Chapter 3
An Opinion Mining Model for Generic Domains

Franco Tuveri and Manuela Angioni

Abstract Online users are talking across social media sites, on public forums and within customer feedback channels about products, services and their experiences, as well as their likes and dislikes. The continuous monitoring of reviews is ever more important in order to identify leading topics and content categories and to understand how those topics and categories are relevant to customers according to their habits. In this context, the chapter proposes an Opinion Mining model to analyze and summarize reviews related to generic categories of products and services. The process, based on a linguistic approach to the analysis of the opinions expressed, includes the extraction of features terms from the reviews in generic domains. It is also capable to determine the positive or negative valence of the identified features exploiting FreeWordNet, a WordNet-based linguistic resource of adjectives and adverbs involved in the whole process.

Keywords Opinion mining · Sentiment analysis · Text categorization · Feature extraction · Opinion summarization

1 Introduction

Reviews are used every day by common people or by companies who need to make decisions. They facilitate to book a hotel or a restaurant, to buy a book, or to taste the market tracing the customer satisfaction about a product. It is evident that the opinion monitoring is essential for listening to and taking advantage of the conversations of

F. Tuveri (✉) · M. Angioni
CRS4, Center of Advanced Studies, Research and Development in Sardinia, Sardinia, Italy
e-mail: tuveri@crs4.it

M. Angioni
e-mail: angioni@crs4.it

C. Lai et al. (eds.), *Distributed Systems and Applications of Information Filtering and Retrieval*, Studies in Computational Intelligence 515,
DOI: 10.1007/978-3-642-40621-8_3,

possible customers in a data-driven decision making process or in order to elaborate strategies of marketing research.

Researches about Opinion Mining, also called Sentiment Analysis, are passing through the simple evaluation of the polarity of the expressed feeling, to a deeper analysis of contents where opinions extracted are context related and the information about products and services are more detailed. Because of the overwhelming amount of information available new automatic tools are even more requested and appreciated especially by large organizations that track not only brands but even consumer preferences and opinions. A Gartner analysis for the 2012-year [1] illustrates the expectations about emerging technologies and how the need for automated methods is growing and social media analytics offers an answer [2], as one of the key themes emerging in the near future.

The last "Sentiment Analysis Symposium", hosted by Seth Grimes in New York City, evidenced the state of research about sentiment analysis, bridging technology and business in discovering business values in opinions and attitudes in social media, news, and enterprise feedback. The symposium gave important indications about how Opinion Mining is gaining ground in several domains of interest ranging from military intelligence to financial markets, where traders build strategies around online and social-media sentiment.

In details, Seth Grimes [3] talked about deep Marketing Research (DeepMR), "enabled by an ensemble of text analytics, sentiment analysis, behavioural analyses, and psychometric technologies—applied to social and online sources, as well as to traditional surveys—with the potential to revolutionize market research".

On the other hand rescues coming from a do-it-yourself marketing research [4] are evident. Without the training to spot marketing research problems it is often possible to waste time and resources developing researches that are essentially worthless.

In this context there is someone, like the Keller Fay Group [5], that push to make the word of mouth (WOM) practice a central part of media planning processes.

Whereas word of mouth was once limited to casual feedback or an informal chat during a work break, reviews expressed by customers, describing experiences and perceptions, are now shared on blogs, web forums and product review sites.

So, new tools are under development in order to provide demonstrable metrics how brand conversations influence purchase behaviour and for how marketing influences conversations about brands products and services [5]. Although Opinion Mining applications currently are not thoroughly able to perform deep extraction and elaboration of information related to reviews of products and services, some existing tools can evidence opinions and produce elaborated cross references of data products with timelines and behavioural outcomes.

In this context the chapter describes the development process of an Opinion Mining model for generic domains. The process, based on a linguistic approach to the analysis and summarization of the opinions expressed in a set of reviews, includes the automatic extraction of features from the reviews people express about a product or a service. The term feature is here used with the same sense given by [6] in their approach to the Opinion Mining.

