

文章编号: 2095-4980(2020)06-1128-05

基于卷积神经网络的旅游推荐模型设计

张佳琳, 柏思佳, 刘爽

(哈尔滨商业大学 计算机与信息工程学院, 黑龙江 哈尔滨 150028)

摘要: 旅游业和网络时代高速发展, 导致旅游信息过载问题日益严重, 旅游推荐方法对解决信息过载问题十分重要。传统推荐算法只针对用户和项目之间的评分和基本属性计算相似度进行推荐, 但行为需求及具有游客情感因素的评论却被忽视。本文利用卷积神经网络(CNN)对文本评论特征提取进行情感分类, 用皮尔逊相似度公式计算相似的用户群体, 用平均绝对误差(MAE)对结果误差进行评价。与传统的协同过滤方法进行对比, 本文提出的模型能有效降低预测误差。

关键词: 信息过载; 旅游推荐; 卷积神经网络; 协同过滤

中图分类号: TP183

文献标志码: A

doi: 10.11805/TKYDA2019288

Design of travel recommendation model based on convolutional neural network

ZHANG Jialin, BAI Sijia, LIU Shuang

(School of Computer and Information Engineering, Harbin University of Commerce, Harbin Heilongjiang 150028, China)

Abstract: The rapid development of tourism and Internet leads to the increasingly serious problem of tourism information overload. Therefore, tourism recommendation method is very important to solve the problem of information overload. Traditional recommendation algorithms only calculate similarity between users and items based on the score and basic attributes, behavioral needs and comments with tourist emotional factors are ignored. In this paper, Convolutional Neural Network(CNN) is utilized to classify the feature extraction of text comments, Pearson similarity formula is adopted to calculate similar user groups, and Mean Absolute Error(MAE) is employed to evaluate the error of the results. Compared with the traditional collaborative filtering method, the experimental results show that the proposed model can effectively reduce the prediction error.

Keywords: information overload; travel recommendation; convolutional neural network; collaborative filtering

在线旅游已成为日常旅游的应用方式, 在很大程度上为人们出行旅游带来了方便。但由于旅游人数的增多和网络信息的爆炸式增长, 导致旅游用户很难在海量的数据中及时找到自己所需的旅游数据, 推荐系统的出现有效地解决了这一信息过载问题。推荐系统^[1]指的是, 通过网站平台挖掘用户的显式信息和隐式信息判断用户的喜好, 选取其中最符合的 n 项产品推荐给用户, 从而帮助用户高效地抉择。为了更加深入地研究推荐系统, 各种各样的推荐算法应运而生。常用的推荐算法有: 基于内容的推荐算法、协同过滤(Collaborative Filtering, CF)推荐算法和混合推荐算法^[2]。基于内容的推荐算法分析用户的历史内容数据, 建立兴趣模型, 将预测的相似兴趣项目推荐给用户, 但该算法最大的问题是推荐领域狭窄。协同过滤算法仅依赖用户的历史数据, 得到兴趣相似的用户群体, 然后把该用户群体中偏好的项目推荐给目标用户^[3]。混合推荐算法将多种算法融合, 在理论上是最能有效解决各算法缺点的一种算法, 但该算法的设计相对复杂。

近几年, 人工智能发展得如火如荼, 作为人工智能领域重要研究方向的深度学习已渗透到各行各业。深度学习可以深层次地挖掘图像、文本和自然语言处理等方面的数据特征, 尤其针对容易忽略的隐式信息, 取得了

收稿日期: 2019-08-13; 修回日期: 2019-10-23

基金项目: 黑龙江省博士后科研启动金资助项目(LBH-Q19028); 教育部人文社会科学基金项目资助项目(18YJAZH128); 黑龙江省哲学社会科学规划项目资助项目(18GLB029)

作者简介: 张佳琳(1979-), 女, 博士, 副教授, 主要研究方向为数据挖掘、大数据分析等。email:pansv@126.com

一定进展^[4]。在传统的旅游推荐模型中，只是根据游客和景点之间的评分进行预测，忽视了游客和景点的特征属性及游客对旅游体验的评论信息。通过对属性的分析及利用卷积神经网络挖掘评论信息可以有效改善旅游推荐的质量。Fan Yang 等^[5]为了建立视频中多模式特征之间的特征表示，提出了一种基于多模态深度学习的多模态推荐算法，有效获取了视频中更多信息。高茂庭等^[6]为了更精准地向用户推荐电影，利用循环神经网络处理用户和电影的时间序列数据，使用长短期记忆单元实现长时间距离的历史状态的积累，最终提高推荐准确度。本文针对旅游业信息过载问题设计了一个基于卷积神经网络的旅游推荐模型。

