International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

Personalized Aspect based Recommender System in Social Networks

M. Venu Gopalachari [1], Kiran Kumar Varaka [2]
[1]Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, India
[2]Department of Computer Science & Engineering., Lords Institute of Engineering and Technology, Hyderabad, India

Abstract— The social network in current age populates opinion on various products, services, persons through ratings and reviews. Unlike ratings, the reviews can help users to elaborate their opinion and share the extent of consumption experience in various aspects. Though some of the existing recommendation systems have been using the user reviews, the interpretability of the models were erroneous. The aspect based sentiments exploited with qualitative model can lead to the qualitative recommendations. This paper proposed a modeled that exploits the correlation among aspects in review using deep neural network. It formulates the preference of users' on aspect category as a bipartite relation, represents it as a location-aspect category bipartite graph, and models the explainable recommendation with the notion of ordered dense subgraph extraction using bipartite core-based and ranking-based approaches, and it evaluates the generated recommendation with three datasets.
Keywords— Aspect Based Sentiments, Social Networks, Recommender system, and spatial data

I. INTRODUCTION

Now a day web services are populating ratings and review texts through facilitating users to provide their experience. On the other hand Location Based Social Networks (LBSN) are been a useful platform to share experience on different factors of interest such as price, accessibility, quality of service, quality of product etc. There exist different polarities of sentiments on different parts of the transaction. For instance, the review "Mobile display is excellent but need huge size of memory" covers two aspects display and memory about the product where first one is positive and second one is negative.

The research on the exploitation of different factors of LBSN for an efficient recommendation has been quite popular in the last decade [1]. Most of the studies have just focused on non-text attributes, such as the categorical, temporal, spatial, and social aspects. Most of the existing systems have also been concealing the reasons behind the recommendation and have been less transparent and less interpretable (i.e. the factors used to get the recommendation are hidden from end users). Contrary to that, some of the studies [2], have already claimed the usefulness of explainability on the persuasiveness of users towards real-world systems. With the abundance of LBSN datasets and affordable computing cost, the research community has anticipated the necessity of incorporation of additional attributes for a generic, scalable, and explainable recommendation. Some of the motivating studies [3] have convincingly emphasized the importance of explanations. The similarity-based approaches have proposed the user-based neighbor style (e.g., users with similar interest have purchased the following items) explanations. The item-based neighbor style (e.g., items similar to you viewed or purchased in the past), influence style (how the users' input have influenced the generation of recommendation), and keyword-style (items that have similar content to purchase history) can be some other variants of explanations.

To the best of our knowledge, this paper is the first to explore the problem of aspect-based personalized explainable POI recommendation. There are many factors that make this problem challenging and interesting. First of all, the aspect extraction from ambiguous and noisy text itself is a difficult task. As there can be many aspect terms, efficiently organizing them into relevant categories (e.g., food, service, etc.) is also nontrivial. The task of personalized recommendation is challenging as we need to model individual user preferences. The aspect-based personalized explanation generation becomes challenging as it needs to deal with the sentiments associated with the aspects, and also the individual user preferences and item features to get the relevant explanation.

The ease of adaptation of arbitrary continuous and categorical attributes in a scalable manner makes the Convolutional Neural Networks (CNN) a good candidate for classification problems (e.g.,[4]. This also makes them ideal for review-aspect category correlation problem. This paper first formulates the problem of review and aspect category correlation using Convolutional Neural Networks. This simplifies the process of mapping the user sentiments to the (location, aspect category) tuples and modeling the users' aspect category preferences as the aspect category-location bipartite relation. We represent such a bipartite relation using a bipartite graph and extract the ordered aspect category preference of a user. The contributions of this paper are to exploit the aspects from review text and categorize them into different aspect categories and to formulate the review-aspect category correlation using deep neural network. The proposed model outperform the existing recommender system w.r.t various evaluation measures.

