

## Chapter 13

# Opinion Mining with SentiWordNet

**Bruno Ohana**

*Dublin Institute of Technology, Ireland*

**Brendan Tierney**

*Dublin Institute of Technology, Ireland*

### ABSTRACT

*Opinion Mining is an emerging field of research concerned with applying computational methods to the treatment of subjectivity in text, with a number of applications in fields such as recommendation systems, contextual advertising and business intelligence. In this chapter the authors survey the area of opinion mining and discuss the SentiWordNet lexicon of sentiment information for terms derived from WordNet. Furthermore, the results of their research in applying this lexicon to sentiment classification of film reviews along with a novel approach that leverages opinion lexicons to build a data set of features used as input to a supervised learning classifier are also presented. The results obtained are in line with other experiments based on manually built opinion lexicons with further improvements obtained by using the novel approach, and are indicative that lexicons built using semi supervised methods such as SentiWordNet can be an important resource in sentiment classification tasks. Considerations on future improvements are also presented based on a detailed analysis of classification results.*

### INTRODUCTION

Opinion information concerns people's expressed beliefs and judgments on a certain topic, and can be an important component used in making more accurate decisions in a number of scenarios. Companies for instance, have a keen interest in finding out what are customers saying about

their products and service offerings. Consumers on the other hand would benefit from accessing other people's opinions and reviews on products they wish to purchase, as recommendations from other users tend to play a part on influencing such decisions. Knowledge of other people's opinions is also important on other realms such as political activism, where for instance it could be of interest to discover the general sentiment towards a new piece of legislation or towards political parties and

DOI: 10.4018/978-1-60960-067-9.ch013

public figures; or in the detection of subjective bias on environments where there should be none, such as in monitoring news coverage.

In recent years, the internet has enabled access to opinions in the form of written text from a variety of sources and in a much larger scale: it is now easier for people to express their opinions on virtually any subject by means of specialized product review websites, discussion forums and blogs. This is in fact a growing trend, as pointed out in research performed by Horrigan (2008) over 30% of internet users have at one time posted a comment or review online about a product or service they've purchased suggesting an ever growing availability of opinion related information on the web. The same research states that, as on 2007, 81% of internet users in the United States have used the internet to perform research on a product they intended to purchase. Further evidence of the importance of opinions in guiding consumer decisions can be seen in a study on the online travel industry from (Akehurst, 2009) highlighting the perceived high credibility of information found in user generated content, and on the study relating transaction feedback posted by users and consumer behavior on online auction services (Dellarocas, 2003).

The internet is quickly becoming a vast repository of publicly available user generated content dedicated to expressing opinions on any topic of interest. However, there are challenges in extracting useful information from large volumes of data. In Horrigan (2008), 58% of internet users reported that researching product information online was either confusing, difficult to find, or have found the volume of information available to be overwhelming. Suggesting information overload issues may be present in online opinion repositories where these resources become too big to be analyzed in a timely fashion, and are poorly utilized as a result (Farhoomand & Drudy, 2002). Automated methods for efficiently extracting opinion knowledge from these resources appear an attractive proposition for both individuals who

would be able to make informed decisions and to companies who could quickly gauge opinions on their products and services, adding this knowledge to their product development processes. These goals are in essence closely related to those of the discipline of knowledge discovery proposed in Fayyad et al. (1996), which concerns computational methods aiming at finding "valid, novel, potentially useful and ultimately understandable patterns from data". In addition, opinions are generally expressed in textual form, making it a rich ground for the application of text mining techniques and natural language processing. Thus the motivating need to analyze large volumes of opinion information, coupled with advances in natural language processing and knowledge discovery methods gave rise to research in the emerging field of *Opinion Mining*.

This chapter introduces the reader to the research field of opinion mining by presenting a review of research literature, an outline of the potential applications of this technology and the intellectual challenges involved in extracting useful knowledge from opinions in text. Particular emphasis is given to the topics of predictive opinion mining and the application of sentiment lexicons to such tasks. The authors also present the results of their research with the SentiWordNet lexicon (Esuli & Sebastiani, 2006) applied to the task of sentiment classification of film reviews. This research presents a unique approach that uses opinion lexicons to build a set of features that can be used to train a classifier, which achieved improved classification results in the experiment. The results obtained are discussed together with findings and opportunities for future development.

## OPINION MINING

The starting point for mining opinions in text is the abstraction that a document can be considered as a collection of objective and subjective statements, where objective statements refer to factual

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/opinion-mining-sentiwordnet/46900](http://www.igi-global.com/chapter/opinion-mining-sentiwordnet/46900)

## Related Content

---

### Multiagent Knowledge-Based System Accessing Distributed Resources on Knowledge Grid

Priti Srinivas Sajja (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains* (pp. 244-265).

[www.irma-international.org/chapter/multiagent-knowledge-based-system-accessing/46899](http://www.irma-international.org/chapter/multiagent-knowledge-based-system-accessing/46899)

### Federated Data Warehouses

Stefan Bergerand Michael Schrefl (2010). *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications* (pp. 82-107).

[www.irma-international.org/chapter/federated-data-warehouses/39589](http://www.irma-international.org/chapter/federated-data-warehouses/39589)

### The Birth of a Survey: Developing an Assessment of Preservice Teachers' Diversity Awareness

Mary Cain Fehr, Mary Frances Agnello, Steven M. Crooks, Fethi Inanand Raymond Flores (2012). *Cases on Institutional Research Systems* (pp. 288-304).

[www.irma-international.org/chapter/birth-survey-developing-assessment-preservice/60855](http://www.irma-international.org/chapter/birth-survey-developing-assessment-preservice/60855)

### Scenarios in Event Bushes: A Formal Proposal

Uwe Wolter (2018). *Dynamic Knowledge Representation in Scientific Domains* (pp. 76-88).

[www.irma-international.org/chapter/scenarios-in-event-bushes/200170](http://www.irma-international.org/chapter/scenarios-in-event-bushes/200170)

### Opinion Mining with SentiWordNet

Bruno Ohanaand Brendan Tierney (2011). *Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains* (pp. 266-286).

[www.irma-international.org/chapter/opinion-mining-sentiwordnet/46900](http://www.irma-international.org/chapter/opinion-mining-sentiwordnet/46900)