Opinion Extraction Applied to Criteria

Benjamin Duthil[1], François Trousset[1], Gérard Dray[1], Jacky Montmain[1], and
Pascal Poncelet[2]

[1] EMA-LGI2P, Parc Scientifique Georges Besse, 30035 Nîmes Cedex, France
`name.surname@mines-ales.fr`
[2] LIRMM, Université Montpellier 2, 161 Rue Ada, 34392 Montpellier, France
`name.surname@lirmm.fr`

Abstract. The success of Information technologies and associated services (*e.g.*, blogs, forums,...) eases the way to express massive opinion on various topics. Recently new techniques known as *opinion mining* have emerged. One of their main goals is to automatically extract a global trend from expressed opinions. While it is easy to get this overall assessment, a more detailed analysis will highlight that the opinions are expressed on more specific topics: one will acclaim a movie for its soundtrack and another will criticize it for its scenario. Opinion mining approaches have little explored this multicriteria aspect. In this paper we propose an automatic extraction of text segments related to a set of criteria. The opinion expressed in each text segment is then automatically extracted. From a small set of opinion keywords, our approach automatically builds a training set of texts from the web. A lexicon reflecting the polarity of words is then extracted from this training corpus. This lexicon is then used to compute the polarity of extracted text segments. Experiments show the efficiency of our approach.

1 Introduction

Web technologies development have made numerous textual records available. The rapid increase of this mass of information requires efficient support system to ease the search of relevant information. Numerous tools are already designed in this way. For instance exhibiting customers opinion on a specific product, searching or automatically indexing documents are contemporary concerns. In particular, numerous tools have been developed for moviegoers to know the global trends of opinions on movies. However aggregated information found on the web does not always reflect the semantic richness provided by the critics. This is highlighted in figure **??** which presents one's opinion expressed on the movie "Rise of the Planet of the Apes". We may notice that this individual estimation gives a score of 1/10 for this movie. It may be explained in the following manner: the character development and the storyline have brought the score down despite the "Great special effects" from Rupert Wyatt noticed in the critic. The aggregated score hides the divergence between the different criteria assessments. Thus this overall score will not reflect the semantic richness of the text. This observation is not specific to the domain of movies. This issue may be found in every domain

when a sentiment analysis is done as in politics ([1]), e-commerce ([2]) or recommender systems ([3])...

In this paper we propose an approach to answer the problem of opinion identification on selected criteria. A first step consists in extracting text segments related to one criterion ([4]). In a second step the polarity of each segment relatively to the criterion is identified. This process is iterated for each selected criterion. The way opinions are expressed may be quite different from one document to another and are often specific to the thematic the document deals with. Thus the vocabulary which is used depends on this thematics ([5]). This vocabulary is then automatically learned for the thematic prior to any text or opinion extraction.

Learning the vocabulary is usually performed using supervised method (e.g., Mindserver Categorization, Thunderstone, ...). This require annotated training sets to extract the vocabulary of opinion specific to a given topic. However, in the web context, creating annotated corpus for each criterion is very expensive and even discouraging.

Indeed, considering the diversity of documents (e.g., blogs, forums, journalistic dispatches), numerous topics of interest (films, news, hi-tech ...), the language levels used may vary significantly from one medium to another. This considerably increases the number of significant words to learn before being able to get meaningful results. Moreover, the mass of data to process makes the manual task of annotation difficult and even impossible. All these facts highlight the interest of being able to automatically build training corpus with minimal human intervention.

The main contributions of this paper are as follows. First, rather than allocating an overall score to the general opinion associated with a document, we propose a more detailed analysis by breaking down this overall assessment on a set a criteria, and giving a polarity of the document for each of them (in the previous example related to movies, the criteria could be story and actor). Furthermore, we propose a method for automatic construction of the training corpus using minimal expertise (a few opinion related keywords).

Based on statistical methods, our approach builds a lexicon of opinion descriptors for the selected thematic. This lexicon is then used to automatically extract polarity of the document for each of the criteria.

