

基于改进鱼群算法与张量分解的社会化标签推荐模型

张 浩^{1,2} 何 杰² 李慧宗³

(淮阴工学院交通工程学院 淮安 223003)¹ (东南大学交通学院 南京 210096)²

(合肥工业大学计算机与信息学院 合肥 230009)³

摘 要 基于大众分类法(folksonomy)的标签应用已逐渐成为一种重要的互联网内容组织方式,但随着数据规模的海量增长,产生了严重的信息过载问题,而传统的基于“用户-项目”二元关系的个性化推荐算法难以有效应对由“用户-项目-标签”所构成的三元关系。通过对基本人工鱼群算法进行改进,提出一种对标签推荐系统初始数据集进行聚类分析的方法,用以降低标签推荐系统的数据分析规模。在此基础上,综合考虑标签推荐系统中的元素权重以及反映用户偏好的评分信息,将元素权重和评分等级进行加权处理,以处理结果作为张量中的元素,建立了一种新的加权张量模型,并利用动态增量更新的张量分解算法进行模型求解,进而完成个性化的推荐。最后在两个真实的实验数据集上对比分析了所提算法(FTA)与另外两个经典标签推荐算法的推荐性能,实验结果表明 FTA 算法在准确率和召回率上均具有较好的表现。

关键词 鱼群算法,聚类分析,张量分解,标签推荐

中图法分类号 C931,TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.030

Social Tagging Recommendation Model Based on Improved Artificial Fish Swarm Algorithm and Tensor Decomposition

ZHANG Hao^{1,2} HE Jie² LI Hui-zong³

(Faculty of Transportation Engineering, Huaiyin Institute of Technology, Huai'an 223003, China)¹

(School of Transportation, Southeast University, Nanjing 210096, China)²

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)³

Abstract Popular classification (Folksonomy) tag application has gradually become an important way of internet content organization, but with the massive increase in the scale of data, the problem of information overload has been produced. On the other hand, the traditional personalized recommendation algorithm based on the relationship between ‘user-item’ is difficult to have effect on the three elements of the “user-item-label”. Based on the improvement of basic artificial fish swarm algorithm, a clustering analysis method was proposed for the initial data set of the tag recommendation system (TRS), which is used to reduce the scale of the data analysis of the TRS. Based on this, through comprehensive consideration of the label recommendation system element weights and the reflection of user preference score information, and by weighted processing of the element weights and grades as the elements in the tensor, a new weighted tensor model was established, and the model was solved by the dynamic incremental updating of the tensor decomposition algorithm, completing the personalized recommendation. Finally, on two real experimental data sets, the proposed algorithm (FTA) and the other two classic tag recommendation algorithms were compared and analyzed. The experimental results show that the FTA algorithm has better performance in the recall rate and precision rate.

Keywords Artificial fish swarm algorithm, Clustering analysis, Tensor decomposition, Tag recommendation

1 引言

传统的个性化推荐系统一般多是利用项目(文档、商品、服务)的属性进行基于内容的推荐,或利用用户间、项目间的相似性进行协同过滤推荐,其主要特征是基于用户-用户、用户-项目、项目-项目间的二元关系,进行数据的过滤、分析与处理,涉及到的相关技术主要有向量空间模型、TF/IDF、K-

means 与机器学习等^[1]。随着 Web2.0 的深入发展,社会化标签逐渐成为一种重要的互联网内容组织方式,由此产生了基于大众分类法(folksonomy)的标签推荐应用,并成为了当前个性化推荐研究领域的热点之一。

社会化标签推荐系统包含用户、项目、标签 3 种主要元素,并将传统的二元关系扩展为三元关系进行分析与建模处理,这为推荐系统提供了更为丰富的实用信息,但同时也出现

到稿日期:2016-06-16 返修日期:2016-08-06 本文受江苏省高校哲学社会科学项目(2014SJB688),国家统计局项目(2014LY058),教育部人文社会科学项目(13YJCZH077)资助。

张 浩(1982—),男,博士生,主要研究方向为计算机应用、交通大数据分析,E-mail:andyhao@seu.edu.cn;何 杰(1973—),男,博士,教授,主要研究方向为交通安全与物流规划;李慧宗(1979—),男,博士,副教授,主要研究方向为社会化推荐系统。

