Verbs Speak Loud: Verb Categories in Learning Polarity and Strength of Opinions

Marina Sokolova and Guy Lapalme

Département d'informatique et de recherche opérationnelle
Université de Montréal
sokolovm@iro.umontreal.ca,
lapalme@iro.umontreal.ca

Abstract. We show that verbs reliably represent texts when machine learning algorithms are used to learn opinions. We identify semantic verb categories that capture essential properties of human communication. Lexical patterns are applied to construct verb-based features that represent texts in machine learning experiments. Our empirical results show that expressed actions provide a reliable accuracy in learning opinions.

1 Introduction

The English saying *Actions speak louder than words* states that we learn more from people's actions than from their words. This statement may express human wisdom and be true for human learning, but could it be also true for machine learning as well? We show that, under certain conditions, people's expressed actions, i.e. verbs that they use, provide for more accurate machine learning of opinions than all words, even when the latter are re-enforced with the history of speakers' opinions. We consider verbs that indicate stronger personal physical actions (speak, write), mental and sensual actions (think, feel), and intentions (could, should, can, will). We apply ideas from communication theory to build semantic verb categories, we then formalize their use by language patterns from which we construct text features.

We apply machine learning techniques (regression and classification) to texts represented by the verb-based features. These texts are debates from the US Congress and consumer-written forum messages. Regression problems for opinion learning have not been studied before. Previous opinion studies mainly focused on binary classification [1], while sometimes solving a three-class classification problem [2]. Our combination of regression and classification learning provides a more detailed opinion analysis and relies only on the data. Other methods achieve a similar accuracy by adding personal information about speakers, e.g. history of previous comments [3]. However, additional information is not often available. Our method's accuracy is close to human-human agreement on positive and negative sentiments, when it is based on verbs [4]. Our results complement opinion and sentiment mining, a research area whose results are in an increasing demand from government, media and business practitioners.

S. Bergler (Ed.): Canadian AI 2008, LNAI 5032, pp. 320–331, 2008.

2 Semantic Verb Categories

Sentiment and opinion analysis are more subjective and difficult to solve than traditional text classification and mining tasks [5]. We propose a method that uses interpersonal aspects of communication and views language as a resource of accomplishing goals within context of social interactions. This approach is reminiscent of the Systemic Functional Linguistics developed by Halliday [6].

We consider that opinion can be emotional or rational. Emotional opinion may be expressed by attitude (enjoy, hate) and, partially, by the perception of the situation (smell, feel). Rational opinion may require the person to list facts such as events (meet, send), the state of affairs (depend, have). Possibility, necessity, politeness or irony can be directly shown by the use of primary modals (can, will) or more conditional secondary modals (could, should) [7].

We also consider that an informal, loosely structured, spoken-like language differs from a formal, structured, written-like one. Verbs denoting activity (play, write, send) and cognition verbs (think,believe) are the two most frequent categories when opinions are expressed in spoken-like language. Activity, the largest among verb categories, is the most frequent in all types of texts. The high frequency of mental verbs is specific for spoken language [8,9]; thus, we separate mental verbs into three categories: perception, attitude and cognition. Verbs denoting process (live, look, stay) often appear in written language, sometimes as often as activity verbs [10]. Table 1 shows the semantic categories we built from the seed verbs given in Leech[7] to which we added synonyms from Roget's Interactive Thesaurus [11].

Table 1. The list of non-modal verb categories and examples of corresponding verbs

Category	Refers to	Examples
cognition	mental state	consider, hope, think, know
perception	activity of the senses	see,feel,hear
attitude	volition and feeling	enjoy,hate,love
activity	a continuing action	read, work, explain
event	happening or transition to another state	become,reply,pay,lose
process	continuing or eventual change of state	change,increase,grow

We generalize the use of the verb categories by means of patterns defined by grammar rules in Figure 1. The semantic categories make up the lower level. The intermediate level forms four groups. Physical action verbs are considered to be more direct in expressing opinions, whereas mental verbs correspond to more hesitation and condition [12], e.g. We played well expresses more confidence than I think we played well. At the highest level, we consider whether the person involves herself in evaluation (*firstPerson*) or projects it on interlocutors (*you*).

