

Automatic Generation of Lexical Resources for Opinion Mining: Models, Algorithms and Applications

Andrea Esuli*

Istituto di Scienza e Tecnologie dell'Informazione

Consiglio Nazionale delle Ricerche

via G. Moruzzi, 1

56124, Pisa - ITALY

andrea.esuli@isti.cnr.it

Opinion mining is a recent discipline at the crossroads of Information Retrieval and of Computational Linguistics which is concerned not with the topic a document is about, but with the opinion it expresses. It has a rich set of applications, ranging from tracking users' opinions about products or about political candidates as expressed in online forums, to customer relationship management.

Functional to the extraction of opinions from text is the determination of the relevant entities of the language that are used to express opinions, and their opinion-related properties. For example, determining that the term **beautiful** casts a *positive* connotation to its subject.

In this thesis we investigate on the automatic recognition of opinion-related properties of terms. This results into building opinion-related *lexical resources*, which can be used into opinion mining applications.

We start from the (relatively) simple problem of determining the *orientation* of subjective terms. We propose an original semi-supervised term classification model [2] that is based on the quantitative analysis of the *glosses* of such terms, i.e. the definitions that these terms are given in on-line dictionaries. This method outperforms all known methods when tested on the recognized standard benchmarks for this task.

We show how our method is capable to produce good results on more complex tasks, such as discriminating *subjective* terms (e.g., **good**) from *objective* ones (e.g., **green**) [3], or classifying terms on a fine-grained *attitude* taxonomy [1].

We then propose a relevant refinement of the task, i.e., distinguishing the opinion-related properties of distinct *term senses*. We present SENTIWORDNET [4], a novel high-quality, high-coverage lexical resource, where each one of the 115,424 senses contained in WORDNET has been automatically evaluated on the three dimensions of positivity, negativity, and objectivity.

We propose also an original and effective use of random-walk models to rank term senses by their positivity or negativity [5, 6]. The random-walk algorithms we present have a great application potential also outside the opinion mining area, for example in word sense disambiguation tasks. A result of this experience is the generation of an improved version of SENTIWORDNET.

*Student of the Doctorate in Information Engineering of the University of Pisa

We finally evaluate and compare the various versions of SENTIWORDNET we present here with other opinion-related lexical resources well-known in literature, experimenting their use in an *Opinion Extraction* application. We show that the use of SENTIWORDNET produces a significant improvement with respect to the baseline system, not using any specialized lexical resource, and also with respect to the use of other opinion-related lexical resources.

Available online: <http://etd.adm.unipi.it/theses/available/etd-04042008-163914/>

References

- [1] S. Argamon, B. Kenneth, A. Esuli, and F. Sebastiani. Automatically determining attitude type and force for sentiment analysis. In *Proceedings of LTC-07, the 3rd Language & Technology Conference*, Poznan, PL, October 2007.
- [2] A. Esuli and F. Sebastiani. Determining the semantic orientation of terms through gloss classification. In *Proceedings of CIKM-05, the ACM SIGIR Conference on Information and Knowledge Management*, pages 617–624, Bremen, DE, 2005. ACM Press.
- [3] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings EACL-06, the 11rd Conference of the European Chapter of the Association for Computational Linguistics*, pages 193–200, Trento, IT, 2006.
- [4] A. Esuli and F. Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC-06, the 5th Conference on Language Resources and Evaluation*, Genova, IT, 2006.
- [5] A. Esuli and F. Sebastiani. Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of ACL-07, the 45th Annual Meeting of the Association of Computational Linguistics*, pages 424–431, Prague, CZ, June 2007. Association for Computational Linguistics.
- [6] A. Esuli and F. Sebastiani. Random-walk models of term semantics: An application to opinion-related properties. In *Proceedings of LTC-07, the 3rd Language & Technology Conference*, Poznan, PL, October 2007.