

Classification of Opinions with Non-affective Adverbs and Adjectives

Marina Sokolova
EHIL,
Children's Hospital of Eastern Ontario
Ottawa, Ontario, Canada
msokolova@ehealthinformation.ca

Guy Lapalme
RALI,
Université de Montréal
Montréal, Québec, Canada
lapalme@iro.umontreal.ca

Abstract

We propose domain-independent language patterns that purposefully omit the affective words for the classification of opinions. The information extracted with those patterns is then used to analyze opinions expressed in the texts. Empirical evidence shows that opinions can be discovered without the use of affective words. We ran experiments on four sets of reviews of consumer goods: books, DVD, electronics, kitchen, and house ware. Our results support the practical use of our approach and its competitiveness in comparison with other data-driven methods. This method can also be applied to analyze texts which do not explicitly disclose affects such as medical and legal documents.

1 Introduction

Opinion and sentiment analysis has recently received much attention from researchers in the Natural Language Processing (NLP) and Machine Learning (ML) communities. Most of the research addresses the primary clues of sentiments in a binary setting, e.g. positive/negative word orientation (good vs evil), subjective/objective statements (The movie was awesome vs We saw the movie yesterday), texts belonging to positive/negative opinion categories (I recommend this camera ... vs This book is awful ...). In this work, we concentrate on learning opinion from complete texts, classifying the texts as positive or negative.

In opinion learning, NLP and ML research mainly concentrates on the emotional *polarity* of texts which is expressed through the use of affective words (This is an excellent view shows positive polarity, excellent is an affective word). In this work, we, however, propose that learning opinions should allow for the use of the word categories other than affective. We present a method which uses non-affective adjectives and adverbs (future, full, perhaps), supplemented by degree pronouns and mental and modal verbs, to determine whether a text bears a positive or a negative opinion label.

The method engineers features by using the intrinsic characteristics of a language and avoids extensive and elaborate computational mechanism. Methods used to

classify complete texts according to opinions and sentiments usually employ automated feature selection, e.g., [17, 15]. Although such methods can be applied to different domains, they sometimes involve complex optimization problems, e.g., *NP*-hard approximation problems [2].

We concentrate on expressions of stance (maybe, necessary), degree (extremely, any), time (ago, now), frequency (rare, again), size (short, high), quantity (many, few), and extent (big). We show that these indicators reliably represent texts in opinion learning. We organize the corresponding word categories – stance/degree/time/frequency adverbs, frequency/size/quantity adjectives, degree pronouns – into a semantic hierarchy. Its lowest level works with words; the middle level generalizes word categories into groups; the highest level applies to the text as a whole. The hierarchy avoids the use of emotionally-charged words. We use the hierarchy to extract lexical features from texts. Next, we use the features to represent texts in a series of machine learning experiments.

Empirical evidence obtained on four data sets shows reliability of our approach. The presented method can be applied to analyze texts which do not explicitly disclose affects, e.g., medical and legal documents. This work extends preliminary studies presented in [22]. The rest of the presentation is organized as follows: we introduce word categories used in the text representation, then the hierarchy is presented, followed by description of the information extraction procedure and empirical results; discussion of related work, results and future work conclude the paper.

2 Text representation

Studies of sentiment and subjectivity analysis mostly concentrate on the use of the affective words in expression of sentiments and opinions. Some opinion studies use topic and domain words and affect-neutral verbs [18, 21]. We propose that words which emphasize quantitative properties (high, some), time (old, yesterday) and confidence in happening (can, necessary, probably) can be used in learning opinions.

