

基于用户评论的深度学习模型预测 旅游景点推荐

张凝琛¹, 李佳臻²

¹澳门科技大学资讯科技学院, 澳门

²对外经济贸易大学信息学院, 北京

收稿日期: 2021年10月12日; 录用日期: 2021年11月11日; 发布日期: 2021年11月19日

摘 要

随着在线旅游交易规模的增长, 通过个性化推荐技术为游客提供更好的服务, 吸引了业界的关注。本文主要探讨在线旅游平台上用户评论在景点推荐过程中的作用, 首先利用用户评论构建景点间社会网络, 并基于深度学习模型LSTM预测用户对景点的情感得分; 然后按照时间序列对情感得分加权作为用户对景点的评分; 最后基于评分、景点社会网络信息和协同过滤算法实现景点的推荐。采用携程网的用户评论数据对推荐效果进行检验, 实验结果表明在线平台的用户评论在景点推荐的多样性和新颖性方面影响效果显著, 但对推荐准确性改善并不显著。

关键词

文本分析, 深度学习, 情感得分, 景点推荐

A Deep Learning Model Based on User Reviews to Predict Tourist Attractions Recommendations

Ningchen Zhang¹, Jiazhen Li²

¹Faculty of Information Technology, Macau University of Science and Technology, Macau

²School of Information Technology and Management, University of International Business and Economics, Beijing

Received: Oct. 12th, 2021; accepted: Nov. 11th, 2021; published: Nov. 19th, 2021

Abstract

With the growth of online tourism transaction scale, providing better services for tourists through

文章引用: 张凝琛, 李佳臻. 基于用户评论的深度学习模型预测旅游景点推荐[J]. 计算机科学与应用, 2021, 11(11): 2725-2730. DOI: 10.12677/csa.2021.1111276

personalized recommendation technology has attracted the attention of industry and academia. This paper mainly discusses the role of user comments in the process of scenic spot recommendation on the online tourism platform. Firstly, the social network between scenic spots is constructed by user comments, and the user's emotional score on scenic spots is predicted based on the deep learning model LSTM; then, the emotion score is weighted according to the time series as the user's score of the scenic spot; finally, the recommendation of scenic spots is realized based on scoring, scenic spot social network information and collaborative filtering algorithm. The data on CTRIP website is used to test the recommendation effect. The experimental results show that user comments of the online platform has a significant effect on the diversity and novelty of scenic spot recommendation, but the improvement of recommendation accuracy is not significant.

Keywords

Text Analysis, Deep Learning, Emotion Score, Scenic Spot Recommendation

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着在线旅游交易规模的增长,利用在线旅游平台上的数据获取用户偏好,并通过个性化推荐技术为游客提供更好的服务,吸引了产业界和学术界的关注[1][2]。数据的稀疏性、冷启动等问题制约着平台推荐算法的性能,学者们开始将用户的时空特征、人口统计特征、用户评论应用到旅游景点的推荐,提升推荐性能[3]。但缺少对用户评论文本情感分析的应用,特别是评论文本隐含的情感倾向与评分值不一致情况的分析。

本文主要探讨在线旅游平台上的评论文本对旅游景点推荐的影响,主要贡献有两点:

- 1) 基于游客的评论文本信息和 LSTM (Long Short-Term Memory)模型预测用户对景点的情感得分,并按照时间序列加权替代用户对景点的评分,降低了评分数据稀疏以及部分评分与评论不一致对推荐算法性能的影响。
- 2) 基于用户的评论数据构建景点间社会网络,利用社会网络图中顶点的中心性计算景点的流行度,并将融入到景点的预测评分计算,进一步降低数据稀疏性对推荐算法性能的影响。

2. 文献综述

推荐技术已被广泛地应用在电子商务、产品服务、旅游、朋友发现等诸多领域。推荐算法大体上分为:基于评分的推荐方法、基于内容的推荐方法以及两种方法的混合推荐方法。

基于评分的推荐算法主要依赖于用户历史评分,代表性的是协同过滤算法[4],其基本思想是相似的用户具有相似的兴趣,主要有基于物品的协同过滤(Item-CF)和基于用户的协同过滤(User-CF)算法。然而,协同过滤算法性能受评分数据的稀疏性影响很大,在线旅游平台上的游客评分数据存在同样的问题。学者们在协同过滤算法中融入的游客和景点信息,基于内容的推荐和混合推荐模型方法被应用到景点推荐[5]。刘海鸥等使用游客基本信息、行为偏好信息和情境信息对游客进行画像,使用基于用户画像的情境化方法提供推荐服务[6]。Ravi 等使用基于位置的社交网络的游客反馈数据中的积极评论构建语言模型,挖掘游客的社会活动和行为,抽取游客的时空属性[7]; Jiang 等从旅游日志中描述游客的兴趣和景点的特征的标签,如游客购买力、时间、季节偏好、景点空间等标签[8]。上述的这些研究主要是将从文本中抽取的各种主题

