ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Easy Domain Adaptation For Sentiment Classification

Divya R[1] , Geetha K[2], Jesima Nilopher A[3], Mrs S. Jayanthi[4]
Department of Computer Science and Engineering
KCG college of technology, Chennai, Tamil Nadu, India

[1]divyaramessh@gmail.com
[2]geethakannan7595@gmail.com
[3]jesimanilopher@gmail.com
[4]jayanthi.jagadeesan@gmail.com

***Abstract*—Sentiment Classification is the task of adapting a sentiment classifier trained on a particular domain (source domain), to a different domain (target domain), without requiring any labelled data for the target domain. A sentiment classifier is developed which extracts aspects from reviews and analyze the sentiment sensitive embeddings and rate it based on the arousal. A cross domain sentiment classifier is proposed which can adapt to different variety of domains without need to do much training on target domain. Domain thesaurus is built which can easily classify the aspects and sentiments associated with it. The task is performed in four steps: POS Tagging for user reviews, Chunking the reviews and Aspect extraction, Building Domain Thesaurus, Sentiment Classification and Service Recommendation. Experimental results using reviews of hotels and electronic products sold online demonstrate the effectiveness of this classification.**

***Index Terms*— Sentiment Classification, Valence, Arousal, Domain Thesaurus**

I. INTRODUCTION

A vast number of products are sold online, it is both costly as well as infeasible to manually annotate reviews for each product type. The ability to correctly identify the sentiment expressed in user-reviews about a particular product is an important task for several reasons. First, if there is a negative sentiment associated with a particular feature of a product, the manufacturer can take detect a negative sentiment associated with a product might result in decreased sales. From the user's point-of-view, in online stores where one can't physically touch and evaluate a product as in a real-world store, the user opinions are the only available subjective descriptors of the product. By automatically classifying the user-reviews according to the sentiment expressed in them, the potential buyers of a product to easily understand the overall opinion about that product can be assisted. Considering the numerous applications of sentiment classification such as opinion mining, opinion summarization, contextual advertising, and market analysis, it is not surprising that sentiment classification has received continuous attention. Sentiment classification can be considered as an instance of text classification where a given review must be classified into a pre-defined set of sentiment classes. In binary sentiment classification, a review must be classified into two classes depending on whether it expresses a positive or a negative sentiment towards an entity. Alternatively, a Review can be assigned a discrete sentiment score (e.g. from one to five stars) that indicates the degree of the positivity or negativity of the sentiment. Once a review has been identified as sentiment bearing further analysis can be performed.

Domain adaptation methods can be classified into two groups: supervised domain adaptation methods [3], [4], [5], and unsupervised domain adaptation methods [6], [7], [8], [9]. In supervised domain adaptation, one assumes the availability of a small labeled dataset for the target domain in addition to the labeled data for the source domain, and unlabeled data for both the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

source and the target domains. On the other hand, unsupervised domain adaptation does not assume the availability of labeled data for the target domain.

A cross domain sentiment classifier is proposed which can adapt to different variety of domains without need to do much training on target domain. A domain thesaurus is built which can easily classify the aspects and sentiments associated with it.

The sentiment classification for two different domains with a single classifier without any training on target domains is demonstrated. We extend the SST model proposed earlier to build the Domain Thesaurus on particular target domains and our classifier is able to give the needed results hassle free. The implementation uses Natural Language Processing Techniques such as POS Tagging [2] for extracting aspects and uses the Domain Thesaurus to classify the Aspects based on the Target Domains. Valance and Arousal will be calculated to calculate rating for the particular aspects in the user Review. Product Reviews as well as Hotel Reviews are used for implementation. Of Which Hotel Domain is extended to give Service recommendation to users based on their requirements. A user-based CF (Collaborative Filtering) algorithm is adopted to generate appropriate recommendations. It aims at calculating a personalized rating of each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services to the customer.

II. RELATED WORK

The ability to correctly identify the sentiment expressed in user-reviews about a particular product is an important task for several reasons. By adapting an existing sentiment classifier to previously unseen target domains, we can avoid the cost for manual data annotation for the target domain. This problem was modelled as embedding learning [1] and three objective functions were constructed that capture: (a) distributional properties of pivots (i.e., common features that appear in both source and target domains), (b) label constraints in the source domain documents, and (c) geometric properties in the unlabeled documents in both source and target domains. Unlike prior proposals that first learn a lower-dimensional embedding independent of the source domain sentiment labels and next a sentiment classifier in this embedding, our joint optimisation method learns embeddings that are sensitive to sentiment classification. The methods such as Structural correspondence learning (SCL), Spectral feature alignment (SFA), Sentiment Sensitive Thesaurus (SST) are used in the classification process. The experimental results are statistically comparable to the current methods for cross domain sentiment classification.

