

基于PSO的协同过滤推荐算法研究

陆 春¹, 洪安邦², 宫 剑¹

LU Chun¹, HONG Anbang², GONG Jian¹

1. 上海财经大学 信息化办公室, 上海 200433

2. 上海财经大学 信息管理与工程学院, 上海 200433

1. Information Office, Shanghai University of Finance and Economics, Shanghai 200433, China

2. School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China

LU Chun, HONG Anbang, GONG Jian. Research on collaborative filtering recommendation method based on PSO algorithm. Computer Engineering and Applications, 2014, 50(5): 101-107.

Abstract: Collaborative filtering is one of the most effective way in the recommended system. The forecast accuracy of recommendation algorithm depends on two key points; the extraction of the nearest neighbors and the calculation of project/user similarity. The paper extracts rated most neighbors with the current project as a nearest neighbor candidate set, and proposes a weighted cosine similarity method to calculate the project/user similarity, then optimizing the weight by the Particle Swarm Optimization(PSO) algorithm. The experimental results show that using these methods can efficiently improve the accuracy of the score predicted, and provide better recommendation results than traditional collaborative filtering algorithms.

Key words: recommended system; Particle Swarm Optimization(PSO); collaborative filtering

摘 要: 协同过滤是推荐系统中最有效的方法之一, 推荐算法评分预测的精确性受到最近邻居的提取以及项目或用户相似度计算的两个关键点的影响。根据用户行为相似性原理, 采用最大交集法提取与当前项目共同评分最多的邻居作为最佳邻居候选集, 同时提出了加权余弦相似性方法对相似度进行计算, 并采用粒子群优化算法(PSO)对权重进行优化求解。实验结果表明, 采用上述方法相对于传统方法来说, 能较好地改善评分预测的精确度, 有效地提高推荐系统的推荐质量。

关键词: 推荐系统; 粒子群算法; 协同过滤

文献标志码: A **中图分类号:** TP393 **doi:** 10.3778/j.issn.1002-8331.1307-0288

1 引言

随着互联网应用的不断发展, Web上的信息呈现爆炸式增长, 如何从这类纷繁冗杂的信息中提取对用户有用的信息, 成为一个非常棘手的问题。

因此, 为用户提供自动推荐的推荐系统应运而生。协同过滤是推荐系统中最有效的方法之一, 已经被广泛应用到许多领域, 甚至有学者将其作为推荐系统的代名词^[1]。协同过滤的基本过程是通过记录和抽取每个用户的个性化信息, 建立相关的用户模型, 然后根据推荐算法建立个性化推荐系统, 主动为用户推荐符合其个性化需求的信息或服务。以看电影为例, 可以根据用户观看

电影的历史信息以及用户本人信息来构建模型, 然后根据特定的算法建立推荐系统, 从而为每个用户推荐符合其自身特点的电影, 解决了这类需求不明确的问题。

协同过滤推荐系统在一定意义上都可归结为一种评分预测过程。在评分预测的过程中, 存在着两个关键问题, 一是最近邻居候选集的提取; 二是用户或项目相似度的计算。

最近邻居候选集的提取, 是直接影响评分预测准确度的一个重要因素。最近邻居候选集提取不当, 一方面会造成得到的邻居都不是当前项目或者用户的邻居, 另一方面提取的邻居之间存在的共同评分很少, 使得相似

基金项目: 上海市科学技术委员会科研计划项目(No.13dz1508402)。

作者简介: 陆春(1980—), 男, 硕士; 洪安邦(1987—), 男, 硕士; 宫剑(1975—), 男, 博士。

收稿日期: 2013-07-22 **修回日期:** 2013-11-11 **文章编号:** 1002-8331(2014)05-0101-07

度的计算不准确,最终都会导致评分预测不准确。其中后者也就是协同过滤的一个共性问题——数据稀疏性问题。本文采用最大交集法来提取邻居,避免了直接使用相似度计算过程中的数据稀疏性问题。

相似度的计算可以采用不同的指标,例如相关相似性、余弦相似性、修正的余弦相似性、欧几里德距离等等。不同的计算方法会直接影响到评分预测的精确度。为了提高预测的精确度,本文提出一种新的相似度计算方法——加权的余弦相似度值,并通过PSO方法来优化计算其中的权重。通过实验表明,该方法有效地改善了预测的精确度。