融入到推荐算法, 缺少评论文本游客情感的分析, 特别是评论文本隐含的对景点的情感倾向与评分值不一致的分析。社会网络分析也成为了提升基于内容推荐、混合推荐模型性能的重要方法, 部分学者们利用社会网络特征信息提升个性化推荐算法的性能, 胡云等基于社会网络分析的用户间的信任关系, 实现了信任关系和社会标签融合的社会网络推荐方法, 解决了推荐算法的评分矩阵稀疏性问题[9]。Yuchen Pan 等在进行 O2O 服务推荐时, 利用网络中心性, 建立频繁服务集网络, 解决了数据稀疏性问题[10]。

随着深度学习模型在自然语言处理展现出的效果, 将基于深度学习模型的文本分析结果融入到推荐算法以提高推荐性能值得深入探讨。长短时记忆网络(LSTM)在捕捉短期序列信息的同时保留较长距离信息的依赖关系, 在语音识别、建模, 情感分析和预测等序列分析中得到广泛的应用[11]。Cheng 等在进行旅游行程推荐时, 认为旅游历史记录和下一个目的地的选择具有相关性, 且相关程度呈指数衰减, 在景点推荐中考虑偏好的序列相关性有助于实现更准确的预测[12]。

受上述研究的启发, 对景点较固定且数量有限、游客广泛的景点推荐问题, 可以利用 LSTM 模型提取游客评论文本的情感信息, 修正游客评分, 采用基于项目的协同过滤算法实现景点推荐。因而本文从基于用户评论构建的景点社会网络信息、游客评论的时间信息、文本情感信息几个视角探讨在线旅游平台的用户评论在景点推荐过程的作用。

3. 融合用户评论的景点推荐算法

3.1. 用户评论文本情感分析

文本的情感分析是对文本包含的情感倾向进行分析, 其一般流程是: 使用 word2vec 等词嵌入模型获取文本中的特征词, 将文本转化为词向量, 整理得到文本矩阵和特征词表; 然后将词向量作为机器学习算法的输入, 得到该文本正向情感或负向情感的概率。LSTM 深度学习模型可以在捕捉短期序列信息的同时保留较长距离的依赖关系, 在文本情感分析展现出了很好的效果[11]。因而, 面对文本长度不同且句式复杂的旅游评论文本, 本文选择 LSTM 深度学习模型对游客评论文本进行情感得分的计算, 并用情感得分代替用户对景点的评分。首先对每一条评论使用 jieba 分词, 去掉常见停止词和数词; 然后将评论文本转为 one-hot 编码并用其训练 word2vec 模型, 将评论中的每个词转化为词向量, 作为 LSTM 模型的输入。根据用户的评分人工标注了部分数据训练 LSTM 模型, 其中 5 分评论作为正向情感, 4 分评论作为中性情感, 3 分及以下作为负向情感。训练模型时按照 70% 的比例抽取训练集, 并在训练集按照数据比例增大中性和负向情感数据的权重, 进行样本均衡, 用剩余 30% 评论文本数据作为测试集检验模型的性能。然后用训练好的 LSTM 模型计算每条评论的情感得分。

由于偏好具有较强的序列相关性, 游客对景点的偏好会随着时间的推移而产生变化, 较为近期的游览记录可以更准确的体现出游客在旅游目的地选择上的偏好, 对游客的游览记录按时间排序, 通过拉普拉斯调整后的指数平滑方法计算游客评论的时序重要性, 即相应评论情感得分的重要性, 如式(1)所示。

$$w_i = \frac{2}{1 + e^{rx_i}} \quad (1)$$

其中, r 为指数平滑的参数, 本文设置 $r = 1$, x_i 代表景点 i 在某游客所有游览记录的顺序。

3.2. 景点社会网络

假设共有 n 个景点, 根据游客评论记录构建无向图表示景点社会网络, 图的顶点表示景点, 顶点间的边表示两个景点被同一用户评论过, 边上的权重 f_{ij} 表示景点 s_i 和 s_j 被同一游客评论(游览)过的频率。为了缓解协同推荐过程中评分数据稀疏性问题, 在景点社会网络的基础上定义景点中心性, 作为景点的流行度, 如式(2)所示。

$$pop_i = \frac{cen_i}{\sum_{j=1}^n cen_j} = \frac{\sum_{j=1}^n f_{ij}}{\sum_{j=1}^n \sum_{p=1}^n f_{jp}} \quad (i \neq j) \quad (2)$$