An object O is an entity that can be a product, person, event, organization, or topic. It is associated with a pair, O: (T, A), where T is a hierarchy or taxonomy of components (or parts), sub-components, and so on, and A is a set of attributes of O. Each component has its own set of sub-components and attributes. Given an object, that could be a service, a person, an event or an organization, the term feature is used to represent a sub-component or an attribute describing the object.

The process makes use of FreeWordNet, a WordNet-based linguistic resource of adjectives and adverbs, which plays a relevant role in the whole process. In FreeWordNet each synset is enriched with a set of domain-related semantic properties and with polarity values helpful in order to determine the positive or negative valence of a review in relation to specific features. Moreover FreeWordNet is involved to perform a WSD for adjectives and adverbs, in the steps of distinction and identification of subjective, objective or factual sentences and contributes in a basic way in the task of contextualization of the features.

The remainder of the chapter is organized as follows: Sect. 2 refers to the state of the art and related works. Section 3 introduces our approach to the Opinion Summarization as part of our Opinion Mining model and examines the work performed, giving some details about FreeWordNet and the feature extraction process. In the same section some details are given about the creation of the matrix of features, a structure that permits us to group the features, automatically extracted from a corpus of reviews, in subsets, follows the description of the chunker analysis and finally the summary presentation. Section 4 draws conclusions.

2 Related Works

Several independent vendors are proposing solutions in web and social media analytics, using their prior experience in business intelligence. Although the proposed solutions come in some cases from leader industries skilled in the business intelligence and in text analytics technologies, in most cases solutions do not provide valid approaches to the problems related to Opinion Mining. More in details the state of the art in the text and social media analytics domains, and more in particular in Opinion Mining, is still away from provide a definitive solution to the deep analysis of contents and from give a real semantic interpretation of the meanings expressed in texts.

In this scenario, whereas several industry leaders propose solutions for the customer and consumer analysis with Opinion Mining technologies, it is necessary to separate solutions that meet industry needs from unresolved research questions and how research faces them with state of the art approaches.

In Opinion Summarization several approaches are based on the use of lexicons of words able to express subjectivity, without considering the specific meaning the word assumes in the text by means of any form of semantic disambiguation. Other approaches consider instead the word meanings as [7], that builds and evaluates a supervised system to disambiguate members of a subjectivity lexicon, or [8], that propose a methodology for assigning a polarity to word senses applying a Word Sense Disambiguation (WSD) process.

Some authors [9] asserted that the introduction of the sense disambiguation in text analysis showed that systems adopting syntactic analysis techniques on extracting opinion expressions tend to show higher precision and lower recall than those which do not adopt this kind of techniques. The result has been obtained by the comparison of six Opinion Summarization systems, concentrating on how the overall sentiment of each feature of a product is summarized.

In our approach we take advantage of [7–9] results by developing Free-WordNet and by performing a WSD of the opinions by means of a deeper syntactic analysis. Feature extraction is a relevant task of the opinion summarization process. Some works about features are based on the identification of nouns through the pos-tagging and provide an evaluation of the frequency of words in the review based on tf-idf criterions and its variation [10], as partially done in the feature extraction method proposed but we perform a deeper syntactic analysis and the WSD of the features.

In [11] a very promising study about Opinion Summarization is proposed. The objective of the study, based on data mining and natural language processing methods, is to provide a feature-based summary of a large number of customer reviews of some products sold online. The developed framework performs a semi-structured feature-based opinion summarization. The summarization task is performed in three steps: the extraction of product features commented on by customers, the identification of opinion sentences and the aggregation and summarization of the opinions for each product feature. The framework aims to visually summarize and compare consumer opinions on different products.

Others researchers [12] proposed a constrained semi-supervised learning method based on the contextualization of reviews grouped in specific domains. The method also try to solve the problem to group feature expressions and to associate them to feature labels using a characterization of the features defined by users. They do not use WordNet for several reasons including the problem of the semantic disambiguation, the lack of technical terms or specific meanings related to the context of use, or yet the differences of synonymy between different context.

