Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments

Caroline Brun, Caroline Hagège

Xerox Research Centre Europe
6, chemin de Maupertuis, 38240, Meylan, France
Caroline.Brun@xrce.xerox.com, Caroline.Hagege@xrce.xerox.com

Abstract. In the context of the development of a feature-based opinion mining system for English, we observed that there is some very interesting information provided by customers but not yet covered by "standard" opinion mining techniques: opinion mining aims at detecting whether comments are positive or negative, but it appears that customers are very often suggesting improvements about what they are reviewing, which is quite different from expressing an opinion. This papers proposes to apply Natural Language Processing techniques in order to address this rather new task of extracting automatically such kind of suggestions for improvement from user's comments.

1 Introduction

With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise for information technologies to seek out and understand the opinions of others. There is a growing interest and activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in texts (see [1]). However, when looking at customer reviews in general, which are one of the core target of opinion mining applications, it appears that there is a wide part of very interesting information provided by customers but not yet covered by standard opinion mining techniques: opinion mining aims at detecting whether comments are positive or negative, but it appears that reviewers are very often suggesting improvements about what they are reviewing, which is quite different from expressing an opinion. At the same time, when someone in a review suggests some improvements for a given product, it means that this person is not fully satisfied with the current version of the product. This gives thus important information regarding the degree of satisfaction of a customer. Suggestions for improvement of a product are done using two discursive figures, either by wishing the presence of a missing feature or component, or by regretting the absence of such a feature or component. This goes beyond the scope of traditional sentiment analysis; however, it complements it and provides very valuable information that is worth collecting and analyzing in the context of business analytics. To illustrate this problem, the following

Caroline Brun, Caroline Hagège

sentences are some instances of suggestions for improvement extracted from a corpus of user's comments about printers:

- "HP should have made the bin deep enough to hold an entire ream (500 sheets)."
- "Why HP didn't opt to throw in a 500-sheet tray with at least the standard 7 sizes (as per the 250-sheet tray) is beyond me."
- "This C4280 has filled the bill in those areas, however, for the price, HP should have considered throwing in more features and lowering the print costs."
- "I think they should have put a faster scanner on the machine, one at least as fast as the printer."
- "My only regret was that this unit could not fax!"
- "If you're considering any HP OfficeJet, I think most people will be happy with the R-40, which is only missing the automatic paper feed."

In this paper, we propose to use Natural Language Processing techniques in order to address the task of extracting automatically such kind of suggestions for improvement in comments. To perform this task, we have used a corpus of customer reviews about printers from different manufacturers.

2   Context of this Work

A fundamental task in opinion mining is classifying the polarity of a given text, sentence or feature/aspect level to find out whether it is positive, negative or neutral. Different methodologies using NLP and machine learning techniques are used for this purpose. Research has also shown that subjectivity or objectivity identification can also help achieving the purpose. However the most fine grained analysis model is the feature based sentiment mining method. Feature based opinion mining aims at determining the sentiments or opinions that are expressed on different features or aspects of entities (e.g. [2],[3], [4]). When a text is classified at document level or sentence level, it might not tell what the opinion holder likes or dislikes. If a document is globally positive about an object, it does not mean that the opinion holder holds positive opinions about all the aspects or features of this object. Similarly, if a document is globally negative, it does not imply that the opinion holder dislikes everything about the object described.

The context of this paper is the development of a feature-based opinion mining system, for English.

Formally, our system adopts the representation of a given opinion as proposed by [1], where an opinion is a five place predicate of the form $(o_j, f_{jk}, s_{ijkl}, h_i, t_l)$, where:

- o_j is the target object of the opinion (the main concept)
- f_{jk} is a feature associated to the object
- s_{ijkl} is the value (positive or negative) of the opinion expressed by the opinion holder about the feature

- h_i is the opinion holder
- t_l is the time when the opinion is expressed.

