

基于泛函网络的组合推荐算法

崔春生^{1,2,3}

(1. 河南财经政法大学 计算机与信息工程学院, 郑州 450046; 2. 中国社会科学院 数量经济与技术经济研究所, 北京 100010; 3. 河南财经政法大学 经济研究中心, 郑州 450046)

摘 要 在研究组合算法的基础上, 提出一种基于泛函网络实现前融合推荐算法. 探讨 Vague 集条件下推荐算法的前融合问题, 给出了基于泛函网络构架实现前融合组合推荐算法的一般过程. 通过推荐系统泛函拓扑结构的建立, 利用泛函神经元的自学习、自组织和自适应能力, 进一步优化推荐结果, 较大地提高了系统的推荐准确度. 最后, 将算法应用于 Movielens 推荐系统中, 计算机仿真实验结果表明, 本文提出的基于泛函网络实现前融合推荐算法是有效的、可靠的.

关键词 推荐系统; 组合推荐算法; 电子商务; 泛函网络; Vague 集

Hybrid recommendation based on functional network

CUI Chun-sheng^{1,2,3}

(1. College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China; 2. Institute of Quantitative & Technical Economics, Chinese Academy of Social Sciences, Beijing 100010, China; 3. Economic Research Center, Henan University of Economics and Law, Zhengzhou 450046, China)

Abstract Based on the research of hybrid recommendation, model hybrid recommendation algorithm is provided with functional network. Meanwhile, how to get the model hybrid recommendation with vague set is discussed, so the general process of model hybrid recommendation using functional networks is studied. The topology of a functional network is established to improve the accuracy of the system's recommendation by using the neuron's self-learning, self-organizing and adaptive ability. Finally, the algorithm is used in Movielens recommender system. Computer simulation and experimental results show that the algorithm is effective and reliability.

Keywords hybrid recommendation; recommendation system; e-commerce; functional network; vague set

1 引言

中国互联网络信息中心 (CNNIC) 发布的《第 31 次中国互联网络发展状况统计报告》显示, 截至 2012 年 12 月, 我国网民规模达 5.64 亿, 普及率达到 42.1%, 保持低速增长, 网民规模较 2011 年底增长 5090 万人, 较 2011 年底提升 3.8%. 截至 2012 年 12 月, 我国网络购物用户规模达到 2.42 亿, 网络购物使用率提升至 42.9%. 与 2011 年相比, 网购用户增长 4807 万人, 增长率为 24.8%. 在网民增速逐步放缓的背景下, 网络购物应用依然呈现快速的增长势头. 中国网上零售市场规模达 13205 亿元, 较 2011 年增长 64.7%, 预计 2013 年全国网络零售市场交易额则有望达到 18155 亿元. 数据显示, 2012 年中国网络零售市场交易规模已占到当年社会消费品零售总额的 6.3%. 而这个比例在 2011 年仅为 4.4%, 这或许意味着电商改变零售业格局的开始. 中国电子商务研究中心还预计, 这一比例将保持扩大态势, 到 2013 年达到 7.4%.

电子商务推荐系统^[1]是 20 世纪末随着互联网和电子商务的发展而迅速发展起来的一种技术, 其直观的作用体现在利用智能推荐技术, 避免用户迷失于海量的产品信息中, 帮助用户选择和推荐需要的产品, 提高产品的销售量. 它是一个涵盖心理学、管理学、信息学等多学科的研究领域^[2], 是解决信息超载问题的有

收稿日期: 2013-06-26

资助项目: 河南科技厅基础与前沿技术研究项目 (132300410011); 河南省科技厅软科学项目 (142400410313); 河南省社科规划办项目 (2013BJJ061); 河南省教育厅科学技术研究重点项目 (14A630013)

作者简介: 崔春生 (1974-), 男, 汉, 河南南阳人, 副教授, 博士后, 研究方向: 电子商务智能推荐, 决策理论与方法, E-mail: traiton@126.com.

效途径之一^[3]。电子商务推荐系统根据用户的信息需求、兴趣等,将用户感兴趣的信息、产品等推荐给用户。与搜索引擎相比,推荐系统通过研究用户的兴趣偏好,进行个性化计算,由系统发现用户的兴趣点,从而引导用户发现自己的信息需求。一个好的推荐系统不仅能为用户提供个性化的服务,还能和用户之间建立密切关系,让用户对推荐产生依赖^[4]。

2 组合推荐算法研究现状

推荐系统包括推荐输入、推荐算法和推荐输出^[5]三个环节^[6],其中推荐算法是整个推荐系统中最核心和关键的部分,在很大程度上决定了推荐类型和推荐质量^[7]。目前,出现的推荐算法有很多,对其分类也没有一个统一的标准,但受到大家公认的推荐算法包括:基于内容的推荐^[8]、协同过滤推荐、基于知识的推荐、基于网络结构的推荐、组合推荐及关联规则推荐^[9],基于内容的推荐和协同过滤推荐是推荐领域中主流研究的两种算法。

相对于其他推荐算法,组合推荐算法(hybrid recommendation)是近年发展起来的一种新的算法。原因在于组合推荐算法是唯一能够使其他算法的缺点得到补偿的一种有效算法^[10]。没有一种推荐算法是完美的,协同过滤存在新产品的问题,内容过滤对交易频繁的产品推荐质量不够好等等,那么很容易就设想将这些推荐算法组合起来,取长补短,这也就是组合推荐的思想^[11]。理论上讲可以有很多种的推荐组合方法,但目前研究和应用最多的组合推荐是基于内容的推荐和协同过滤推荐的组合。文献[2]根据不同的组合性质,将组合推荐方法分为后融合、中融合、前融合三大类型。2009年结束的Netflix推荐大赛中,最终的顶级小组并没有使用单个的推荐算法,而使用了组合算法,通过组合多种不同的算法来提升预测的性能。如冠军小组BellKor用的Pragmatic Chaos算法即为多级集成学习技术,多次使用的Bagging、Stacking等集成学习技术,用神经网络对多个预测模型建模进而得到最终的模型。

