北京大学学报(自然科学版) 第 52 卷 第 1 期 2016 年 1 月
Acta Scientiarum Naturalium Universitatis Pekinensis, Vol. 52, No. 1 (Jan. 2016)
doi: 10.13209/j.0479-8023.2016.011

基于评论主题的个性化评分预测模型

马春平[1,2] 陈文亮[1,2,†]

1. 苏州大学计算机科学与技术学院, 苏州 215006; 2. 软件新技术与产业化协同创新中心, 苏州 215006;
†通信作者, E-mail: wlchen@suda.edu.cn

摘要 针对现有基于评论分析的推荐算法没有充分考虑个性化的问题, 通过对评论进行主题分析, 挖掘用户的喜好, 分别建立基于用户和物品的个性化评分预测模型。在真实数据集上进行实验验证, 结果表明该模型有效地提高了推荐系统的评分预测性能。

关键词 个性化推荐; 推荐系统; 评分预测; 评论信息
中图分类号 TP391

Personalized Model for Rating Prediction Based on Review Analysis

MA Chunping[1,2], CHEN Wenliang[1,2,†]

1. School of Computer Science and Technology, Soochow University, Suzhou 215006;
2. Collaborative Innovation Center of Novel Software Technology and Industrialization, Suzhou 215006;
† Corresponding author, E-mail: wlchen@suda.edu.cn

Abstract Existing recommender systems do not take full advantage of personalization. To address this problem, a novel approach is proposed to mine the opinions and preference of users to build a personalized model for each user or item. Experimental results generated from a real data set show that the proposed approach can improve the accuracy of rating prediction.

Key words personalized recommendation; recommender system; rating prediction; review comment

传统的推荐算法主要根据用户的历史行为分析用户喜好, 并做出推荐。随着互联网的普及和电子商务的发展, 用户成为互联网主动的参与者, 并产生大量的用户个性化内容。这些内容包括用户评论、地理位置、好友关系等。传统的推荐算法主要利用用户的评分或者物品的描述性特征。相对于评分, 评论能够更加具体、准确地表达用户对物品的喜好。与用户和物品的描述信息相比, 用户评论的内容也更加丰富, 更能体现用户的个性化特征, 可以在此基础上构建更准确的个体画像。以评论“这家店虽然环境一般, 但味道正宗, 老板人也很好。”为例, 该用户给该饭店打 5 分满分。从用户的角度可以分析出, 该用户比较注重服务和口味, 对环境要求较低; 而从商户的角度可以分析出, 该商户的口味、服务一流, 但环境欠佳。可见评论的信息量远远大于评分。如果能从评论中精确地分析出用户的喜好和物品的特征, 必定能在很大程度上提高推荐精度。

但是, 评论属于非结构化文本, 由网络用户自由表达, 极具个性化特征, 不具有语法严谨、表达明确、逻辑清楚等特点, 处理起来难度较大。近年来, 情感分析和意见挖掘领域的研究者做了大量工作(具代表性的工作如文献[1–3]), 有效地从评论中挖掘出评论的情感和潜在主题。如何充分利用评论分析得到的评论情感和主题来改进推荐系统, 已成为一个重要的研究课题[4–8]。

本文首先对评论进行主题分析, 构建用户和物品的画像。在此基础上, 提出一种新的个性化评分

国家自然科学基金(61203314, 61373095)资助
收稿日期: 2015–06–07; 修回日期: 2015–08–19; 网络出版日期: 2015–09–29

预测模型。该模型针对各个用户和物品进行建模，较大程度地进行个性化处理。在大众点评数据集上进行实验验证，结果表明本文提出的模型可以有效地提高推荐系统的评分预测性能。

1 相关工作

协同过滤系统是最先得到广泛应用的个性化推荐系统，分为基于用户的协同过滤(User-based Collaborative Filtering[9])和基于物品的协同过滤(Item-based Collaborative Filtering[10])。但是，这些方法没有利用用户或者物品的语义信息，导致推荐系统的性能较低。