The opinion detection system we have built relies on a robust deep syntactic parser, (c.f. [5]), as a fundamental component, from which semantic relations of opinion are calculated. Once syntactic relations are extracted by a general dependency grammar, we use the robust parser by combining lexical information about word polarities, sub-categorization information and syntactic dependencies to extract these semantic relations. The polarity lexicon has been built using existing resources and also by applying classification techniques over large corpora, while the semantic extraction rules are handcrafted, see [6] and [7] for the complete description of these different components. There exist other systems, such as the one described in [3], that use syntactic dependencies to link the source and target of the opinions. Our system belongs to this family, since we believe that the syntactic processing of complex phenomena (negation, comparison and anaphora) is a necessary step in order to perform feature-based opinion mining. Another characteristic of our system is that it respects a two-level architecture; it relies on a generic level, applicable to all domains and corpora, and on a domain-dependent level, adapted for each sub-domain of application.

Additionally, our system includes a semantic mapping between polar vocabulary and the features it corresponds to. For instance, the opinion word "fast" is mapped to the feature "speed", the word "expensive" to the feature "price", the word "clunk" to "noise" and so on. This mapping enables us to further exploit the comments of the user by referring to specific product characteristics.

Some examples follow that show what the system would output considering certain input sentences extracted from customer reviews about printers:

1. *Input*: "I like this printer, but I think it is too expensive."
   Output: OPINION(printer, _, positive, I, _),
   OPINION(printer, price,negative, I, _).

   In this example, the system identifies that the user expresses a negative opinion on the feature "price" of the printer although the rest of its characteristics satisfies him.

2. *Input*: "The problem of this printer is the fuser."
   Output:
   OPINION(printer, fuser, negative, _, _).

   Here, the system identifies that the user expresses a negative opinion about the feature "fuser" of the printer.

In this paper, we focus on the integration of a new component to this pre-existing system, which is a component aiming at detecting suggestions for improvement within customer reviews and that goes beyond the scope of traditional opinion detection.

3 Related Work

While there is a lot of literature regarding opinion mining and more precisely feature-based opinion-mining, we didn't find many works related to the automatic extraction of suggestions within comments, except the following paper by [8] about extracting wishes, which has some connections with the approach described here. The cited paper presents a work whose goal is related to our goal: it aims at detecting wishes in documents using NLP techniques combined with Machine Learning techniques. Their tool relies on very simple pattern matching (about 13 patterns like "I wish _", "if only _", "would like _ if", ...) combined with pattern-based sentence classification (using Support Vector Machines). They train their system using a corpus of New Year's wishes made by people around the world (see http://www.timessquarenyc.org/events/new-years-eve/wishing-wall/index.aspx). The goal they address is wider than the one proposed here on one hand, and more restricted on the other. As they extract all kind of wishes, the scope is wider than ours, since suggestions for improvement can be viewed as subset of wishes: for example, from our customer printer's reviews, their method would extract "I would like to share my experience", "what I would like to stress here in this review", (these expressions are wishes indeed) while our method would filter out these sentences since they are not suggestions for improvement (and not very useful in a business analytics context). Moreover, as they are using a corpus containing exclusively (very general) new year's wishes, their system is not covering all the other way people are expressing suggestions, i.e. by regretting the absence of a feature or a component: our method would extract "My only regret was that this unit could not fax!", "The lack of any PC fax status software is incredible", which are implicit suggestions for improvement, while their proposal is not covering these cases.

4 Detecting Suggestions

As said in introduction, suggestions appearing in user comments are expressed using "wishes" or "regret". Therefore, our suggestion detection component relies on syntactico-semantic patterns that capture these types of expressions. The global architecture is showed on figure 1.

The system relies on four main components:

- a structured terminology of the target topic;
- a thesaurus of vocabulary related to expressions of suggestions;
- a fine grained linguistic parser;
- an extractor for suggestions for improvement (SfI).

