text{Record}_{ik}^j = 1, c = 1 \\ a * T_{ik}^j + b * H_{ik}^j, \text{Record}_{ik}^j = 0, c = 1 \end{cases} \quad (14)$$

参数说明:取 $c = 1$,选择 a 、 b 的值,使 $(a * T_{ik}^j + b * H_{ik}^j)$ 取值范围为 $[0,1]$,则 P_{ik}^j 取值范围为 $[0,2]$ 。

1.5 模糊神经网络推荐算法分析

该模糊神经网络由 4 层网络构成,结构如图 2 所示。

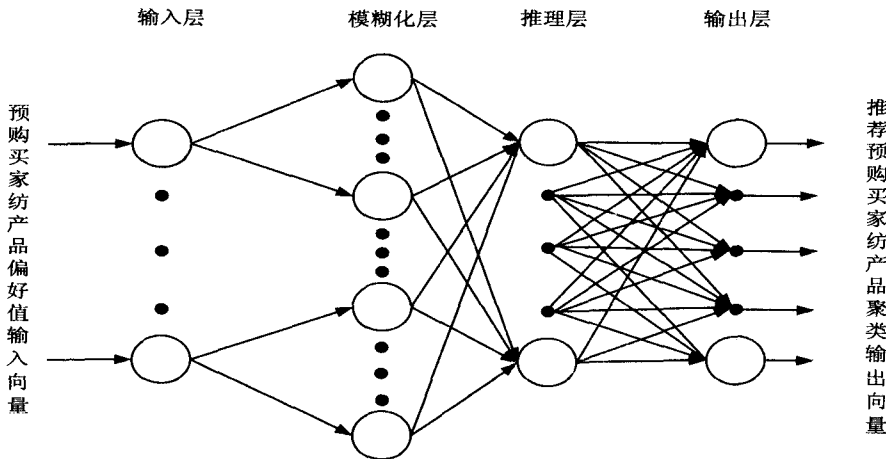


图2 模糊神经网络结构

第1层:对 P_{ik}^j 家纺产品偏好值的输入向量,是 $[0,2]$ 之间的连续的值。

$$O_i^{(1)} = I_i^{(1)} = P_{ik}^j \quad (15)$$

参数说明: $O_i^{(1)}$ 代表第一层第 i 个神经元的输出, $I_i^{(1)}$ 代表第一层第 i 个神经元的输入。

第2层:模糊化层,对输入的 P_{ik}^j 进行模糊化处理。结合推荐信系统本身的特性,这里选择鲁棒性比较好,分辨率比较低的梯形隶属度函数。则模糊化层函数如(16)式所示:

$$O_{ik}^{(2)} = \begin{cases} \frac{O_i^{(1)} - a_{ik}}{b_{ik} - a_{ik}}, O_i^{(1)} \in [a_{ik}, b_{ik}] \\ 1, O_i^{(1)} \in [b_{ik}, c_{ik}] \\ \frac{d_{ik} - O_i^{(1)}}{d_{ik} - c_{ik}}, O_i^{(1)} \in (c_{ik}, d_{ik}] \\ 0, \text{other} \end{cases} \quad (16)$$

参数说明: a_{ik} 、 b_{ik} 、 c_{ik} 、 d_{ik} 分别表示为第 i 个输入对第 k 个模糊词集梯形函数的 a 、 b 、 c 、 d 的值, $O_{ik}^{(2)}$ 表示第二层第 i 个神经元经过第 k 个模糊词集运算的输出。

第3层:推理层,每个神经元代表一条模糊规则,结合推荐系统本身的特点,选用最大化规则推荐策略。

$$O_i^{(3)} = I_i^{(3)} = \bigcup_j^n O_{jk}^{(2)} \quad (17)$$

参数说明: $O_i^{(3)}$ 代表第三层第 i 个神经元的输出, $I_i^{(3)}$ 代表第三层第 i 个神经元的输入, $\bigcup_j^n O_{jk}^{(2)}$ 代表对 $O_{jk}^{(2)}$ 取并集运算。

第4层:输出层。可被视为线性神经元层,采用加权平均法,即

$$I_i^{(4)} = \sum_j W_j O_j^{(3)}, y_{ik}^j = O_i^{(4)} = \frac{I_i^{(4)}}{\sum_j O_j^{(3)}} \quad (18)$$

参数说明: $I_i^{(4)}$ 表示第四层第 j 个神经元的输入, W_j 表示第四层第 i 个神经元的权值, $O_i^{(4)}$ 表示第四层第 i 个神经元的输出, y_{ik}^j 表示用户 i 对第 j 系列家纺对产品的第 k 类别的偏好值的初始预测值。

该推荐系统的模糊神经网络的算法学习采用比较成熟基于变尺度 (DFP) 优化学习算法的改进型学习算法 (MDFP)^[4,5]。该算法的特点是学习速度快、收

敛速度快、稳定性好。

1.6 推荐结果输出

聚类处理完成后,检查各聚类群组中各用户的历史浏览记录,取出该群组中最热门(Top-N)的信息项目,得到初始的推荐列表(recommendation list),形成初始推荐结果。最终推荐的家纺产品表达式如(19)所示:

$$R_{ik}^j = y_{ik}^j + \theta_{ik}^j \quad (19)$$

参数说明: R_{ik}^j 表示用户*i*对第*j*系列家纺对产品的第*k*类别的偏好值的最终预测值; θ_{ik}^j 表示用户*i*对第*j*系列家纺对产品的第*k*类别的产品流行款式关注度,是一个修正值,根据具体的情况取值。

