Journal of
Information Technology & Software Engineering

Lazhar and Yamina, J Inform Tech Soft Engg 2012, 2:2
DOI: 10.4172/2165-7866.1000108

Research Article | Open Access

Identification of Opinions in Arabic Texts Using Ontologies

Farek Lazhar* and Tlili Guiassa Yamina

Computer Science Department, University of Annaba, Algeria

Abstract

A powerful tool to track opinions in forums, blogs, e-business sites, etc., has become essential for companies, politicians as well as for customers, and that because of the huge amount of texts available which make the manual exploration more and more difficult and useless. In this paper, we present our approach of identification of opinions based on an ontological exploration of texts. This approach aims to study the role of domain ontologies and their contributions in the identification phase. In our approach, domain ontology and sentiments lexicon are needed as pre-requirements.

Keywords: Identification; Opinion mining; Sentiment; Arabic; Ontologies; Texts; Subjectivity; Classification

Introduction

The views available on the internet have a significant impact on users, for example, if users have already researched opinions on a product, they are willing to pay more for a product whose opinion is more favourable than another, and the product will be more marketed than another whose opinion is less favourable [1].

Companies, politicians, and customers need a powerful tool to track opinions, sentiments, judgments, and beliefs that people can express in blogs, comments, or in the form of texts, toward a product, a service, a person or an organization, etc. [2].

In opinion mining area, the use of expressions as a "bag of sentiment words" to detect the semantic orientation of the overall content of a text needs to give values to those expressions as positive, negative or neutral towards a given topic [3].

Generally, research works in this area can be grouped into four main categories:

- Development of linguistic and cognitive models for opinion mining where all approaches based on dictionary or corpus are used automatically or semi-automatically to extract opinions based on the semantic orientations of words and phrases [4];
- Opinions extraction from texts, where all the local opinions are aggregated to determine the overall orientation of a text [4-6];
- Features based opinion mining, where all the opinions expressed towards the characteristics of a product or an object are extracted and summarized [7-9].

  This article focuses on identification and classification of opinions in Arabic texts, which aims to calculate the semantic orientation of the entire content of a text as positive or negative toward a subject or an object from the subjective expressions carrying the semantic orientations of the different features, but the key questions that we should ask are:
- How to get this set of features?
- What features are related to each other?
- What model of knowledge representation to be used to produce an understandable summary for the studied domain?

To answer these questions, we propose in this paper to study the role of ontologies used in opinion mining, and more specifically, our goal is to study how domain ontology can be used to:

- Structure the features;
- Extract explicit and implicit features from the texts;
- Produce summaries based on reviews and user comments.

The paper is organized as follows: We present in Section 2, state of the art of the main approaches used in the field and the motivations of our work. We present in the next section, our approach and the general architecture of opinions identification process.

State of the Art

Related work

Overall, two main types of work are distinguished, those that are based on simple features extraction from the texts, and those who organize features into a hierarchy using taxonomies or ontologies. The extraction process mainly concerns explicit features. We can distinguish two main families:

Opinion mining without knowledge representation models: All approaches that do not use knowledge representation models are based on the use of algorithms to discover the different characteristics of a product or an object. Only the expressions of opinions (adjectival and adverbial) are extracted, then a summary is produced to show for each characteristic, the positive and the negative opinions and the total number of these categories [4,8].

The main limitation of these approaches is that there is a large number of extracted features and a lack of organization. In addition, similar concepts are not grouped

(For example, in some domains, the words "appointment" and "rendezvous" which have the same meaning "appointment"), and possible relationships between the features of an object are not recognized (example: "coffee" is a specific term of "drink"). Thus, analysis of polarity (positive, negative or neutral) of the text is done by assigning the dominant polarity of opinion words, regardless of the polarities associated with each feature individually [10].

***Corresponding author:** Farek Lazhar, Computer Science Department, University of Annaba, Algeria, E-mail: lazhar3000@hotmail.com

Received February 13, 2012; **Accepted** March 27, 2012; **Published** March 29, 2012

Citation: Lazhar F, Yamina TG (2012) Identification of Opinions in Arabic Texts Using Ontologies. J Inform Tech Soft Engg 2:e108. doi:10.4172/2165-7866.1000108

Copyright: © 2012 Lazhar F, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Opinion mining with knowledge representation models: The family itself can be divided into two subfamilies:

I. Use of taxonomies

This kind of approaches does not seek a list of features, but rather a hierarchical organized list by the use of taxonomies. We recall that a taxonomy is a list of terms organized hierarchically through a sort of "is a kind of". In [7] the author use pre-defined taxonomies and semantic similarity measures to automatically extract the features and calculate the distances between concepts.

Generally, the use of taxonomies is coupled with a classification technique; the sentences corresponding to the leaves of the taxonomy are extracted. At the end of the process, a summary that can be more or less detailed is produced.

