



Sentiment Analysis Approaches for Product Evaluation Using Social Media Data

Bachelorarbeit

zur Erlangung des akademischen Grades „Bachelor of Science (B.Sc.)“ im Studiengang
Wirtschaftswissenschaft der Wirtschaftswissenschaftlichen Fakultät der
Leibniz Universität Hannover.

Vorgelegt von:

Nachname: Zmuda

████████████████████

Vorname: Nicole

██████████

Prüfer: Prof. Dr. rer. nat. Michael H. Breitner

Hannover, den 11. August 2016

List of contents

List of contents.....	I
Abstract.....	III
List of figures.....	IV
List of tables.....	V
List of abbreviations.....	VI
1 Introduction	1
1.1 Motivation and Relevance.....	1
1.2 Objectives of Research.....	2
1.3 Research Design.....	2
2 Conceptual Knowledge and the Implementation of an Algorithm in R.....	3
2.1 Text Mining.....	3
2.2 Opinion Mining for Product Evaluation.....	4
2.3 Feature-Based Opinion Mining and Ranking Algorithm Implementation in R.....	7
2.3.1 Motivation of the Researchers and Authors Eirinaki et al. (2011).	7
2.3.2 System Architecture and Opinion Mining and Ranking Algorithms	8
2.3.3 An Experimental Implementation of an Algorithm in R.	13
3 Analysis of existent Sentiment Analysis Algorithms and Basic Approaches for the Optimization with Social Media Data.....	16
3.1 Related Work aiming at the Mining of Product Features.....	16
3.1.1 Systematic and Computational Treatment of Explicit Features	16
3.1.2 Systematic and Computational Treatment of Implicit Features	23
3.1.3 The Importance of Mining Comparative Opinions	34
3.2 Discussion and Basic Approaches for the Optimization of Sentiment Analysis Algorithms.....	40

4	Critical Considerations of Sentiment Analysis with the help of Mining Algorithms.....	45
4.1	Recapitulation and Limitations of the Gained Perceptions.....	45
4.2	Recommendations for Further Research.....	49
	Conclusion.....	53
	References.....	54
	List of Attachments.....	VII
	Ehrenwörtliche Erklärung.....	XII

1 Introduction

1.1 Motivation and Relevance

“Rich does not mean anything unless there is poor, happy unless there is sad. [...] It is not “content” that determines meaning, but “relations“.”

(Arthur Asa Berger,
American professor emeritus in broadcast and electronic communication arts, 2014)

An increasing amount of people are buying products online.¹ The revenue generated from e-commerce was rapidly growing in the last years and is forecasted to increase furthermore.² As long ago as the end of 1996, Intel Chairman Andrews Grove was explaining that an enterprise needs to define its business as more than just the building and selling of goods. It is needed to be described as the delivery of information and lifelike interactive experiences, according to Grove.³

With the purpose of enhancing customer satisfaction and the online shopping experience, the enabling of reviewing or judging a product has become a customarily technique of online merchandising businesses.⁴ Almost one in two global online shopper is influenced in their buying decision by reading social media reviews, comments and feedbacks.⁵ In the year 2011, a study was conducted in a period of approximately one month by interviewing 1299 people in Germany on whether product reviews are important or not. 81,4 percent of them did agree on the hypothesis that product reviews are indeed meaningful.⁶

Plenty of new challenges and opportunities, regarding the method how information about a product is searched, are accompanied by the proliferation of blogs and social networks.⁷ It is estimated that there will be around 2.55 billion social network users around the globe in 2018, up from 1,87 billion in 2014.⁸ It is expected that 69,3 percent of worldwide online population will be accessing social networks, up from 64 percent in 2014.⁹ Therefore, the internet is an instrument for making it possible to get, in some cases, more than hundreds of product reviews, comments and feedbacks from people, a potential customer have never heard of before.^{10,11} Moreover, different languages and, on top of that, different letters like, for example, Chinese ideographs can be reasons for complicating the process of understanding what potential customers or employees want to know about a specific product. Since the number of product reviews is increasing rapidly, it becomes difficult to extract the helpful and interesting product reviews from the ones that do not contain personally valuable information. Several product reviews can contain personal experiences or lists of pros and cons referring to different and perhaps some personally uninteresting features of the specific product.¹²

¹ Cf. Hu and Liu, 2004a, p. 168

² Cf. Emarketer, 2016a

³ Cf. Centre of Retail Research, n.d.