Such words constitute detailed, *specific*, description of an object or action [4, 9]. We organize them in the

following word categories:

1. pronouns of degree (**everybody**);
2. adverbs
 - (a) time (**yesterday**),
 - (b) frequency (**often, rarely**),
 - (c) degree (**only**),
 - (d) stance (**necessary**);
3. adjectives
 - (a) size/quantity/extent (**large**),
 - (b) time (**old**),
 - (c) relational (**different**);
4. comparative and superlative adjectives of the listed above categories (**largest, older**);
5. order words
 - (a) ordinal numbers (**third**)
 - (b) cardinal numbers (**two**);
6. stance verbs
 - (a) modal verbs (**could**)
 - (b) mental verbs (**believe**).

We use word categories 1 – 5 to build the entry level for the hierarchical text representation. To be less domain and topic-dependent, we ignore subcategories closely related to the text topics, e.g., derived (**useless**), topical (**royal, economic**), affiliative (**American**), foreign (**ersatz**) [4]. We purposefully omit evaluative/emotive adjectives (**excellent, disgusting**) while constructing the lexical level of hierarchy. This omission allows emphasis on the role of quantitative description in text.

3 Semantic Hierarchy

In this section, we introduce a hierarchy of text representation. Starting from the bottom, the hierarchy defines the word categories used in detailed descriptions, then groups the categories into four types of comments, and finally combines the types into direct and indirect detailed categories. The levels of the hierarchy are represented by a set of rules which capture the essential characteristics of their language indicators. The rules have the following form:

non-terminal \rightarrow alternative₁ | alternative₂ | ...

where *non-terminal* must be replaced by one of the alternatives. Alternatives are composed of other non-terminals and **terminals** which are the pieces of the final lexical string. The lowest, lexical, level presents terminals for the word categories discussed in Section 2. The middle level organizes word categories into

semantic groups. The highest, the most general, level is concerned with text pragmatics.

We determined the list of terminals by finding seed words for these word categories in [4, 6] and added their synonyms from an electronic version of Roget's Interactive Thesaurus [20]. To accommodate negative comments, we added the negation rule. There are 303 rule terminals, not counting the negation terminals. Figure 1 shows the rules for finding detailed descriptions in text; within a hierarchy level, the rules are listed in alphabetical order.

We now list some implications of the rules presented by Figure 1:

direct Details presents primary clues of quantitative evaluation and attributes of the discussed issues. Two rules of the middle level provide factual information through the word categories of the lowest level:

Estimation lists the reference attributes: physical parameters, relative and absolute time.

Quantification expresses the broadness of the discussed reference by specifying its multiplicity, frequency and extent.

indirect Details presents secondary clues of the issue evaluation. Two rules of the middle level define indirect evaluation through the word categories of the lowest level:

Comparison presents a comparative evaluation of the discussed issues, their qualities and relations among them.

Confidence reflects on the certainty about the happening of events;

4 Feature Set Construction

We hypothesize that expressed opinions can be accurately learned from non-affective adverbs and adjectives. To evaluate our hypothesis, we apply the hierarchy to find and extract features for text representation. Our core assumption is the following: the hierarchy rule terminals emphasize important characteristics of the discussed issues.

Grammatically, the rule terminals are modifiers, amplifiers and identifiers. In sentences, such words usually precede their references, especially in conversational text [4]. The extraction of words which follow the rule terminals results in the set of words most emphasized in the text. The extraction procedure is presented on Figure 2; it has only one adjustable parameter *h*, whose value is determined during the empirical step.

To build the set of words most emphasized in the text, we look for words with a high probability of appearance on the right side of the rule terminal. We

