Mining Multilingual Opinions through Classification and Translation

James G. Shanahan, Gregory Grefenstette, Yan Qu, David A. Evans

Clairvoyance Corporation
5001 Baum Boulevard, Suite 700
Pittsburgh, PA 15213-1854
{jimi,grefen,yqu,dae}@clairvoyancecorp.com

Abstract

Today, much of product feedback is provided by customers/critiques online through websites, discussion boards, mailing lists, and blogs. People trying to make strategic decisions (e.g., a product launch, a purchase) will find that a web search will return many useful but heterogeneous and, increasingly, multilingual opinions on a product. Generally, the user will find it very difficult and time consuming to assimilate all available information and make an informed decision. To date, most work in automating this process has focused on monolingual texts and users. This extended abstract describes our preliminary work on mining product ratings in a multilingual setting. The proposed approaches are automatic, using a combination of techniques from classification and translation, thereby alleviating human-intensive construction and maintenance of linguistic resources.

Introduction

Market researchers rely on consumer feedback for various business decisions (such as forming marketing strategies, adapting products, and creating new products). Today, much of that feedback can be provided by customers/critiques online through vendor websites, discussion boards, mailing lists, and blogs. When trying to make strategic decisions (e.g., a product launch, a purchase), people typically find that a general web search will return many useful but heterogeneous and, increasingly, multilingual opinions. Generally, the user will find it very difficult and time consuming to assimilate all available information and make an informed decision. Several attempts have been made to make this task easier. Most of these rely upon manually aggregating opinions. An example of such a web service is *Ewatch.com*. There have been several automatic attempts at aggregating customer feedback data (Morinaga et al. 2003; Sano 2003). Some of these systems rely upon a human providing linguistic resources (Subasic and Huettner 2000; Liu et al. 2003; Sano 2003), though there has been some work on automatically identifying them (Wiebe 2000). For example, pioneering work by Subasic and Huettner (2000) on creating computational models of affect was based upon a human constructed affect lexicon that required a linguist to characterize each word along two dimensions: *centrality* (indicating the degree a word belongs to an affect category) and *intensity* (representing the strength of the affect level described by that entry). Liu et al. (2003) explore the use of the Open Mind Commonsense database as means of constructing a model for measuring the affective qualities of email. This model is based upon a six-state affect lexicon that is manually constructed. These approaches, while being interesting, are labor intensive and can be vulnerable to error and high maintenance costs. More recently, research has focused on generating systems automatically for affect and opinion modeling. A n example of this includes Pang et al.'s (2002) work on classifying movie ratings. Another example was provided by Dave et al. (2003), where various machine learning and information retrieval approaches were compared for the task of product ratings classification. Das and Chen (2001) used a classifier on investor bulletin boards to see if apparently positive ratings are correlated with positive stock price. Commercial efforts in this area include Justsystem's CB Market Intelligence system that organizes feedback data in an affect map (Sano 2003). Another example of a commercial system is NEC's SurveyAnalyzer, which mines the reputations of products (Morinaga et al. 2003).

The approaches presented above all focus on capturing affect and opinion in monolingual applications.. Here, in this paper, we focus on opinion mining in a multilingual setting using automatic approaches based upon translation and machine learning. A monolingual opinion classifier is first constructed from a monolingual corpus. Opinion classifiers in target languages are then constructed using translation (of the classifier or source documents). The learnt opinion classifiers are then used to classify documents from their respective languages. These opinion ratings can then be pooled to mine market perception for a product at an international level.

In the next section, we describe the proposed approaches. Subsequently, we define our experimental setup and some initial results. We finish with intermediate conclusions.

Approach

We describe three approaches for automatically mining multilingual product reviews. The first two approaches are based upon classification and translation technologies, while the third is based exclusively on classification only. Below, the approaches are outlined and then the feature selection and classification construction components are subsequently detailed.

Figure 1 presents a schematic of the first approach. This approach begins with training a classifier from a labeled ratings database, where each document is written in language *1* or the source language and is associated with a rating. In our experiments, we use a support vector machine (SVM) classifier (see comments below for some background on this). For the purposes of our studies, we limited ratings to positive and negative ratings. In this case, training results in a single SVM model. This model, linear in nature, consists of list or vector of terms and associated weights. The terms are words or multiword expressions from the source language. A corresponding model in a target or second language (e.g., Japanese) is generated using a translation system such as http://babel.altavista.com/, or a dictionary-based lookup, commonly used in cross lingual information retrieval (CLIR) (Qu et al. 2002). This approach can be used to generate models in any number of target languages. Subsequently, these ratings classifiers can be used to classify documents in their respective languages and in mining ratings.