⁴ Cf. Hu and Liu, 2004b, p. 168

⁵ Cf. PricewaterhouseCoopers, 2016

⁶ Cf. Dr. Schengber & Friends, 2011

⁷ Cf. Eirinaki et al., 2011, p. 1175

⁸ Cf. Emarketer, 2016b

⁹ Cf. Emarketer, 2016c

¹⁰ Cf. Hu and Liu, 2004b, p. 775

¹¹ Cf. Pang and Lee, 2008, p.1

¹² Cf. Hu and Liu, 2004b, p. 168

At this point, the question of whether automatic evaluations of the opinions contained in product reviews that can be found would be advantageous for all involved in the business, i.e. for the potential customers as well as the manufacturer and thus the provider of the product.

1.2 Objectives of Research

As yet, how many researchers did concentrate on the programming of an algorithm that analyzes the opinions expressed in product reviews in order to automatize the evaluation of them? What differences can be identified by comparing a range of these algorithms? For what reasons optimizing aspects have to be considered in order to implement an efficient algorithm that is mining opinions included in product reviews? What potential tasks for further research can be identified and why? The why and wherefore is focused answering the above-mentioned questions.

1.3 Research Design

On the one hand, this bachelor thesis is presenting a theoretical background about the relevant technical terms “Text Mining”, “Opinion Mining” and “Sentiment Analysis” in chapter 2, whereby the latter is used as a synonym for Opinion Mining in the following. This section is including definitions, objectives as well as current challenges referring to the above-mentioned terms. Furthermore, the second chapter of this thesis is summarizing the paper titled “Feature-based opinion mining and ranking” written by Eirinaki, Pinal and Singh in the year 2011. The second chapter of this bachelor thesis will refer closely to one of the algorithms programmed by the three above-mentioned authors. One popular open source software, that has made budget-friendly software available to text miners, is called R. Since the language has a wide acceptance for statistical and applied mathematical applications, the chapter 2 will focus on using that programming language as an approach in order to recreate one of the algorithms.¹³ The third chapter is focusing on the presentation of related work published by many different researchers in the last years. The systematical and computational treatment of explicit features will be distinguished from the treatment of implicit features as well as the mining of comparative opinions. By discussing the differences of the analysis algorithms as well as other relevant aspects presented in the essays, various basic approaches for the optimization of sentiment analysis algorithms are formulated in chapter 3. On the one hand, chapter 4 is focusing on the illustration of a range of limitations of the discussed algorithms. On the other hand, the second part of chapter 4 is presenting selected recommendations for further research by including a marketing research approach for investigating technological acceptance. Finally, chapter 5 is concluding this bachelor thesis by giving a short summary.

¹³ Cf. Walker, 2010, p. 1

5 Conclusion

The presented thesis has been studied various sentiment analysis approaches for customer product reviews, comments and feedbacks. The analysis was realized to pursue the demonstration of basic approaches for optimizing opinion mining systems. Besides, limitations and opportunities for further research tasks were shown.

The analysis based on the fact that the number of evaluations of products in an online form is forecasted to increase, since a rising amount of people are buying products online and the willingness to experience from other people's opinion has always been an aspect of the human nature. In the light of the dynamism and the popularity of social networks, a rapid spread of information is guaranteed. Next to the uniqueness of the language and the tools in social networks, these facts are presenting immense challenges for mining opinions.

Through the experimental implementation of a feature-based ranking algorithm in R, threats and challenges could be found, like the inability of mining comparative opinions or identifying colloquial language terms and further grammatical and stylistic specialties of language. Afterwards, these perceptions were further studied on as well as added by examining additional opinion mining systems focusing on several difficulties in the field of opinion mining.

Based on these findings, opportunities and challenges for optimizing sentiment analysis approaches, formulated by various researchers, have been stated. Further research tasks were recommended, for example the involvement of different social media tools and functions in the opinion mining system's frameworks as well as basic approaches for a scientific structure of a marketing research study pursuing the target of verifying the acceptance of sentiment analysis approaches of different target groups.