MEANING OF SENTIMENT ANALYSIS FOR COMPANIES

Menno van Zaanen

Department of Communication and Information Sciences School of Humanities

Tilburg University, The Netherlands

mvzaanen@tilburguniversity.edu

Lydia Mutiara Dewi

Department of Accounting, Faculty of Economics

Parahyangan Catholic University, Indonesia

l.m.d.abigail@gmail.com

Abstract

People often ask others for product advice. Once, word-of-mouth (WOM) was, due to practical limitations, shared locally. Nowadays, WOM is shared online (eWOM), which has a much larger reach. As eWOM is publicly accessible (unlike WOM), it can be used as information on brand attitude. eWOM can be aggregated and assessed using sentiment analysis (identifying positive/negative messages). The assumption is that sentiment analysis illustrates people's brand perception. We investigate the relationship between sentiment analysis and brand perception. We collected tweets with sentiment information of eight brands in Indonesia using Twitter's built-in sentiment analysis over a week. Using these tweets, aggregated sentiment scores were computed. The scores were correlated with brand perception collected using questionnaires. 206 participants attributed scores to seven properties: Complaint handling, Design, Friendliness, Information, Marketing, Service, and Overall score. Either insignificant or correlations close to zero were found, so online sentiment does not correspond to offline brand perception.

Keywords: *word-of-mouth, artificial intelligence in business, sentiment analysis*

JEL Classification: *M29*

INTRODUCTION

People often want to make an informed choice when buying products. In the past, people could ask other people in their direct surroundings for their experiences. Before buying a product, one would ask friends or family for their opinion. This type of information sharing is called word-of-mouth (WOM). Due to the physical transfer of information, only a small group of people has a major influence on the perception of products or brands.

Nowadays, with the availability of social media, sharing information has become very easy. One may write a message and share it with the entire world. Product or brand experiences can also be shared using social media and, in fact, people distribute their opinions online. Sharing this type of information corresponds to an electronic form of WOM, hence called eWOM.

An interesting side-effect of eWOM is that everybody can access this information. Not only do people use this information, companies may also access this information. In fact, many companies are active on social media in different ways. For example, they interact with customers, resolve complaints, answer questions, perform online marketing, and perhaps even influence eWOM by adding positive remarks or reduce the impact of negative messages.

Companies also collect the information that people share online about the company. The amount of information people post (especially) about the more well-known companies is huge. This means that opinions about the company will be present in this data, which may be extracted and aggregated. The idea is that there is a relationship between opinions found in eWOM and what people really think about the company. If true, collecting opinions from eWOM is a relatively cheap way to collect people's opinions of the company.

The idea that brand attitude information can be extracted from social media data, relies on the assumption that a relationship between the social media sentiment and the actual brand attitude exists. Here, we investigate what this relationship looks like. We will analyse eWOM sentiment and relate this to different properties of companies in order to answer our research question: What is the relationship between online sentiment and properties of companies?

BACKGROUND INFORMATION

Word-Of-Mouth and Electronic Word-Of-Mouth

By definition, word-of-mouth (WOM) is the process of delivering information from one individual to another. In a business context, WOM information may involve opinions, reactions, or details which consumers share concerning a product or a service (Jansen et al., 2009). WOM is primarily based on connections and trust within close social networks such as families and friends (Jansen et al., 2009). Kotler & Amstrong (2016) indicate that consumers feel more confident with what their families, friends and associates recommend about a product than what commercial sources inform by applying various modes of advertisements. Both authors even confirm that 92% of consumers take into account suggestions from families and friends more than that of any form of advertising. Hence, the WOM influence provides a major impact in consumer buying behavior. The WOM influence itself can be described as how the personal recommendations of families, friends, other consumers, and associates give a contribution to the consumer buying behavior (Kotler & Armstrong, 2016).

The development of social networks allows consumers to easily share and receive information in a large scope. They can exchange opinions about products, services, or companies on an online basis (Duan et al, 2008). This new type of word-of-mouth is known as online word-of-mouth (oWOM) or electronic word-of-mouth (eWOM) (Jansen et al., 2009).

eWOM, people collecting and sharing opinions through an online channel, has been on-going since the introduction of Web 2.0 (van de Kruis, 2013). Web 2.0, a term invented by Dale Dougherty (O'Reilly, 2007), also called a read-write web (Gillmor, 2004), allows people to interact with each other, to collaborate, to contribute to creating online material, to personalize a website for their personal use, or to express their opinion (Hew & Cheung, 2013). Several services are available, allowing people to easily and promptly share information.

