Exploring Corporate Reputation based on Sentiment Polarities That Are Related to Opinions in Dutch Online Reviews

R. E. Loke[a] and J. Vergeer
Centre for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Keywords: Aspect Based Sentiment Analysis (ABSA), Opinion Mining, English Texts, Dutch Texts.

Abstract: This research demonstrates the power and robustness of the vocabulary method by Hernández-Rubio et al. (2019) for aspect extraction from online review data. We showcase that this algorithm not only works on the English language based on the CoreNLP toolkit, but also extend it on the Dutch language, specifically with aid of the Frog toolkit. Results on sampled datasets for three different retailers show that it can be used to extract fine-grained aspects that are relevant to acquire corporate reputation insights.

1 INTRODUCTION

As evidenced by GfK (2019), the number of online purchases in the Netherlands has continuously risen in recent years. Along with this rise in purchases, the importance of online reviews is also evident as over 90% of online shoppers research into an online retailer through customer reviews (Statista, 2019). In addition, it is well known that the number of online reviews that a retailer has, affects its sales (Chevalier & Mayzlin, 2006). Following these trends, the increasing importance of online reviews might suggest that online reviews form a relevant company asset that, at least to some extent, give an indication of an organization's reputation. At the moment little research has been done on this subject; the literature on corporate reputation that does exist mainly employs surveys which is not always the most readily available data. If it were possible to partly base corporate reputation on online reviews it could logically provide meaningful insights to stakeholders like online retailers to boost their performance. Thus, corporate reputation retained from online reviews could prove to be an intangible yet important resource for an organization. Intangible resources are in general known to be important due to the business value they can potentially create and the difficulty of replication by competitors (Roberts & Dowling, 2002).

With regard to processing capabilities of online reviews that are related to corporate reputation, a comprehensive analysis using unsupervised techniques and the automatic extraction of aspects could be useful and might be worth researching into as there are often no explicit ground thruths available for online review data. Our research will therefore focus on this angle. Hastie, Tibshirani and Friedman (2009) describe unsupervised learning as attempting to infer properties from data without having a ground truth available: unsupervised learning techniques lack a clear cut measure of success and are generally used when there is no ground truth available in the data. It is, at least, an interesting alternative to expensive supervised solutions that require example output labels along with online review data to work properly. In addition, the performance of supervised systems is often limited due to a certain practical imbalance that is commonly observed in data in terms of the number of positive reviews available in comparison to the number of negative reviews. The vocabulary method by Hernández-Rubio et al. (2019) is a candidate unsupervised technique that allows to extract fine-grained aspects from texts and facilitates automatic labeling of their sentiment polarities. It depends on dependency parsing and, with the method, already excellent results have been obtained in the English language on general texts. In this paper, we research into applying and extending this precise method to online reviews in Dutch.

[a] https://orcid.org/0000-0002-7168-090X

Loke, R. and Vergeer, J.
Exploring Corporate Reputation based on Sentiment Polarities That Are Related to Opinions in Dutch Online Reviews.
DOI: 10.5220/0011285500003269
In *Proceedings of the 11th International Conference on Data Science, Technology and Applications (DATA 2022)*, pages 423-431
ISBN: 978-989-758-583-8; ISSN: 2184-285X

Table 1: Definitions on the three dimensions that comprise corporate reputation according to Lange et al. (2011, p. 155).

Dimension	Definition
Being known	"*Generalized awareness or visibility of the firm; prominence of the firm in the collective perception*"
Being known for something	"*Perceived predictability of organizational outcomes and behavior relevant to specific audience interests*"
Generalized favorability	"*Perceptions or judgements of the overall organization as good, attractive and appropriate*"

In the rest of the paper, we review relevant literature in Section 2, describe our processing framework as well as the online review data that we obtained in Section 3, give results in Section 4, and discuss and conclude the work in Section 5.

2 LITERATURE REVIEW

In this section, an overview of recent scientific literature on corporate reputation and online reviews is given. Thereafter, we also focus on two relevant generic text mining methods, being topic extaction and sentiment analysis.

