An Aspect-Lexicon Creation and Evaluation Tool for Sentiment Analysis Researchers*

Mus'ab Husaini[1], Ahmet Koçyiğit[1], Dilek Tapucu[1,2], Berrin Yanikoglu[1], and Yücel Saygın[1]

[1] Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Turkey
{musabhusaini,akocyigit,dilektapucu,berrin,ysaygin}@sabanciuniv.edu
[2] Dept. of Computer Engineering, Izmir Institute of Technology, Izmir, Turkey

Abstract. In this demo paper, we present SARE, a modular and extendable semi-automatic system that *1)* assists researchers in building gold-standard lexicons and evaluating their lexicon extraction algorithms; and *2)* provides a general and extendable sentiment analysis environment to help researchers analyze the behavior and errors of a core sentiment analysis engine using a particular lexicon.

1 Introduction

In sentiment analysis, domain aspect (also called feature) extraction is crucial for gaining a deeper and more refined understanding of the opinions expressed in a given document [3, 2]. Without domain-specific aspects, the sentiment analysis process remains prone to generalizations and dilution of opinions. A *domain aspect lexicon* consists of a set of *aspects* that are broad features of the domain; and for each aspect, a set of aspect-related *keywords*. For example, in the *hotel* domain, "room quality" might be one such aspect and the terms "furniture" and "size" could be keywords associated with this aspect.

Several automatic and semi-automatic methods have been proposed in the literature to extract a domain aspect lexicon from a given domain corpus, such as those cited in [5]. In evaluating their methods, researchers either compare the coverage of the extracted lexicon to that of a hand-built one considered to be the gold standard; or they compare the performance of a baseline sentiment analysis system using the generated lexicon versus some other available lexica. The gold-standard lexicon mentioned in the former case is obtained through one of the following ways: *a)* by manually tagging words from a domain corpus; *b)* by one or more domain experts choosing aspects and keywords without the use of a corpus; or *c)* using review sets that have already been annotated with aspects and keywords by the original reviewers. The first approach is naturally rather tedious as domain corpora are usually too large to be manually processed. The second one is vulnerable to generalization error since the experts' vocabulary tends to be narrower compared to the broader vocabulary of a mass of reviewers. Finally,

---

the third approach is not always possible, since such review sets are not available in all cases. It is also difficult to verify and evaluate a hand-built lexicon to make sure that it contains all the relevant words and only the relevant words. There are no tools freely available, to the best of our knowledge, that deal with all of these challenges and assist the researcher with these tasks. We developed *SARE* (Sentiment Analysis Research Environment) to fulfill this need.

2   Overview and Main Contributions

In this demo paper, we present SARE, which is designed to support researchers in constructing and analyzing their sentiment analysis systems addressing the issues raised in Sect. 1. Currently, SARE consists of two modules: *SARE–Lex* (Lexicon Creation Tool) provides a semi-automatic method for developing gold-standard domain lexica by reducing the size of the corpus that needs to be processed manually. *SARE–Core* (Core System Results), takes a review corpus, a sentiment analysis engine, and a domain lexicon; and displays aspect-based as well as overall sentiment summary for each review in the corpus. The second module can take as input a lexicon generated by the first or a pre-generated lexicon obtained as output from a separate feature-extraction algorithm. In either case, the goal is to help the researcher analyze the behavior and errors of a core sentiment analysis engine using a particular lexicon. Thus, while SARE is intended to be a general sentiment analysis research environment, its two modules can be used in creating a hand-built lexicon and assessing the performance of any given lexicon.

Our main contribution is a publicly-available and open-source tool for producing gold-standard domain aspect lexica. This tool is primarily intended for use by sentiment analysis researchers, and to a certain extent, practitioners. For demonstrating SARE, we will present an interactive scenario for creating a lexicon from scratch given a corpus; then we will show the capabilities of the system for analyzing the output of the sentiment analysis module. Our future work involves extending SARE, which is designed to be highly modular and extendible, to support research in various sentiment analysis sub-problems by introducing pluggable components such as custom sentiment analysis engines and polarity lexica. In this work, we only deal with features that are explicitly expressed using nouns such as the ones mentioned above; the task of extracting implicit features such as those expressed using adjectives and adverbs is left for future work. We also plan on extending the system to add support for languages other than English which only requires substituting the NLP process and polarity lexicon.