1) 后融合方法(result hybrid recommendation)

后融合推荐方法是融合两种或两种以上的推荐方法得到的推荐结果列表,融合列表的结果决定最后推荐的产品。如何从众多推荐结果中选择用户需要的推荐结果成为该算法的一个重要研究点。例如文献[12]使用了评分结果的线性组合,文献[13]使用了投票机制来组合推荐结果,采用一定的标准对两者产生的推荐结果判断。另外也可以分别考察两个推荐列表,判断使用其中的哪个推荐结果,如Daily Learner system^[14]计算推荐结果的可信度,之后选择一个列表的结果。后融合一般采用加权、混合等融合方式。

2) 中融合方法(process hybrid recommendation)

中融合推荐方法是以一种推荐方法为基本框架,融合另一种推荐方法。目前中融合的组合推荐主要有两种。一种是以协同过滤的方法为框架融合基于内容的方法。为了克服协同过滤的稀疏问题,这种方法把用户当作对象,使用基于内容的特征提取方法将用户本身的特征如年龄、工作情况等人口统计学特征使用到相似度计算中,而不是仅仅依赖用户的点击行为。Good等人^[15]引入多种不同的用户描述符来归类用户,挖掘用户的内在联系,从而得到更好的推荐效果。文献[16]使用独立的基于内容的特征来补偿用户提供的简单的rating。另一种是以基于内容的方法为框架,融合协同过滤的方法。该方法利用降维技术把基于内容的对象特征进行精简,如文献[17]使用LSI(latent semantic indexing)算法,在基于内容的框架中使用精化的用户特征向量。中融合一般采用层叠、特征组合、特征扩充、Mata-level等融合方式。

3) 前融合方法(model hybrid recommendation)

前融合推荐方法是直接融合各种推荐方法,如将基于内容和协同过滤的方法整合到一个统一的框架模型下。近年来这类推荐方法最受学者的关注,例如协同推荐的框架内混合基于内容的推荐(或相反)、基于协同推荐的框架内混合基于网络结构的推荐,社会网络分析法的推荐框架内混合基于内容的推荐,基于网络结构的推荐和基于社会网络分析法的推荐的混合等。文献[18]中,研究者把用户的年龄和电影的类型放到一个统一的分类器中训练学习。另外一种方法^[19]使用了贝叶斯混合效果回归模型,并通过马尔可夫蒙特卡罗方法得到这个模型的参数。将用户和产品的特征放在一个统计模型下来计算推荐函数,定义用户属性 z 、产品属性 w 及其交互关系 x 来计算效用 R ,产品 s_i 对于用户 c_u 的推荐值 R_{ui} 可以表示为:

$$R_{ui} = x_{ui}\mu + z_u\gamma_i + w_i\lambda_u + e_{ui},$$

这里, $e_{ui} \sim N(0, \sigma^2)$, $\gamma_i \sim N(0, \tau)$, $\lambda_u \sim N(0, \Lambda)$ 。

其中, 3种正态分布的变量分别用于描述数据的噪声、产品属性的异质性和用户属性的异质性。该式说

明推荐值是由这几个因素共同决定的. 这 3 种分布的 3 个参数由马尔可夫蒙特卡罗方法估算得到前融合一般变换等融合方式.

国内外已有的研究中, 大多采用中融合方法 [10,20]. 这种方法以推荐方法 A 为基础, 融合另一种推荐方法 B, 因此 A 方法的缺点得到补偿, 但是 B 方法的优点不能发挥. 也有采用后融合方法, 目的在于解决协同过滤中的冷启动 [21–22]、数据稀疏 [23–24] 和内容推荐算法中的推荐质量不高等问题, 但容易诱发新的问题, 如新资源以及海量数据等.

3 泛函网络的提出

泛函网络 [25–26] 是由西班牙学者 Enrique Castillo 提出的, 是对神经网络的一种有效推广. 1998 年 Enrique Castillo 给出了泛函网络的基本模型及其构成元素, 介绍了泛函网络与人工神经网络的区别, 提出了可分离泛函网络模型与广义可结合泛函网络模型, 并将泛函网络应用于回归问题. 之后泛函网络又被应用于非线性系统辨识 [27], 线性及非线性回归 [26], 微分、差分, 泛函方程求解 [28] 以及混沌时间序列预测 [29] 等领域.

国内对泛函网络也进行了一定的研究. 1996 年, 李洪兴 [30] 教授通过对人工神经网络和泛函网络的分析指出, 泛函连接网络中的函数形式仍然可以采用神经元函数来表示. 2001 年李春光给出非线性系统辨识的泛函网络方法 [31], 2003 年李卫斌等提出了三类具有可分离交换性的泛函网络模型 [32], 2005 年戴祯杰等人提出了多项式泛函网络模型 [33]. 2005–2008 年, 周永权 [34–35] 发表了“层次泛函网络整体学习算法”、“一种复值可分离的泛函网络学习算法”、“新型 Sigma-Pi 泛函网络模型”、“泛函网络神经元函数类型优化”等一系列文章, 极大地推动了泛函网络的发展. 近几年, 泛函网络的应用拓展到海洋、农业、控制、机械、气候、生物等多个领域 [36].

泛函网络是对人工神经网络的一种扩展和拓广 [34], 其组成部分 [37] 包括五个部分 (如图 1 所示).

