A Formal Study of Classification Techniques on Entity Discovery and their Application to Opinion Mining

Shadi Banitaan*
Department of Computer Science
North Dakota State University
Fargo, ND 58102, USA
shadi.banitaan@ndsu.edu

Saeed Salem
Department of Computer Science
North Dakota State University
Fargo, ND 58102, USA
saeed.salem@ndsu.edu

Wei Jin
Department of Computer Science
North Dakota State University
Fargo, ND 58102, USA
wei.jin@ndsu.edu

Ibrahim Aljarah
Department of Computer Science
North Dakota State University
Fargo, ND 58102, USA
ibrahim.aljarah@ndsu.edu

ABSTRACT

Entity discovery has become an important topic of study in recent years due to its wide range of applications. In this paper, we focus on examining the effectiveness of various classification techniques on entity discovery and their application to the opinion mining task. The initial and most important step in opinion mining is to identify and extract highly specific product related and opinion related entities from product reviews. We formulate this problem as a classification task and present a comprehensive study of classification techniques on identifying entities of interest. The impacts of linguistic features such as part-of-speech (POS), and context features such as surrounding contextual clues of words on the classification performance are carefully evaluated. The experimental results show that good classification performance is closely related to the use of classification techniques, linguistic features, and context features. The evaluation is presented based on processing the online product reviews from Amazon.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

General Terms

Experimentation

*Corresponding Author.

SMUC'10, October 30, 2010, Toronto, Ontario, Canada.

Keywords

Entity Discovery, Opinion Mining, Sentiment Analysis

1. INTRODUCTION

With the rapid increase of e-commerce, customers share their reviews about the products they purchase or use. As a result, the number of reviews on thousands of online products has become huge which makes it very hard for a potential customer to read them to make an informed decision on whether to purchase a product. In addition, the large number of reviews makes it hard for e-commerce companies to keep track of customer opinions on their products. To be beneficial, these reviews need to be cleaned and summarized. Many researchers studied this problem which is called opinion mining or sentiment analysis.

Two main research directions are explored in opinion mining, document level opinion mining and feature level opinion mining. At the document level, the aim is to classify the review as positive or negative. At the feature level, the aim is to determine the opinion orientation for each feature in the review. There are three main tasks in the feature level opinion mining as defined by Hu and Liu [5]:

1. Identify and extract object features that have been commented on by an opinion holder.
2. Determine whether the opinions on the features are positive or negative.
3. Group feature synonyms.

In this paper, we are focusing on feature level opinion mining and our goal is to design a framework that is capable of extracting highly specific product related and opinion related entities from reviews. Our system first identifies potential product and opinion related entities, and then recognizes opinion sentences in the reviews. The last step is to predict whether the opinion for each recognized product entity is positive or negative.

We formulate the problem of extracting both product related entities and opinion related entities as a classification task. The goal of classification is to build a model that describes a set of predetermined classes such that each sample is assumed to belong to a predefined class. The resulting model is then used to predict class labels for the testing instances, where the values of the features are known, but the value of the class label is unknown. Classification techniques have been successfully used in many fields such as computer vision, drug discovery, speech recognition, handwriting recognition, natural language processing, document classification, pattern recognition, and bioinformatics.

To summarize, we make the following key contributions in this work:

- We use various popular classification techniques, such as Conditional Random Fields (CRF), Naive Bayes, Bayesian network, random forests, and support vector machines (SVMs), to extract highly specific review entities. Their respective classification performance on entity discovery is presented.
- We integrate the tasks of opinion sentence identification and opinion orientation detection with various classification models which further evaluate the effectiveness of each classification technique on the opinion mining application.
- We study and report the impact of using different feature sets on the classification accuracy by different models.

The rest of the paper is organized as follows: section 2 discusses related work. Section 3 describes our data representation framework. Section 4 presents each of the proposed techniques in details. The experimental evaluation and discussion are presented in Sections 5 and 6, respectively. Section 7 concludes the paper.

2. RELATED WORK

There has been much research studying opinion mining at the document level. Turney (2002) [16] introduced a simple unsupervised learning algorithm for rating a review; their algorithm has three steps: (1) extracting phrases containing adjectives or adverbs; (2) estimating the semantic orientation of each phrase by using point-wise mutual information (PMI) and (3) classifying the review based on the average semantic orientation of the phrases. Pang et al. (2002) [11] applied several machine learning techniques to classify movie reviews into positive and negative and in [10], they studied another machine learning technique based on subjectivity detection and minimum cuts in graphs for sentiment classification of movie reviews.