Microblogging

An example of a Web 2.0 technology is microblogging (Grosseck & Holotescu, 2008). Microblogging is a type of communication that allows people to share their information in relatively short messages. Several websites provide microblogging services. Twitter is one of the most popular microblogging tools[1] (Java et al., 2007).

On Twitter, people can share messages known as tweets. The maximum length of each tweet is 140 characters (Go et al., 2009). Tweets have several specific attributes. It is possible to direct messages directly to other users, add hashtags, which allow for the categorising of tweets or simplifying search for topics, retweeting resends an existing tweet, which allows for, for example, forwarding the tweet to other users. Also, tweets often contain emoticons, which are used to indicate people's feelings. Examples of emoticons are ':)' and ':(', which represent positive and negative feelings respectively (van de Kruis, 2013; Aisopos et al., 2011).

Reviews on Twitter are different from posts on review websites such as Cnet[2] (van de Kruis, 2013). Due to the length limitations, people shorten words, e.g., 'great' is shortened to 'gr8' and 'congratulations' to 'congratz'. Additionally, many tweets do not follow standard spelling or grammar (Aisopos et al., 2011). These special attributes may influence the performance when analysing tweets.

To indicate the amount of information found on Twitter, 500 million tweets are sent on average every day by 317 million monthly active users in January 2017 (Aslam, 2017). In 2016, on average five businesses are followed by each Twitter user and 80% of the users mentioned a brand in their tweets. Between 2014 and 2016, the number of conversations between users and companies increased 2.5 times, and if the company gave a friendly service, 76% percent of the users are likely to recommend the brand (Smith, 2016).

A huge number of tweets contain opinions about products and brands and hence are interesting for companies. However, analysing each individual tweet manually is too time-consuming. Analysing huge amounts of tweets is possible sentiment analysis tools, such as

---

[1] https://www.twitter.com/
[2] https://www.cnet.com/

Engagor[3], that automatically detect people's opinions. Relevant tweets are analysed to identify the users' opinions, which are aggregated providing information about the valence contained in the tweets. These results can be useful for the company, for example, for customer care or for deciding the next marketing strategies (van de Kruis, 2013).

Sentiment Analysis

Liu (2010) defines sentiment analysis or opinion mining as 'the computational study of people's opinions, appraisals, and emotions toward entities, events and their attributes' (p. 1). Sentiment analysis research can be applied to many internet platforms (van de Kruis, 2013), such as: blogs, social media sites (including Facebook, Youtube, Twitter, and Pinterest), virtual worlds (such as Second Life and Everquest) (Kotler & Armstrong, 2016), websites that feature reviews (Dave et al., 2003; Hu & Liu, 2004), chat rooms, news articles (Morinaga et al., 2002; Nasukawa & Yi, 2003), and message boards (Kotler & Armstrong, 2016; Das & Chen, 2007). These tools analyse the text provided by users and provide information on the valence, or the sentiment, of the users, often in the form of a value on the positive versus negative scale.

While sentiment analysis tools aim to detect people's opinions on social media sites and represent the opinions as sentiment scores, it is still not clear how well the tools perform and how they represent people's real opinion about a certain brand (van de Kruis, 2013). Moreover, Kotler and Armstrong (2016) mention that even though huge amounts of opinions can be collected on the Internet and specifically on social media, yet 90 percent of the opinions are shared in the traditional way: WOM (Kotler and Armstrong, 2016). Therefore, we need to compare the (online) sentiment scores with (offline) brand attitude to investigate how they relate to each other.

Brand Attitude

Brand attitude is the overall evaluation of a brand regarding its ability to meet the consumer's expectation (Percy and Rossiter, 1992). This evaluation can be either positive or negative (Mitchell and Olson, 1981), although there are many scales to measure brand

---

[3] http://cxsocial.clarabridge.com/

attitude in the marketing literature. For example, Bruner (2012) uses nine-point Likert-type items to measure brand attitude (Bruner, 2012), whereas van de Kruis (2013) uses ten-point-Likert scales. Here, we rely on the definition of attitude by Krech et al. (1962): an 'enduring system of positive or negative evaluations, emotional feelings, and pro or con action tendencies with respect to social objects' (1962, p. 139). This definition can be converted into properties that, combined, measure brand attitude (van der Kruis, 2013).