2.1 Corporate Reputation

Definitions of corporate reputation can differ when looking at multiple sources. Barnett, Jermier & Lafferty (2006, p. 34) define corporate reputation, after reviewing the definitions used in a multitude of other sources, as the "Observers' collective judgments of a corporation based on assessments of the financial, social, and environmental impacts attributed to the corporation over time". This is, at least in part, in line with Hall's (1993) findings where an organization's reputation is usually built over multiple years, yet also fragile. Gotsi and Wilson (2001) present two schools of thought, one where corporate reputation is synonymous with corporate image and one where the two concepts are seen as different, but generally are also seen as interrelated. Barnett et al. (2006) argue that we should move away from seeing corporate reputation and image as being one and the same, which is in line with the second school of thought of Gotsi and Wilson (2001). This definition, where corporate reputation and image are interrelated yet not synonymous, would be logical to adopt for this research with online reviews as the other vocal point of interest. Online reviews encompass a customer's experience with a company and not necessarily how that company is seen by said customer.

Table 2: The seven dimensions that make up corporate reputation according to Fombrun et al. (2015).

Dimension	Definition
Products & Services	Whether the company's offerings are considered of high quality, value, and service, and if customer needs are met.
Innovation	Whether or not a company is innovative and adaptive.
Workplace	Perceptions of the way a company shows concern for its employees along with rewarding and treating them fairly.
Governance	Perceptions of how a company is held regarding ethics, fairness and transparency.
Citizenship	How a company is perceived regarding the environment, whether it supports good causes and its contributions to society.
Leadership	Whether leaders within a company are perceived as visionary and if they endorse their companies.
Performance	Perceived financial performance, profitability, and growth.

Lange, Lee & Dai (2011) found that corporate reputation can be categorized along three different dimensions; see Table 1. Given the context of this research, the being known for something dimension seems to be most relevant as specific aspects like delivery and product quality would intuitively be the most prevalent in reviews. Generalized favorability might be difficult to assess on individual review level but could be seen along the lines of overall sentiment on a company. Lange et al. (2011) do underline that corporate reputation should be viewed along more than one dimension. While useful in conceptualizing corporate reputation, Lange et al. (2011) really focus on defining corporate reputation and its underlying dimension as opposed to quantifying it. They do not mention any specific variables that could be used to model the dimensions. This seems to occur often as other research into corporate reputation, like Lange et al. (2011), generally put a heavy focus on defining corporate reputation and the role it plays in an organization's success instead of attempting to, for example, quantify it. One of the articles that does lean more in this direction is the article by Fombrun, Ponzi and Newburry (2015). This article attempts to validate a system that assesses an organization's reputation from the perspectives of different stakeholders. They do this by validating the dimensions that this system uses, namely: "Products & Services", "Innovation", "Workplace", "Governance", "Citizenship", "Leadership" and "Performance". For definitions on these dimensions, see Table 2. For our research the possibility of extracting all seven dimensions from online reviews does not seem likely. A dimension like, for example, "Leadership" is not likely to be relevant in an online customer review. A dimension like "Products & Services" and perhaps "Innovation" could be relevant, however.

Unlike Lange et al. (2011), Fombrun et al. (2015) do mention specific variables that make up their found dimensions, see Table 3. These variables show that this specific model has been made to be based on the perspectives of multiple stakeholders. The "stands behind" variable of the Products & Services dimension focuses on the company while the other variables of that dimension focus more on the products and services themselves. This is relevant since in this research, the online reviews from Trustpilot generally encompass the views of the customer stakeholder group.

Table 3: Variables that make up the seven corporate reputation dimensions of Fombrun et al. (2015).