Some existing research on identifying product features that have been commented on by reviewers include (Hu and Liu [5]; Liu, Hu, and Cheng [9]; Popescu and Etzioni [12]; Carenini, Ng and Zwart [2]). Jin and Ho (2009) [6] proposed a novel machine learning framework using lexicalized HMMs. Their framework identified complex product related entities and opinion expressions

In this paper, we propose to formulate the problem of product and opinion entity identification as a classification task, and thus various classification techniques can be employed. The conditional random field technique, which has been shown effective in part-of-speech tagging [8], Biomedical Named Entity Recognition [13], and shallow parsing [14], is first applied in the opinion mining task. Other existing classification techniques such as Naive Bayes, Bayesian network, random forests, and support vector machines techniques have also been examined. The machine learning methods have also been applied to document level opinion mining. Pang et al. [11] recently classify movie reviews into two classes, positive and negative. They used Naive Bayes, maximum entropy, and support vector machines.

3. DATA REPRESENTATION FRAMEWORK

We follow Jin and Ho (2009) [6] and adopt a hybrid tag representation framework for our training and testing data. This framework defines four entity categories to represent each word in the review. Table 1 shows these categories information (a digital camera is used as an example). The basic tag set given in Table 2 is used to identify each entity category and the pattern tag set given in Table 3 is defined to capture the phrases' internal formation patterns. Both the basic tag set and pattern tag set are used to describe each word's specific roles in a sentence (referred to as a hybrid tag representation). An entity can be a single word or a phrase. So, a word may take one of the following roles:

- Independent entity,
- The beginning of an entity,
- The middle of an entity,
- The end of an entity.

The following example illustrates the hybrid tag representations of the sentence "The quality of the photos was bad."

<BG>The</BG><PROD-FEAT-BOE>quality
</PROD-FEAT-BOE><PROD-FEAT-MOE>of
</PROD-FEAT-MOE><PROD-FEAT-MOE>the
</PROD-FEAT-MOE><PROD-FEAT-EOE>
photos</PROD-FEAT-EOE>
<BG>was</BG><OPINION-NEG>
bad</OPINION-NEG>

Table 1: Definitions of entity categories and examples.

Components	Physical objects of a camera including the camera itself, e.g., LCD, viewfinder, battery.
Functions	Capabilities provided by a camera, e.g., movie playback, zoom, auto focus.
Features	Properties of components or functions, e.g., color, speed, size, weight, clarity.
Opinions	Ideas and thoughts expressed by reviewers on product features, components, or functions.

4. THE PROPOSED TECHNIQUES

Given a sequence of words, $w = \{w_1, w_2, \cdots, w_n\}$, and the corresponding part-of-speech tags, $s = \{s_1, s_2, \cdots, s_n\}$, we aim to generate an appropriate sequence of tags, or labels, $y = \{y_1, y_2, \cdots, y_n\}$. We define the set of possible label

Table 2: Basic tag set and its corresponding entities.

Tags	Corresponding entity.
<PROD-FEAT>	Feature entity.
<PROD-PARTS>	Component entity.
<PROD-FUNCTION>	Function entity.
<OPINION-POS>	Positive Opinion entity.
<OPINION-NEG>	Negative Opinion entity.
<BG>	Background word.

Table 3: Pattern tag set and its corresponding patterns.

Pattern tags	Corresponding entity.
<>	Feature entity.
<BOE>	Component entity.
<MOE>	Function entity.
<EOE>	Positive Opinion entity.

values as a combination of basic tags and pattern tags proposed by [6] as given in Tables 2 and 3. Given w, s, and y, we construct a data set $D = \{(x_i, y_i)|y_i \in y\}$, where x_i is a set of features that represent the neighborhood of the word w_i which can be a combination of w and s, and y_i is the class label of word w_i. For example, the class labels that we use to represent product feature entities are: PROD-FEAT, PROD-FEAT-BOE, PROD-FEAT-MOE, and PROD-FEAT-EOE.

The following table shows the class labels for the sentence: " The quality of the photos was bad ."

Word	Class label.
The	BG
quality	PROD-FEAT-BOE
of	PROD-FEAT-MOE
the	PROD-FEAT-MOE
photos	PROD-FEAT-EOE
was	BG
bad	OPINION-NEG

4.1 Feature Extraction

Feature extraction is an essential part of building an accurate classification model. We examine the impact of linguistic features such as part-of-speech (POS), and context features such as surrounding contextual clues of the word on the classification performance. The following is a set of features that we use to build our data set.

