

Learning Opinionated Patterns for Contextual Opinion Detection

Caroline Brun¹

(1) Xerox Research Centre Europe, 38240 Meylan, France
Caroline.Brun@xrce.xerox.com

ABSTRACT

This paper tackles the problem of polar vocabulary ambiguity. While some opinionated words keep their polarity in any context and/or across any domain (except for the ironic style that goes beyond the present article), some other have an ambiguous polarity which is highly dependent of the context or the domain: in this case, the opinion is generally carried by complex expressions (“patterns”) rather than single words. In this paper, we propose and evaluate an original hybrid method, based on syntactic information extraction and clustering techniques, to learn automatically such patterns and integrate them into an opinion detection system.

TITLE AND ABSTRACT IN FRENCH

Apprentissage de patrons polarisés pour la détection contextuelle d’opinions

Cet article se penche sur le problème de l’ambiguïté du vocabulaire de polarité. Alors que certains mots conservent la même polarité dans n’importe quel contexte ou domaine (à l’exception du registre ironique qui va au-delà du présent article), d’autres ont une polarité ambiguë dépendante du contexte ou du domaine : dans ce cas l’opinion est portée par des expressions complexes (patrons) et non des mots isolés. Dans cet article, nous proposons et évaluons une méthode hybride originale, utilisant de l’information syntaxique et des techniques de « clusterisation », pour apprendre automatiquement de tels patrons et les intégrer à un système de détection d’opinions.

KEYWORDS: opinion detection, polar vocabulary ambiguity, hybrid method

KEYWORDS IN FRENCH: détection d’opinions, ambiguïté du vocabulaire de polarité, méthode hybride

Introduction

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. The most fine grained analysis model is the feature based sentiment mining method. Feature based opinion mining aims at to determining the sentiments or opinions that are expressed on different features or aspects of entities (e.g. [Bloom et al. 2007]).

The context of this paper is the development of a feature-based opinion mining system, for French. One of the essential tasks in the course of this development is the acquisition of polar vocabulary, for which one encounters almost immediately the problem of polarity ambiguity. In the present paper, we try to address this particular problem: while some opinionated words keep their polarity in any context and/or across any domain (except for the ironic style that goes beyond the scope of the present article), some other have an ambiguous polarity and are highly dependent of the context or the domain. In this case, the opinion is generally carried by complex expressions rather than single words. Let's illustrate this problem with some French examples:

- An adjective like “hideux” (hideous) can be considered to have a negative polarity in any context and any domain;
- An adjective like “merveilleux” (wonderful) can be considered to have a positive polarity in any context and any domain
- On the contrary, an adjective like “frais” (fresh) in French might have different polarities depending on context and domain :
 - In the context “avoir le **teint frais**” (to have a healthy glow), “frais” has a positive connotation
 - In the context « un **accueil** plutôt **frais** » (a rather cool reception)... “frais” has a negative connotation
 - In the context un “**poisson** bien **frais** (a fresh fish) « frais » has a positive connotation
- An adjective like “rapide » (rapid, fast) in French might also have different polarities depending on context and domain :
 - In the context “l'**impression** est **rapide**” (the printing is fast), “rapide” has a positive connotation
 - In the context “un **résumé** **rapide**” (a short summary), “rapide” is rather neutral.
- Etc.

When building an opinion detection system, it is necessary to be able to disambiguate these polar expressions and associate them the adequate polarity, i.e. positive or negative, according to the context. In this paper, we focus on the extraction of contextual patterns that carry a given polarity. In other terms, we try to automatically detect the polarity of a term according to the context, i.e. learn contextual polarity patterns, for ambiguous polar adjectives.

After a short review of the related work, we briefly describe our feature based opinion detection system, and then we present the methodology we propose to acquire opinionated patterns, which is based on syntactic information extraction combined with simple clustering techniques. We then show how we have integrated the learned patterns into our opinion detection system, and finally evaluate the benefits of this integration.