Dimension	Variables
Products & Services	High Quality; Good Value; Stands Behind; Meets Customer Needs
Innovation	Innovative; First to Market; Adapts to Change
Workplace	Rewards Employees Fairly; Concern for Employees' Well-Being; Equal Opportunities in Workplace
Governance	Open and Transparent; Behaves Ethically; Fair in Doing Business
Citizenship	Protects Environment; Supports Good Causes; Positive Influence on Society
Leadership	Strong and Appealing Leader; Clear Vision of Future; Well Organized; Excellent Managers
Performance	Profitable; Good Financial Results; Strong Growth Prospects

In conclusion, corporate reputation is defined as a collective view on a corporation. The reputation is built over time and can be fragile. In the context of online reviews, it likely can only be measured in the "being known for something" and "generalized favorability" dimensions of Lange et al. (2011). These dimensions are quite broad however and were not designed to be used directly as quantification dimensions. The model by Fombrun et al. (2015) could be used, but likely only in part as it concerns more stakeholder groups than that single-faceted research would logically focus on. The amount of work done in attempting to quantify corporate reputation remains quite limited and the models that do exist are likely not completely applicable.

2.2 Online Reviews

Online reviews can be a valuable resource for insights in an organization's performance. Wang, Lu and Zhai (2010) state that with Web 2.0 it is possible for large groups of people to express their opinions on many things like products and services. They continue in saying that these opinions contribute to both other users and retailers as they enable both types of parties to gather information and make educated decisions.

Xing and Zhan (2015) do note some flaws in this, the first being that with every single person being able to post their opinions, the quality of these opinions cannot be guaranteed. They note some examples like online spammers and fake reviews. The second flaw Xing and Zhan (2015) note is that a ground truth in the context of, for example, online reviews is a "tag of a certain opinion" and not necessarily an established truth. It thus can be useful for gauging the opinions of a customer base, but they do not necessarily display an accurate description of the quality of a product/service. This is exemplified by Korfiatis, García-Bariocanal and Sánchez-Alonso (2012) who state that the experience that multiple consumers have with the same product can be different due to, for example, differences in expectations. It can however be argued that the aggregated opinions of a customer base are a strong indicator of quality, making the true truthfulness of online reviews somewhat context reliant.

2.3 Text Mining

The analysis of textual data, also known as text mining, can have a different definition depending on the area in which it is being applied (Hotho, Nürnberger and Paaß, 2005). The latter authors focus on the area which is also most relevant for this research and as such define text mining as "the application of algorithms and methods from the fields machine learning and statistics to texts with the goal of finding useful patterns" (Hotho et al., 2005, p. 4).

The paper by Hotho et al. (2005) mentions multiple methods that fall under text mining, of which some are relevant to this research and need to be discussed and explored further.

To analyze text data it first is in general in step (1) necessary to find structure in them, this can be achieved with techniques like topic modelling (step 1A) or through the classification of syntactic dependencies (step 1B). After that, in step (2), the data can be analyzed on aspects like the prevalent sentiments in the text. As the granularity of the data increases, however, the difficulty of the analysis will also increase.

The success of working with textual data can be reliant on the language used in the data and on data preparation (pre) processing that is commonly done in step (0). Relatively simple operations like stemming (reducing words to their root (Lovins, 1968)) and removing stop words (words that carry a grammatical function but do not carry a meaning of the document's content (Wilbur & Sirotkin, 1992)) might largely still be successful but more refined operations like lemmatization (changing words to their basic forms (Korenius, Laurikkala, Järvelin, & Juhola, 2004)) can be more difficult. The Stanford CoreNLP toolkit (Manning, et al., 2014) carries a lemmatize option but does not support the Dutch language. An alternative is the Frog toolkit (Van den Bosch et al., 2007). Frog was specifically made for the Dutch language and can, among others, lemmatize words, add part of speech tags (POS) and add syntactic dependency tags.

Below we describe topic extraction (step 1A) and sentiment analysis (step 2). Modeling of syntactic relationships (step 0) and classification of syntactic dependencies (step 1B) will be described in Section 3.2 and Section 3.3.

2.3.1 Topic Extraction

The first step besides data preparation often is to find the most important groups of aspects that can be found in review data. By applying LDA and BTM, underlying topics in the available reviews can be extracted to gain insight into the aspects these reviews focus on. These aspects could in our application that is directed to social listening from Trustpilot online reviews for example be subjects like product quality or delivery time.

