JID:IJA AID:8141 /FLA [m3G; v1.224; Prn:6/11/2017; 16:12] P.1 (1-15)
International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A fuzzy-based strategy for multi-domain sentiment analysis ☆

Mauro Dragoni [a], Giulio Petrucci [a,b]

[a] Fondazione Bruno Kessler, Trento, Italy
[b] University of Trento, Trento, Italy

ARTICLE INFO

Article history:
Received 16 March 2017
Received in revised form 20 October 2017
Accepted 24 October 2017
Available online xxxx

Keywords:
Sentiment analysis
Fuzzy logic
Multi-domain environment

ABSTRACT

Multi-domain sentiment analysis aims to compute text polarity by leveraging on information extracted from domain-specific models. In this scenario, one of the main challenge to solve is to infer the polarity of texts belonging to domains which are different from those used for building the opinion model. Indeed, a common issue of several approaches presented in the literature is their poor capability in classifying such documents. This paper presents an approach that takes advantage of the possible conceptual domains' overlaps for build general models able to compute the polarity of texts belonging to arbitrary domains. Given the uncertain flavor of the information managed by the system, fuzzy logic is used for representing the polarity learned from either training sets or a training set. Such learned information is integrated with further conceptual knowledge gathered from two resources widely used in the sentiment analysis field, namely the SenticNet and the General Inquirer vocabulary. The presented approach is then validated by experiments that follow the DRANZIERA protocol and the results demonstrate the effectiveness of the presented strategy and set a starting point for further investigations.

1. Introduction

Sentiment analysis [1] is a natural language processing (NLP) task which aims at classifying documents according to the opinion they express about a given object [2]. Generally speaking, the goal of the sentiment analysis is to determine the attitude of a subject with respect to the overall tonality of the content he/she produces. Recently, the usage of the World Wide Web for marketing purposes has gained more and more importance, fostering the widespread of sentiment analysis based approaches in processing user generated content. Examples of contexts in which sentiment analysis have been applied are product reviews, opinion about political or sport events, and so on.

Liu and Zang [3] formalize the opinion mining problem as follows. An "opinion" or "sentiment" may be represented as a tuple of five elements:

$$\langle o_x, f_{xy}, so_{zxyv}, h_z, t_v \rangle \tag{1}$$

where o_x is the object targeted by the opinion, f_{xy} is an object's feature (also called "aspect"), h_z is the subject providing the opinion, generally called "opinion holder", and t_v is a timestamp annotation indicating when the opinion has been expressed. Finally, so_{zxyv} is the judgment (hereafter called "polarity") that the opinion holder h_z provided about the feature

☆ This paper is part of the Virtual special issue on Uncertainty Reasoning for the Web, Edited by Fernando Bobillo, Kenneth J. Laskey, Trevor Martin, Matthias Nickles.

E-mail addresses: dragoni@fbk.eu (M. Dragoni), petrucci@fbk.eu (G. Petrucci).

https://doi.org/10.1016/j.ijar.2017.10.021
0888-613X/

Please cite this article in press as: M. Dragoni, G. Petrucci, A fuzzy-based strategy for multi-domain sentiment analysis, Int. J. Approx. Reason. (2017), https://doi.org/10.1016/j.ijar.2017.10.021

f_{xy} at time t_v. The polarity value so_{zxyv} can be of three general types: positive, negative, or neutral. However, a more fine-grained representation can be provided depending on the level of granularity that it has to be implemented into the system.

The sentiment analysis problem has been mostly investigated without considering the domain which a document belongs to. With the term "domain", we refer to the definition provided by [4] considering a "domain" as a "different category of product entities managed by the classifier".

The aspect of not considering domains can penalize the overall effectiveness of the algorithms because they would be unable to integrate, into the sentiment models, information concerning the different domain-based contexts from which each concept have been extracted. This challenge can be better summarized through the examples provided below. In the first example, we present two sentences where the adjective "hot" is used in different "emotion-based" contexts:

1. I think that you should investigate on this topic in the next months, it is definitely "hot".
2. This drink is too "hot", it is not the best choice today.

In the second example, instead, we consider two sentences, describing different products, where the adjective "big" has been used in two different "subjective-based" contexts:

1. This smartphone is too "big" and I am not able to put it in my pocket.
2. The trunk of my new car is very "big". Finally, I will have no problems when I will travel with my family.

The first couple of sentences present two "emotional" contexts where the adjective "hot" has been used with different polarities: a positive polarity when the opinion holder of the first sentence refers to the topic on which a person should work, and a negative polarity when the opinion holder of the second sentence refers to the drink.

A different scenario is instead presented in the second couple of sentences, here, the two sentences share the adjective "big" referring to two different domains: "Electronics" and "Automotive". The polarity of the adjective "big" is "negative" in the first sentence, because it is highlighted the issue about having a big smartphone. While, in the second sentence, the adjective "big" is used with a positive polarity by acclaiming the big trunk size as a solution to a specific problem.

From the examples provided above, it can be observed how the analysis of information coming from different domains has to be properly analyzed with approaches able to be effective and general as well. The investigation on sentiment analysis by considering the multi-domain environment started recently [4]. The approaches presented in the literature, and discussed in Section 2, leverage on transferring information of the created models across domain pairs [5] in a so-called transfer learning fashion. These strategies demonstrated of being effective when they transfer polarity information from specific pairs of domains, i.e. the ones used for building the sentiment model. However, there are a couple of issues that prevent their generalization: (i) every time a new domain has to be analyzed, a new transfer model has to be designed for preserving system effectiveness, and (ii) the transfer capability of the system is limited to the domains included into the built model.

The contribution presented in this paper aims at addressing the issue of working with information sources coming from different domains. The described approach exploits linguistic overlaps between domains for building models. These models permit the inference of polarities for documents belonging to domains that have not been used for training the model integrated within the inference engine. The originality of the approach presented in this paper, with respect to the transfer learning paradigm, consists of having a data-provenance-independent system for building the model, i.e. the domains. This aspect makes the presented strategy innovative in the field of multi-domain sentiment analysis.

The article is structured as follows. Section 2 reviews the sentiment analysis literature in both the single and multi-domain contexts. Section 3 presents a brief introduction of the fuzzy set theory, where fuzzy sets have been used for representing polarity values, and introduces the resources integrated into the approach presented in Section 4. In Section 5, we provide the evaluation of the system performed by following the "DRANZIERA" [6] protocol. Finally, Section 6 concludes the paper.

2. Related work

The topic of sentiment analysis has been extensively studied in the literature. Examples are the work of Liu and Zang [3], Cambria [7], and Dashtipour [8]. In these works, several techniques have been presented and validated.

Sentiment analysis strategies applied in a multi-domain environment have been discussed from different points of view. Approaches presented in the literature can be classified in two main categories: (i) transferring of learned sentiment models across different domains [4,5,9,10], and (ii) labels propagation through graph structures [11–15].

These approaches, applied in a multi-domain environment, demonstrated their effectiveness. However, they suffer from limitations when they are used in domains different from those used for building the sentiment model. Differently from the works cited above, our approach aims to build sentiment models able to infer the polarity of documents coming from other domains. This aspect supports the use of built sentiment models in scenarios where the domain of the document content is unknown.