

# 神经网络推荐系统协同过滤算法研究

龙国虎, 朵琳

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650500)

**摘 要:** 针对传统协同过滤算法中的冷启动和推荐准确度问题, 提出一种在神经网络基础上的算法。利用卷积神经网络对用户以前的历史行为和评价项目属性建立卷积神经网络模型, 然后, 提取用户和其项目属性的特征, 通过全连接的方式对评分进行拟合, 其次, 运用关联和推理思想, 利用关联矩阵计算方法, 结合用户使用信息的评测和系统内的关联项目进行对接, 并自动生成关联矩阵所对应的信息项目集。最后, 算出实际用户与其他用户之间的评分信息相似性大小, 依据评分的相似性大小。依照最相似邻居的评分信息算出实际用户没有评过的项目预测大小值, 接着根据预测值排名在前面项目对用户进行推荐, 此推荐算法通过在数据集 MovieLens 采用对比试验进行验证, 得出该研究的算法解决了冷启动问题并同时提高了推荐的准确率。

**关键词:** 神经网络; 推荐系统; 协同过滤; 推荐算法; 关联规则; 遗忘函数; 冷启动; 推荐准确率

中图分类号: TP393

文献标识码: A

文章编号: 1006-7973 (2024) 03-0040-03

随着人工智能和大数据时代的来临, 各类推荐系统如雨后春笋一般的涌现出来, 如当前大家最熟悉的电影领域、音乐领域、电商领域、社交领域和图书领域的都有各自成熟的推荐算法。传统的推荐过程中, 以用户的浏览频次进行由高到低的排列, 以排名靠前的信息优先推荐给客户, 在基于神经网的协同过滤算法推荐系统的计算过程中, 能够通过仿真实验和举证建模等算法模型, 有效弥补传统协同过滤算法过程中的种种缺陷, 通过实践检验和用户的实际使用情况, 都表现出基于神经网络推荐系统的优先计算方式和人工智能推荐模式, 能快速准确的实现用户所需信息的推进过程, 深受当前广大用户的欢迎和热烈追捧。

## 一、相关技术理论

### 1. 卷积神经网络模型

卷积神经网络模型从用户的相关信息和电影的相关参数为着眼点, 在嵌入层中逐渐进入到连接层, 进行电影向量的有效拼接, 针对用户的浏览量, 点击数关键词搜索和相关影视的评论打分等相关操作, 智能匹配用户的兴趣和爱好, 实现信息库中电影类型与用户个人爱好的及时对接, 最后在输出层中按照用户的兴趣和喜欢程度进行由高到低的排序。卷积神经网络模型框架如图 1 所示。

### 2. 推荐系统中的算法

#### (1) 余弦相似性 (cosine)

余弦相似性度量模式是通过计算分布向量之间的余弦角来度量客户之间的距离, 即相似度。客户-项目得分分布矩阵可以看作是  $n$  维分布空间中的分布向量, 对于未得分的投资项目, 将得分值调制为 0, 假设分布向量  $u$  和  $v$  依次表示客

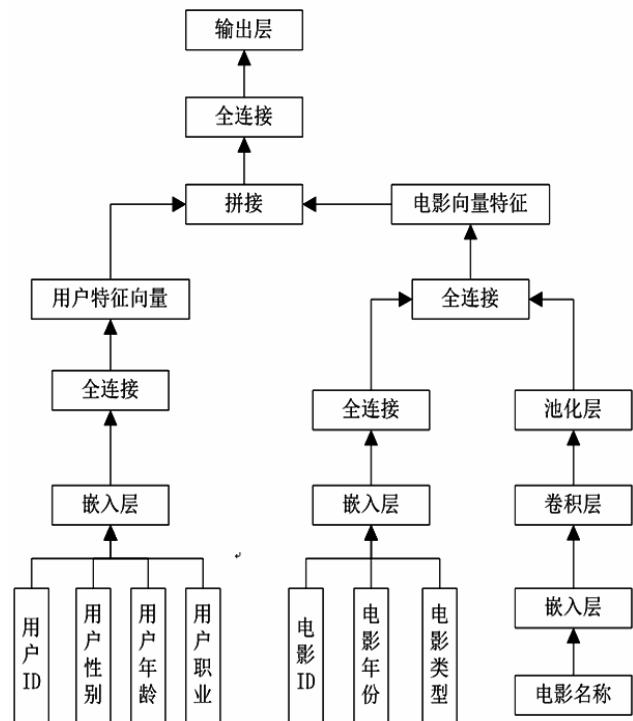


图 1 卷积神经网络模型框架

户  $u$  和客户  $v$  在  $n$  维分布空间中的得分, 就利用根据协同过滤的电子网络商务市场交易差异化举荐算法分析研究客户  $u$  和客户  $v$  相互之间的相似作用程度是:

$$\text{sim}(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| * |\vec{v}|} \quad (1)$$

