Engineering Applications of Artificial Intelligence 47 (2016) 122–139

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Jointly identifying opinion mining elements and fuzzy measurement of opinion intensity to analyze product features

Haiqing Zhang [a,*], Aicha Sekhari [a], Yacine Ouzrout [a], Abdelaziz Bouras [a,b]

[a] DISP Laboratory, University Lumière Lyon 2, 160 Bd de l'Université, Bron Cedex, Lyon 69676, France
[b] Computer Science Department—Qatar University, ictQATAR, Box 2731, Doha, Qatar

ARTICLE INFO

Available online 29 June 2015

Keywords:
Opinion mining
Dependency relations
Fuzzy sets and logic
Opinion degree intensifiers
Feature-by-feature analysis

ABSTRACT

Opinion mining mainly involves three elements: feature and feature-of relations, opinion expressions and the related opinion attributes (e.g. Polarity), and feature–opinion relations. Although many works have emerged to achieve its aim of gaining information, the previous researches typically handled each of the three elements in isolation, which cannot give sufficient information extraction results; hence, the complexity and the running time of information extraction is increased. In this paper, we propose an opinion mining extraction algorithm to jointly discover the main opinion mining elements. Specifically, the algorithm automatically builds kernels to combine closely related words into new terms from word level to phrase level based on dependency relations; and we ensure the accuracy of opinion expressions and polarity based on: fuzzy measurements, opinion degree intensifiers, and opinion patterns. The 3458 analyzed reviews show that the proposed algorithm can effectively identify the main elements simultaneously and outperform the baseline methods. The proposed algorithm is used to analyze the features among heterogeneous products in the same category. The feature-by-feature comparison can help to select the weaker features and recommend the correct specifications from the beginning life of a product. From this comparison, some interesting observations are revealed. For example, the negative polarity of video dimension is higher than the product usability dimension for a product. Yet, enhancing the dimension of product usability can more effectively improve the product.

1. Introduction

The widely used Web communication on mobile and web-based technologies has dramatically changed the way individuals and communities express their opinions. More and more reviews are posted online to describe customers' opinions on various types of products. These reviews are fundamental pieces of information needed to support both firms and customers to make good decisions. The features and attributes of a product extracted from online customer reviews can be used in recognizing the strengths and weaknesses of the heterogeneous products for firms. While customers do not always have the ability to wisely choose among a variety of products in the market, they commonly seek product information from online reviews before purchasing a new product.

However, the number of reviews grows rapidly, so it becomes impractical to analyze them by hand. In addition, the inherent characteristics of the reviews are diverse and complex. Firms tend to portray the products in different ways, which makes the products more distinguishable and prevents the products from being substituted for each other easily. The heterogeneous products in the same category have slightly different functions, features, and physical characteristics. By doing this, more products become competitive; price alone is not the most important factor to be successful in competing products any more. This internal rule of releasing products causes the number of reviews to increase fast. If the related opinions towards features can be obtained from the massive reviews, the firms will greatly benefit by using the extracted information to evaluate how and where to improve the product through the product development process. Hence, extracting information from the online reviews is academically challenging and has practical use.

Identifying the opinions in a large-scale document of customer reviews is an opinion mining issue, which is a sub-division of information extraction that is concerned with: the features, with the opinion it expresses, and the relations between features and opinion expressions. An opinion is a positive or negative sentiment or attitude about an entity or an aspect of the entity from an opinion holder. An opinion is defined as a quintuple in Liu and Zhang (2012). We extend the quintuple into a sextuple by adding the relations among features and opinions, which is shown as (e_i, f_{ij}, oo_{ijkl}, r_{ijkl}, h_k, t_l), where e_i is the name of an entity; f_{ij} is a feature of e_i; oo_{ijkl} is the opinion expression on feature f_{ij} of entity e_i; r_{ijkl} is the sets of feature–opinion relation

* Corresponding author.
E-mail addresses: haiqing.zhang.zhq@gmail.com (H. Zhang), aicha.sekhari@univ-lyon2.fr (A. Sekhari), yacine.ouzrout@univ-lyon2.fr (Y. Ouzrout), abdelaziz.bouras@qu.edu.qa (A. Bouras).

http://dx.doi.org/10.1016/j.engappai.2015.06.007
0952-1976/

extraction, feature–feature relation extraction, and opinion–opinion relation extraction; h_k is the opinion holder; and t_l is the time when the opinion is expressed by h_k. This definition can provide a basis for transforming unstructured text to structured data. For example, the review presented in Fig. 1 has six constitutes: product name (e_i), product features (f_{ij}), product opinions (oo_{ijkl}), relations (r_{ijkl}), opinion holder (h_k), and opinion post time (t_l). The sextuple resolves the unstructured text data into a formalized structured text. The added attribute r_{ijkl} can be used to summarize the overall attitude of the whole review and reflect the opinions with respect to a specific feature. For instance, in the following Fig. 1, the information "Ronald J. Magdos (opinion holder) has only "good" (product opinions) things to say about "Canon Powershot SX510HS's" (entity) in regards to the "photos" (product features) since he has discovered this product ⟨photos good⟩(relations) on "October 9, 2014" (post time)" will obtain parameters e, f, h, t, oo, and r based on the indices $i=1$, $j=1$, $k=1$, and $l=1$.