3   Process and Modules

SARE is a self-contained web application that can be deployed to any Java-based web server. The system allows for a domain corpus to be imported in a variety of formats including text and XML. This corpus is then analyzed using the method described in Sect. 3.1 to obtain the most informative documents from which the user extracts domain aspects and related keywords to
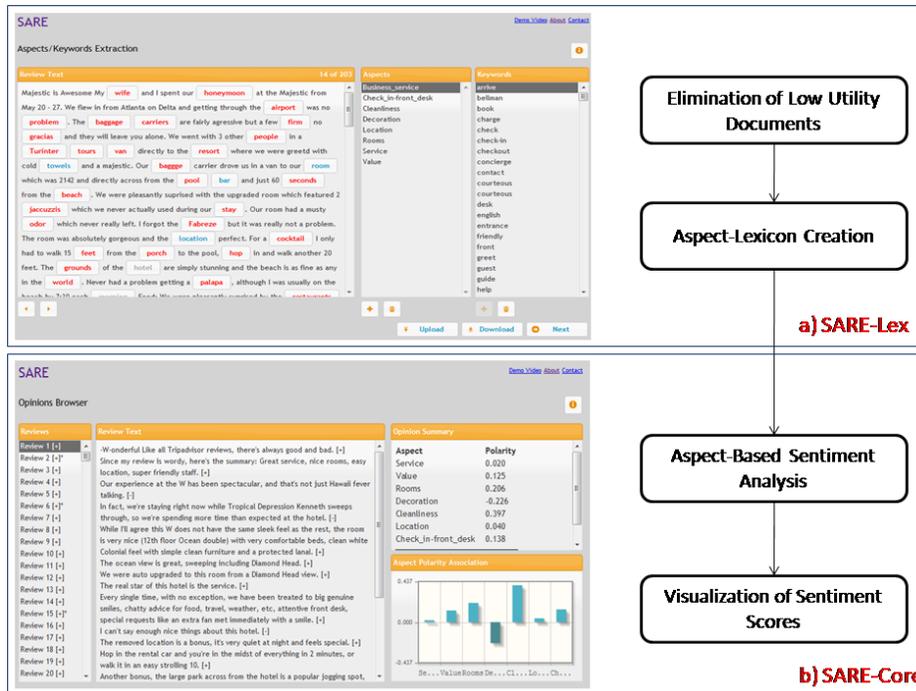
Fig. 1. SARE: (a) SARE-Lex module (b) SARE-Core module

create the aspect lexicon. Alternatively, a pre-generated lexicon can be uploaded by the user as well. The lexicon, manually extracted or otherwise, is then used by the sentiment analysis engine to process the corpus and results are displayed to the user for further analysis. Figure 1 illustrates this workflow with the help of screenshots. The system currently uses a fixed sentiment analysis engine and polarity lexicon, but will be extended to accept pluggable versions in the near future. This application is publicly accessible online by visiting `http://ferrari.sabanciuniv.edu/sare`, and a demo video explaining its usage can be downloaded from `http://ferrari.sabanciuniv.edu:81/public/videos/sare-demo.wmv`. It should be noted that SARE is developed on the principles of REST architecture, which allows other systems to easily access it, use its output independent of the website, and extend it as desired. The list of necessary API calls for this purpose can be provided on demand. In the following subsections, we will discuss the major components of the system.

3.1 SARE-Lex: Lexicon Extraction Module

This module deals with the problem of aspect lexicon extraction by corpus summarization and user annotation. We approximate aspect keywords with corpus nouns and apply a variation of the *Greedy Set Cover* algorithm that we developed called *Eagerly Greedy Set Cover* algorithm to find the minimum set of

documents that cover all of the nouns in the corpus. Conceptually, each noun is assigned a utility score equal to its corpus frequency, and the utility score of each document is calculated as the cumulative utility score of the nouns it contains. However, since we have formulated this as a set cover problem, a given noun can only lend its score to one document, and therefore we find the best placement for each noun where it maximizes document utility. For large enough corpora, this generally results in a utility distribution where most of the documents have very low utility score. By ignoring documents with utilities in the lower percentile (a percentile threshold), the number of documents requiring annotation can be significantly reduced. This threshold is a parameter of the module and the tool assists the user in selecting it by displaying a utility distribution. The reduced set of documents is then displayed sequentially to the user for collecting new aspects and associated keywords. As each document is displayed, the interface emphasizes document nouns for better visibility. The user can then interact with these emphasized nouns to mark them as aspects or keywords of an aspect.

3.2 SARE-Core: Sentiment Analysis Module

The objective of this module is to use the domain aspect lexicon from the previous module for calculating aspect-based and overall sentiment scores for each review, and present a summarized result. If reviews in the provided corpus were labeled as positive or negative, then the interface also indicates erroneous classifications for each review. To calculate sentiment scores, word polarities are first obtained from the SentiWordNet polarity lexicon [1]. A polarity-placement algorithm is then used to calculate score values for each aspect and the overall review. Using syntactic dependencies obtained through the Stanford NLP Parser [4], polarity values can be transferred from the polarity word to the aspect keyword, and consequently to the aspect. The tool also displays polarities of polar words, sentences, and other intermediate results. Currently, only the default engine can be used. As future work, we plan to introduce the option to plug in any sentiment analysis engine provided as a web service. We are also planning an extension to allow the user to provide their own polarity lexicon instead of the default one.

References

1. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proc. of LREC. vol. 6, pp. 417–422 (2006)
2. Liu, B.: Sentiment analysis and subjectivity. Handbook of Natural Language Processing, pp. 627–666 (2010)
3. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)
4. Toutanova, K., Klein, D., Manning, C., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. Proc. of the 2003 Conf. of the NAACL 1(June), 173–180 (2003)
5. Zhai, Z., Liu, B., Xu, H., Jia, P.: Clustering product features for opinion mining. In: Proc. of the 4th ACM International WSDM Conf. pp. 347–354. ACM (2011)