了一个新的问题,即传统的推荐算法难以直接处理这种情况。为此,有学者提出将用户、项目和标签三者间的关系看成一个三部图模型,然后在进行标签系统推荐时将其分解为3个二部图模型,再结合传统的推荐算法(如协同过滤算法)进行标签或项目的推荐^[2]。廖志芳等人试图寻找一种兼顾最大程度保留标签系统三维关系信息与降低高维稀疏数据处理难度的方法,提出了一种基于三部图的三维张量分解推荐算法,并通过具体的应用案例对比验证了所提算法的有效性^[3]。三部图模型的核心思想是把社会化标签系统的三维关系转化为二维关系来进行分析与处理,虽简化了系统的复杂度,但也丢失了标签系统中元素间潜在的间接信息。

另外,标签数据维度的加入使得推荐系统的数据规模与稀疏性的问题也进一步加剧。文献[4]提出将张量分解理论应用于标签推荐系统,并利用张量分解方法进行标签预测,用以降低数据稀疏程度。文献[5,6]则通过对标签系统元素聚类的方法来应对用户随意性标注所带来的数据规模与噪声的问题。除此之外,国内外众多学者还从关联规则挖掘、链接结构分析以及概率模型等角度对社会化标签系统进行了丰富且卓有成效的研究^[7]。本文的创新主要体现在:首先利用一种新型的群智能算法——人工鱼群算法^[8]来对标签系统元素进行聚类分析,以降低社会化标签推荐模型的初始输入数据规模,然后利用增量式的张量分解算法进行标签推荐系统建模,在满足推荐性能的基础上,最大程度地保留标签系统中元素间的潜在关系。相关工作对于提升社会化标签推荐系统的推荐质量以及为用户提供高质量的推荐服务等都具有重要的理论和应用价值。

2 基于改进人工鱼群算法的数据聚类

2.1 人工鱼群算法及其改进

人工鱼群算法是根据鱼群活动行为特征提出的一种随机自治体寻优模式,其基本思想是:在一片水域中,鱼群数目最多的地方一般是该水域中营养物质最丰富的地方,依据这一假设来模拟鱼群的觅食、聚群、追尾等行为,从而实现全局寻优目标^[8]。本文将对基本人工鱼群算法进行改进,用以对标签推荐系统的初始数据集进行聚类分析,降低标签推荐系统的数据分析规模,并将其作为下一阶段利用增量式张量分解算法进行个性化推荐建模的数据输入。关于人工鱼群算法的参数描述与几种典型行为如下(以求极大值为例):

(1)参数描述:设人工鱼总数为 N ,人工鱼个体状态为 $X=(x_1, x_2, \dots, x_n)$,其中, $x_i (i=1, 2, \dots, n)$ 为待寻优变量,人工鱼在当前位置的食物浓度(即适应度函数)为 $Y=f(X)$ (Y 为目标函数值),人工鱼视野感知距离为 $Visual$, δ 为拥挤度因子, Try_number 为试探次数,人工鱼个体 i 和 j 间的距离为 $d_{ij}=|x_i - x_j|$,移动步长为 $Step$, $Rand()$ 产生 $(0, 1)$ 间的随机数。

(2)觅食行为: $Prey()$, 设人工鱼的当前状态为 X_i , 当前的适应度为 Y_i , 在其视野感知距离内随机选择一个状态 X_j , 若 $Y_i < Y_j$, 则向该方向前进一步,更新当前状态。反之,重新选择随机状态 X_j , 判定状态更新条件,若如此反复尝试 Try_number 次后仍不满足条件,则随机移动一步,其觅食规则定义如式(1)所示。

$$X_i^{t+1} = X_i^t + \frac{X_j - X_i^t}{\|X_j - X_i^t\|} \cdot Step \cdot Rand(), Y_i < Y_j \quad (1)$$