1. Overall appreciation: evaluation of a certain brand or product.

2. Rating of the design: evaluation of the design of a product.

3. Rating of the service: general perception of how the company provides service (i.e., handling sales transactions properly, providing proper after sales service, and so forth).

4. Rating of customer friendliness: evaluation of the friendliness of the personnel.

5. Rating of complaint handling: evaluation of how complaints are handled.

6. Rating of information provision: evaluation of the product, service, or company procedure information provided by the company.

7. Rating of the marketing and communication: evaluation of how the company communicates and positions its products (i.e., an attractive advertisements, etc.).

METHODOLOGY

To investigate the relationship between the offline and online sentiment of companies, we ask people their opinion of certain properties of the companies under consideration and collect sentiment data about these companies from social media. The offline sentiment is collected through questionnaires, whereas the online sentiment comes from sentiment analysis of tweets.

As we do not know which properties people refer to when sharing the sentiment online, we asked people in the questionnaire to indicate their opinion on several properties: overall appreciation of the brand or product (overall), product design (design), service of the brand (service), customer friendliness (friendly), complaint treatment (complaint), provision of information (information), and marketing and communication (marketing). In total, 206 people participated in the questionnaire.

During the same period as collecting the offline sentiment information, we collected tweets of the same companies. Tweets that contain the name of the company were downloaded using the Twitter API (https://dev.twitter.com/). We separated these tweets into positive and negative categories based on the basic sentiment analysis tool of Twitter, which relies on the identification of positive and negative emoticons (':)' and ':('). Tweets that originated from the companies are not taken into account.

Based on the counts, we computed sentiment scores in six different ways. For the 'pos' and 'neg' metric, we only use the absolute counts of the positive and negative tweets, respectively. The 'sum' metric is the sum of the 'pos' and 'neg' metrics, whereas the 'diff' metric is the difference of the 'pos' and 'neg' metrics. These metrics denote absolute counts, but the number of tweets may be different per company, so we also compute two normalized metrics: 'pos_norm', which is the 'pos' metric divided the 'sum' metric, whereas 'diff_norm' is the 'diff' metric divided by the 'sum' metric.

Remember that we expect there to be a relationship between the sentiment score based on the tweets and the questionnaire results. To be able to compute correlations, we need several measurements, which are obtained by collecting sentiment values (offline and online) of several companies. We selected eight well-known companies that are active in Indonesia: Starbucks, JCo, Telkomsel, XL Axiata, TIKI, JNE, Honda, and Toyota. These companies were selected because most people know them and we expect there to be enough tweets to reach a reliable sentiment score.

RESULTS

The results of the questionnaires can be found in Table 1. The cells contain the mean and standard deviation values (between brackets) for each of the properties and brands. The values are in the range between one and ten, where ten is best and one is worst. We see that in general, the brands receive positive scores for all properties. Also, the standard deviations are low, indicating that people tend to agree on these values.

Table 1. The numbers denote averages and standard deviations (within brackets) for each of the properties for each of the brands. The numbers fall in the range between one and ten with ten being best.

Brand	Complaint	Design	Friendly	Information	Marketing	Service	Overall
Starbucks	7.26 (1.47)	7.79 (1.31)	7.96 (1.29)	7.49 (1.45)	8.04 (1.41)	7.92 (1.25)	8.00 (1.25)
JCo	6.92 (1.37)	7.92 (1.17)	7.37 (1.25)	7.06 (1.37)	7.17 (1.47)	7.44 (1.26)	7.79 (1.18)
Telkomsel	7.02 (1.75)	7.34 (1.44)	7.52 (1.45)	7.38 (1.58)	7.78 (1.36)	7.34 (1.69)	8.12 (1.16)
XL Axiata	6.76 (1.50)	7.29 (1.47)	7.02 (1.38)	7.05 (1.40)	7.41 (1.36)	7.06 (1.46)	7.25 (1.39)
TIKI	6.67 (1.34)	7.17 (1.15)	6.87 (1.21)	6.77 (1.39)	6.62 (1.58)	7.10 (1.29)	7.38 (1.15)
JNE	6.78 (1.58)	7.61 (1.27)	7.06 (1.37)	7.13 (1.40)	7.25 (1.45)	7.44 (1.30)	7.93 (1.17)
Honda	7.60 (1.27)	8.15 (1.17)	7.78 (1.16)	7.70 (1.27)	8.05 (1.20)	7.86 (1.19)	8.32 (1.07)
Toyota	7.62 (1.30)	8.03 (1.12)	7.82 (1.13)	7.76 (1.25)	8.05 (1.18)	7.96 (1.16)	8.22 (1.10)