<i>direct Details</i>	→	[<i>negation</i>](<i>Estimation</i> <i>Quantification</i>)
<i>indirect Details</i>	→	[<i>negation</i>](<i>Comparison</i> <i>Confidence</i>)
<i>Comparison</i>	→	<i>adjComparat</i> <i>adjRelation</i> <i>adjSuperlat</i> <i>numOrdinal</i>
<i>Confidence</i>	→	<i>advStance</i> <i>verbCognition</i> <i>verbModal</i>
<i>Estimation</i>	→	<i>adjPhysical</i> <i>adjTemp</i> <i>advTime</i>
<i>Quantification</i>	→	<i>adjDegree</i> <i>advDegree</i> <i>advFrequen</i> <i>numCardinal</i> <i>pronDegree</i>
<i>adjComparat</i>	→	longer smaller older ...
<i>adjDegree</i>	→	full rare usual ...
<i>adjPhysical</i>	→	deep long small ...
<i>adjRelation</i>	→	same different ...
<i>adjSuperlat</i>	→	highest longest oldest ...
<i>adjTemp</i>	→	belated future new ...
<i>advDegree</i>	→	extremely only roughly ...
<i>advFrequen</i>	→	often rarely sometimes ...
<i>advStance</i>	→	necessarily perhaps probably ...
<i>advTime</i>	→	ever now yesterday ...
<i>numCardinal</i>	→	one two ...
<i>numOrdinal</i>	→	first second ...
<i>pronDegree</i>	→	everybody few some ...
<i>verbCognition</i>	→	believe know think
<i>verbModal</i>	→	can could may might ...
<i>negation</i>	→	not no can't none ...

Fig. 1: Rules for the identification of detailed comments in text. “|” separate alternatives, square brackets indicate optional parts and parenthesis are used for grouping. Terminals are written in this font.

Step 1 build a bigram model of the data:
1. for sequences $w_{j-1}w_j, j = 1, \dots, m$, calculate the probabilities of their occurrence in the data;
2. disregard the sequences with the probability of 0;
Step 2 find words appearing on the right side of the terminal:
1. for each $t_i \in T$, extract bigrams t_iw_j where the pattern terminals appear on the left side;
2. build the unigram model of the extracted bigrams;
3. remove the terminal unigrams;
Step 3 find frequently modified and intensified words:
1. determine the parameter h ;
2. keep w_j with $n(w_j t_i) > h$.

Fig. 2: The procedure for finding and extracting frequently modified and intensified words in text. The procedure uses the same notations as equations (1) and (2). The adjustable parameter h is determined empirically.

estimate this probability $P(w_j|T)$ by computing:

$$P(w_j|T) = \sum_{t_i \in T} P(w_j|t_i), \quad (1)$$

$$P(w_j|t_i) = \frac{n(w_j|t_i)}{\sum_{j=1}^m n(w_j)} \quad (2)$$

where w_j is a word, T is the set of all terminals, t_i is a terminal, $w_j|t_i$ is the event where the word w_j appears after t_i in text ($t_i w_j$), m is the size of the data vocabulary, $n(x)$ is the number of occurrences of x in the data.

The idea behind the search procedure is the following: two-word sequences $t_i w_j$ - bigrams - which have terminals on their left side capture the modified and intensified words. After extracting such bigrams, we find modified and intensified words. By calculating the probability of the word occurrence after a terminal, we can find most frequently modified and intensified words. Concentrating on one-side bigrams prevents the multiple extraction of the same word.

In supervised learning experiments, each text is represented by a vector x_1, \dots, x_n, y , where x_i is a number of occurrences a word w_i , a feature, appearing in the text, and y is the opinion label. As a weighting factor, we use normalization of the vector attributes with respect to the number of words in the text. It eliminates the bias introduced by the length of the text. Based on the rule terminals and the extracted words, we construct three feature sets for text representation:

I terminals of *direct Details* rules enhanced by personal pronouns; h was determined by frequencies of personal pronouns;

II terminals of all the hierarchy rules enhanced by the most frequent extracted words; h was determined by frequencies of personal pronouns;

III the terminals and all the words extracted by the procedure presented in Figure 2; the cut-off threshold $h = 5$ was determined by using Katz smoothing to ensure reliability of data representation.

5 Empirical results

We ran experiments on data introduced in [5]. There are four sets of reviews of different consumer goods: books, DVD, electronics, kitchen and houseware. Each data set has 2000 labelled examples, all evenly split on 1000 positive and 1000 negative examples. Blitzer et al. deleted reviews they considered as having ambiguous opinions. A typical review contained abundance of information assigned to several fields: (i) product name, (ii) product type, (iii) unique id which often summarized the review contents, (iv) product rating,

(v) review helpfulness rating, (vi) the review title, (vii) the date, (viii) the review text, etc.

