Incorporating Author Preference in Sentiment Rating Prediction of Reviews

Subhabrata Mukherjee
IBM India Research Lab
subhabmu@in.ibm.com

Gaurab Basu
IBM India Research Lab
gaurabas@in.ibm.com

Sachindra Joshi
IBM India Research Lab
jsachind@in.ibm.com

ABSTRACT

Traditional works in sentiment analysis do not incorporate author preferences during sentiment classification of reviews. In this work, we show that the inclusion of author preferences in sentiment rating prediction of reviews improves the correlation with ground ratings, over a generic author independent rating prediction model. The overall sentiment rating prediction for a review has been shown to improve by capturing facet level rating. We show that this can be further developed by considering author preferences in predicting the facet level ratings, and hence the overall review rating. To the best of our knowledge, this is the first work to incorporate author preferences in rating prediction.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing, Linguistic Processing

General Terms

Algorithms, Experimentation

Keywords

Author Preference, Aspect, Rating, Sentiment Analysis

1. INTRODUCTION

Sentiment analysis attempts to find the opinion orientation of a piece of text. A review may have multiple facets with a different opinion about each facet. For example, the movie review "*The film failed to make an impact despite the powerful performance of the lead actor due to sloppy story telling*" has the facets *actor, film* and *story*. The opinion with respect to *actor* is *positive* and that with respect to *film* and *story* is *negative*. The overall review polarity is a weighted function of the facet specific polarities.

The initial works in sentiment analysis focused on determining the overall sentiment of a review as *positive* or *negative* [1]. This gave way to a more fine-grained approach that aims to predict a *rating* for a given review in a pre-defined scale [2]. These works attempted to find the overall rating directly from the given text. A more recent work [3] showed that the overall rating prediction can be improved by first considering the facet level ratings and then aggregating them.

Consider the following review that has been assigned a rating *+4* by an author, *The hotel has a nice$^+$ ambience and comfortable$^+$ rooms. However, the food is not that great$^-$*. Now consider another review by a different author who assigned it a rating *+5*, *The hotel has an awesome$^+$ restaurant and food is delicious$^+$. However the rooms are not too comfortable$^-$*. Both the reviews involve the same facets *ambience*, *rooms* and *food* with a different opinion about each facet. It is obvious from the second review, that its author prefers *food* quality over everything else, which makes him assign the highest rating to the hotel despite having not-so-comfortable *rooms*; whereas the author of the first review prefers the *ambience* and *room* quality over *food*, which makes him assign it a high rating despite his dissatisfaction about *food*. Clearly, the facet specific preference of the author influences the overall rating. Given a set of known facets and a set of reviews by an author with overall ratings, the objective is to learn the author preference about each facet from the reviews. Now, given a new review the target is to predict its overall rating as a function of the facet specific opinions weighed by the author's facet-specific preference.

2. FACET RATING PREDICTION

Let us consider a review *R* with a set of known facets $t_i \in T$ (Example: *value, food, atmosphere, service etc.*) with respect to which the review (of say, a *hotel*) is to be evaluated. So every opinion expressed in the review has to be associated to one of the known facets. For example the opinion, *the dishes are awesome*, relates to the facet *food*, whereas the opinion, i*t's a very large and yet peaceful place*, relates to *atmosphere*. Thus the first subtask is to find a rating for the individual facets.

Let the review *R* consist of *n* sentences, S_i ($i=1...n$) where each sentence has an opinion about a feature present in the sentence. Initially all the *Nouns* in any sentence are considered to be potential features, which are identified by a POS-tagger, and added to the candidate feature set F_i. Let O_{ij} ($j=1...|F_i|$) be the opinion about the feature $f_{ij} \in F_i$ present in the sentence. In order to find the association between $f_{ij} \in F_i$ ($i=1...n$, $j=1...|F_i|$) and $t_k \in T$ ($k=1...|T|$), we use a WordNet-based similarity metric. The Wu & Palmer measure [4] calculates relatedness between two concepts by considering their depths in the WordNet taxonomies, along with the depth of their *Lowest Common Subsumer* (LCS). The Wu-Palmer similarity between two concepts s_1 and s_2 is given by $2*depth(lcs) / (depth(s_1) + depth(s_2))$. Consider $|T|$ clusters C_k, where t_k is the clusterhead of C_k. Each feature f_{ij} is assigned to the cluster C_{k*}, where $k^*=argmax_k$ $Wu\text{-}Palmer(f_{ij},t_k)$. If the similarity score is less than some *threshold*, the feature is ignored.

In order to find the opinion O_{ij} in the sentence S_i with respect to a given feature f_{ij} we use the *dependency parsing* based feature specific sentiment extraction approach in [5]. If $f_{ij} \in C_k$ ($k=1...|T|$) the opinion O_{ij} relates to the facet t_k ($k=1...|T|$), and ignored otherwise. All the opinions about the facet t_k across all sentences S_i are aggregated, and mapped to a numeric rating in the scale 1-5. The rating is expressed as a function of the *positive* and *negative* opinion words, present in the *aggregated* opinion about t_k, which are identified with the help of a lexicon [6].

WWW 2013 Companion, May 13–17, 2013, Rio de Janiero, Brazil.
ACM 978-1-4503-2038-2/13/05.

