

Textual Variations Affect Human Judgements of Sentiment Values

Phoey Lee Teh¹, Paul Rayson², Irina Pak¹, Scott Piao², Jessica Sze Yin Ho³, Andrew Moore² and Yu-N Cheah⁴

¹Sunway University, Sunway City, Selangor 47500 Malaysia.

²Lancaster University, Lancaster, LA1 4YW United Kingdom.

³Heriot-Watt University Malaysia, Putrajaya, Selangor 63100 Malaysia.

⁴Universiti Sains Malaysia, Pulau Pinang, 11800 Malaysia.

Corresponding author: Phoey Lee Teh (e-mail: phoyleet@sunway.edu.my).

Electronic word-of-mouth communication in the form of online reviews influences people's product or service choices. People use text features to add or emphasise feelings and emotions in their text. The text emphasis can come in as capital letters, letter repetition, exclamation marks and emoticons. The existing literature has not paid sufficient attention to the effects of such textual variations on human text interpretation. This paper presents an analysis of text variations that can affect the interpretation of a text. A total of 1,041 online comments were collected, in which seven types of the most used textual variations were identified and simulated for hypothesis testing. Sentiment scores from 500 participants were collected to rate the value expressed for each of the textual variations. Statistical analysis showed that collected ratings are significant for the accurate calculation of sentiment values for short comments. Furthermore, the performance of ten existing sentiment tools was analysed based on seven textual variations. Results indicate that those tools should consider these textual variations to fully reflect a human interpretation on the text variations.

Keywords: Sentiment analysis, text analysis, classification algorithms, text mining, NLP tools, computer mediated cues (CMC), punctuation.

This work was supported by the Sunway-Lancaster Grant SGSSL-FST-DCIS-0115-11 and Universiti Sains Malaysia Research University Grant 1001/PKOMP/8014002

Textual Variations Affect Human Judgements of Sentiment Values

1 Introduction

Research on Electronic word-of-mouth (EWoM) or social recommendation has become a hot topic in data science, market research and social media listening. EWoM or social recommendation in this paper refers to the comments and reviews posted by consumers online on social media. Recent research beyond natural language processing has revealed the importance of analysing EWoM communication. For instance, Lin and Shen (2011) specified that an analysis of consumers' buying behaviour is important to help explain: 1) the emotional motivation of the consumer, 2) the rational motivation as a means of understanding cost-effective and affordable pricing, and 3) the purchase motivation when a consumer has unique preferences. EWoM, therefore, provides a significant value in understanding the purchasing decisions for products and services.

Multiple general emotion classifications and labelling frameworks are available for mark-up in corpora (Mori et al., 2016), however, this paper focuses on the sentiment dimension, which has received much attention in the computational linguistics literature. Sentiment analysis is defined as a common task in Natural Language Processing (NLP) and a process of identifying the evaluative nature of the text (Alharbi and de Doncker, 2019; Kiritchenko et al., 2014; Tang et al., 2020). To date, most of the social media text analysis techniques exploited by commercial organisations are simplistic and fail to take account of the entire linguistic complexity of the text. Other studies discussed those textual variations as well. For instance, Carey (1980) in his study of paralanguage, initially referred to these different forms as “grammatical markers”, Burgers et al. (2012) in their communication study of message irony referred to them as “typographic markers”, while Kalman and Gergle (2009) referred to them as computer-mediated communication (CMC) cues or “nonverbal cues”.

Moreover, “grammatical markers” and “typographic markers” can include letter capitalisation, letter repetition, emoticons and different types of punctuation. In a more recent study of (Jianqiang et al., 2018), term “syntactic features” was used for exclamation marks and emoticons and so on. Moreover, it was stated that researchers aim to determine sentiment value though extracting those features (Jianqiang et al., 2018).

Unlike verbal or face-to-face communication, emotions and feelings in online reviews are conveyed through the text form alone. In the text, the use of these textual variations replaces non-verbal communication cues which would otherwise appear as body language, facial expressions or tone of voice.

In an online review setting, the analysis of short, informal and unformatted text can be complex because people tend to use multiple textual variations to highlight or emphasise different meanings of expressions. For human readers, the capitalisation of letters in a text can signify extra attention or voice volume, repeated letters tend to strengthen or emphasise the emotion, and emoticons can help visualise the facial expressions of the person. All of these features can affect the understanding of the sentiment of a given text. Presently, no detailed study on the scale of the effects of the textual variations on human readers in the context of sentiment analysis has been conducted.

This study proposes that the sentiment value of comments from social media should not be normalised before being analysed by automatic systems and should preserve the original text form to avoid losing key information related to sentiment. Textual variations and additions that are expressed in various forms of text markers (e.g. symbols, punctuations, capital letters and emoticons) carry extra information that should be taken into account to accurately determining sentiment values. This work demonstrates this effect through an experiment with 500 participants who manually rated the value of scale in agreeing on the expression level for different textual variations. In addition, this study gauges the extent to which the various text

markers, when used in combination with plain text, can affect the sentiment and value of comments regarding different online sentiment analysis tools. To test if current tools are able to sufficiently take textual changes into account when considering its sentiment, ten sentiment tools available online were tested with a set of non-emphasised textual comments and simulated seven types of variations, which contain capital letters, emoticons, exclamation marks and repeated letters. For example, “I loooove this dress” and “I love this dress!!!!!!” in comparison with its non-emphasised version “I love this dress”. The scoring for the text with different text variations was compared and analysed in all tools.

The rest of the paper is structured as follows. Section 2 summarises the previous work on textual variations on CMC, including how they have been discussed in the sentiment analysis literature. In Section 3, data collection and survey design were described, and the method in which the existing sentiment tools were used in our experiments were explained. Section 4 shows the survey outcome and the comparative evaluation of the performance of the sentiment tools. Finally, the closing concludes with a summary and future work.

2 Literature review

Textual formats play an important role in communication in psychology, linguistics, and natural language processing. Studies on textual variations and emphases predate the use of non-standard varieties in online contexts. In historical (Early Modern English) texts before the standardisation of spelling and publication of dictionaries, there were many reasons for spelling differences, including alignment and justification of the page by typesetters and printers. Archer et al. (2015) performed the first large-scale quantitative survey of spelling variations in Early Modern English. They showed that a majority of word types were non-standard in 16th century English. Later, Carey (1980) carried out one of the first studies of variation patterns in CMC, including the use of capital letters, letter repetitions, vocal (phonetic) spellings,

punctuations and so on, and referred to different textual variations as “grammatical markers”. People use them in (what were called at the time) computer conferencing systems to emphasise or highlight statements and opinions within the text. Burgers et al. (2012), on the other hand, referred to some textual variations (e.g. capital letters, emoticons, punctuation and so on) as “typographic markers”. They stated that typographic markers could draw attention to ironic statements. In many sentiment analysis systems (Naradhipa and Purwarianti, 2011), these variations tend to be treated as a problem that needs to be fixed through pre-processing before sentiment classification on the “cleaned” text can begin. Moreover, Kim et al. (2014) hypothesised that users tend to use words in capital letters to emphasise a particular point or word. They observed that capital letters, emoticons, abbreviations and exclamation marks are used in online chat situations to give non-verbal signals. From the natural language processing approach, Neviarouskaya et al. (2011) stated the interpretation (meaning) behind text messaging in online communication created a model which could be used to infer the emotional state of the person writing the message.