The rules at the top of Figure 1 define the expression of an author's involvement, either in the form of closeness or distancing.

```
closeness        → firstPerson (logic | physicalAction | mentalAction | state)
distancing       → you (logic | physicalAction | mentalAction | state)

logic            → primaryModal | secondaryModal
physicalAction   → [modifier] (activity | event | process)
mentalAction     → [modifier] (cognition | perception | attitude)
state            → [modifier] havingBeing

firstPerson      → I | we
primaryModal     → can | may | will | shall | have to | must
secondaryModal   → could | might | should | would
activity         → read | work | explain | ...
event            → become | reply | pay | send | ...
process          → change | increase | stay | ...
cognition        → believe | consider | hope | ...
perception       → feel | hear | see | smell | taste
attitude         → enjoy | fear | like | love | hate
havingBeing      → have | be | depend | consist | ...
modifier         → negation | adverb
```

Fig. 1. Grammar rules generalizing the use of verb categories. | separate alternatives, [] indicate optional parts and parentheses are used for grouping.

We now outline some involvement implications for each rule:

closeness uses `I` or `we` to indicate a direct involvement of the author. Its subrules indicate different degrees of the author's involvement:

- ***logic*** expresses permission, possibility, and necessity as the representation of logic, and superiority, politeness, tact, and irony as the representation of practice:
  - ***primaryModals*** such as can and may express direct possibility, permission or necessity of an action.
  - ***secondaryModals*** use a more polite, indirect and conditional pattern than a primary modal and indicate more hypothetically and tentatively the author's intentions.
- ***physicalAction*** denotes an author's goal-oriented actions (*activity*), actions that have a beginning and an end (*event*) and a series of steps towards a defined end (*process*). The pattern corresponds to a direct and active involvement of the author.
- ***mentalAction*** uses mental action verbs, being more polite and tentative, that are a common face-saving technique and that mark openness to feedback.
- ***state*** indicates personal characteristics and corresponds to actions without definite limits and strong differentiations.

distancing uses second person pronouns and shows how an author establishes distance from the matter.

3 Feature Engineering

To validate our hypothesis on actions, we looked at various types of information provided by verbs. We considered a general information provided by the use of verb categories, including verb past and continuous forms which reflect uncertainty of speakers [13], specific information resulting from the use of individual verbs, and information enhanced by words collocated with the pattern terminals.

We constructed three feature sets based on the pattern terminals (Figure 1):[1]

I The first feature set generalizes the use of word categories, separating their use in present, past and continuous forms. We are interested in density and diversity of the words in each category. For a text T, for each category C_j, the number of word tokens $N_j(T) = \sum_{t_i \in C_j} n(t_i)(T)$ and the number of word types $V_j(T) = \sum_{t_i \in C_j} I(t_i)(T)$ estimate these two parameters respectively; $n(x)$ denotes the number of occurrences of x; t_i is a terminal token; $I(x)$ equals 1 if x appears in T and 0 otherwise.

As a result, for each non-modal verb category we built six features. To represent modal verbs, we built four features: two – for primary modals, two – for secondary modals [2]. Altogether, there are 40 features with numerical attributes.

II The next set has individual terminals as its features. Each terminal is represented by its occurrences in the text: $N_i(T) = \sum_{t_i \in t} n(t_i)(T)$. There are 301 features.

III The third feature set expands the pattern terminals with words appearing with a high probability after or before a terminal. We estimate this probability by computing:

$$P(w|t) = \frac{\sum_{t_i \in t} n(w, t_i)}{\sum_{j=1}^{m} n(w_j)} \tag{1}$$

t is the set of all terminals; w, t_i is the event where the word w appears immediately after or before t_i; m is the size of the data vocabulary. In practice, we find collocated words with the following extraction procedure:

Step 1: we build the bigram model of data $w_{k-1}w_k$; bigrams are used because they avoid multiple extraction of the same word with respect to the same terminal;

Step 2: we extract bigrams $t_i w_k$ where the pattern terminals appear on the left side; this captures modified and intensified words appearing on the bigram's right side;

[1] Note that only negations preceding a terminal will appear in text representation.