Finally another important work is [13], that worked on the explicit features in noun phrases.

3 Opinion Mining Model

The Opinion Mining model analyzes and summarizes reviews related to generic categories of products and services and their aspects or features. The process, based on a linguistic approach, includes the automatic extraction of features from the reviews people express about a product or a service and determines the positive or negative valence of the opinions in relation to a specific feature. Figure 1 provides a representative view of the model we developed in our activities depicting the structural elements and their relations.

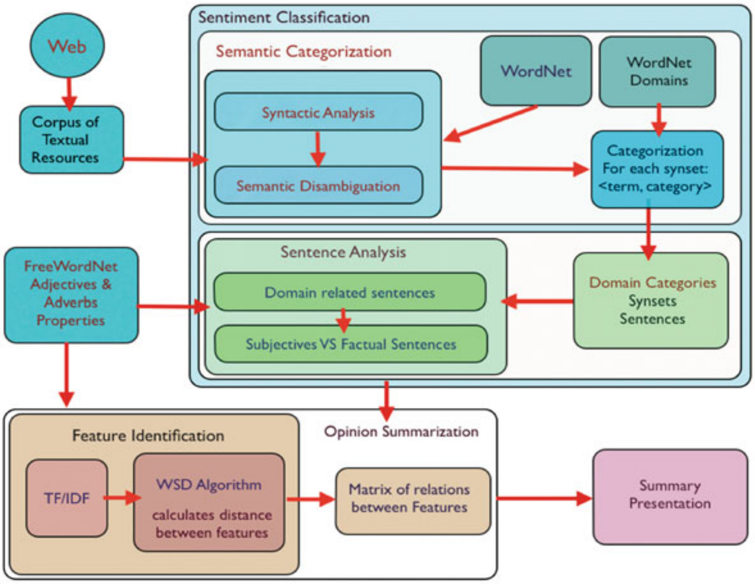

Fig. 1 The schema of the feature extraction process

The opinions are analyzed by two sub-modules that define the Sentiment Classification module: the Semantic Classifier and the Sentence Analyzer. The Sentiment Classification module provides the results to the Opinion Summarization Module.

In Opinion Mining the Opinion Summarization is the process of detection and summarization of the opinion related to relevant terms or expressions in a specific domain. The summarization of opinions is the end point of the whole process of an Opinion Mining system.

As described in [14], the aspect (or feature) based Opinion Summarization is the most common type of Opinion Summarization technique able to generate opinion summaries around a set of aspects or features.

The proposed Opinion Summarization system follows the three main steps of the aspect-based summarization technique: feature identification, sentiment prediction, and summary generation [14]. The main goal of the feature identification step is the identification of main topics within the opinions to be summarized. Sentiment prediction or sentiment classification allows for the discovery of a positive or negative valence about the feature. The summary generation step uses the results of feature discovery and sentiment prediction to generate and present the final opinion summaries in an effective and easy to understand format.

The Opinion Summarization system we developed is based on a process of analysis of opinions built on an automatic method for the extraction of the features from the reviews and based on a linguistic approach to the analysis of the opinions. As described in [15], FreeWordNet, a WordNet-based linguistic resource of adjectives and adverbs plays a relevant role in the whole process. In FreeWordNet each synset is enriched with a set of domain-related semantic properties and with polarity values helpful in order to perform a WSD for adjectives and adverbs, in the steps of distinction and identification of subjective, objective or factual sentences and contributes in a basic way in the task of contextualization of the features. A proper interface helps users to understand the details of opinions based on the information extracted by the method and based on their real needs.

3.1 FreeWordNet

As said, FreeWordNet is a lexical database of synsets in which a number of WordNet adjectives and adverbs have been enriched with a set of properties, with a positive, negative or neutral value associated. The addition of information given by the properties associated to each synset helps to better identify the sentiment expressed in relation to the features giving more details about them.