随着 Web2.0 的发展，在线评论逐渐进入推荐系统研究者的视野[4−7,11−16]。Ganu 等[4]通过人工标注评论的主题和情感训练 SVM 模型，将评论按角度和情感进行分类，最后综合评论中的正面评价和负面评价，做出评分预测。Qu 等[5]提出意见袋(bag-of-opinions)的概念，用来表示评论中的每条评论意见的评价词根、修饰词和否定词，利用意见袋模型和评分训练线性模型进行评分预测。这些算法都是根据用户对物品的评论，预测用户对物品的评分，还不能直接用于推荐系统。McAuley 等[7]提出利用 HFT (hidden factors as topics)将评分和评论信息结合，构建特征矩阵，利用 SVD 算法来做推荐，但无法同时考虑评论信息中的用户角度和物品角度。Wang 等[8]提出 LARA (latent aspect rating analysis)算法，首先利用自举算法(boot-strapping)获得与物品各个主题相关的情感词，然后利用 LRR (latent rating regression)算法分析用户对物品每个角度的情感和各个角度所占权重。

Zhang 等[6]利用 LDA (latent dirichlet allocation)算法[17]对评论进行主题分析，生成主题词表。根据评论中是否含有主题词，将一条评论表示成一组向量，根据用户或者物品分类，通过将这些向量平均、归一化，得到用户特征和物品特征。同时利用向量和对应的评分，通过机器学习模型训练，得到用户对物品不同的主题的权重。与上述工作相比，本文方法的最大不同之处是针对每个用户或者物品分别建模，进行个性化处理。

2 基于评论主题的个性化模型

2.1 相关定义

用户评论数据中包含 m 个用户组成的用户集合 $U = \{u_1, \ldots, u_m\}$和 n 个物品组成的物品集合 $I = \{i_1, \ldots, i_n\}$。用户–物品评分数据集可以用 $m\times n$ 阶矩阵 $\boldsymbol{R}$ 表示，R_{ui} 表示用户 u 对物品 i 的评分。C_{ui} 表示用户 u 对物品 i 的评论。C_u 表示用户 u 所有的评论集合，C_i 表示用户对物品 i 所有的评论集合，$|X|$表示集合 X 中元素的个数。$\bar{u}$ 表示用户 u 的对所有物品评分的平均值。

2.2 评论主题分析

为了分析用户评论所表达的潜在主题，本文使用 LDA 算法对用户评论进行主题分析。LDA 是一种主题模型，属于无监督学习算法，可以将文档集中每篇文档的主题按照概率分布的形式给出，并且对于每一个主题均可以找出一些词语来描述。大众点评数据集在 LDA 实验结果中的主题分布如表 1 所示，其中主题词按在该主题下的概率由大到小排列。实验主题数设置为 6，每个主题的主题词个数设置为 20，超参数 α 设置为 0.2，β 设置为 0.1，迭代次数为 1000。根据评论是否涉及各个主题，将评论表示成一组 K 维向量(K 是主题个数)，分析结果将在 2.3 节中被用于推荐系统。

2.3 用户和物品的特征表示

根据评论分析结果，对评论进行特征表示。评论 C_{ui} 的特征表示为 θ_{ui}:

$$\theta_{ui} = [\theta_{ui1}, \cdots, \theta_{uik}], k \in [1, K], \tag{1}$$

其中 K 是实验设置的主题的个数，θ_{uik} 表示用户 u 对物品 i 的评论第 k 个特征值。特征值的计算方式如下:

$$\theta_{uik} = \sum_{t=1}^{n} \theta_{uikt}, \tag{2}$$

其中，n 是各个主题下主题词的个数。若评论中包含主题词 t，则 θ_{uikt} 是主题词在 k 主题下的概率；反之，若评论中不包含任何主题词，则 θ_{uikt} 为 0。

然后，生成用户的特征表示 p_u 和物品的特征表示 q_i。用户 u 第 k 维特征用 p_{uk} 表示:

$$p'_{uk} = \frac{\sum_i \theta_{uik}}{|C_u|}, \tag{3}$$

$$p_{uk} = \frac{p'_{uk}}{\sum_k p'_{uk}}, k \in [1, K]。 \tag{4}$$

式(4)是对相应的特征进行归一化。同样，定义物品 i 第 k 维特征 q_{ik}:

表 1　基于 LDA 的主题分布

Table 1　Topics distribution based on LDA

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
好	好	好	菜	服务员	好
味道	味道	味道	味道	好	不错
甜	不错	小	不错	菜	环境
蛋糕	好吃	面	好	店	店
不错	饭	不错	鱼	团	菜
小	小	汤	辣	差	味道
奶茶	大	店	好吃	态度	感觉
面包	牛肉	好吃	香	东西	朋友
好吃	鱼	鸡	虾	钱	价格
茶	锅	大	大	少	小
家	店	肉	肉	小	地方
店	火锅	家	汤	大	大
巧克力	肉	饭	鸡	味道	口味
咖啡	套餐	辣	小	慢	贵
大	牛	价格	嫩	老板	东西
冰	酱	粥	酸	生意	服务员
感觉	新鲜	馄饨	锅	质量	餐厅
奶	羊肉	感觉	油	客人	楼
口味	汤	东西	感觉	环境	老板
香	感觉	牛肉	甜	量	家

说明：数据来自大众点评网(www.dianping.com)。

$$q'_{ik}=\frac{\sum_u \theta_{uik}}{C_i}, \tag{5}$$

$$q_{ik}=\frac{q'_{ik}}{\sum_k q'_{ik}}, k\in[1,K]。 \tag{6}$$

2.4　基准模型

本文以 Zhang 等[6]提出的模型为基准模型。Zhang 等通过对评论进行主题分析生成主题词表。根据评论中是否含有主题词，将每条评论表示成一组向量，然后生成用户和物品的特征向量表示。利用向量和对应的评分，通过机器学习模型训练得到用户对物品不同主题的权重。在对饭店评论例子中，经过上述工作得到的权重可以理解为大多数用户对饭店不同角度的偏好，比如大多数用户对饭店菜品的口味要求比较高，那么所得口味主题的权重会比较大。

但是，这种分析没有充分考虑不同用户的个性化需求，比如用户 A 为高收入者，对环境要求比较高；用户 B 为低收入者，对价格比较敏感。如果对用户 A 和 B 用同样的主题权重去预测评分，给出推荐，必然影响推荐的精准度。

2.5　个性化模型

针对基准模型的不足，本文提出个性化评分预测模型。个性化评分预测模型可以分为用户个性化(User-based)和物品个性化(Item-based)，其中用户个性化可以解释为向用户推荐其喜欢的物品，而物品个性化可以解释为为物品寻找对其感兴趣的用户。在评分预测阶段，利用用户 u 对物品 i 的评分以及用户 u 对物品 i 的评论的特征表示，可以通过线性回归模型训练特征权重，公式如下：

$$\text{User-based: } r_{ui}=\boldsymbol{W}_u^{\mathrm{T}}\boldsymbol{\theta}_{ui}+\varepsilon_u, \tag{7}$$

$$\text{Item-based: } r_{ui}=\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{\theta}_{ui}+\varepsilon_i, \tag{8}$$

其中 r_{ui} 是用户 u 对物品 i 的评分，W_u 和 ε_u 是基于用户的个性化模型训练后所得各特征权重和误差偏置，W_i 和 ε_i 是基于物品的个性化模型训练后所得各

特征权重和误差偏置。然后，对于给定的目标用户 u 和目标物品 i，由式(4)产生的用户特征和式(6)产生的物品特征模拟目标用户 u 对目标物品 i 的评论特征表示为

$$\theta'_{uik} = p_{uk} q_{ik}, \tag{9}$$

$$\hat{\theta}_{uik} = \frac{\theta'_{uik}}{\sum_k \theta'_{uik}}, k \in [1, K]。 \tag{10}$$

根据线性回归得到的权重和误差偏置以及模拟的评论特征表示，使用以下公式计算目标用户 u 对物品 i 的评分。

$$\text{User-based: } \hat{r}_{ui} = \boldsymbol{W}_u^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{ui} + \varepsilon_u, \tag{11}$$

$$\text{Item-based: } \hat{r}_{ui} = \boldsymbol{W}_i^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{ui} + \varepsilon_i。 \tag{12}$$