[2] Modal verbs do not have past and continuous forms.

Step 3: we find $n(w_k|t_i)$ – occurrences of words appearing on the right side of terminals; the resulting $n(w_k|t_i)$ shows what words w_k were modified and intensified most (recall that rule terminals can be synonyms; in this case they may modify and intensify the same words);

Step 4: we keep w_k with $n(w_k|t_i) > 5$; the occurrence threshold 5 is chosen based on the language modeling characteristics.

Attributes are normalized to eliminate the bias introduced by the text length.

4 Data

We experimented with two types of data sets. One, consumer-written product reviews posted on the web, representing loosely-edited, free structured texts, presumably written by general population. The other, records of the US Congress debates, are structured, edited and professionally written. There are some commonalities between the data sets: each set covers several topics, i.e., several products reviewed by consumers and multiple legislations debated by congressmen; for each data set, its records come from hundreds of contributors. Both characteristics ensure that our empirical results will not be confined to a specific group of people or events.

We use consumer reviews data set introduced in [14]. Consumer reviews are posted on a web site dedicated to consumer goods evaluation. They are written by users of consumer goods. Although with some restrictions, reviews satisfy the following criteria of spoken language: they are spontaneous, loosely structured and socially interactive. The data set in our experiments consist of 314 reviews evaluating consumer electronics. The set size is 71,711 words (tokens) and 6,908 distinct words (types). Some text segments, but not all, have been manually tagged by Hu and Liu according to positive or negative opinions expressed by the reviewers. The following excerpt – from a positive review of a digital camera g3 – has a positive score 3:

> this `is` *my* first digital camera , and what a 'toy' *it* `is`! *i* `am` a software engineer and `am` very keen into technical details of everything *i* `buy`, *i* `spend` around 3 months before `buying` the digital camera; [3] and *i* `must say`, g3 worth every single cent . . . .

For the regression problem, three numerical labels are computed for each text for learning the strength of opinions:

- the number of positive tags; its range: 0 – 24;
- the number of negative tags; its range: −18 – 0;
- a signed sum of the two numbers; the range is −13 –24.

In the classification problem for learning the strength of opinions, we apply unsupervised equal-frequency discretization to each numerical label [15]. This makes fine distinctions between data entries that are close (e.g., with 4-6 positive opinion labels) and ignores big differences among data entries that are far apart (e.g., with 18-24 positive opinion labels).

We also used 1,117 Congress debate records [3]. Congress data are recorded speeches made by congressmen during legislation debates. Congress debates speeches are usually prepared in advance by congressmen and their assistants and read by congressmen during debate time. They are non-spontaneous, well-structured, and often close to immediate interaction. Each text is a recorded speech of the member of the US Congress, that either supports or opposes a proposed legislature. The debate record size is 1,139,470 words (tokens) and 21,750 distinct words (types). Thomas et al. labeled texts by numerical polarity scores, computed by SUPPORT VECTOR MACHINE. SVM builds a decision surface that separates positive and negative texts. The distance from a text to the surface defines the text's score value. The text's position with respect to the surface defines whether the score is positive or negative. We keep their scores as the data labels. The following excerpt has a positive score of 0.72:

> *we* `have known` that small businesses and working families `need` tax relief, and *we* `have fought` hard to `make` that happen so that *we* `see` the opportunity right there ....

For the data, the opinion labels range from -1.56 to 1.74. For classification purposes, we use the score signs as the data labels.

5 Language Pattern Distribution

In Figure 1 of Section 2, we outlined implications for the use of verb categories. To compare their use in consumer reviews and Congress debates, we computed the rule distribution reported in Table 2.

The upper part of Table 2 reports percentage held by the rules and the subrules for consumer reviews data. 100% is the total use of two pattern rules. To simplify the table, we combined results for pronouns `I` and `we` and for patterns with and without modifiers. In consumer reviews the use of mental verbs prevails over the use of physical action verbs. In fact, patterns with cognition verbs are the most frequent among all action patterns. Cognition verbs are common as a face-saving technique, thus their frequent use moderates the level of the author's involvement.