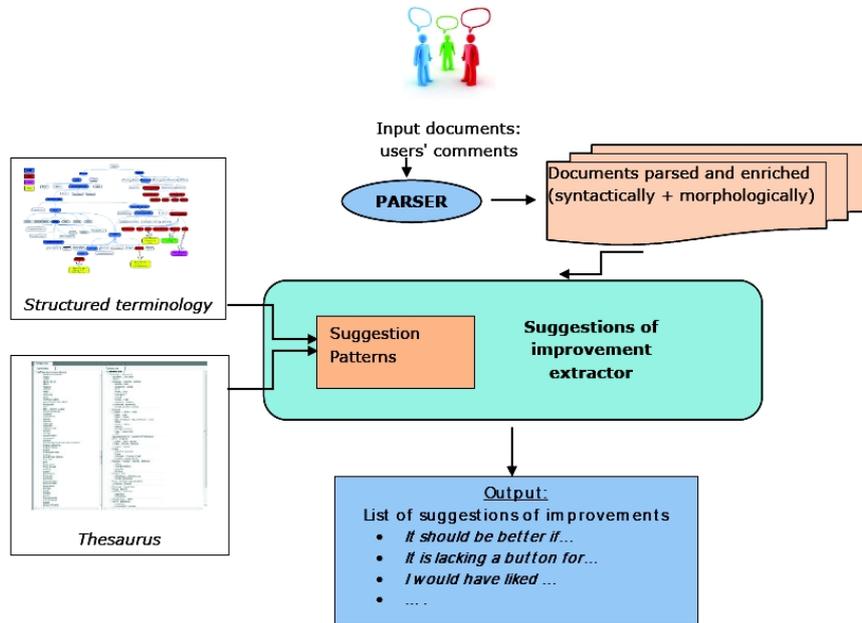These different components are detailed in the following sections.

Fig. 1. Architecture of the system

4.1 Structured Terminology

The structured terminology consists in a description of the topic in focus. Taking for example a manufactured product like a printer, the terminology encodes concepts and associated terms such as the "part-of" concepts of the product (for instance *paper tray*), the manufacturer of the product (for instance *Xerox*), physical characteristics (for instance *weight*, *aspect*) and commercial/technical characteristics (for instance *speed*, *price*, *noise*). The terms found in the terminology are potential targets on which suggestions for improvement may apply. These terminologies can be part of already existing ones if available (e.g. in an existing opinion mining system) or can be acquired from processing texts of the domain as very simple ontological relations are needed. In our case, the structured terminology was already part of our opinion detection system, since this information is also needed to perform feature-based opinion mining. To extract this domain-dependent terminology, we have used a method partly similar to what is proposed in [9], who seek meronymy relationships (part-whole). In our case, the main concepts discussed in reviews about printers are denoted by the following vocabulary : *printer*, *copier*, *scanner*, *machine*, and *product*. We have used our parser to extract, from the corpus, the most frequent nouns modifying these main concepts, i.e. matching the MODIFIER syntactic relations and calculate frequencies of these relations.

4.2 Thesaurus of Expressions of Suggestion

The thesaurus of expression of suggestions is encodes expressions used in sentences denoting wishes, regrets and misses. Choosing as seeds the words *wish*, *regret*, *miss* and *lack* we explored existing thesauri and gather nouns and verbs that will potentially trigger expressions of suggestion. Existing thesauri can be found for instance in the following links : http://thesaurus.com/, http://www.macmillandictionary.com/, etc.. We finally consider two different semantic classes for verbs and predicative nouns. The class *absence* which contains verbs such as *lack*, or noun like *absence*, the class *belief* which contains verbs like *expect* or *believe*. We gather a total of 16 lexical entries of classes *absence* and *belief* and also consider isolated lexical entries such as *wish* considered as verb or noun, the verb *like* and the verb *prefer*.

4.3 Linguistic Processor

As said in section 2, the linguistic processor we used is a robust syntactic parser, that performs a fine grained linguistic analysis of input text. This analysis contains the morphological features of the words (in particular information about modality and aspect) but it also gives us dependency links between the linguistic units in the text. These dependencies are labeled with the main syntactic functions such as *SUBJECT*, *OBJECT*, *MODIFIERS* (see [5] for more details). Automatic detection of named entities is also handled by this system, and covers the extraction of standard named entity type such has PERSON, COMPANY, PRODUCT NAMES, PLACES, DATES ...

