Intelligent Information Systems 2008
ISBN 978-83-60434-44-4, pages 213–217

Automated classification of product review sentiments in Polish

Aleksander Buczyński and Aleksander Wawer

Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland

Abstract

Automated sentiment polarity prediction from text is a challenging problem addressed in this paper. We present results of a comparison of pure "Bag of Words" approach against a mixed method, extended by detecting opinion patterns using shallow-parsing techniques. We utilize two resources for the analysis: Spejd shallow parsing engine and Zetema dictionary of sentiment in Polish. The performance of both approaches has been evaluated on online product review database.

Keywords: sentiment analysis, shallow parsing

1 Introduction

Although the work presented here is a pioneering effort for Polish, the problem of automatically recognizing product review sentiment polarities has been widely recognized in English. Most of recent works can be divided into one of two wider approaches. The first one focuses around operating on the level of individual phrases or expressions and extracting sentiments about specific properties or dimensions of a product. Examples include Jaynce Wiebe's work on expressions of subjectivity on phrase-level (Wilson et al, 2005) and in the context of corpora (Wiebe, 2000). The second approach considers product review as a whole, typically assigning a numerical score (like the number of stars) to certain properties of texts under investigation (Turney, 2002), often with a help of machine learning classifiers (Pang et al, 2002). In this paper we take an intermediate way by confronting numerical, human evaluations, encoded into the number of stars assigned to product review as a whole, with meta-lexical features, namely presence of positive or negative vocabulary and syntactic structures recognized by the shallow parsing engine.

2 Data Overview

The dataset consists of 3830 product or service reviews downloaded from various Polish e-commerce websites and Internet shops. Each review has a corresponding numeric score (number of starts) ranging from 0 to 10. In case of some reviews we had to rescale from a system natively based on 5 or 6 point scale to obtain a common, coherent 10 point metrics.

Reviews were typed in a rather loose manner, sometimes omitting Polish diacrits which does increase ambiguity on top of the "natural" ambiguity of language. A dedicated procedure has been applied to guess the missing diacrits, which improved the detection ratio of identified positive sentiment words by 5% and the number of negative words by 3%.

3 Bag of Words

Bag of Words methods imply a representation which disregards grammar and word order. In our approach, we do not account for presence of a particular lexeme, but presence of a specific category of lexems. Such an abstraction originates in content analysis systems, most notably the classic General Inquirer (Stone et al, 1966), but is not limited to those, as sense generalizations can be obtained using other means like the WordNet library. Lexical categories used in this work include two distinctive sets of words and phrases: 1870 negative and 1580 positive ones, created by Zetema and validated against IPI Corpus.

4 Shallow Parsing of Polish

For detecting multiword opinion patterns we decided to use Spejd — a tool for simultaneous morphosyntactic disambiguation and shallow parsing (Przepiórkowski and Buczyński, 2007). The Spejd formalism is essentially a cascade of regular grammars. Unlike in the case of other shallow parsing formalisms, the rules of the grammar allow for explicit morphosyntactic disambiguation statements, independently or in connection with structure-building statements, which facilitates the task of the shallow parsing of ambiguous and/or erroneous input.

An example of a simple Spejd rule is:

```
Match: [pos~~prep][base~"co|kto"];
Eval:  unify(case,1,2);
       group(PG,1,2);
```

The rule means: **1)** find a sequence of two tokens such that the first token is an unambiguous preposition, and the second token is a form of the lexeme CO 'what' or KTO 'who'; **2)** if there exist interpretations of these two tokens with the same value of case, reject all interpretations of these two tokens which do not agree in case; **3)** if the above unification did not fail, mark thus identified sequence as a syntactic group of type `PG` (prepositional group), whose syntactic head is the first token and whose semantic head is the second token.

5 Sentipejd

Altough Spejd was originally designed for morphosyntactic disambiguation, it is also highly flexible. Therefore we extended the morphosyntactic tagset with a semantic category, expressing properties of positive or negative sentiment. For the purpose of this work, we called this hybrid approach Sentipejd.

The rules for sentiment extraction were created semi-automatically with the help of statistical methods of collocation extraction. First, a list of word bigrams with the highest value of Frequency biased Symmetric Conditional Probability (Buczyński, 2006) was created, to find collocations which are both common in the corpora and strongly dependent. A simple heuristics was used to discard proper names from the results. Then, the collocations were manually generalised into rules and assigned sentiment values.

As a result, we found the following types of simple sentiment structures:

Affirmation — an expression of positive sentiment, usually an adverb confirming the sentiment of a positive word; such expresion is very unlikely to be negated, and should be treated as strong indication of the sentiment.[1]

Negation — as simple as the difference between "polecam" ('I recommend') and "nie polecam" ('I do not recommend').

```
Match: [orth~nie/i] [base~"być"]? [sentiment~"spos"];
Eval:  word(3, neg:sneg, "nie " base);
```

Nullification — expressing lack of a certain quality or property (usually of negative sentiment), for example "nie mam zastrzeżeń" ('I have no objections') or "zero wad" ('zero defects').

```
Match: ([base~"bez|brak|zero|żaden"] | [orth~nie/i] [base~mieć])
       [sentiment~sneg];
Eval:  word(2, spos, );
```

