Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 199 (2022) 871–878

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021)

Movie Recommendation Algorithm Based on Sentiment Analysis and LDA

Yilin Zhang[a,b], Lingling Zhang[a,b,1]

[a] School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China
[b] Key Laboratory of Big Data Mining and Knowledge Management,Chinese Academy of Sciences, Beijing, 100190, China

Abstract

Traditional recommendation algorithms have problems such as data sparseness and not paying attention to the diversity of recommendation results. In this paper, we use LDA to extract topics of comments about movies, and identify the emotional tendencies related to topics. As a result, we enrich user interest model and product feature model based on emotional tendencies to improve content-based recommendation algorithms. Most of prior work on applying sentiment classification to recommendation systems only consider the use of sentiment dictionaries to judge polarity, and adopt pattern matching methods to identify features. This paper uses BERT to train sentiment classification models and uses LDA to extract topics. The algorithm is run on the movie review database crawled from Douban, and the experimental result showed that the diversity of recommendation lists had been significantly improved.

Keywords: Sentiment Analysis; Topic Extraction; Content-based Recommendation; Movie Reviews

1. Introduction

With the rapid development and popularization of the Internet, users are facing the problem of information overload. The existence of a large amount of redundant information has seriously affected people's accurate selection of information. Helping users to screen out invalid information and quickly providing users with content that may be of interest has become an important requirement in recent years. Traditional search engines presented a single search result based on static keywords, which can not meet individual needs of users. In order to effectively filter information and provide content that is more in line with users' personalized characteristics, the academia had introduced recommendation technology.

Personalized recommendation systems use artificial intelligence, data mining and other related technologies on the Internet to mine data related to users' past behaviors, recommend products that predicted to be of great interest to users [1]. With the development of artificial intelligence, big data and cloud computing technologies, limitations of traditional recommendation algorithms had become increasingly obvious, many recommendation algorithms that incorporate emerging technologies had emerged.

*Corresponding author. Tel.: +86-010-8268-0676 ; fax: +86-010-8268-0676 .
E-mail address: zhangll@ucas.ac.cn.

10.1016/j.procs.2022.01.109

There are two main traditional recommendation algorithms: content-based recommendation and collaborative filtering recommendation. Content-based recommendation generally includes the following three steps: establish a product description model, establish a user interest description model, and generate a recommendation list; the term "collaborative filtering" first appeared in the literature written by Goldberg D et al. [2], the first collaborative filtering recommendation system, Tapestry, is introduced in detail. The collaborative filtering method calculates the similarity between users or items and combines historical data to predict users' evaluations of unrated items.

With the development of machine learning, more and more recommendation algorithms introduced machine learning to build the core recommendation model. Models such as logistic regression, factorization machines and gradient boosting trees had appeared after collaborative filtering algorithm were widely used in recommendation systems. In 2016, with the introduction of a series of deep learning recommendation models such as Deep Crossing, Wide&Deep, FNN and PNN, the field of recommendation has fully entered the era of deep learning. The recommendation model constructed based on deep learning has strong expressive ability and flexible structure, which makes up for the shortcomings of traditional recommendation models, it had gradually become the research focus of industry and academia.

Various recommendation models are suitable for different application scenarios and data characteristics. Therefore, choosing a suitable core recommendation algorithm according to the application scenario is the key to building a successful recommendation system. At the same time, how to efficiently utilize user information, scene information, and historical behaviors to improve the efficiency of feature engineering is an important upstream task for training recommendation models.

Movie recommendation is one of the most classical application areas of recommendation algorithms. Traditional movie recommendation algorithms make full use of ratings and extend feature model based on partial user information. However, traditional algorithms usually only use tag data to describe movie features, ignoring the massive user evaluation information collected by UGC (User-generated content) websites. User reviews are unstructured characteristics, and emotional tendencies contained in them are usually associated with movie features. Recognizing the emotional tendencies of users about products' features helps enrich the feature engineering.