(3)聚群行为: $Swarm()$, 设人工鱼的当前状态为 X_i , 搜索当前邻域内 ($d_{ij} < Visual$) 的伙伴数目 n_f 以及中心位置 X_c 。若 $Y_c / n_f > \delta Y_i$, 则表明伙伴中心有较多食物且不太拥挤,向伙伴的中心位置方向移动一步,移动规则如式(2)所示,否则执行觅食行为。

$$X_i^{t+1} = X_i^t + \frac{X_c - X_i^t}{\|X_c - X_i^t\|} \cdot Step \cdot Rand(), Y_c / n_f > \delta Y_i \quad (2)$$

(4)追尾行为: $Follow()$, 设人工鱼的当前状态为 X_i , 搜索当前邻域内 ($d_{ij} < Visual$) 的伙伴中适应度 Y_j 值最大的伙伴 X_{max} , 若 $Y_j / n_f > \delta Y_i$, 则表明伙伴 X_{max} 的当前状态具有较高的食物浓度且周围并不拥挤,向伙伴 X_{max} 的方向前进一步,移动规则如式(3)所示,否则执行觅食行为。

$$X_i^{t+1} = X_i^t + \frac{X_{max} - X_i^t}{\|X_{max} - X_i^t\|} \cdot Step \cdot Rand(), Y_j / n_f > \delta Y_i \quad (3)$$

(5)行为策略与公告板,人工鱼会根据自身的状态和周围的环境进行当前最优行为策略的选择,以便到达食物浓度更高的位置,从而提高寻优的效率。同时算法中设置了一个公告板来记录最优人工鱼的状态及周围食物浓度值,用于每条人工鱼在行动一次后与公告板中的状态进行比较并及时更新相应状态。

人工鱼群算法具有良好的跳出局部极值及实现全局自适应搜索的能力,在参数优化、数据聚类、信号处理等多领域表现出了良好的适应性^[8]。但人工鱼群算法运行后期会存在搜索盲目性较大、收敛速度下降和寻优精度低的问题。为此,借鉴文献[9,10]提出的思想,采用自适应的感知视野和步长来降低算法后期的搜索盲目性及提高收敛速度,并通过行为策略选择和公告板来优化人工鱼的位置更新策略,具体改进方法如下:

(1)对人工鱼视野与步长进行自适应改进。人工鱼的视野和步长直接影响算法的搜索空间的大小与收敛速率,为此可以在算法运行前期选取较大的视野和步长,以增加算法的全局搜索能力并加快收敛速度;在算法运行后期由于比较靠近最优结果,可以减少其视野及步长,使其局部搜索能力得到加强,避免出现求解震荡,提高算法的求解精度,相应的调整规则如式(4)所示。

$$\begin{cases} Visual = Visual \times \exp(-30 \times (\frac{t}{T})^\mu) + Visual_{min} \\ Step = Step \times \exp(-30 \times (\frac{t}{T})) + Step_{min} \\ X_{next} = X + \frac{X_n - X}{\|X_n - X\|} \cdot |1 - \frac{Y}{Y_n}| Step \end{cases} \quad (4)$$

其中, t 为当前迭代次数; T 为最大迭代次数; $\mu \in [1, 10]$; X_{next} 为人工鱼当前状态 X 和搜索的下一状态 X_n 共同作用下的移动规则;为平衡算法的局部搜索和全局搜索矛盾,设定 $Visual$ 和 $Step$ 为3分段函数,初期保持较大值,逐渐变小,最后保持在 $Visual_{min}$ 和 $Step_{min}$ 。

(2)移动策略的改进。人工鱼的随机移动行为会造成搜索的不确定以及解的退化问题,为此可以通过公告板的引入在寻优过程中始终保持向较优解的方向移动,以加快算法收敛的速度并提高精度,防止群体中最优解的退化。同时引入生物竞争机制^[11],采用消亡算子 $\Gamma_d (X \in R^n, \Gamma_m(X)) = rand() * (H - L) + L$,其中 H, L 分别为变量范围的上界,下界,

在算法运行过程中淘汰掉一些目标函数值较小的个体,将其重新初始化,用以保持种群的多样性,增强算法的全局寻优能力。

2.2 基于改进人工鱼群算法的聚类分析

在建立用户-项目评分矩阵以及项目-标签权重矩阵的基础上,将每个用户视为一条人工鱼,利用用户的 Pearson 相关系数^[1]建立人工鱼聚类的目标函数,并通过人工鱼的 4 种行为达到寻优目的,完成用户的聚类分析过程。

