

基于蝙蝠优化聚类控制的网络数据协同过滤推荐

杨玲玲

(河南工业贸易职业学院,信息工程学院,河南省郑州市,450064)

摘要 为了进一步提高网络数据特征提取能力,设计了一种基于蝙蝠优化聚类控制的网络数据协同过滤推荐方法。以协同过滤推荐算法为基础并型建立细粒度偏好模型,利用蝙蝠优化算法达到聚类效果。研究表明:在加权系数处于[0.1,0.5]范围内时,平均绝对误差都会随着加权系数的增加而发生降低,符合用户实际评价要求,使推荐结果得到明显改善。协同过滤推荐方法在提高预测精度下,能够减小邻近邻域搜索区间,提高结果实时性。

关键词 网络数据;蝙蝠优化;偏好矩阵;序列化数据

中图分类号:G354 文献标识码:B

文章编号:1008-0899(2025)08-0031-02

近年来,随着互联网技术的快速进步,针对大数据开展的研究日益受到了重视。内容推荐与协作过滤两种方法具有显著区别,其中内容推荐不需要对用户评论行为进行分析,而是采用目标特征参数组成推荐指标^[1]。还有一些学者^[2]将LDA模式和Word2Vec模式进行联合对以前所关注的音乐进行研究。根据以上条件,文献^[3]提出了一种基于深度学习与矩阵分解的混合建模方法,使之能够在稀疏数据环境下更好地挖掘出用户与商品的相关性参数。而文献^[4]采用渐进式更新协作滤波法与语义解析法进行融合后构造出一种混合型推荐算法。欧朝荣等^[5]提出融合显式和隐式反馈的协同过滤推荐模型,在覆盖不同领域数据集上实验,有效地融合显式和隐式反馈。

根据以上分析,本文提出了一种基于用户选择特点的协作过滤推荐方法,并通过实验验证该方法的有效性。

1 算法模型

1.1 用户类型偏好矩阵

在构建参数模型时,设定了指标类别特性,能够精确反映使用者喜好,可以精准地为使用者提供信息支持。本研究采用“使用者-项目评价”和“项

目-类别”相组合的方式建立了基于参数分类方法,能够对各类商品的多种类别喜好程度准确预测,由此建立精细化的“使用者-项目”类别选择矩阵。

而在个性化参数推荐方面,每个条目都包含了不同的特性,因此条目种类要少于条目数目。每个使用者对于各类商品的喜好存在一定的差异。在进行评估和计算的时候,均采用实际评估参数进行分析。

得分比率表示特定使用者u对e类总得分 $R_{u,i}$ 相对该指标总得分 R_u 的比例,建立以下计算式:

$$P_{u,e} = RP_{u,e} \times IIP_{i,e} = \frac{\sum_{i \in I_e} R_{u,i}}{\sum R_u} \times \ln \frac{\text{card}(I)}{1 + \text{card}(I_e)} \quad (1)$$

反向项目比例(IIP)为e类项目数量 $\text{card}(I_e)$ 相对总项目数 $\text{card}(I)$ 的比例,可以有效防止e类热门项目所导致的用户偏好变化。为防止出现e类项目个数等于0的情况,本研究将 $1 + \text{card}(I_e)$ 作为算式分母。通过计算式(1)求解不同类别使用者对各类项目的喜好性,并构造用户-项目种类偏好矩阵。

1.2 蝙蝠优化聚类控制流程

选择蝙蝠优化用户模糊聚类算法开展分析时,先利用蝙蝠寻优算法来获取最佳初始聚类中心,再对用户进行模糊C均值聚类处理,对每个蝙蝠参数进行处理的过程中,把每个蝙蝠都视为簇中心矩阵C。

第1步:以适应度函数建立输入算式,细粒度使用者-项目种类偏好矩阵,蝙蝠种群数量规模h、簇数、迭代上限T。先通过随机方法构建初始变幅群体C,再对群体中个体 C_i 速度值、位置 x_i 、脉冲辐射速

作者简介:杨玲玲(1985~),女,河南焦作温县人,硕士,讲师,研究方向:电子信息工程。

率 r_i 进行初始化,同时完成群体的音量 A_i 和频率 f_i 初始化。

第2步:进行隶属度计算。

第3步:对群体中各蝙蝠个体的适合度进行统计,并对其进行排序,最后确定最佳适合度的蝙蝠个体。

第4步:调整单个蝙蝠的定位与速度。

第5步:设置随机数 $rand1$,遍历整个群体的蝙蝠族群体,在符合 $rand1 < r_i$ 的条件下,产生新的 x_{new} ,并求解获得适应度 $fit(x_{new})$ 。