A variant on this strategy is presented schematically in Figure 2. Here, the approach begins with translating each source language document in a labeled ratings database into the target language using translation tools such as http://babel.altavista.com/. Having translated the ratings database into a target language, a corresponding SVM classifier is trained for that language. The learnt language-specific classifiers can then be used as previously described.

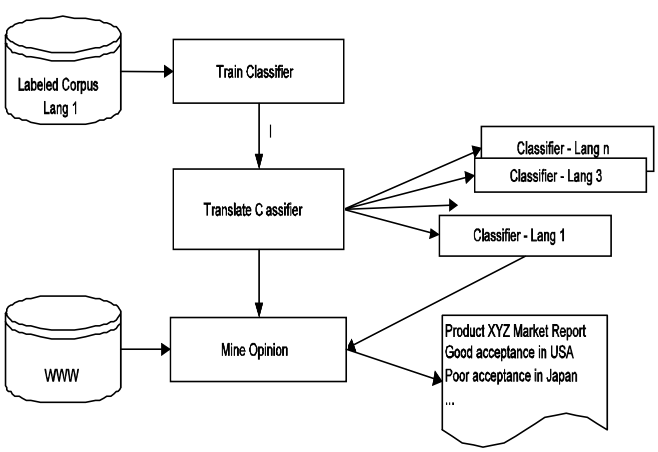

Figure 1: Multilingual Opinion Mining via Translated Models

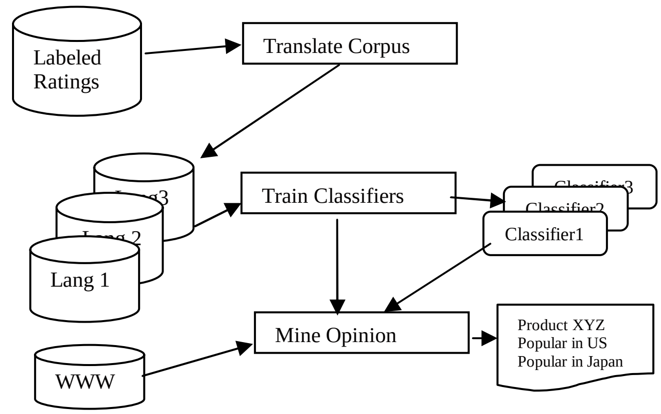

Figure 2: Multilingual Opinion Mining via Translating Documents and then Learning the Model.

The third strategy for mining multilingual product reviews relies on rated documents existing in both the source language and target language(s). Here, if enough data exist, a monolingual opinion classifier can be learned for each language. The learnt classifiers can then be applied to classify documents in either the source or target language(s). If insufficient data exist for any target language, then source language documents can be translated into this target language and, subsequently, a model can be learned from the original documents and the translated documents.

Feature Selection

Document representation is a critical step in text classification. Here, we resorted to three different representations of documents. The first representation uses features occurring in the affect lexicon constructed by Subasic and Huettner (2000). The second set of features correspond to the set of commonly used English words (approximately 80,000 terms) adapted from the Clarit Toolkit (Evans et al. 1995). The third representation uses all available words. Two versions of terms are used: a stemmed version, corresponding to stemming the term with the Porter Stemming algorithm); and an un-stemmed (surface form) version.

Classifier Construction

We chose to represent our classifier using a Support vector machines (SVM). SVMs were originally introduced by Vapnik in 1979 and have provided state-of-the-art performance for a variety of learning problems (and in some cases better than state-of-the-art). They have only recently gained popularity in the text retrieval and classification community (Vapnik 1995; Pang et al. 2002; Dave et al. 2003).–For our current study, we implemented a version of SVM that uses a variation of the dual space SMO (Sequential Minimal Optimization) learning algorithm (Platt 1998; Keerthi et al. 1999). We extended the SVM learning algorithm by adjusting the threshold of the SVM to relax the conservative nature of SVMs (Shanahan and Roma 2003).

Experimental Setup

The parameter settings explored in the experiments reported in this paper are summarized in Table 1. Most parameters are self-explanatory, apart from how a document is represented and choice of *n*, the number of folds used in cross validation. We represent documents as a vector of terms that are derived as follows: we replace all numerical and punctuation characters by spaces and eliminate stop-words such as articles and prepositions, etc.; each term is associated with a *TFxIDF* weight, where *TF* denotes the frequency of a term in a document, and *IDF* is calculated based on the distribution of the term in the training corpus (following (Salton 1983)). In experiments, document vectors were normalized to unit length.