1 相关工作

以往旅游推荐中很少将游客的评论作为考虑的因素，但评论中含有游客对此次旅游的情感因素，这些因素已成为其他用户评判的重要依据。本文将卷积神经网络和协同过滤方法相结合，通过卷积神经网络对游客文本评论中的因素特征提取并进行分类，利用协同过滤方法对用户项目评分进行预测。其中，卷积神经网络对评论中的特征进行挖掘可以有效缓解协同过滤方法中的数据稀疏性问题，并且该方法可以保留协同过滤方法中共享经验、个性化的优点。

1.1 协同过滤推荐算法

作为应用最广泛的推荐算法，协同过滤的核心思想是计算用户或项目之间的相似度实现推荐。基于用户的协同过滤推荐算法是计算用户之间的相似度，建立近邻用户群，然后将近邻用户群中兴趣产品推荐给目标用户；基于项目的协同过滤推荐算法计算的是项目间的相似度；基于模型的协同过滤推荐算法结合一些智能模型对数据进行训练和测试，以便更有效地获取用户的偏好信息实现推荐^[3]。

为了找到近邻的用户或项目，相似度计算是关键。计算相似度方法中最常用的是皮尔逊方法及余弦方法^[7]。其中皮尔逊相似度方法更适用于文档数据的计算，且对于不同用户的考虑更加全面，该公式表示为：

$$sim(a,b) = \frac{\sum_{i \in I_{a,b}} (R_{a,i} - \bar{R}_a)(R_{b,i} - \bar{R}_b)}{\sqrt{\sum_{i \in I_{a,b}} (R_{a,i} - \bar{R}_a)^2 \sum_{i \in I_{a,b}} (R_{b,i} - \bar{R}_b)^2}} \quad (1)$$

式中： $sim(a,b)$ 表示用户 a,b 间的相似度； $I_{a,b}$ 为用户 a,b 共同用户特征； $R_{a,i}$ 为用户 a 的第 i 个特征值^[7]， $R_{b,i}$ 为用户 b 第 i 个特征值； \bar{R}_a, \bar{R}_b 分别为用户 a,b 的平均特征评分。

Jianrui Chen 等^[8]采用一种新的动态进化聚类算法将用户和项目集划分为 K 个聚类，并对相似度较高的个体进行聚类。在 K 个聚类各自中应用协同过滤技术预测，根据聚类所得预测等级对目标用户进行推荐预测。Linhui XU 等^[9]利用组合协同过滤推荐算法，采用离线测试方法提高了团购网站的精准利用率。

1.2 卷积神经网络

卷积神经网络(CNN)是目前深度学习中研究的重点之一，它在图像处理领域的应用已经相当成熟，也广泛用于文本评论的处理。作为一种前馈神经网络，CNN 基本结构与其他神经网络相似，都是由输入层、隐含层及输出层进行数据特征提取。CNN 的隐含层包括卷积、池化及全连接，共 3 层^[10]。其中，卷积层是最重要的，它将输入信号与含有特征提取器的卷积核进行卷积计算，深层次地挖掘数据内部信息^[11]；池化层通过池化函数将提取特征进行过滤筛选，对数据降维，保留更重要的特征数据；全连接层将上层处理过的特征数据整合到一起，再传至输出层。CNN 结构如图 1 所示。

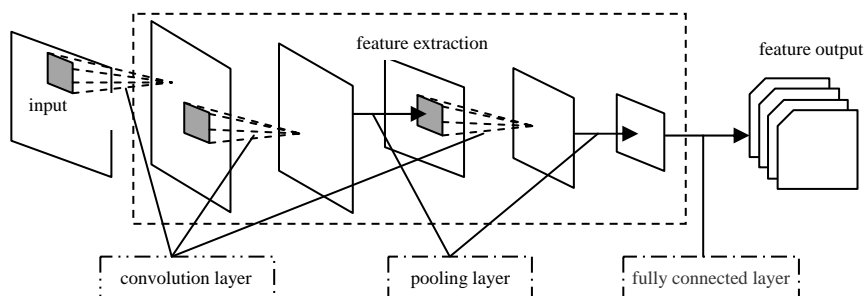


Fig.1 Structure of convolutional neural network
图 1 卷积神经网络结构

Nan Zhou 等^[12]利用卷积神经网络中卷积层和池化层的非线性特征过滤微博文本内容，确保搜索信息安全。Ashu Abdul 等^[13]将深度卷积神经网络和加权特征提取的输出结合，实现用户数据和音乐之间的相关性提取。