第6步:产生随机数 $rand2$,再对群体中所有蝙蝠样本进行遍历,并对响度参数和脉冲辐射速率进行调节。

第7步:如果迭代的数目比 T 更小,或者没有达到簇中心收敛性要求,那么就返回步骤2继续迭代。相反,在最佳适应度下,以此组成最佳聚类中心,对聚类中心进行重新划分。

将基于协作式推荐算法的引入用户偏好的最优聚类计算中,构建了基于分类特征的个性化分类模型,采用Batman算法进行分类,最终依据用户权重设置预测分值。

2 结果与分析

2.1 数据集

本实验选择的测试参数是由美国 GroupLens 团队构建的,从MovieLens网站上获取,测试前对这些数据进行了预处理。将ML-100K中的样本按照5:1的数量比进行随机分类,由此获得训练样本和测试样本。再对5次试验后得到的平均数据进行五重交叉校验,以此得到最后的评价指标。

2.2 参数结果分析

为获取本文建立的CF-UPC算法最佳加权系数,再通过上述优化方式得到10个簇和40个近邻;在没有设置加权因子的时候,得到图1的试验体系平均绝对误差分配参数。

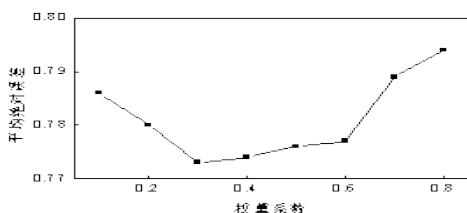


图1 不用权重系数下系统平均绝对误差分布
通过对图1数据进行分析可知,在加权系数处

于 $[0.1, 0.5]$ 范围内时,无论簇数目多大,平均绝对误差会随着加权系数的增加而发生降低,这说明设定了用户项偏爱程度的情况下使用CF-UPC算法进行评价时,可以使其符合用户实际评价要求,从而使推荐结果得到明显改善。当权重系数介于 $0.5 \sim 0.9$ 时,逐渐提高权重系数时,户的项目类型选择相似度已达到很高的比例。因此会对推荐效果产生不利影响。

2.3 实时性的验证

将其设置成40个近邻,通过该节点的占有率 R_2 求出最近用户重复次数 R_1 与检索速率 SR ,不同 R_2 , R_1 、 SR 分配参数见表1。当 $R_2=40\%$ 时,其最近邻用户的搜寻速率 SR 最小等于1,并且当 R_2 等于40%时达到了最大。通过研究可知,与UBCF相比,本文提出的CF-UPC方法在提高得分预测精度的情况下,能够减小邻近邻域的搜索区间,从而提高了计算结果的实时性。

表1 不同 R_2 下 R_1 和 SR 分布

R_2	0.2	0.4	0.6	0.8	1.0
R_1	0.21	0.46	0.81	0.93	0.99
SR	2.46	2.83	2.79	1.55	1.23

3 结论

本文开展基于蝙蝠优化聚类控制的网络数据协同过滤推荐分析。在加权系数处于 $[0.1, 0.5]$ 范围内时,平均绝对误差会随着加权系数的增加而发生降低,符合用户实际评价要求。协同过滤推荐方法在提高预测精度下,能够减小邻近邻域搜索区间,提高结果实时性。

参考文献

- [1] 王兴茂,张兴明,吴毅涛,等.基于启发式聚类模型和类别相似度的协同过滤推荐算法[J].电子学报,2016,44(7):1708-1713.
- [2] 周航帆,周莲英.基于会话记录的Word2Vec音乐推荐算法研究[J].通信技术,2019,52(04):850-857.
- [3] 胡思才,孙界平,琚生根,等.基于深度神经网络和概率矩阵分解的混合推荐算法[J].四川大学学报(自然科学版),2019,56(06):1033-1041.
- [4] 刘辉,万程峰,吴晓浩.基于增量协同过滤和潜在语义分析的混合推荐算法[J].计算机工程与科学,2019,41(11):2033-2039.
- [5] 欧朝荣,胡军.融合显隐式反馈的协同过滤推荐模型[J].控制与决策,2024,39(03):1048-1056.