There are numerous other studies that address the importance of sentiment analysis. For instance, Pang et al. (2002) noted the value of sentiment classification and reviewed the type of machine language classifiers such as Naive Bayes, Maximum Entropy and Support Vector Machines in comparing movie reviews. Thelwall and Wilkinson (2010) analysed social media comments and highlighted the importance of emotional expressions in the informal text in blogs and online forums. In an experiment with MySpace comments, they predicted positive and negative emotions using the SentiStrength tool and described the role of emotions in free text language. Furthermore, Prabowo and Thelwall (2009) proposed an approach of dividing text using sentiment analysis based on consumer reviews, feedback and comments. They highlighted the importance of classifying text within a three-way sentiment model with positive, negative and neutral categories. Their findings showed that Statistics-Based Classifier

had improved the effectiveness of hybrid classification, which was not the case with General Inquirer-Based Classifier. Also, Thelwall et al. (2012) analysed free language text by extracting its sentiment value using SentiStrength, a revised version of the tool that can identify the positive or negative value of the text. Lochter et al. (2016) studied short text opinion detection in online social networks and surmised that the mining of online opinions may be helpful for marketing and sales departments in companies.

They managed noisy texts containing spelling variations by normalising the text using a dictionary look-up approach and then applying ensemble sentiment methods to clean text. Such normalisation resulted in losing information from the textual variations. In another study (He et al., 2012), text reviews were collected and analysed to observe the behaviour of users who are in an enterprise network; the usefulness of such analysis is to monitor and detect the misuse of the workplace network. Likewise, Zin et al. (2012) also built a behavioural analysis tool to study consumer behaviour. Cobb et al. (2013) conducted experiments to understand how a consumer's decision-making process depends on online reviews and messages. They discovered that more positive thoughts about products lead to higher demand for the product. Xue et al. (2014) proposed a dictionary model that can classify text into positive, negative or neutral categories. The sharing of opinions online is widely evident in hotel, movie and restaurant reviewing sites (Kotelnikov and Pletneva, 2016). In a survey study of 47 academic papers, Mohey and Hussein (2016) highlighted the challenges in sentiment analysis, such as the nature of the topic and the review structure. They have determined that one of the main challenges in sentiment analysis is domain dependence. The solution they proposed for this problem is to find the relationship between the proportion of sentiment techniques usage in theoretical and technical types of challenge (Mohey and Hussein, 2016).

Further, they reflected the average results on each type of challenges (Mohey and Hussein, 2016). In a more recent study (Smetanin, 2020), it was mentioned that sentiment

analysis could measure the reaction to certain news or event. Liang and Dai (2013) discussed the significance of analysing microblogs using sentiment values due to microblogging sites such as Twitter and Facebook. They developed an automated system that is able to extract messages from social media and classify sentiment into positive and negative values. The main issue they have tried to solve is to deal with the short social media messages, which was a challenging task for systems to handle (Liang and Dai, 2013). Table 1 summarises studies that examined different elements of variation in text-based communication.

Table 1 Different Elements Of Variation In Text-Based Communication

Variation	Example	Reference
Capitalisation	THIS PHONE IS AMAZING This phone is AMAZING	(Almunawar et al., 2013; Byron, 2008; Carey, 1980; DeWald and Geyer, 1975; Ledbetter and Larson, 2008; Pak and Teh, 2018; M. A. Riordan and Kreuz, 2010; Schomaker and Bulacu, 2004; Sesa-Nogueras and Faundez-Zanuy, 2012; Vandergriff, 2013)
Exclamation marks	This phone is amazing!!! This phone is amazing!!!!!!	(Godin et al., 2013; Kalman and Gergle, 2014; Mohammad et al., 2013; Teh et al., 2015b; Zavattaro et al., 2015)
Letter repetition	This phone is amaaaazing	(Brody and Diakopoulos, 2011; Derks et al., 2007; Garrison et al., 2011; Huang et al., 2008; Kalman and Gergle, 2014; Mohammad et al., 2013; Pak et al., 2018; Tossell et al., 2012)
Emoticons	This phone is amazing :-) This phone is amazing :-((Brody and Diakopoulos, 2011; Byron, 2008; Carey, 1980; Derks et al., 2007; Garrison et al., 2011; Hogenboom et al., 2013; Huang et al., 2008; Kalman and Gergle, 2014; Mohammad et al., 2013; Ted et al., 2010; Teh et al., 2016; Tossell et al., 2012; Urabe et al., 2013)

2.1 Capitalisation

In the studies on letter capitalisation, the use of capital letters is found to indicate vocal or/and emotional expression. According to Byron (2008), the use of capital letters indicates emotional intensity. Vandergriff (2013) used a pragmatics perspective to analyse the use of emotions in computer communication of advanced foreign language learners. According to the microanalytic approach (Ledbetter and Larson, 2008), letter capitalisation might be used to indicate “increasing volume” (equivalent to raising one’s voice in verbal communication). The use of capital letters can underline meaning and characterise signs of sarcasm. The authors

stated that fully capitalised text is interpreted as shouting and often unwelcome. Similarly, M. a. Riordan and Kreuz (2010) have also discussed the possibility of capitalisation to reflect a different meaning in a text compared to lower case text. The study by Mohammad et al. (2013) also acknowledged the presence of one of the textual variations in their system, they were able to detect the number of words which are capitalised. However, whether or not the usage of capitalisation expresses a happy mood or unwelcome emotions, it has an impact on the reader's opinions when a comment or product review is read online. In conclusion, letter capitalisation is clearly used to reinforce the expression and underline the meaning of texts (Pak and Teh, 2018).