3.3. 景点相似性计算

本文采用 Jaccard 相似性度量景点间的相似性, 如式(3)所示, $|u_{s_i}|$ 表示游览过景点 s_i 的游客数, $|u_{s_i} \cap u_{s_j}|$ 表示同时游览过景点 s_i 和景点 s_j 的游客数。

$$sim_{s_i, s_j} = \frac{|u_p \cap u_q|}{|u_p|} = \frac{f_{ij}}{f_{ii}} \quad (3)$$

3.4. 推荐算法

输入: 旅游评论数据集;

输出: 推荐的 k 个景点列表

Step 1: 用部分标注的用户评论文本训练 LSTM 模型, 计算用户 u 对景点 s_i 评论的情感得分 $senscore_{u, s_i}$;

Step 2: 对评论文本数据集中每个用户 u , 构造其游览过的景点集合 V_u , 并将其评论按时间倒序排序, 按照式(2)计算评论情感权重, 构建景点社会网络;

Step 3: 计算景点的流行度 pop_i 和景点间的相似性;

Step 4: 对待推荐的用户 u , 读取该游客游览过的景点集合 V_u , 把景点网络中与 V_u 有边相连但用户 u 没有游览过的景点作为推荐的候选集合 $V_{u, R}$ 。用式(4)计算用户 u 对候选集合 $V_{u, R}$ 中每个景点 s_i 的预测评分, 并降序排序, 将排在前 N 的景点推荐给用户 u 。

$$sscore_{u, s_i} = \sum_{s_j \in V_u} w_j \times senscore_{u, s_j} \times sim_{s_i, s_j} \times (1 + pop_i) \quad (4)$$

若数据稀疏性较大, 则景点 s_i, s_j 的相似性 sim_{s_i, s_j} 可能为 0, 因此在式(4)中引入景点的流行度, 以降低数据稀疏性对推荐的影响。

4. 实验及结果分析

本文使用的实验数据来自携程网, 是截止 2020 年 3 月 2451 个景点、21,616 游客的 592,178 条用户评论文本及其评分数据 $score_{u, s_i}$, 评分数据取值范围是[1, 5]。

4.1. 推荐评价指标

本文采用推荐精确率/Precision、召回率/Recall、F1 值/F1-Score 和多样性评价推荐算法的性能, 计算公式如式(5)~(8)所示。

$$\text{Precision} = \frac{N_k}{k} \quad (5)$$

$$\text{Recall} = \frac{N_k}{N_{used}} \quad (6)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

其中, k 为推荐算法推荐的景点数量, N_k 为 k 个推荐的景点中用户游览的景点数量, N_{used} 为所有游客实际游览的景点中被准确推荐的景点数量。

通过内部列表距离定义的个体多样性描述推荐算法的多样性[13], 描述了推荐列表中所有推荐商品两两相似性的平均, 其计算公式如式(8)所示。

$$ILS(R_u) = \frac{2 \sum_{i=1}^k \sum_{j=1}^k sim_{s_i, s_j}}{k(k-1)} \quad (8)$$

其中, R_u 表示用户 u 的推荐列表, s_i 和 s_j 表示该用户推荐列表中的第 i 和第 j 个被推荐的景点, $i=1, 2, \dots, k$ 且 $i \neq j$, k 为推荐列表的长度, 即为每个游客推荐的景点数量。

4.2. 情感分析模型 LSTM 性能验证

首先对每一条评论文本使用 jieba 进行分词, 并去掉常见停止词和数词; 然后将评论文本转为 one-hot 编码并用其训练 word2vec 模型, 将评论中的每个词转化为词向量, 作为神经网络的输入。在超参数的设置上, 使用默认的学习率, 使用小批量梯度下降, 梯度大小为 32。LSTM 层设置输出维度为 50, activation = “tanh”, inner_activation = “hard_sigmoid”, 优化算法使用 adam 算法。训练模型时首先按照 70% 的比例抽取训练集, 剩余 30% 评论文本数据作为验证集。按 5 分评论作为正向情感, 4 分评论作为中性情感, 3 分及以下作为负向情感, 按照数据比例增大中性和负向情感数据的权重, 对训练样本集强制进行样本均衡, 训练 LSTM 模型。最终得到模型的训练准确率为 0.91, 预测准确率为 0.89。