- Part-of-speech (POS) features: Each word w_i is classified into one of the part-of-speech tags: noun, verb, adverb, etc. we do the same for neighboring words in a $[-2, +2]$ window. We use the Part-Of-Speech tagger designed by the Stanford NLP Group [15].

- Word features: We use the surrounding words (in a $[-2, +2]$ window) of a particular word as features.

Let w_i be the word at position i and p_i be the part-of-speech tag of w_i. Note that w_0 represents the current word and i represents the i^{th} position on the right of the current word and $-i$ represents the i^{th} position on the left of the current word. Table 4 shows the representation of the sentence " The quality of the photos was bad." using the feature set $w_{-2}, w_{-1}, w_0, w_1, w_2$ and Table 5 shows the representation of the same sentence using the feature set $p_{-2}, p_{-1}, p_0, p_1, p_2$. Missing values are represented by a single question mark.

Table 4: The representation of the sentence using the feature set $w_{-2}, w_{-1}, w_0, w_1, w_2$.

w_{-2}	w_{-1}	w_0	w_1	w_2	Class label
?	?	The	quality	of	BG
?	The	quality	of	the	PROD-FEAT-BOE
The	quality	of	the	photos	PROD-FEAT-MOE
quality	of	the	photos	was	PROD-FEAT-MOE
of	the	photos	was	bad	PROD-FEAT-EOE
the	photos	was	bad	?	BG
?	was	bad	?	?	OPINION-NEG

Table 5: The representation of the sentence using the feature set $p_{-2}, p_{-1}, p_0, p_1, p_2$.

p_{-2}	p_{-1}	p_0	p_1	p_2	Class label
?	?	DT	NN	IN	BG
?	DT	NN	IN	DT	PROD-FEAT-BOE
DT	NN	IN	DT	NNS	PROD-FEAT-MOE
NN	IN	DT	NNS	VBD	PROD-FEAT-MOE
IN	DT	NNS	VBD	JJ	PROD-FEAT-EOE
DT	NNS	VBD	JJ	?	BG
?	VBD	JJ	?	?	OPINION-NEG

In the following we will describe each of the classification techniques we employ in our entity discovery task.

4.2 Classification Techniques

The main task that we would like to solve using classification algorithms is as follows: Given a Data set D, the goal is to learn a classifier that is best able to predict y for unseen data. Here we briefly present the classification techniques that we use in our work.

4.2.1 Conditional Random Fields

Conditional Random Fields are undirected statistical graphical models, a special case of which is a linear chain that corresponds to a conditionally trained finite-state machine. Following Lafferty et al. [8], the conditional probability of a sequence of labels y given a sequence of words w is given by:

$$P(y|w) = \frac{1}{Z_w} exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_j f_j(y_{i-1}, y_i, w)\right) \quad (1)$$

$$Z_w = \sum_{y} exp\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_j f_j(y_{i-1}, y_i, w)\right) \quad (2)$$

Where Z_w is a normalization factor of all state sequences. $f_j(y_{i-1}, y_i, w)$ is one of m functions which describes a feature and λ_j is a learned weight for each feature function. To train a CRF, the objective function to be maximized is the conditional log likelihood of labeled sequences in a training dataset D_{train} given the observed sequences.

4.2.2 Naive Bayes Classifier

Naive Bayes is a simple probabilistic classifier that makes the naive assumption that all the feature variables are independent. A classification problem can be written as the problem of finding the class with maximum probability given a set of observed features values. This probability is the posterior probability of the class given the data, and is computed using the Bayes theorem, as:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{3}$$

Where C is a Class variable, X is represented by a vector of features, $X = \{x_1, x_2, \cdots, x_m\}$, while $P(C)$ and $P(X|C)$ are the prior probability of the class and the conditional probability of the feature variables given the class. When computing $P(X|C)$, the feature variables are assumed to be independent which implies that the joint probability can be written as a product of probabilities:

$$P(X|C) = P(x_1, x_2, \cdots, x_m|C) = \prod_{j=1}^{m} P(x_j|C) \tag{4}$$

The class label of X is predicted as the class C which has the highest $P(C|X)$.