2.2 Exclamation Marks

One of the most widely used punctuation characters is exclamation marks. There are quite a number of studies which stated that exclamation marks express different emotional value. For instance, Carey (1980) also noted that grammatical markers, in this case, exclamation marks, help convey the tone of voice and indicate pauses and emphasis. However, the study by (Carey, 1980) was conducted in 1980, and there were no mechanisms available at that time for automatically identifying grammatical markers. In the study of (Teh et al., 2015b), the number of exclamation marks has been tested in online sentiment tools to identify if they can consider this in their sentiment scoring. Twelve sentiment tools were tested, and only two of them were found to consider the difference in the number of exclamation marks when calculating their sentiment scores. However, the results showed that the sentiment value is higher (more positive) for positive comments if more exclamation marks are used. In contrast, negative comments have a lower (more negative) sentiment value when they contain more exclamation marks. In the studies by (Kalman and Gergle, 2014, 2009) on the repetition of punctuation, they observed that the repetition of exclamation marks has the most extensive use among other observed punctuation in their dataset. They provided examples that show

exclamation marks being used multiple times in the same context (Kalman and Gergle, 2009). For instance, “Have a great trip!!!!!!” and “neat! Don’t worry; I’m excited to cook!!!! I’ll try to do it in advance!!!! but if i can’t, it’ll need to cook at your place!!!! that ok?!?!?!!!!!!! see you soon!!!! Best”. It can be concluded that exclamation marks can strengthen the sentiment value of the text (Kalman and Gergle, 2009). The study of Lin Su (2018) stated that exclamation marks could indicate danger and red flags. Moreover, Lin Su (2018) also noted that punctuation marks, including exclamation marks, have their own tonality and different semantic and mood expressions. Babii et al. (2020) mentioned that number of exclamation marks may serve as a marker for identifying emotion. A more recent study of Sidi et al. (2021) examined repeated exclamation marks in a simulation, where they presented email with repeated exclamation marks and asked participants to interpret them. They concluded that females in their study tended to express their emotions using repeated punctuations more than males in their cohort (Sidi et al., 2021). Moreover, the level of perceived sender happiness was higher for the participants who received the message with exclamation marks (Soderlund et al., 2021).

2.3 Letter Repetition

Carey, (1980) pointed out that the repetition of vowels may express an individual dialect or regional accent, for instance, “Weeeell”. That means letter repetition can represent prolonged pronunciations in some dialects. In other research, the formatting of different variations of text has been studied. For example, Kalman and Gergle (2014) conducted a study of the repetition of alphabetical letters by analysing Twitter microblogs and concluded that users tend to repeat punctuation characters in a text to express personalisation and more intention of messages. In addition, they mentioned that the usage of capital letters, asterisks, blank spaces or character repetitions, as well as combinations of these “gimmicks”, enhance and enrich the text. Meanwhile, the study of Mohammad et al. (2013) referred to words containing letter repetition

as elongated words. Their system was able to detect words with letters repeated more than two times, for example, “o” for the elongated word “soooo”. Brody and Diakopoulos (2011) discovered that the repetition of single letters occurs roughly in every six tweets on Twitter. They concluded that Twitter users tend to emphasise sentiment by repeating letters. There is a need to design methods that can consider letter repetition when calculating sentiment scores of text. Kalman and Gergle (2014) also stated that letter repetition indicates the stretching of a word to simulate the spoken conversation, for example, “It is sweeeeeet”. Letter repetition can also signify a change in pitch or lowering of one’s voice, for example, “Yeeeeeeehaaaw!!!!!!!!!!”, “Sshhhhhh.....let’s keep it between us”.

Moreover, they pointed out that letter repetition may represent musical intonation, such as “Happy birthday to youuuu Happy birthday to youuuu Happy birthday dear”. It is also postulated that letter repetition may replace vocal noises and sounds in face-to-face communication such as guttural sounds and laughter. For example, “WOOOOOOOOOOHOOOOOOOOOOOOOOOO, Daddy’s getting a new Blue Wave Bay boat!!!! WOOOHOOOO”, “And pffffff, he is away”, and “Heeeeeheeee!” (Kalman and Gergle, 2014). Several studies on letter repetition agreed that it could represent different vocal expressions.

In summary, studies have shown that letter repetition signifies emphasis or features which would alter the reader’s interpretation of sentiment associated with the text.

2.4 Emoticons

While letter repetition can signify vocal noises or dialectal emphasis in text, emoticons can symbolise facial expressions alongside the text and studies in [Table 1](#), emphasise that. Emoticons represent non-verbal face-to-face expressions. For instance, Derks et al. (2007) studied the use of emoticons in online interactions and found that the frequency of using positive emoticons is higher than the negative ones. Huang et al. (2008) pinpointed that

Deleted:

emoticons are a valuable addition to communication methods. They found that increased emoticons' usage tends to positively affect personal interactions, perceived usefulness, and information richness. Urabe et al. (2013) discussed how emoticons could effectively express non-verbal emotional tones and body language and proposed an emoticon recommendation system that helps users express their emotions using emoticons. Moreover, the studies of Hogenboom et al. (2013) and Teh et al. (2016) suggested that emoticons cannot only convey the sentiment nature of the text, but they also have the potential to reverse the polarity of a statement. The study of Novak et al. (2015) went further and generated an Emoji Sentiment Ranking to identify emotional content behind more than 700 frequently used emojis. Moreover, Tossell et al. (2012) highlighted that users tend to use emoticons to include socio-emotional contexts in their messages. Garrison et al. (2011) claimed that emoticons are a meaningful linguistic unit, and Byron (2008) and Pfeifer et al. (2022) stated that emoticons are equivalent to emotions and facial expressions in face-to-face communication. Emoticons such as emoji, stickers, GIFs and other animated elements convey the emotional component of user feedback (Artemenko et al., 2020).

From the literature above, it can be concluded that it is essential to analyse all these different types of variations in the text. People tend to use textual variations to express various forms of emphasis or vocal signals. Furthermore, it can be seen from previous studies that these variations in the written forms will impact readers' interpretation of the text.

3 Method

Our data collection stage consisted of two phases. Phase 1 focused on collecting comments on products and identifying the textual variations in the comments (positive and negative). Phase 2 focused on assessing the textual variations for sentiment values based on the different text variants identified in Phase 1.

3.1 Phase 1: Comments Collection and Identification of Textual Variants

Phase 1 of our study involved the collection of comments on commercial items from social media websites. A total of 1,041¹ comments containing 105,437 words from 10 different product categories were collected from a variety of social media platforms, including Amazon, Facebook, E-bay and GSM Arena. To ensure a wide comment coverage, the 10 product categories were selected from the main categories displayed on the Yahoo! homepage (at the time of data collection), namely “Mobiles and Tablets”, “Fashion”, “Jewellery and Watches”, “Cameras”, “Home Appliances”, “Consumer Electronics”, “Computers”, “Beauty and Health”, “Toys and Kids” and “Sporting Goods”. The existing dataset is a commonly cleaned version without variations. Using the Wmatrix corpus analysis system (Rayson, 2008) and text variation layers derived from (Teh et al., 2015a), the collected comments were examined to determine the judgement from participants on the variation of text from the most frequently used positive and negative terms. [Table 2](#), This process is performed with the tool Wmatrix (Rayson, 2008).