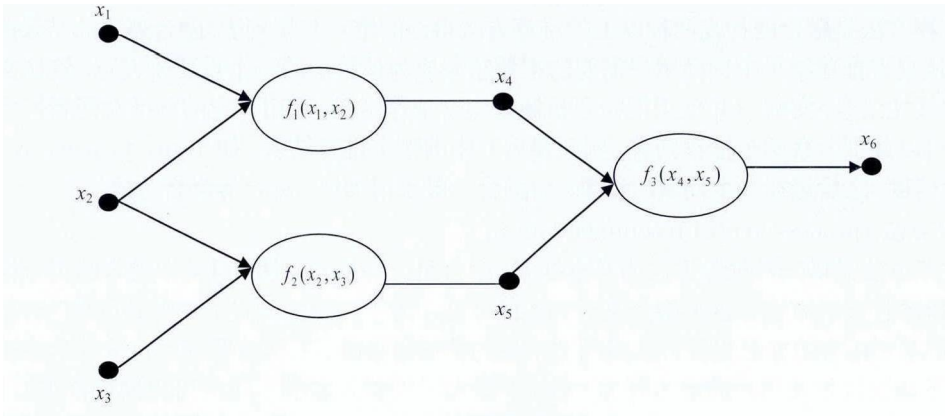


图 1 泛函网络结构图

①输入层. 其功能与人工神经网络一样, 也是输入信息, 输入单元以带有相应名字的实心圆来表示. 在图 1 中输入层包含 $\{x_1, x_2, x_3\}$.

②输出层. 它输出网络计算结果, 是网络中的最后一层单元, 输出单元也以带有相应名字的实心圆来表示, 图 1 中为 $\{x_6\}$, 其输出信息为:

$$x_6 = f_3(x_4, x_5) = f_3(f_1(x_1, x_2), f_2(x_2, x_3)).$$

③中间存储单元层. 它存储由神经元产生的信息, 中间存储单元层可以有多层, 中间存储单元也以带有相应名字的实心圆来表示. 在图 1 中只有一层, 包含 $\{x_4, x_5\}$, 存在关系: $x_4 = f_1(x_1, x_2)$, $x_5 = f_2(x_2, x_3)$.

④泛函神经元 (或称处理单元). 泛函网络中每个神经元都是一个独立的计算单元, 它计算来自输入单元或前一层神经元的输入值, 并为下层输出单元或神经元提供数据结果. 计算单元之间用连接线相互连接, 各神经元的输出可作为下一神经元或输出单元输入的一部分. 系统一旦获取输入值, 输出便由神经元的类型来确定. 在图 1 中有两层泛函神经元, 分别为 $\{f_1, f_2\}$ 和 $\{f_3\}$.

⑤连接线. 它们连接输入层、中间层神经元和输出层, 在图 1 中用带箭头的直线表示, 箭头的方向表示信息流的方向.

类似于人工神经网络, 泛函网络也需要通过学习获得知识. 两者的差别在于, 人工神经网络是对权值的

学习, 而泛函网络是对网络参数和结构的学习神经元不是固定的, 具有自学习的功能. 泛函网络通常是一些给定的基函数簇的线性组合, 在实际应用中可根据所求问题的物理背景知识和结构学习评价准则, 选择不同的基函数簇实现用较小的网络规模获得更满意的泛化特性^[38]. 泛函网络的学习有两种, 一种是根据具体问题进行泛函网络的构建和简化; 另一种是神经元的学习. 泛函网络的学习方法根据不同的精确程度分为近似学习和精确学习. 前者是依据给定的样本评价神经元函数的质量; 后者是确定泛函网络对应泛函方程的解函数^[39].

可见, 泛函网络对模型的整合, 恰好迎合了推荐组合的思想, 泛函网络的输入层正是组合推荐算法中的组合要素, 输出层正是推荐系统希望得到的推荐结果, 中间存储单元层表征了推荐模型之间的融合与整合, 泛函神经元为推荐系统的优化提供了可能, 也为推荐系统的自学习提供了理论基础和实现可能, 而连接线则让整个推荐过程形成一个有机的整体. 同时, 泛函网络中神经元的自学习能力, 又为推荐系统质量的提高提供了可能性, 这种自学习能力不仅可以有效地提高模型的推荐质量, 也是推荐过程得以实现的基础.

4 泛函网络用于组合推荐的一般过程

采用泛函网络解决推荐组合算法问题, 首先要将被融合的推荐算法用泛函拓扑结构表示. 这里以内容推荐和协同过滤推荐的融合为例, 探讨基于泛函网络的组合推荐算法.

基于内容的推荐算法^[8]的基本思想是根据已知产品和待推荐产品的属性值, 确定产品间相似性, 并以用户对已知产品的评分作为权重, 得到待推荐产品的推荐值. 图 2 表示了基于内容推荐算法的基本思路.

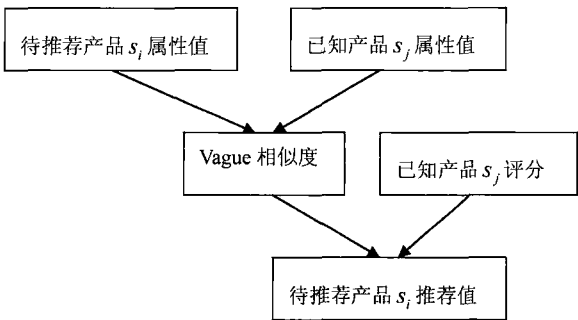


图 2 基于内容推荐算法结构图

根据泛函网络的建模思想, 得到内容推荐算法的泛函网络图, 如图 3 所示.

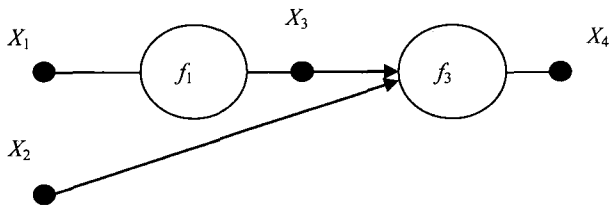


图 3 基于内容推荐算法泛函网络图

其中,

$X_1 = S = \{s_i\}, i = 1, 2, \cdots, m$, 表示所有产品 (product) 的集合.