For this study, we extracted the review texts; see Figure 3 for samples of the extracted reviews; in those texts we have marked the features presented by the hierarchy (Figure 1) and constructed through the procedure (Section 4). Correspondence among information provided by different fields is left for future work. Review texts are long enough to provide meaningful communication and lexical information; Table 1 lists the descriptive statistics of the extracted data. These four sets allow us to compare our empirical results with those obtained by other methods. In order to establish how a speaker's detailed descriptions are related to her opinion, we apply *supervised learning* techniques that construct a function on a set of input and output pairs (\vec{x}, y) where \vec{x} represents a text and y is its opinion label (training data). This function is then used to predict opinion labels on previously unseen examples (testing data).

We want to establish a link between information extracted by patterns and the text opinion categories, e.g. positive or negative. We expect non-linear dependencies between the terminals' appearance in texts and a speaker's opinion. We applied decision-based C4.5, prototype-based K-NEAREST NEIGHBOR and kernel-based SUPPORT VECTOR MACHINE (SVM). SVM performed considerably better than other algorithms on all the four data sets. The algorithm is known for its high accuracy in text classification. SVM does not make any assumption about the data distribution and could work with non-linear dependencies, albeit on one level of learning. Further we report only the SVM's results. We use the Weka's implementation¹. Classification measures use the following counts:

Data class	Classified as <i>pos</i>	Classified as <i>neg</i>
<i>pos</i>	<i>tp</i>	<i>fn</i>
<i>neg</i>	<i>fp</i>	<i>tn</i>

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn} \quad (3)$$

evaluates the overall performance of SVM; tp and tn provide a detailed analysis of the algorithm's performance on positive and negative classes. We use ten-fold cross-validation to compute the three measures because of its generalization accuracy and the reliability of its results.

We compare text representations built on the three levels of rules presented by Figure 1. Table 2 reports learning results obtained on the three representations introduced in Section 4. As the baseline, we apply SVM on texts represented by the feature set *I*. All 62 selected words appear frequently in the data and provide substantial information about texts. These features include adverbs of degree and adverbs of frequency which were used by [3] in sentiment classification. Adding all *II* features makes a statistically significant difference in accuracy (paired t-test,

¹ <http://www.cs.waikato.ac.nz/ml/weka/>

Book Reviews	
Extracted review sample	Label
<i>This</i> thing sat on <i>my</i> shelf, half-read for <i>the</i> longest time. Only <i>the</i> notice of <i>the</i> upcoming release <i>this</i> November of Pynchon's next got <i>me</i> motivated enough to dig into <i>it</i> again . <i>It's not</i> that <i>it</i> isn't brilliant. No one else around can dazzle <i>you</i> with so much wit and wonder. The first encounter with <i>the</i> talking dog is <i>as</i> magical <i>as</i> anything <i>you'll</i> ever read. <i>And</i> it's not like this is <i>the</i> only Pynchon novel that takes some effort to get into.	pos
<i>I</i> picked up <i>the</i> first book in this series (The Eyre Affair) based purely on <i>its</i> premise and was left somewhat underwhelmed. Still , <i>the</i> potential for the series seemed so large that <i>I</i> went ahead and <i>read</i> this second one too ...	neg
Kitchen and Houseware Reviews	
Extracted review sample	Label
<i>i</i> absolutely love <i>this</i> product. <i>my</i> neighbor has four little yippers <i>and</i> <i>my</i> shepard/chow mix was antagonized by <i>the</i> yipping on <i>our</i> side of <i>the</i> fence. <i>I</i> hung <i>the</i> device on <i>my</i> side of <i>the</i> fence <i>and</i> <i>the</i> noise keeps <i>the</i> neighbors dog from picking "arguments" with <i>my</i> dog. all barking <i>and</i> fighting has ceased. all <i>the</i> surrounding neighbor <i>as well</i> as <i>me</i> can get a good nights sleep now	pos
<i>He</i> just looks away from where <i>the</i> spray emits—and barks again! <i>It</i> also doesn't work 100 % of <i>the</i> time...and <i>we're</i> not sure why. When <i>we</i> fill <i>it</i> , <i>it</i> seems to work fairly well right after but <i>it</i> either does not have as many sprays as <i>it</i> is supposed to, or <i>it</i> isn't working very long . <i>It</i> does work <i>well</i> for <i>my</i> other small dog who is not such a persistent barker. Terriers are just too stubborn to care if <i>they're</i> getting sprayed, <i>I</i> guess.	neg