Established in 2005, Douban is famous for its evaluation and communication of book, audio, and video. It is a typical UGC website. So far, Douban has collected hundreds of thousands of movies, has tens of millions of users and massive movie review data. This paper proposes a movie recommendation algorithm based on sentiment analysis and LDA topic model, using about 285,000 movie evaluation data of 640 movies crawled from Douban to evaluate the performance of the proposed algorithm.

2. Related research

2.1. Sentiment analysis

2.1.1. Definition of text sentiment analysis

In 2002, Bo Pang et al. [3] first proposed the concept of emotional polarity classification, marking the official start of systematic research in this field.

Sentiment analysis is also called opinion mining, opinion analysis, subjective and objective analysis, etc. [4]. Sentiment analysis of texts can not only mine the valuable information hidden in the text, but also assist decision-making and optimize the algorithm.

Comment refers to the positive or negative evaluation of the product generated by users in the form of text, voice and other forms through the Internet, it is an important source of information for obtaining user preferences [5]. Sentiment analysis of online reviews refers to using text mining and other technologies to analyze the meaning and structural characteristics of text, so as to identify the emotional polarity. Studies have shown that emotional factors play an important role in determining users' behavior and preferences. In the process of recommending information flow, users' emotional tendency and state should be fully taken into account to better meet the different needs of users and realize personalized services [6].

2.1.2. Main methods of text sentiment analysis

At present, there are two methods for text sentiment analysis. The first one is based on a structured dictionary. The vocabulary in the dictionary usually includes sentiment words, degree adverbs, and negative words. The sentiment score of the text is finally calculated by marking words' part of speech and emotional polarity. The second one is based on machine learning, which uses the idea of text classification to process sentiment analysis problems, and finally obtains a model that can predict the emotional polarity[7].

The sentiment analysis method based on sentiment dictionary is relatively simple, it identifies positions of sentiment words, negative words and degree adverbs, and combines the sentiment scores of phrases to get the polarity of sentences.

The main problem of the sentiment analysis method based on sentiment dictionary is that the capacity of sentiment dictionary is always limited. If there are words in the text that do not appear in the sentiment dictionary, the accuracy of sentiment classification results will be reduced. The update speed of online vocabulary is very fast, and the cost of maintaining sentiment dictionary in time is also very high. Therefore, the accuracy of the sentiment analysis method based on the dictionary is usually low.

Methods based on machine learning can be divided into supervised machine learning methods and weakly supervised deep learning methods. The most commonly used machine learning algorithms are SVM and NB, labeled corpus is used to train the model, and then we use models to classify emotions.

Traditional machine learning methods require manual labeling of a large number of emotional words and samples, which will consume a lot of time. The efficiency and quality can't meet the work requirements in the big data era. Deep learning methods have relatively less demand for labeled data sets, and can discover more abstract high-level features of text through neural networks, and ultimately improve the accuracy of text classification.

The current deep learning models include CNN, RNN, LSTM, BiLSTM, GRU and attention mechanism, etc. [8]. Sentiment analysis models trained by deep learning algorithms have good performance on various data sets. In sentiment analysis tasks, LSTM (Long Short-Term Memory) is more widely used.

In 2018, Devlin J [9] from Google proposed the BERT model, which is a pre-trained language model based on a multi-layer two-way Transformer. Compared with traditional word vector tools such as Word2vec, BERT can be integrated into specific downstream NLP tasks in addition to providing word vectors as features.

BERT achieves the best performance on a large number of sentence-level and Token-level tasks, and is superior to many systems with task-specific architectures, including tasks in the field of sentiment analysis [10]. Therefore, since the BERT pre-training model was proposed, it is often applied to sentiment analysis tasks. The proposed algorithm is based on Google's open source BERT model, we used movie reviews on Douban (5000 positive reviews, 2969 neutral reviews, and 5000 negative reviews) as training set, finally obtained the sentiment classification model applied in the movie field.