Since reviews tell about the experience of the authors with consumer goods, we could expect a frequent and diverse use of the *activity* verb categories. However, this is not the case: *closeness* patterns of the *physicalAction* verbs use more frequently *process* verbs, but not *activity* or *event* ones. Among distancing patterns, consumer reviews use straightforward *physicalAction* more often than any other sub-rule. The Congressional debate patterns are distributed differently than the ones in consumer reviews (the lower part of Table 2). Frequent use of physical action patterns demonstrates a stronger level of involvement. This also can be seen through a frequent use of activity verbs. In contrast with consumer reviews, congressmen do not use second-person pronoun often: only 4.37% of the rule patterns belongs to distancing while it is 40.90% in consumer reviews.

Comparison of extracted patterns supports the assumption that patterns vary across different communication environments. In Congressional debates a combined share of the most straightforward *primaryModal* and *activity* patterns

Table 2. Percentage held by the patterns in the consumer reviews and the US Congress debates data. 100% is the total number of the used rules in each data set. Results are combined for `I` and `we` and for patterns with and without modifiers. The largest percentage is given in **bold**, the second largest – in ***bold italic***, the smallest – in *italic*.

Consumer review data					
Rules	%	Subrules	%	Verb categories	%
closeness	**59.10**	*logic*	***17.77***	*primaryModal*	**14.51**
				secondaryModal	3.26
		mentalAction	***17.77***	*cognition*	9.27
				attitude	7.14
				perception	1.36
		physicalAction	13.04	*process*	8.42
				activity	3.46
				event	1.16
		state	10.52	*havingBeing*	***10.52***
distancing	*40.90*	*physicalAction*	**19.43**	*event*	7.72
				process	6.21
				activity	5.50
		logic	11.27	*primaryModal*	8.27
				secondaryModal	3.00
		mentalAction	9.58	*attitude*	5.32
				cognition	2.40
				perception	1.86
		state	*0.62*	*havingBeing*	*0.62*
Congress debates data					
Rules	%	Subrules	%	Verb categories	%
closeness	**95.63**	*logic*	**31.15**	*primaryModal*	**26.57**
				secondaryModal	4.58
		mentalAction	***29.16***	*cognition*	***17.76***
				attitude	10.43
				perception	0.97
		physicalAction	28.60	*process*	13.91
				activity	13.07
				event	1.62
		state	6.72	*havingBeing*	6.72
distancing	*4.37*	*logic*	3.81	*primaryModal*	3.29
				secondaryModal	0.52
		physicalAction	0.33	*event*	0.28
				process	*0.05*
		mentalAction	*0.23*	*perception*	0.23

is 39.64% of all the rules, whereas in consumer reviews their combined share is 17.97%. Patterns for *distancing* substantially vary across two data: they are frequently present in consumer reviews (40.90%) and rarely found in Congress debates (4.37%).

To illustrate the difference between the verb distributions we projected them with respect to *closeness* vs *distancing* axes. The plot on Figure 2 shows the