Fig. 3: Samples of the extracted reviews. Terminals of rules in Figure 1 are in **bold**; the words found by procedure in Figure 2 are in italics; negations are ignored if they do not appear before the rule terminals. For each data set, the upper sample has a positive opinion label that labels the whole text whereas the lower sample has a negative opinion label.

Table 1: Customer-written reviews from Amazon.com pre-processed by J. Blitzer et al (2007). Texts (from all four data) they considered as ambiguous opinions were deleted.

Data	# examp	# pos	# neg	Tokens	Types	Aver length
Books	2000	1000	1000	349530	39811	175
DVD	2000	1000	1000	337473	39776	169
Electronics	2000	1000	1000	222862	20664	111
Kitchen	2000	1000	1000	188137	17296	99

$P = 0.010$). When we use all the extracted words, the statistically significant difference increases (paired t-test, $P = 0.003$). For each representation, the classification results are close across the data sets. However, there is a remarkable difference between the Electronics data and the three other sets. Let's consider true classification of the positive and negative reviews. For Books, DVD, Kitchen sets, the positive reviews are always classified more correctly than the negative reviews (the only exception is a tie for Books on the II representation). The Electronics set provides the opposite results: the negative reviews are always classified more correctly than the positive ones.

It is interesting to observe that only 303 rule terminals, i.e., the II features, already provide an opinion accuracy of 74% – 78%. These are reliable results for opinion learning, since human agreement on whether a message provides a positive or a negative opinion about the discussed topic could be 78% for

positive opinions (tp) and 74% for negative opinions (tn) [16]. On the four data sets, all extracted features provide accuracy of more than 80%. There are 1999 terminals and extracted words used. Reported tp and tn show that the extracted features provide a well-balanced classification of the classes: on all the four data sets difference between tp and tn is $< 10\%$. This can be only achieved if the algorithm is successful in learning both positive and negative classes.

6 Related work

Opinion and sentiment analysis that focuses on whether a text is subjective, bears positive or negative opinion or expresses the strength of an opinion has received a vast amount of attention in the recent years. In this section, we discuss only research with application to *complete texts*, which sometimes are referred to

Table 2: SVM’s classification accuracy of positive and negative opinions, in per cent. Accuracy (*Acc*) shows how effective is the approach in prediction of previously unseen examples, *tp* – prediction of positive examples, *tn* – prediction of negative examples. All the reported values are obtained with near-linear kernels ($\exp = 0.75, \dots, 0.92$) and a small error penalty for misclassification of training examples ($C = 1.05, \dots, 1.21$).

Data	Text Features								
	I			II			III		
	<i>Acc</i>	<i>tp</i>	<i>tn</i>	<i>Acc</i>	<i>tp</i>	<i>tn</i>	<i>Acc</i>	<i>tp</i>	<i>tn</i>
Books	68.70	69.00	68.40	74.75	74.70	74.80	80.20	81.05	79.35
DVD	70.80	73.25	68.35	73.80	76.70	70.90	80.50	84.10	76.80
Electronics	69.50	66.50	72.50	75.70	74.25	76.95	82.40	76.85	87.85
Kitchen	69.10	70.05	68.15	76.50	78.25	74.75	85.20	88.20	82.20

as documents. We omit research on sentiment/opinion analysis of terms, phrases, sentences and other text segments; references and discussion can be found in [7].

