CHAPTER X

A LEXICON-BASED APPROACH TO SENTIMENT ANALYSIS: THE ITALIAN MODULE FOR NOOJ

S. Pelosi – A. Maisto

Abstract

In this paper we present a lexicon–based method for the automatic analysis of opinionated documents. We built with Nooj a Sentiment Lexicon and a grammar network of Contextual Valence Shifters and we tested them on a dataset of multi-domain customer reviews, reaching an average Precision of 74% and Recall of 97% on the document-level classification task.

Introduction

In the last decade, the rise of online commerce, the growth of user generated contents; the phenomenon of costumer empowerment and the increasing impact of the online word-of-mouth [Vollero 2010] made it necessary for companies to automatically extract, analyse and summarise not only factual information, but also opinions, expressing the people’s positive or negative affect, appreciation and judgment regarding any king of product or service they offer [Liu 2010; Bloom 2011]. An opinion can be a positive or negative appraisal about a topic, stated by an opinion holder. It can be represented as a quintuple \((o_j, f_{jk}, oo_{ijkl}, h_i, t_i)\), where \(o_j\) is the object, \(f_{jk}\) is the feature, \(oo_{ijkl}\) is the opinion orientation, \(h_i\) is the opinion holder and \(t_i\) is the time when the opinion is expressed. Undeniably, an appropriate management of the online corporate reputation requires a careful monitoring of the new digital environments that strengthen the stakeholders’ influence and independence and give support during the decision-making processes.
For all these reasons, software able to transform unstructured texts written in natural language into structured data, liable to be stored and queried in database tables, are strongly required.

In the present paper we focus on the document-level sentiment polarity classification, which means classifying an opinionated document as expressing a positive or negative opinion on an object. The whole document is considered as the basic information unit and its Semantic Orientation (SO) is calculated on its base. SO is a subjectivity and opinion measure that weights the polarity (positive/negative) and the strength (intense/weak) of opinions.

Section 2 summarises the most used techniques for the Sentiment Lexicon development and for the Sentiment Classification task; Section 3 describes the Sentiment lexical and syntactic resources we built with Nooj; Section 4 presents the software DOXA, an opinion mining application that applies the Nooj resources to different kinds of opinionated documents and the evaluation of the tool; Section 5 briefly underlines the strengths and the weaknesses of our tool and introduces the future lines of action that our research will take.

State of the art

Many techniques have been discussed in literature to perform the Sentiment Polarity Classification. Lexicon-based approaches always start from this basic assumption: the text sentiment orientation comes from the semantic orientations of words and phrases contained in it. Thus, the SO identification task requires the determination of the individual words polarity [Taboada 2006]. The most commonly used SO indicators are adjectives or adjective phrases [Hatzivassiloglo 1997; Hu 2004; Taboada 2006]; but recently it became really common the use of adverbs [Benamara 2007], nouns [Vermeij 2005; Riloff 2003] and verbs [Neviarouskaya 2009] as well. There are several methods used to build and test those dictionaries, such as Latent Semantic Analysis [Landauer 1997], bootstrapping algorithms [Riloff 2003]; graph propagation algorithms applied on the web [Velikovich 2010; Kaji 2007] distributional similarity [Wiebe 2000]; the use of conjunctions (e.g. “and” or “but”), or morphological relations between adjectives [Hatzivassiloglou 1997]; the context coherency [Kanayama 2006]; the Word Similarity [Mohammad 2009]; and, in the end, the Pointwise Mutual Information (PMI) based on Seed Words [Turney 2002; Velikovich 2010]. Learning and statistical methods usually make use of Support Vector Machine classifiers. Pang et
al. (2002) use Support Vector Machine, Naive Bayes and Maximum Entropy classifiers, with diverse sets of features, such as unigrams, bigrams, binary and term frequency feature weights and others. In the end, as regards the hybrid methods, must be cited the works of Andreevskaia (2008) and Dasgupta (2009). In order to obtain a good Accuracy in results, it is not enough to dispose of Sentiment Lexicons; indeed, the local context often makes the polarity of the sentences change. That is the case of Negation [Choi 2008, Benamara 2012], Intensification [Kennedy 2006; Polanyi 2006], Irrealis markers and Conditional tenses [Taboada 2011; Narayanan 2009]. Rule-based approaches that take into account the syntactic dimension of the Sentiment Analysis are the ones used by Mulder (2004) and Nasukawa (2003).