4.2.3 Bayesian Network Classifier

A Bayesian network [3] is a probabilistic graphical model that represents a set of random variables and their conditional independences via a directed acyclic graph (DAG). The nodes of the DAG represent a set of random variables. The edges represent conditional dependencies between the random variables. Two random variables are conditionally dependent if there is an edge between them and they are conditionally independent if there is no edge between them. Each node is associated with a probability distribution table. Formally, given a class variable C and a vector of feature variables, $X = \{x_1, x_2, \cdots, x_m\}$, a Bayesian network classifier defines the joint probability:

$$P(C, x_1, x_2, \cdots, x_m) = P(C)\prod_{i=1}^{m} P(x_i|Pa(x_i)) \tag{5}$$

Where $Pa(x_i)$ is the set of parents of the feature x_i, this means the feature x_i is conditionally dependent on the set of its parents.

4.2.4 Random Forests Classifier

Random forests [1] is an ensemble classifier that generates many decision trees. Each decision tree within the forests is built using a different bootstrap sample from the data set. To classify a new object, we put the object down in each of the trees in the forest. Each tree gives a class label, and we say the tree votes for that class label. The random forests classifier selects the class label that has the most votes. Figure 1 shows the construction of the random forests. A convenient method to build the ensembles is by random vectors which are generated via random selection procedure from integrated training set.

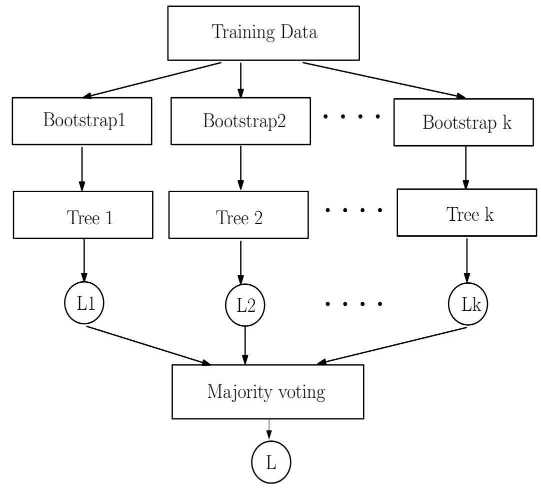

Figure 1: *Construction of random forests.*

4.2.5 Support vector machines

A support vector machine is a classification method which is based on maximum margin linear discriminants. The SVM maps the data from the original space into a higher dimensional space. The SVM then tries to find a linear separation between the classes in the new space. SVM utilizes the kernel function approach to compute the dot product of the points in the new space without having to map every point to its image in the new space. A linear discriminant function in d-dimensions is given by a hyperplane, $h(x) = w^T x + b$, where w is a d-dimensional weight vector, and b is a scalar, called the bias. For points that lie on the hyperplane, $h(x) = w^T x + b = 0$. To classify any given point x, h(x) is used as a linear classifier, which predicts the class label y for any point x such that the predicted class label is $+1$ if $h(x) > 0$ and it is -1 if $h(x) < 0$.

4.3 Opinion Sentence Extraction

In the previous step, we have identified potential product and opinion entities from the reviews. This step will recognize opinion sentences in the reviews for further investigation. Opinion sentence in our work is defined as a sentence that expresses an opinion on some product related entities. For example, the sentence “The quality of the photos was bad.” is an opinion sentence because it contains both an opinion entity (“bad”) and a product entity (“The quality of the photos”) . However, the sentence “I wanted a camera that the whole family could use.” is not an opinion sentence because it does not contain any opinion entities. In other words, opinion sentence is a sentence that contains the pair (O,P) where O is an opinion entity that takes on the label OPINION-POS or OPINION-NEG, P is a product entity that takes on one of the labels PROD-FEAT, PROD-PARTS, or PROD-FUNCTION.

4.4 Predicting Opinion Orientation

Opinion orientation for each identified product entity is further predicted as positive or negative in the subsequent step. For each product entity recognized in the opinion sentences, we search for the nearest opinion entity. The opinion orientation for the corresponding product entity is predicted as positive if the nearest opinion entity is positive; otherwise negative as the nearest opinion entity is negative.

Table 6: Experimental results on entity extraction.