II. RELEATED WORK

The problem of aspect extraction from review text has been used in various applications such as rating prediction [5], recommendation and aspect sentiment summarization [6].
Yang et al. exploited the tips from Foursquare to extract user preferences. They relied on a sentiment lexicon (e.g., SentiWordNet)-based approach and defined the preferences based on the tips, check-ins, and social relation to generate recommendations. They did not fully exploit the preferences at aspect level and also had no provision of recommendation explanation. Wang et al. [8] exploited multi-modal (i.e. text, image, etc.) location semantic similarity. The topics extracted using LDA [9] was used to find similar locations. Their model also did not focus on the aspect level preference modeling, and recommendation explanation.

Zhang et al. [7] exploited user opinions from the tips and fused tip polarities, social links, and geographical information for POI recommendation. Though their fused model was claimed efficient for polarity prediction of tips, and for location recommendation, the recommendation was not generated for individual aspects, and had no provision for the explanation. Covington et al. [10] used DNN for Youtube video recommendation. They first applied a module to filter out potential candidates and then used a deep network for the recommendation. They incorporated different factors, such as users' activity history, demographics, etc., but did not incorporate the opinions from user comments, and also did not have any provision of recommendation for each aspect category. Manotumruksa et al. [11] used word embedding to model the users and venues. Though they used context attributes from review texts, provision for additional attributes (such as location category, check-in time, etc.), and sentiment polarity remained unexplored.

Recently, Zheng et al. adapted to exploit the user reviews and to map the user and item feature vectors into same space and estimated the user-item rating. Our model has following major differences from other works (i) it uses the sentiment polarity of reviews at the sentence level rather than the whole review text, (ii) it learns to classify each review sentence into aspect categories, and models users and places using these aspect categories and embedding of additional features (e.g., the location category, check-in time, etc.), and (iii) it efficiently exploits bipartite core extraction and ranking method to segregate the aspect categories and relevant locations for explainable recommendation.

Chen et al. [12] personalized ranking based tensor factorization model and used phrase-level sentiment analysis across multiple categories. They extracted aspect-sentiment pairs from review text, and used Bayesian Personalized Ranking to rank the features mentioned by a user in her review text. Finally, the feature wise preference of a user was derived using the user-item-feature cube and rank of the feature. A similar study used matrix factorization to estimate the missing values and the recommendation was made by matching the most favorite features of a user with the matching properties of the items. They used simple text templates to generate a feature-based explanation of positive and negative recommendations for an item. However, incorporation of additional features (for instance, location category) was not explored in their research.

Lawlor et al. [13] exploited sentiment wise explanation to explain why a place might or might not be interesting to a user. For every aspect, they also compared the recommended place to the alternatives and provided the explanation (for instance, better (worse) than 90% (20%) of alternatives for room quality (price), etc.). However, they simply relied on the frequency of aspects of locations and the users to get such relation and the incorporation of additional features for location prediction remained unexplored. He et al. [14] modeled the user-item-aspect as a tri-partite graph and used the graph-based ranking algorithm to find the most relevant aspects of a user that match with the relevant aspects of places. The common relevant aspects were used in the explanation.

Most of the studies were tightly coupled to the aspects and their sentiments. They analyzed the influence of all the aspects together. The influence of aspects among each other can have some adverse impact on the personalized recommendation. For instance, a place that is good in “memory” category might be opposite in “display” category. A user who just cares about “memory” aspect might ignore some “display” related problems in that place. So we need to minimize the influence of aspects among each other. This work attempt to fill this gap by the concept of bipartite graph and bipartite core extraction. For a user, the most dense subgraph represents the most preferred aspect and the places that are most popular for this aspect. The dense sub graph extraction is followed by disconnecting the edges within the dense subgraph.