Related Work

In the literature about opinion mining, there is a considerable number of works aiming at associating polarity to single words. For example SentiWordnet (Baccianella et al. 2010) is a resource aiming at associating polarity scores to WordNet synsets. Many works try to classify polar adjectives, like for example (Vegnaduzzo 2004) who proposes a distributional method to classify polarity adjective using a small seed of polar adjectives. For French, (Vernier and Monceaux 2010) present a learning method relying on the indexing of Web documents by a search engine and large number of linguistically motivated requests automatically sent. There is considerably less attempts to address the problem of associating polarities to larger expressions, and in particular pairs of words in a given syntactic relation, as we propose here. (Wilson et al. 2005), noticed that polar vocabulary have a “prior polarity” that can change according to the context (negation, diminishers such as “little”, “less”, etc). They learn such contexts by performing classification using various features and an annotated corpus. In the present paper, we focus on different kind of patterns (noun-adj) and also use a different methodology since we only use the marks given to reviews by users and data automatically annotated with our rule-based system to perform the clustering step. (Riloff et al. 2003) propose a bootstrapping process that learns linguistically rich extraction patterns for subjective (opinionated) expressions. High-precision classifiers label are used on un-annotated data to automatically create a large training set, which is then given to an extraction pattern learning algorithm. The learned patterns are then used to identify more subjective sentences. The bootstrapping process learns many subjective patterns and increases recall while maintaining high precision. While it as some similarities with the work proposed in this paper, is also quite different since they try to learn opinionated syntactic patterns while we try to learn opinionated pairs of words, contextually dependent in a given syntactic relation. They also make use of annotated data, while we only use the marks given to reviews by users and data automatically annotated with our rule-based system in order to perform the clustering step.

Our Opinion Detection System

The opinion detection system we build relies on a robust deep syntactic parser, c.f. (Ait-Mokhtar et al. 2002), as a fundamental component, from which semantic relations of opinion are calculated. Having syntactic relations already 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 the 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 (Brun 2011) for the complete description of these

different components. At this step of development of the system for the French language, we have built generic rules for extracting opinion relations and a generic polar lexicon containing elements that can be considered as non ambiguous in terms of polarity. The work described in this paper aims at enriching this system with patterns that disambiguate ambiguous polar terms according to their context of appearance.

Learning Opinionated Patterns

As said in introduction, our goal is to try to automatically detect the polarity of a term according to the context, i.e. learn contextual polarity patterns, for ambiguous polar adjectives. We focus on NOUN-ADJ expressions, where the adjective is qualifying the noun and that can be mainly found in texts within two types of expressions, adjectives in modifier (1) or attribute (2) position:

- (1) « un **accueil sympathique** », ... "a sympathetic reception")
- (2) « la **cuisine est inventive** », le **service est lent**, ... (« the cooking is inventive », « the service is slow »)

To perform this task, we first collect a large corpus of customer reviews from the web, where such opinionated patterns can be found. We then use a robust syntactic parser to extract the candidate patterns, i.e. the modifier and attribute relationships presented above. We apply clustering techniques to group automatically the pattern according to their polarities. These different steps are detailed in the remaining of this section.

Corpus Selection

We have extracted a large corpus of online user's reviews about restaurant in French, extracted from the web site (<http://www.linternaute.com/restaurant/>). The reviews in html format have been cleaned and converted into xml format. Here's an example of such review, which contains a title (the name of the restaurant), and one or more user reviews containing the user rating of the restaurant and a free text comment:

```
<review>
<title> Brasserie André, restaurant gastronomique à Lille</title>
<userreview>
<rating>3</rating>
<comment> Très bonne adresse, les salades sont copieuses, le coin retiré de la circulation, rapport qualité
prix très correct. </comment>
(Very good place, salads are substantial, the place is far from traffic, value for money quite correct.)
</userreview>
</review>
```

The corpus we have collected contains 99364 user's reviews about 15473 different restaurants, i.e. 260 082 sentences (3 337 678 words). The repartition of the reviews according to the rating given by the users is shown on table 1. We consider that reviews rated from 0 to 2 are negative and that reviews rated from 3 to 5 are positive.

| | | | | | | | |
|-------------------|------|------|------|-------|-------|-------|-------|
| User's rating | 0/5 | 1/5 | 2/5 | 3/5 | 4/5 | 5/5 | total |
| Number of reviews | 2508 | 8810 | 7511 | 14142 | 41382 | 25011 | 99364 |

