Shallow parsing in sentiment analysis of product reviews

Aleksander Buczyński, Aleksander Wawer

Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
Aleksander.Buczynski@uw.edu.pl, axw@ipipan.waw.pl

Abstract

The article discusses a practical application of shallow parsing to sentiment polarity analysis of product reviews in Polish. Examples on how partial parsing can help the task on different levels are presented, ranging from disambiguation between mophosyntactic interpretations with different sentiment polarity, through detection of structures expressing negation or lack of a certain sentimentally polarised property, to capturing idioms. All the stages are expressed and implemented in the same, coherent shallow parsing formalism.

1. Introduction

The article presents an attempt to apply shallow parsing to improve the accuracy of automatic recognition of product review sentiment polarity (Turney, 2002) in Polish. Examples of application on various levels are presented, ranging from disambiguation, through detection of negation, to capturing idioms. All the stages are expressed and implemented in the same, coherent shallow parsing formalism.
Section 2. contains the data overview, including the preprocessing technique. Section 3. briefly introduces Spejd, the shallow parsing formalism and engine used for experiments. Section 4. presents the connection between marking consistency grammar structures and disambiguating sentiment polarity. Section 5. shows how to improve the results by identifying constructions describing negation or lack of a certain sentimentally polarised property. Finally, section 6. describes an attempt to cover idiomatic expressions in the same formalism.

2. Data Overview

The evaluation dataset consists of 4175 product or service reviews downloaded from various Polish e-commerce websites and Internet shops. Reviewed products included books, games, printers, monitors, cameras, phones, cosmetics, tools, holidays.
Each review has a corresponding numeric score (number of stars), assigned by the review's author. Most of the websites have scores ranging from 0 (worst) to 10 (best), but some are based on 5 or 6 point scale. For evaluation purposes, in order to obtain a common, coherent metrics, we decided to rescale all scores into three categories: negative (-1), neutral (0), positive (+1).
As it often happens with data collected from the Internet, reviews were typed in a rather loose manner, sometimes omitting Polish diacrits which futher increases ambiguity of the input 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%.
As a baseline, we took the bag of words approach, disregarding grammar and word order. We do not account for presence of a particular lexeme, but rather 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). Lexical categories used in this work include two sets of lexemes (dictionaries): 1580 positive and 1870 negative ones. A string is considered to have a positive/negative sentiment if at least one of its morphosyntactic interpretations belongs to a positive or negative dictionary, respectively.[1] After the addition of missing diacrits, the application of the lexicons resulted in sentiment tags for 19370 words in the reviews (13768 positive and 5602 negative).
Baseline accuracy was calculated by running a C5.0 classifier (a commercial successor of C4.5 (Quinlan, 1993)) on the tagged reviews, taking as input variables the number of positive and negative tags in a review. Such a classifier predicted the sentiment of the reviews with accuracy of 74,9%.

3. Shallow Parsing of Polish

For detecting syntactic structures we decided to use Spejd — a tool for simultaneous morphosyntactic disambiguation and shallow parsing (Przepiórkowski, 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 the identified sequence as a syntactic group of type PG (prepositional group), whose syntactic

[1] Therefore it is theoretically possible for the same string to have both positive and negative sentiment.

Table 1: Examples of pairs of words in Polish with different sentiment polarity, which have common forms. Tags :spos and :sneg denote `sentiment` value. Lack of tag means neutral sentiment.

obraz (image)	obraza (insult:sneg)
ok (ok:spos)	oko (eye)
płytka (tile, CD/DVD)	płytki (shallow:sneg)
lub (or)	lubić (to like:spos)
kupa (poo:sneg)	kupić (to buy)
wina (guilt:sneg)	wino (wine)

head is the first token and whose semantic head is the second token.

Although Spejd was originally designed for morphosyntactic disambiguation, it is also highly flexible. Therefore we extended the morphosyntactic tagset with a semantic category (`sentiment`), expressing properties of positive or negative sentiment (`spos` and `sneg` respectively). We called this hybrid approach Sentipejd.