Deleted:

Table 2 Top 20 Commonly Used Emotional Phrases (10 Positives and 10 Negatives)

Positive	Negative
I love it	Some serious abuse
I like it	I did hit very well
I am very happy	Very disappointed
I am glad	I don't care
I am big fan	I hate it
My favourite	It is really annoying
Hours of fun	I boot it
Very satisfied	Too much trouble
I prefer it	I have to worry
Really enjoy	Totally fierce

To find out if different text variations could impact the level of expression of those emotional words, the top 20 commonly used emotional phrases were samples for later analysis. These

¹Collected comments are available at <https://github.com/UCREL/HumanJudgementsOfSentimentValues>

words are extracted from the original terms with their own concordance and texts are displayed in

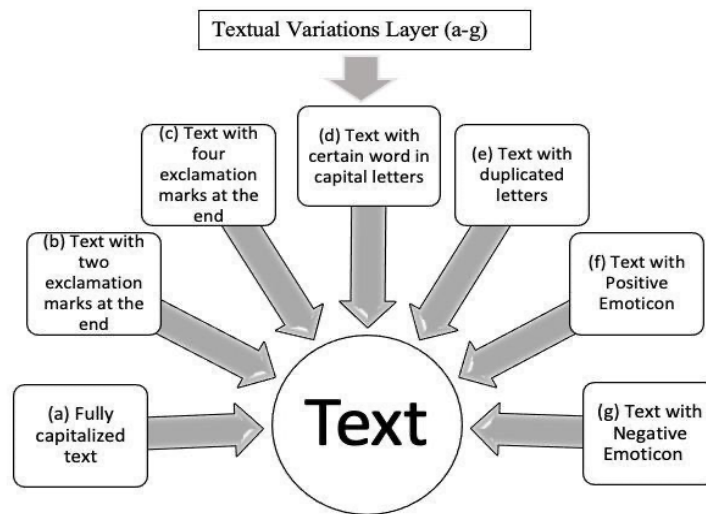


Figure 1 Textual Variations Layer

Twenty statements from [Table 2](#), were simulated in the questionnaire design described in Phase 2. Seven versions of each comment were derived, containing only standard text, using the seven major types of textual variations based on the model in Figure 1. The seven types of textual variations were categorised based on the manual analysis and observation of these comments.

As shown in Figure 1, these variations include:

- a) Fully capitalised text,
- b) Text with two exclamation marks at the end,
- c) Text with four exclamation marks at the end,
- d) Text with a word in capital letters,
- e) Text with duplicated letters,
- f) Text with positive emoticon,
- g) Text with negative emoticon

Deleted:

Our hypothesis is that the sentiment values of these seven types of variations differ. In other words, the textual variations augment the sentiment value of standard text comments, for both positive and negative comments (Table 2). This is further explained below:

1) H1a-H1g: For positive comments, the sentiment values expressed in layers (a)(b)(c)(d)(e)(f)(g) from the model in Figure 1 vary significantly from the sentiment values of comments without textual variations.

2) H2a-H2g: For negative comments, the sentiment values expressed in layers (a)(b)(c)(d)(e)(f)(g) from the model in Figure 1 vary significantly from the sentiment values of comments without textual variations.

3.2 Phase 2: Assessment of Sentiment Values

A questionnaire was designed to verify the human judgment of sentiment values of phrases extracted in Table 2. The reason for this step is to gather the opinions of what a reader may understand from the sentiment expressed written within the comments. The judgements provided by the participants were to rate within the scales from 1 for “Strongly dislike” to 7 for “Strongly like”. The ratings were performed for each of the variations in the textual variations layer. The questionnaire consists of three parts. In part one, the respondents were asked for general information about themselves. Such information was required to understand whether the respondents had any experience with online comments. Table 3, below summarises the respondents experience.

Table 3 Demographics

Variables		Number	%
Native English speaker	Yes	277	55.4%
	No	223	44.6%
Age	18 and below	117	23.4%
	19-35	381	76.2%
	36-50	0	0
	51 and above	2	0.4%
Gender	Male	243	48.6%
	Female	257	51.4%
Do you comment (give feedback/review) on global social product online?	Yes	310	62.0%
	No	190	38.0%

Which Social Media Platform have you commented?	Facebook.com	347	69.4%
	eBay.com	77	15.4%
	Amazon.com	34	6.8%
	Twitter.com	21	4.2%
	Other	21	4.2%
What is your purpose/intention of commenting (giving feedback/review) on products online?	Sharing experience	225	45.0%
	Giving warning to others	102	20.4%
	Expressing yourself	82	16.4%
	Giving suggestion	72	14.4%
	Other	15	3%
What is the category of product that you usually comment (giving feedback/review) ONLINE?	Mobile and Tablets	136	27.2%
	Fashion	148	29.6%
	Home Appliance	91	18.2%
	Computer	52	10.4%
	Jewellery and Watch	30	6.0%
	Toys and Kids	24	4.8%
	Camera	4	0.8%
	Consumer Electronics	4	0.8%
	Beauty and Health	11	2.2%

A total of 277 (55.4%) people responded that they are native English speakers, the rest 223 (44.6%) of respondents are not. The majority of respondents are between 19-35 years old with a frequency of 381 (76.2%), the rest 117 (23.4%) and 2 (0.4%) are 18 and below and 51 and above respectively. Among them, 243 (48.6%) are females and the rest are males. A total of 310 (62%) respondents agreed that they have experience in commenting on global social products online, and a lower number of people 190 (38%) responded that they do not comment. The most visited social media platform is Facebook with a frequency of 347 (69.4%), followed by eBay 77 (15.4%) and Amazon 34 (6.8%) and the same lowest number of 21 (4.2%) was recorded for Twitter and Other.

Among all the 500 respondents, the most popular reason for commenting is to share experience by providing feedback/review on products, which constitute a total of 45%, followed by 30.4% in providing warnings to others. Expressing themselves was recorded as the third most popular reason which constitutes a total of 16.4%. Finally, the remaining reasons are 14% to give suggestions, and 3% with other reasons and 4 respondents left questions unanswered. Three categories such as Fashion, Mobile and Tablets and Home Appliance were

selected as the top categories of product that people usually commented on, with frequencies of 148(29.6%), 136(27.2%) and 91(18.2%) respectively. The least popular categories are on Camera and Consumer Electronic with the frequency of 4(0.8%).

In order to test all the phrases with all the different variations, we require human raters to provide judgement, therefore, we gather from the participants the different ratings of judgements in terms of the variety of text format used for each of the phrases in [Table 2](#). We wish to discover if the usage of different text variations could actually affect the sentiment level of expressions. In the questionnaire, the sequencing of the positive and negative sets of phrases was randomly arranged to eliminate straight-lining and respondent bias when answering the questions, this method is named as the intercept method (Cohen et al., 2011). Each question was divided into eight sub-questions and simulated as follows:

- Unformatted text (e.g. I love it).
- A whole phrase in capital letters (e.g. I LOVE IT).
- A phrase that ends with two exclamation marks (e.g. I love it!!).
- A phrase that ends with four exclamation marks (e.g. I love it!!!!).
- A phrase with specific word(s) in capital letters (e.g. I LOVE it).
- A phrase with repeated letters (e.g. I loooooove it).
- A phrase with a positive emoticon (e.g. I love it :)).
- A phrase with a negative emoticon (e.g. I love it :().