TABLE 1 – Repartition of reviews according to user's rating

Pattern Extraction

In order to extract the patterns we aim at classifying as positive or negative, we use the robust syntactic parser presented in section 2, which detects such relations (attribute modifier relations between noun and adjectives). Moreover, as this work aims at improving an opinion detection system, we also use the opinion detection component we have developed on top of this robust parser (see (anonymous_reference)). We filter out patterns that are already marked as positive and negative by the opinion detection system (because they contain single polar terms that are already encoded in the polar lexicon of the system) and keep only the patterns that do not carry any information about polarity. The parser outputs syntactic relations among which we select the noun-adj modifiers and noun-adj attributes. We then count the number of occurrences of these relations within reviews rated 0, 1, 2, 3, 4 and 5. Moreover, we use the existing opinion detection system presented previously in order to also count the number of time a given pattern co-occurs with positive opinions and with negative opinions, on the whole corpus of reviews. Some examples of the results are shown in table 2:

| Review rating \ Noun,adj patterns | 0/5 | 1/5 | 2/5 | 3/5 | 4/5 | 5/5 | Frequencies of co-occurring positive opinions | Frequencies of co-occurring negative opinions |
|-----------------------------------|----------------------------------------|-----|-----|-----|-----|-----|-----------------------------------------------|-----------------------------------------------|
| | Frequencies of patterns within reviews | | | | | | | |
| <i>addition, convenable</i> | 0 | 0 | 0 | 1 | 1 | 0 | 6 | 0 |
| <i>estomac, solide</i> | 2 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| <i>service, minimum</i> | 1 | 4 | 5 | 3 | 0 | 0 | 21 | 11 |
| <i>service, lent</i> | 30 | 87 | 71 | 71 | 64 | 10 | 707 | 399 |
| <i>service, rapide</i> | 0 | 1 | 2 | 2 | 6 | 6 | 55 | 7 |

TABLE 2 – Frequency counting for some example noun-adj patterns

We end up with a list of 29543 different NOUN-ADJ patterns together with their number of occurrences per type of reviews as well as the number of co-occurring positive and negative opinions within the whole corpus.

Clustering

In this step, we aim at clustering together the patterns to group them according to their polarity. We use the frequencies per type of review and the number of co-occurring positive and negative opinions previously extracted as features for clustering algorithms. We use the Weka software (Hall et al. 2009) that embeds a collection of machine learning algorithms for data mining tasks, among which clustering algorithms. We tested several algorithms and choose to use the Kmeans¹ algorithm. We experimentally try several numbers of clusters as target for the algorithm, as we have a relatively large number of data to cluster (~30 000 patterns). We needed to have a trade-off between number of clusters and precision of the results: a too small number of clusters gives imprecise results, a too large number of clusters is difficult to evaluate and useless (for example starting from N=60 clusters, a lot of clusters contain only 1 element, which is not interesting). We found this trade-off with a number of 50 clusters, that we reorder from the smallest to the largest, since the smallest clusters are the more accurate and contain the most frequent elements. Here is the content of the very first clusters (with the associated numerical features):

Cluster1 (5 elements) :

| | |
|----------------------------------------------|-------------------------|
| prix,élevé,41,77,45,57,62,15,541,321 | <i>(high,price)</i> |
| service,lent,33,107,92,95,80,13,707,399 | <i>(service,slow)</i> |
| attente,long,31,69,70,50,60,14,521,342 | <i>(wait,long)</i> |
| service,long,69,280,233,255,218,37,1637,1012 | <i>(service,long)</i> |
| accueil,froid,35,95,53,33,29,3,297,223 | <i>(reception,cool)</i> |

Which is clearly a cluster of expressions with negative polarity;

Cluster2 (9 elements) :

| | |
|---------------------------------------------------|------------------------------|
| cuisine,simple,4,25,56,225,362,109,1910,133 | <i>(cooking,simple)</i> |
| restaurant,petit,8,26,32,213,608,244,2286,182 | <i>(restaurant,small)</i> |
| produit,frais,7,24,45,246,1049,637,5138,324 | <i>(product,fresh)</i> |
| prix,abordable,3,11,17,102,363,250,2117,101 | <i>(price,affordable)</i> |
| service,rapide,22,72,117,478,1180,433,5920,514 | <i>(service,fast)</i> |
| cuisine,original,2,10,23,115,451,210,1949,115 | <i>(cooking,original)</i> |
| service,efficace,7,19,31,142,451,140,2337,177 | <i>(service,efficient)</i> |
| resto,petit,4,7,30,152,404,187,1739,98 | <i>(resto,small)</i> |
| cuisine,traditionnel,5,12,28,161,427,169,1814,108 | <i>(cooking,traditional)</i> |