4.3. 实验结果分析

选择 Item-CF、Item-CF + 景点社会网络分析、本文的推荐方法, 推荐列表长度 $k=10$, 实验结果如表 1 所示。其中, Item-CF 算法中用户对景点评分采用的是网站上获取的评分数据, 并用式(9)计算用户的预测评分; Item-CF + 景点社会网络分析是在 Item-CF 算法中融入景点的社会网络信息, 用式(10)计算用户的预测评分。

$$sscore_{u, s_i} = \sum_{s_j \in V_u} score_{u, s_j} \times sim_{s_i, s_j} \quad (9)$$

$$sscore_{u, s_i} = \sum_{s_j \in V_u} score_{u, s_j} \times sim_{s_i, s_j} \times (1 + pop_i) \quad (10)$$

Table 1. Performance comparison of recommendation algorithms

表 1. 推荐算法性能对比

评价指标	Item-CF	Item-CF + 网络分析	本文算法
精确率	0.0095	0.0460	0.0584
召回率	0.0071	0.0343	0.0435
F1-score	0.0082	0.0393	0.0500
多样性	0.7004	0.6828	0.7625

从表 1 数据可以发现, 基于物品的协同过滤算法(Item-CF)几乎无法做出任何准确的推荐。这是由于在极度稀疏的数据集上, 作为协同过滤算法核心的相似度计算难以实现。而本文提出的推荐算法在精确率、召回率、F1-score、多样性上都具有较好的表现, 说明网络分析方法和文本分析方法在解决数据稀疏性上具有较好效果, 且情感的提取有助于分析游客的兴趣, 并提供令游客惊喜的推荐。

5. 总结

个性化推荐可以为用户提供更好的信息服务, 本文主要探讨在线旅游平台上的评论文本对旅游景点

推荐的影响,发现用户评论数据对个性化推荐性能有显著影响,特别是在多样性新颖性推荐方面有较好的表现。所以旅游平台应该更好地利用平台上的评论数据,实现更加精准的用户偏好分析和目的地景点推荐,为用户的出游活动提供有力支持,激活疫情后的旅游业。

参考文献

- [1] 李广丽,朱涛,袁天,滑瑾,张红斌.混合分层抽样与协同过滤的旅游景点推荐模型研究[J].数据采集与处理,2019,34(3):566-576.
- [2] 郑淞尹,谈国新,史中超.基于分段用户群与时间上下文的旅游景点推荐模型研究[J].数据分析与知识发现,2020,4(5):92-104.
- [3] Xin, M. and Wu, L. (2020) Using Multi-Features to Partition Users for Friends Recommendation in Location Based Social Network. *Information Processing & Management*, **57**, Article ID: 102125. <https://doi.org/10.1016/j.ipm.2019.102125>
- [4] 陈思,田敬阳.基于协同过滤算法的旅游景点推荐模型研究[J].现代电子技术,2020,43(11):132-135.
- [5] Burke, R. (2002) Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, **12**, 331-370. <https://doi.org/10.1023/A:1021240730564>
- [6] 刘海鸥,孙晶晶,苏妍嫒,张亚明.基于用户画像的旅游情境化推荐服务研究[J].情报理论与实践,2018,41(10):87-92.
- [7] Ravi, L., Subramaniaswamy, V., Vijayakumar, V., et al. (2019) Hybrid Location-Based Recommender System for Mobility and Travel Planning. *Mobile Networks & Applications*, **24**, 1226-1239. <https://doi.org/10.1007/s11036-019-01260-4>
- [8] Jiang, S., Qian, X., Mei, T., et al. (2016) Personalized Travel Sequence Recommendation on Multi-Source Big Social Media. *IEEE Transactions on Big Data*, **2**, 43-56. <https://doi.org/10.1109/TBDATA.2016.2541160>
- [9] 胡云,张舒,李慧,施珺,仲兆满.融合社会标签与信任关系的社会网络推荐方法[J].数据采集与处理,2018,33(4):704-711.
- [10] Pan, Y. and Wu, D. (2019) Personalized Online-to-Offline (O2O) Service Recommendation Based on a Novel Frequent Service-Set Network. *IEEE Systems Journal*, **13**, 1599-1607. <https://doi.org/10.1109/JSYST.2018.2883214>
- [11] Ji, J., Luo, C., Chen, X., Yu, L. and Li, P. (2018) Cross-Domain Sentiment Classification via a Bifurcated-LSTM. *Advances in Knowledge Discovery and Data Mining*, 681-693. https://doi.org/10.1007/978-3-319-93034-3_54
- [12] Cheng, C., Yang, H., Lyu, M.R., et al. (2013) Where You Like to Go Next: Successive Point-of-Interest Recommendation. *Twenty-Third International Joint Conference on Artificial Intelligence*, Beijing, 3-9 August 2013, 2605-2611.
- [13] Ziegler, C.N. and Lausen, G. (2010) Making Product Recommendations More Diverse. *Bulletin of the Technical Committee on Data Engineering*, **32**, 23-33.