2 相关工作

协同过滤系统源于20世纪90年代中期^[2],经历了20年的研究与应用,协同过滤系统得到了快速的发展。

Sarwar^[1]等根据协同过滤系统所使用的关联对象不同,将其分为基于用户的协同过滤和基于项目的协同过滤。基于用户的协同过滤假设用户之间存在的相关关系,通过用户之间的相似性来做出推荐;而基于项目的协同过滤假设项目之间的关联,通过这种关联可以进行相关性推荐。

Breese^[3]等则根据协同过滤中算法的不同,将其分为基于模型的协同过滤(Model-based)和基于内存的协同过滤(Memory-based)两种。基于模型的协同过滤主要是系统通过数据挖掘算法或统计方法来学习和训练训练数据集中的相关复杂模型,然后将其用于实际数据的预测中。基于内存的协同过滤首先计算两个用户或者项目的相似度或者权重,而后通过相似度或权重计算当前用户的评分,选出预测评分最高的 n 个项目作为推荐对象。在基于内存的协同过滤中,相似度的计算是一项非常重要的工作,通常计算相似度的指标有皮尔逊相关系数,余弦相似性,修正的余弦相似性^[4-6]等等。

协同过滤推荐中,评分预测通常有两个用途:其一,通过预测评分来确定待推荐对象;其二,通过预测评分来缓解用户-项目评分矩阵的稀疏性^[7-8]。最常用的评分预测是加权平均法^[9],通过找到当前用户或者项目的 n 个最近邻居,按其当前用户的相似度为权重来计算预测评分。加权平均法是协同过滤中最基本的方法,但是其面临的是最近邻居确定过程中数据稀疏性的问题,同时存在着最近邻居选取的问题。

邓爱林,朱扬勇^[7]针对用户评分数据极端稀疏情况下传统相似性度量方法的不足,提出了一种基于项目评分预测的协同过滤推荐算法,根据项目间的相似性初步预测用户对未评分项目的评分,降低了用户-评分矩阵的稀疏性,并在此基础上,采用相似性度量方法计算目

标用户的最近邻居。

张锋,常会友^[8]采用BP神经网络预测用户对项的评分,减小候选最近邻数据集的稀疏性。该算法避免了降维法和智能Agent法的缺点,但使用BP神经网络,训练模型的时间较长,并且得到的模型稳定性也不是很好,这样使得预测评分结果精确性受到一定影响。

一般来说,解决数据的稀疏性问题最常用的方法包括聚类分析和矩阵分解。其中,聚类分析将“最近邻居”的搜索对象限制在最相近的聚类中,或者利用聚类的质心提取推荐结果,后一种方法虽然能提高推荐速度,但降低了推荐质量,并未从根本上解决问题^[10]。矩阵分解法利用矩阵的单值分解原理,对用户-产品矩阵进行分解,从而降低矩阵的维数,抽取主要信息。矩阵分解是一种重要的降维技术,可利用的分解方法较多。Rennie^[11]等人将最大边际矩阵分解(Maximum Margin Matrix Factorization, MMMF)用于协同过滤矩阵降维,Decoste^[12]指出MMMF存在局部极小及噪声影响等潜在问题,他采用集成方法对其进行改进。Wu^[13]将正则化矩阵分解、非负矩阵分解和最大边际矩阵分解三种算法的输出结果进行集成和平均,以减少各算法的不一致。

Choonho Kim^[14]等人提出采用多级关联规则挖掘方法解决数据稀疏性问题。多级关联规则用于计算由于数据稀疏而无法用单级关联规则来计算的预测得分。杨磊^[15]构建了一种最大熵模型,通过把rating信息和tag信息作为模型输入,经过最大熵模型处理后,把当前用户对于目标项目的评分作为模型输出结果,从而形成推荐。

近年来,粒子群算法(PSO)在协同过滤算法中得到了广泛应用,取得了较好的效果。文献[16]提出了一种新的推荐系统,该系统采用粒子群算法学习用户概貌,并为每个用户提供合适的推荐,实验结果证明了模型的有效性。

文献[17-18]将一种差分上下文松弛模型(Differential Context Relaxation model, DCR)用于统计项目与内容之间的链接关系,并用布尔的离散的粒子群算法(BPSO)优化得到适当的松弛因子,取得了较好的效果。