$X_2 = R = \{R_{ui}\}, u = 1, 2, \cdots, n, i = 1, 2, \cdots, m$, 表示用户 (user) c_u 对产品 s_i 评分的集合.

$X_3 = f_1(X_1) = Sim^V(s_i, s_j)$ 是所有产品 Vague 相似度的集合.

$X_4 = f_3(X_3, X_2) = X_3 X_2 = R(s_i)$ 是所有待推荐产品 s_i 的推荐值.

显然, 该网络是一个多输入单输出的泛函网络.

进而得到产品 s_i 相对用户 c_u 的推荐值为:

$$R(s_{ui}) = f_3(f_1(X_1), X_2) = X_2 f_1(X_1) = \sum_{\substack{j=1 \\ j \neq i}}^m R_{uj} Sim^V(s_i, s_j), \quad i = 1, 2, \cdots, m$$

(1)

产品 s_i 相对系统的推荐值为:

$$R(s_i) = f_3(f_1(X_1), X_2) = X_2 f_1(X_1) = \sum_{\substack{j=1 \\ j \neq i}}^m \sum_{u=1}^n R_{uj} Sim^V(s_i, s_j), \quad i = 1, 2, \cdots, m$$

(2)

协同过滤推荐算法的基本思想是根据系统中用户的历史评分, 确定用户间相似性, 从而确定待推荐用户的最邻近用户, 并以用户对已知产品的评分作为权重, 得到待推荐产品的推荐值. 图 4 表示了协同过滤推荐过程.

根据泛函网络的建模思想, 得到协同过滤推荐算法的泛函网络图, 如图 5 所示.

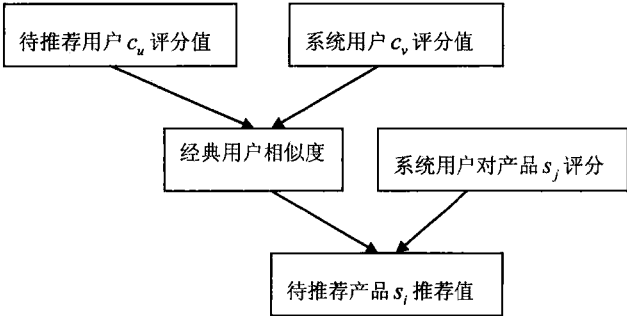


图 4 协同过滤推荐算法结构图

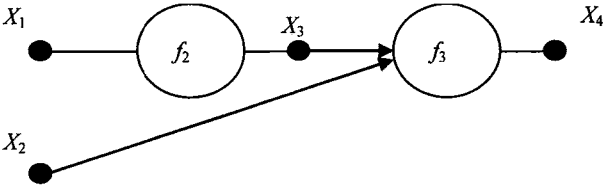


图 5 协同过滤推荐算法泛函网络图

其中,
 $X_1 = C = \{c_u\}, u = 1, 2, \cdots, n$, 表示所有用户 (user) 的集合.
 $X_2 = R = \{R_{ui}\}, u = 1, 2, \cdots, n, i = 1, 2, \cdots, m$, 表示用户 (user) c_u 对产品 s_i 评分的集合.
 $X_3 = f_1(X_1) = Sim(c_u, c_v)$ 是所有用户经典相似度的集合.
 $X_4 = f_3(X_3, X_2) = X_3 X_2 = R(s_i)$ 是待推荐产品的 s_i 推荐值.
显然, 该网络也是一个多输入单输出的泛函网络.
得到产品 s_i 相对用户 c_u 的推荐值为:

$$R(s_{ui}) = f_2(f_1(X_1), X_2) = X_2 f_2(X_1) = \sum_{\substack{u=1 \\ u \neq v}}^n R_{ui} Sim(c_u, c_v), \quad i = 1, 2, \cdots, m$$

(3)

产品 s_i 相对系统的推荐值为:

$$R(s_i) = f_3(f_2(X_1), X_2) = X_2 f_2(X_1) = \sum_{v=1}^n \sum_{\substack{u=1 \\ u \neq v}}^n R_{ui} Sim(c_u, c_v), \quad i = 1, 2, \cdots, m$$

(4)

5 基于泛函网络的推荐系统学习过程研究

根据泛函网络学习过程, 得到前融合推荐系统的模型学习过程如下:

- ①问题陈述: 前融合推荐算法的基本思想是将内容推荐算法和协同过滤算法融合为一体, 用统一的产品评分值, 得到待推荐产品的推荐值.
- ②初始结构: 根据泛函网络的建模思想, 得到前融合推荐算法的泛函网络如图 6 所示.

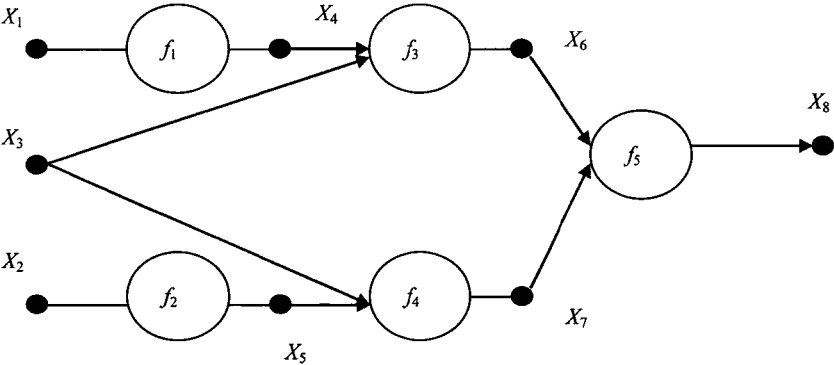


图 6 前融合推荐算法泛函网络图

其中,

$X_1 = S = \{s_i\}, i = 1, 2, \dots, m$, 表示所有产品 (product) 的集合.