通过用户初始聚类并结合用户-项目评分矩阵来确定不同用户簇具有的偏好项目,并进一步可以得到项目所对应的标签聚类,进而完成对社会化标签系统中元素的初始聚类分析,得到密集的初始分析数据集,减少数据冗余和降低数据规模。与传统的 K-means 聚类算法受初始聚类中心以及聚类中心间距最小等参数影响较大相比,人工鱼群算法具有对初值和参数值不敏感、算法简单易于实现和鲁棒性强的特点,具有较好的全局搜索性能及收敛性^[8],其算法实现的步骤如下:

(1)数据预处理,对初始数据集进行数据清洗、数据转化和归一化处理,建立用户的空间向量模型,设用户集合 $U = \{u_i | u_i = (u_{i1}, u_{i2}, \dots, u_{im}), i = 1, 2, \dots, m\}$ 是 m 个维度为 n 的用户数据集, u_{ij} 为用户 i 的第 j 个特征值, u_i 表示人工鱼的个体状态。设置算法的基本参数,如 δ , Try_number 、初始聚类数 k 、随机选择 k 个数值确定初始聚类中心等。

(2)建立目标函数,设目标函数 $f(u)$ 为 u_i 与聚类中心 z_j (食物浓度)的距离接近程度函数,具体定义为:

$$\min f(u) = \sum_{u_i \in z_j} \sum_{j=1}^k (sim(u_i, z_j) \theta + \frac{1}{D(u_i, z_j)} (1 - \theta)) \quad (5)$$

其中, $sim(u_i, z_j)$ 为利用 Pearson 相关系数确定的用户同所属聚类中心的相似度, $D(u_i, z_j)$ 采用经典的欧氏距离定义, θ 为调节系数。

(3)计算人工鱼的初始食物浓度,并对公告板赋初值,同时采用 2.1 节的公告板更新策略来优化人工鱼的位置状态信息。

(4)比较人工鱼当前食物浓度与视域范围内的最优伙伴所在区域的食物浓度,并按照式(4)计算自适应的人工鱼视野与步长,执行觅食、聚群、追尾等行为,并与初始聚类中心状态进行比较,更新公告板状态或生成当前最优聚类中心。

(5)判断是否到达最大迭代次数或满足终止条件,如符合则输出最优解,否则转入步骤(1)直到满足条件。

3 张量建模与推荐算法

3.1 张量与张量分解

在代数几何中,常以一个定义在向量空间和对偶空间的笛卡尔积上的多重线性函数系数来定义张量。一般意义上,零阶张量是标量,向量类似一阶张量,矩阵类似二阶张量,三阶张量类似立方体矩阵。在标签系统推荐中,可以将用户、项目、标签三元组构成一个三阶张量,借以描述标签推荐系统中的元素特征和三元关系特征。

张量分解是把高阶张量进行降阶处理的一种手段,其目的是得到一个在规模上小于原张量的稠密化近似张量,并最大化地保留原张量的信息,进而提供个性化的推荐服务。常用的张量分解方法有 CP 分解与 Tucker 分解^[12]。CP 分解是把张量分解为有限个秩一张量,是传统矩阵分解的推广; Tucker 分解是把张量近似于一个核张量 G 与因子矩阵的 mode- n 乘积,对于一个由用户、项目和标签构成的三阶张量 $Y \in R^{I \times J \times K}$ 的 Tucker 分解模型,其形式化定义如下:

$$Y \approx \hat{Y} = G \times_1 U \times_2 V \times_3 W \quad (6)$$

$$r_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R u_{ip} u_{jq} w_{kr} g_{pqr}$$

其中, $G \in R^{P \times Q \times R}$ 是核心张量, $U \in R^{P \times I}$, $V \in R^{Q \times J}$, $W \in R^{R \times K}$ 是 3 个维度上的特征矩阵,张量 $\hat{Y}$ 的秩为 (P, Q, R) , $\times_i$ 为 mode- i 张量-矩阵乘法, r_{ijk} 为张量 $\hat{Y}$ 的第 (i, j, k) 个元素, Tucker 分解的原理如图 1 所示,更详细的张量分解算法步骤可以参见文献^[13],在此不再赘述。