2 仿真实验的结果和分析

2.1 实验结果

在实验中,随机抽取日志文件中5个、10个、15个、20个、30个用户的数据,其中用户预购买的产品的类别数为20,产品的种类数为10。经过对已购买家纺产品的用户日志文件由上面采样算法的处理,产生训练集,然后由训练集在推荐算法中进行学习,形成相应的家纺产品聚类,由这些聚类形成基准集。用偏好值表示的基准集为: $\{R_{11}^1, \dots, R_{ik}^j, \dots, R_{NL}^M\}$,且*N*、*M*、*L*的取值分别为:*N*=30, *M*=20, *L*=10,预购买家纺产品的用户在基准集与推荐算法的作用下,完成家纺产品的购买过程,通过对已购买的家纺产品用户日志文件的分析,以及算法处理,得出推荐集,用偏好值表示的推荐集为: $\{(R_{11}^1), \dots, (R_{ik}^j), \dots, (R_{NL}^M)\}$ 。

推荐系统的 MAE 值的计算公式为^[10,11]:

$$MAE = \frac{\sum_{i=1}^N |R_{ik}^j - (R_{ik}^j)'|}{N} \quad (20)$$

分别取 *N*=5,10,15,20,30,得出如图3所示的实验结果。

2.2 实验分析

分析图3实验结果,得出以下几点结论:

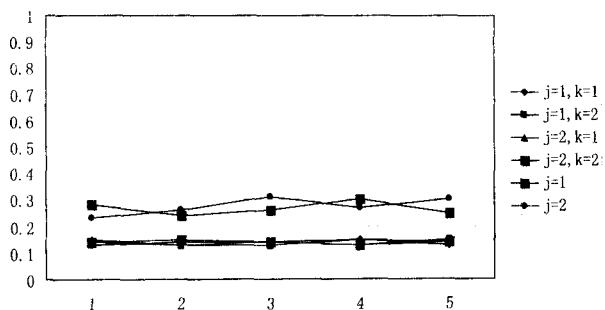


图3 算法计算的实验结果

(1)在推荐同一系列同一类别家纺产品的情况

下,随着用户人数的增加,MAE 的值很小,且 MAE 值变化不是太明显,证明了推荐系统算法在同一系列同一类别的家纺产品的情况下,推荐算法具有很好的一致性、稳定性,且具有很好的聚类特性;

(2)在推荐同一系列不同类别家纺产品的情况下,随着用户人数的增加,MAE 的值虽然比较第一种情况,有点变大,但是依然保持比较小的值,且 MAE 值变化不是太明显,证明了推荐系统算法在同一系列不同类别的家纺产品的情况下,推荐算法具有很好的一致性、稳定性,且具有很好的聚类特性。

3 结束语

文中通过分析用户的浏览行为,判断用户对家纺产品的偏好度,通过模糊神经网络的聚类学习,给用户推荐相关聚类的家纺产品。理论分析与实验结果证明该算法的科学性与合理性。这一算法的合理应用,也是模糊神经网络在推荐系统中的有力尝试。为今后与其它推荐系统算法的比较研究,奠定了一定的基础。

参考文献:

- [1] Kosko B. Fuzzy system as Universal Approximators[C]// IEEE Fuzzy'92. [s.l.]:[s.n.],1992:1153-1162.
- [2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. IEEE Trans. on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [3] 林倩瑜,冯少荣,张东站. 基于神经网络和模式匹配的股票预测研究[J]. 计算机技术与发展, 2010, 20(5): 17-20.
- [4] 张乃尧,阎平凡. 神经网络与模糊控制[M]. 北京:清华大学出版社,1998.
- [5] 李士勇. 模糊控制·神经控制和智能控制论[M]. 哈尔滨:哈尔滨工业大学出版社,2004.
- [6] 朱晓铭,王士同. 单体模糊神经网络的学习规则及其收敛性研究[J]. 计算机研究与发展, 2001, 38(9): 1057-1060.
- [7] 刘云辉,李钟慎. 改进型模糊神经网络模型的构造[J]. 华侨大学学报(自然科学版), 2010, 31(3): 256-259.
- [8] 王宁. 一种基于BP神经网络的即时在线推荐系统[J]. 计算机技术与发展, 2009, 19(7): 230-233.
- [9] Balabanovic M, Shoham Y. Fab: Content-Based, collaborative Recommendation[J]. Communications of ACM, 1997, 40(3): 66-72.
- [10] 李晓昀,余颖. 基于隐性反馈的自适应推荐系统研究[J]. 计算机工程, 2010, 36(16): 270-275.
- [11] 潘宇,林鸿飞,杨志豪. 基于用户聚类的电子商务推荐系统[J]. 计算机应用与软件, 2008(4): 25-26.
- [12] 刘韵毅,梁樑. 基于用户偏好的文献推荐系统[J]. 情报理论与实践, 2007, 30(1): 61-64.