3. AUTHOR SPECIFIC RATING PREDICTION

The previous section discusses a rule-based approach to find the facet-specific rating from a given review, which has been assumed to be *author independent*. This assigns a rating (say *+5*) to the facet *atmosphere* in the given example, where the reviewer states that *the hotel has a nice ambience and comfortable rooms*. This is a generic rating which simply indicates how *good* or *bad* a facet is no matter who wrote the review. However, it is essential to find out what it means to a particular author, *i.e.* the generic facet specific rating has to be weighed by the author specific preference for the given facet.

Let us consider a review *r* written by an author *a*. The overall rating $P_{r,a}$ of the review *r* is given by, $P_{r,a} = \Sigma_t\ h_{r,t} \times w_{t,a}$, where $w_{t,a}$ is the preference of author *a* for facet *t*, and $h_{r,t}$ is the rating assigned to the facet *t* in the review *r*. The problem can be posed as a linear regression formulation to learn the author preferences, from all the reviews written by him, which is given by $P_{R\times A} = H_{R\times T} \times W_{T\times A}$ *or* $W = (H^T H)^{-1} H^T P$

4. EXPERIMENTS

Trip advisor [7] is used to collect *1526 reviews* for our experiments. It contains the profile of many authors with reviews on different topics, as well as overall review ratings. For our experiments, we chose *restaurant* as the topic and a list of 9 authors who are the *top contributors* in the forum, each of whom had a minimum of 100 restaurant reviews along with their ratings. *Table 1* shows the data statistics per author. For each author, 80% of the reviews were used for training and the remaining for testing. For the topic *restaurant*, we chose the four known facets as *value*, *atmosphere*, *food* and *service*.

Table 1. Dataset Statistics for 9 Authors

Authors	1	2	3	4	5	6	7	8	9
Reviews/Author	152	102	322	383	169	100	100	100	100
Avg. Words/Review	40.4	150	181	52	108	242	113	84	56.4

The first baseline for our work is taken as a simple linear aggregation of all the opinions in the review. This baseline does not take into account the facet specific ratings but the simple majority opinion about all the facets in the review. For the second baseline, the facet weights are learnt over the entire corpus, over all authors. This indicates how much a facet is important, in general, independent of the review author. Pearson's Correlation Co-efficient (PCC) [8] is used to find the correlation of the predicted ratings with ground ratings. *Table 2* shows the comparison of author specific prediction model with baselines.

Table 2. PCC Score Comparison of Different Models

Majority Voting over All Facets	Facet Specific, General Author Preference	Facet and Author Specific Preference
0.550	0.573	0.614

Figure 1 shows the facet specific preferences of each author in the corpus. Overall, the facet preferences have been found to be in the order *Service > Food > Value > Atmoshpere*.

It is observed that the simple majority voting of opinions in the review achieves the lowest correlation with the ground ratings. The performane is improved by considering the overall rating to be a function of facet specific ratings, where the facet ratings are weighed by the general importance of the facet to the reviewers. However, the best correlation is achieved by considering each author's preference for a given facet, which is learnt from the reviews of the given author.

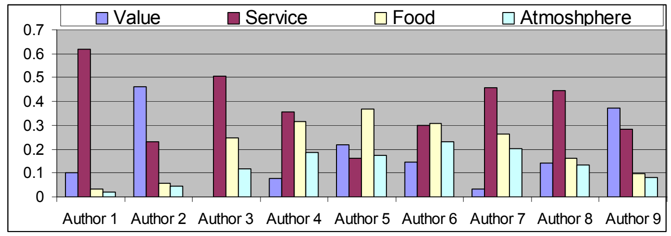

Figure 1. Facet Specific Preferences of Different Authors

5. CONCLUSIONS AND FUTURE WORK

In this work, we have considered the following traditional models for rating prediction of reviews: 1. Models that do not use facet ratings to obtain the overall rating of the review. 2. Models that learn the facet preferences or weights over the entire corpus independent of the author. We proposed a third model that learns the author specific facet preferences from reviews written by a particular author. We have shown that the proposed approach obtains the best correlation with ground ratings over the second model which, again, performs better than the first one.

In facet rating prediction, we have assumed the set of seed facets (like *value, atmosphere, service* and *food*) to be known. Every opinion expressed in the review has been assumed to be associated to one of those known facets and irrelevant otherwise. Thus the performance of facet rating prediction is constrained by the similarity metric used (which does not capture all associations well), as well as the feature specific opinion extraction module. Instead of performing all these sub-tasks separately, which are inter-related, a generative model for jointly discovering the *topics of interest* of the authors*,* their *topic specific preferences* and *opinion*, and hence the *overall review rating* can be used.

6. REFERENCES

[1] Pang, B., Lee, L., and Vaithyanathan, S., 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proc. of EMNLP 2002*

[2] Pang, B., Lee, L., and Vaithyanathan, S., 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proc. of ACL 2005*

[3] B. Snyder and R. Barzilay. 2007. Multiple Aspect Ranking Using the Good Grief Algorithm. In Proc. of HLT-NAACL

[4] Wu, Zhibiao and Martha Palmer. 1994. Verb semantics and lexical selection. In Proc. of ACL, 1994

[5] Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Feature specific sentiment analysis for product reviews. Part 1, Lecture Notes in Computer Science, Springer 7181:475–487.

[6] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

[7] http://www.tripadvisor.com/. 2012. Website

[8] http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient. 2012. Website