



## Feature-based opinion mining through ontologies



Isidro Peñalver-Martinez<sup>a</sup>, Francisco Garcia-Sanchez<sup>a</sup>, Rafael Valencia-Garcia<sup>a,\*</sup>,  
Miguel Ángel Rodríguez-García<sup>a</sup>, Valentín Moreno<sup>b</sup>, Anabel Fraga<sup>b</sup>, Jose Luis Sánchez-Cervantes<sup>b</sup>

<sup>a</sup>Departamento de Informática y Sistemas, Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain

<sup>b</sup>Computer Science Department, Universidad Carlos III de Madrid, Av. Universidad 30, Leganés, 28911 Madrid, Spain

### ARTICLE INFO

#### Keywords:

Opinion mining  
Ontology  
Sentiment analysis  
Feature extraction  
Part of speech tagging  
Polarity identification

### ABSTRACT

The idiosyncrasy of the Web has, in the last few years, been altered by Web 2.0 technologies and applications and the advent of the so-called Social Web. While users were merely information consumers in the traditional Web, they play a much more active role in the Social Web since they are now also data providers. The mass involved in the process of creating Web content has led many public and private organizations to focus their attention on analyzing this content in order to ascertain the general public's opinions as regards a number of topics. Given the current Web size and growth rate, automated techniques are essential if practical and scalable solutions are to be obtained. Opinion mining is a highly active research field that comprises natural language processing, computational linguistics and text analysis techniques with the aim of extracting various kinds of added-value and informational elements from users' opinions. However, current opinion mining approaches are hampered by a number of drawbacks such as the absence of semantic relations between concepts in feature search processes or the lack of advanced mathematical methods in sentiment analysis processes. In this paper we propose an innovative opinion mining methodology that takes advantage of new Semantic Web-guided solutions to enhance the results obtained with traditional natural language processing techniques and sentiment analysis processes. The main goals of the proposed methodology are: (1) to improve feature-based opinion mining by using ontologies at the feature selection stage, and (2) to provide a new vector analysis-based method for sentiment analysis. The methodology has been implemented and thoroughly tested in a real-world movie review-themed scenario, yielding very promising results when compared with other conventional approaches.

© 2014 Elsevier Ltd. All rights reserved.

### 1. Introduction

User participation is the primary value driver in Web 2.0 applications. Such a simple idea has had a tremendous impact on the way in which users interact with the Web. While on traditional Web 1.0 sites, companies published content and users were merely information consumers, in the Web 2.0 era users play a more active role in Web interactions, and have become not only consumers but also producers of information and media. In this new context, namely the Social Web, the number of online reviews in which people freely express their opinions with regard to a whole variety of topics is constantly increasing. The content that is produced as a direct consequence of this user participation on Websites such as Amazon

(<http://www.amazon.com>), booking (<http://www.booking.com>) or IMDB (<http://www.imdb.com>) is useful for both information providers and readers. Opinion mining is a recent sub-discipline at the crossroads of information retrieval and computational linguistics. The focus of opinion mining does not concern the topic that a text is about, but rather what opinion that text expresses (Esuli & Sebastiani, 2005). It determines whether the comments in online forums, blogs or comments relating to a particular topic (product, book, movie, etc.) are positive, negative or neutral. Opinions are very important when someone wishes to hear another's views before making a decision.

Although various approaches have been applied to opinion mining, two major methodologies can be distinguished. On the one hand, there are approaches for opinion mining that aim to classify entire documents as having a positive or negative polarity, while on the other there are segment level opinion analysis approaches whose aim is to distinguish sentimental from non-sentimental sections. Both methodologies have some drawbacks because an entire document or a single sentence could contain different opinions

\* Corresponding author. Tel.: +34 868888522; fax: +34 868884151.

E-mail addresses: [ipmartinez1@gmail.com](mailto:ipmartinez1@gmail.com) (I. Peñalver-Martinez), [frgarcia@um.es](mailto:frgarcia@um.es) (F. Garcia-Sanchez), [valencia@um.es](mailto:valencia@um.es) (R. Valencia-Garcia), [miguelangel.rodriguez@um.es](mailto:miguelangel.rodriguez@um.es) (M. Ángel Rodríguez-García), [vmpelayo@inf.uc3m.es](mailto:vmpelayo@inf.uc3m.es) (V. Moreno), [afraga@inf.uc3m.es](mailto:afraga@inf.uc3m.es) (A. Fraga), [jose Luis.s.cervantes@alumnos.uc3m.es](mailto:jose Luis.s.cervantes@alumnos.uc3m.es) (J.L. Sánchez-Cervantes).

about different features of the same product or service (Cambria, Schuller, Liu, Wang, & Havasi, 2013). In fact, classifying opinions at the document or sentence level does not indicate what the user likes and dislikes. A positive document on an object does not mean that the user has positive opinions on all aspects or features of that object. Likewise, it is impossible to ensure that a negative document signifies that the user dislikes everything about the object. In a document (e.g., a product review), the user typically writes about both the positive and negative aspects of the object, although the general sentiment toward that object may be positive or negative (Ahmad & Doja, 2012). To obtain such detailed aspects, it is necessary to perform feature-based opinion mining in an attempt to identify the features in the opinion and classify the sentiments of the opinion for each of these features (Feldman, 2013). The feature-based opinion mining of product reviews is a difficult task, owing to both the high semantic variability of the opinions expressed, and the diversity of the characteristics and sub-characteristics that describe the products and the multitude of opinion words used to depict them (Balahr & Montoyo, 2010).