In sextuple, the features f_{ij}, opinion expression oo_{ijkl}, and opinion description r_{ijkl} are the necessary pieces of information which are difficult to obtain. Two fundamental problems of mining such information are opinion features extraction and opinion words locating.

1) Opinion features are characteristics of the products in which the opinion has been described. Two issues are generated in product feature extraction.
   a) *One is that synonyms are often occurring in extraction of features.* For example, 'image' and 'photo' refers to the same product feature in camera reviews.
   b) *The other one is some product features are combined by several nouns*. The obtained 'wifi feature' in Fig. 1 is an example. Hence, feature-of relation is used to record the synonyms of features and rebuild the noun terms to more accurately represent product features.
2) Opinion expressions are the opinion words that the reviewers have adopted to describe their opinions on the related features.
   a) *Opinion expressions are commonly composed by an opinion pattern involving adjectives, adverbs, and verbs instead of a single opinion word.* Thus, opinion-of relation extraction is adopted to keep the opinion patterns. For example, 'simple to use', 'very user friendly', and 'easy to use' in Fig. 1.
   b) *Opinion expressions also need to express the evaluation for correct targets.* For example, "I had a nice trip to Yellowstone. I took some very nice pictures and videos by using this camera", "nice" describes the 'trip', but "trip" is not the feature of a digital camera. Therefore, "nice" does not need to be extracted in this case.
   c) The feature–opinion relation extraction is necessary to be proposed to express the opinion expressions corresponding with the related opinion features.

An interesting observed phenomenon from the obtained sextuple is that the degree of success about the heterogeneous products in the same category is vastly different. Some features are key impact factors and have a great influence on decisions for purchasing (customers), as well as product development strategies (firms). Reducing the weakness of important features can increase consumption quantities and enhance the product's reputation effectively. A straightforward solution for important, but weaker, feature identification is to select all of the important features that have the most negative comments as the weak features. However, customers' opinions on the worst features may not have the greatest influence on their overall opinions, and thus will not influence the general consumption quantities. Moreover, a feature's weakness may affect other features, causing negative comments, emphasis, and improvements to be focused on the features that are not the root cause of the issue. Thus the worst features could lead to firms spending more effort into maintaining features and cannot gain the maximum benefits from them. Hence, a comprehensive analysis strategy for feature-by-feature comparison has to be able to identify the important features that are currently weak.

This work identifies the strengths and weaknesses of heterogeneous products in the same category; selects the most successful type as the benchmark product; then recommends improvements for the appropriate weaker features of the weaker products. The aim is to make more heterogeneous products more successful during the

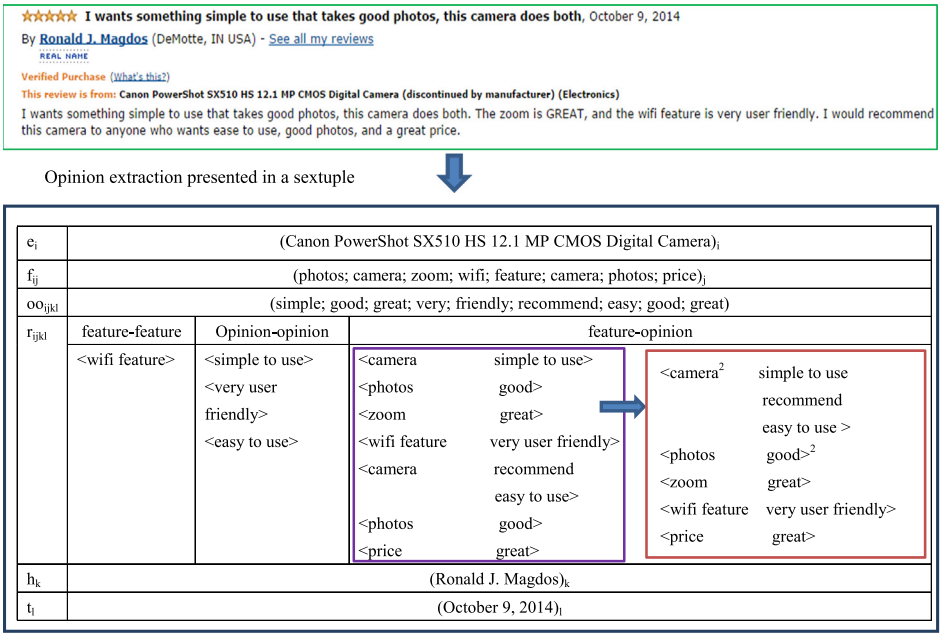

Fig. 1. Extracted opinions and features showing in a sextuple.