Method and Resources

In the present work we present a hand-tagged Sentiment Lexicon that has been built with Nooj. Adjectives, verbs and nouns contained in the Nooj Italian dictionaries of simple words have been manually evaluated by excluding the words with a neutral meaning, and by weighting the Prior Polarity [Osgood 1957] of the words endowed with a positive or negative SO. The polarity of the adverbs has been automatically derived from the adjectives of sentiment, using a Nooj morphological grammar that will be described below. In order to obtain two separate scales for the evaluation of the strength (intense/weak) and of the polarity (positive/negative), every entry of the lexicon of sentiment has been weighed combining four tags: +POS (positive), +NEG (negative), +FORTE (intense) and +DEB (weak). This way we created an evaluation scale from -3 to +3 and a strength scale from -1 to +1. Contextual Valence Shifters have been taken into consideration thanks to a network of syntactic grammars that computes the words’ Prior Polarity, making it consistent with the real textual context of words.

In detail, the Nooj dictionary of Sentiment Adjectives contains 5300+ entries. Examples of the evaluation and the strength scales are reported in Table 1. Thanks to the morphological grammar showed in Figure 1, it has been possible to derive the dictionary of Sentiment Adverbs from the Adjectives one. All the adverbs contained in the Italian dictionary of simple words have been put in a Nooj text and the above-mentioned grammar has been used to quickly populate the new dictionary by extracting the ones ending with the suffix -mente, “-ly” and by making
such words inherit the adjectives’ polarity. The Nooj annotations have been manually checked, producing a set of 3600+ adverbs of sentiment.

<table>
<thead>
<tr>
<th>Entries</th>
<th>Translation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>meraviglioso,A+FLX=N88+DRV=ISSIMO:N88+POS+FORTE</td>
<td>“wonderful”</td>
<td>+3</td>
</tr>
<tr>
<td>divertente,A+FLX=N79+DRV=ISSIMO:N88+POS</td>
<td>“funny”</td>
<td>+2</td>
</tr>
<tr>
<td>accettabile,A+FLX=N79+DRV=ISSIMO:N88+POS+DEB</td>
<td>“acceptable”</td>
<td>+1</td>
</tr>
<tr>
<td>insapore,A+FLX=N79+DRV=ISSIMO:N88+NEG+DEB</td>
<td>“flavourless”</td>
<td>-1</td>
</tr>
<tr>
<td>cafone,A+FLX=N88+DRV=ISSIMO:N88+NEG</td>
<td>“bumpkin”</td>
<td>-2</td>
</tr>
<tr>
<td>disastroso,A+FLX=N88+DRV=ISSIMO:N88+NEG+FORTE</td>
<td>“disastrous”</td>
<td>-3</td>
</tr>
<tr>
<td>straripante,A+FLX=N79+DRV=ISSIMO:N88+FORTE</td>
<td>“overflowing”</td>
<td>+1</td>
</tr>
<tr>
<td>episodico,A+FLX=N87+DRV=ISSIMO:N88+DEB</td>
<td>“episodic”</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 1 – Extract of the Sentiment dictionaries

![Diagram of morphological grammar](image)

Figure 1 – Extract of the morphological grammar used to automatically populate the dictionary of Sentiment Adverbs

The verbs chosen for our sentiment lexicon are the Psychological Semantic Predicates [Gross, 1981, Gross 1995] belonging to the Italian Lexicon-grammar classes 41, 42, 43 and 43B. Between these verbs, a list of 600+ entries has been evaluated and hand-tagged with the same labels used to evaluate the adjectives. The nominalizations of these predicates have been used to manually build the Sentiment dictionary of nouns that comprehends 1000+ entries.

In the end, 500+ Italian frozen sentences containing adjectives [Vietri, 1990; 2011] have been evaluated and then formalised with a pair of dictionary-grammar. Among the idioms considered there are the comparative frozen sentences of the type \( N_0 \ Agg \ come \ C_1 \), described by [Vietri 1990]

1 Other idioms included in our resources are \( N_0 \ essere (Agg + Ppass) \ Prep \ C_1 \) (e.g. Max è matto da legare, “Max is so crazy he should be locked up”); \( N_0 \ essere \ Agg \ e \ Agg \) (e.g. Max è bello e fritto, “Max is cooked”); \( C_0 \ essere \ Agg \ (come \ C_1 + E) \) (Mary ha la coscienza sporca ↔ La coscienza è sporca, “Mary has a guilty conscience” ↔ “The conscience is guilty”), \( N_0 \)
Although a great part of the works on Sentiment Analysis focuses on the simple lexical valence of negative or positive words, it must be noticed that, in many cases, the sentence or the discourse context shifts the valence of individual terms [Polanyi 2006].

In order to find the Semantic Orientation of real sentences written in natural language, a grammar net that computes the polarity of the opinion lexicon has been built with Nooj (Figure 2). In our grammar, adjectives, adverbs, nouns and verbs have been treated separately in four dedicated metanodes. A fifth metanode has been dedicated to domain-independent sentiment expressions that are not built around specific sentiment words, but must be considered opinion indicators as well. The Sentiment Pattern Extraction and the consequent text annotation are performed using six different nodes (Figure 3) which are enclosed in every metanode of the main graph.