The process of multicriteria opinion extraction involves two steps:

- The first step consists in identifying portions of text related to a given criterion. This step is performed with the approach called *Synopsis* proposed by [4]. From a limited number of words defining the criteria, *Synopsis* automatically builds a learning corpus of texts from document extracted from the Web. From this corpus *Synopsis* will then automatically learn the descriptors that will characterize the criterion.
- The second step establishes the polarity (positive/negative) of each text segments extracted by *Synopsis* for the given criterion. This step of opinion detection is the main contribution of this paper. In this step, the approach will automatically constructs a new corpus of texts but this time dedicated to the opinion extraction. It then automatically creates a new lexicon of opinion terms. This is done

by studying their semantic relations (proximity) to generic opinions words (multithematic).

The paper is organized as follows. Section 2 describes the three stages of the approach: Training Corpus Acquisition, Opinion Words Learning and Opinion Extraction. Section 3 presents the conducted experiments. Related works are presented in section 4. And finally, conclusions and perspectives are formulated in section 5.

2 Description of the Opinion Mining process

In this section we present the three main phases of the opinion mining process. The first one consists in the automatic acquisition of the corpus of texts from web document and is presented in section 2.1. The second one explains how learning opinion words is done leading to the creation of the lexicon of opinion descriptors. This process is described in 2.2. Finally, the third step consisting in the opinion extraction is presented in 2.3.

2.1 Acquiring the training corpus automatically

The main objective is to automatically acquire a set of texts of opinions to form the training corpus. This corpus is needed to learn and classify descriptors denoting a positive or negative opinion. Opinion descriptors are specific to the thematic in which there are used ([5]) and their interpretation may drastically change from one thematic or context to another. For example, let us consider the two following sentences: "*The picture quality of this camera is high*" and "*The ceilings of the building are high*". The fist one is related to the thematic of movies while the second focuses on architecture. The adjective *high* expresses a positive opinion in the first sentence and is neutral in the second one. This highlights the unavoidable construction of an opinion training corpus that is related to the thematic of interest.

Moreover, as said in the introduction (cf. section 1), our goal is to identify opinion expressed in various kinds of documents. The set of documents constituting the training corpus must then reflect the heterogeneity of the modes of expression (language levels). To construct such a corpus, the document are taken from the Web.

To ensure the presence of positive and negative opinions in the texts constituting the training corpus, the approach focuses on texts containing at least one opinion word. These words are taken from two sets of given opinion words P (positive ones) and N (negative ones). Those words are called *seed words* and are widely used ([6]):

$P = \{good, nice, excellent, positive, fortunate, correct, superior\}$;
$N = \{bad, nasty, poor, negative, unfortunate, wrong, inferior\}$;

By considering that a document containing at least one word from P (resp. N) but none from N (resp. P) contains positive (resp. negative) data, we can semantically attach documents to each seed word expressing

positive (rep. negative) opinions. Documents are obtained by using a Web search engine and specifying the required thematic and seed word.

Thus, for each seed word g from P (resp. N) concerning the given thematic $\mathcal{T}$ (*movie*), K documents are collected. They express opinions related to g, with same polarity and in the context of $\mathcal{T}$. We thus get 14 corpus. Each of them contains K documents. Let us call S_g the corpus associated to the seed word g.

2.2 Learning opinion descriptors

The objective of the second phase consists in identifying the descriptors carrying opinion in collected documents. They are adjectives and specific "expressions". We call "expression" the concatenation of an adjective and all of the adverbs preceding it in the text. For example: "the ridiculously uneducated", "all bad", "very very good", "very nice", "simply not good", "so very good".

In the documents collected in all the corpus, the approach will now search for adjectives and expressions which are correlated with the seed word associated with the document. Both adjectives and expressions which are carrying opinion are merged into the same concept called *descriptor*. The purpose of this learning phase is to enrich the original sets of seed words with opinion descriptors having same polarity.

To do that task, we consider the following assumption: the more a descriptor is correlated to a seed word (i.e. it is close to the seed word), the more it is likely to have the same polarity as this seed word. In opposite a "distant" descriptor (far from any seed word) is considered irrelevant to this seed word.