2.2. LDA

Blei of Princeton University proposed the LDA model [11], which is a document topic generation model. The LDA has three layers: word layer, topic layer, and document layer. LDA is an unsupervised machine learning model, as well as a domain-independent and language-independent topic extraction algorithm. Compared with traditional keyword extraction algorithms, LDA can describe the content of documents more comprehensively and accurately.

LDA is fast and accurate in topic extraction tasks of short texts. In movie recommendation system, reviews are usually short texts of 20-200 words, which are short in length and serious in colloquialism. It is not suitable for traditional classification algorithms utilizing machine learning. Massive review data also puts forward higher requirements of effectiveness of the extraction method, so it is very proper to choose LDA as the topic extraction strategy of reviews in movie recommendation.

2.3. Recommendation algorithms

2.3.1. Traditional recommendation algorithms

The description content of the item is usually an important basis for recommendation, and the content-based recommendation algorithm can effectively solve the cold start problem [12]. There are many manifestations of content, description in natural languages of features is the most common method. How to automatically extract item's features and user interests from natural expressions is the key to improving recommendation efficiency.

Collaborative filtering recommendation algorithms can be divided into two categories, namely user-based collaborative filtering algorithm and item-based collaborative filtering algorithm.

Content-based recommendation and collaborative filtering recommendation have the problems of data sparseness, cold start, lack of domain knowledge, difficulty in balancing timeliness and recommendation quality, and not paying attention to the diversity of recommendation. Sentiment analysis of comments or other text data existing in the application scenarios can not only help improve the quality of recommendation, but also solve limitations of traditional recommendation algorithms to a certain extent.

2.3.2. Recommendation algorithms based on machine learning

Logistic regression model is the most classic machine learning model. Compared with collaborative filtering recommendation that only uses the interactive information of users and items, logistic regression model can integrate characteristics of user, item and scene to generate more comprehensive recommendation results. The LR model sorts items by predicting the probability of being a positive sample, and converts recommendation problem into a click-through rate estimation problem. However, the LR model has a weak expressive ability and can only use a single feature. The POLY2, FM, and FFM models try to do feature crosses to generate high-dimensional combined features.

Since 2016, Google, Alibaba, Microsoft and other well-known Internet companies have proposed a series of deep learning recommendation models. The deep learning recommendation model draws on and integrates the achievements of deep learning in image, speech, natural language processing, etc., based on application scenarios and data characteristics, it can flexibly adjust the model structure to adapt to the recommendation needs in different business contexts. The Deep Crossing model [13] is a complete application of the deep learning framework in the recommendation system. It uses the classic deep learning framework of "Embedding layer + multiple hidden layers + output layer" to pre-complete automatic deep crossover of features. The Wide&Deep model [14] proposed by Google in 2016 is a hybrid model composed of a single-layer Wide part and a multi-layer Deep part. Since its proposal, it has exerted a huge influence in the industry. Alibaba introduced the attention mechanism into the deep learning network, and successively proposed the DIN model [15] and the DIEN model [16].

2.3.3. Application of sentiment analysis in recommender system

Existing recommendation algorithms usually do not consider sentiment factors and only consider quantitative scoring data, but this type of data is sparse when there are a large number of products and users. In various systems that apply recommendation technology, user reviews construct a component that cannot be ignored. Quantifying user reviews through sentiment analysis technology makes up for the biggest shortcoming of existing recommendation algorithms.

The core of collaborative filtering recommendation is to accurately define the similarity of interest between users. On the basis of traditional algorithms, Wang Wei et al. [5] considered the impact of online comments on user similarity recognition. For large-category product recommendation, only the emotional polarity of reviews was identified, for similar products, feature-emotional polarity pairs were extracted to identify user's emotional tendency for each feature. Lei Ming et al. [17] used movie feature mining and sentiment analysis on user's reviews to calculate the user's emotional tendency for different movie features, and more accurately measured interest similarity, the recommendation results better met users' needs. Hongyan Liu et al. [18] considered the difference between user ratings and user opinions, and recommended according to users' preferences for different features of items, the proposed PORE algorithm achieved good results on the online review data set of hotels.