Input texts (i.e users reviews) are processed by the linguistic processor which will calculate the syntactic relations holding between the words and will give also information about verbal tenses, modality and verbal aspect. For instance, the following sentence *I would have liked a better resolution for my pictures*, the linguistic analysis provides the following information.

```
MOD_PRE(resolution,better)
SUBJECT(like,I)
VMODAL(like)
VASPECT_PERFECTIVE(like)
VTENSE_PAST(like)
OBJECT(like,resolution)
PRINTER_FONCTION(resolution)
```

The system has detected that the verb *like* in this sentence is used together with a modal form and is employed in a perfective aspect. It also has detected that the object complement of *like* is the word *resolution* which is a printer function. Finally, the system also detects that the subject of *like* is the pronoun *I*.

4.4 Extractor of Suggestions for Improvement

The core of the extractor of suggestions for improvement consists in a set of general syntactico-semantic patterns that will then be matched against exist-

ing analyzed reviews. These patterns express general formulas of suggestion in which arguments are elements of the domain terminology. These patterns also use suggestion expressions that have been extracted from the thesaurus. Finally, linguistic information such as dependency relations, verbal modality and aspect is also enclosed in the patterns. For instance, in the context of printers' reviews, the following syntactico-semantic pattern uses both linguistic and terminological information and expresses a suggestion for improvement.

$$
\begin{bmatrix}
\text{SUBJECT} & \begin{bmatrix} \text{SEMANTIC-TYPE} = \text{MANUFACTURER} \end{bmatrix} \\
\\
\text{PREDICATE} & \begin{bmatrix} \text{TENSE} = & \text{PAST} \\ \text{ASPECT} = \text{PERFECTIVE} \\ \text{M}odal = & \text{YES} \end{bmatrix}
\end{bmatrix}
$$

This pattern can be rephrased as "a manufacturer entity, which is subject of a modal verb used in the past tense and perfective aspect". This pattern, which corresponds semantically to a wish, extracts the pertinent sentence *HP should have made the bin deep enough*, but does not extract the non-pertinent sentence *HP should be ashamed*, as *ashamed* is not a verb in the perfective form.

The following pattern can be rephrased as "a product-function term, which is subject of a verb of the semantic class absence".
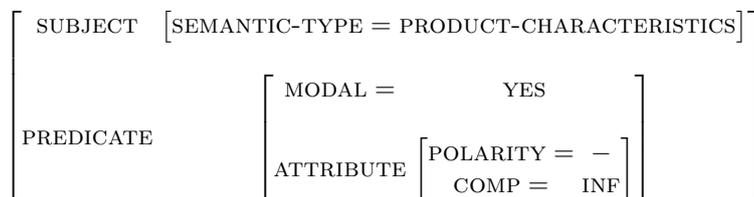
$$
\begin{bmatrix}
\text{SUBJECT} & \begin{bmatrix} \text{SEMANTIC-TYPE} = \text{PRODUCT-FUNCTION} \end{bmatrix} \\
\\
\text{PREDICATE} & \begin{bmatrix} \text{SEMANTIC-CLASS} = \text{ABSENCE} \end{bmatrix}
\end{bmatrix}
$$

This pattern does not correspond to a wish as in the previous example. It just states that something is missing in the described product, which implicitly means that the user would have liked to have this component. It extracts (among others) the pertinent sentence *But the ability to kill jobs is really lacking, often requiring hard reset when needed.*

Such patterns are applied after the analysis of the customer reviews by our robust parser. When a given pattern matches a sentence analysis result, the sentence is considered as relevant. The output of the system consists in a list of sentences expressing either customer wishes regarding the products or absence of components or functionalities. This result can then be manually or semi-automatically processed in order to detect what are the main claims of the customers without having to send them any questionnaire.