At the end of this process, each seed word g gets an associated set of correlated descriptors which is called its *class* C_g. In the opposite, the sets of descriptors that have been found "distant" from the seed word is called its *anti-class* AC_g. Thus the *class* is considered to have the polarity of its associated seed word and the *anti-class* to have opposite polarity. The basic idea carried by this discrimination is to identify relevant and irrelevant descriptors by studying their frequency in the *class* and the *anti-class*. If a descriptor is more common in the *class* than in the *anti-class* it will be considered relevant and with same polarity as the *class*. In the opposite, if a descriptor is more common in the *anti-class* than in the *class* it will be considered relevant with the polarity of the *anti-class*. Other descriptors having similar correlation to the *class* and to the *anti-class* are simply said irrelevant.

To compute the proximity of descriptors to seed words, we introduce the notion of window F of size sz. During this phase, windows are all centered on seed words g found in a document t belonging to S_g. Such a window is a set of words m and is defined as follows:

$$F(g, sz, t) = \{m \in t / d_{JJ}^t(g, m) \leq sz\} \tag{1}$$

where $d_{JJ}^t(g, m)$ is the distance between the word m and the seed word g in text t. It is the number of adjectives (JJ) counted between m and g plus one.

The learning phase

Let $S_P = \bigcup_{g \in P} S_g$ and respectively $S_N = \bigcup_{g \in N} S_g$. Thus S_p (resp. S_N) is the corpus of text obtained from all positive (resp. negative) seed words. Similarly we construct $C_P = \bigcup_{g \in P} C_g$, $C_N = \bigcup_{g \in N} C_g$, $AC_P = \bigcup_{g \in P} AC_g$ and $AC_N = \bigcup_{g \in N} AC_g$ which are sets of descriptors.

The learning phase is based on a discrimination technique *class/anti-class* as explained previously but is done on all corpus from S_P (resp. S_N) instead of doing it separately for each seed word g. For any descriptor M a frequency $X(M)$ in the class C_P (resp. C_N) is computed as follows using windows of size 1 ($sz = 1$):

Let $\mathcal{O}(w, txt)$ be the set of occurrences of descriptor w (word, seed word or expression) in the text txt (it may be a full document or any portion of text taken from any corpus).

$$X(M) = \sum_{g \in \mathcal{P}} \sum_{t \in S_g} \sum_{\gamma \in \mathcal{O}(g,t)} |\mathcal{O}(M, F(\gamma, sz, t))| \qquad (2)$$

This score cumulates the frequencies of a descriptor M over all the windows in all the texts belonging to S_P (resp. S_N).

The frequencies $\overline{X}(M)$ in the anti-classes AC_P (resp. AC_N) are computed in a similar way using windows of size 2 ($sz = 2$). In this case, only descriptors outside of all windows are taken into account.

Let $\overline{F}(g, sz, t) = \{m \in t / d^t_{JJ}(g, m) > sz\}$. Thus we get $\overline{F}(g, sz, t) = t \setminus F(g, sz, t)$ (it is the complement of $F(g, sz, t)$ in the text t). Then $\overline{X}(M)$ may be computed as follows:

$$\overline{X}(M) = \sum_{g \in \mathcal{P}} \sum_{t \in \mathcal{T}(g)} |\mathcal{O}(M, \bigcap_{\gamma \in \mathcal{O}(g,t)} \overline{F}(\gamma, sz, t))| \qquad (3)$$

As shown in table **??**, the computed frequencies for $X(M)$ and $\overline{X}(M)$ already provides information about the polarity of the descriptors. These polarities are formally computed using the following discrimination function $Sc(M)$:

$$Sc(M) = \frac{(X - \overline{X})^3}{(X + \overline{X})^2} \qquad (4)$$

In table **??**, a positive value means that the polarity of the descriptor is similar to the one of germs $g \in P$ (resp $g \in N$). Negative values do not actually mean that the descriptor are of opposite polarity but only that they do not have the same polarity: in fact they may be neutral. Thus we only keep the descriptors having positive values. In the example, we may infer that *good* and *very nice* are positive as well as *wrong* and *incorrect* are negative. However it is impossible to take a decision with regard to the polarity of *dramatic* and *not important*. The only information we can attest is that *dramatic* is not positive and that *not important* is not negative.