The key to introducing emotion factor is to accurately identify product features and determine the emotional orientation of the evaluation. Lei Ming et al. [17] used the PMI algorithm to extract movie features in the article, and calculated the emotional orientation of short reviews based on dictionary. Wang Wei et al. [5] used a feature extraction algorithm based on pattern matching [19], obtained product feature-viewpoint pairs, and judged emotional polarity based on opinion words.

The recommendation algorithm proposed in this paper used LDA to extract the evaluation topics of comments. Compared with the traditional pattern matching method, the accuracy rate is higher, and there is no need for syntactic and lexical analysis. At the same time, based on the open source code of the Google BERT model, the sentiment classification model for film reviews was trained, the accuracy reached 79.6%, avoiding the problem of low coverage of emotional words when labelled manually.

3. Algorithm

3.1. Topic extraction

First, use LDA to extract topics of approximately 285,000 reviews in the movie review database one by one, set the number of generated topics to 1, and count the frequency of occurrence of each topic word. The word cloud of themes about films is drawn in Figure 1.

Fig. 1: Word cloud of themes

Filter the subject words with a word frequency greater than 0.2%, and filter out vocabulary that has nothing to do with the evaluation angle, such as nouns (movies and films), adjectives (like, good-looking, good, classic), and thematic words describing movie categories (love, life, and war).

The subject words with similar meanings were grouped into one category, and finally 25 evaluation themes were summarized in Table 1.

Table 1: Topic words

ID	Topic	The number of occurrences
1	故事	10451
2	剧情、情节	7374、3112
3	演员	4042
4	结局、结尾	5362、4325
5	演技、表演	4218、2690
6	导演	3879
7	年代、时代	1476、1758
8	节奏	2944
9	画面、场面	3308、1136
10	镜头	2554
11	角色、人物、主角	2183、1356、1524
12	男主、男主角	2152、1644
13	编剧、剧本	2104、1037
14	配乐、音乐	1886、2051
15	细节	1826
16	女主、女主角	1368、1154
17	原著	1070
18	题材	1489
19	特效	2411
20	风格	1159
21	台词	778
22	逻辑	1258
23	背景	870
24	摄影、拍摄	759、683
25	开头	956

The frequency of the topic word ”story” with the highest word frequency in the database is 3.7%, and the frequency of the topic word ”beginning” with the lowest word frequency is 0.27%.

3.2. Judging the emotional tendency

This paper used a three-category sentiment analysis corpus of movie reviews (5000 positive, 5000 negative, and 2969 neutral) as the training set, utilized the pre-trained BERT Chinese word vector model released by Google to train the sentiment classification model. The training parameters were set as Table 2 indicated.

Table 2: Training parameters

Parameter	Value
max_eq_length	128
train_batch_size	8
learning_rate	1e-5
num_train_epochs	10

Based on the trained sentiment classification model, the emotional tendency of the text in the recommendation system can be divided into positive, negative and neutral. We also evaluated dictionary-based method and classificaiton model trained by NB algorithm, the performance of three models is displayed in Table 3.

Table 3: Performance of models

Indicator	Polarity	Dictionary	NB	BERT
Precision	negative	0.502	0.677	0.832
	positive	0.488	0.727	0.831
	neutral	0.246	0.729	0.714
Recall	negative	0.459	0.818	0.841
	positive	0.756	0.787	0.815
	neutral	0.046	0.395	0.725
F1-score	negative	0.48	0.741	0.836
	positive	0.593	0.756	0.823
	neutral	0.077	0.513	0.72

Due to the small number of neutral evaluations in the training set, the accuracy and recall of the neutral category were lower than positive and negative, but the accuracy rate that is higher than 70% has basically met the requirement of the movie recommendation system.