According to Hu, Boyd-Graber, Satinoff and Smith (2014), the problem of dealing with large volumes of unstructured textual data remains a persistent problem. To gain insight in underlying themes that can be found in texts, topic modeling can be a potential solution. A well-known and often used topic modeling technique is "Latent Dirichlet Allocation" (LDA) by Blei, Ng and Jordan (2003). The idea behind this technique is that a text document is composed of a finite number of underlying "topics" or themes. In the article where they originally presented their findings Blei et al. (2003) test the effectiveness of the topics found by LDA by using them as features in a classification model. They report that not only did LDA reduce the dimensionality of the model (it was tested against a model using word occurrence), but it also achieved high performance. As such it can be seen as a method that can bring structure to text data which could solve a problem with the analysis of text data mentioned by Hotho et al. (2005), namely that text data is unstructured and as such cannot easily be processed by a computer.

It must be noted however that topic models are not perfect. According to Hu et al. (2014), it can very well occur that a found topic is nonsensical, practically a duplicate of another topic or a combination of multiple topics that logically speaking should be separate.

Another method for extracting topics is the "biterm topic model" (BTM) by Yan, Guo, Lan, and Cheng (2013) whose authors state that conventional topic models like LDA may not work well on short texts due to data sparsity in these documents. Weng, Lim, Jiang, and He (2010), for example, attempt to remedy this sparsity by aggregating the texts that they analyze based on author. Instead of assuming that a text is comprised of a mixture of topics, BTM assumes that the entire corpus is a mixture of topics where the cooccurrence of two words (a biterm) is independently related to a topic. With BTM it could be interesting to extract topics from the reviews as these are generally shorter than, say, a chapter of a book.

Since LDA and BTM are being unsupervised learning techniques (there is no predefined ground truth available) a potential outcome is a collection of nonsensical topics. To assess topic models and find the optimal number of topics, the perplexity metric is commonly used as was done in Blei et al. (2003).

It is important to mention that topic modeling approaches (step 1A) proved to be unsuccessful for our application goals. Therefore, we adopted step 1B instead.

2.3.2 Sentiment Analysis

After that text data has been structured it can be more readily analyzed, for example through sentiment analysis. Sentiment analysis is analyzing the sentiment that a group of people hold towards specific entities like products, services, or organizations (Xing & Zhan, 2015; Liu, 2012). According to Liu (2012) there are, broadly speaking, three levels of sentiment analysis: document level, sentence level and entity level. On the document level, sentiment is assessed over the entire text, on sentence level, sentiment is assessed per sentence. Entity level is the most fine-grained level where sentiment is really specified on the opinions themselves. Liu does note that as the analysis level gets more specific, the level of difficulty of the analysis also rises. This is in line with Nguyen, Shirai and Velcin (2015) who state that sentiment analysis on social media texts is difficult due to problems like short text lengths, spelling errors and uncommon grammar.

Given the focus of our research, sentiment analysis is likely to play an important role. There are multiple ways to approach it, Farhadloo and Rolland (2013), for example, use a machine learning approach, whereas Vashishtha and Susan (2019) use a rule-based approach. The common factor of both approaches is that they use sentiment scoring in some shape or form. These scores represent the sentiment of the analyzed aspect(s) be they individual words, sentences, or entire texts.

3 DATA AND METHODOLOGY

In this section, we describe the online review datasets that we obtained, the modeling of syntactic relationships in the review texts as well as the aspect extraction and sentiment labeling that we applied in our processing framework.

3.1 Online Review Datasets

The primary data that this research employs concerns online reviews collected from trustpilot.com as this website by far contains the largest number of online reviews. Trustpilot.com is an often-used website for writing and sharing online reviews, it is in the top 1% most popular websites and boasts over 97 million reviews on over 420,000 companies (Trustpilot, 2020).

Our specific data concern reviews on 3 different online retailers operating in the Netherlands: Bol.com, Agradi, and ABOUT YOU. The number of reviews, being exactly 100 for each company, is relatively low and is only meant for the purpose of demonstrating the use of any techniques found to be performing well. In future work, once techniques have been completely finetuned properly, we obviously plan to process larger datasets for diverse stakeholders and our partners.