Some linguistic resources are built considering three properties: subjectivity, orientation, and strength of term attitude. For example, ‘good’, ‘excellent’, and ‘best’ are positive terms while ‘bad’, ‘wrong’, and ‘worst’ are negative terms. ‘Vertical’, ‘yellow’, and ‘liquid’ are objective terms. ‘Best’ and ‘worst’ are more intense than ‘good’ and ‘bad’.

Our analysis concentrates instead mainly on the qualitative adjectives, able to specify for instance colour, size, smell, and on the adverbs classified by their meaning, they position or their strength. We have thus extended the properties of the semantic network of WordNet focusing on the characteristics of adjectives and adverbs. We have classified about 2.300 pairs of adjectives/synsets and about 480 pairs of ad-verbs/synsets. FreeWordNet has been built for version 3.0 of WordNet and maintains an interconnection between the languages: Italian, English, Spanish and Catalan, using the data retrieved by FreeLing [16, 17]. We build ex novo a set of about 11,000 Italian terms, that in future will be made available freely online.

For each adjective and adverb, all the possible synsets available on WordNet has been considered and, for each of the meaning expressed by a synset, a property and a polarity valence has been associated.

The characteristics identified for the adjectives provide additional information about the content of the sentences, regarding for instance personal, moral, ethical or even aesthetical aspects. Some of these categories allow a polarization that can be used by Opinion Mining algorithms. In other cases it is immediately obvious that adjectives contain meanings intrinsically related to geographic, to time or to weather aspects. In our opinion, the use of such qualities associated both to adjectives and adverbs is useful to identify a first level of contextualization about objective and subjective phrases allowing referring to things, people, places and weather conditions that can be contextualized on specific features.

Adverbs are useful too for the identification of the sentiment into the Opinion Mining process. We concentrate on some adverbs associating to each of them specific synsets as made for the adjectives. Based on their characteristics we have considered adverbs of manner, adverbs of place, adverbs of time, adverbs of quantity or degree, of affirmation, negation or doubt, adverbs as intensifiers or emphasizers and adverbs used in adversative and in consecutives sentences. Only the adverbs of manner may be positive, negative or neutral (objectives). The adverbs of degree give the idea about the intensity with which something happens or have an impact on sentiment intensity. Other adverbs, related to categories of places and time, give additional information to the analysis related to the location, the direction and the time.

The introduction of the synsets instead of considering only the words as keywords, extending in future work a similar evaluation to nouns and verbs, allows to have immediately the same qualities and values for the languages whose mapping between synsets is available.

3.2 Sentiment Classification

The creation of the corpus of reviews related to a specific domain is the first step of the process. The reviews are gathered considering only syntactically correct sentences, selected and inserted in the corpus in order to avoid introducing errors and to facilitate the syntactic parser activities. Sentences having orthographic errors are corrected or discarded.

The sentences of the corpus are analyzed by a set of two modules including, at a top level, a Semantic Classifier and a Sentence Analyzer. The first module, the Semantic Classifier, identifies the domain of the corpus by means of a set of categories and their associated weights. During this step, the Semantic Classifier also evaluates the categories and the weights for each sentence, useful to establish if a sentence is relevant, comparing them with the categories describing the domain of the corpus. In a first step it performs a thorough syntactic analysis of the sentences. The TreeTagger [18] parser and chunker executes a phrase chunking process, annotating the text with part-of-speech tags and lemma information and identifying into each sentence its sub-constituents. A Java class wraps the evaluation provided by TreeTagger and, analyzing the parts of speech, identifies the associations between nouns and their related information. Such analysis is used in the semantic categorization process of the corpus of reviews.

The text categorization process provides as result a set of categories and weights able to define the domain for the corpus of reviews. For example, considering a set of reviews about a hotel, the domain is characterized by categories such as *Tourism*, *Person*, *Gastronomy*, and by their weights. The Semantic Classifier also classifies the corpus of reviews evaluating the categories and the weights for each sentence. These categories and weights are used to establish if a sentence is relevant, comparing them with the categories describing the domain of the corpus. For example, analyzing reviews about tourism and especially reviews about hotels, we expect to examine sentences containing opinions about geographical locations, buildings, rooms, staff and food.