在本文中,在对用户-项目评分矩阵进行评分预测时,利用基于项目的协同过滤和利用最大交集法提取最近邻居,对数据稀疏性问题有所克服。

3 基于PSO算法的协同过滤推荐

为了提高推荐系统的推荐质量,本文基于协同过滤算法的相关原理,提出了对推荐系统进行优化的一种手段。整个评分预测算法大致过程如下:

(1)建立项目配置文件(或者称项目概貌),根据项

目的特征建立包含用户基本信息、项目信息、评分等多个特征值的项目配置文件,项目配置文件是项目特征的综合反映。

(2)最近邻居候选集的提取。采用最大交集法,利用矩阵运算来提取与当前项目有最多共同评分用户的项目作为最近邻居候选集。

(3)最优项目相似度计算。文献[18]中使用PSO算法对差分上下文权值(Differential Context Weighting, DCW)的优化进行了解决。本文的思路是,首先设定项目配置文件中项目属性的每个特征值的权重初值,而后建立当前项目和邻居项目之间的加权余弦相似性函数,并通过PSO优化算法在训练集中对权重值进行优化,当评分预测的误差达到最小时,获得最近权重值,根据最近权重值计算当前项目和邻居项目的最近相似度解。由于项目配置文件包含许多属性,其中有许多数据比较稀疏或不完整,为了克服这些问题,本文使用随机和启发式模型加快和完善用户概貌匹配的质量,通过将粒子群优化算法应用项目配置文件的匹配,将此类问题的解决方法又进行了改善。

(4)当前项目的评分预测。根据最优邻居项目受到的评分和两部项目之间优化过后的这个特征值得到的最近相似度解来预测实验集中现有项目的评分。在预测的过程中,先选取对当前项目没有评分而对邻居项目有评分的用户,根据先前确立的最近邻居候选集,以当前项目配置文件和邻居项目配置文件之间的最近相似度值为权重来预测当前项目的评分。

(5)做出推荐。在对用户-项目评分矩阵进行填补后,选取某一用户预测得分较高的项目对当前用户做出推荐。

3.1 建立项目配置文件

项目的配置文件是进行项目之间距离和相似度计算的基础,当配置文件建立后,需要得到最近邻居候选集,然后才进行项目之间的相似度关系的计算。

项目 j 的配置文件表示为 $profile(j)$, 每个配置文件具有 K 列,分别表示项目配置文件的 K 个特征值。每部项目配置有两个部分:一个可变部分,包括了评分用户的特征,如用户的职业、性别、年龄以及每个用户对此的评分;和一个固定部分,包括了项目 j 的类型划分。定义 $profile(j, i)$ 为 i 用户对 j 项目评分的项目配置文件,如图1所示便是一个标准的项目配置文件样式。

第一项代表了项目评分;第二项代表年龄,为消除量纲的影响,可将其进行标准化,如年龄可按式(1)进行标准化:

$$\frac{age(i) - \min(age)}{\max(age) - \min(age)} \quad (1)$$

1	2	3	4	...	n
Rating	Age	Gender	Occupation		18Genre frequencies
5	23	0	45		10001100100

图1 项目配置文件样式

在本文的研究中,项目配置文件中的定量变量都按上述公式进行了归一化。

3.2 最近邻居候选集的选取与邻居最近相似度的确定

最近邻居的选择是个性化推荐系统中非常重要的一个步骤,协同过滤系统的推荐质量很大程度上取决于算法寻找邻居的效果。本文中,最近邻居的选择包含了三部分的工作,其一,采取了最大交集法选取最近邻居候选集;其二,采用PSO优化算法来确定邻居的最近加权余弦相似度;其三,为了保持推荐的效率和实时性,将项目模块和用户模块的建模进行离线运算,并将相关数据结果进行存贮。

(1)项目配置文件候选集的选择

在确定邻居选择的数据集时,国内外学者大多是对整个数据库进行遍历计算与选取,即对数据库中的所有配置文件与当前项目配置文件之间进行搜索,选取相似度最高的作为最近邻居。遍历选取存在如下问题,其一是数据稀疏性,有些项目与当前项目之间很少有共同评分,从而导致二者相似度无法计算;其二是当项目和用户数据非常巨大时,进行全遍历搜寻是不现实的。本文根据用户行为相似性原理选取,假设相同项目之间受到共同用户评价越多,说明二者在用户行为上可能具有更好的相似度,因此将与当前项目有共同用户评分的邻居按降序排列,选择其中 n 项作为最近邻居选取的候选集,如式(2)所示。其中,代表当前项目 i 和邻居项目 j 之间交集的大小, $R(i, :)$ 、 $R(j, :)$ 分别代表了项目 i 和 j 的评分向量。求得 r 个后,然后对其进行排序,选择其值最大的前 n 个作为邻居候选集集合。