3 实验结果与分析

3.1 数据集

本文实验采用大众点评网(www.dianping.com)的数据集。大众点评网是中国最大的独立第三方消费点评网站。本文使用的数据集为中文数据集，全部来自上海地区，包含自大众点评 2003 年成立至 2013 年 9 月，60 万个用户对 5 万个商户的 360 万条评论。评论信息包含用户名、商户名、总体评分、评论时间以及评论文本内容。由于本文的目标是建立针对用户或者物品的个性化模型，考虑到不同用户或者不同商户的评论数量差异对实验的影响，需要对实验数据设置过滤值。例如，进行用户个性化实验时，设置过滤值为 10，表示只取数据集中评论商户数超过 10 的用户的评论；进行物品个性化实验时，设置过滤值为 10，表示只取数据集中拥有 10 条以上用户评论的商户的评论。本试验在进行基于个性化的实验以确定最佳过滤值时，按 7:1:2 的比例，随机将数据分为训练集、开发集和测试集。根据过滤值对数据过滤时，只对训练集进行过滤，保持开发集和测试集不变。

3.2 评价指标

本文采用平均绝对偏差(mean absolute error, MAE)评价算法的预测准确程度，MAE 的计算公式如下:

$$\text{MAE} = \frac{1}{n} \sum_{(u,i) \in T} |\hat{r}_{ui} - r_{ui}|, \tag{13}$$

其中 T 是测试集，$\hat{r}_{ui}$ 表示系统对目标用户 u 在物品 i 上的预测评分，r_{ui} 是真实评分，n 表示预测的次数。显然 MAE 的值越低，算法的预测精度越高。

3.3 参数选定

基于用户个性化的开发集实验结果如表 2 所示，第一列为过滤值，第二列为过滤后开发集数据的实验结果，为了与基准模型做比较，第三列为开发集总的数据集的实验结果。总的数据集实验结果生成方式为：针对目标用户 u，如果用户 u 在过滤后的数据集中，则使用个性化预测结果，反之则使用非个性化结果(基准模型的结果)。从表 2 看出，随着过滤值的增大，过滤后的数据 MAE 降低，因为用户的评论数增长有利于训练用户的特征权重。但是，过滤值的增大也导致数据集中未建模的用户数增大，因此在总的数据集上实验效果呈现 MAE 先降低后升高的趋势，而且升高的趋势越来越大。基于物品个性化的开发集实验结果如表 3 所示，实验结果 MAE 变化趋势与表 2 相同。根据实验结果，两个模型都在过滤值为 5 时取得最小值。

3.4 实验结果分析

根据上述在开发集上的实验结果，本文将个性化模型的过滤值选为 5，在测试集上的实验结果如表 4 所示。实验 1 使用 Zhang 等[6]提出的非个性化方法，实验结果 MAE 为 0.6765；实验 2 基于用户的个性化模型，实验结果 MAE 为 0.6418；实验 3 基于物品的个性化模型，实验结果 MAE 为 0.6359。本文主要讨论的两种基于个性化的模型，其异同点分析如下：基于物品个性化的实验结果与基于用户个性化实验结果趋势类似，随着过滤值的增大，过滤后的数据实验结果MAE下降，总的数据集上实验

表 2 基于用户个性化的开发集实验结果

Table 2 Development data's experimental result of User-based model

用户数	MAE	
	过滤后	未过滤
2	0.6091	0.6395
3	0.5944	0.6393
5	**0.5850**	**0.6389**
10	0.5673	0.6392
15	0.5556	0.6393
20	0.5518	0.6396
25	0.5486	0.6398
30	0.5485	0.6396
35	0.5489	0.6402

说明：粗体表示最佳实验结果。

表 3　基于物品个性化的开发集实验结果
Table 3　Development data's experimental result of Item-based model

物品数	MAE	
	过滤后	未过滤
2	0.5986	0.6383
3	0.5897	0.6381
5	**0.5802**	**0.6377**
10	0.5624	0.6380
15	0.5511	0.6382
20	0.5438	0.6388
25	0.5425	0.6393
30	0.5423	0.6397
35	0.5424	0.6409

说明：粗体表示最佳实验结果。

表 4　主要实验结果
Table 4　Main experimental result

实验序号	系统	MAE
1	Baseline	0.6765
2	User-based Personalized Model	0.6418
3	Item-based Personalized Model	0.6359