Evaluation Corpus

The proposed approaches are currently being evaluated on a ratings corpus that was generated by downloading reviews from Amazon.com for a variety of books. Though the Amazon's rating scale is 1 through 5, for the purposes of our experiments, we reduced the scale to two classes: a positive class corresponding to scalar values of 3 (3 stars), 4, and 5; and a negative class (1 and 2 stars). The breakdown for each rating is shown in Table 2. We are currently downloading ratings in Japanese and French from Amazon.co.jp and Amazon.fr respectively. These rated opinions will be used to test the approaches proposed here.

Results

To date, we have generated classifiers only in English and tested their validity on English ratings documents. We used the Amazon.com ratings corpus in conjunction with 5-fold cross-validation. Results are presented in Table 3.

Conclusions

In this paper, we have described our automatic approaches to mining multilingual product reviews. We treat the case where there is not much data in the target language. We compare two approaches: the first approach builds a model in a source language and translates the model into a target language; and the second approach translates the documents in the source language and builds a model in the target language using the translated documents. Our ongoing experiments are comparing both of these approaches. At the time of writing, we have presented the monolingual results in our experimental setup. We plan to present the results of multilingual experiments in the final version of this paper.

Table 1: Learning Decision Variables and Explored Values.

Decision Variable	Purpose	Explored Values
Learning Algorithm	SVM Learning	SMOK2
C	Lagrange multiplier upper bound for SMO	1
Tolerance	Numerical precision	0.001
n	Number of folds used for cross validation	5
Kernel type	SVM kernel	Linear
Sampling Ratio	Ratio of positive to negative examples	Use all training data
Input features	Term types	White space delimited tokens with numbers, punctuation, and stopwords removed
Term Selection	Learning	Use {affect | Common | All} terms
Input feature values	Term weighting	TF*IDF

Table 2: Ratings Class Breakdown for Amazon.com Database.

Rating Type	Count
Negative	513
Positive	959

Table 3: Results for the Amazon.com Ratings Corpus Test.

Approach	$F_{beta=0.5}$	Precision	Recall
CC SVMs	0.75	0.67	0.76

References

Das, S. R. and Chen, M. Y. (2001) Yahoo! for Amazon: Sentiment parsing from small talk on the web. In Proceedings of the 8th Asia Pacific Finance Association Annual Conference.

Dave, K., Lawrence, S., and Pennock, D. (2003) Mining the Peanut Gallery: Opinion Extraction and Semantic Classification for Product Reviews. In Proceedings of the International World Wide Web Conference, pages 519–528.

Evans, D. A. and Lefferts, R. G. (1995) CLARIT-TREC Experiments. Information Processing and Management, 31(3):385–395.

Keerthi, S. S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K. (1999) Improvements to Platt's SMO algorithm for SVM classifier design. Technical report, Dept of CSA, IISc, Bangalore, India.

Liu, H., Lieberman, H., and Selker, T. (2003) A model of textual affect sensing using real-world knowledge. In Proceedings of the Seventh International Conference on Intelligent User Interfaces, pages 125–132.

Morinaga, S., Yamanishi, K., Tateishi, K., and Fukushima, T. (2002) Mining product reputations on the web. In Proceedings of KDD.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86.

Porter, M. F. (1980) An algorithm for suffix stripping. Program, 14(3):130–137.

Qu, Y. Grefenstette, G., Evans, D. A., (2002) Resolving Translation Ambiguity using Monolingual Corpora. In Peters, C., Braschler, M., Gonzalo, J., and Kluck, M. (eds.): Advances in Cross-Language Information Retrieval, Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002. Revised Papers, Lecture Notes in Computer Science 2785, Springer.

Teufel, S. and Moens, M. (2002) Summarizing Scientific Articles -- Experiments with Relevance and Rhetorical Status. In *Computational Linguistics*, 28 (4).

Salton, G., (1983) Introduction to Modern Information Retrieval, Verlag: McGraw Hill, New York.

Sano, M. (2003) An Affect-based Text Mining System for Qualitative Analysis of Japanese Free Text, Submitted to IPS Japan DBWeb 2003.

Shanahan, J. G., and Roma, N. (2003) Improving SVM Text Classification Performance through Threshold Adjustment, European Conference on Machine Learning (ECML), pages 361-372.

Subasic, P. and Huettner, A. (2000) Affect Analysis of Text Using Fuzzy Semantic Typing, FUZZ-IEEE San Antonio, May.

Vapnik, V. (1995).The Nature of Statistical Learning Theory, Springer-Verlag.

Wiebe, J. (2000) Learning subjective adjectives from corpora. In Proceedings of AAAI/IAAI 2000, pages 735–740.