II. Use of ontologies

These approaches aim to organize the features using elaborated representation models. Unlike taxonomies, ontology is not restricted to a hierarchical relationship between concepts, but can describe other types of paradigmatic relations such as synonymy, or more complex relationships such as relations of composition or spa-tial relationships.

Generally, the extracted features correspond exclusively to terms contained in the ontology. The feature extraction phase is guided by a domain ontology, built manually [11], or semi-automatically [9,12], which is then enriched by a process of automatic extraction of terms, corresponding to new features identification.

Similar features are grouped together using semantic similarity measures.

Ontologies have also been used to support polarity mining. For example, in [13], the authors manually built ontology for movie reviews and incorporated it in the polarity classification task which substantially improved the performance of their approach.

Ontology based opinion mining

In [2], the use of a hierarchy of features improves the performance of features based identification systems. However, works using domain ontologies exploit the ontology as a taxonomy using only "is a" relations between concepts. They do not really use all data stored in ontology, such as the lexical components and other types of relationships. We believe that we can get several advantages in the domain of opinion mining by the full use of domain ontology capabilities:

- Structuring of features: Ontologies are tools that provide a lot of semantic information. They help to define concepts, relationships, and entities that describe a domain with an unlimited number of terms;

- Extraction of features: Relationship between concepts and lexical information can be used to extract explicit an implicit features.

Our Approach

Description

For each studied domain, our approach requires three basic elements:

- A domain ontology O, where each concept and each property is associated to a set of labels that correspond to their semantics;

- A lexical resource L of opinion expressions;

- A set of texts T as comments and views.

Based on the conceptual model described in [3], and on the definition described in [14] which define an elementary discourse unit (EDU) as a clause containing at least an elementary opinion unit (EOU) or a sequence of clauses that address a rhetorical relation to a segment expressing an opinion. Note that an EOU is an explicit opinion expression composed of an explicit noun, an adjective or a verb with its possible modifiers (negation and adverbs).

In a review, the opinion holder comments a set of features of an object or a product using opinion expressions. Each feature corresponds to a concept or a property in the ontology O.

For each extracted EDU, the system:

- Extracts EOUs using an approach based on rules;

- Extracts features that correspond to the process of terms extraction using the do-main ontology;

- Associates, for each feature within the EDU, the set of opinion expressions;

We detail below, these steps:

Extraction of elementary opinion units: Nouns, adjectives or verbs may be associated with certain modifiers such as words of negation and adverbs. For example, "excellent", "not good" are EOUs (Figure 1).

For example in the following comment, the EDUs are between square brackets, the EOUs are underlined, and the characteristics of the object are in bold. There is an inverse relationship between the EDUa and the EDUb, representing the review ex-pressed in the EDUd.

Features extraction: This step aims to extract for the comment all the labels of the ontology. As each concept is an explicit feature, we simply project the lexical components of the ontology on the text to obtain, for each EDU, all the features. To extract the implicit features, ontology properties are used. We recall that these properties are to define the relationships between concepts of the ontology. For example, the property "drive" links the concepts "conductor" and "car".

Linking opinions expressions with extracted features: In this step, extracted opinions expressions in step (a) have to be linked to the features extracted in step (b), i.e. we should associate with each EDUi the set of pairs (fi, OEi). During this step, we distinguish the following cases:

I. Known opinionated features and known opinions expressions: In this case, opinionated features match to the used opinions expressions. For example, if our lexicon contains the concept "nature", and sentiments lexicon contains the word "amazing", from the EDU "amazing nature", it is easy to extract the couple (nature, amazing) from the text.

II. Known opinionated features and unknown opinion expressions: Expressions, as in the EDU "acceptable result", where the

[Yesterday, I purchased a **phone**]$_{a,}$ [Even if the **phone** is <u>excellent</u>]$_{b,}$ [the **design** is <u>very basic</u>]$_{c,}$ [which is <u>disappointing</u> in this mark]$_{d.}$

Figure 1: Example showing EOUs Extraction.

Citation: Lazhar F, Yamina TG (2012) Identification of Opinions in Arabic Texts Using Ontologies. J Inform Tech Soft Engg 2:e108. doi:10.4172/2165-7866.1000108

opinion word "acceptable" was not extracted in step (a) (see section 3.1). In this case, the lexicon of opinions can be automatically updated with the recovered opinion word.

III. Unknown opinionated features and unknown opinion expressions: As in the EDU "wonderful rainforest" where the feature "rainforest" has not been extracted in step (b) (see section 3.1), in this case, the domain ontology can be updated by adding a new concept or a new property in the right place.