E-Commerce is growing rapidly, more products are sold on the web and more people are buying the products online. All customer reviews of a product were mined and summarized to enhance customer satisfaction [2]. The summarization is done in three main steps: (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. These steps are performed in multiple sub-steps.

INFERENCE

The ability to correctly identify the sentiment expressed in user-reviews about a particular product is an important task for several reasons. By adapting an existing sentiment classifier to previously unseen target domains, the cost for manual data annotation for the target domain can be avoided.

NLP techniques are used to perform pre processing of user reviews. The NLP linguistic parser is used to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc). The process also identifies simple noun and verb groups. Each sentence is saved in the review database along with the POS tag information of each word in the sentence. Aspect extraction is done by parsing all the reviews and the nouns and phrasal nouns are extracted and recorded as aspects.

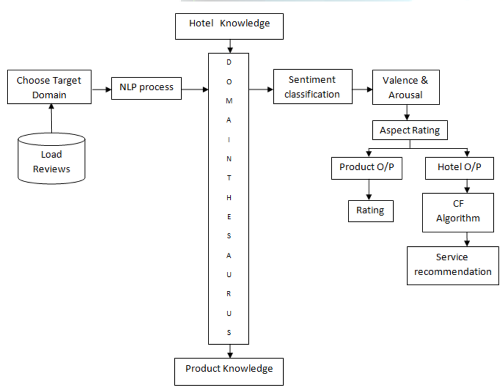
III. PROPOSED SYSTEM

The system proposed in this paper is the sentiment classification for two different domains with a single classifier without any training on target domains.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

A cross domain sentiment classifier which can adapt to different variety of domains without need to do much training on target domain. A domain thesaurus is built which can easily classify the aspects and sentiments associated with it. Our proposed method is different from SCL and SFA in that, we consider not only the unlabeled data but also labelled data for the source domain when constructing the representation.

The implementation uses Natural Language Processing Techniques for extracting aspects and uses the Domain Thesaurus to classify the Aspects based on the Target Domains. Valance and Arousal will be calculated to calculate rating for the particular aspects in the user Review. Product Reviews as well as Hotel Reviews are used for Implementation. Of Which Hotel Domain Sentiment classification can be extended to give Service recommendation to users based on their requirements.



A user-based CF algorithm is adopted to generate appropriate recommendations. It aims at calculating a personalized rating of each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services to him/her.

IV. EXPERIMENT

The following modules like POS tagging for user reviews, chunking the review and aspect extraction, building domain thesaurus on target domain, sentiment classification and service recommendation are identified in this system.

a) POS Tagging for User Reviews

Huge Collection of data is retrieved from open source datasets that are publicly available from web applications like Trip Advisor and Amazon. The Data's are in CSV or TSV Format. The CSV(Comma separated values) files were read and manipulated using Java API that itself developed by us which is developer friendly ,light weighted and easily modifiable. The User review for two different domains were loaded as a CSV or TSV file ,parsed using api and then each review by each customer is processed sequentially. The reviews were given one by one to POS Tagger which splits each word in the review and tags it based on the Parts of Speech the word belongs.

b) Chunking and Aspect Extraction

Chunker Process is done on each and every review of all and the products. The chunker Process will take POS tagged output as input for grouping the Words based on meaning of the Review. Chunker Process is done so that the sentiment embeddings associated with the Aspects of the particular review can be easily extracted. The meaningful words that should be read continuously for proper understanding of the review are marked with square bracket. Now the Aspects in each review are extracted from the POS Tagger result. The Noun and Phrasal Verbs are the key Attributes in any sentence. So those things were extracted from the tagged reviews and marked as Aspects of the particular review by a user. Now mappings are done to properly annotate the user review and associated Aspects with the Chunks in it.