Feature	Technique	Precision	Recall	F-score
$F1 = \{w_{-2}; w_{-1}; w_0; w_1; w_2\}$	CRF	0.737	0.413	0.529
	SVM	0.876	0.442	0.588
	Naive Bayes	0.646	0.647	0.646
	Bayesian Net	0.618	0.655	0.636
	Random Forests	0.684	0.701	0.692
$F2 = \{p_{-2}; p_{-1}; p_0; p_1; p_2\}$	CRF	0.646	0.272	0.383
	SVM	0.386	0.300	0.338
	Naive Bayes	0.499	0.555	0.525
	Bayesian Net	0.502	0.554	0.527
	Random Forests	0.526	0.555	0.540
$F3 = \{p_{-2}; p_{-1}; p_0; w_0; p_1; p_2\}$	CRF	0.786	0.364	0.498
	SVM	0.709	0.447	0.548
	Naive Bayes	0.617	0.655	0.635
	Bayesian Net	0.626	0.663	0.644
	Random Forests	**0.687**	**0.708**	**0.697**
$F4 = \{w_{-2}; w_{-1}; p_{-1}; w_0; p_0; w_1; p_1; w_2\}$	CRF	0.886	0.435	0.584
	SVM	0.861	0.477	0.614
	Naive Bayes	0.644	0.679	0.661
	Bayesian Net	0.652	0.682	0.666
	Random Forests	0.682	0.700	0.691
$F5 = \{w_{-2}; w_{-1}; w_0; p_0; w_1; w_2\}$	CRF	0.868	0.424	0.570
	SVM	0.878	0.462	0.605
	Naive Bayes	0.666	0.657	0.661
	Bayesian Net	0.652	0.677	0.664
	Random Forests	0.681	0.697	0.689
$F6 = \{(p_{-2}, p_{-1}); (p_{-1}, p_0); (p_0, p_1); (p_1, p_2); (w_{-1}, w_0); (w_0, w_1); w_{-2}; w_{-1}; p_{-1}; w_0; p_0; w_1; p_1; w_2\}$	CRF	0.875	0.449	0.593
	Naive Bayes	0.643	0.665	0.654
	Bayesian Net	0.644	0.668	0.656
	Random Forests	0.665	0.672	0.668
$F7 = \{(p_{-2}, p_{-1}); (p_0, p_1); (p_2, p_3)\}$	CRF	0.509	0.123	0.198
	Naive Bayes	0.438	0.522	0.476
	Bayesian Net	0.467	0.519	0.492
	Random Forests	0.425	0.470	0.446
$F8 = \{(w_{-2}, w_{-1}); (w_0, w_1); (w_2, w_3)\}$	CRF	0.342	0.222	0.269
	Naive Bayes	0.517	0.448	0.480
	Bayesian Net	0.510	0.462	0.485
	Random Forests	0.463	0.377	0.416

5. EXPERIMENTS

We use the online product reviews from Amazon as our evaluation data set and around 100 reviews for Camera Canon were manually tagged by experts. All the experiments were run on a 2.1GHz PC machine with 4GB RAM running Linux OS. To obtain unbiased evaluation results, we performed a 5-fold cross-validation. We developed our CRF model using CRF++ package [7]. Table 6 shows the precision, recall, and F-score of entity extraction using different feature sets. We used unigram features in the feature sets $F1$ to $F5$. A combination of unigram and bigram features in the feature set $F6$ and bigram features in the feature sets $F7$ and $F8$, where (w_0, w_1) is a bigram feature that consists of both the current and the next word. We found that the best result in terms of F-sore for CRF model is achieved by using the feature set $F6$, but the recall was relatively low compared with high precision obtained. For training the Naive Bayes, Bayesian Network and Random Forests, we used the Weka data mining tool [4]. We developed our SVM model using YamCha tool (Kudo and Matsumoto, 2001). The F-score results for entity extraction using the feature sets $F1$ to $F5$ by different models are presented in Figure 2. Table 7 and Table 8 show the results of opinion sentence extraction and opinion orientation predictions using the proposed methods. We used the feature set that gives the best result for each classifier.

Table 7: Experimental results on opinion sentence extraction.