IV. Opinion expressions only: As in the EDU "It's slow". This kind of EDU expresses an implicit feature. In this case, we use the ontology properties to retrieve the associated concept in the ontology

V. Features only: An EDU with features alone can also be an indicator of the presence of an implicit opinion expression towards the feature as in "the park became a haven for perverts", which express a negative opinion towards "the park".

Architecture of our approach

In this section, we present the general architecture of our approach and the different modules constituting our system (Figure 2).

As indicated in the last figure, our system contains the following modules:

Texts edus segmentation: Generally, extraction of elementary discourse units (EDUs), depends on the use of delimiters such as ".", "," "?" "!";

Eous extracting: Elementary opinions units EOUs and semantic orientations are usually extracted using a lexicon of emotions specific to domain of study;

Features extraction: Features can be extracted by a simple projection of the ontology on the elementary discourse units (EDUs);

Associating eous to Features: Each extracted feature should be associated to one or more elementary opinions units in order to extract its semantic orientation;

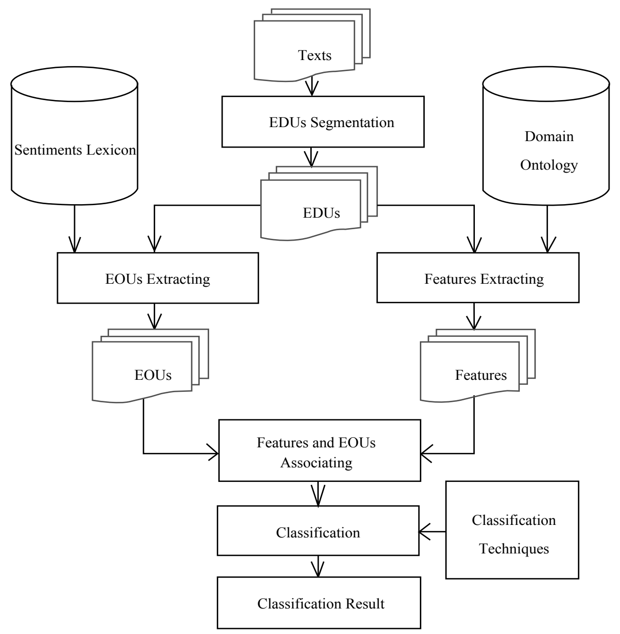

Figure 2: General architecture of our approach.

Classification: The last phase of our work is to classify the identified opinions into positive or negative classes using supervised classification techniques.

Conclusion

In this paper we presented our approach based on an ontological exploration of Arabic texts. Our method is promising because the use of ontologies improves the ex-traction of features and facilitates the association between opinions expressions and opinionated features of the object. On the one hand, domain ontology is useful within its list of concepts which carry much semantic data in the system. The use of ontology concepts labels can recognize terms that refers to the same concepts and provides a hierarchy between these concepts. On the other hand, ontology is useful to its list of properties between concepts that can recognize the opinions expressed on the implicit features.

References

1. Gilot S (2010) Fouille d'opinions, Rapport de stage.
2. Cadihac A (2010) Ontolexical resources for feature based opinion mining: a case study. Beijing.
3. Lazhar F (2009) Identification d'opinions dans les textes arabes, IC.
4. Turney PD, Littman ML (2002) Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus. National Research Council, Institute for Information Technology, Technical Report ERB-1094, (NRC-44929).
5. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment Classification using Machine Learning Techniques. Proceedings of the conference on EMNLP.
6. Kim SM, Hovy E (2006) Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In Proceedings of Workshop on Sentiment and Subjectivity in Text. Sydney, Australia.
7. Carenini G, Ng RT, Ed Z (2005) Extracting Knowledge from evaluative text. In Proceedings of the 3rd international conference on Knowledge capture.
8. Hu M, Liu B (2004) Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.
9. Cheng X, Xu F (2008) Fine-grained Opinion Topic and Polarity Identification. In Proceedings of the Sixth International Language Resources and Evaluation (LREC' 08), Marrakech, Morocco.
10. Lazhar F (2009) Identification d'opinions dans les textes arabes, IC.
11. Zhao L, Li C (2009) Ontology Based Opinion Mining for Movie Reviews. In Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management.
12. Feiguina O (2006) Résumé automatique des commentaires de Consommateurs. Mémoire pré-senté à la Faculté des études supérieures en vue de l'obtention du grade de M.Sc. en infor-matique, Département d'informatique et de recherche opérationnelle, Université de Mon-tréal.
13. Chaovalit P, Zhou L (2005) Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches. HICSS '05 Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) 4.
14. Asher N, Lascarides A (2003) Logics of Conversation. The press Syndicate of the University of Cambridge, Cambridge, United Kingdom. University Press.
15. Pak A, Paroubek P (2011) Classification en polarité de sentiments avec une représentation textuelle à base de sous-graphes d'arbres de dépendances, TALN 2011,Montpellier, 27 juin – 1er juillet.