III. METHODOLOGY

The proposed system handles the issues with aspects and try to figure out the sentiments using a sequence of operations involving review processing, aspect extraction, categorization, training data preparation, model building and recommendation generation. As part of review preprocessing, for every user, the review texts posted for every location are segregated by the ratings and are organized in separate text files. As the user rating is for whole review text rather than a single sentence, we use VADER [15] to correct the sentiment polarity of individual review sentences. This approach is reasonable because every review sentence might deal with different aspects and might have different sentiment. The pre-processed review sentences are fed to the aspect extraction module to extract the aspects defined in two tasks. The first task filters out the nouns and noun phrases using some experimentally set frequency threshold. It has been found that most of the reviews focus on a set of topics and this approach can easily capture such topics. The second task is to use a rule-based approach that adapts the dependency parsing [17] to handle the aspects missed in the previous step. For the sake of ease of computation, it is necessary narrow down numerous aspects to few well-known aspect categories. The next step is preparation of training data for sentence based aspect category, in which, the review text after aspect extraction is labeled by the aspect category which has the nearest neighbors' w.r.t. its aspects. The distance between aspect words and these aspect category words from the WordNet are used to assign the closest possible label. The sentences with multiple aspect terms get multiple aspect categories.

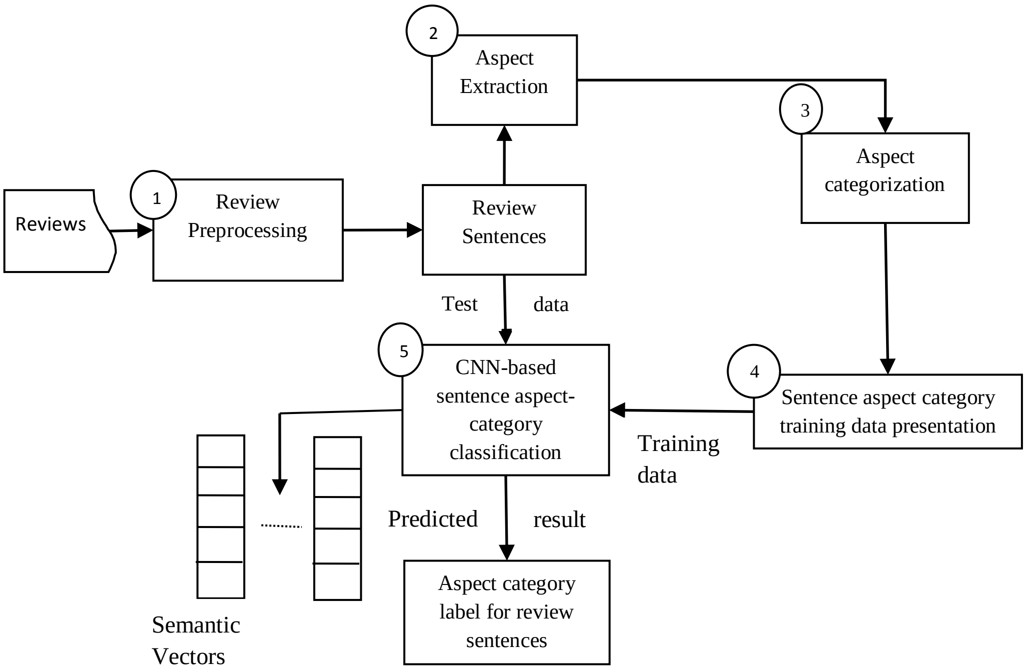

Figure 1: Architecture of proposed Model

The proposed system used a classifier which is a simple binary classifier that learns to classify a review sentence into different aspect categories. As we have more than two aspect categories, we used a simple CNN based sentence classifier for each of the aspect categories. This binary classifier can classify a review sentence to an aspect category or not to that aspect category. The input to this classifier is word embedding, which maps every word to a uniform size vector in a latent feature space, of review sentences [18]. For every user, we get a set of sentence feature vectors which is simply an embedding of her preferred aspects. Similarly, for every location, the sentence feature vector is simply an embedding of the aspects which were specified in its reviews. As a place might be positively or negatively mentioned for some aspects, we also use the sentence polarity to identify the user's sentiment polarity for the location's feature vector. The outcome of this is a bipartite relation between user review and the aspect categories. This bipartite relation can be used to define the location-aspect category tuples and user-aspect category tuples. Finally the recommendation model is used to generate set of suggestion based on location data, categories etc. along with review text. For every user, such vectors are concatenated to the review vector of the place, obtained from the CNN-based classifier. The same process is repeated for every location. Every user