In this context, new approaches based on both domain-dependent corpora and Semantic Web technologies for feature-based opinion mining have therefore appeared in the last few years (Cambria et al., 2013).

Semantic Web technologies are currently achieving a certain degree of maturity (Shadbolt, Berners-Lee, & Hall, 2006). The Semantic Web was conceived with the aim of adding semantics to the data published on the Web (i.e., establish the meaning of the data), thus allowing machines to be able to process these data in a way similar to that of humans (Berners-Lee & Hendler, 2001). Semantic technologies have been successfully applied in several natural language processing tasks such as information retrieval (Gladun, Rogushina, Valencia-García, & Béjar, 2013), knowledge acquisition from texts (Ochoa, Valencia-García, Perez-Soltero, & Barceló-Valenzuela, 2013) or semantic annotation (Rodríguez-García, Valencia-García, García-Sánchez, & Samper-Zapater, 2014). In this context, we believe that the already mature Semantic Web technology may be a valuable addition to traditional opinion mining approaches. More concretely, ontologies constitute the standard knowledge representation mechanism for the Semantic Web and can be used to structure information. The formal semantics underlying ontology languages enables the automatic processing of the information in ontologies and allows the use of semantic reasoners to infer new knowledge. In this work, an ontology is viewed as “a formal and explicit specification of a shared conceptualization” (Studer, Benjamins, & Fensel, 1998). Ontologies provide a formal, structured knowledge representation, with the advantage of their being reusable and shareable. They also provide a common vocabulary for a domain and define, with different levels of formality, the meaning of the terms’ attributes and the relations between them. In this work, the Web Ontology Language (OWL), the W3C standard used to represent ontologies in the Semantic Web, has been used to represent the concepts and features of the application domain. Knowledge in OWL ontologies is mainly formalized by using five kinds of components: classes, datatype properties, object properties, axioms and individuals.

However, most of the studies on opinion mining deal exclusively with English documents, perhaps owing to the lack of resources in other languages (Martín-Valdivia, Martínez-Cámara, Perea-Ortega, & Ureña-López, 2013). Since the Spanish language has a much more complex syntax than many other languages, and is currently the third most spoken language in the world, we are of the firm belief that the computerization of Internet domains in this language is of great importance.

The work presented herein is principally motivated by the need to develop new feature-based opinion mining approaches which leverage knowledge technologies that can be applied to different

languages (such as Spanish) and different domains while operating in an automated fashion. During the first stage of the proposed methodology, Semantic Web techniques are used to identify features from a domain ontology. Once these features have been identified, new polarity identification and opinion mining approaches are used to obtain an effective and efficient sentiment classification. The proposed approach has been validated with a use case scenario in the movie review domain.

The remainder of the paper is organized as follows: related work is expounded in Section 2; the proposed method is explained in Section 3; a validation of this ontology-guided approach in the movie review domain is performed in Section 4; and finally, conclusions and future work are put forward in Section 5.

## 2. Related work

Opinion mining (aka sentiment analysis or sentiment classification) is gaining momentum in the current Social Web landscape. Quite a number of researchers from both the information retrieval and computational linguistics fields have focused their investigations on this particular subject. Some of the most prominent research works on this topic are reviewed in this section.

Most current opinion mining approaches classify words into two categories (positive or negative) and provide an overall positive or negative score for the text. If a document contains more positive than negative terms, it is assumed to be a positive document, and is otherwise considered to be negative. These classification mechanisms are useful and improve the effectiveness of a sentiment classification but are not sufficiently precise to help determine what the opinion holder liked and disliked about each particular product feature/aspect.

There are also studies that not only determine the polarity of a text (positive or negative) but also indicate the level of polarity as being high/medium/low positive/negative, such as the work presented in Baccianella, Esuli, and Sebastiani (2009) in which several aspects of the problem in the product domain are explored, with special emphasis on how to generate vectorial representations of the text by means of POS tagging, sentiment analysis, and feature selection for ordinal regression learning.

Other approaches, such as that of (Gamon, 2005), make use of a complete set of statistical feature selection methods and also apply machine learning techniques. The experiments conducted by the aforementioned researchers demonstrate that machine learning techniques do not achieve an overall satisfactory performance level in sentiment classification.

The application of Semantic Web technology, and more concretely ontologies, to feature-based opinion mining has been introduced in some recent works (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013; Zhao & Li, 2009; Zhou & Chaovalit, 2008). The main goal of these works is to calculate the polarity by taking into consideration the features of a concept. In these works the authors propose that the domain ontologies which model the content of the texts in the corpus should first be built, and that the adjectives appertaining to the features in the texts from these ontologies should then be extracted. In Zhou and Chaovalit (2008), the polarity of a text is calculated on the basis of feature weights, and the value of each feature’s polarity is estimated using the “maximum likelihood estimate” approach. In Zhao and Li (2009), the authors obtain the text’s semantic orientation from the ontology hierarchy. The method proposed is also capable of obtaining the positive, negative and neutral orientation of the text.

Finally, several opinion mining methods have been proposed in the last few years. Most of the current opinion mining techniques are focused on either Chinese or English. Table 1 shows some of the most recent sentiment classification systems that have been

Download English Version:

<https://daneshyari.com/en/article/382905>

Download Persian Version:

<https://daneshyari.com/article/382905>

[Daneshyari.com](https://daneshyari.com)