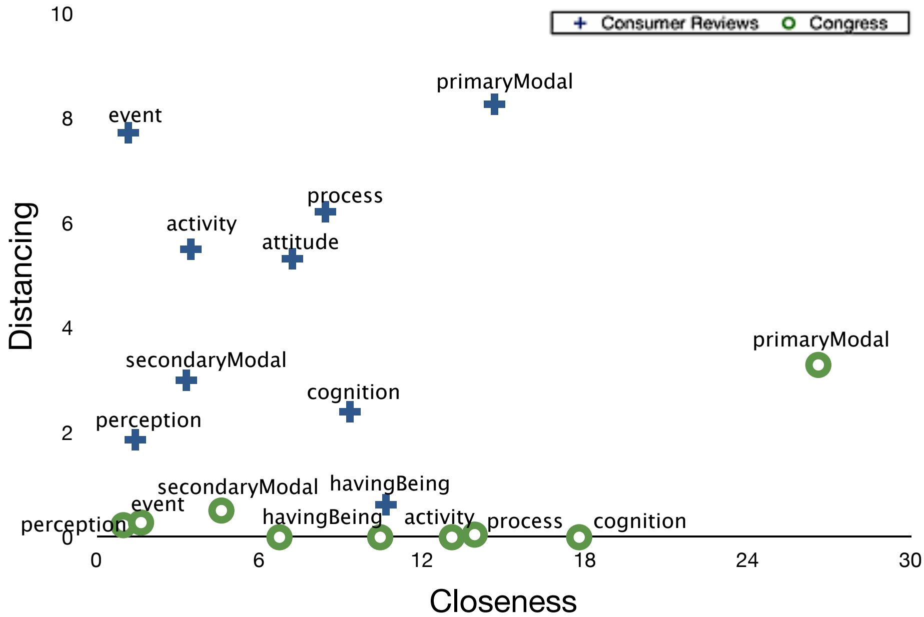

Fig. 2. Distribution of verb categories in Congress debates and Consumer data sets. The horizontal axis estimates *closeness* (in per cent), the vertical axis – *distancing* (in per cent). Crosses denote Consumer reviews categories, circles – those in Congress debates.

resulting two clusters, crosses indicate Consumer review and circles – Congress debates. Each cluster has only one outlier: *havingBeing* – for consumer reviews, *primaryModal* – for Congress debates. The clusters do not overlap, meaning that the category distribution differs across *closeness* and *distancing* dimensions.

The difference can be attributed to the impact of the social environment of the discourse. In consumer reviews the authors try to appeal to the audience, thus, using second person pronoun more often than congressmen who focus on presenting their own opinion during debates. On the other hand, congressmen emphasize their active involvement in the discussed matters more than consumer reviewers. We could partially attribute this to the fact that congressmen should show their involvement to the constituency whereas authors of consumer reviews remain anonymous to readers.

6 Learning Results

We solved four regression problems to learn how the expressed actions estimate the strength of opinions and four classification problems – for learning positive and negative opinions. For the Congress data set, we learned opinion scores and positive and negative opinion labels. For the consumer review data set, the number of positive evaluations in a review (# pos), the number of negative evaluations

Table 3. Smallest *RelativeAbsoluteError* and *RootRelativeSquaredError* obtained by the algorithms. Rows report results for each algorithm. Columns report results for each problem. For each problem, the smallest *RAE* is in *italic*.

Algorithms	Consumer reviews						Congress	
	positive		negative		overall		debates	
	RAE	*RRSE*	*RAE*	*RRSE*	*RAE*	*RRSE*	*RAE*	*RRSE*
kNN	91.19	87.97	90.77	88.70	93.56	96.50	78.74	86.60
SVM	80.98	84.15	89.33	96.71	91.38	94.38	90.89	94.80
BM5P	*80.26*	82.21	*87.21*	85.81	*89.82*	96.61	*73.73*	78.84

in a review (# neg), the overall score define three regression problems; their equal-frequency splits define three classification problems. We used ten-fold cross-validation to estimate the quality of learning. Ten-fold cross-validation is chosen because of its high generalization accuracy and reliability of results.

We ran algorithms available on Weka [16]. Our goal was to tackle regression (*quantitative*) problems. These are new problems for opinion learning. So far, machine learning experiments of opinion detection and prediction concentrated on classification (*qualitative*) tasks. Because of the novelty of this application, we wanted to try different types of learning paradigms. We chose kNN, a prototype-based algorithm, an optimization algorithm, SVM, and M5 Trees, a decision-based one. We applied Bagging (bootstrap aggregating) to assess the influence of training data. Bagging allows an algorithm to train its classifiers on randomly selected training subsets and then choose the best performing classifier. The more a bagged algorithm improves its performance, the more the choice of training data is important. In our experiments, Bagging improved performance of M5 Trees, but not kNN nor SVM. An exhaustive parameter search was applied to every algorithm, separately on every problem. The search is necessary because of performance variance, especially high for Nearest Neighbor and Support Vector Machine.

Table 3 reports smallest relative absolute error *RAE* and corresponding root relative squared error *RRSQ* obtained by the algorithms. The best performance, with the smallest error, was obtained on the Congress data set. Positive consumer opinions were learned better than negative and overall opinions. An interesting phenomenon emerges when comparing algorithm performance – in terms of the learned correlation coefficients. The best performing algorithm in terms of accuracy is Bagged M5 Trees. Since better accuracy implies that the algorithm learns dependencies between opinions and expressed actions better than other algorithms, we conclude that the output decision trees provide a reliable model of the data sets.