The data was collected using the web scraping framework Scrapy in the Python programming language. The ratings on the Trustpilot website that come along with the reviews were scraped as well to test several validity assumptions. In general, ratings can give a sense of the overall sentiment of a review but do not say much on the more fine-grained topics that corporate reputation is assessed on.

3.2 Modeling of Syntactic Relationships

An interesting way that turns out to be quite useful of finding structure in unstructured textual data is to look at the syntactic relations in texts. According to Qiu, Liu, Bu and Chen (2011) the syntactic relation between two words (A and B) can be defined as: A depends on B, or B depends on A. These dependencies can differ in nature, an adjective might be dependent on a noun but that same noun might be dependent on a verb. One relation might specify the subject of a sentence, others might specify a modification of the meaning of that sentence (adding a sentiment or modifying the intensity of that sentiment). Applying this definition to a full sentence, a syntactic tree can be made, visualizing relationships between words. For an example of such a visualization, see Figure 1. Annotations of these relations between words are usually not readily available (especially in user generated content like online reviews) and as such an often-used solution is a so-called dependency parser like the one available in the Stanford CoreNLP toolkit (Manning, et al., 2014; Chen & Manning, 2014).

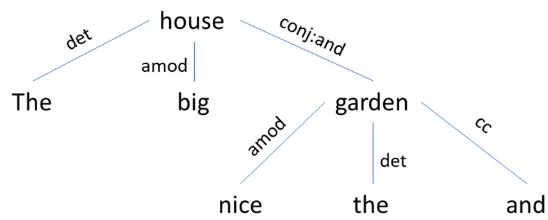

Figure 1: Syntactic tree of the sentence: “The big house and the nice garden”, classified by the Stanford CoreNLP toolkit.

Syntactic dependencies can be extracted in more languages than just English, the Stanford CoreNLP toolkit can work with five different languages (Manning, et al., 2014). The Dutch language is not supported, however; an alternative is the Frog toolkit by Van den Bosch, Busser, Canisius and Daelemans (2007) which was designed specifically for the Dutch language. Frog is capable of many of the tasks CoreNLP can perform, one of which being able to extract syntactic dependency tags, as evidenced in Figure 2.

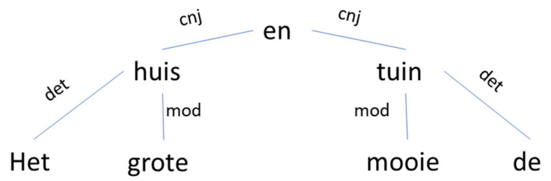

Figure 2: Syntactic tree of the sentence: "Het grote huis en de mooie tuin", classified by the Frog toolkit.

In terms of syntactic dependency tags, CoreNLP and Frog largely convey the same information. Looking at the syntactic trees in Figures 1 and 2 there is at least one structural difference, however. The way conjunctions like "and" in Figure 1, and "en" in Figure 2, are noted makes for different types of relations between words, between English and Dutch. Individually this might be of minimal impact, but it is something to take note of when using an algorithm that takes syntactic dependency tags as input.

3.3 Aspect Extraction

There are multiple approaches that can be taken to extract aspects or entities from a text. The (unsupervised) vocabulary-based method described by Hernández-Rubio, Cantador, & Bellogín (2019, p. 404) is based on the syntactic dependency and POS tags generated by the Stanford CoreNLP toolkit and extracts sentiment polarities that are related to aspects on entity level in English texts. This algorithm generates outputs in the form of:

$$(noun,\ adjective,\ modifier,\ isAffirmative) \quad (1)$$

where *noun* is the extracted aspect with *adjective* specifying a positive, neutral, or negative opinion (sentiment polarity whereto the adjective adheres to, see next subsection how the sentiment polarity is being determined) which is potentially modified by *modifier*. The variable *isAffirmative* specifies whether the sentiment polarity must be inverted due to the word "not" being used.