tends to mention some opinion on preferred aspects in her reviews and every place is mentioned about the aspects it was reviewed for. A positive review rating implies positive experience on relevant aspects and a negative review rating implies the opposite. After getting the place-aspect category bipartite relation from CNN-based classifier, we can represent the user-aspect category preference as a bipartite graph and can generate the recommendation explanation by extracting the densest subgraphs from this bi-partite graph.

Finding the subsets of aspect categories with highest similarity score not only facilitates explanation of recommendation but also provisions clustering of the users who have similar preferences on aspects (even in case of absence of explicit social links) and generate a group recommendation. It can be used to generate preference wise recommendation. This can also facilitate the clustering of places that are preferred for similar aspects (for instance, the cluster of mobile phones that are popular for some set of aspects).

IV. RESULT AND DISCUSSION

The real-world datasets, as show in table 1, used (see Table 1) to evaluate the proposed models. Table 1 shows that in all three datasets, most of the users tend to give high (positive) ratings to the places. The top-10 terms of different aspect categories are illustrated in Table 2. The evaluation is made on individual components such as, aspect extraction and categorization, Sentence-aspect-category classification, in terms of accuracy.

Table 1: Statistics of the data sets

Attributes	Yelp	Trip Advisor	AirBnB
Reviews	2,225,813	246,399	570,654
Users	552,339	148,480	472,701
Places	77,079	1,850	26,734
Words	302,979,760	43,273,874	54,878,077
Sentences	18,972,604	2,167,783	284,1004
Average Sentence/review	8.53	8.79	4.98
Average Word/review	136.15	175.62	96.16
Average Reviews/user	4.03	1.66	1.20
Average Reviews/place	28.87	133.18	21.34
4, 5 stars	591,618 and 900,940	78,404 and 104,442	479,842
1, 2 stars	260,492 and 190,048	15,152 and 20,040	5,766

Table 2: Top-10 terms in different aspect categories

Category	Terms
Price	cash, redeem, cheap, expensive, afford, refund, skyrocket, economize, reimburse, discount
Food	cappuccino, buffet, mushroom, cranberry, salami, healthy, shell, croissant, sushi, broccoli
Pet	mew, swan, cat, fish, ant, pony, dog, bird, duck, purr
Service	friendly, repair, employment, discount, servings, safari, checkouts, cleansing, sightseeing, attitude
Amenities	breakfast, massage, yoga, housekeeping, excursion, gamble, sightseeing, paraglide, exercise, television

The performance of the recommendation generated by our proposed model is evaluated by comparing with other existing models. In word-embedding approach, the review sentences from a user and the one for an item are mapped to a latent space using the word embedding [18]. For a user, the K-nearest neighbors in the space were considered as the top-K recommendations. Latent Dirichlet Allocation approach [9]- In this model, we extract the topics relevant to a user and the topics relevant to places. The user-place tuples with most common topics are used for the recommendation. DeepConn [13] is a Convolutional Neural Network based model users the review embedding but ignores other features' embedding and the polarity of reviews.