$X_2 = C = \{c_u\}, u = 1, 2, \dots, n$, 表示所有用户 (user) 的集合.

$X_3 = R = \{R_{ui}\}, u = 1, 2, \dots, n, i = 1, 2, \dots, m$, 表示用户 c_u 对产品 s_i 评分的集合.

$X_4 = f_1(X_1) = Sim^V(s_i, s_j)$ 是所有产品 Vague 相似度的集合^[6].

$X_5 = f_2(X_2) = Sim(c_u, c_v)$ 是所有用户经典相似度的集合.

$X_6 = f_3(X_3, X_4) = X_3 X_2 = R_1(s_i)$ 是待推荐产品 s_i 的推荐值 (基于内容推荐算法).

$X_7 = f_4(X_3, X_5) = X_3 X_5 = R_2(s_i)$ 是待推荐产品 s_i 的推荐值 (协同过滤推荐算法).

$X_8 = f_5(X_6, X_7) = R(s_i)$ 是待推荐产品 s_i 的推荐值 (组合推荐算法).

③网络学习: 根据基于内容推荐算法和协同过滤算法的泛函网络可知, 以上网络拓扑结构为最简网络结构.

显然, 该网络仍然是一个多输入单输出的泛函网络. 网络中, 上部分表示基于内容推荐算法, 下部分表示协同过滤推荐算法. 通过两种模型的结合, 实现了两种推荐算法的前融合.

④确定表达式: 由网络拓扑结构可知, $X_8 = f_5(X_6, X_7) = R(s_i)$, 且通过基于内容的推荐算法和协同过滤算法可得:

$$X_6 = R_1(s_i), \quad X_7 = R_2(s_i).$$

则产品 s_i 相对用户 c_u 的推荐值为:

$$R(s_{ui}) = f_5(X_6, X_7) = f_5(R_1(s_{ui}), R_2(s_{ui})) \quad (5)$$

产品 s_i 相对系统的推荐值为:

$$R(s_i) = f_5(X_6, X_7) = f_5(R_1(s_i), R_2(s_i)) \quad (6)$$

将 (1) 和 (3) 代入 (5) 式, 将 (2) 和 (4) 代入 (6) 式, 可得组合推荐方法中产品 s_i 相对用户 c_u 的推荐度以及产品 s_i 相对系统推荐值, 即

$$R(s_{ui}) = f_5 \left(\sum_{\substack{j=1 \\ j \neq i}}^m R_{uj} Sim^V(s_i, s_j), \sum_{\substack{u=1 \\ u \neq v}}^n R_{ui} Sim(c_u, c_v) \right) \quad (7)$$

$$R(s_i) = f_5 \left(\sum_{\substack{j=1 \\ j \neq i}}^m \sum_{u=1}^n R_{uj} Sim^V(s_i, s_j), \sum_{u=1}^n \sum_{\substack{i=1 \\ i \neq j}}^m R_{ui} Sim(c_u, c_v) \right) \quad (8)$$

⑤数据收集: 以收集到的 MovieLens 中的 6040 用户对 3900 部电影的评价值作为数据源. 数据源包括三个文件, 其中 ratings.dat 列表收集了 UserID、MovieID、Rating、Timestamp 四项内容, users.dat 列表收集了 UserID、Gender、Age、Occupation、Zip-code 五项内容, movies.dat 列表收集了 MovieID、Title、Genres 三项内容. 鉴于以上问题求解的需要, 这里采用 ratings.dat 数据进行网络学习. ratings.dat 中共有四项内容, 其中 UserID 的范围介于 1 和 6040, MovieID 的范围从 0 到 3592, Rating 分为 5 个等级, Timestamp 以秒为单位记录评分结果返回的时间, 系统中每个用户至少为 20 个电影评分.

采用 Matlab 编制程序, 以函数 generateBtn_Callback, 按照指定的用户数目和产品数目, 从 rating 中选择评分数据作为训练集.

编制函数 trainBtn_Callback, 根据推荐公式计算 $R_1(s_{ui})$, $R_2(s_{ui})$, 并抽取真实评分值 R_{ui} 训练泛函神经元.

编制函数 recommendBtn_Callback, 选择目标用户, 利用神经元计算相应的推荐值.

程序的主界面如图 7 所示.

⑥参数学习: 借助通过线性组合方法得到 (7) 和 (8) 的参数表达形式:

$$R(s_{ui}) = \lambda R_1(s_{ui}) + (1 - \lambda) R_2(s_{ui}) = \lambda \sum_{\substack{j=1 \\ j \neq i}}^m R_{uj} Sim^V(s_i, s_j) + (1 - \lambda) \sum_{\substack{u=1 \\ u \neq v}}^n R_{ui} Sim(c_u, c_v) \quad (9)$$

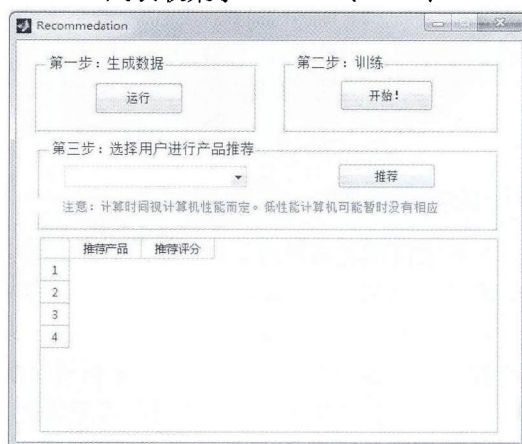


图 7 程序主界面

$$R(s_i) = \lambda R_1(s_i) + (1 - \lambda)R_2(s_i) = \lambda \sum_{\substack{j=1 \\ j \neq i}}^m \sum_{u=1}^n R_{uj} Sim^V(s_i, s_j) + (1 - \lambda) \sum_{u=1}^n \sum_{\substack{j=1 \\ j \neq i}}^m R_{ui} Sim(c_u, c_v) \quad (10)$$

其中 $\lambda(0 \leq \lambda \leq 1)$ 为融合系数, 若认为系统中得到评分的产品数量比较多, 此时基于内容的推荐结果准确率较高, 可以取 $\lambda \geq 0.5$; 同样, 若认为系统中参与评分的用户数量比较多, 此时协同过滤推荐结果准确率较高, 可以取 $\lambda \leq 0.5$.