The vocabulary method of Hernández-Rubio et al. (2019, p. 404) is in their paper available as pseudo-code. This code has been implemented from scratch in Python, using the results gained from applying the Stanford CoreNLP toolkit for English texts and the Frog toolkit for Dutch texts. As has been mentioned before, in comparison to CoreNLP, Frog uses different syntactic dependency tags, but largely conveys the same information.

3.4 Sentiment Classification

Different approaches can be adopted to map adjectives to sentiment polarities.

A labor-intensive possibility for finding ground truths would be to manually tag adjectives in reviews according to their sentiment.

Fang and Zhang (2015) take the approach of calculating the overall sentiment on sentence level using predefined sentiment scores by looking at the used words. With this approach it may be possible to assess the correctness of the sentiment polarities by aggregating the polarities on review level to be compared with the review ratings. However, to the best knowledge of the authors, there is no general applicable list of Dutch words available wherein every word has a sentiment score attached.

There are lists of positive and negative words available, positive, and negative Dutch lexicons, however (Chen & Skiena, 2014). Using these lists, adjectives can be roughly classified as positive or negative. A neutral sentiment only gets assigned if an adjective has conflicting polarities and on average is neither positive nor negative. This is the approach that we have adopted in our processing framework. Note that we only make use of a binary sentiment classification based on word use, and that there is not any scoring involved. This makes that the modification of sentiments, e.g., the use of the word "very", is not accounted for in the current implementation.

4 RESULTS

For initial testing of the combination of Frog along with the vocabulary method, the 30 testing sentences used by Hernández-Rubio et al. (2019) were translated into Dutch and processed by the algorithm. The workings of the method in combination with CoreNLP had been tested in an earlier stage and were comparable to the original results of Hernández-Rubio et al. (2019).

Table 4: Aspects extracted from simple sentences.

The hotel staff was friendly.	('staff-3', 'friendly-5', '-', True)
Het hotelpersoneel was vriendelijk.	('hotelpersoneel', 'vriendelijk', '-', True)

Table 5: Aspects extracted from slightly more complicated sentences.

The hotel had friendly and efficient staff.	('staff-7', 'efficient-6', '-', True), ('staff-7', 'friendly-4', '-', True)
Het hotel had vriendelijk en efficiënt personeel.	('efficient', 'vriendelijk', '-', True), ('hotel', 'vriendelijk', '-', True)

Table 6: Aspects extracted from simple sentences (inverted sentiments).

The hotel staff was not friendly.	('staff-7', 'friendly-6', '-', False)
Het hotelpersoneel was niet vriendelijk.	('hotelpersoneel', 'vriendelijk', '-', False)

Table 7: Aspects extracted from slightly more complicated sentences (inverted sentiments).

The hotel staff was not friendly and efficient.	('staff-3', 'efficient-8', '-', True), ('staff-3', 'friendly-6', '-', False)
Het hotelpersoneel was niet vriendelijk en efficient.	('efficient', 'vriendelijk', '-', False), ('hotelpersoneel', 'vriendelijk', '-', False)

Tables 4, 5, 6 and 7 show the results of the aspect extraction. The algorithm for the Dutch language seems competent on simple, straight to the point sentences; see Table 4. Slightly more complicated sentences, namely those that involve conjunctions, work suboptimal; see Table 5. This is likely due to Frog's labeling of conjunctions structurally differing from Stanford's CoreNLP. The algorithm can pick up inverted sentiments ("not friendly") in simple phrases; see Table 6. The Dutch word "efficiënt" had to be transformed to "efficient" due to Frog registering "ë" as a special character which heavily influences how it processes the rest of the sentence; see Table 7. This version is also classified as a noun while it should be an adjective.

Figures 3, 4 and 5 show results of the aspect extraction and sentiment classification for the online review datasets for the retailers Bol.com, Agradi, and ABOUT YOU that were scraped from Trustpilot.

Note that use of personal pronouns such as, e.g., "ik" seems to be prevalent in all figures; these aspects give little insight without their contexts and can in daily practice easily be filtered out so that they do not occlude more interesting and more relevant aspects. This observed personal pronoun problem is not exclusive to the Dutch language however as Hernández-Rubio et al. (2019) mention in their paper that their implementation suffered the same issue.