In the following we will only focus on the descriptors which polarity has been confirmed (positive values in table **??**). For practical reasons, in the

lexicon, the polarity of descriptors in C_P get positive values and those in C_N get negative ones. Table **??** gives an excerpt from such a lexicon. We can notice that *dramatic* and *not important* get no entry in the lexicon.

2.3 Opinion mining

Once the lexicon is built, opinion can be automatically extracted from a text or an excerpt of text dealing with thematic $\mathcal{T}$. A window is again introduced. But this time, instead of centering the windows on seed words and propagating the polarity of the seed word to the descriptor, we reverse the process: windows are centered on every adjective and their polarity is inferred from the ones of all the descriptors it contains. This is done by computing a score for each window f as follows:

$$WSc(f) = \sum_{M \in f} Sc(M) \tag{5}$$

The polarity of a text t is then identified by studying the sign of the score of the text computed as follows:

$$Score(t) = \frac{\sum_{f \in t} WSc(f)}{|\{f \in t\}|} \tag{6}$$

The polarity of the text t is given by the sign of its score: if $Score(t) < 0$ then t is negative and if $Score(t) > 0$ the t is positive.

3 Experiments

We have chosen to validate our approach against the test corpus proposed by [7] designed on the thematic of cinema. This corpus is a compilation of film critics coming from the *Internet Movie DataBase* [3]. Each text has got a human expertize and is labeled positive or negative. The process of validation consists in extracting parts of these texts relative to criteria *actor* and *scenario* (using *Synopsis*) and then identifying their polarity with our approach.

Our results are then compared to those obtained using the well known *SenticNet* ([8]) resource for opinion mining. *SenticNet* gives a set of concepts with associated polarity in $[-1, 1]$ which constitute a semantic network. The lower values are given to higher negative polarity and higher values to higher positive polarities. Zero values means neutral. The polarity of a text is then obtained by computing for each sentence an average of all the polarities of the words it contains. The final polarity is obtained by aggregating those of its sentences.

Traditional indicators as precision, recall and FScore are then used to assess the performance of the classifier.

We consider "relevant" as positive or negative in our case.

---

[3] http://reviews.imdb.com/Reviews

Validation of the approach for text classification

In a first step we compare our opinion mining approach with *SenticNet* one on the whole text without using *Synopsis*. We just learn descriptors relative to the thematic of cinema and compute a global opinion on the whole texts. The results are given in table **??**.

This highlights that our approach is efficient in a context of document classification: its *FScore* is better than the one of *SenticNet*. We may notice some weakness of our approach concerning the detection of positive texts. But this is widely compensated by the one of negative texts for which we obtain a recall of 0.63 which is more than twice the one obtained by *SenticNet* (0.25). These results may indicate that non specific opinion vocabulary may be sufficient to get a global evaluation of a document. We can also notice that in this text classification our approach gives similar results as another global approach. That means that the vocabulary which is automatically learned is relevant for this kind of task and that the sole learning of the adjectives and expression appears sufficient. Furthermore we must point out that our approach is completely automated contrary to the one of *SenticNet* which is fully supervised. On our side we just need a minimal expertise to tell the system what means positive and negative through a small set of seed words.

Validation upon criteria

The second step of validation consists in evaluating the opinion mining task on chosen criteria excerpts. To do that we have chosen the two concepts *actor* and *scenario*. Once again we use the annotated corpus proposed by [7]. First we extract all the segments of text belonging to any of the two criteria *actor* or *scenario*. This was done by using *Synopsis* [4]. We then extract all the opinions of excerpts related to a criterion in a text and aggregate them to assess the global polarity of the text relatively to each criterion. We simply uses an average as aggregation operator. This is done similarly using the lexicon learned with our approach and the one from *SenticNet*. Results are compared in table **??**.

Analysis of the results shows that opinions formulated on those critics highly depend on criteria. We can notice that criterion *actor* is more in line with the overall opinion expressed in the critics than the *scenario* one. This is highlighted by the lower score obtained for criterion *scenario* than for *actor*. In a multicriteria approach this might correspond to allocate a lower weight to criterion *scenario* than to *actor* one and to check the overall score.

Results obtained in tables **??** represent an average of all the opinions expressed in the whole text. By comparing them to the ones obtained in table **??** we can deduce the well known fact that human mind uses a complex preference aggregator from various criteria rather than a simple average to identify the opinion that rises from the text.