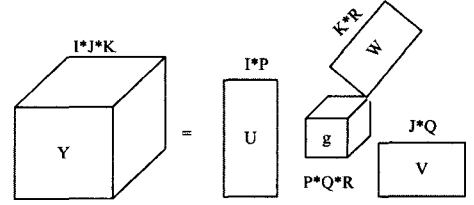


图1 Tucker分解原理

3.2 一种动态增量更新的张量分解方法

Tucker 分解模型在标签推荐系统中的应用体现出了良好的推荐性能,但对于目前大多数的网络应用系统,新用户和新项目的不断加入使得以传统张量分解算法为基础的推荐系统需要重新对张量进行反复求解运算,极大地增加了计算资源的消耗。为解决这个问题,可以采用一种动态增量更新的张量分解方法^[14],在原始张量分解的基础上结合新增加的用户、项目构成新的张量,分部进行张量分解运算,并对原始张量分解结果进行动态更新,以降低算法的复杂度,提升算法的适应性。具体思路为:设当前张量为 Y ,新增张量为 Y' ,则 $Y^* = (Y + Y')$,对于老用户可以使用原来的分解结果进行推荐,对于新用户只需在原有模型基础上更新张量分解模型部分参数,获取 Y' 的预测值,进而完成推荐过程,而不需要对整个张量 Y^* 进行重复分解计算。具体求解过程如下:设现有数据为 $X^{old} (X_1, X_2, \dots, X_{N_k})$,新用户数据为 $X^{new} (X_{N_k+1}, X_{N_k+2}, \dots, X_{N_k+l})$, W 从 $N_k \times R$ 变成 $(N_k + l) \times R$, l 为新增用户数,将 W 分解为 $W = (W^{old}, W^{new})$,其中 W^{old} 保持不变,此时只需求解 W^{new} 即可,相应的目标函数为: $\min J = \|X^{new} - G \times_1 U \times_2 V \times_3 W^{new}\|$,保持 (G, U, V) 不变, W^{new} 的最优解为: $(X^{new})^T A (A^T A)^{-1}$,结合上式可得 X^{new} 的最优解 Y^{new} ,即 $Y^{new} = \sum_{k=1}^l \sum_{r=1}^R U G^r V^T (W^{new})$ 。

3.3 张量建模与推荐算法

在社会化标签推荐算法中,依据“用户-项目-标签”三元关系建立张量模型进而进行张量分解寻优,其中张量的表示和建模方法主要有“0/1”和“+/-”正反例两种表示形式^[15],这种做法虽然易于理解和处理,但忽视了标签系统中元素的重要性。为此基于“用户-项目-标签”的三元组关系,提出综合考虑标签推荐系统中的元素权重以及反映用户偏好的评分信息,将元素权重和评分属性进行加权处理后作为张量中的元素以建立一种新的加权张量模型,并利用 3.2 节中提出的基于增量更新的张量分解算法进行模型求解,进而完成个性化的推荐采用第 2 节所提出的基于改进鱼群算法的数据聚类方法来完成数据预处理。

以为用户进行标签推荐为例,对于一个给定的用户 (i) 、项目 (j) 、标签 (k) 三元组关系,抽取出其标签的属性,采用欧氏距离的方法定义其与对应项目的标签聚类中心的距离为 $D(t_k, t_c)$,将 $w_{ijk} = 1/D(t_k, t_c)$ 定义为三元组中标签的权重值,

同时定义 R_{ij} 为用户 i 对项目 j 的评分(0~5),则得到一个三元组(用户-项目-标签)的权值 $W_{ijk} = R_{ij} * w_{ijk}$,然后利用三元组权值 W_{ijk} 作为张量元素建立张量模型,并采用 3.2 节提出的增量更新的张量分解算法进行模型求解,最后依据权值的大小为某一用户在项目上产生最优 TOP-N 的标签推荐列表集,在用户推荐和项目推荐应用中亦可以采用类似的方法。

4 实验设计与结果分析

本节选取两种不同的测试数据集作为实验分析样本,对比分析本文所提算法(FTA)、经典标签推荐算法(FolkRank)和张量分解算法(Tensor)等在推荐性能上的表现,推荐质量评价指标采用常用的准确率和召回率,用以验证所提算法的有效性。