The second module, the Sentence Analyzer, manages the categorization of each sentence of the reviews in order to distinguish between subjective and objective sentences, with or without orientation, and in particular in order to detect factual sentences having polarity value. In this phase two sets of categories related to the synsets are used: the semantic one, performed automatically by the Semantic Classifier, and the human one, given by the properties of FreeWordNet. The first set of categories allows excluding sentences not belonging to the domain of the corpus. As said, the properties of FreeWordNet related to the Moral/Ethic or Emotional sphere imply subjective values, while others identifying e.g. Chronologic or Shape properties imply factual valence. In such a way, we consider only subjective sentences or factual sentences having polarity valence. The Sentence Analyzer allows distinguishing between the following cases:

- "The room had the classic moldy smell" is a factual sentence with negative valence.
- "I went with my older sister to Cagliari" is a factual sentence without valence.
- "Our room was modern and spacious" is a subjective sentence with positive valence.

The pre-processing of the corpus of textual resources has been performed in order to acquire different levels of information, related to the whole corpus, to the sentences or to each term. All the information involved in the categorization process is still used in the feature extraction phase in order to perform the disambiguation of the terms and to extract relations between features, adjectives and adverbs.

3.3 The Feature Identification

The feature extraction process consists of two main phases. The first step involves the application of a term frequency—inverse document frequency function (tf-idf) to the nouns contained in the corpus of sentences having polarity orientation, obtaining as result a first list of candidate features.

The number of candidate features is then reduced excluding the features not belonging to the domain. The categories of each feature, resulting by the mapping of each synset on the WordNetDomains categories, are compared with the domain categories.

In the second step the WSD algorithm processes the resulting feature terms in order to perform their disambiguation, excluding synonyms and terms not referred to the domain categories.

The features are now identified by their synsets.

The WSD algorithm calculates the semantic distance between the synsets related to the features using the semantic net of WordNet and is based on the measure of similarity proposed by [19]. The algorithm assigns the most probable meaning of each term in a given domain starting by the evaluation of the minimum distance between the different senses of the term itself and the senses of other features belonging to the domain. The measure is function of the length of the path linking the synsets in the WordNet semantic net by using IS-A relations. The idea behind is that the closer they are, the more they are semantically related.

The algorithm evaluates the semantic distance between each sense of a feature f1 and each sense of a feature f2 by the application of the formula

$$Sim\,(f1, f2) = \max\left[-\log\frac{Np}{2D}\right]$$

where:

- *D* is the maximum depth of the WordNet hierarchy. We assume that the maximum depth D of the noun taxonomy is 18 for WordNet 2.0 considering the presence of a unique root node, as defined in [20].
- *Np* is the number of nodes in path p in the semantic net of WordNet from f1 to f2.

The formula returns the maximum value of similarity calculated between all possible pairs of synsets belonging to two features.

The algorithm also verifies the existence of common categories between the synsets of each pair of features and provides a weight to each synset based on the number of synsets related to each term. In such a way, the algorithm defines a matrix of all the possible relations between the synsets of the features. The rows and columns of said matrix are the disambiguated synsets of the extracted features. The matrix contains as weights the values of distance that measure the strength of the relations existing between two features. The higher the weight, the stronger the relation. By means of the values in the matrix, the system is able to group the features using the strength of their relations.

3.4 Referring Adjectives and Adverbs to Features

The summarization of the opinions is performed considering the association between the features and the adjectives and adverbs included in the sentences.

The wrapper implements a set of rules, based on the sequences of chunks depicted in the graph. The chunks are related to the parts of speech identified by the parser in order to have a precise association between the features and their related information.