To collect a broader scale of values, the answers were set to a 7-point scale with the following preferences: 1) strongly dislike, 2) dislike, 3) slightly dislike, 4) neutral, 5) slightly like, 6) like, and 7) strongly like. Results from a 7-point scale are considered to be more accurate than those from a 5-point scale (M. a. Riordan and Kreuz, 2010). Using the intercept methods (Cohen et al., 2011), 500 university students were randomly and personally approached by respondents (Teh et al., 2015a). To participate in this study, the respondents must fit the criteria of having

Deleted:

prior experience in using social media and being fluent in English. The respondents were asked to rate the customer comments in the simulated variations based on whether the comments showed a liking or disliking for a product. Results for this experiment are presented in Section 4.1.

3.3 Phase 3: Experiment with Sentiment Tools

Ten² freely accessible online sentiment tools were evaluated in this study. Sentiment analysis tools are a type of Natural Language Processing (NLP) tool (Wu et al., 2007) which examine the value of text and determines the overall emotional level. In general, each tool produces a score for the text based on the different methods, algorithms, and lexicons used, and so on. For instance, some tools rate the text based on polarity categories (e.g. positive, negative or neutral); others produce numeric scores that can be varied in range. The criterion for selecting those tools is that they can process short text messages similar to online comments/reviews. The software code that we have written can automatically extract the score from these tools and is now available online. Other tools needed to be downloaded, and a licence required. Different sentiment tools have different scales as displayed in Table 4 and illustrated Figure 2.

Table 4 Tool And Its Scoring Domain

Sentiment tool	Scoring Domain
SentiStrength	-5 to +5
Repustate	-1 to 1
Text analysis online	-1 to 1
Text processing	-1 to 1
WordNet	-1 to 1
MPQA	-2 to 2
Opinionlexicon	2 to 2
IMDB	-1 to 1
SentiWordNet	-3 to 3
Vader	-1 to 1

² The code associated with these experiments can be found at the following URL: <https://github.com/UCREL/WSSExtract>

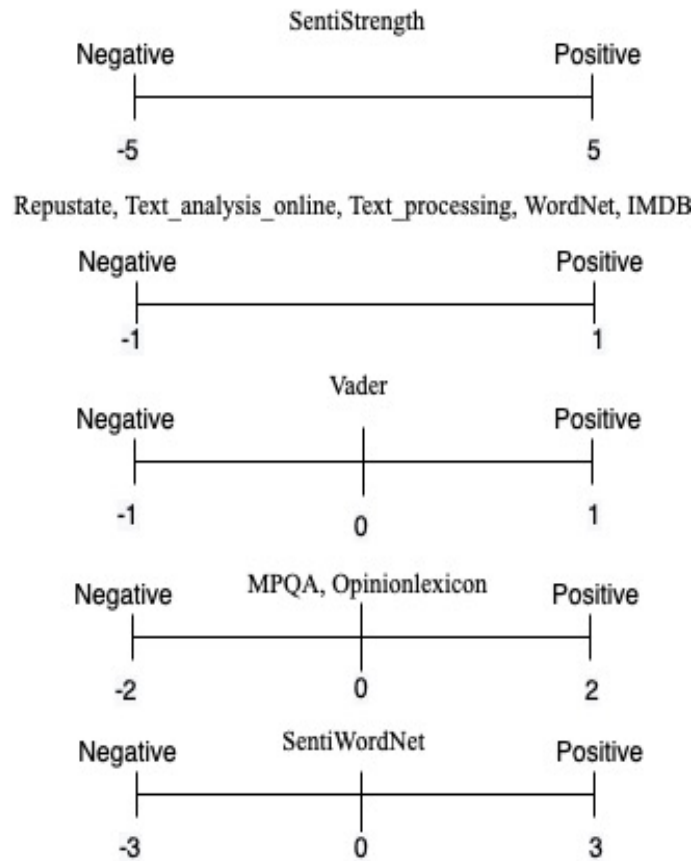


Figure 2 Scale of score for sentiment tools

The results of this study had to be scaled for consistent comparison and benchmarking purposes. For instance, the score for Opinionlexicon ranges from -2 to 2, which includes 0 in between, SentiStrength (Thelwall et al., 2010) ranges from -5 to +5 without having 0, Vader (Hutto and Gilbert, 2014) ranges from -1 to 1, including 0, and SentiWordNet (Esuli et al., 2006) ranges from -3 to 3 including 0. Without scaling, benchmarking cannot be done, and tools for different domains cannot be compared. Should there be a concern that normalisation could cause information loss, we argue that it has not affected the results of this study. The reason is that normalisation provides a standard scaling for comparison to be made and to demonstrate the ability of the tools in analysing the textual variation layer. The result of the comparison has no impact on our measurement. Table 5 summarizes the notations used in this model. Table 5 summarizes the notations used in this model.

Table 5 Notations used in the paper.

Symbol	Description
z	Normalized score in the range of -1 to 1
x	Sentiment score collected from sentiment tools
g	Mean absolute error (MAE)
y_i	Prediction
x_i	True value
n	Total number of data points
m	Root mean squared error (RMSE)
i	Variable
j	Number of non-missing data points
x_i	Actual observation time series
y_i	Is estimated time series
δ	Percent error
y_e	Expected value
u	Root relative squared error (RRSE)
P_{kj}	Value predicted by the the individual model
k	Record j out of n records
T_j	The target value for the record j

The normalisation scoring range was set from -1 to 1, with 0 indicating neutral. Normalisation was performed manually by applying the following general rescaling formula (Li and Liu, 2011):

$$z = ((x_i - \min(x)) / (\max(x) - \min(x))) \quad (1)$$

where $\min x = -1$ and $\max x = 1$, and $\min x$ and $\max x$ refer to an arbitrary range of sentiment scores.

To automatically test these ten tools, a prototype tool was built for accessing the sentiment tools. For each input text, the tool simultaneously produced ten sentiment analysis results using the ten tools. A total of fifty positive and fifty negative text samples were tested, with each sample simulated with the seven variations as in Figure 1. Result for each simulated sample was recorded for further analysis.

One hundred samples (real comments) from social media were selected randomly and tested on sentiment tools. Those samples are different from the previously mentioned 40

samples used for the survey. Each of the samples were simulated using the model shown in Figure 1. A total of eight variations of one sample was used, including unformatted text.

The sentiment score data generated with the sentiment tools were analysed using Weka, a data mining tool (Hall et al., 2009). One classifier was selected, in light of a previous study, to produce five statistical measures to compare the sentiment tools' performance (Pak and Teh, 2016). Each tool was selected as a class to produce its corresponding outcome.

The five statistical measures such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), and the difference between MAE and RMSE were used for analysis.