极端情况下, 若取 $\lambda = 1$, 则 $R(s) = R_1(s_i)$, 组合推荐方法退化为基于内容的推荐算法; 若取 $\lambda = 0$, 则 $R(s) = R_2(s_i)$, 组合推荐方法退化为协同过滤推荐算法.

在 rating 数据表中, 以 0.8 : 0.2 的比例抽取训练集和测试集, 采用函数 trainBtn_Callback 训练神经元, 表 1 给出了训练集中部分 $R_1(s_{ui})$, $R_2(s_{ui})$ 及真实评分值 R_{ui} .

⑦模型确认: 利用一组数据集, 测试泛函网络进行, 确定所选的函数簇是否充分逼近神经元函数.

这里采用平均绝对偏差 MAE 计算公式: $MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{n}$, 抽取用户的预测值与评分值进行模型检验, 得到 $MAE = 0.57625$. 该值相对较小, 模型可以投入运用.

⑧模型运用: 若测试的结果满意, 则模型投入使用.

图 8 给出了通过泛函网络模型得到的用户 901 的电影推荐结果.

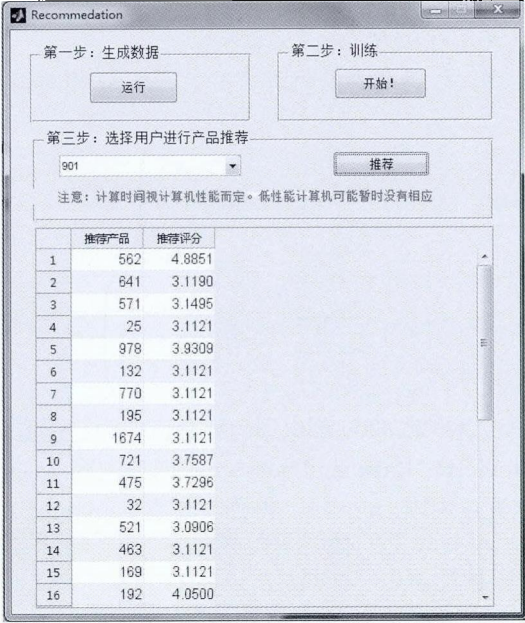


图 8 用户 901 推荐结果

表 1 训练集用户评分预测值

| $R_1(s_{ui})$ | $R_2(s_{ui})$ | R_{ui} |
|---------------|---------------|----------|
| 4.248476 | 3.422755 | 4 |
| 2.39641 | 3.060121 | 3 |
| 4.838056 | 4.173378 | 5 |
| 2.148168 | 2.040247 | 3 |
| 2.770084 | 3.010046 | 3 |
| 1 | 1.083603 | 1 |
| 2.820009 | 2.077495 | 3 |
| 3.611575 | 1.476447 | 2 |
| 4 | 3.172929 | 4 |
| 2.183167 | 1.706274 | 2 |
| 4.607254 | 1.842854 | 3 |
| 4.330175 | 2.44669 | 3 |
| 3.855835 | 4.126197 | 4 |
| 4.178984 | 2.225601 | 3 |
| 4.450638 | 1.721533 | 3 |
| 1.051724 | 4.338832 | 3 |
| 2.858732 | 1.181442 | 2 |
| 3.473198 | 1.706274 | 3 |
| 3.455285 | 4.373229 | 4 |
| 1.680736 | 0.539372 | 2 |
| 2.031104 | 0.538284 | 1 |
| 2.224747 | 1.481242 | 2 |
| 1.915849 | 3.445353 | 2 |
| 2.949208 | 4.85039 | 4 |
| 3 | 2.4731 | 3 |
| 4.814617 | 4.355581 | 5 |

6 结论与展望

论文借助泛函网络实现了电子商务的组合推荐, 取得了以下有效的结论. 第一, 论文建立了组合推荐的泛函拓扑结构. 一方面通过对泛函网络基本特点的研究, 找到了泛函网络与推荐系统的结合点, 建立了内容推荐算法和协同过滤推荐算法的泛函拓扑结构. 另一方面根据前融合推荐算法的思想, 将两个拓扑结构组合成前融合泛函拓扑结构图. 第二, 根据组合推荐的拓扑结构构造了组合推荐数学模型. 通过拓扑结构得到的组合推荐模型既体现了泛函网络的思想又突出了“融合”的理念, 模型中 λ 值的确定体现了组合推荐算法的灵活性. 第三, 通过神经元的学习得到有效的组合推荐泛函方程, 充分利用泛函神经元自学习的特点和推荐系统中庞大的历史数据, 通过 Movielens 中的推荐数据实现神经元的训练, 最终得到组合推荐系统的泛函方程.