3.3. Construct description models

Different themes have obviously different importance to users. According to the main idea of TF-IDF, the importance of a topic word increases proportionally with its frequency in user reviews, and at the same time decreases in inverse proportion with its frequency in the entire evaluation corpus. This paper introduced TF-IDF method to describe the importance of the topic, and the calculation method is as following:

$$TF_{iw} = \frac{n_{iw}}{n_i} \tag{1}$$

$$IDF_w = log\frac{N}{n_w + 1} \tag{2}$$

$$TF - IDF_{iw} = TF_{iw} \times IDF_w \tag{3}$$

n_{iw} represents the total score related to $topic_w$ in the movie evaluation records written by U_i (the emotional tendency of the clause related to the topic term is positive, then the score is 1, neutral is 0.2, negative is 0.6). n_i represents the total number of evaluations of U_i, N represents the total number of short reviews in the corpus, and n_w represents the number of short reviews related to $topic_w$.

$TF-IDF_{iw}$ can express the importance of $topic_w$ in the user's interest, and the normalized topic word weight vector can be used to describe the user's interest.

$$C_i = (I_{i1}, I_{i2}, ..., I_{iw}, ..., I_{itot}) \tag{4}$$

$$I_{iw} = \frac{TF-IDF_{iw}}{\sum_{j=1}^{tot} TF-IDF_{ij}} \tag{5}$$

tot represents the total number of topic words, I_{ij} represents the weight of $topic_j$ in the interest of U_i, and C_i is U_i's interest description vector.

The description vector of product feature corresponds to the user interest description vector. The movie feature description vector MC_j corresponding to C_i is calculated similarly by replacing U_i by $Movie_i$. When calculating n_{iw}, the positive score is 1, the neutral is 0.2, and the negative is -1.

3.4. Generate recommendation list

The target user is U_i, the number of movies to be recommended is k, and the similarity between I_j and U_i is defined as $sim(U_i, I_j)$.

$$sim(U_i, I_j) = \frac{\sum_{k=1}^{tot} C_{ik} \times MC_{jk}}{\sqrt{\sum_{k=1}^{tot} C_{ik}^2} \times \sqrt{\sum_{k=1}^{tot} MC_{jk}^2}} \tag{6}$$

Finally, similarity scores are sorted, and the k movies with the highest similarity to the target user are selected as the recommendation list.

3.5. Empirical Research

The description vector of user's interest was constructed by taking the user ”Long Xiandai” who had more evaluation records in the evaluation corpus as an example. The interest description vector of the user ”龙如今” is shown in Table 4. Then we calculated the feature description vectors of movies.

Table 4: The interest description vector

User name	Interest description vector
龙如今	(0.010, 0.177, 0, 0.044, 0, 0.062, 0.044, 0, 0.576, 0,0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.086, 0, 0, 0, 0)

The cosine similarity between the product feature vector and the user's interest vector is calculated to measure the degree of user's interest in specific product. The Top-3 recommendation list generated for user ”龙如今” is shown in Table 5.

Table 5: Top 3 recommendation list

Movie Name	Similarity
流感	0.925
摩纳哥王妃	0.920
火烧圆明园	0.915

The traditional content-based recommendation algorithm only pays attention to movie tags, and ignores the evaluation themes that appear repeatedly in comments such as ”story”, ”plot” and ”lines”. Topics frequently appearing in comments can reflect the degree of attention on different features of movies paid by users.

For user ”龙如今”, he had made reviews about feature movies, science fiction movies, and comedies. Based on traditional algorithms, only movies of those types will be recommended to the user. According to his reviews, he cared a lot about plots and scenes. The recommendation list generated by our proposal contained ”The Flu”, ”Grace of Monaco”, and ”Burning of Imperial Palace”, which adapted to his concerned features about movies.