2 旅游推荐模型构建

2.1 问题描述

本文研究目的是设计个性化的旅游推荐模型，针对基于用户协同过滤推荐算法很难挖掘深层次信息的数据稀疏性问题，选用卷积神经网络对旅游数据的评论信息深入抽取进行分类。游客对旅游的文本评论中包含了对旅游过程的情感因素，分为正面和负面两类。将旅游用户的基本信息、行为信息及情境信息相结合构成旅游用户数据，解决数据单一问题。用户数据的具体分类如下：

- 1) 用户的基本信息 P_i 包括：姓名(ID)、年龄(age)、性别(sex)、所在地(place)等属性，其中姓名(ID)起唯一标识作用，不参与数据的分析过程，即 $P_i = \{name_i, age_i, sex_i, place_i\}$ 。
- 2) 行为信息 S_i 分为需求(demand)、兴趣偏好(prefer)及消费情况(consume)，即 $S_i = \{demand_i, prefer_i, sex_i, consume_i\}$ 。
 - a) 需求信息是指用户的食、住、行以及旅游方式。旅游方式包括：家人游、密友游、独游、情侣/夫妻游；本地游、周边游、异地游(指离所在城市较远)等。
 - b) 兴趣偏好包括喜欢什么样的环境、什么风格的景点以及什么活动等。
 - c) 消费情况指用户的价格偏好，如出行方式、门票价格、住宿价格等。
- 3) 情境信息 Q_i 包括所处时间(time)、用户所在位置(location)、用户在网络平台上的会话及互动信息(interact)，其中会话和互动信息是指用户对景点的访问时间、访问次数、收藏记录等， $Q_i = \{time_i, location_i, interact_i\}$ 。

2.2 模型构建

本文的旅游推荐模型主要考虑用户数据中的基本信息、行为信息和情境信息三大要素，将向量化的旅游用户评论数据矩阵发送到卷积神经网络^[14]中训练，得到游客对项目的正负面情感，再将处理后的旅游数据进行相似度计算。旅游推荐模型结构如图 2 所示。

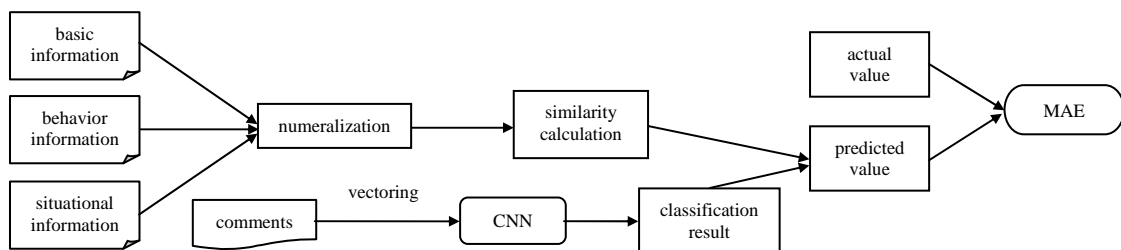


Fig.2 Structure of recommended model
图 2 推荐模型结构

旅游用户的基本信息表示为：

$$P = \alpha_1 age + \alpha_2 sex + \alpha_3 place \tag{2}$$

旅游用户的行为信息表示为：

$$S = \beta_1 demand + \beta_2 prefer + \beta_3 consume \tag{3}$$

旅游用户的情境信息表示为：

$$Q = \gamma_1 time + \gamma_2 location + \gamma_3 interact \tag{4}$$

所以，旅游推荐特征信息表示为：

$$TR = \alpha P + \beta S + \gamma Q \tag{5}$$

式中： α 为基本信息的权重， $\alpha = [\alpha_1, \alpha_2, \alpha_3]$ ； β 为行为信息的权重， $\beta = [\beta_1, \beta_2, \beta_3]$ ， γ 为情境信息的权重， $\gamma = [\gamma_1, \gamma_2, \gamma_3]$ 。

对上面提到的非结构化数据进行处理，其中年龄在 18 岁以下记为 1，18 岁~25 岁记为 2，26 岁~35 岁记为 3，36 岁~45 岁记为 4，46 岁~55 岁记为 5，56 岁及以上记为 6；性别为男记为 0，女记为 1；所在地相同记为 1，不同记为 0；按照上面旅游方式的排序分别用 1~7 表示；消费在 1 000 元以下记为 1，1 000~2 000 元记为 2，2 001~3 000 元记为 3，3 001~4 000 元记为 4，4 001~5 000 元记为 5，5 000 元以上记为 6；所处时间按照春