MAE measures the average magnitude of errors in a set of predictions without considering their direction. It is the average over the test samples of the absolute difference between prediction and actual observation where all individual differences have equal weight (Chai and Draxler, 2014). The formula for MAE is below:

$$g = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2)$$

RMSE is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of squared differences between prediction and actual observation (Chai and Draxler, 2014). The formula for RMSE is below:

$$m = \sqrt{\frac{\sum_{i=1}^j (p_i - y_i)^2}{j}} \quad (3)$$

RAE is very similar to the relative squared error in the sense that it is also relative to a simple predictor, which is the average of the actual values (Subramanian et al., 2016). Mathematical equation of RAE is below:

$$\delta = \left| \frac{x_i - y_e}{y_e} \right| \cdot 100\% \quad (4)$$

RRSE is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative

squared error takes the total squared error and normalises it by dividing the total squared error of the simple predictor (Subramanian et al., 2016). The mathematical equation of RRSE is below:

$$u = \frac{\sqrt{\sum_{j=1}^n (P_{kj} - T_j)^2}}{\sqrt{\sum_{j=1}^n (T_j - \bar{T})^2}} \quad (5)$$

The formula for \bar{T} is below:

$$\bar{T} = \frac{1}{n} \sum_{j=1}^n T_j \quad (6)$$

4 Results

4.1 Statistical Analysis Results from Phases 1 And 2

Before subjecting our survey data to hypothesis testing using one sample t-test, our data was examined for sample size adequacy, data distribution and the threat of common method bias. Two outliers were removed due to participants not having experience in online shopping. Results from the sample size assessment using G*Power post-hoc analysis showed the appropriateness of the sample size for this study (n=498), with power = 1.000 based on 0.15 effect size. Further, the normality of the data distribution was assessed based on (Kline, 2005) skewness (± 3) and kurtosis (± 10) thresholds. To avoid common method bias, (Podsakoff et al., 2003) procedural remedy of retaining the survey respondents' anonymity during data collection was applied. The reliability of our measures based on the seven types of variants used in online comments was assessed. Results from Cronbach's alpha analysis (Table 6) verified the reliability of measurements used in this study, as all Cronbach's alpha values are recorded at >0.7 (Nunnally and Bernstein, 1979). Table 6 also shows the mean of all rankings provided by the participants based on the variety of the different sentences using the same typographical elements.

Table 6 Mean, Standard Deviation and Cronbach's Alpha for each Evaluated Variation for Positive and Negative Text

	Construct	Mean	S.D.	Cronbach's Alpha
Positive Statements	Unformatted Text	4.667	0.77	0.873
	Capital Letters	5.492	0.926	0.908
	2 exclamation marks at the end	4.921	0.87	0.894
	4 exclamation marks at the end	5.533	1.127	0.943
	Certain Words in Capital	4.872	0.941	0.905
	Repeated Letters	5.149	1.36	0.951
	Positive Emoticons	5.205	1.184	0.932
Negative Statements	Negative Emoticons	3.461	0.924	0.901
	Unformatted Text	3.648	0.851	0.817
	Capital Letters	2.879	1.279	0.875
	2 exclamation marks at the end	3.266	1.092	0.878
	4 exclamation marks at the end	2.693	1.393	0.914
	Certain Words in Capital	2.999	1.074	0.866
	Repeated Letters	2.783	1.07	0.850
Positive Emoticons	3.514	0.881	0.816	
Negative Emoticons	2.587	0.989	0.850	

The paired sample t-test results in Table 7 show that the means for all seven variations of sentiment format are significantly different from the unformatted positive statement with unformatted text.

Table 7 Results Of The Hypothesis Tests ($p < 0.01$, ** $p < 0.05$, * $p < 0.1$)*

Hypothesis	Paired Comparison	Mean	S.D.	Std. error	Confidence Interval		t	Result
					Lower	Upper		
H1a	Positive_PlainText - Positive_AllCap	.825	.881	.039	.747	.902	20.88***	Supported
H1b	Positive_PlainText - Positive_Exc2	.253	.935	.042	.171	.336	6.041***	Supported
H1c	Positive_PlainText - Positive_Exc4	.865	1.160	.052	.763	.968	16.65***	Supported
H1d	Positive_PlainText - Positive_MCap	.204	1.129	.051	.105	.304	4.04***	Supported
H1e	Positive_PlainText - Positive_RepLetter	.482	1.462	.065	.353	.610	7.355***	Supported
H1f	Positive_PlainText - Positive_EmoPos	.538	1.303	.058	.423	.653	9.216***	Supported
H1g	Positive_PlainText - Positive_EmoNeg	-1.207	1.149	.051	-1.308	-1.106	-23.445***	Supported
H2a	Negative_PlainText - Negative_AllCap	-.769	.930	.042	-.851	-.687	-18.451***	Supported
H2b	Negative_PlainText - Negative_Exc2	-.382	.871	.039	-.459	-.305	-9.79***	Supported
H2c	Negative_PlainText - Negative_Exc4	-.955	1.152	.052	-1.056	-.854	-18.507***	Supported
H2d	Negative_PlainText - Negative_MCap	-.650	1.063	.048	-.743	-.556	-13.645***	Supported
H2e	Negative_PlainText - Negative_RepLetter	-.866	1.193	.053	-.971	-.761	-16.19***	Supported
H2f	Negative_PlainText - Negative_EmoPos	-.135	1.367	.061	-.255	-.014	2.201**	Supported
H2g	Negative_PlainText - Negative_EmoNeg	-1.061	1.026	.046	-1.151	-.971	23.086***	Supported

All hypotheses are supported with mean differences varying from 0.204 to -1.207 (t-values ranging from 4.04 to 23.455, p-value <0.01***, 05**) and 95% confidence intervals that do not straddle the value of zero. Similarly, all seven variations of negative statements correlate with negative statements with normal text.

Texts using a mix of capital letters and two exclamation marks at the end makes the least mean difference when compared to standard positive statements (0.204 and 0.253, respectively). When positive comments are given in all capital letters and four exclamation marks at the end, the mean differences are 0.825 and 0.865, close to a one-unit increase in sentiment values. Repeated letters and positive emoticons used in positive statements also increase the sentiment value of positive text to 0.5 unit. The role of negative emoticons in changing the sentiment values of positive statements is the highest (more than one unit) compared to the other variations of text (mean difference = -1.207).