The sentiment scores that are computed based on tweets for each of the brands can be found in Table 2. We collected tweets for one week. The numbers of tweets that contain sentiment are found in the columns 'pos' for positive and 'neg' for negative. The 'sum' and 'diff' columns contain the sum (total number of tweets containing sentiment) and difference (between the positive and negative sentiment) respectively. Additionally, normalized counts, found in the 'pos_norm' and 'diff_norm' columns, are presented. These two columns are investigated to make sure that the differences in absolute numbers do not have a major influence on the sentiment value.

Table 2. The numbers represent absolute counts of tweets per brand (in the pos, neg, sum, and diff columns) or normalized scores (in the norm_pos and norm_diff columns).

Brand	pos	neg	sum	diff	pos_norm	diff_norm
Starbucks	10	10	20	0	0.500	0.000
JCo	3	6	9	-3	0.333	-0.333
Telkomsel	4481	71	4552	4410	0.984	0.969
XL Axiata	3	1	4	2	0.750	0.500
TIKI	31	9	40	22	0.775	0.550
JNE	212	32	244	180	0.869	0.738
Honda	57	11	68	46	0.838	0.676
Toyota	20	1	21	19	0.952	0.905

Given that we expect there to be a correlation between the online and offline sentiment scores, we compute the Pearson's product moment correlation between the offline and online sentiment scores. This should result in significant positive correlations which means that if the online sentiment is high, the offline sentiment should also be high and vice versa for low sentiment.

Computing the correlations, we see that this only leads to significant correlations for some of the combinations. All significant correlations ($p<.05$) can be found in Table 3. We see that even when a significant correlation can be found, the correlation coefficient is very close to 0, meaning that no linear correlation between the variables can be found.

Table 3. Pearson's correlations between the online (columns) and offline (rows) sentiment results. Missing numbers indicate that no significant correlation was found.

Property	pos	neg	sum	diff	pos_norm	diff_norm
Complaint	-	-	-	-	-	-
Design	-.093	-.090	-.093	-.093	-.053	-.053
Friendly	-	-	-	-	-	-
Information	-	-	-	-	.068	.068
Marketing	.057	-	.057	.057	.072	.072
Service	-.049	-	-.049	-.049	-	-
Overall	.077	.091	.077	.077	.079	.079

DISCUSSION AND CONCLUSION

In this research, we investigate the relationship between sentiment collected from online, social media networks, like Twitter, and offline sentiment, which corresponds to the brand attitude of people 'in real life'. We collected sentiment values for a week from tweets for eight well-known brands in Indonesia and at the same time asked people for their brand attitudes of the same brands through questionnaires.

We expected there to be a correlation between the online and offline sentiment values, although the nature of this correlations was unknown. Therefore, we asked people to provide

their brand attitude for seven different properties of brands. This allowed us to investigate potential correlations between the online sentiment and each of the brand properties.

The results showed that no correlation between the online sentiment values and any of the seven offline sentiment values can be found. There may be several explanations for this. Firstly, the data collection may have introduced problems. The results of the questionnaire do not show much variation, so finding a correlation is difficult. Also, the results of the online sentiment may be problematic. For example, the identification of the sentiment of a tweet might have been unreliable and more advanced sentiment analysis tools for Indonesian may be required. Alternatively, one week might not have been enough time to collect reliable data.

Another problem may be that correlations do exist, but are not consistent over the brand properties between companies. In other words, it might be the case that for one company the online sentiment scores correlate with one property and for another brand they correlate with another. If we want to measure this, we will need to collect data for each of the different companies (both online and offline) over a longer period of time, allowing us to compute correlations over time. The fact that different types of companies are used in this research may have influenced this.