Which is clearly a cluster of expressions with positive polarity;

Cluster3 (10 elements):

| | |
|------------------------------------------|--------------------------------|
| poisson,frais,2,5,8,44,155,82,775,71 | <i>(fish,fresh)</i> |
| ambiance,familial,2,1,5,43,155,88,719,48 | <i>(atmosphere,family)</i> |
| cuisine,fin,3,4,10,58,309,152,1336,61 | <i>(cooking,delicate)</i> |
| oeil,fermé,1,3,1,13,119,170,1150,54 | <i>(eyes,shut)~blindfolded</i> |

¹ There might be alternative clustering algorithms, we use this one because it was accurate and fast and gave.

| | |
|------------------------------------------|--------------------|
| choix,grand,1,3,15,49,233,70,924,43 | (choice,large) |
| plat,original,3,6,19,60,198,104,1067,85 | (dish,original) |
| choix,large,3,10,9,59,194,66,865,50 | (choice,large) |
| salle,petit,11,18,22,93,191,59,1129,180 | (room,small) |
| service,discret,2,6,19,51,191,77,1143,74 | (service,discreet) |
| carte,varié,1,13,18,82,288,123,1273,65 | (menu,varied) |

Which is clearly a cluster of expressions with positive polarity; etc.

We validated the first 14 clusters, by counting the number of elements of the cluster that have the polarity of the whole cluster. We stopped evaluating at this stage since the accuracy started to be low as well as the corpus frequencies of the elements of the clusters. Thanks to this validation, we end up with a list of 151 positive patterns and 118 negative patterns, i.e. a total of 269 opinionated frequent NOUN-ADJ patterns.

Integration within the Opinion Detection System

At the end of the previous step, we have collected and validated clusters of patterns and associated them a positive or negative polarity. We then inject these results in our rule-based opinion extractor by automatically converting these patterns into rules (in the dedicated format of our robust parser). For example a pattern like “service,lent”, which belongs to a negative cluster (cluster1 showed before), is automatically converted into the following rule:

```
|#1[lemma:"lent", negative=+ |
If ( ATTRIB(#2[lemme : « service »],#1) | ADJMOD(#2[lemme : « service »],#1))
~
```

This rule assigns the semantic feature « negative » to the adjective “lent”(#1) (“slow”), **if and only if** this adjective is in attribute or modifier relation with the noun “service”(#2), (“service”). Then, the opinion detection component that is applied afterward benefits from these polar rules to extract opinion relations accordingly.

Using these rules, if the input sentence is: “Le service est lent.” (*the service is slow*), the system extracts a negative opinion relation : OPINION[negative](service,lent). While if the input sentence is: “La cuisson doit etre lente.” (*the cooking should be slow*), the system does not extract any opinion relation, because the association “cuisson, lente” is rather neutral.

It is quite straightforward to convert automatically the clustered validated patterns into this kind of rules that then can be applied on top of the parser, and integrated into the opinion detection module. This specific parsing component contains 269 such rules.

Evaluation

In order to evaluate the impact of the learned opinionated rules on the overall performance of the opinion detection system, we compare the application of the system to review’s classification task, with and without including the new resource. The corpus we have collected can be considered as annotated in terms of classification, since the user gives an explicit mark: 0, 1, 2 = negative and 3, 4, 5 = positive. We use the

relations of opinions extracted by our system to train a SVM binary classifier (SVMLight, Joachims 1999) in order to classify the reviews as positive or negative. The experimental setup² consists in 25000 reviews extracted randomly from the initial corpus to train the SVM classifier, 3500 reviews extracted randomly for validation and 3500 reviews extracted randomly for testing. The SVM features are the relations of opinion on a given *target concept* and their values are the frequencies of these relations within a given review, e.g. OPINION-POSITIVE-on-SERVICE:2, OPINION-NEGATIVE-on-CUISINE:1 , etc. Using this information, we evaluate the system ability to classify reviews according to an overall opinion, and we run exactly the same test with the same data, respectively with and without the integration of our new learned resource of opinionated patterns. The following table shows the results we obtain on the test set.