受囿于时间和实验条件的限制, 本文尚存在一些有待进一步改进或深入研究的内容. 第一, 论文对于协同过滤算法中的相似度采用经典的方法, 而内容推荐算法中的相似度则采用 Vague 集方法, 两种方法混用虽

然得到了满意的结果,但是产生了计算复杂度和一致性问题.如果能够完成基于 Vague 集的协同过滤推荐算法研究,这一问题将会得到有效的解决.第二,推荐算法已有的研究中均是基于有限数据库的计算,在大数据时代,优化推荐算法值得关注.在 Internet 普及,数据海量增加的条件下,推荐算法的运算性、实时性、可扩展性、准确性等如何保证.第三,用户兴趣漂移问题的研究.在推荐系统中,还应考虑时间维度.因为事物特征,用户兴趣,都可能随时间的推移而变化,尤其是用户兴趣的漂移.用户的兴趣不是一成不变的,兴趣随着市场环境、生活环境、工作环境、支付能力、生理年龄、需求层次等因素的变化都会发生变化,已有推荐系统的研究都是基于用户兴趣固定这样一个假设.实质上,大部分长期兴趣是固定不变的;而与之对应的短期需求是经常发生变化的.

参考文献

- [1] Resnick P, Varian H R. Recommender systems[J]. Communications of the ACM, 1997, 40(3): 56-58.
- [2] 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.
Xu Hailing, Wu Xiao, Li Xiaodong, et al. Comparison study of internet recommendation system[J]. Journal of Software, 2009, 20(2): 350-362.
- [3] 程岩. 电子商务中面向延迟购买行为的易逝品动态捆绑策略 [J]. 系统工程理论与实践, 2011, 31(10): 1892-1902.
Cheng Yan. Delay buying behavior-oriented perishable products dynamic bundling strategy in e-commerce setting[J]. Systems Engineering — Theory & Practice, 2011, 31(10): 1892-1902.
- [4] 王国霞, 刘贺平. 个性化推荐系统综述 [J]. 计算机工程与应用, 2012, 48(7): 66-76.
Wang Guoxia, Liu Heping. Survey of personalized recommendation system[J]. Computer Engineering and Applications, 2012, 48(7): 66-76.
- [5] 崔春生. 基于集团序方法的推荐系统输出 [J]. 系统工程理论与实践, 2013, 33(7): 1845-1851.
Cui Chunsheng. Output of recommender systems based on aggregative rank[J]. Systems Engineering — Theory & Practice, 2013, 33(7): 1845-1851.
- [6] 崔春生, 吴祈宗. 基于 Vague 集的内容推荐算法研究 [J]. 计算机应用研究, 2010, 27(6): 2108-2110.
Cui Chunsheng, Wu Qizong. Research on content-based recommendation based on Vague sets[J]. Application Research of Computers, 2010, 27(6): 2108-2110.
- [7] 崔春生, 齐延信, 田哲, 等. 基于 Vague 值的电子商务推荐系统及其相似度研究 [J]. 图书情报工作, 2012, 56(7): 130-134.
Cui Chunsheng, Qi Yanxin, Tian Zhe, et al. Recommender Systems of EC and its similarity based on vague value[J]. Library and Information Service, 2012, 56(7): 130-134.
- [8] 崔春生, 李光, 吴祈宗. 基于 Vague 集的电子商务推荐系统研究 [J]. 计算机工程与应用, 2011, 47(10): 237-239.
Cui Chunsheng, Li Guang, Wu Qizong. Research on recommender systems of electric commerce based on Vague sets[J]. Computer Engineering and Applications, 2011, 47(10): 237-239.
- [9] 李杰, 徐勇, 王云峰, 等. 面向个性化推荐的强关联规则挖掘 [J]. 系统工程理论与实践, 2009, 29(8): 144-152.
Li Jie, Xu Yong, Wang Yunfeng, et al. Strongest association rules mining for personalized recommendation[J]. Systems Engineering — Theory & Practice, 2009, 29(8): 144-152.
- [10] 洪文兴, 翁洋, 朱顺彪, 等. 垂直电子商务网站的混合型推荐系统 [J]. 系统工程理论与实践, 2010, 30(5): 928-935.
Hong Wenxing, Weng Yang, Zhu Shuzhi, et al. Hybrid recommender system for vertical e-commerce website[J]. Systems Engineering — Theory & Practice, 2010, 30(5): 928-935.
- [11] 朱岩, 林泽楠. 电子商务中的个性化推荐方法评述 [J]. 中国软科学, 2009(2): 183-192.
Zhu Yan, Lin Zenan. A review of e-business recommendation system[J]. China Soft Science, 2009(2): 183-192.
- [12] Claypool M, Gokhale A, Miranda T, et al. Combining content-based and collaborative filters in an online newspaper[C]// Proceedings of ACM SIGIR Workshop on Recommender Systems. Menlo Park: AAAI Press, 1999: 1-11.
- [13] Pazzani M. A framework for collaborative, content-based, and demographic filtering[J]. Artificial Intelligence Review, 1999, 13(5-6): 393-408.
- [14] Billsus D, Pazzani M. User modeling for adaptive news access[J]. User Modeling and User-Adapted Interaction, 2000, 10(2-3): 147-180.
- [15] Good N, Schafer J B, Konstan J A, et al. Combining collaborative filtering with personal agents for better recommendations[C]// AAAI/IAAI. Menlo Park: AAAI Press, 1999: 439-446.
- [16] Melville P, Mooney R J, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C]// AAAI/IAAI. Menlo Park: AAAI Press, 2002: 187-192.
- [17] Soboroff I, Nicholas C. Combining content and collaboration in text filtering[C]// Proceedings of the IJCAI. Menlo Park: AAAI Press, 1999, 99: 86-91.
- [18] Basu C, Hirsh H, Cohen W. Recommendation as classification: Using social and content based information in recommendation[C]// AAAI/IAAI. Menlo Park: AAAI Press, 1998: 714-720.

