2011 International Conference on Advancements in Information Technology
With workshop of ICBMG 2011
IPCSIT vol.20 (2011) © (2011) IACSIT Press, Singapore

Cross-Product Sentiment Analysis via Ensemble SVM Classifiers

Saksham Varma+

[1] Department of Information Technology, National Institute of Technology Kurukshetra

Abstract. Sentiment analysis in context of product reviews involves automatically classifying whether users have expressed positive or negative opinions regarding a product. In this paper we define the novel problem of cross-product sentiment analysis, which involves predicting the sentiment polarity of a new product's reviews using training data available for a different set of products. Our work has high practical utility, since while learning user feedback of newly released products is often most important, training data for them is seldom available. Our approach based on an ensemble of SVM classifiers combined using a vocabulary overlap heuristic, outperforms a baseline SVM based method.

Keywords: Cross Product Sentiment analysis, transfer learning, SVM ensemble, opinion mining

1. Introduction

With the rise of web 2.0, the internet is being largely dominated by user generated content. A sizeable chunk of this data is in the form of user provided reviews. These represent invaluable user feedback for products or services. One critical analysis that can be performed on such reviews is to analyze whether they largely reflect a positive or negative sentiment about some product. Since numerical ratings aren't always available or are difficult to automatically extract, considerable attention has been focused towards building automated sentiment classifiers.

The general trend in automated sentiment classifiers is to utilize supervised machine learning methods that use labeled data to train statistical models for binary classification. While this works well in general, the framework is not suitable for new products for which very little training data would be available. Classifiers trained on training data for one product do not necessarily work well on another due to differences in vocabulary. However, it is usually early on in a product life cycle that a company wants to quickly assess popular sentiment towards a product. Under such circumstances, the only option available is to manually label a large number of product reviews to generate training data; a costly endeavor.

In this paper we explore some solutions to this problem. In particular we analyze the extent to which classification models trained on one set of products can be used for analysis of reviews of a different product. In particular we explore appropriate strategies for combining multiple classification models trained on different products. We also present a Support Vector Machine (SVM) [1] ensemble method which appropriately combines multiple SVM classifiers trained over a set of different products using a feature overlap heuristic.

The rest of the paper is organized as follows. We present the related work on sentiment classification and a formal definition of the problem in section two. We then provide the details of our solution approach along with the manner in which features were constructed in section three. Section four presents a description of our dataset along with the experiments performed and the results obtained. Finally in section five we provide the conclusion.

+ Corresponding author. Tel.: +919896911153.
E-mail address: saksham.nitk@gmail.com.

2. Background

2.1. Related Work

While sentiment classification has been around for a while [2], training classifiers to work across products has been a relatively recent area of research. In recent related works [3] stress on classifying only the subjective terms which they consider as the terms which are either always positive or always negative (polarity) in their occurrence. John Blitzer et al [4] investigated the domain adaptation of various sentiment classifiers by identifying a set of domains over which classifiers could be trained such that the resulting classifier could provide reasonably accurate results over a large horizon of test set data. Due to limitations in space, it is not possible to cover other related works in the area. Interested reader may look at [2] for more details. While these attempts focus on finding a generic vocabulary or set of domains that could be used to train a single classifier to work in general, our work focuses on the method of combining classifiers. Rather than trying to build a single generic classifier or vocabulary, we provide a framework in which any available classifier could be incorporated into an existing classifier ensemble.

2.2. Problem Definition

We define the problem of cross-product sentiment analysis as follows. Given a significant number of labeled reviews for products P_1 to P_n we want to learn a classification model for a new product P_{n+1} for which only a small number of labeled reviews are provided.

Let D_i represent a set of labeled product reviews for product i. Then each element of D_i is a two tuple (r_{ij},l_{ij}), where r_j is the jth review for product P_i and l_j is its sentiment label (either positive:1 or negative:0). Our goal is to use the available data in sets $D_1 \ldots D_n$ to achieve a high prediction accuracy on the test set of reviews D_{n+1} for product P_{n+1}.

3. Method

3.1. Feature Definition

Each review r_{ij} in our dataset is represented via a set of word features. Let V be the vocabulary of all the words present in all the reviews in our dataset. Each word in the vocabulary was considered a unique feature. The value of this feature in a review was given by taking a log of the number of occurrences in the review. Taking log had the effect of normalizing the wide variations in word frequencies. Since SVMs, tend to work best when all feature values are nearly within the same range.

3.2. Ensemble of SVM

Support vector machines have been widely successful for various text classification tasks [5]. We therefore start with SVM as our baseline model. Our SVM ensemble approach is based on the following intuition. Given two different product domains, one could obtain an estimate of their similarity by looking at how well the feature vectors of their reviews, or in our case the vocabulary of their reviews, overlaps. While predicting the sentiment polarity of a review for a novel product, we would often prefer to use the classifiers that were trained on other similar products, than those trained on very different ones. Thus once domain similarity is calculated, one could assign different product classifiers weights proportional to their similarity scores. Mathematically, we formalize this idea as follows:

Given a set of labelled product specific reviews D_1 through D_n, we train n SVM classifiers C_1 through C_n, one per product. Let the prediction score of classifier C_i on some review r for the target product P_{n+1} ($r \in D_{n+1}$) be given by $C_i(r)$ where:

$C_i(r) > 0$: If review sentiment is positive

$C_i(r) < 0$: If review sentiment is negative

Also let the vocabulary of product P_i be given by V_i. We define vocabulary overlap between products P_i and P_{n+1} to be :

$$VO(i, n+1) = (|V_i \cap V_{n+1}|) \div (|V_i|)$$

where i varies from 1 through n. We further define normalized vocabulary gap to be:

$$NormVO(i,n+1) = VO(i,n+1) \div (\min_{j=1\ to\ n}\{VO(j,n+1)\})$$

Intuitively, the vocabulary overlap between two products provides the proportion of training domain product's words that are common to the two products. Note that this makes the metric non-symmetric, i.e. VO(i,j) $\neq$ VO(j,i). Normalized vocabulary overlap on the other hand is given by normalizing by the minimum available vocabulary overlap for the products P_1 through P_n.