Bol.com customers seem to be divided on whether they are positive or negative on Bol's customer service ("service", "klantenservice"); see Figure 3.

Delivery seems to be the aspect that is most often used in Agradi reviews, for the majority being positive; see Figure 4. Agradi customer service is quite often neutral, signifying that diverging sentiments are prevalent on this aspect in individual reviews. When price is mentioned in this sample of Agradi reviews it is only done so in a positive context.

Relatively speaking, delivery is mentioned quite often in ABOUT YOU reviews; this occurs in positive contexts, see Figure 5. The extracted aspects seem, for the vast majority, to be used in a positive context. This could be due to the sample, but it is clear that ABOUT YOU, on average, is regarded to be the most positive on the Trustpilot samples out of the three companies. This could, perhaps in part, explain the overwhelming general positivity of customers for ABOUT YOU, but more data should be analyzed.

5 DISCUSSION & CONCLUSION

Although the vocabulary method described by Hernández-Rubio et al. (2019) was originally designed for the English language with the CoreNLP toolkit, it works relatively well with the Frog toolkit on individual Dutch sentences for extracting aspects. The overall results of the vocabulary method on a sample of online Dutch reviews gives some interesting insights on the broad sentiments prevalent in different aspects. These results demonstrate the potential of this algorithm and show that it can extract fine-grained aspects.

The outputs generated by the vocabulary method are fine-grained aspects; according to Liu (2012, p. 81), the next step would be to group these aspects in "synonymous aspect categories". The need for this is evidenced in Figures 3 and 4 where two different Dutch words are used to denote the category customer service. It is recommended that further research is done on this as Liu (2012, p. 82) calls it critical for opinion analysis.

For practical reasons like data availability the focus of this research and any demonstrations of found solutions concern reviews on online retailers. There is also the possibility of using alternative sources of information instead of online reviews. Dijkmans, Kerkhof, & Beukeboom (2015) for example report that social media engagement of airlines positively affects perceived corporate reputation. Although more engaged customers are also exposed to posts with a negative sentiment, Dijkmans et al. (2015) find that the net effect remains positive. Social media messages might very well be an interesting addition or alternative to online reviews. The solutions developed during this research are likely to be applicable to online reviews concerning other stakeholders and other textual data like social media messages. Though the scope of this specific research concerns the e-commerce stakeholder its applications by no means must necessarily be limited in the same manner.

Assessing corporate reputation on aspects generated by this method, using for example the model described by Fombrun et al. (2015), might very well be possible but would require more research as the current results are too fine-grained to adequately base estimations on.

Overall, the vocabulary-based method seems to be a very interesting technique that can extract aspects from reviews: the fine-grained aspects give some insights into what topics are prevalent in these reviews and the sentiment can be assessed. More research will have to be done in the future to draw concrete conclusions in relation to corporate reputation, but the demonstrations done in this research do give a positive signal.

	aspect	n	positive	negative	neutral
0	ik	30	28	2	0
1	service	16	8	7	1
2	dit	14	8	6	0
3	zijn	13	11	1	1
4	het	13	13	0	0
5	klant	12	12	0	0
6	klantenservice	12	4	6	2
7	dag	11	9	2	0
8	boek	9	5	3	1
9	pakket	8	7	1	0

Figure 3: Top 10 most used aspects for Bol.com.

	aspect	n	positive	negative	neutral
0	levering	24	20	4	0
1	ik	19	16	3	0
2	product	18	16	1	1
3	service	17	10	7	0
4	klantenservice	15	7	3	5
5	prijs	14	13	0	1
6	het	13	9	4	0
7	zijn	12	8	3	1
8	bestelling	10	6	3	1
9	veel	10	8	2	0

Figure 4: Top 10 most used aspects for Agradi.

	aspect	n	positive	negative	neutral
0	levering	32	31	0	1
1	service	20	18	1	1
2	ik	14	13	1	0
3	prijs	10	10	0	0
4	leveren	10	9	0	1
5	site	8	8	0	0
6	verpakken	8	7	0	1
7	kleding	7	7	0	0
8	kwaliteit	7	7	0	0
9	aanbieding	7	6	0	1

Figure 5: Top 10 most used aspects for ABOUT YOU.