We defined 21 patterns for our SfI. It is worth mentioning that some of the patterns also make use of polarity lexicon defined for the opinion-mining application. For instance the following pattern states that if a negative adjective which is used with a comparative of inferiority is also an attribute of a product characteristics (terminology information), and if this product characteristics is the subject of a modal verb, then we extract the sentence. This pattern will

cover examples as *fax could be less noisy.*

$$
\begin{bmatrix}
\text{SUBJECT} & \begin{bmatrix} \text{SEMANTIC-TYPE} = \text{PRODUCT-CHARACTERISTICS} \end{bmatrix} \\
\\
\text{PREDICATE} & \begin{bmatrix} \text{MODAL} = & \text{YES} \\ \\ \text{ATTRIBUTE} & \begin{bmatrix} \text{POLARITY} = & - \\ \text{COMP} = & \text{INF} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

Finally, it is also important to note that suggestions for improvement are not necessarily correlated with positive/negative opinions. Suggestions can be expressed in texts where the general opinion is either positive or negative. Even at the sentence level, suggestions can appear in a positive sentence like *But the lack of this feature isn't a complaint, it's a request for the next model*, even if they appear more often in neutral or negative comments as in *Unfortunately many of the manuals are only available on-line as I would have liked a printed main manual*. At the same time, suggestions for improvement can be considered as slightly negative (if someone suggests an improvement for a product, it means that he/she is not completely satisfied with the product). As a result, integrating the suggestions for improvement module in a more general opinion-mining tool reflects more precisely the customer's opinion, and complete it, providing useful information to the manufacturer.

5   Evaluation

5.1   Corpus

In order to build our opinion detection system, we used a corpus of reviews available on the website "Epinion"[1]. This is a general site compiling millions of user reviews about products, movies, books, etc. These reviews are semi-structured and contain the following information:

- The product name;
- the overall score (from 0 to 5 stars): we consider that reviews ranked from 0 to 2 are globally negative and that reviews from 3 to 5 stars are globally positive;
- the review title;
- the creation date;
- the sections "Pros", "Cons" and "Bottom Line";
- the content of the review in free text, with the assessment: "Recommended": "yes" or "no".

As our first target application deals with consumer reviews about printers, we extracted a corpus of about 3,500 printer reviews from this site. These reviews in html format have been cleaned and converted into xml format, ready for processing.

---

[1] http://www.epinions.com/

5.2 Experimental Results

We selected an unseen part of the corpus previously described in order to perform the evaluation of the SfI detector prototype we developed. The main problem was that this corpus was not annotated in terms of suggestions from the user. We have therefore extracted randomly 60 reviews from our initial corpus and annotated them manually. After running the prototype against this sub-corpus we obtained the following measures (table 1):

Table 1. Evaluation results

Precision	Recall	F-measure
77%	70%	73%

These results are quite satisfying regarding the complexity of the task, as the approach we adopted frequently catches suggestions from users which can lead to product improvements.

However, regarding precision, some erroneous suggestions were extracted. Among them, we can find ironic suggestions like in *Perhaps they should have stuck with making sewing machines and typewriters* or in *I soon discovered, however, that for the money they spent on this jammin' little number, they should have spent a couple of nights on the beautiful Oregon Coast and pampered themselves at expensive restaurants instead.*. At the moment, we are unable to automatically detect irony, and, apart from this, these sentences match perfectly our formal criteria.

Furthermore, our system also detects erroneously complex expressions where a suggestion is present just to reinforce customer satisfaction, as for instance in *...in my opinion, the lack of high end features is a very fair trade-off for the outrageously low price tag.* Our system also detects suggestions expressed in beliefs like in *Because of previous HP experiences, I knew the printer would work great, but I figured the other features would lack in quality.*, which can not be considered as a suggestion for improvement.

Regarding recall, our system sometimes misses interesting suggestions for improvement mentioned by the users in the reviews. Here are some examples of this kind of errors.