In the Fig. 2 N stands for Noun, ADJ for Adjective, ADV for Adverb, V for Verb, PC for Prepositional Chunk, and SENT is the symbol used to indicate the conclusion of the sentence. The set of rules produces better performances in the definition of the relations between adjective, adverbs and the related features and makes easier the production of a feature-based summary of opinions.

Adjectives and adverbs related to the features need to be disambiguated.

The following example shows the WSD of the adjectives. Given the sentence "The arid climate is characterized by a high evaporation and lack of rainfalls" the result of the semantic categorization identifies the most relevant categories (Meteorology 75 %, Psychology 25 %).

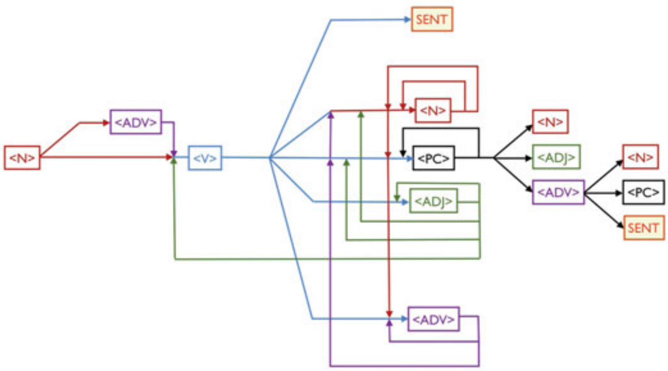

Fig. 2 The possible patterns of chunks

The algorithm calculates, as first parameter, the distance between the nouns identified by the pertainym relationship with the adjective "arid" and the noun "climate". As second parameter the algorithm considers the matching of the most relevant categories with the categories of both the glosses of the adjective "arid". The higher value determines the choice of the synset related to the adjective arid.

3.5 Summary Presentation

The visualization of the opinions and the task of summarization are based on the in-formation the system is able to extract by the method of analysis of the opinions described in previous sections. In particular, during the process, the features related to the reviews are extracted and a matrix of weighted relations between couples of features is generated in order to establish the strength of their relations.

Figure 3 is realized by means of the JavaScript InfoVis Toolkit. The data, stored in a static JSON tree, are loaded into a Squarified Treemap. Figure 3 depicts the visualization of such result referred to a corpus of reviews about a hotel in Cagliari (Sardinia, Italy), where the features extracted by the opinions expressed by users are grouped by means of the strength of their relations. Such screenshot represents the first step of the visualization, aiming at giving the user a general and complete idea of the domain, allowing in further steps to refine the search about the information.

The creation of the matrix allows evidencing the relations between features like Restaurant, Bar, Buffet and Breakfast, grouping them under the same main feature Dining and evidencing this relation by different shades of the same color (fuchsia in this case). Other color shades indicate that other features, such as Balcony or TV, are related to the same feature Rooms. Moreover, the visualization facilitates the user to identify the exact information he is looking for by means of the images the interface is enriched by. The images are extracted by the reviews expressed about the hotel and

Fig. 3 The representation of the features

are related to the specific feature, giving an immediate representation of the same feature the user might be interested. Clicking on each image it is possible to read one or more reviews related to the hotel and referred to the specific feature.

Furthermore, the system allows a summarization of the reviews filtering them through the setting of different parameters and crossing them in order to obtain more refined information, as showed in Fig. 4.

The user can select the initial and the final date of the journey, or can decide to visualize only the reviews related to a specific feature. The search of the precise information is allowed even by the selection of the polarity, choosing between only