ACKNOWLEDGEMENTS

This paper has been inspired on the MSc master project of Joël Vergeer who was involved via the master Digital Driven Business at HvA. Thanks go to Selmar Meents as well as several anonymous reviewers for providing some useful suggestions to an initial version of this manuscript. Rob Loke is assistant professor data science at CMIHvA.

REFERENCES

Barnett, M. L., Jermier, J. M., & Lafferty, B. A. (2006). Corporate reputation: The definitional landscape. Corporate reputation review, 26-38.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 993-1022.

Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser Using Neural Networks. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 740-750.

Chen, Y., & Skiena, S. (2014). Building sentiment lexicons for all major languages. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 383-389.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 345-354.

Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. Tourism management, 58-67.

Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data.

Farhadloo, M., & Rolland, E. (2013). Multi-class sentiment analysis with clustering and score representation. 2013 IEEE 13th international conference on data mining workshops (pp. 904-912).

Fombrun, C. J., Ponzi, L. J., & Newburry, W. (2015). Stakeholder tracking and analysis: The RepTrak® system for measuring corporate reputation. Corporate Reputation Review, 3-24.

GfK. (2019, September 19). Quarterly number of online purchases in the Netherlands from 1st quarter 2015 to 2nd quarter 2019 (in million units) [Graph]. Retrieved April 8, 2020, from Statista: https://www.statista.com/statistics/556353/quarterly-number-of-online-purchases-in-the-netherlands/

Gotsi, M., & Wilson, A. M. (2001). Corporate reputation: seeking a definition. Corporate communications: An international journal, 24-30.

Hall, R. (1993). A framework linking intangible resources and capabiliites to sustainable competitive advantage. Strategic management journal, 607-618.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. New York: Springer.

Hernández-Rubio, M., Cantador, I., & Bellogín, A. (2019). A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. User Modeling and User-Adapted Interaction, 381-441.

Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A brief survey of text mining. Ldv Forum, 19-62.

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2014). Interactive topic modeling. Machine learning, 423-469.

Korenius, T., Laurikkala, J., Järvelin, K., & Juhola, M. (2004). Stemming and lemmatization in the clustering of finnish text documents. Proceedings of the thirteenth ACM international conference on Information and knowledge management, 625-633.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. Electronic Commerce Research and Applications, 205-217.

Lange, D., Lee, P. M., & Dai, Y. (2011). Organizational reputation: A review. Journal of management, 153-184.

Liu, B. (2012). Sentiment analysis and opinion mining. Morgan & Claypool Publishers.

Liu, Y., Liu, Z., Chua, T. S., & Sun, M. (2015). Topical word embeddings. Twenty-Ninth AAAI Conference on Artificial Intelligence.

Lovins, J. B. (1968). Development of a stemming algorithm. Mechanical Translation and Computational Linguistics, 22-31.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, 55-60.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 9603-9611.

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. Computational linguistics, 9-27.

Roberts, P. W., & Dowling, G. R. (2002). Corporate reputation and sustained superior financial performance. Strategic management journal, 1077-1093.

Statista. (2019). Online Reviews. Retrieved April 8, 2020, from Statista: https://www.statista.com/study/50566/online-reviews/

Trustpilot. (2020, August 8). Over Trustpilot. Retrieved from Trustpilot: https://nl.trustpilot.com/about

Van den Bosch, A., Busser, B., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. LOT Occasional Series, 191-206.

Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Systems with Applications, 112834.

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: a rating regression approach. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, (pp. 783-792).

Weng, J., Lim, E. P., Jiang, J., & He, Q. (2010). Twitterrank: finding topic-sensitive influential twitterers. Proceedings of the third ACM international conference on Web search and data mining, (pp. 261-270).

Wilbur, W. J., & Sirotkin, K. (1992). The automatic identification of stop words. Journal of information science, 45-55.

Xing, F., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. Proceedings of the 22nd international conference on World Wide Web, (pp. 1445-1456).