c) Building Domain Thesaurus

A Domain Thesaurus is built depending on the Keyword Candidate List and Candidate Services List. Keyword Candidate List is the Candidate Services List are interdependent on the Target domains and it can be prepared before porting the classifier to Target domain. Expert Knowledge should be given for preparing the domain Thesaurus. The Domain Thesaurus can be Updated Regularly to get accurate Results of the Recommendation System. Now the Aspects extracted are subjected to domain groping based on the target domain.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

d) Sentiment Classification and Service Recommendation

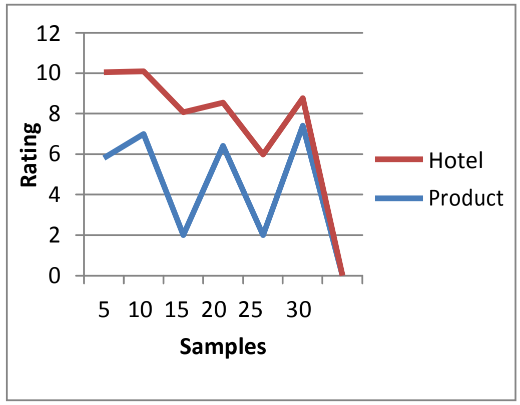
The Chunked Reviews of the User is retrieved and the Keywords (Aspects) corresponding to the User is Analyzed for its Valence and Arousal. Valence Means whether the Keywords Means a positive or Negative thing and Arousal answers, how much it is. Ratings are given for each domain in Target based on the Valence and Arousal for each User of each review.

For product reviews the Overall Rating is now manipulated by taking average values of each rating of several users of a particular product.

In Hotel Domain ranking is extended to give personalized service recommendation to user based on requirements to user. Ranking is done for all hotels based on Ratings by similar users using CF (Collaborative Filtering) and will be sorted based on Bubble Sort Algorithm to have the most appropriate personalized Recommendation for the User.

V. RESULT

A cross domain sentiment classifier which can adapt to different variety of domains without need to do much training on target domain is been proposed. A domain thesaurus is built which can easily classify the aspects and sentiments associated with it. Expert Knowledge is been given for preparing the domain Thesaurus. The Domain Thesaurus is Updated Regularly to get accurate Results of the Recommendation System. Valence and Arousal are calculated to calculate rating for the particular aspects in the user Review. Product Reviews as well as Hotel Reviews are used for implementation, of which Hotel Domain Sentiment classification can be extended to give Service recommendation to users based on their requirements. A user-based CF algorithm is adopted to generate appropriate recommendations.



VI. FUTURE WORK

The Natural Language Processing is implemented to analyze the reviews of the previous user. The NLP Process Comprises Tokenizing a Sentence or a word, POS (Parts of Speech) Tagging, Extraction of Nouns and Verbs, Synonym Retrieval and Spell Check of Extracted Keywords using Word Net Dictionary .Valence and Arousal is implemented for calculating Ratings of Aspects of a Hotel. In future work, further research is how to deal with the case where term appears in different categories of a domain thesaurus from context and how to distinguish the positive and negative preferences of the users from their reviews to make the predictions more accurate.

VII. REFERENCES

[1] Danushka Bollegala, Tingting Mu, John Yannis Goulermas. 'Cross-Domain sentiment classification Using Sentiment Sensitive Embeddings', IEEE Trans. Knowl. Data Eng., 2016, VOL.28, pp.398-410.
[2] M. Hu and B. Liu, 'Mining and summarizing customer reviews', in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2004, pp.168–177.
[3] H. Daum_e III, A. Kumar, and A. Saha, "Co-regularization based semi-supervised domain adaptation," in Proc. Adv. Neural Inf. Process. Syst. 23, 2010. pp. 478–486.
[4] D. Lopez-Paz, J. M. Hernandez-Lobato, and B. Scholkopf, "Semisupervised domain adaptation with non-parametric copulas," in Proc. Adv. Neural Inf. Process. Syst. 25, 2012, pp. 674–682.

[5] H. Daume III, "Frustratingly easy domain adaptation," in Proc. 45th Annu. Meeting Assoc. Comput. Linguistics, 2007, pp. 256–263.
[6] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in Proc. Conf. Methods Natural Language Process., 2006, pp. 120–128.
[7] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in Proc. 45th Annu. Meeting Assoc. Comput. Linguistics, 2007, pp. 440–447.
[8] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 751–760.
[9] D. Bollegala, D. Weir, and J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," IEEE Trans. Knowl. Data Eng., vol. 25, no. 8, pp. 1719–1731, Aug. 2013.