| Test set | positive reviews | negative reviews | Total reviews |
|------------------------------------------------------------------------------------------|------------------|------------------|---------------|
| Number | 1750 | 1750 | 3500 |
| Accuracy of the classification : system without the learned resources (~baseline) | 81,6% | 78.6% | 80.1% |
| Accuracy of the classification : system including the learned resources | 85.7% | 83.1% | 84.4% |

TABLE 3 – Results on review classification task

Both results are in line with state of the art results, obtained for similar classification tasks, cf. (Pang et al. 2002) or (Paroubek et al. 2007), but the patterns, once encoded into our system, improve the classification task accuracy of about 3.3%, which is a quite satisfying result.

Conclusion and Perspectives

In this paper, we propose an original hybrid method to cope with the problem of ambiguous polar vocabulary, by automatically learning contextual patterns and encode them into an opinion detection system. The learning step consists in syntactic pattern clustering using frequencies extracted thanks to the ratings given by the user in review’s comments, and frequencies about co-occurring opinions extracted by an opinion detection system. This system is then enriched with the new learned patterns. The evaluation on the task of review classification provides encouraging results. We plan to pursue this work along three perspectives. This first one will be to investigate other types of syntactic patterns for example SUBJECT or OBJECT relations between verbs and nouns or MODIFIER relation between nouns and nouns, in order to enrich the opinion detection system new opinionated patterns. The second is to apply the methodology to opinion detection in English. The last perspective is to improve the clustering step by investigating methods to automatically detect the optimal number of cluster, as for example proposed in (Pham et al. 2005) or (Arthur 2007).

² We constrained a 50% repartition of positive and negative reviews on the train, validation and test corpora.

References

- Ait-Mokthar, S., Chanod, J.P. (2002). Robustness beyond Shallowness: Incremental Dependency Parsing. *Special Issue of NLE Journal*.
- Arthur, D. and Vassilvitskii, S. (2007). k-means + +: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1027-1035.
- Baccianella, S., Esuli, A. and Sebastiani. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. ELRA.
- Bloom, K. Navendu G., Argamon S. (2007). Extracting Appraisal Expressions. In *Proceedings of HLT-NAACL*, Rochester, USA.
- Brun. C. (2011). Detecting Opinions using Deep Syntactic Analysis. In *Proceedings of RANLP 2011 (Recent Advances in Natural Language Processing)*. Hissar, Bulgaria.
- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)*, Seattle, Washington, USA.
- Hall, M., Eibe, F., Holmes, G. , Pfahringer, B., Reutemann, P. and Witten I. H. (2009); The WEKA Data Mining Software: An Update. *SIGKDD Explorations, Volume 11, Issue 1*.
- Joachims T. (1999). Making large-Scale SVM Learning Practical. *Advances in Kernel Methods – Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT Press.
- Paroubek, P., Berthelin J.B., El Ayari S., Grouin C., Heitz T., Hurault-Plantet M., Jardino M., Khalis Z., Lastes M. (2007). Résultats de l'édition 2007 du Défi Fouille de Textes, In *Proceedings of DEFT'07*, GRENOBLE.
- Pang B., L. Lee, S. Vaithyanathan. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86.
- Pham, D. T., Dimov, S. S. and Nguyen C. D. (2005) Selection of K in K -means clustering. In *Proceedings of the I MECH E Part C Journal of Mechanical Engineering Science*, Vol. 219, No. 1. , pp. 103-119.
- Riloff, E. and Wiebe J. Learning extraction patterns for subjective expressions. (2003) In *Proceedings of the 2003 conference on Empirical Methods in Natural Language Processing*, Morristown, NJ, USA: Association for Computational Linguistics, p. 105–112.
- Vegnaduzzo, S. (2004). Acquisition of subjective adjectives with limited resources. In *Proceedings of the AAI spring symposium on exploring attitude and affect in text: Theories and applications*, Stanford, US.
- Vernier M. et Monceaux L. (2010) Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques. *Revue TAL volume 51 (1)*, pp. 125-149.

Wilson, T., Wiebe, J. and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 347-354.