We may notice that the *SenticNet* approach gives lower results than our approach in all cases. It highlights as expected that a vocabulary of opinion specific to the studied thematic gives better results when applied to criterion of this thematic rather than using a universal vocabulary as *SenticNet*. The lower results obtained in the previous experiment (table

??) can be explained as follows: opinions from thematic specific descriptors were covered by other common language opinion words in the first experiment. This significantly decreases the performance of the system. Finally, using our method, these results are very easy to reproduce on any thematic and criteria. As both thematic and opinion extraction methods used in this paper are unsupervised, the required human expertise is minimal. It only consists in giving a very small set of words to *Synopsis* [4] to specify what the criteria are. In the opinion mining task once again we only need a few words to identify the thematic and the polarities This is not the case for approaches like *SenticNet* which need human expertise for all of the descriptors.

4   State of the art

This work focuses on the detection of opinions and feelings in textual data. Today opinion mining concerns are also shared both with natural language processing and information retrieval approaches. It uses lot of techniques from both of these disciplines as text analysis, machine learning, semantic analysis, text mining...

Mainly we can identify two kinds of approaches designed to extract and classify opinions. The binary classification ones that identify two classes of documents (positive and negative ones) and the multi-classes ones that offer several degrees of opinion (example: strongly positive, positive, neutral, negative, strongly negative). The major part of the works has focused on binary classification of documents especially when they are dedicated to recommender systems or point of view confrontation [9]. Opinion mining techniques are now widely used in various area especially in the domain of e-commerce to recommend products after comparing them [10, 11]. Each product supplies a lot of feature that may be evaluated but only part of them are of interest to the consumer [12, 13]. Opinion mining process of interest take account of user needs to produce user customized opinion extraction. This allows sophisticated opinion detection to compare customer opinions on concurrent products. Another opinion mining investigation field concerns sentiment summary [14, 15]. For instance [10] uses sentiment summarization to analyze users behaviors and preferences to detect flagship products. Sentiment summarization aims at identifying polarity of critics as well as the intensity of this polarity and provides the elements of the document that support this polarity. [16, 17] are interested in opinion detection with arguments (*Opinion Mining Reason*) which are subjective sentences extracted from the document that consolidate the opinion assessment.

Other works focus on determining the intensity of expressed opinions. This may be achieved by using preselected sets of seed words and/or linguistic heuristics. Some of them have highlighted that restricting the analysis to sole adjectives improves the performance [18]. However if this fact is true, it has been proved that it is not sufficient. Adverbs must be taken in account. Some nouns and verbs also carry opinions and shall be considered too. Automated text annotation is mainly done in two ways: either based on training corpus or on dictionaries. The first ones mostly

rely on syntactic analysis or on word co-occurrences. The second ones take advantage of the semantics orientation of words given by dictionaries like *WordNet* (http://wordnet.princeton.edu/) to deduce the polarity of sentences and texts [9]. Some of these methods may also use similarity measures computed between candidate words opinion words to deduce the candidate opinion [19].

Even if the results obtained by all of these approaches are very close to ours, they mainly differ at a very important point. In fact all of them need a huge human expertise to be applied (dictionaries, human annotated corpus, language semantics...). In the opposite we think that human expertise is a major constraint that interferes in the objective of having wide used, very flexible and highly adaptive opinion mining methods.

5 Conclusion

This paper presents a novel approach to automatically extract opinions from texts. The orientation given to this work was to minimize the human expertise to be provided to obtain relevant results. Furthermore we relate opinion extraction to multicriteria analysis. Following this guideline, our approach automatically builds the training corpus from which it learns the polarity of descriptors and automatically builds its lexicon lately used for opinion extraction. We have shown that the obtained lexicon pertinence is similar to human annotated ones like *SenticNet* when used in a general context. We have also demonstrated that the descriptors may bring different opinions in different contexts and that our criteria based method is able to automatically construct such a context oriented lexicon. Using such a lexicon in proper context significantly enhances the results. This work illustrates the complexity of human mind in the process of extracting the opinion carried by a text or a sentence. It also shows that this process is not based on an overall assessment but rather on local opinions expressed in terms of criteria which are lately aggregated. This consideration complicates cognitive automation of sentiment analysis task in the way that opinion mining should be combined with thematic extraction to enhance the results.