4. Sentiment disambiguation

Since both morphosyntactic tagging and partial constituency parsing involve similar linguistic knowledge, shallow parsing can be a powerful tool for simultaneous morphosyntactic disambiguation, as discussed in (Przepiórkowski, 2007). But different morphosyntactic interpretations often imply also different semantic interpretations, including sentiment polarity (especially when it comes to disambiguating between different base forms). Therefore, a tool for disambiguating between various morphosyntactic interpretations can also help to disambiguate the sentiment polarity of an interpreted unit.

For example, strings like obraz and obrazy can be forms of the word obraza (insult), which has a definitely negative sentiment polarity, as well as obraz (image, painting), with no sentiment connotations at all. Table 1 shows a few more examples of such ambiguities in Polish.

For testing the application of shallow parsing to sentiment disambiguation, we used a preliminary shallow grammar of Polish, developed at the Polish Academy of Sciences, Institute of Computer Science. The grammar is written in the Spejd formalism, allowing to encode structure building and disambiguation in the same rules. It contains 58 rules for syntactic group identification. Among these, the rules identifying noun groups turned out to be particularly useful for sentiment analysis, because of their case unification or strict requirements.

Let us examine the sentence:

Najlepszy obraz uzyskamy, podłączając go do cyfrowego wyjścia karty graficznej.[2]

(The best image we achieve, connecting it to a digital output of graphic card.)

Najlepszy is definitely a superlative form of singular adjective 'good', but can be assigned four possible combinations of case and gender:

- NOM:M1
- NOM:M2
- NOM:M3
- ACC:M3

The string obraz can be:

- either a form of the word OBRAZ (image) in nominative or accusative case,
- or a genitive of the word OBRAZA (insult).

After the grammar identifies "Najlepszy obraz" as a noun group, it enforces case, number and gender unification between the words constituting the group. The result is:

Najlepszy	obraz
~~good:adj:sg:nom:m1:sup:spos~~	image:subst:sg:nom:m3
~~good:adj:sg:nom:m2:sup:spos~~	image:subst:sg:acc:m3
good:adj:sg:nom:m3:sup:spos	~~insult:subst:pl:gen:f:**sneg**~~
good:adj:sg:acc:m3:sup:spos	

Although the interpretations are still somewhat ambiguous (the parser has not yet decided whether the phrase is nominative or accusative), for sentiment analysis it is important that the invalid interpretation of "obraz" as a genitive form of 'insults' has been discarded, therefore removing the only interpretation with negative sentiment polarity, which could lead to a wrong conclusion about the sentiment of the whole review.

Let us consider another example:

Trochę zajmuje mu odczyt płytki.

(Some [time] takes him reading the CD.)

The word "płytki" has eight interpretations, four with the base form PŁYTKA:SUBST:

- SG:GEN:F
- PL:NOM:F
- PL:ACC:F
- PL:VOC:F

and four with base form PŁYTKI:ADJ:

- SG:NOM:M1
- SG:NOM:M2
- SG:NOM:M3
- SG:ACC:M3

Spejd has identified the pronoun "mu" as a NG, and following that — part of the sentence "odczyt płytki" as a NG with a genitive postmodifier. The rule identifying the latter group has decided to discard all non-genitive interpretations for "płytki", correctly leaving only PŁYTKA:SUBST:SG:GEN:F.

On the test data the aforementioned shallow parsing rules allowed to assign correct, unambiguous sentiment tags to 144 semantically ambiguous segments, generating no false positives or negatives. Although comparing to the total number of sentiment tags it may not seem much, one has to take into consideration that the grammar used is indeed very shallow and still in developement.

[2] Diacrits were missing in the original input and have been added in preprocessing.

5. Sentiment Phrases

5.1. Rules extraction

Our 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 — if all occurences of both words forming a collocation started with a capital letter, the pair of words was considered a proper name. Then, the remaining collocations were manually generalised into two kinds of rules — sentiment reversibility and feature extraction.