$$S_{i,j} = R(i, :) * R(j, :)' \quad (2)$$

(2)最优加权余弦相似度计算

最近邻居候选集选择完成后,随后进行加权余弦相似度计算。首先,测量现有项目和被选择项目之间的相似度。一般采用的指标有配置文件之间的余弦相似性或修正的余弦相似性以及皮尔逊相关系数。本文利用配置文件之间的加权余弦相似性,在确定最近的权重值时,采用PSO算法在训练集中优化得出。

在使用协同过滤算法时,一般的做法是直接采用用户的评分信息作为衡量项目或者用户之间的相似度。但是在相似度计算中,两个项目或者用户之间表现出相似性并非仅仅用评分就可以衡量,除了项目本身的因

素,还受到用户的个人特征影响。因此在计算相似度时,需要综合考虑用户的信息和项目的信息。同时由于每个因素对评分的影响程度不同,例如对于用户,也许性别的因素比职业更加影响评分,那么性别相对于职业可能权重更高。本文在每个影响特征前加上一个权重(表示特征 f 的权重),并在实验的过程中,通过 PSO 算法来优化这个权重,使其达到最优解。

设 $W(A)$ 表示了当前项目配置文件特征权重值的一个潜在解, A 被表示成为了一个权重的集合。其中表示特征 f 的权重,每个特征集合包含了 K 个权重值,代表了粒子在一个 K 维的空间里的位置。当粒子在这个 K 维的空间中运动时,这些特征值也将不断地进行调整,直到找到最近描述活跃用户偏好的权重集合。本文采用调整的余弦函数来比较项目配置文件之间的相似度或者距离。 $similarity(A, j)$ 表示了当前项目 A 和邻居项目 j 之间的相似度,如式(3)表示:

$$similarity(A, j) = \sum_{i=1}^n \frac{AW \cdot jW}{|AW| \times |jW|} \quad (3)$$

其中, A 代表当前项目配置文件; j 代表项目配置文件选择过程中选择出来的邻居项目配置文件,并且 $A \neq j$; W 是一个 $K \times K$ 对角阵,对角阵上每个值代表了每个特征值的权重, n 代表两部项目之间拥有共同评分用户的总数。

(3)最近项目配置文件收集与存储

当得到 $similarity(A, j)$ 后,选取项目配置文件的最近余弦距离存储于系统之中。由于系统中项目在一段时间内保持稳定,因此本文采用了离线存储余弦距离的方式,当需要对用户进行评分预测时,只要调用这个值,并在一定时间内对这个相似度值进行更新。此外,对于大型的或者时常更新的系统,诸如信息发布、电子商务等场合,由于用户的评分相当频繁,离线存储并调用的方式会对推荐的性能造成一定影响,常常采用在线增量的研究方法。为适应用户兴趣变化,邢春晓等人^[19]提出了基于时间的数据权重和基于资源相似度的数据权重,以使基于项目的协同过滤能够及时反映用户的兴趣变化。其中前者反映用户近期访问资源的重要性以捕捉用户当前兴趣,后者则通过计算用户已访问资源与用户当前兴趣相似度以捕捉有价值的早期访问数据;李聪^[20]提出了一种适应用户兴趣变化的项目相似性增量更新机制,通过计算提交评分后相应项目相似性数据的增量值,以较小计算量实现系统对项目相似性的实时更新,同时可以使得最新的项目相似性数据进入推荐处理,从而使推荐服务能够适应用户兴趣偏好的动态变化,弥补了传统离线计算项目相似性方法难以反映用户兴趣漂

移的不足。

3.3 基于 PSO 算法的权重优化

本文采用了 PSO 算法用于实现当前项目配置文件特征值权重的优化,以寻找到符合项目的特点相似度值。

(1)动态粒子

本文选择传统的 PSO 算法。如上文所述,本文的粒子表示的是各特征值的权重,每个粒子处于 K 维空间中,拥有一个位置和速度。通过其他粒子在空间的移动,同时进行迭代,当粒子所处的空间位置所得的适应函数值达到最优时,便将这个位置确定为群体的全局最优位置。