The use of positive emoticons in negative statements does not clearly affect the value of negative statements mean = -0.135). When negative statements are expressed with two exclamation marks at the end (mean difference = - 0.382), the sentiment values change the

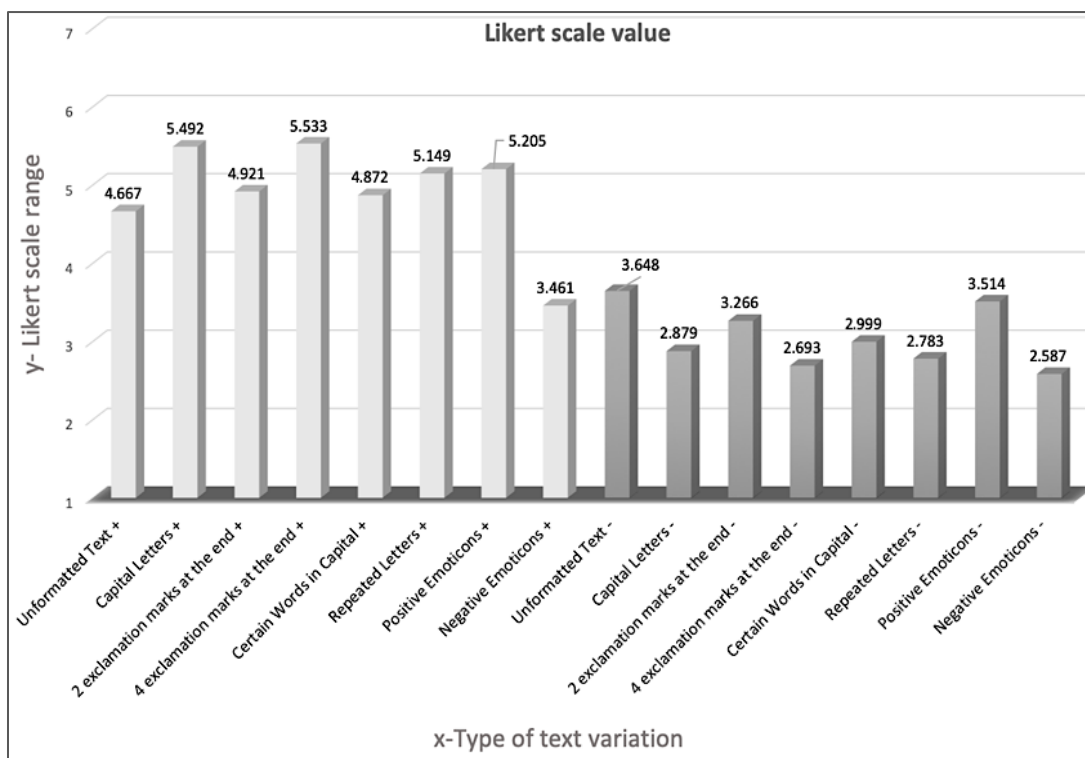


Figure 3 Average scores from collected survey for positive and negative statements.

least. It is worth noticing that for all other forms of variations in negative statements (all capital letters, four exclamation marks at the end, a mix of capital letters, and repeated letters), the magnitude of the mean difference exceeds more than 0.5 unit. As for positive statements, the role of negative emoticons in changing the sentiment values of negative statements is the highest (more than one unit) compared to the other variations of text (mean difference = -1.061).

Error! Reference source not found.³ shows the bar graph comparing positive and negative statements with their respective sentiment variations. When comparing the use of variations for both statements, the magnitude of change in negative statements is larger compared to positive statements. The use of exclamation marks reinforces the sentiment value of both positive and negative statements. The magnitude of change is significant when negative emoticons are used in both positive and negative statements. Negative emoticons reduce positive value for positive text. Whereas, the use of negative emoticons in negative text increase their negative value.

Deleted:

4.2 Sentiment Tools' Results from Phase 3

Findings show that different textual variations affect the human judgement of sentiment values. In this section, further analysis of the impacts of the textual variations is compared with existing/available automatic sentiment analysis tools. Specifically, automated sentiment analysis and detection were investigated by testing the text samples with variations in ten existing tools listed in Table 8 and Table 9. The differences in tool performances were examined with and without the inclusion of variation.

Table 8 Statistical Results From Sentiment Tools For Positive Text

Sentiment tool	Mean absolute error (MAE)	Root mean squared error (RMSE)	Difference between MAE and RMSE	Relative absolute error (RAE) (%)	Root relative squared error (RRSE) (%)
SentiStrength	0.0480	0.2191	0.1711	23.9005	48.8881
Repustate	0.1791	0.2378	0.0587	57.8568	67.8421
Text_analysis_online	0.1736	0.247	0.0734	60.0137	69.6591

Text_processing	0.2104	0.3978	0.1874	32.8216	53.6056
WordNet	0.0502	0.0858	0.0356	33.6884	48.942
MPQA	0.0280	0.0415	0.0135	33.3069	37.0264
Opinionlexicon	0.0122	0.0210	0.0088	28.7018	39.3466
IMDB	0.2021	0.5663	0.3642	60.5541	81.5634
SentiWordNet	0.0223	0.0431	0.0208	54.3788	71.0491
Vader	0.2015	0.2757	0.0742	86.7501	90.9473

Table 9 Statistical Results From Sentiment Tools For Negative Text

Sentiment tool	Mean absolute error (MAE)	Root mean squared error (RMSE)	Difference between MAE and RMSE	Relative absolute error (RAE) (%)	Root relative squared error (RRSE) (%)
SentiStrength	0.2514	0.5014	0.2500	27.0341	51.9932
Repustate	0.2046	0.2766	0.0720	55.4582	68.9762
Text_analysis_online	0.1618	0.2372	0.0754	47.8368	56.1917
Text_processing	0.2327	0.3980	0.1653	53.5585	72.7421
WordNet	0.0417	0.0930	0.0513	23.7107	35.8019
MPQA	0.0153	0.0323	0.0170	23.3125	39.6119
Opinionlexicon	0.0118	0.0201	0.0083	29.6166	38.6066
IMDB	0.1149	0.1641	0.0492	49.1558	56.7412
SentiWordNet	0.0144	0.0203	0.0059	46.2401	50.8546
Vader	0.2950	0.3696	0.7460	82.4103	87.8858

Table 8 presents the statistical results for positive text, and compares the sentiment tools' performances. MAE and RMSE can be used together to diagnose the error variation in the study. The range for MAE and RMSE can be from 0 to 1. They are negatively oriented scores wherein lower values are better. Based on the MAE results in Table 8, the lowest value recorded by Opinionlexicon indicates that it has the lowest variance in individual errors. Moreover, the difference between MAE and RMSE for that tool is also the lowest followed by MPQA, SentiWordNet and SentiStrength, while IMDB shows the highest percentages for MAE, RMSE and the difference between the MAE and RMSE. It follows that the sentiment analysis based on IMDB failed the test for positive text. The sentiment algorithm based on WordNet in a study by Medagoda et al. (2016) reported an accuracy of 60%. Although in our study, RAE for WordNet is 33.69%, the remaining 66.31% of data were accurately classified, the accuracy close to that of (Medagoda et al., 2016).

RRSE aggregates the magnitude of errors in multiple predictions into a single measure of predictive power. While RRSE is a good measure of accuracy, it is scale-dependent (Rong

et al., 2015), and therefore it only compares forecasting errors of different models for a particular variable—not between variables. As such, smaller values are better and values greater than 100% indicate that the model is doing worse than just predicting the mean. RAE is computed similarly.