Finally we define the ensemble SVM classifier as follows:

$\sum_{i=1}^{n} NormVO\ (i,n+1)\ \ C_i(r) > 0$: Classify review r's sentiment as positive

$\sum_{i=1}^{n} NormVO\ (i,n+1)\ \ C_i(r) < 0$: Classify review r's sentiment as negative

Overall the ensemble SVM classifier combines the predictions of all n SVM classifiers and weighs them based on their domain's similarity with the target product. If the resulting prediction score is greater than 0, we classify the review as positive otherwise we classify it as negative. In the subsequent section we describe the dataset used and present our results.

4. Experiments and Results

4.1. Dataset Description

We built an entirely new dataset for studying transfer learning on multiple domains from the product reviews available on Amazon.com. The dataset has four domains viz. books, games, covers and cases of various electronics and kindle based products. The following table shows the number of reviews in each domain.

Table 1: Dataset Description

Name of the domain	**Number of reviews**
Books	2014
Covers and Cases	1824
Games	556
Kindle based products	6124

Each review node had mainly three attributes, namely the original rating of the review on a five star scale (as rated by the reviewer), the title of the review and the main body of the review. The true polarity of a review was decided by this rating of the review. The ones with rating greater or equal to 3 stars were considered as positive ones and the ones below 3 stars were considered as negative in polarity.

4.2. Evaluation Criteria

For evaluation criteria we use classification accuracy. It is the percentage of total reviews in test set whose sentiment polarity was correctly predicted by the classifier. Mathematically, accuracy is the ratio of the sum of the true positives and the true negatives to the total number reviews taken in the test set i.e the sum of total positives and total negatives in the test set.

$$Accuracy = (True\ \ Positives + True\ \ Negatives) * 100 / Total\ Test\ \ \mathrm{Re}views$$

4.3. Results

We look at the performance results for a baseline SVM classifier directly trained over all training reviews and compare it to the performance of our ensemble SVM (Table 2). In order to evaluate our method, we conducted experiments alternately choosing one out of four products as the test product and the remaining three for training. We observe that while training a combined SVM classifier over all reviews generally performs better than individual cross product classifiers, our ensemble approach outperforms it for three out of four domains.

Table 2: Performance comparison between baseline SVM and ensemble of SVM classifiers

Training Domains	**Test Domain**	**SVM Ensemble Accuracy (%)**	**Baseline SVM Accuracy (%)**	**% Improvement**
Games, Kindle,	Books	70.80	64.62	+6.18%

Cases or Covers				
Books, Games, Kindle	Cases or Covers	82.35	78.30	+4.05%
Books, Cases or Covers, Kindle	Games	80.94	79.36	+1.58%
Books, Cases or Covers, Games	Kindle products	84.50	89.32	-4.82%

In order to understand in more detail why baseline SVM outperformed our ensemble approach for kindle products, we look at the vocabulary overlap values for all product pairs. Table 3 shows these values. We observe that is fairly high for all pairs with Kindle as the test product, thus suggesting that all the three training domains are fairly useful for predicting Kindle reviews. Thus in such cases directly training a classifier over all available training reviews is likely to be effective as well, which is what we observe. On the other hand for all other test products at least one of the vocabulary overlap values was reasonably low resulting in the performance of the baseline SVM classifier being lower. Thus the observation also suggests that Vocabulary Overlap could potentially be a useful metric which given a set of test/train domains, could help us decide whether to train a single cumulative classifier or an ensemble.

Table 3: Vocabulary Overlap values for different test/train pairs. Recall that VO(i,j) is asymmetric.

Test Domain	**Training Domain**	**Vocabulary Overlap**
Books	Cases or Covers	0.646
Books	Games	0.618
Books	Kindle	0.396
Kindle	**Books**	**0.525**
Kindle	**Cases or Covers**	**0.766**
Kindle	**Games**	**0.638**
Games	Books	0.340
Games	Cases or Covers	0.503
Games	Kindle	0.264
Cases or Covers	Books	0.258
Cases or Covers	Games	0.365
Cases or Covers	Kindle	0.230

Finally note that our approach has two main advantages in that it does not require any labeled examples from the target domain, and does not require new classifiers to be trained for every domain. Instead we just need to obtain the vocabulary overlap values, which are relatively easy to calculate.

5. Conclusion

In this paper we presented our work on cross-product sentiment classification. Our experiments were based on a newly crawled dataset of nearly 10K reviews across four different products. Results show that the proposed method of utilizing an ensemble of SVM classifiers combined via the vocabulary overlap heuristic is indeed useful for improving cross classification performance. In future we hope to extend our work by experimenting on a bigger dataset and developing novel classification models tuned specifically towards cross-product sentiment classification.

6. References

[1] Vladimir N. Vapnik, The Nature of Statistical Learning Theory. Springer, 1995.

[2] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.

[3] A. Esuli and F. Sebastiani. Determining the semantic orientation of terms through gloss classification. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05), Bremen, DE, pp. 617-624.

[4] J. Blitzer, M. Dredze, and F. Pereira. Biographies, Bollywood, Boom-boxes, and Blenders: Domain Adaptation for Sentiment Classification. Association for Computational Linguistics - ACL 2007

[5] T. Joachims. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In Proc. of ECML, pages 137-142, 1998.