5.2. Sentiment reversibility

We paid special attention to structures expressing reversion or cancellation of sentiment polarity. Although the work presented here is a pioneering effort for Polish, the problem of recognising phrase level sentiment polarity reversal has been addressed in English (Whitelaw et al, 2005). Several of the rules presented below could be implemented using a window based approach, but the precision of such techniques can be problematic in inflected languages.
For our experiments we used the following types of sentiment modifying structures:

Negation — reversing the polarity as from "polecam" ('I recommend') to "nie polecam" ('I don't recommend'). The example generic rule captures also statements including the optional verb 'to be' (`[base~być]?`), like "nie jest dobry" ('isn't good'):

```
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'). An example of a nullification rule[3] is:

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

The second, optional specification in match (`[base~żaden]?`) serves capturing typical double negative constituents, expressing a single negation. Negative concord is quite common in Polish (Przepiórkowski, 1997), and also in the product reviews, for example "nie miałem żadnego problem" ('I didn't have no problem')[4]

Limitation — a limiting expression tells us that an expression of positive or 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|1"]
       [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'), "nieprzyjemne doświadczenie" ('unpleasant experience')[5]

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

5.3. Feature extraction

Among the captured collocations, many were product specific, like "wysoki kontrast" ('high contrast') or "duży wyświetlacz" ('large display'). We have chosen to ignore these, to make the rules product-independent. However, there seem to be sentiment polarised features that are common for many different products. The following features were included in the rules:

- high/low price,
- high/low quality,
- easy/difficult to use.

5.4. Captured Structures

The set of rules described above captured 1774 structures. Table 2 presents the most common of them. Although the structures provide only less than 10% of all sentiment tags, they often change the polarity of tags, therefore having a significant impact on the results. As shown in Table 3, the structures increased the classification accuracy by up to 2,5% comparing to the baseline bag of words approach, using the same c5.0 classifier.

[3] The rule is very generic and does not force any case requirements or unification. It performed well on the reviews data, but for other aplications it may be more suitable to split the rule into a few more sophisticated ones.

[4] In the early version of the system, double negation used to cancel itself, therefore not giving any improvement over the bag of words approach.

[5] It is worth noting that the structures captured by the opposite of the rule, ie. a positive modifier of a negative subject, are very hard to assign an unambiguous sentiment polarity.

Table 2: Most commonly applied Sentipejd rules.

Value	Count	%
negation of positives	493	27,8%
negation of negatives	341	19,2%
nullification of negatives	320	18,0%
feature: ease of use	147	8,3%
nullification of positives	146	8,2%
limitation of negatives	119	6,7%

6. Idioms

This section deals with multiword entries, including phraseological units and idioms, with non compositional sentiment value. The calculation of sentiment present in such structures in then a matter of accurate and efficient recognition, which can be reduced to a more general question - of how to encode and recognize multiword entries in Polish, an inflected language.

In principle, encoding of multiword expressions for natural language processing falls in two general groups (Moszczyński, 2006): encode them in an existing formal grammar, such as Debusmann (Debusmann, 2004) or use a specialized formalism such as IDAREX (Segond, 1995) or Phrase Manager (Pedrazzini, 1994). A formal grammar approach makes the lexicon of multiword sentiment expressions heavily dependant on a particular grammar, which might make its reusability questionable. The expressive power of such grammars might be largely unused in the context of sentiment analysis. The other method, specialized formalisms, seems more promising, but the overview of existing approaches proves that none of them meets the requirements of Polish. IDAREX, which is a regular grammars based formalism, does not allow for handling expressions that have a very variable word order and allow many modifications. Expressing sentiment-bearing idioms in IDAREX has to include all the possible variations which leads to a description that suffers from overgeneration. Moreover, IDAREX does not support unification. This fact alone renders it unsuitable for any reliable recognition, as Polish requires to enforce agreement between constituents of a phraseological expression. Phrase Manager is not suitable for Polish multiword structures as it enforces membership of such a structure to predefined syntactic classes, which in turn leads to an unnecessary overhead and classes proliferation.

We found out that the sentiment carrying idioms can be conveniently described in the same Spejd formalism as the polarity reversing structures. Examples of commonly used non-compositional sentiment phrases used in dialogs involve popular euphemistic expressions such as 'have somebody somewhere':

```
Match: [base~mieć]
       [base~"to|ty|ten" && case~acc]
       [pos~~adv]?
       [base~gdzieś];
Eval:  leave(case~acc, 2);
       set(qub:sneg, , 4);
```

where somewhere is an euphemism for a more abusive "arse", the negative meaning being only recognizable in the context of the whole expression.

Sentiment carrying non-compositional expressions are very infrequent in the reviews, nevertheless a careful examination revealed several such multiword structures. The two most common examples are presented below:

- 'Almost makes a great difference' (meaning: to fail to meet some requirements):