Some of this work relied on a list of characteristics of reviewed products. Hu and Liu extracted features based on association rule mining algorithms in conjunction with frequency to extract main product characteristics [10]. These characteristics are then used to extract adjacent adjectives which are assumed to be opinion adjectives. Later, these opinion adjectives are used to find product characteristics that are mentioned only once or few times. In contrast, we opted for a domain-independent method that does not involve the use of the domain’s content words. Popescu and Etzioni (2005) extracted product characteristics from noun phrases in the data and matched them with known product features. In contrast, we opted for a domain-independent method that does not involve the use of the domain’s content words.

For automating recognition and the evaluation of the expressed opinion, complete texts are represented through *N*-grams or patterns and then classified as opinion/non-opinion, positive/negative, etc. [19]. In [5], the authors combine supervised and semi-supervised structural correspondence learning to classify the four data sets. They use fully automated feature selection based on frequency and the mutual information of words. However, the difference in evaluation technique does not allow us to directly compare the obtained results.

Syntactic and semantic features that express the intensity of terms are used to classify the text opinion intensity [23]. Benamara et al. studies the impact of combining adverbs of degree with adjectives for the purpose of opinion evaluation [3]. Our approach deals instead with opinion analysis which is broader than the analysis of sentiments. We focus on the formalization and utilization of non-emotional lexical features.

Except sentiment analysis, machine learning is used to study opinions from the point of view of predictive power [12], strength [23], and also in summarization and feature extraction studies [8]. Although Kim and Hovy (2007) generalized bi- and trigrams found

in texts (e.g. NDP will win and Liberals will win became Party will win), they did it bottom-up, without providing theoretical background. We, instead, used a top-down hierarchical approach based on pragmatic and lexical rules. Wilson et al (2006) concentrated on learning subjective vs objective sentences and sentence clauses. Their initial manual clues included verbs of judgment (*reprove*, *vilify*); in the final text representation they use syntax clues. In contrast, to represent texts, we look for non-affective adverbs and adjectives.

Consumer and expert-written product reviews are intensively studied by psychology, marketing, etc. [14]. Kamakura et al. analyzed movie reviews written by 46 experts [11]. They were one of the first to do research on the information contained in unguided reviews and built a model linking the expert’s history of movie evaluations and quantitative information found in reviews. However, the authors did not actually analyze the texts or language cues contained in the reviews. Their results showed that experts are not uniformly informative across movies that they scored differently. This study supports our views that specific information in product reviews relates to the speaker’s opinion, although their results were obtained on expert-written reviews.

The research listed above does not consider a hierarchy of opinion disclosure. A pragmatic-lexical hierarchy of semantic verb categories was proposed in our previous work [21]. We showed that the hierarchy worked well in environment where negative opinions were expressed indirectly, without the use of negative adjectives or adverbs, e.g. debates in the US federal Senate. In contrast, in the current work, we concentrated on the use of non-affective adverbs and adjectives and degree pronouns.

7 Conclusion and future work

We have proposed a hierarchical text representation (the highest level that considers a text as a whole) and derived rules (the middle level), as well as the rules’ terminal categories (the lowest level that works with words). The terminals were used to extract the

emphasized information from the text. Our goal was to build domain-independent rules that do not rely on domain content words and emotional words. Further, the additionally extracted words and the rule terminals were used to represent texts in learning experiments.

In our experiments, we used four data sets which texts were gathered in different domains. We studied the relevance of detailed, specific words to the learning of positive and negative opinions. Our empirical results show that the corresponding lexical features are effective in learning opinions.