To summarize, companies use online sentiment analysis tools to measure brand attitude. However, we have shown that the interpretation of the sentiment scores is more complex than simply assuming that this is indicative of the overall brand attitude. Either more research into the reliability of the sentiment analysis tools needs to take place, or a more in depth analysis of what people write in tweets is required to link the results to specific brand properties. Online sentiment analysis may provide a useful tool, but care has to be taken when used.

REFERENCES

Aisopos, F., Papadakis, G., and Varvarigou, T. (2011), Sentimengt Analysis of Social Media Content using N-Gram Graphs. In Proceedings of the 3rd ACM SIGMM International Workshop on Social Media, ACM pp. 9-14

Aslam,S. (2017, 1 24). Retrieved from Omnicore: https://www.omnicoreagency.com/twitter-statistics/

Bruner, G.C. (2012), Marketing Scales Handbook: A Compilation of Multi-item Measures for Consumer Behavior & Advertising Research, Vol. 6, GCBII productons, Carbondale.

Das, S. R., & Chen, M. Y. (2007), Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management Science, 53(9), pp. 1375-1388

Dave, K., Lawrence, S., and Pennock, D. M. (2003), Mining The Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceeding of the 12th International Conference on World Wide Web, ACM, pp. 519-528

Duan, W., Gu, B., and Whinston, A. B. (2008), Do Online Reviews Matter? - An Empirical Investigation of Panel Data. Decision Support Systems, 45(4), pp. 1007-1016

Gillmor, D. (2004), We the Media: Grassroots Journalism by the People, for the People, O'Reilly Media, Sebastopol, California.

Go, A., Bhayani, R., and Huang, L. (2009), Twitter Sentiment Classification using Distant Supervision. CS224N Project Report, Stanford, pp. 1-12

Grosseck, G., and Holotescu, C. (2008), Can We Use Twitter for Educational Activities? In 4th International Scientific Conference, eLearning and Software for Education, Bucharest, Romania.

Hew, K. F., and Cheung, W. S. (2013), Use of Web 2.0 Technologies in K-12 and Higher Education: TheSearch for Evidence-Based Practice. Educational Research Review, 9, pp. 47-64

Hu, M., and Liu, B. (2004), Mining and Summarizing Customer Reviews. In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp. 168-177

Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. (2009), Micro-blogging as Online Word of Mouth Branding. In Proceeding of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, ACM, pp. 3859-3864

Java, A., Song, X., Finin, T., and Tseng, B. (2007), Why We Twitter: Understanding Microblogging Usage and Communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, pp. 56-65

Kotler, P., and Armstrong, G. (2016), Principles of Marketing, 16th ed., Pearson, Boston.

Krech, D., Crutchfield, R. S., and Ballachey, E. L. (1962), Individual in Society, McGraw-Hill, New York.

Kruis, S. van de. (2013), Sentiment Analysis on Tweets Measuring the Correctness of Sentiment Analysis on Brand Attitude. Master Thesis Communication and Information Sciences Tilburg University, pp. 1-58

Liu, B. (2010), Sentiment Analysis: A Multi-Faceted Problem. IEEE Inteeligent Systems, 25(3), pp. 76-80

Mitchell, A. A., and Olson, J. C. (1981), Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude? Journal of Marketing Research, pp. 318-332

Morinaga, S., Yamanishi, K., Tateishi, K., And Fukushima, T. (2002), Mining Product Reputations on The Web. In Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 341-349

Nasukawa, T., and Yi, J. (2003), Sentiment Analysis: Capturing Favorability using Natural Language Processing. In Proceeding of the 2nd International Conference on Knowledge Capture. ACM, pp. 70-77

O'Reilly, T. (2007), What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. COMMUNICATIONS & STRATEGIES, 65(1), pp. 17-37

Percy, L., and Rossiter, J. R. (1992), A Model of Brand Awareness and Brand Attitude Advertising Strategies. Psychology & Marketing, 9(4), pp. 263-274.

Smith, K. (2016, 5 17). Brandwatch. Retrieved from Brandwatch Blog: https://www.brandwatch.com/blog/44-twitter-stats-2016/