4.1 测试数据集选取

实验数据源主要下载自国外公开的用于测试推荐算法性能的真实数据集,包括 MovieLens([www. movielens. org](http://www.movielens.org))和 Last. fm(www. last. fm),具体的样本规模及特征如表 1 所列。

表 1 实验样本集

Data Sets	Users	Items	Tags	Ratings
MovieLens	525	834	2804	569
Last. fm	756	728	3183	405

4.2 算法评价指标

准确率-召回率(Precision-Recall):准确率反映系统推荐的标签集中用户实际上偏好的标签数所占的比例,召回率描述有多少比例的用户偏好标签包含在最终的推荐列表中。本文对其进行了一些调整,具体定义如下:假设 $P(u)$ 为待预测的测试用户的偏好标签, $T(u)$ 为系统为用户提供的 TOP-N 推荐列表, $R(u)$ 为系统交互中用户的实际偏好的标签,则 $Precision = \frac{T(u) \cap P(u)}{N}$, $Recall = \frac{T(u) \cap R(u)}{P(u) \cap R(u)}$ 。

4.3 实验结果与分析

考虑到性能比对的一致性,选取 TOP-N=[6,7,8,9,10],分别对上述的 5 种情形进行 10 次重复实验,然后对其最优值进行平均化处理后作为最终的评估结果进行统计,目的是尽量降低数据分布与参数调整对算法性能的影响,实验结果如图 2 和图 3 所示。

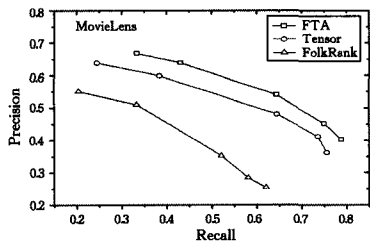


图 2 MovieLens 数据集上不同算法的对比

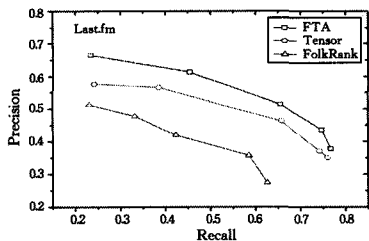


图 3 Last. fm 数据集上不同算法的对比

实验结果表明,本文提出的 FTA 算法对两种不同的数据集均表现出了较好的适应性,推荐质量也好于其它两种经典的算法。另外,从实验结果的对比可以看出:1)随着 TOP-N 的增大,Precision 和 Recall 呈现出矛盾的关系,即 N 较小时,预测结果的 Precision 较大,但 Recall 则较低;而随着 N 的增大,预测结果的 Precision 开始降低,但 Recall 更高,这主要是由评价指标 Precision-Recall 的性质所决定的,但具体到每种不同的算法,其在 Precision-Recall 指标上均呈现出各异的结果,因此 Precision-Recall 仍不失为一种衡量算法推荐质量的评价标准。2)传统的 Tensor 算法作为一种经典的张量分解算法,在一般情况下能够表现出良好的性能,但在处理具有较大稀疏性的数据集时往往不能很好地控制其复杂度,但推荐质量仍优于 FolkRank 算法。3)FTA 算法通过对初始数据集进行预处理可以有效地降低数据分析的规模,同时采用增量更新的张量分解策略一方面可以缩短算法的处理时间,另一方面也能够较好地保留三元组元的隐含关系特征,在推荐质量上相对于其它两种算法表现更好。

结束语 社会化标签系统的广泛应用为用户提供了一个更为广阔的信息分享和开放平台,同时也带来了信息过载的难题。传统的推荐系统多是基于用户-用户、用户-项目、项目-项目间的二元关系来为用户提供个性化的推荐服务,但难以有效应对社会化标签系统中的用户-项目-标签的三元关系的处理。本文通过引入一种改进的人工鱼群算法并结合张量建模的手段建立了一种新型的社会化标签推荐模型,首先通过人工鱼群算法对标签系统中的元素进行聚类分析,可以有效降低社会化标签推荐模型的初始输入数据规模;然后利用一种增量式的张量分解算法进行标签推荐系统建模,在满足推荐性能的基础上,最大程度地保留标签系统中元素间的潜在关系;最后在两个真实的实验数据集上对比分析了本文所提算法与另外两种经典标签推荐算法的推荐性能,实验结果表明本文所提算法在准确率和召回率上均具有较好的表现。但在研究过程中也存在一些问题,如可以再进一步地降低算法的复杂度,在算法评价方面还需要尝试新颖性、多样性等一些其它的评估指标,这些问题将是下一步研究的方向。