Our approach can be applied to analyze language in texts which traditionally lack emotive and affective words. Medical and legal domains provide us with such texts. It is worth note that these two domains attract more and more attention of Text Data Mining community as evidenced by many publications, for example, in the Journal of the American Medical Informatics Association ² and the International Journal of Law and Information Technology ³. Another venue can be a joint analysis of information contained in different fields of reviews.

In this study, we applied supervised learning algorithms that required labeled data, which is usually restricted. Our future work will be to incorporate unlabeled data and apply semi-supervised approaches, e.g., a framework of learning predictive structures [1]. A notable drawback of this framework is the need to find “good” features for data representation. We can overcome this problem by using the results of the current study. Another direction worth trying is to incorporate more pragmatic knowledge into feature construction, e.g., likelihood of the use of features in written or spoken language obtained from the British National Corpus [13].

Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Ontario Centre of Excellence.

References

- [1] R. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853, 2005.
- [2] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira. Analysis of representations for domain adaptation. In *Neural Information Processing Systems*, 2006.
- [3] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than the adjectives alone. In *Proceedings of International Conference on Weblogs and Social Media (ICWSM’2007)*, 2007.
- [4] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan. *Longman Grammar of Spoken and Written English*. Longman, 1999.
- [5] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 440–447. Association for Computational Linguistics, 2007.
- [6] D. Bolinger. *Degree Words*. Mouton De Gruyter, 1972.
- [7] E. Breck, Y. Choi, and C. Cardie. Identifying expressions of opinion in context. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-2007)*, pages 2683–2688, 2007.
- [8] O. Feiguina and G. Lapalme. Query-based summarization of customer reviews. In *Proceedings of the 20th Canadian Conference on Artificial Intelligence (AI’2007)*, pages 452–463, 2007.
- [9] M. Hamilton. Extending an information processing model of language intensity effects. *Journal of Language and Social Psychology*, 17:109–143, 1998.
- [10] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’04)*, pages 168–177, 2004.
- [11] W. Kamakura, S. Basuroy, and P. Boatwright. Is silence golden? an inquiry into the meaning of silence in professional product evaluations. *Quantitative Marketing and Economics*, 4:119–141, 2006.
- [12] S.-M. Kim and E. Hovy. Crystal: Analyzing predictive opinions on the web. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1056–1064, 2007.
- [13] G. Leech, P. Rayson, and A. Wilson. *Word Frequencies in Written and Spoken English: based on the British National Corpus*. Longman, 2001.
- [14] B. Loken. Consumer psychology: Categorization, inferences, affect, and persuasion. *Annual Review of Psychology*, 57:453–485, 2006.
- [15] V. Ng, S. Dasgupta, and S. M. N. Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 611–618. Association for Computational Linguistics, 2006.
- [16] K. Nigam and M. Hurst. Towards a robust metric of opinion. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pages 98–105. AAAI, 2004.
- [17] B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, pages 115–124. Association for Computational Linguistics, 2005.
- [18] A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Proceedings of HLTC/EMNLP 2005*, pages 339 – 346, 2005.
- [19] E. Riloff, S. Patwardhan, and J. Wiebe. Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 440–448. Association for Computational Linguistics, 2006.
- [20] roget. Roget’s interactive thesaurus, 2006. <http://thesaurus.reference.com/>.
- [21] M. Sokolova and G. Lapalme. Verbs speak loud: Verb categories in learning polarity and strength of opinions. In *Proceedings of the 21st Canadian Conference on Artificial Intelligence (AI’2008)*, pages 320–331. Springer, 2008.
- [22] M. Sokolova and G. Lapalme. Learning opinions without using emotional words. In *Proceedings of the 22nd Canadian Conference on Artificial Intelligence (AI’2009)*, pages 253–256. Springer, 2009.
- [23] T. Wilson, J. Wiebe, and R. Hwa. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99, 2006.

² <http://www.jamia.org/>

³ <http://ijlit.oxfordjournals.org/>