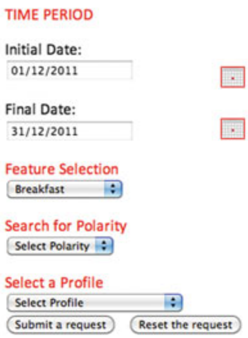

Fig. 4 The filtering options

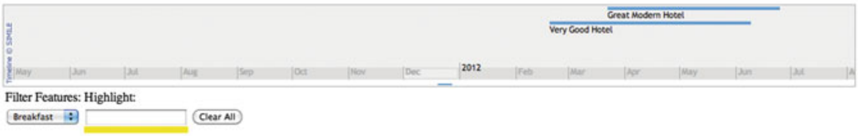

Fig. 5 The reviews selected by time period and filtered by the feature breakfast

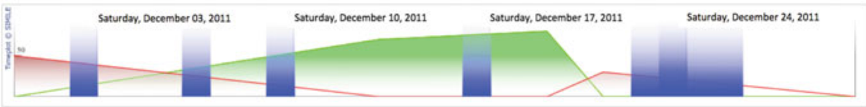

Fig. 6 The time plot with events and opinions related

positive or negative reviews in relation to a specific feature. Finally the system permits to select a profile, such as mature couple, young couple, business traveller, etc.

The selection of the specific parameters produce as result a timeline showing only the reviews that match the choices, as depicted in Fig. 5 where the reviews are filtered by time period and by the matching of the feature Breakfast.

Figure 6 shows a graph describing the time plot series. The graph points out the events in the selected time period and the opinions related to the feature of interest. It is possible in such a way to put in relation customer approval, special offers and various occurring events by dates.

Figures 5 and 6 are graphical representations of the data extracted by the reviews in the corpus and are realized by means of the SIMILE Timeline and Timeplot web widgets.

Figure 7 shows the bar and the radar graphs representing the features and the weights extracted from some reviews in the selected time period. Both the graphs represent the positive and negative valence related to each feature and the weights associated.

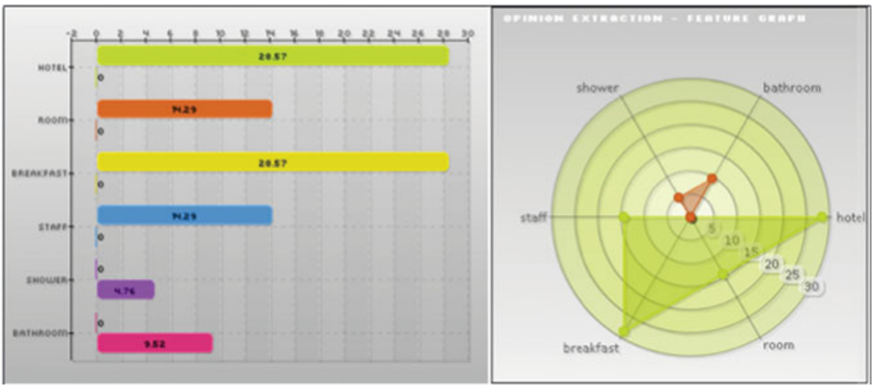

Fig. 7 The bar and the radar graphs of the features

4 Conclusions and Future Works

Online users talk across social media sites, on public forums and within customer feedback channels about products, services and their experiences, as well as their likes and dislikes.

The continuous monitoring of reviews is ever more important in order to identify leading topics and content categories and how those topics and categories are relevant to customers according to their habits. In this scenario, several independent vendors are proposing solutions in web and social media analytics, using their prior experience in business intelligence.

Although the proposed solutions come in some cases from leader industries skilled in business intelligence and in text analytics technologies, in most of the cases solutions do not provide valid approaches to the problems related to Opinion Mining.

More in details the state of the art in Opinion Mining is still away from have a definitive solution to the deep analysis of contents and from give a complete semantic interpretation of the meanings expressed in texts. In this context, the chapter proposes an Opinion Mining model to analyze and summarize reviews related to generic content categories. The process of analysis and summarization includes the extraction of features from the reviews people express about a product or a service and to determine the positive or negative valence of the reviews in relation to a specific feature. The proposed approach, taking advantage of several methods previously described, exploits the definition of FreeWordNet, a linguistic resource, an algorithm for the WSD and the generation of a matrix, that establishes the strength of the relations between features. A representation of data extracted and elaborated is showed as result of the Opinion Summarization step of the discussed Opinion Mining approach.