效果 MAE 呈现先降低后升高的趋势。由于数据集中商户的数量远少于用户的数量，平均每个商户拥有的评论数量远大于用户的平均评论数，更有利于训练特征权重，因此基于物品个性化的实验结果比基于用户个性化的实验结果 MAE 更低。图 1 显示在测试集上基于用户个性化和基于物品个性化实验的结果比较。在实际应用中，用户的增长远比物品的增长快得多，因此基于物品的个性化模型在解决扩展性和数据稀疏等问题上有一定的优势。

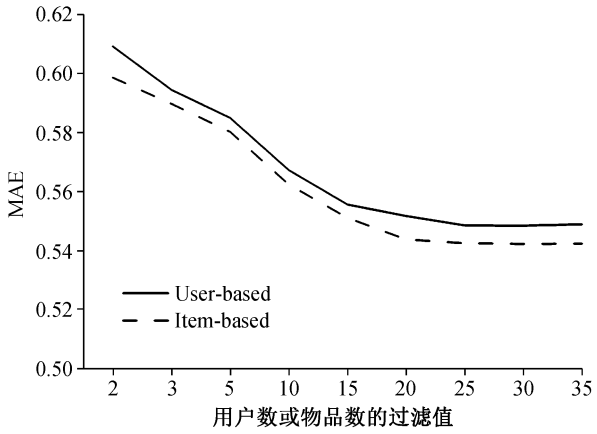

图 1　两种个性化方法实验的结果比较
Fig. 1　Comparision of two personalized approaches

4　总结

本文在对评论进行主题分析的基础上，针对用户和物品分别建立不同的个性化解决方案。经过大规模的数据实验，结果表明该方法显著地提高了评分预测的预测精度。通过比较，基于物品的个性化预测方法效果更好，并且在解决扩展性和数据稀疏等问题上有一定的优势。

参考文献

[1] Titov I, McDonald R T. A joint model of text and aspect ratings for sentiment summarization // Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies. Ohio, 2008: 308–316

[2] Brody S, Elhadad N. An unsupervised aspectsentiment model for online reviews // Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, 2010: 804–812

[3] Jo Y, Oh A H. Aspect and sentiment unification model for online review analysis // Proceedings of the fourth ACM International Conference on Web Search and Data Mining. Hong Kong, 2011: 815–824

[4] Ganu G, Elhadad N, Marian A. Beyond the stars: improving rating predictions using review text content // The 12th International Workshop on the Web and Databases. Providence, Rhode Island, 2009: 1–6

[5] Qu Lizhen, Ifrim G, Weikum G. The bag-of-opinions method for review rating prediction from sparse text patterns // Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, 2010: 913–921

[6] Zhang Rong, Gao Yifan, Yu Wenzhe, et al. Review comment analysis for predicting ratings // The 16th International Conference on Web-Age Information Management. Qingdao, 2015: 247–259

[7] McAuley J, Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text // Proceedings of the 7th ACM conference on Recommender systems. Hong Kong, 2013: 165–172

[8] Wang Hongqing, Lu Yue, Zhai Chengxiang. Latent aspect rating analysis on review text data: a rating regression approach // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, 2010: 783−792

[9] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews // Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work. Chapel Hill, 1994: 175−186

[10] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms // Proceedings of the 10th International Conference on World Wide Web. Hong Kong, 2001: 285−295

[11] Dave K, Lawrence S, Pennock D M. Mining the peanut gallery: opinion extraction and semantic classification of product reviews // Proceedings of the 12th International Conference on World Wide Web. Budapest, 2003: 519−528

[12] Devitt A, Ahmad K. Sentiment polarity identification in financial news: a cohesion-based approach. // Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies. Prague, 2007: 984−991

[13] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques // Proceedings of the ACL-02 Conference on Empirical methods in Natural Language Processing — Volume 10. Philadelphia, 2002: 79−86

[14] Pang B, Lee L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales // Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, 2005: 115−124

[15] Goldberg A B, Zhu X. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization // Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing. New York, 2006: 45−52

[16] Snyder B, Barzilay R. Multiple aspect ranking using the good grief algorithm // Human Language Technologies: The 2007 Annual Conference of the North American Chapter of the Association for Computational Linguistics. New York, 2007: 300−307

[17] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation. The Journal of Machine Learning Research, 2003, 3(1): 993−1022