Limitation — a limiting expression tells us that an expression of positive and negative sentiment has only a very limited extend, therefore hinting that the general sentiment of the review is the opposite of the expression. Examples: "jedyny problem" ('the only problem'), "jedyna zaleta" ('the only advantage').

```
Match: [base~"jeden|jedyny"] [sentiment~"sneg"];
Eval:  agree(case number gender, 1, 2); word(2, spos, );
```

Negative modification — an adjective of negative sentiment preceeding a noun of usually positive sentiment, for example "koszmarna jakość" ('nightmarish quality')[2]

```
Match: [sentiment~sneg && pos~adj] [sentiment~spos && pos~subst];
Eval: agree(case number gender, 1, 2); word(2, sneg, base);
```

Review collocations — we tried to avoid product specific collocations, like "wysoki kontrast" ('high contrast') or "duży wyświetlacz" ('large display'). However, there seem to be categories that are common for many different products, for example high/low price, high/low quality, easy/difficult to use. These were included in the scoring as well.

[1]Of course, it is still possible to use such an expression in a sarcastic manner. However, we have not found an evidence of such usage in the review database.

[2]The opposite of the rule, ie. a positive modifier of a negative subject, did not work so well.

6 Results

Human-assigned numerical scores were recoded into a triple-valued dependent variable (negative|neutral|positive). Then, the recoded scores along with outputs from Bag of Words and Sentipejd were submitted into two different classification models, namely C5.0, the commercial successor of C4.5 (Quinlan, 1993), and Logistic Regression. We did not follow the multistep method of splitting the data into learning and training partitions - instead, we focused on separability. Results of the experiment are presented on Table 1.

Table 1: Accuracy of selected classifiers on Bag of Words and Sentipejd

Method	C5.0	Logistic Regression
Bag of Words	74.8%	71.2%
Bag of Words (with diacrit guesser)	75.1%	71.2%
Sentipejd (with diacrit guesser)	78.2%	76.5%

Of the two used classifiers, C5.0 proved superior than the Logistic Regression. Introducing Sentipejd on diacrit guessed input raised the C5.0 classification accuracy by 3%, compared to the baseline of non diacrit guessed Bag of Words. To grasp the effect of Sentipejd we decided to repeat the classification experiment with C5.0, dividing the reviews into three groups of various word length, as in Table 2.

Table 2: Accuracy of c5.0 classifier on reviews of various lengths

Method	Length of review in words		
	1—5	6—20	21+
Bag of Words	81.9%	73.6%	74.7%
Bag of Words (with diacrit guesser)	81.6%	74.0%	74.4%
Sentipejd (with diacrit guesser)	82.2%	77.4%	76.4%

The impact of Sentipejd is the most noticeable when applied to the medium (3% improvement) and long (2% improvement) reviews. In the case of more wordy reviews, all tested methods performed worse. This can be explained by the fact that adding more words often introduces more ambiguity, as positive aspects of a product or service under review are discussed along with negative ones. The final decision, reflected in a numerical score, is the result of an interplay between (or weighting) various pros and cons.

To a certain extent, the approach proposed in this article follows the ideas of contextual valence shifters proposed for English in (Polanyi and Zaenen, 2006), where the authors consider a number attitudinal valence modifying structures, arguing that focusing on lexeme level attitudes is incomplete and in some cases renders wrong results. The work presented in this paper fully supports such a conclusion.

7 Conclusions and Future Work

While the results prove that both shallow parsing methods and diacrit guessing are steps toward high accuracy sentiment detection, further work is required to investigate false matches and improve accuracy. Nevertheless, we find shallow parsing methods a promising tool for sentiment analysis, in more ways than one. First, they allow to detect sentiment modifiers and valence shifters. Furthermore, by eliminating interpretations that do not fit into syntactic structures, shallow parsing can also contribute by disambiguating and correcting errors in the input (Buczyński and Wawer, 2008).

References

Buczyński A. (2006) Wybrane zastosowania programu Kolokacje do badań lingwistycznych. W: A. Duszak, E. Gajel, U. Okulska, Korpusy w angielsko-polskim językoznawstwie kontrastywnym. Teoria i Praktyka. Kraków 2006, pp. 427-448.

Buczyński A., Wawer A. (2008) Shallow parsing in sentiment analysis. *Proceedings of the workshop on Partial Parsing – Between Chunking and Deep Parsing.* Forthcoming.

Pang B. L., Lee S., Vaithyanathan (2002) Thumbs up? Sentiment classification using Machine Learning Techniques. *Proceedings of the EMNLP conference.*

Polanyi L., Zaenen A. (2006) Contextual Valence Shifters. In: Shanahan J. G., Qu Y., Wiebe J. (eds), Computing Attitude and Affect in Text: Theory and Applications. *Springer.*

Przepiórkowski A., Buczyński A. (2007) Shallow Parsing and Disambiguation Engine. *Proceedings of the 3rd Language and Technology Conference, Poznań, Poland.*

Quinlan, J. R. (1993) C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers.*

Stone, P. J., Dunphy D. C., Smith M. S., Ogilvie D. M. (1966) The General Inquirer: A Computer Approach to Content Analysis. *MIT Press.*

Turney, P. (2002) Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the ACL*, pp. 417-424.

Wiebe, J. M. (2000) Learning subjective adjectives from corpora. *Proceedings of the 17th AAAI conference.*

Wilson T., Wiebe J., Hoffman P. (2005) Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing.*