For Congressional debates, all output tree models agree that `demand, has, have` are the most important features, followed by `should, would`. We only report here the results of the best performing algorithms. Since this implies that the algorithms model dependencies better than other algorithms, we conclude that the strong language verbs have a positive correlation with attitude toward proposed legislations. On the consumer review data set, bagged trees placed `can, are, find` as the most important features for learning the overall opinions.

Table 4. *Accuracy* (per cent) and corresponding *Recall* (per cent) obtained by SVM. Rows report results for each feature set; B means *binarized*. Columns report results for each problem. For each problem, the largest accuracy is reported in **bold**. Baselines are the majority class accuracy: for the consumer data set – 52.22, for Congress – 59.76.

Features	Consumer reviews						Congress	
	positive		negative		overall		debates	
	Acc	*Recall*	*Acc*	*Recall*	*Acc*	*Recall*	*Acc*	*Recall*
Categories	74.52	74.50	63.64	61.50	66.24	67.30	65.70	67.90
Terminals	76.12	75.80	66.56	67.20	70.06	74.50	69.63	72.00
Terminals-B	76.43	75.70	67.83	73.20	73.60	75.20	70.61	73.40
Collocations	77.75	79.00	68.33	69.50	73.82	78.90	75.18	77.60
Collocations-B	**78.87**	80.10	70.95	71.40	75.21	79.70	**78.14**	81.10

Somehow expectedly, `like` was among most decisive features for learning positive opinions. Learning negative opinions relied on `be, am, would, should` more than on other verbs.

To better display abilities of our approach, we performed a more traditional task of opinion classification (Table 4). We chose Support Vector Machine for solving classification problems. SVM is well-known for a high accuracy of text classification. Also, its use enabled us to directly compare our results with those of [3], obtained on the Congress debate data set.

Their reported test accuracy for positive/negative classification starts from 66.05, obtained on the data set that we used for the current work. To increase accuracy to 76.16, Thomas et al. linked each data entry with previous speeches of the same speaker. Our Congress results have a better accuracy, although we did not use previous records of speakers or other data reinforcements; the results are reported in the right part of Table 4. The obtained results show that the expressed actions do speak loud. Under certain conditions, they reveal more than the previous history of the same speaker. For consumer reviews, learning positive opinions was easier than learning negative and overall opinions.

7 Related Work

Opinion and sentiment analysis that focuses on whether a text, or a term is subjective, bears positive or negative opinion or expresses the strength of opinion has received a vast amount of attention in recent years. Application of learning algorithms - through classification - has been pioneered by Lee et all [1] and successfully used in works of many others. The authors of this much-cited, pioneering work used machine learning algorithms on reviews written by only four professional critics. This means that the algorithms were trained and tested on overly specific, undiversified data. It is not surprising, perhaps, that to achieve a comparable accuracy on the Congress data, the same team had to enhance the data set by using previous speeches of speakers. Our goal, instead, is to work with a large diverse group of data contributors and seek general enough features that reliably represent the resulting diversity of data.

Consumer review data set in our experiments has been used in summarization and feature extraction studies [17]. Some of the listed publications relied on a list of characteristics of reviewed products [14]. Popescu [18] extracted these characteristics from noun phrases and matched them with known product features. We opted for a domain-independent method that does not involve the use of domain's content words.

For automating recognition and evaluation of the expressed opinion, texts are represented through N-grams or patterns and then classified as opinion/non-opinion, positive/negative, etc. [19]. Syntactic and semantic features that express the intensity of terms are used to classify opinion intensity [2]. The listed works do not consider hierarchy of opinion disclosure. We, however, built the pragmatic-lexical hierarchy of the use of semantic categories. The hierarchy allows machine learning models, which are formulated in the lexical terms, to be interpreted in terms of the text pragmatics.