The recommendation algorithm based on sentiment analysis and topic model proposed in this paper effectively utilized comment data and mined a lot of hidden information from online reviews.

4. Conclusion

Based on the regularity of feature words in the movie review field, this paper used review corpus from Douban and the LDA model to extract topics, finally obtained 25 high-frequency evaluation topic categories in this field, which provided a reference for feature recognition in movie recommendation tasks.

This paper proposed a movie recommendation algorithm based on sentiment classification and topic extraction. It used the topic model and sentiment analysis techniques to enrich user interest description model and product feature description model, and generated recommendation lists by ranking similarity between user interest and product features. The algorithm proposed in this paper organically combined text mining with recommendation, and used part of the neglected text information to improve recommendation accuracy.

However, the traditional similarity calculation method does not consider feature intersection, and ignores the importance of high-dimensional combination features in user's interest and product's characteristics. In the future, it is possible to consider the emotional polarity toward a specific topic as a feature, enriching the input layer of the deep learning recommendation model, and enhancing the expression ability of the recommendation model.

Acknowledgements

The work is supported by the National Natural Science Foundation of China (No. 72071194). The authors are very grateful to the anonymous reviewers for their constructive comments and suggestions that have led to an improved version of this paper.

References

[1] M. Balabanović, Y. Shoham, Fab: content-based, collaborative recommendation, Communications of the ACM 40 (3) (1997) 66–72.
[2] D. Goldberg, D. Nichols, B. M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, Communications of the ACM 35 (12) (1992) 61–70.
[3] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques, arXiv preprint cs/0205070 (2002).
[4] W. Li, Sentiment analysis and cognition, Computer Science 37 (7) (2010) 11–15.
[5] W. Wang, H. Wang, Y. Meng, Research on collaborative filtering recommendation algorithm: Considering the emotional tendency of online reviews, System Engineering Theory and Practice 34 (12) (2014) 3238–3249.
[6] Z. Jiang, Y. Jin, et al., Research on recommendation algorithm combined with review sentiment analysis, Application Research of Computers (05) (2016) 1312–1314.
[7] K. Wang, R. Xia, A survey of automatic construction methods of sentiment dictionary, ACTA Automatica Sinica 42 (4) (2016) 495–511.
[8] A. Yadav, D. K. Vishwakarma, Sentiment analysis using deep learning architectures: a review, Artificial Intelligence Review 53 (6) (2020) 4335–4385.
[9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[10] A. Adhikari, A. Ram, R. Tang, J. Lin, Docbert: Bert for document classification, arXiv preprint arXiv:1904.08398 (2019).
[11] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, the Journal of machine Learning research 3 (2003) 993–1022.
[12] W. Zhang, J. Wang, A collective bayesian poisson factorization model for cold-start local event recommendation, in: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015, pp. 1455–1464.
[13] Y. Shan, T. R. Hoens, J. Jiao, H. Wang, D. Yu, J. Mao, Deep crossing: Web-scale modeling without manually crafted combinatorial features, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 255–262.
[14] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, et al., Wide & deep learning for recommender systems, in: Proceedings of the 1st workshop on deep learning for recommender systems, 2016, pp. 7–10.
[15] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, K. Gai, Deep interest network for click-through rate prediction, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1059–1068.
[16] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu, K. Gai, Deep interest evolution network for click-through rate prediction, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 5941–5948.
[17] M. Lei, M. Zhu, Application of sentiment analysis in movie recommendation system, Computer Engineering and Applications 52 (10) (2016) 59r63–107.
[18] H. Liu, J. He, T. Wang, W. Song, X. Du, Combining user preferences and user opinions for accurate recommendation, Electronic Commerce Research and Applications 12 (1) (2013) 14–23.
[19] H. Song, Y. Fan, X. Liu, D. Tao, Extracting product features from online reviews for sentimental analysis, in: 2011 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), IEEE, 2011, pp. 745–750.