Table3: Comparison of Proposed model with existing models on various datasets

Models	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
Data Set	**Yelp**		**TripAdvisor**		**Airbnb**	
LDA	0.51	0.53	0.49	0.65	0.51	0.60
DeepConn	0.52	0.79	0.58	0.73	0.60	0.67
Proposed Model	0.69	0.85	0.63	0.77	0.64	0.74

V. CONCLUSION

This paper formulated the review-aspect category correlation using deep neural network and represented the place aspect category bipartite relation as a bipartite graph. We proposed a bipartite-core based, ranking-based, and a shingles-based method to extract the dense subgraphs that represented the users' ordered preferences on aspect categories. The aspect categories from the dense subgraphs were used to explain the recommendation. The evaluations on three real-world datasets and the presented case studies clearly demonstrate the efficiency of proposed models. The deep neural networks require lots of training data for efficient results. The prediction relies on the data used in offline training and might need further tunings to handle real- time data. The model relied on single approach to extract the aspects. The model can also be extended to exploit phrase level aspect and sentiment analysis which can capture the multi-aspect multi-sentiment relation within a sentence. In future, we would like to cluster the users based on their preference order on aspect categories, incorporate more aspect categories, evaluate the explanations with additional metrics, and exploit the dynamic user- aspect preferences.

REFERENCES

[1] W. Zhang, Q. Yuan, J. Han, and J. Wang, "Collaborative multi-level embedding learning from reviews for rating prediction." in IJCAI, Pp. 2986–2992, 2016.
[2] J. Vig, S. Sen, and J. Riedl, "Tagsplanations: explaining recommendations using tags," in Proceedings of the 14th international conference on Intelligent user interfaces, ACM, Pp. 47–56, 2009.
[3] F. Gedikli, D. Jannach, and M. Ge, "How should i explain? a com- parison of different explanation types for recommender systems," International Journal of Human-Computer Studies, Vol. 72, no. 4, Pp. 367–382, 2014.
[4] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.
[5] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in Proceedings of the 7th ACM conference on Recommender systems. ACM, Pp.165–172, 2013.
[6] S. Moghaddam and M. Ester, "Ilda: interdependent lda model for learning latent aspects and their ratings from online product re- views," in Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, Pp. 665–674, 2011.
[7] J.-D. Zhang, C.-Y. Chow, and Y. Zheng, "Orec: An opinion-based point-of-interest recommendation framework," in Proceedings of the 24th ACM International on Conference on Information and Knowl- edge Management. ACM, Pp. 1641–1650, 2015.
[8] X. Wang, Y.-L. Zhao, L. Nie, Y. Gao, W. Nie, Z.-J. Zha, and T.- S. Chua, "Semantic-based location recommendation with multi- modal venue semantics," IEEE Transactions on Multimedia, vol. 17, no. 3, pp. 409–419, 2015.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," Journal of machine Learning research, vol. 3, no. Jan, pp. 993–1022, 2003.
[10] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in Proceedings of the 10th ACM Conference on Recommender Systems. ACM, Pp. 191–198, 2016.
[11] J. Manotumruksa, C. Macdonald, and I. Ounis, "Modelling user preferences using word embeddings for context-aware venue recommendation," arXiv preprint arXiv:1606.07828, 2016.
[12] X. Chen, Z. Qin, Y. Zhang, and T. Xu, "Learning to rank features for recommendation over multiple categories," in Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, Pp. 305–314, 2016.
[13] A. Lawlor, K. Muhammad, R. Rafter, and B. Smyth, "Opinionated explanations for recommendation systems," in Research and Development in Intelligent Systems XXXII. Springer, Pp. 331–344, 2015.
[14] X. He, T. Chen, M.-Y. Kan, and X. Chen, "Trirank: Review-aware explainable recommendation by modeling aspects," in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, Pp. 1661–1670, 2015.
[15] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in Eighth International AAAI Conference on Weblogs and Social Media, 2014.
[16] L. Zhang and B. Liu, "Aspect and entity extraction for opinion mining," in Data mining and knowledge discovery for big data. Springer, Pp:1–40, 2014.
[17] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit." in ACL (System Demonstrations), Pp. 55–60, 2014.
[18] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their com- positionality," in Advances in neural information processing systems, Pp. 3111–3119, 2013.