- [19] Ansari A, Essegaier S, Kohli R. Internet recommendations systems[J]. *Journal of Marketing Research*, 2000, 37(3): 363–375.
- [20] 孙立莹. 基于组合推荐技术的个性化学习资源推荐的研究 [D]. 大连: 大连海事大学, 2010.
Sun Liying. Research on learning materials personalized recommendation based on combination recommended technology[D]. Dalian: Dalian Maritime University, 2010.
- [21] 李聪, 梁昌勇. 基于 n 序访问解析逻辑的协同过滤冷启动消除方法 [J]. *系统工程理论与实践*, 2012, 32(7): 1537–1545.
Li Cong, Liang Changyong. Cold-start eliminating method of collaborative filtering based on n -sequence access analytic logic[J]. *Systems Engineering — Theory & Practice*, 2012, 32(7): 1537–1545.
- [22] 杨毅, 王晓荣, 胡迎春. 基于客户/项目的聚类协同过滤组合推荐算法研究 [J]. *广西工学院学报*, 2011, 22(4): 74–78.
Yang Yi, Wang Xiaorong, Hu Yingchun. Researches of collaborative filtering recommendation algorithm based on user and item clustering combination[J]. *Journal of Guangxi University of Technology*, 2011, 22(4): 74–78.
- [23] 郁雪, 李敏强. 基于 PCA-SOM 的混合协同过滤模型 [J]. *系统工程理论与实践*, 2010, 30(10): 1850–1854.
Yu Xue, Li Minqiang. Effective hybrid collaborative filtering model based on PCA-SOM[J]. *Systems Engineering — Theory & Practice*, 2010, 30(10): 1850–1854.
- [24] 姜维, 庞秀丽. 面向数据稀疏问题的个性化组合推荐研究 [J]. *计算机工程与应用*, 2012, 48(21): 21–26.
Jiang Wei, Pang Xiuli. Research on personal hybrid recommendation overcoming data sparse problem[J]. *Computer Engineering and Applications*, 2012, 48(21): 21–26.
- [25] Castillo E. Functional networks[J]. *Neural Processing Letters*, 1998, 7: 151–159.
- [26] Castillo E, Cobo A, Manuel Gutierrez J. Functional networks with applications[M]. Kluwer Academic Publisher, 1999.
- [27] Castillo E, Hadi A S, Lacruz B, et al. Semiparametric nonlinear regression and transformation using functional networks[J]. *Computational Statistics & Data Analysis*, 2001, 52(4): 2129–2157.
- [28] Castillo E, Cobo A, Manuel Gutierrez J, et al. Working with differential, functional and difference equation using functional networks[J]. *Applied Mathematical Modeling*, 1999, 23(2): 89–107.
- [29] Castillo E, Manuel Gutierrez J. Nonlinear time series modeling and prediction using functional networks. Extracting information masked by chaos[J]. *Physics Letters A*, 1998, 244(1): 71–84.
- [30] 李洪兴. 数学神经网络 (I) —— 神经网络的插值机理 [J]. *北京师范大学学报: 自然科学版*, 1996, 32(4): 452–459.
Li Hongxing. Mathematical neural networks (I) — Interpolation mechanism of mathematical neural network[J]. *Journal of Beijing Normal University: Natural Science*, 1996, 32(4): 452–459.
- [31] 李春光, 廖晓峰, 何松柏. 非线性系统辨识的一种泛函网络方法 [J]. *系统工程与电子技术*, 2001, 23(11): 50–53.
Li Chunguang, Liao Xiaofeng, He Songbai. Functional network method for the identification of nonlinear systems[J]. *Systems Engineering and Electronics*, 2001, 23(11): 50–53.
- [32] 李卫斌, 焦李成. 三类可分离交换性泛函网络模型 [J]. *上海大学学报: 自然科学版*, 2003, 9(4): 347–350.
Li Weibin, Jiao Licheng. Three models of commutable and separable functional network[J]. *Journal of Shanghai University: Natural Science*, 2003, 9(4): 347–350.
- [33] 戴桢杰, 农正, 周永权. 多项式泛函网络运算模型及应用 [J]. *计算机工程与应用*, 2005, 41(21): 49–51.
Dai Zhenjie, Nong Zheng, Zhou Yongquan. Polynomial functional networks computing model and application[J]. *Computer Engineering and Applications*, 2005, 41(21): 49–51.
- [34] 周永权. 泛函网络理论及其学习算法研究 [D]. 西安: 西安电子科技大学, 2006.
Zhou Yongquan. Functional network theory and learning algorithms[D]. Xi'an: Xidian University, 2006.
- [35] 周永权, 赵斌. 泛函网络神经元构造理论与方法 [J]. *计算机科学*, 2008, 35(7): 122–125.
Zhou Yongquan, Zhao Bin. Functional network neurons construct theory and method[J]. *Computer Science*, 2008, 35(7): 122–125.
- [36] 崔强, 武春友, 匡海波. BP-DEMATEL 在空港竞争力影响因素识别中的应用 [J]. *系统工程理论与实践*, 2013, 33(6): 1471–1478.
Cui Qiang, Wu Chunyou, Kuang Haibo. Influencing factors research of air ports competitiveness based BP-DEMATEL model[J]. *Systems Engineering — Theory & Practice*, 2013, 33(6): 1471–1478.
- [37] 罗洪方, 周永权, 谢竹诚. 一种基于进化泛函网络的建模与函数逼近方法 [J]. *计算机科学*, 2010, 37(7): 200–204.
Luo Hongfang, Zhou Yongquan, Xie Zhucheng. Modeling and function approach based on evolutionary functional networks[J]. *Computer Science*, 2010, 37(7): 200–204.
- [38] 周永权, 赵斌. 泛函网络模型及应用研究综述 [J]. *电子科技大学学报*, 2010, 39(6): 803–809.
Zhou Yongquan, Zhao Bin. Progress of functional networks and their applications[J]. *Journal of University of Electronic Science and Technology of China*, 2010, 39(6): 803–809.
- [39] 吕咏梅. 泛函网络新模型及其学习算法研究 [D]. 南宁: 广西民族大学, 2008.
Lü Yongmei. New models and learning algorithms for functional network[D]. Nanning: Guangxi University for Nationalities, 2008.