(2)适应函数

在实验中,假定每部项目特征值权重是稳定的,因此在优化的过程中,当前项目只存在一个权重值的优化解,即只存在一个单一的适应函数。这个适应函数,将其定义为两部项目之间评分预测误差的平均值。这样,优化的目标便是计算出每部当前项目的 W (每个 W 对应一个适应函数)。

适应函数的表达式如式(4)所示,其中 n 为 A 与 j 之间存在的共同评分的用户数目。 A 代表当前项目, j 代表 A 的一个邻居, $mean_A$ 代表 A 受到评分的平均值, $similar(A, j)$ 为 A 与 j 的相似度, $vote(j, i)$ 代表 i 用户对项目 j 的评分, $mean_i$ 代表 i 用户所有评分的均值, $vote(A, i)$ 代表 i 用户对 A 项目的实际评分。

$$fitness = \frac{1}{n} \sum_{i=1}^n \{ (mean_A + k \sum_{j=1}^m similar(A, j) \times (vote(j, i) - mean_i) - vote(A, i)) \} \quad (4)$$

(3)PSO 优化过程

PSO 优化分为三个过程,首先进行粒子速度和位置的初始化,本文实验采用一个随机的过程进行初始化;其次,按照粒子动态移动规则,确立新的粒子位置和粒子速度进行迭代;而后,确立每步局部最优和全局最优的粒子,通过每个粒子的位置,以及这个粒子位置得到的适应函数值,决定是否更新局部最优和全局最优。

3.4 用户评分预测与推荐

在确立了最近邻居项目和当前项目配置文件的最近权重后,可以开始评分的预测。

(1)评分预测。评分预测公式如式(5)所示,其中 $P_vote(A, i)$ 为当前项目的预测得分, $mean_A$ 为所有 A 项目受到的评分的均值, $vote(A, i)$ 为 i 用户对 A 项目评分的配置文件, j 为 i 用户对 j 项目评分的配置文件, $similarity(A, j)$ 为训练集中优化得出的项目配置文件 A, j 之间最近余弦相似度值, $vote(j, i)$ 为 i 用户对 j 项目的评分, $mean_i$ 为 i 用户对所有项目评分的均值, n 为

选取的邻居个数, k 为标准化参数, 余弦距离之和为1。

$$P_vote(A, i) = mean_A + k \sum_{j=1}^n (similarity(A, j) \times (vote(j, i) - mean_i)) \quad (5)$$

由于当前项目的评分在实验集中是已经存在的, 因此可以根据平均绝对误差(Mean Absolute Error, MAE)确定预测结果的准确性, 如式(6), n 代表实验集中当前项目和邻居项目具有共同用户评分的个数, Y_{im} 代表预测得分, S_{im} 代表实际得分。

$$MAE = \frac{1}{n} \sum_{m=1}^n |Y_{im} - S_{im}| \quad (6)$$

(2) 进行推荐。推荐系统进行推荐的依据因算法的不同有所区别, 主要的方法有两种: 其一, 在提取出当前用户兴趣最相似的最近邻居时, 将邻居用户感兴趣的项目推荐给当前用户; 其二, 在存在评分的推荐系统中, 首先提取最近邻居, 然后根据邻居的评分预测当前项目或用户的评分, 然后将当前用户的预测评分进行排名, 提取评分高的几项进行推荐。

4 实验与分析

4.1 实验数据集及预处理

本文实验采用MovieLens的电影评价数据集(<http://www.movielens.umn.edu>)。该数据集包含 943 用户和 1 682 部项目(此处为电影)的细节信息以及这些用户对项目的 10 万次评分。本文使用了这个数据集中的共 22 个特征: 包括了项目评分, 以及年龄, 职业, 性别三个用户特征和 18 个项目类型频率: 动作, 冒险, 动画, 儿童, 喜剧, 犯罪, 纪录片, 戏剧, 幻想, 黑色项目, 恐怖, 音乐, 神秘, 浪漫, 科幻, 惊悚, 战争, 西部。为了使得在寻找邻居项目文件时, 二者有更多的相同评分用户, 实验将少于 50 个用户评分的项目进行剔除, 得到 83 715 条评分数据。并且将年龄按照式(1)进行标准化, 以消除量纲的影响。