Feature set	Technique	Precision	Recall	F-score
$F8$	CRF	0.977	0.512	0.672
$F4$	SVM	0.977	0.500	0.661
$F4$	Naive Bayes	0.655	0.429	0.518
$F4$	Bayesian Net	0.595	0.595	0.595
$F3$	Random Forests	0.957	0.524	0.677

6. DISCUSSION

In terms of relative performance of the proposed techniques for entity extraction, CRF tends to perform the worst for almost every feature set and Random forests tends to

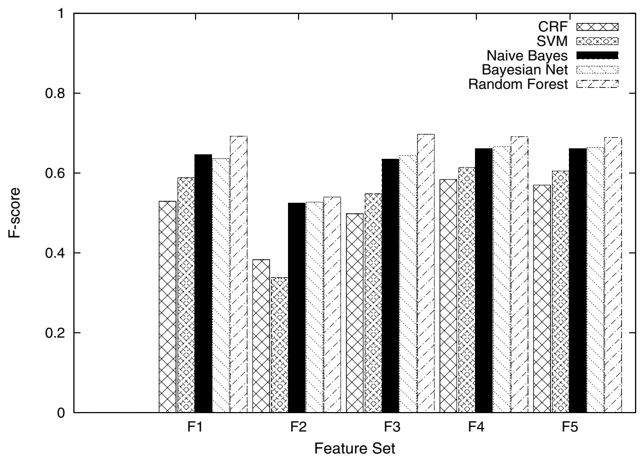

Figure 2: *Results of entity extraction.*

Table 8: Experimental results on opinion orientation prediction.

Feature set	Technique	Precision	Recall	F-score
$F8$	CRF	0.950	0.232	0.373
$F4$	SVM	0.976	0.488	0.650
$F4$	Naive Bayes	0.608	0.378	0.466
$F4$	Bayesian Net	0.535	0.458	0.494
$F3$	Random Forests	0.955	0.506	0.662

perform the best. The CRF has achieved the best precision but the recall value is much low. The best F-score is achieved by the Random forests classifier using the feature set $F3$. We experimented with a combination of unigram and bigram features in the feature set $F6$ but the performances of Naive Bayes, Bayesian Net, and Random forests are not improved by adding bigram features to the unigram features. The bigram features only improve the results of the CRF. For opinion sentence extraction and opinion orientation prediction, the best results have also been obtained by using the random forests classifier.

7. CONCLUSION

In this paper, we present a formal study of classification techniques on identifying entities of interest in product reviews, and propose to integrate the classification techniques into solving a new emerging problem, opinion mining and extraction. Various classification models have been examined and compared. Our evaluation shows that the Random forests classifier outperforms Naive Bayes, Bayesian network, Conditional Random fields, and support vector machines. The experimental results also demonstrate good classification performance is closely related to the use of linguistic features such as part-of-speech (POS), and context features such as surrounding contextual clues of words.

Future directions include the development of more sophisticated model to address opinion orientation classification that can combine various evidence sources. We are looking at integrating the statistical method, such as association rule mining, and linguistic resources (e.g., grammatical rules, specific language phenomena) to formulate this joint model. We are also expanding data sets and tagging new data sets from multiple domains to further examine the feasibility and effectiveness of classification techniques in the opinion mining application.

8. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments and suggestions. This publication was made possible by NIH grant number P20 RR016471 from the INBRE program of the National Center for Research Resources.

References

[1] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

[2] G. Carenini, R. T. Ng, and E. Zwart. Extracting knowledge from evaluative text. In *K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture*, pages 11–18, New York, NY, USA, 2005. ACM.

[3] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Mach. Learn.*, 29(2-3):131–163, 1997.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software. 2009. Software available at http://www.cs.waikato.ac.nz/ml/weka/.

[5] M. Hu and B. Liu. Mining and summarizing customer reviews. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, New York, NY, USA, 2004. ACM.

[6] W. Jin and H. H. Ho. A novel lexicalized hmm-based learning framework for web opinion mining. In *ICML*, page 59, 2009.

[7] T. Kudo. Crf++: Yet another crf toolkit, 2005. Software available at http://crfpp.sourceforge.net.

[8] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[9] B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 342–351, New York, NY, USA, 2005. ACM.

[10] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL*, pages 271–278, 2004.

[11] B. Pang, L. Lee, and S. Vaithyanathan. 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, 2002.

[12] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

[13] B. Settles. Biomedical named entity recognition using conditional random fields and rich feature sets. In N. Collier, P. Ruch, and A. Nazarenko, editors, *COLING 2004 International Joint workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP) 2004*, pages 107–110, Geneva, Switzerland, August 28th and 29th 2004. COLING.

[14] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 134–141, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[15] K. Toutanova. A part-of-speech tagger, 2008. Software available at http://nlp.stanford.edu/software/tagger.shtml.

[16] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417–424, Morristown, NJ, USA, 2002. Association for Computational Linguistics.