In this work, the metanodes basically work as “boxes” for the Sentiment Expressions, that receive the same label if they are embedded in the same Sentiment box. We will describe below in detail the Contextual Valence Shifters that have been taken into account in the present work: negation, intensification, modality and some kinds of comparative constructions.

As regards negation, we included in our grammar negative operators (e.g. *non*, “not”, *mica, per niente, affatto*, “not at all”), negative quantifiers (e.g. *nessuno*, “nobody” *niente*, *nulla*, “nothing”) and lexical negation (e.g. *senza*, “without”, *mancanza di, assenza di, carenza di*, “lack of”) [Benamara 2012]. Negation indicators not always change a sentence polarity in its positive or negative counterparts (e.g. *La Citroen non produce auto valide*, “Citroen does not produce efficient cars”, Negative sentence); they often have the effect of increasing or decreasing the sentence score (e.g. *Grafica non proprio spettacolare*, “The graphic not quite spectacular”, Weakly Negative sentence). That is why we prefer to talk about valence “shifting” rather than “switching”.

In order to take Intensification into account, we firstly combined in the grammar the words belonging to the strength scale with the sentiment words listed in the evaluation scale. Besides, also the repetition of more than one negative or positive words, or the use of absolute superlative affixes have the effect of increasing the words’ Prior Polarity and, for this reason, have been included into the grammar net. Intensification and

*essere C1 Agg (Mary è una gatta morta, “Mary is a cock tease”). Just because of the higher frequency, we only included C0 essere Agg (come C1 + E) into the idioms’ grammar using its transformation N0 avere C0 Agg.*
negation can also appear together in the same sentence, e.g. *Personale alla reception non sempre [Negative_Operator] gentile [AVV+FORTE] “Not always kind desk clerks.”*, Weakly Negative sentence.

Figure 2 – Main graph of the Contextual Valence Shifters grammar

Figure 3 – Using metanodes as boxes for the Sentiment Pattern Extraction

Modality also impacts the Semantic Orientation of sentiment expressions. We focused on a particular modality type, among the ones analysed by Benamara [2012], the epistemic category, that refers to the Opinion Holder’s personal beliefs and affects the strength and the certainty degree of opinion expressions. It can be expressed by both doubt or necessity adverbs and modal verbs such as dovere (“have to”) e potere (“may/can”). In the present work the just mentioned adverbs have been respectively considered as downtoners and intensifiers and have been, thus, registered in the strength dictionary.

As far as the comparative sentences are concerned, we considered in this work the already mentioned comparative frozen sentences of the type *N0 Agg come C1*; some simple comparative sentences that involve the expressions meglio di, migliore di, “better than”, peggio di, peggiore di, “worse than”, superiore a, “superior to” inferiore a, “less than”; and the comparative superlative.

In the end, in the Fifth metanode of the Sentiment grammar (Figure 2) are listed and computed many cases in which expressions that do not imply the words contained in our dictionaries are sentiment indicators as well. For simplicity, in the present work we put in this node of the grammar the sentences that imply the use of frozen, semi-frozen expression and words that, for the moment, are not part of the dictionaries.
Experiment and results

The dataset used to evaluate our tools has been built using Italian opinionated texts in the form of users’ review and comments found on e-commerce and opinion websites. It contains 600 texts units (50 positive and 50 negative for each product class) and refers to six different domains, for all of which different websites (such as www.ciao.it; www.amazon.it; www.mymovies.it; www.tripadvisor.it) have been exploited. Using the command-line program noojapply.exe, we built a prototype written in JAVA by which users can automatically apply our resources to every kind of text, getting back a feedback of statistics that contain the opinions expressed in each case (Figure 6). With it, we summed up the values corresponding to every sentiment expression and, then, we standardized the result for the total number of sentiment expressions contained in the review. Doxa compared this value with the stars that the Opinion holder gave to his review and provided statistics about the opinions expressed in every domain. In Figure 6 we report the analysis made up on the domain of hotels reviews.

Figure 2 – Sentiment Analysis with Doxa

Because our lexical and grammatical resources are not domain-specific, we observed their interaction with every single part of the corpus, which is composed of many different domains, each one of them characterised by its own peculiarities. Moreover, in order to verify the performances of every part of speech (and of the expressions connected to them) we checked, as shown in Table 2, the Precision and the Recall applying separately every single metanode (A, ADV, N, V, D-ind) of the main graph of the sentiment grammar. The values marked by the asterisks
have been reported to be thorough, but they are not really relevant because
of the small number of concordances on which they have been c
alculated.