4.2 实验设计

在本文实验过程中, 为了验证评分预测的精确性, 将整个数据集随机分成 A、B、C 三个部分, 进行 3 次轮流交叉实验。在每次实验中, 选取两部分的评分项目作为实验的训练集, 用于训练加权余弦相似度的权重和修正的欧几里德距离的权重; 剩余部分评分项目作为实验集以验证优化结果的有效性。

具体来说, 本文采用随机选取的方法从数据集中选取当前项目, 依次以 A、B; A、C; B、C 为训练集, C、B、A 为实验集进行三次交叉实验, 每次实验分成两组如下:

(1) 随机提取 10 部项目作为当前项目(active item),

采用最大交集法提取最近邻居候选集的项目数目为 $n(n=10, 11, \dots, 30)$ 。

(2) 随机提取 10 部项目作为当前项目(active item), 随机提取最近邻居候选集的项目数目为 $n(n=10, 11, \dots, 30)$ 。

通过实验 1 和实验 2 中最近邻居候选集提取的不同, 来验证最大交集法提取最近邻居候选集时, 对评分预测精度有所提高。

通过实验 1 和实验 2 中最近邻居提取的数目不同, 来确定最近的邻居提取规模。

在训练集中, 对 PSO 的参数设置相同, 如下所示:

(1) 粒子数目: $N=40$; (2) 学习因子: $c1=2$; (3) 学习因子: $c2=2$; (4) 惯性权重: $W=0.5$; (5) 最大迭代次数: $M=1\ 000$ 。

4.3 对比实验设计

本文设计了两个对比实验分别验证加权余弦相似度预测评分和利用 PSO 优化的权重预测评分的有效, 由于 Ujjin 在文献[16, 19-20]中验证了在使用加权的欧几里德距离预测得分时优于相关相似性指标, 因此在本文的对比实验设计中, 省略以相关相似性为指标做预测实验。

(1) 修正欧几里德距离预测评分的实验

为了评估加权余弦相似度预测评分的优异性, 设置了利用加权的欧几里德距离预测评分作为对比实验。实验的过程与采用加权余弦相似性实验相同, 仅在测定相似度时采用加权的欧几里德距离如:

$$euclidean(A, j) = \sqrt{\sum_{i=1}^z \sum_{f=1}^{22} w_f \times diff_{i,f}(A, j)^2} \quad (7)$$

其中, A 代表当前项目; j 代表项目配置文件选择过程中选择出来的项目, 并且 $j \neq A$; w_f 代表当前项目特征 f 的权重; z 代表共同评分的用户数目; $diff_{i,f}(A, j)$ 代表了 A 和 j 项目的对于 i 用户之间项目配置文件特征值 f 的差异。

在实验中, 保持当前项目、最近邻居候选集与基准实验相同, 保持 PSO 算法参数的设置一致。在实验预测的过程中, 同样利用训练集优化后的权重值作为参数, 建立计算预测评分的方程(式(3)), 仅将其计算相似度部分由加权余弦改成加权的欧几里德距离。最后是评价指标, 同样利用式(4)中平均绝对误差。

(2) 传统的余弦相似性预测评分

为了评估用 PSO 优化的加权余弦相似度预测评分中 PSO 优化权重的有效性, 设置了利用传统余弦相似性预测评分作为对比实验。实验同样分为两组, 在训练集中, 沿用基准实验的当期项目和最近邻居候选集, 并以余弦相似性指标来测定相似度值, 在实验集中, 建立相关计算方程进行评分预测。

4.4 实验结果及分析

图2分别代表依次以A、B、A、C、B、C为训练集,C、B、A为实验集进行三次交叉实验;三次实验中,每次随机选取10部当前电影,并比较按照最大交集法选取 n 位最近候选邻居和随机选取 n 位最近候选邻居所得的评分预测的结果。在图中,横坐标表示最近邻居候选集规模从5到25,纵坐标表示10部当前项目评分预测的一个综合平均误差值MAE。实验结果表明,最大交集法选取最近邻居时得到的评分预测精确度总体高于随机选取结果,并且在候选集规模分别为14、16、17时,预测精度效果趋于平稳。



图2 最大交集法和随机提取法实验比较结果

图3分别代表了三次交叉实验中利用PSO优化的加权的余弦相似度, PSO优化的加权欧几里德距离和余弦相似度(无加权)预测评分的结果。横坐标代表了10部当前项目,纵坐标代表了10部项目的评分预测,在此利用最大交集法选取最近邻居候选集,候选集规模为16。