FreeWordNet is involved in the steps of distinction and identification of subjective, objective or factual sentences and contributes in a basic way in the task of features contextualization. The set of properties associated to synsets and the polarity values brings relevant benefit in the analysis of opinions.

The proposed model is valid for generic domains and is based on linguistic resources, such as WordNet and FreeWordNet not specialized for specific contexts.

Future works include the extension of WordNet evaluating the definition and the use of structured information about specific domains in the model.

References

1. Gartner (ed.): Gartner's 2012 Hype Cycle for Emerging Technologies Identifies "Tipping Point" Technologies That Will Unlock Long-Awaited Technology Scenarios. http://www.gartner.com/it/page.jsp?id=2124315 (2012)
2. Hexagon, C. (ed.): Listen, Understand, Act. How a listening platform provides actionable insight. www.crimsonhexagon.com/PDFs/Crimson_Hexagon_Listen_Understand_Feb_2009.pdf (2009)

3. Grimes, S., DeepMR: Market Research Mines Social Sentiment. http://www.greenbookblog.org/2012/05/01/deepmr-market-research-mines-social-sentiment/ (2012)
4. Jordan S.J.: 5 Dangers of DIY Research. http://www.researchplan.com/blog/?p=51 (2012)
5. Fernandez, J.: Conversations worth talking about. http://www.research-live.com/features/conversations-worth-talking-about/4007565.article (2012)
6. Ding, X., Liu, B., Yu, P.S.: A Holistic Lexicon-Based Approach to Opinion Mining. WSDM '08 Proceedings of the international conference on Web search and web data mining, pp. 231–240. ACM, New York (2008)
7. Akkaya, C., Wiebe, J., Mihalcea, R.: Subjectivity word sense disambiguation. In: Conference on Empirical Methods in Natural Language Processing, Singapore, pp. 190–199. The Association for Computational Linguistics (2008)
8. Rentoumi, V., Giannakopoulos, G.: Sentiment analysis of figurative language using a word sense disambiguation approach. In: International Conference on Recent Advances in Natural Language Processing (RANLP 2009), Borovets, Bulgaria, pp. 370–375. The Association for Computational Linguistics (2008)
9. Lee, D., Jeong, O.R., Lee, S.: Opinion mining of customer feedback data on the web. In: ICUIMC '08 Proceedings of the 2nd International Conference on Ubiquitous Information Management Communication, pp. 230–235 (2008)
10. Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C.: Red Opal: product-feature scoring from reviews. In: ACM Conference on Electronic Commerce, 2007, pp. 182–191 (2007)
11. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM Press (2004)
12. Zhai, Z., Liu, B., Xu, H., Jia, P.: Grouping Product Features Using Semi-Supervised Learning with Soft-Constraints. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010), Beijing, China, pp. 1272–1280 (2010)
13. Popescu, A., M., and Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing, pp. 339–346 (2005)
14. Kim, H.D., Ganesan, K., Sondhi, P., Zhai, C.X.: Comprehensive Review of Opinion Summarization. UIUC Technical Report, USA (2011)
15. Tuveri, F., Angioni, M.: A Linguistic Approach to Feature Extraction Based on a Lexical Database of the Properties of Adjectives and Adverbs, Global WordNet Conference (GWC2012), pp. 365–370. Matsue, Japan (2012)
16. Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., Padró, M.: FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), ELRA. Genoa, Italy, pp. 48–55. http://nlp.lsi.upc.edu/freeling (2006)
17. Vossen, P. (ed.): EuroWordNet: A Multilingual Database with Lexical Semantic Net-works. Kluwer Academic Publishers, Dordrecht (1998)
18. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: Proceedings of the International Conference on New Methods in Language Processing, pp. 44–49 (1994)
19. Leacock, C. and Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: Fellbaum, pp. 265–283 (1998)
20. Michelizzi, J.; Semantic relatedness applied to all words sense disambiguation. Thesis submitted to the University of Minnesota, Duluth. Major: Computer science (2005)