夏秋冬划分，其中 3 月~5 月记为 1，6 月~8 月记为 2，9 月~11 月记为 3，12 月~2 月记为 4。利用式(1)计算用户之间特征的相似度。由于不同的用户特征对用户相似度的影响不同，因此用户的相似度的计算需要对不同的特征赋予相应的权值，通过式(6)找到相似的用户群。

$$SIM(a,b) = \alpha sim(a_p, b_p) + \beta sim(a_s, b_s) + \gamma sim(a_o, b_o) \tag{6}$$

向量数据输入卷积神经网络后，在卷积层中与卷积核进行内积计算，卷积层中含有多个卷积核，每一个组成卷积核的属性元素都具有各自的权重系数和偏量，当处理后的旅游向量数据尺寸小于原旅游向量数据时，可以采用 zero 填充样本边界。卷积神经网络的特征值可以通过式(7)计算得到。

$$c_i = f(w \cdot vec_i + b) \tag{7}$$

式中：w 为卷积权重值；vec_i 为单词向量；b 为偏置项。

选择常用非线性 sigmoid 激活函数完成卷积计算和权值偏量的调节，激活函数为：

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

经卷积计算后的数据进入池化层，选用最大池化法得到结果。在池化层中不含有单独的权重系数。假设卷积层得到的旅游用户数据是 $U = \{u_1, u_2, \dots, u_S\}$ ，则池化结果 P 为：

$$P = max(U) = max\{u_1, u_2, \dots, u_S\} \tag{9}$$

池化结果到达全连接层后，通过 ReLU 激活函数，得出最终旅游用户隐含特征信息 TR 及调节后的权重系数及偏差，然后输出结果，计算公式如下：

$$TR = ReLU(w \cdot vec_i + b) \tag{10}$$

式中 $ReLU(x) = max(0, x)$ 。

卷积神经网络通过对游客评论中是否含有表达情感因素的直接分类词对评论的正负面进行分类，根据正负面词向量所占比例，由高到低进行排序，并对其进行打分。评论中正面词汇达 100% 的记为 5 分，99% 的记为 4.9 分，以此类推，直到 1 分为止，然后根据排列顺序为预测用户进行推荐。

3 实验验证

3.1 实验数据及处理

本文利用八爪鱼爬取马蜂窝上北京 5A、4A 级 32 个景点的游记数据集及携程网站上评分数据集。首先将数据中的重复数据、空白数据及无关景点的无效数据删除；然后对数据进行噪声过滤，即去除非文字符号；再将成句的评论数据利用 Python 的 jieba 对其进行分词处理并用 stopwords 对停用词过滤。卷积神经网络对输入的向量长度有一定的要求，因此要统一分词后的句子长度。处理后的分词数据长度大多数在 50 左右，因此选用其中长度在 50 左右的句子，长度不足或过长可以采用填充或采样处理。将各个游客及景点对应的数据进行合并；再把合并后的分词数据利用 word2vec 向量化；将向量化的多维数据发送到卷积神经网络中提取特征。

3.2 评价指标

实验评价指标选用平均绝对误差(MAE)，MAE 的值越小，代表误差越小，结果越准确。其计算公式为：

$$MAE = \frac{\sum_{u,i \in n} |r_{u,i} - r'_{u,i}|}{n} \tag{11}$$

式中：r_{u,i} 表示游客 u 对景点 i 的实际评分；r'_{u,i} 表示游客 u 对景点 i 的预测分数；n 表示测试数量。

3.3 实验结果与分析

为了更加直观地体现出卷积神经网络与协同过滤方法结合的有效性，本实验将该方法与只考虑用户和项目之间评分的传统协同过滤方法进行对比，实验结果见图 3。从图 3 可以看出，本文方法的 MAE 值要小于协同过滤方法的 MAE 值，说明考虑游客的行为信息、情境信息及评论中的情感因素可以有效提高推荐的准确率。随着测试集数据量的增多，虽然存在 MAE 值上升的情况，但整体呈现下降趋势，说明数据量较多时，更能充分地验证模型的质量。