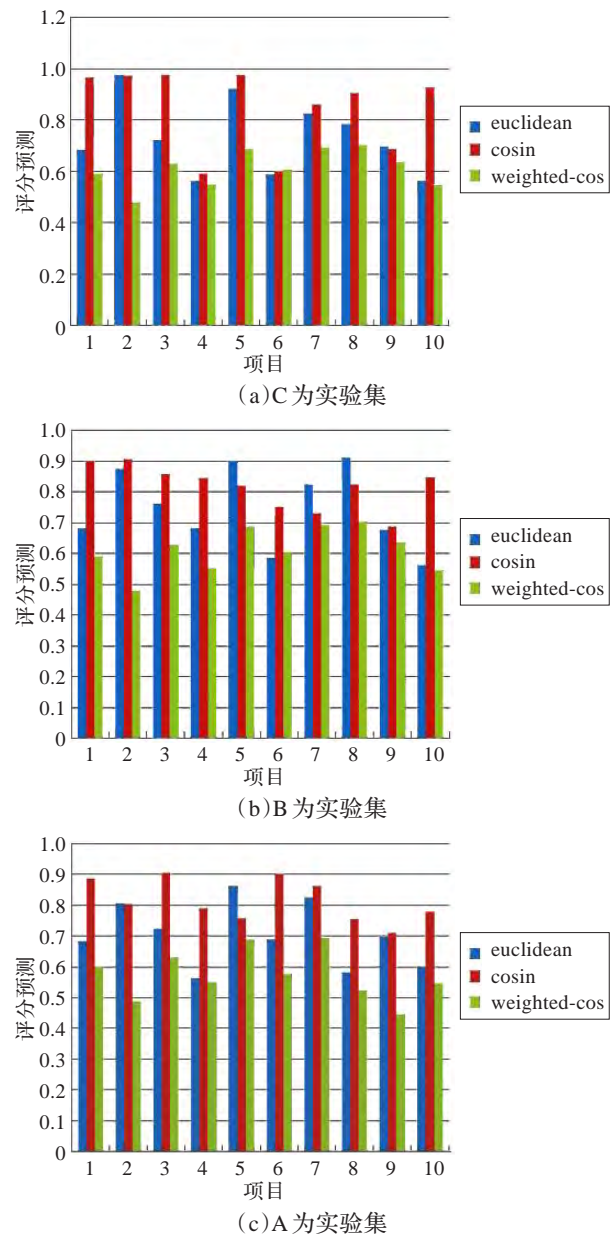


图3 三种相似度指标测算值对比结果

图3的实验结果表明, PSO优化加权的欧几里德距离相对 PSO优化的加权的余弦相似度在预测评分时精确度总体较差。PSO优化的加权余弦相似度预测评分效果优于普通余弦相似度(无加权)预测效果。

综合实验结果表明,利用最大交集法和加权的余弦相似度改进了评分预测的精确度,从而可以改善系统推荐的推荐质量。

参考文献:

- [1] Sarwar B, Karypis G, Riedl J. Item-based collaborative filtering recommendation algorithms[C]//Proc 10th Int'l World Wide Web Conference. New York: ACM Press, 2001: 285-295.
- [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 Transactions on Knowledge and Data Engineering,2005,17(6).
- [3] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998:43-52.
- [4] 项亮.推荐系统实践[M].北京:人民邮电出版社,2012.
- [5] 孙小华.协同过滤系统的稀疏性与冷启动问题研究[D].杭州:浙江大学,2005.
- [6] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques[J].Advances in Artificial Intelligence,2009.
- [7] 邓爱林,朱扬勇,施伯乐.基于项目聚类的协同过滤推荐算法[J].软件学报,2003(9):1621-1628.
- [8] 张锋,常会友.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(4):667-672.
- [9] Vucetic S, Obradovic Z. Collaborative filtering using a regression-based approach[J].Knowledge and Information Systems,2005,7(1):1-22.
- [10] 赵亮,胡乃静.个性化推荐算法设计[J].计算机研究与发展,2002(8):986-991.
- [11] Rennie J D M, Srebro N. Fast maximum margin matrix factorization for collaborative prediction[C]//Proceedings of the 22nd International Conference on Machine Learning. New York:ACM Press,2005:713-719.
- [12] DeCoste D. Collaborative prediction using ensembles of maximum margin matrix factorizations[C]//Proceedings of the 23rd International Conference on Machine Learning. New York:ACM Press,2006:249-256.
- [13] Wu M. Collaborative prediction via ensembles of matrix factorizations[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2007:43-47.
- [14] Kim Choonho, Kim Jnntae. A recommendation algorithm using multi-level association rules[C]//Proceedings of the IEEE/WIC International Conference on Web Intelligence (wT03),2003.
- [15] 杨磊.基于最大熵模型解决协同过滤稀疏性问题研究[D].合肥:中国科学技术大学,2011.
- [16] Ujjin S, Bentley P J. Particle swarm optimization recommender system[C]//Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003), Indianapolis, Indiana,2003:124-131.
- [17] Zheng Y, Burke R, Mobasher B. Optimal feature selection for context-aware recommendation using differential relaxation[J].ACM RecSys,2012,12.
- [18] Zheng Y, Burke R, Mobasher B. Recommendation with differential context weighting[M]//User modeling, adaptation, and personalization. Berlin, Heidelberg: Springer, 2013:152-164.
- [19] 邢春晓,高风荣,战思南,等.适应用户兴趣变化的协同过滤推荐算法[J].计算机研究与发展,2007,44(2):296-301.
- [20] 李聪.电子商务推荐系统中协同过滤瓶颈问题研究[D].合肥:合肥工业大学,2009:83-96.
- [21] 刘玮.电子商务系统中的信息推荐方法研究[J].情报科学,2008(2):300-303.
- [22] 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.

(上接100页)

5 结束语

本文通过在训练模式中加入摄动的方法研究基于爱因斯坦 t -模的模糊联想记忆网络的鲁棒性。从理论上证明了在噪声是正向的情况下该学习算法满足鲁棒性,并用实验验证了该性质;而噪声存在负向影响的情况下,该算法不满足鲁棒性。使用鲁棒性定义更好地研究和评价了基于爱因斯坦 t -模模糊联想记忆学习算法的价值。并且通过对比观察到噪声为正向时训练模式最大摄动与输出模式最大摄动的关系与噪声含负向时的关系有明显不同。在今后的工作中,可以加入对其他网络的研究和对比,以更好、更全面地评价一个模糊联想记忆网络和学习算法。

参考文献:

- [1] Kosko B. Fuzzy associative memories[M]//Fuzzy expert systems reading. MA, USA: Addison Wesley, 1987.
- [2] Kosko B. Bi-directional associative memory[J]. Transactions on System, Man and Cybernetics, 1988, 18(1):49-60.
- [3] Kosko B. Neural networks and fuzzy systems[M]. [S.l.]: Prentice-Hall Inc, 1992.
- [4] 陈松灿,高航.一个改进的指数双向联想内存及性能分析[J].计算机学报,1998,21(增刊):159-162.
- [5] 范周田,钟义信.模糊双向联想记忆网络的收敛性[J].电子学报,2000,28(4):127-130.
- [6] Belohlavek R. Fuzzy logical bidirectional associative memory[J]. Information Science, 2000, 128(1):91-103.
- [7] 程思蔚,徐蔚鸿,王勇,等.基于爱因斯坦 t -模的模糊联想记忆的学习算法[J].计算机工程与应用,2006,42(15):40-44.
- [8] 曾水玲,徐蔚鸿.基于Lukasiewicz t -模的模糊双向联想记忆网络的有效学习算法[J].计算机应用,2006,26(12):2988-2990.
- [9] 曾水玲,杨静宇,徐蔚鸿.模糊双向联想记忆网络的有效学习算法[J].微电子学与计算机,2007,24(12):70-76.
- [10] 曾水玲,杨静宇,徐蔚鸿.基于三角模的模糊双向联想记忆网络的性质研究[J].计算机科学,2009,36(2):238-240.
- [11] 徐蔚鸿,宋鸾姣,李爱华,等.训练模式对的摄动对模糊双向联想记忆网络的影响及其控制[J].计算机学报,2006,29(2):337-344.
- [12] 徐蔚鸿,陈国平,杨静宇,等.规则摄动时模糊蕴涵算子对模糊推理的鲁棒性的影响[J].计算机学报,2005,28(10):1700-1707.