参考文献

- [1] Lü Lin-yuan, Medo M, Yeung C H, et al. Recommender systems [DB/OL]. [2012-02-06]. <http://arxiv.org/abs/1202.1112>
- [2] Chen Chao, Zhang Ying-chao, Miao Jin. A collaborative filtering recommender algorithm based on tripartite network[J]. Journal of Nanjing University of Information Science and Technology (Natural Science Edition), 2010, 2(4): 337-339(in Chinese)
陈超,张颖超,缪进.一种基于三部图网络的协同过滤算法[J].南京信息工程大学学报(自然科学版),2010,2(4):337-339
- [3] Liao Zhi-Fang, Li Ling, Liu Li-Min, et al. A Tripartite Decomposition of Tensor for Social Tagging[J]. Chinese Journal of Computers, 2012, 35(12): 2625-2632(in Chinese)
廖志芳,李玲,刘丽敏,等.三部图张量分解标签推荐算法[J].计算机学报,2012,35(12):2625-2632
- [4] Symeonidis P, Nanopoulos A, Manolopoulos Y. A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis[J]. IEEE Transactions on Knowledge and Data Engineer, 2010, 22(2): 179-192
- [5] Sun Ling-fang, Li Shuo-peng. Social tagging recommendation system based on K-means cluster and tensor decomposition [J].

孙玲芳,李烁朋. 基于 K-means 聚类与张量分解的社会化标签推荐系统研究[J]. 江苏科技大学学报(自然科学版),2012,26(6):597-601

- [6] Wang Long, Wang Jia-lun, Cheng Zhuan-li, et al. Personalized Medicine Recommendation Based on Tensor Decomposition[J]. Computer Science, 2015, 42(5): 225-229(in Chinese)
王龙,王嘉伦,程转丽,等. 基于张量分解的药品个性化推荐[J]. 计算机科学, 2015, 42(5): 225-229
- [7] Fouzia J, Shah K, Amna M, et al. Semantics discovery in social tagging systems: A review[J]. Multimedia Tools and Applications, 2016, 75(1): 573-605
- [8] 江铭炎,袁东风. 人工鱼群算法及其应用[M]. 科学出版社,2012
- [9] Peng Yong, Tang Guo-lei, Xue Zhi-shun. Optimal operation of cascade reservoirs based on improved artificial fish swarm algorithm[J]. Systems Engineering-Theory & Practice, 2011, 36(6): 1118-1125(in Chinese)
彭勇,唐国磊,薛志春. 基于改进人工鱼群算法的梯级水库群优化调度[J]. 系统工程理论与实践, 2011, 36(6): 1118-1125
- [10] Liao Yu-lei, Liu Peng, Wang Jian, et al. Control parameter optimization for the unmanned surface vehicle with the improved artificial fish swarm algorithm[J]. Journal of Harbin Engineering University, 2014, 35(7): 800-806(in Chinese)