As far as the document-level performance is concerned, we calculated
the precision twice by considering in the first case as true positive the
reviews correctly classified by Doxa on the base of a polarity attribution
that corresponds to the one specified by the Opinion Holder; in the second
case by considering as true positive the documents that received by our
tool exactly the same stars specified by the Opinion Holder.

<table>
<thead>
<tr>
<th>Precision (%)</th>
<th>Cars</th>
<th>Smartphones</th>
<th>Movies</th>
<th>Books</th>
<th>Hotels</th>
<th>Videogames</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>83</td>
<td>84,5</td>
<td>51</td>
<td>77</td>
<td>90,5</td>
<td>82</td>
<td>78</td>
</tr>
<tr>
<td>ADV</td>
<td>78,2</td>
<td>75,8</td>
<td>58,2</td>
<td>84,6</td>
<td>92</td>
<td>50*</td>
<td>73,1</td>
</tr>
<tr>
<td>N</td>
<td>70,4</td>
<td>71,4</td>
<td>43,3</td>
<td>63</td>
<td>79,4</td>
<td>71,4*</td>
<td>66,5</td>
</tr>
<tr>
<td>V</td>
<td>88,2*</td>
<td>57,1*</td>
<td>67,2</td>
<td>73,7</td>
<td>57,1*</td>
<td>100*</td>
<td>73,9</td>
</tr>
<tr>
<td>D-ind</td>
<td>79,3</td>
<td>83,5</td>
<td>64,8</td>
<td>70</td>
<td>87,5</td>
<td>89,4</td>
<td>81,8</td>
</tr>
<tr>
<td>Average</td>
<td>79,9</td>
<td>74,4</td>
<td>56,9</td>
<td>73,7</td>
<td>81,3</td>
<td>78,6</td>
<td>74,1</td>
</tr>
<tr>
<td>Document level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polarity</td>
<td>71,0</td>
<td>72,0</td>
<td>63,0</td>
<td>74,0</td>
<td>91,0</td>
<td>72,0</td>
<td>74,0</td>
</tr>
<tr>
<td>Intensity</td>
<td>32,0</td>
<td>45,0</td>
<td>25,0</td>
<td>33,0</td>
<td>49,0</td>
<td>34,0</td>
<td>36,3</td>
</tr>
</tbody>
</table>

Table 2 – Sentence-level Precision

<table>
<thead>
<tr>
<th>Recall (%)</th>
<th>Cars</th>
<th>Smartphones</th>
<th>Movies</th>
<th>Books</th>
<th>Hotels</th>
<th>Videogames</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence-level</td>
<td>72,7</td>
<td>79,6</td>
<td>64,8</td>
<td>65,7</td>
<td>72,1</td>
<td>58,8</td>
<td>69,0</td>
</tr>
<tr>
<td>Document-level</td>
<td>100</td>
<td>98,6</td>
<td>100</td>
<td>96,1</td>
<td>98,9</td>
<td>91,2</td>
<td>97,5</td>
</tr>
</tbody>
</table>

Table 3 – Recall on sentence-level and document level performances

As we can see in the last two rows of Table 2, the latter seems to have
a very low Precision, but at a deeper analysis we discovered that is really
common for the Opinion Holders to write texts that next to never
correspond to the stars they specified. That increases the importance of a
software like Doxa, that does not stop the analysis on the structured data,
but enters the semantic dimension of texts written in natural language.

The Recall pertaining to the sentence-level performance of our tool has
been manually calculated on a sample of 150 opinionated documents (25
from each domain), by considering as false negatives the sentiment
indicators which have not been annotated by our grammar. The document-
level Recall, instead, has been automatically checked with DOXA, by
considering as true positive all the opinionated documents that contained
at least one appropriate sentiment indicator, so the documents in which the
Nooj grammar did not annotate any pattern were the false negatives that
we took into account. Considering the F-measure, the best results were
achieved with the smartphone’s domain (77.0%) in the sentence-level task and with the hotel’s dataset (94.8%) into the document-level performance.

Conclusion

We conclude this work anticipating the future line of action that our research will take: we will improve the performances of our system by enlarging the dictionary of verbs and nouns; by building a sentiment dictionary of bad words; by providing the dictionary of multiword expressions with annotation of sentiment and, in the end, by building a grammar of sentiment expression that is specific for each domain.

Irony (e.g. Quel tocco di piccante (...) è gradevole[A+POS] quanto lo sarebbe una spruzzata di pepe su un gelato alla panna, “And the touch of piquancy (...) is as pleasant as a spattering of pepper on a cream flavoured ice-cream”) and cultural stereotypes (e.g. La nuova fiat 500 è consigliabile[A+POS] molto di più ad una ragazza, “The new Fiat 500 is recommended a lot more to a girl”) still remain open problems for the NLP in general and for the sentiment analysis. For the moment we decided to give up with them, but we do not exclude that in the next feature we will try to face also these challenges.

References