```
Match:
  [base~prawie]
  [base~robić & person~ter]
  [base~"duży|spory|wielki"]
  [base~różnica && case~acc];
Eval:
  unify(case number gender,3,4);
  leave(case~acc, 4);
  word(qub:sneg,
    "prawie robi dużą różnicę");
```

- 'Nothing to add, nothing to lessen' (meaning: perfectly, accurately):

```
Match: [base~nic]
       [base~dodać && pos~inf]
       ([pos~interp] ns?)?
       [base~nic]
       [base~ująć && pos~inf];
Eval: word(qub:spos,
        "nic dodać nic ująć");
```

The recognition of sentiment carrying multiword structures or idioms might not introduce substantial improvements on product review sentiment recognition accuracy, as the reviews language is very simple, and idioms are very rare (few occurences per idiom in the sample). However, it seems that the recognition of sentiment-bearing idiomatic expressions can contribute to sentiment analysis in other language domains, like informal dialogs or literary language. Once properly encoded, the same set of rules for idiom recognition can be used across multiple domains. However, introducing new domains may require extending the idiom set.

7. Results

For evaluation purposes, we grouped the shallow parsing rules into four disjoint sets: disambiguation (described in 4.), sentiment reversibility (5.2.), feature extraction (5.3.) and idioms (6.). Table 3 displays impact of each set on the classifier accuracy, including 'bag of words' (empty ruleset, baseline approach) and all rules (all sets combined).

On the test data, disambiguation actually lowered the accuracy of the classification. An examination of the results revealed that the disambiguation rules are sometimes too strict, not taking into account errors often made by the reviewers. The most significant improvement was achieved by detecting sentiment reversing structures.

Table 3: Accuracy of c5.0 classifier on reviews, depending on the shallow parsing ruleset used.

Method	Accuracy
Bag of words	74,49%
Disambiguation	74,47%
Reversibility	77,01%
Feature extraction	74,56%
Idioms	74,49%
All rules	77,05%

8. Conclusions and Future Work

In the paper we presented an approach to improve automatic sentiment polarity extraction from noisy and ambiguous product reviews in Polish by shallow parsing techniques. We demonstrated that shallow parsing can affect the accuracy of sentiment classification compared to a baseline bag of words approach. The most significant improvement has been achieved by detecting negation-like structures which reverse sentiment polarity. On less noisy data morphosyntactic disambiguation of the phrases can also help by removing certain sentiment ambiguities. Finally, although idioms with clear sentiment polarity were rare in the data set under consideration, they can be described and recognised in the same formalism. Further reasearch is needed to investigate the usability of the formalism in idioms recognition on different types of corpora.

The work so far focused on considering product review as a whole, assigning general sentiment polarity to a product. It remains an open question how shallow parsing can contribute to extracting attitudes towards specific properties or dimensions of a product.

9. References

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

R. Debusmann. 2004. Multiword expressions as dependency subgraphs. In Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing, Barcelona, Spain.

R. Moszczyński. 2006. Formalisms for encoding Polish multiword expressions. Report 994 of IPI PAN (Institute of Computer Science, Polish Academy of Sciences).

S. Pedrazzini. 1994. Phrase Manager: A System for Phrasal and Idiomatic Dictionaries. Georg Olms Verlag, Hildeseim, Zurich, New York.

A. Przepiórkowski, A. Kupść. 1997. Negative Concord in Polish. Research Report 828 of IPI PAN. Institute of Computer Science, Polish Academy of Sciences.

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

J. R. Quinlan. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

F. Segond, E. Breidt. 1995. IDAREX: Formal description of German and French multi-word expressions with finite state technology. Technical Report MLTT-022, Rank Xerox Research Centre, Grenoble.

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

P. Turney. 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.

C. Whitelaw, N. Garg, S. Argamon. 2005. Using Appraisal Groups for Sentiment Analysis. Proceedings of the 14th ACM international conference on Information and knowledge management. Bremen, Germany.