收稿日期: 2022-10-20

作者简介: 龙国虎 (1996-), 男, 昆明理工大学信息工程与自动化学院, 硕士生, 主要研究方向为智能信息处理。

通讯作者: 朵琳 (1974-), 女, 昆明理工大学信息工程与自动化学院, 副教授, 主要研究方向为智能信息处理, 下一代网络。

基金项目: 国家自然科学基金地区基金项目 (61962032); 省科技厅优秀青年项目 202001AW07000。

## (2) 修正的余弦相似性 (Adjusted cosine)

令  $P_{uv}$  表示反映顾客  $u$  和顾客  $v$  的得分的项目集,  $P_u$  和  $P_v$  依次表示反映顾客  $u$  和顾客  $v$  的得分的项目集,  $R_{ua}$ ,  $R_{va}$  依次表示顾客  $u$  和顾客  $v$  对项目  $a$  的得分, 并且可以依次表示反映顾客  $u$  和顾客  $v$  的平均得分, 运算方程式是:

$$\text{sim}(u, v) = \frac{\sum_{a \in P_{uv}} (R_{ua} - \bar{R}_u)(R_{va} - \bar{R}_v)}{\sqrt{\sum_{a \in P_{uv}} (R_{ua} - \bar{R}_u)^2} \sqrt{\sum_{a \in P_{uv}} (R_{va} - \bar{R}_v)^2}} \quad (2)$$

## (3) 皮尔森相似性 (Pearsansimilarity)

令  $P_{uv}$  表示反映客户  $U$  和客户  $V$  的分数的项目集,  $R_{ua}$ ,  $R_{va}$  依次反映客户  $U$  和客户  $V$  在项目  $A$  上的分数, 并且依次反映客户  $U$  和客户  $V$  的平均分数, 运算公式是:

$$\text{sim}(u, v) = \frac{\sum_{a \in P_{uv}} (R_{ua} - \bar{R}_u)(R_{va} - \bar{R}_v)}{\sqrt{\sum_{a \in P_{uv}} (R_{ua} - \bar{R}_u)^2} \sqrt{\sum_{a \in P_{uv}} (R_{va} - \bar{R}_v)^2}} \quad (3)$$

## (4) 修正的 Pearson 系数

修正后的 Pearson 相关系数由传统 Pearson 相关系数和前部分对后部分的修正系数组成。

$$\text{sim}_{ij} = \frac{2|U|}{|U_i| + |U_j|} \frac{\sum_{u \in U} (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{ui} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{uj} - \bar{r}_j)^2}} \quad (4)$$

## 3. 评价指标

平均绝对误差代表了对预测的准确性的最后结果的推荐准确度。MAE 公式为:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |t_i - p_i| \quad (5)$$

## 二、算法流程

## 1. 环境设定

假定  $u$  用户作为系统的新用户, 没有记录过该  $u$  用户的相关信息, 而  $u$  用户在网上的视频、购物、商务和社交网站上有相应的历史行为记录, 可以通过与其他网站的授权方式了解该  $u$  用户的喜好程度和跨其他网站的其他历史行为。

## 2. 用户关系网络的构建

运用历史信息对比该用户的相关信息, 对其可能的兴趣爱好构建关系网络, 并按照其出现的评分数值进行社区划分, 所构建的关系网络计算出相似公式, 公式如下:

$$\text{sim}_{ij} = \frac{2|U|}{|U_i| + |U_j|} \frac{\sum_{u \in U} (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{ui} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{uj} - \bar{r}_j)^2}} \quad (6)$$

## 3. 社区划分

社区划分, 也就是运用模块化指标最大化的参数为依托, 以主模块最大化计算方式, 其计算公式如下:

$$Q = \frac{1}{2m} \text{Tr}(S^T B S), \text{ 其中 } B = A - \frac{dd^T}{2m} \quad (7)$$

## 4. 选择相似邻居策略的改变

在用户的选择过程中, 需要对比相似度最大的邻居用户

$k$ , 由于相关信息和数据采集的局限性, 只分析目标用户所在的社区进行用户选择, 计算公式具体如下:

## 三、实验设计与结果分析

## 1. 实验数据准备

本测试实验的测试实验数据信息采集自 MovieLens 服务功能网站, 数据信息集大小为 1000,000, 包含 943 个客户所有的顾客都对至少 20 部电影评价, 客户所打的评分分为 1, 2, 3, 4, 5 这样 5 个级别, 而所有客户最少对 20 部电影评分过。测试实验选用所采集到的数据信息集的 80.0% 作为训练集 (base 集), 剩下的 20.0% 作为测试集 (test 集)。

## 2. 实验设计

本节中总共进行了两组实验, 在第二组实验中, 用皮尔森和方向向量余弦、改正的方向向量余弦相似符合程度运算模式一, 根据遗忘分布作用曲线的相似符合程度运算模式展开对照, 通过对邻居数的转换, 验证了算法的有效性, 并以 MAE 为参考对测试实验的最终结果进行了评价。在第二测试实验中, 用皮尔森类似和根据遗忘分布作用曲线的相似符合程度运算, 经过转变相似符合程度极限数值来论证文中指出的算法的高效性。由此可知经过运算预计的客户打分和真实打分的误差来衡量预计打分的精确性, MAE 值越小, 说明推荐质量更好。

## (1) 实验一: 算法实现:

输入: 选择一个训练集  $U_{base1}$ , 该训练集为  $m \times n$  的用户评分信息矩阵, 其是  $m$  个用户  $U = \{u_1, u_2, u_3, \dots, u_m\}$ ,  $T_1$  部电影  $M = \{m_1, m_2, m_3, \dots, m_n\}$ ,  $r_{ui}$  为用户  $u$  对电影  $i$  的评分。

输出: MAE 值

Step 1: 对原始矩阵的评分  $r_{ui}$  进行减弱 (不在比较实验中)。

Step 2: 分别利用公式 1、公式 2 和公式 3 计算  $m \times n$  矩阵用户之间的相似度, 分别得到  $\cos$  相似度、 $\text{ACOS}$  相似度和  $\text{pcc}$  相似度, 得到目标用户的最近邻居集合。

Step 3: 利用公式 5 对  $U_{ti}$  进行评分预测产生预测集  $\{t_1, t_2, \dots, t_n\}$ , 比对 test 集中  $U_{ti}$  的评分, 计算 MAE 值。

## (2) 实验二:

输入: 选择一个训练集  $U_{base1}$ , 训练集为  $m \times n$  的用户评分信息矩阵, 其是  $m$  个用户  $U = \{u_1, u_2, u_3, \dots, u_m\}$ ,  $n$  部电影  $M = \{m_1, m_2, m_3, \dots, m_n\}$ ,  $r_{ui}$  为用户  $u$  对电影  $i$  的评分。

输出: MAE 值

Step 1: 衰减原始矩阵的分数 (在比较实验中跳过此步骤)。

Step 2: 使用公式 3 计算用户之间的相似性。

Step 3: 将步骤 2 的结果与相似度阈值进行比较, 将那些大于这个阈值的放入目标用户  $U_{ti}$  最近的邻居集合中。

Step 4: 使用公式 5 来评分  $U_{ti}$  和预测以生成预测集合  $\{t_1, t_2, \dots, t_n\}$ , 比较 test 集合  $U_{ti}$  中的分数, 并计算出 MAE 值。

### 3. 实验结果分析

#### (1) 实验一

在这组测试中,为了比较,对最近的邻居数进行了修剪。邻居的预期增长是从 10 到 100,结果如图 2 所示,为传统型皮尔森相似度算法和改进型相似度算法的测试实验的最终结果(其中红色丝线实线为传统型算法,蓝色虚线为改进型算法,以下同)。从图 2 能够得知,在邻居数等同于 30 时,两算法的 MAE 数值大概相同,然而当邻居数不断增长到 100 时,改善的相似作用程度算法 MAE 数值更小,说明其作用效果更加良好一些

图 3 是传统类型的方向矢量余弦和根据遗忘分布作用曲线的方向矢量余弦相似性算法的测试实验最终结果。根据示意图很容易得知随着近邻数的不断增加,改进算法的方向向量余弦的 MAE 值比传统的方向向量余弦小,表明预测更加准确。