We did not consider that interrogative forms like *Why not just merge into one easy-to-use program?* could sometimes correspond to suggestions for improvement. In this sentence, although the work *program* belongs to the terminology related to the printer and the adjective *easy-to-use* is detected as a positive adjective related to *program*, we have no suggestions template stating that interrogations around a positive aspect of a printer attribute can correspond to a suggestion. This kind of patterns can be added however to the system. Another source of errors is related to the capability of the parser to deal with colloquial language as for instance in *it woulda been nice if this model had an ethernet link*

built in. Finally, we noticed that our system was not able to deal with vagueness in suggestions: sometimes previous textual context makes clear what the user wants, but locally, at sentence level, the improvement needed remains very vague. This is for example the case of *Backlinghting on the LCT screen would help this thing so much.*

5.3 Potential Extension

As said before, we had not corpora annotated in term of suggestions from the users. However, many companies nowadays propose users to express what they want to be improved in products, either by filling some forms[2] or by interacting on some dedicated web forums[3].
This shows how useful automatic suggestion detection would be for these companies, but also potentially constitute a very valuable corpus for the development of such an application. One of the perspective of our work is to use this kind of user-generated corpus to simulate a manually annotated corpus in term of suggestions (hopefully with a minimal human intervention).
From such a corpus, we plan to develop an automatic classifier for sentences containing suggestions for improvement. The idea is to use the different patterns presented in this paper as features for the classifier. We have previously experimented Support Vector Machines based on linguistic features for positive or negative review classification, so we plan to apply the same kind of technique for suggestion classification.

6 Conclusion and Perspectives

We have developed an approach that automatically detects suggestions for improvement expressions from customer reviews. This approach relies on linguistic knowledge and is combined with an opinion mining system. It enriches the opinion mining module if we consider that when a user suggests improvements for a product it means that (s)he is not entirely satisfied with this product. Furthermore, analysis of suggestions from customers can be of great help for the product manufacturers. And additionally, such suggestions, when automatically extracted from customer reviews can also be very useful to improve recommendation systems (see [10]).
Our future perspective is to enrich the patterns for suggestions for improvement, in particular to be able to deal with suggestions expressed in the interrogative form. Furthermore, as this detection module is integrated in a feature-based opinion detection system, we also want to further refine the extraction rules, in order to precisely detect the targets of the suggestions. The plan is also to build an automatic classifier of sentences containing suggestions using these linguistic patterns as features.

---

[2] e.g. http://www.sonycreativesoftware.com/support/productsuggestion.asp
[3] e.g. http://community.norton.com/t5/Product-Suggestions/bd-p/Suggestions

References

1. Liu, B.: Sentiment analysis and subjectivity. Handbook of Natural Language Processing **2nd ed** (2010)
2. Bloom, K., Garg, N., Argamon, S.: Extracting appraisal expressions. In: HLT-NAACL 2007, Rochester, NY (2007) 308–315
3. Kim, S.M., Hovy, E.: Identifying and analyzing judgment opinions. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. HLT-NAACL '06, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 200–207
4. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: KDD 04, ACM (2004) 168–177
5. Aït-Mokhtar, S., Chanod, J.P., Roux, C.: Robustness beyond shallowness: incremental deep parsing. Nat. Lang. Eng. **8** (2002) 121–144
6. Brun, C.: Detecting opinions using deep syntactic analysis. In: RANLP. (2011) 392–398
7. Brun, C.: Learning opinionated patterns for contextual opinion detection. In: Proceedings of the 24th Conference on Computational Linguistics. COLING 2012, Mumbai, India (2012)
8. Goldberg, A.B., Fillmore, N., Andrzejewski, D., Xu, Z., Gibson, B., Zhu, X.: May all your wishes come true: a study of wishes and how to recognize them. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. NAACL '09, Stroudsburg, PA, USA, Association for Computational Linguistics (2009) 263–271
9. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT '05, Stroudsburg, PA, USA, Association for Computational Linguistics (2005) 339–346
10. Stavrianou, A., Brun, C.: Opinion and suggestion analysis for expert recommendations. In: Proceedings of the Workshop on Semantic Analysis in Social Media, Stroudsburg, PA, USA, Association for Computational Linguistics (2012) 61–69