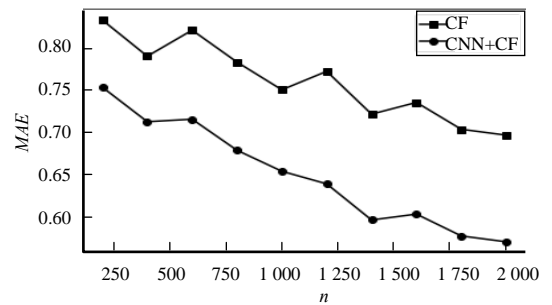


Fig.3 Comparison of experimental results
图 3 实验结果对比图

4 结论

针对传统旅游推荐的不足,提出一种基于卷积神经网络的旅游推荐模型。该模型通过对旅游用户的基本信息、行为信息、情境信息及游客评论数据分析研究,利用卷积神经网络对评论数据中所含正负面的情感进行分类,并计算出游客的相似群体,最终得到预测分值与实际评分的差值。该模型在爬取的马蜂窝及携程数据上进行实验,实验结果表明,本文提出的模型能有效提升推荐效果。

参考文献:

- [1] 王磊,高宸,周蓓,等. 基于深度学习与社交感知的地点推荐[J]. 太赫兹科学与电子信息学报, 2019,17(3):502-508. (WANG Lei,GAO Chen,ZHOU Bei,et al. Deep learning based social-aware location recommendation[J]. Journal of Terahertz Science and Electronic Information Technology, 2019,17(3):502-508.)
- [2] 杨博,赵鹏飞. 推荐算法综述[J]. 山西大学学报(自然科学版), 2011,34(3):337-350. (YANG Bo,ZHAO Pengfei. Review of the art of recommendation algorithms[J]. Journal of Shanxi University(Natural Science Edition), 2011,34(3):337-350.)
- [3] SUN P,LI Z Y,HAN Z Y,et al. An overview of collaborative filtering recommendation algorithm[J]. Advanced Materials Research, 2013(756-759):3899-3903.
- [4] 张荣,李伟平,莫同. 深度学习研究综述[J]. 信息与控制, 2018,47(4):385-397,410. (ZHANG Rong,LI Weiping,MO Tong. Review of deep learning[J]. Information and Control, 2018,47(4):385-397,410.)
- [5] YANG Fan,XIE Hao,LI Huxiong. Video associated cross-modal recommendation algorithm based on deep learning[J]. Applied Soft Computing Journal, 2019(82):105597-1-9.
- [6] 高茂庭,徐彬源. 基于循环神经网络的推荐算法[J]. 计算机工程, 2019,45(8):198-202,209. (GAO Maoting,XU Binyuan. Recommendation algorithm based on circular neural network[J]. Computer Engineering, 2019,45(8):198-202,209.)
- [7] 荣辉桂,火生旭,胡春华,等. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014,35(2):16-24. (RONG Huigui, HUO Shengxu,HU Chunhua,et al. User similarity-based collaborative filtering recommendation algorithm[J]. Journal on Communications, 2014,35(2):16-24.)
- [8] CHEN J R,WEI L D,ZHANG L. Dynamic evolutionary clustering approach based on time weight and latent attributes for collaborative filtering recommendation[J]. Chaos,Solitons and Fractals:The Interdisciplinary Journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena, 2018(114):8-18.
- [9] 徐林会,鲁新新. 基于协同过滤算法的团购个性化推荐系统[J]. 辽宁工程技术大学学报(自然科学版), 2017,36(7):761-766. (XU Linhui,LU Xinxin. Collaborative filtering recommendation system based on interest drifting and similar cloud[J]. Journal of Liaoning Technical University(Natural Science), 2017,36(7):761-766.)
- [10] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报, 2017,40(6):1229-1251. (ZHOU Feiyan,JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017,40(6):1229-1251.)
- [11] 刘健,袁谦,吴广,等. 卷积神经网络综述[J]. 计算机时代, 2018(11):19-23. (LIU Jian,YUAN Qian,WU Guang,et al. Review of convolutional neural networks[J]. Computer Era, 2018(11):19-23.)
- [12] ZHOU Nan,DU Junping,XUE Zhe,et al. Security topics related microblogs search based on deep convolutional neural networks[J]. Neurocomputing, 2020(395):199-211.
- [13] ABDUL Ashu,CHEN Jenhui,LIAO Huayuan,et al. An emotion-aware personalized music recommendation system using a convolutional neural networks approach[J]. Applied Sciences, 2018,8(7):1103-1-16.
- [14] WANG Peng,XU Bo,XU Jiaming,et al. Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification[J]. Neurocomputing, 2016(174):806-814.