廖煜雷,刘鹏,王建,等. 基于改进人工鱼群算法的无人艇控制参数优化[J]. 哈尔滨工程大学学报,2014,35(7):800-806

- [11] Chen Guang-zhou, Wang Jia-quan, Li Chuan-jun, et al. An improved artificial fish swarm algorithm and its applications [J]. Systems Engineering, 2009, 27(12): 105-110(in Chinese)
陈广洲,汪家权,李传军,等. 一种改进的人工鱼群算法及其应用[J]. 系统工程, 2009, 27(12): 105-110
- [12] Weng S S, Lin B S, Chen W T. Using contextual information and multidimensional approach for recommendation [J]. Expert Systems with Applications, 2007, 36(2): 1268-1279
- [13] Li Gui, Wang Shuang, Li Zheng-yu, et al. Personalized Tag Recommendation Algorithm Based on Tensor Decomposition[J]. Computer Science, 2015, 42(2): 267-273(in Chinese)
李贵,王爽,李征宇,等. 基于张量分解的个性化标签推荐算法[J]. 计算机科学, 2015, 42(2): 267-273
- [14] Zou B Y, Li C P, Tan L W, et al. Social recommendations based on user trust and tensor factorization[J]. Journal of Software, 2014, 25(12): 2852-2864(in Chinese)
邹本友,李翠平,谭力文,等. 基于用户信任和张量分解的社会网络推荐[J]. 软件学报, 2014, 25(12): 2852-2864
- [15] Milicevic A K, Nanopoulos A, Mirjana Ivanovic. Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions[J]. Artificial Intelligence Review, 2010, 33(1): 187-209

(上接第 138 页)

使训练样本在高维特征空间中与原点的时间隔尽可能大,同时使目标样本在上近似超平面尽可能紧密,充分利用训练样本的类内结构增强分类性能。

在实验中发现,不同目标函数的形式以及参数 v, p, σ, β 的取值对算法性能有较大影响。因此,如何设计合理的目标函数和选取算法中的参数将是未来的研究重点。

参考文献

- [1] Tax D M J, Duin R P W. Support Vector Data Description [J]. Machine Learning, 2004, 54(1): 45-66
- [2] Campbell C, Bennett P. A Linear Programming Approach to Novelty Detection [M]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2001
- [3] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443-1471
- [4] Lin C F, Wang S D. Fuzzy Support Vector Machine [J]. IEEE Transactions on Neural Networks, 2002, 13(2): 464-471
- [5] Xu Yi-tian, Liu Chun-mei. A rough margin-based one class support vector machine[J]. Neural Comput & Applic, 2013(22): 1077-1084
- [6] Bishop C M. Pattern Recognition and Machine Learning [M]. Cambridge: Springer, 2007: 291-320
- [7] Jun G, Chung F, Wang S. Matrix pattern based minimum within-class scatter support vector machines[J]. Applied Soft Computing, 2011, 11(8): 5602-5610
- [8] Zhang J, Wang K, Zhu W, et al. Least Squares Fuzzy One-class Support Vector Machine for Imbalanced Data[J]. International Journal of Signal Processing, Image Processing and Pattern Re-

cognition, 2015, 8(8): 299-308

- [9] Jeong Y S, Kang I H, Jeong M K, et al. A New Feature Selection Method for One-Class Classification Problems [J]. IEEE Transactions on Systems, Man, and Cybernetics—part c: Applications and Reviews, 2012, 42(6): 1500-1509
- [10] Zhu W, Zhong P. A new one-class SVM based on hidden information[J]. Knowledge-Based Systems, 2014, 60(2): 35-43
- [11] Asharaf S, Shevade S K, Narasimha murty M. Rough support vector clustering[J]. Pattern Recognition, 2005, 38(10): 1779-1783
- [12] Xiao Y, Wang H, Xu W. Parameter Selection of Gaussian Kernel for One-Class SVM[J]. IEEE transactions on cybernetics, 2015, 45(5): 927-939
- [13] An W, Liang M. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises [J]. Neurocomputing, 2013, 110(6): 101-110
- [14] Wang L, Jia H D, Li J. Training Robust Support Vector Machine with Smooth Ramp Loss in the Primal [J]. Neurocomputing, 2008, 71: 3020-3025
- [15] Junejo I N, Bhutta A A, Foroosh H. Single-class SVM for dynamic scene modeling [J]. SIVIP, 2013, 7(1): 45-52
- [16] Mygdalis V, Iosifidis A, Tefas A, et al. Exploiting subclass information in one-class support vector machine for video summarization[C]// 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015: 2259-2263
- [17] Wang Lei, Yang Yi-fan, Zhou Qi-hai. Rough Set based One-class Support Vector Machine[J]. Computer Science, 2009, 36(9): 242-245(in Chinese)
王磊,杨一帆,周启海. 粗糙 one-class 支持向量机[J]. 计算机科学, 2009, 36(9): 242-245
- [18] <http://archive.ics.uci.edu/ml/datasets.html>