Prospects related to this work are numerous. Among all, we intend to extend the learning phase to nouns and verbs because they should allow moderating analyzed opinions. We have already worked to obtain imprecise evaluations of expressed opinions using fuzzy set techniques with relevant results. Our goal is to get enhanced multicriteria recommender systems. In parallel our method has successfully been tested on other thematics like restaurant assessment and further experimentations are still in progress. The main task is to construct human annotated test corpus to validate our approach. This is a very expensive task that still consolidates our point of view which tends to leave human outside the learning process as much as possible.

References

1. Thomas, M., Pang, B., Lee, L.: Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In:

In Proceedings of EMNLP. (2006) 327–335

2. Castro-Schez, J.J., Miguel, R., Vallejo, D., López-López, L.M.: A highly adaptive recommender system based on fuzzy logic for B2C e-commerce portals. Expert Systems with Applications **38**(3) (2011) 2441–2454

3. Garcia, I., Sebastia, L., Onaindia, E.: On the design of individual and group recommender systems for tourism. Expert Systems with Applications **38**(6) (2011) 7683–7692

4. Duthil, B., Trousset, F., Roche, M., Dray, G., Plantié, M., Montmain, J., Poncelet, P.: Towards an automatic characterization of criteria. In: DEXA (1). (2011) 457–465

5. Harb, A., Plantié, M., Dray, G., Roche, M., Trousset, F., Poncelet, P.: Web opinion mining: how to extract opinions from blogs? International Conference on Soft Computing as Transdisciplinary Science and Technology (2008)

6. Turney, P.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In Proceedings of 40th Meeting of the Association for Computational Linguistics (2002) 417–424

7. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP). (2002)

8. Cambria, E., Speer, R., Havasi, C., Hussain, A.: Senticnet: A publicly available semantic resource for opinion mining. Artificial Intelligence (2010) 14–18

9. Xu, K., Liao, S.S., Li, J., Song, Y.: Mining comparative opinions from customer reviews for Competitive Intelligence. Decision Support Systems **50**(4) (2011) 743–754

10. Bai, X.: Predicting consumer sentiments from online text. Decis. Support Syst. **50** (March 2011) 732–742

11. He, Y., Zhou, D.: Self-training from labeled features for sentiment analysis. Information Processing and Management **In Press, Corrected Proof** (2010) –

12. Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T.: Mining product reputations on the web. In: In ACM SIGKDD 2002, ACM Press (2002) 341–349

13. Taboada, M., Gillies, M.A., McFetridge, P.: Sentiment classification techniques for tracking literary reputation. In: In Proceedings of LREC 2006 workshop "Towards Computational Models of Literary Analysis? (2006) 36–43

14. Ku, L.W., Lee, L.Y., Wu, T.H., Chen, H.H.: Major topic detection and its application to opinion summarization. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. SIGIR '05, New York, NY, USA, ACM (2005) 627–628

15. Beineke, P., Hastie, T., Manning, C., Vaithyanathan, S.: Exploring Sentiment Summarization. In Qu, Y., Shanahan, J., Wiebe, J., eds.: Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text Theories and Applications. Volume 07., AAAI Press (2004) 1–4

16. Cardie, C., Wiebe, J., Wilson, T., Litman, D.: Combining low-level and summary representations of opinions for multi-perspective question answering. In: In Working Notes - New Directions in Question Answering (AAAI Spring Symposium Series. (2003) 20–27
17. Gopal, R., Marsden, J.R., Vanthienen, J.: Information mining ? Reflections on recent advancements and the road ahead in data, text, and media mining. Decision Support Systems (2011)
18. Andreevskaia, A., Bergler, S.: Mining wordnet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In: In Proceedings EACL-06, Trento, Italy. (2006)
19. Kamps, J., Marx, M., Mokken, R.J., de Rijke, M.: Using wordnet to measure semantic orientations of adjectives. In: In Proceedings of LREC-04, 4th international conference on language resources and evaluation, Lisbon, PT. Volume 4. (2004) 1115–1118