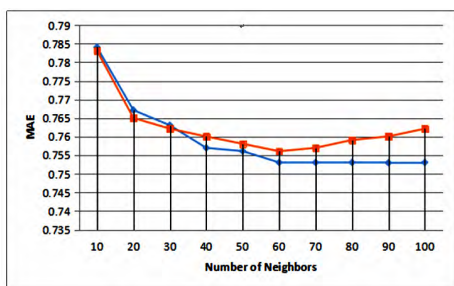


图 2 传统的与改进后的皮尔森相关系数的比较

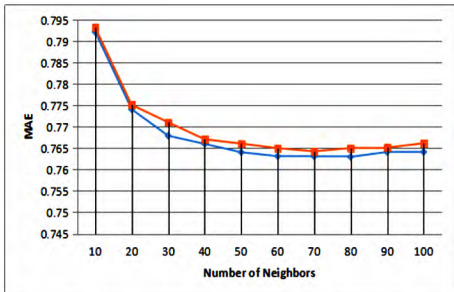


图 3 传统的与改进后的矢量余弦相似度的比较

图 4 是传统的修正方向矢量余弦和根据遗忘曲线改进的相似匹配度算法的测试实验的最终结果。从示意图可以看出,改进算法的 MAE 值始终低于传统的修正方向矢量余弦。

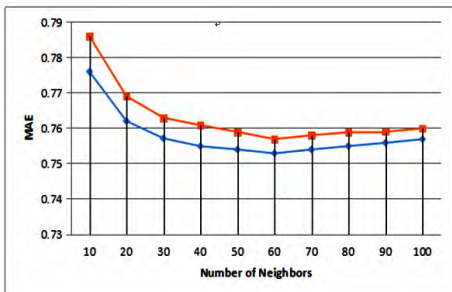


图 4 传统的与改进后的修正矢量余弦相似度的比较

通过以上三个实验,证明了文中提出的算法的有效性。

在计算用户之间的相似度和重合度时,应考虑自然遗忘引起的顾客对物质资源评价的变化。客户的爱好和兴趣会随着时间的推移而变化。通过模拟遗忘过程,提高了预测的准确性。

#### (2) 实验二

用皮尔森类似和根据遗忘分布作用曲线的相似符合程度运算,经过转变相似符合程度极限数值来论证文中指出的算法的高效性。由此可知,运算预计的客户打分和真实打分的误差来衡量预计打分的精确性,MAE 值越小,推荐质量越好。限制值的操作范围从 0.1 逐渐增加到 0.9,增加的总量为 0.1。设置限制值后,使用的邻居和客户相似,且符合度高于此闭合值,之后依次展开测试实验。当极限数值高于 0.4 时,传统类型的皮尔森有关参数的 MAE 的有效数值要显著超过根据遗忘分布作用曲线而改善的算法的 MAE 的有效数值,而低于 0.4 的时候,传统类型的模式要好很多。换言之,当极限数值高于 0.4 的时候,改善的算法预计的更精确一些。

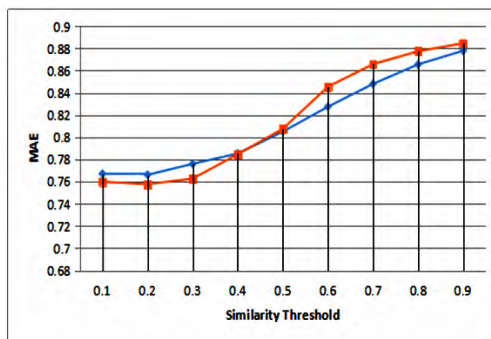


图 5 传统的与改进后的皮尔森相关比较

总而言之,经过两组实验的最终实验结果对比分析,整体上来看,文中提出的以神经网络为基础的推荐系统协同过滤算法能够有效的解决冷启动问题并且显著的提升了预测的精确度,与其他传统的推荐算法相比下来,该算法在相应的评价指标上有着不错的结果,优化了推荐结果同时提高了推荐的准确率。

#### 参考文献

- [1] 薛亮,徐慧,冯尊磊等.一种改进的协同过滤的商品推荐方法[J].计算机技术与发展,2022,32(07):201-207.
- [2] 周迅,刘超慧,周克萍等.融入惩罚因子的协同过滤图书推荐算法[J].江苏科技信息,2022,39(17):76-80.
- [3] 万玉龙,顾峰,柏小颖等.冷链物流智能配载协同过滤推荐算法的设计研究[J].中国储运,2022,(06):93-94.
- [4] 王冠宇,钟婷,冯宇等.基于矢量量化编码的协同过滤推荐方法[J].计算机科学,2022,49(09):48-5.
- [5] 吕淑君.基于卷积神经网络的高等数学资源推荐算法[J].中国新通信,2022,24(07):43-44+57.