8 Conclusion

This study has shown that, in opinion mining, the difference between structured well-edited and loosely composed texts can be important. To support our claim, we studied the relations between expressed actions and opinions on samples that exhibit different qualities. We built language patterns using modal, event, activity, process, cognition, perception, state verbs and personal pronouns. We extracted and analyzed the resulting patterns and applied machine learning methods to establish quantitative relations between the use of verb categories and opinions.

Defining and solving *regression* problems is a new type of problem for opinion, subjectivity and sentiment analysis. Previous studies stated their problems either as binary classification or multi-class classification problems. Unlike a regression problem, that predicts *quantitative* output, a classification output is *qualitative*. Our combination of quantitative and qualitative learning allows a more detailed opinion analysis.

Our empirical results were obtained on two data sets: consumer-written product reviews [14] and the US Congress debate records [3]. For consumer reviews, the most frequent and diverse action patterns coincided with the passive involvement from the authors. For Congressional records, the most frequent and diverse immediacy patterns coincided with an active involvement from the speakers. Regression problems were successfully learned by Bagged M5 Trees. SVM obtained a reliable accuracy in classification problems.

Learning from verbs is then justified and desirable when the context dictates the avoidance of negative adjectives and adverbs, because empirical results showed that negative adjective and adverbs discriminate better between positive and negative opinions than those with a positive affect. In the future, we intend to analyze the use of different types of verb modifiers (always, never). We are also interested in studying the relations between opinions and the pragmatics of communication, e.g. intensity, immediacy.

Acknowledgments. This work has been supported by *Natural Sciences and Engineering Research Council* of Canada. The authors thank Fabrizio Gotti for technical support of experiments.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proc Empirical Methods of Natural Language Processing EMNLP 2002, pp. 79–86 (2002)
2. Wilson, T., Wiebe, J., Hwa, R.: Recognizing strong and weak opinion clauses. Computational Intelligence 22(2), 73–99 (2006)
3. Thomas, M., Pang, B., Lee, L.: Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 327–335 (2006)
4. Kim, S.M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the of the 20th international conference on Computational Linguistics (COLING 2004), pp. 1367–1373 (2004)
5. Sokolova, M., Lapalme, G.: Performance measures in classification of human communication. In: Orgun, M.A., Thornton, J. (eds.) AI 2007. LNCS (LNAI), vol. 4830, pp. 159–170. Springer, Heidelberg (2007)
6. Halliday, M., Matthiessen, C.: An Introduction to Functional Grammar, 3rd edn. Arnold (2004)
7. Leech, G.: Meaning and the English Verb. Longman (2004)
8. Leech, G., Svartvik, J.: A Communicative Grammar of English, 3rd edn. Longman (2002)
9. Sokolova, M., Szpakowicz, S.: Language patterns in the learning of strategies from negotiation texts. In: Sattar, A., Kang, B.-h. (eds.) AI 2006. LNCS (LNAI), vol. 4304, pp. 288–299. Springer, Heidelberg (2006)
10. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E.: Longman Grammar of Spoken and Written English. Longman (1999)
11. Roget's interactive thesaurus (2006), `http://thesaurus.reference.com/`
12. Perkins, M.: Modal Expressions in English. Ablex Publishing Corporation (1983)
13. Sherblom, J., Rheenen, D.V.: Spoken language indices of uncertainty. Human Communication Research 11, 221–230 (1984)
14. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: Proceedings of Nineteeth National Conference on Artificial Intelligence (AAAI 2004), AAAI Press, Menlo Park (2004)
15. Boulle, M.: Optimal bin number for equal frequency discretizations in supervised learning. Intelligent Data Analysis 9(2), 175–188 (2005)
16. Witten, I., Frank, E.: Data Mining. Morgan Kaufmann, San Francisco (2005)
17. Feiguina, O., Lapalme, G.: Query-based summarization of customer reviews. In: Orgun, M.A., Thornton, J. (eds.) AI 2007. LNCS (LNAI), vol. 4830, pp. 452–463. Springer, Heidelberg (2007)
18. Popescu, A., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of HLTC/EMNLP 2005, pp. 339–346 (2005)
19. Kim, S.M., Hovy, E.: 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), pp. 1056–1064 (2007)