According to the RAE results, the sentiment tools SentiStrength and Opinionlexicon recorded the lowest values at 23.90% and 28.70% respectively. These show that the two tools outperformed the rest of the sentiment tools based on RAE results. Overall, the performances of Opinionlexicon and SentiStrength are considered to be the best according to the RAE and indicated RRSE values.

Table 9 presents the results for negative text. Opinionlexicon is shown to record the lowest value for MAE (0.0118). However, Opinionlexicon has the second-lowest difference between MAE and RMSE (0.0083). Similarly, SentiWordNet has low values of MAE (0.0144) and the lowest difference between MAE and RMS (0.0059). SentiStrength, on the other hand, has the highest values of MAE and the difference between MAE and RMSE is 0.2514 and 0.5014, respectively. Therefore, based on MAE and RMSE, it can be concluded that Opinionlexicon and SentiWordNet perform the best for negative text testing, while SentiStrength is the poorest.

This study leads to the conclusion that human expressions in the form of textual variations are equally important for analysis. This work can be a starting point for other studies to propose a method or framework to calculate sentiment by taking those textual variations into account. In addition, a previous study of Chaudhary et al. (2018) has also listed some possible weaknesses of current sentiment approaches which should be addressed. For instance, lexical affinity does not work well with negated sentences and is biased towards the text of a particular category. Overall, it can be seen that, even though the normalised set of data was used for

comparison, there is a difference in sentiment scoring between the available sentiment tools and human ratings.

5 Conclusion and Future Work

This paper investigated the issue of how textual variations affect human judgment, and the extent to which sentiment analysis tools fail to accurately reflect human judgement. Our main contribution in this research was to demonstrate that such textual variations have a significant impact on the accuracy of measuring sentiment values. More specifically, a novel large scale human rating experiment was carried out from which manual scores have been collected³. The findings of our study further suggest that existing automatic sentiment analysis tools should take into account textual variations in text in comparison with human judgement.

Our results show that sentiment analysis algorithms/tools should not remove or filter textual variations from the analysis as it conveys vital sentiment information. These variations can be used to judge the sentiment values expressed in social media and online forums besides positive and negative ratings. Ignorance of textual variations is not “bliss” in this case, as previous research has also revealed their important implications, especially in linguistic studies. And our study uses technology and mathematics to prove the need to hold the value. This value is helpful in opinion mining for business, commerce, relevant text and data science. The implication for both industry practitioners and social media analysts is that the sentiment of comments in social media should not be judged at face value, as the “emotion” embedded within the textual variations hold the important sentiment. However, there are still more to be discovered, such as if emoticons can reverse the polarity of positive or negative comments (Teh et al., 2016), and how would it differ when it comes to textual variations? Future research should take that into account.

³ The manual scores can be found at the following URL: at <https://github.com/UCREL/HumanJudgementsOfSentimentValues>

- Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 7, 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chaudhary, M., Kumar, H., Kaushal, S., Sangaiah, A.K., 2018. The case analysis on sentiment based ranking of nodes in social media space. *Multimed. Tools Appl.* 77, 4217–4236. <https://doi.org/10.1007/s11042-017-4700-3>
- Cobb, N.K., Mays, D., Graham, A.L., 2013. Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. *J. Natl. Cancer Inst. Monogr.* 47, 224–230.
- Cohen, S.A., Chui, K.K.H., Naumova, E.N., 2011. Measuring disease burden in the older population using the slope-intercept method for population log-linear estimation (SIMPLE). *Stat. Med.* 30, 480–488.
- Derks, D., Bos, A.E.R., Grumbkow, J. Von, 2007. Emoticons and social interaction on the Internet: the importance of social context 23, 842–849. <https://doi.org/10.1016/j.chb.2004.11.013>
- DeWald, C.G., Geyer, L.H., 1975. An operations research approach to the modeling and analysis of different feature sets proposed for human perception of capital letters. *Comput. Oper. Res.* 2, 61–70. [https://doi.org/10.1016/0305-0548\(75\)90009-X](https://doi.org/10.1016/0305-0548(75)90009-X)
- Esuli, A., Sebastiani, F., Moruzzi, V.G., 2006. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining, in: 5th Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), Genoa, Italy, pp. 417–422. <https://doi.org/10.1.1.61.7217>
- Garrison, A., Remley, D., Thomas, P., Wierszewski, E., 2011. Conventional Faces: Emoticons in Instant Messaging Discourse. *Comput. Compos.* 28, 112–125. <https://doi.org/10.1016/j.compcom.2011.04.001>
- Godin, F., De Neve, W., Van de Walle, R., 2013. Towards fusion of collective knowledge and audio-visual content features for annotating broadcast video. *Proc. 3rd ACM Conf. Int. Conf. Multimed. Retr. - ICMR '13* 329. <https://doi.org/10.1145/2461466.2461530>
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software. *ACM SIGKDD Explor.* 11, 10–18. <https://doi.org/10.1145/1656274.1656278>
- He, G., Deng, Y., Xiaochun Wu, 2012. Analysis of web-surfing behavior of enterprise network users, in: 2nd International Conference on Computer Science and Network Technology. IEEE, Changchun, China, pp. 916–920.
- Heise, D.R., 2014. Cultural variations in sentiments. *Springerplus* 3, 1–11. <https://doi.org/10.1186/2193-1801-3-170>
- Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., Kaymak, U., 2013. Exploiting emoticons in sentiment analysis, in: SAC'13, the 28th Annual ACM Symposium on Applied Computing. ACM New York, NY, USA ©2013, Coimbra, Portugal., pp. 703–710. <https://doi.org/10.1145/2480362.2480498>
- Huang, A.H., Yen, D.C., Zhang, X., 2008. Exploring the potential effects of emoticons. *Inf. Manag.* 45, 466–473. <https://doi.org/10.1016/j.im.2008.07.001>
- Hutto, C.J., Gilbert, E., 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, in: Eighth International Conference on Weblogs and Social Media. Ann Arbor, MI, pp. 1–10.
- Jianqiang, Z., Xiaolin, G., Xuejun, Z., 2018. Deep Convolution Neural Networks for Twitter Sentiment Analysis. *IEEE Access* 6, 23253–23260. <https://doi.org/10.1109/ACCESS.2017.2776930>
- Kalman, Y.M., Gergle, D., 2014. Letter repetitions in computer-mediated communication: A unique link between spoken and online language. *Comput. Human Behav.* 34, 187–193.

- Kalman, Y.M., Gergle, D., 2009. Letter and Punctuation Mark Repeats as Cues in Computer-Mediated Communication, in: 95th Annual Meeting of the National Communication Association in Chicago.
- Kim, D., Frank, M.G., Kim, S.T., 2014. Emotional display behavior in different forms of Computer Mediated Communication. *Comput. Human Behav.* 30, 222–229. <https://doi.org/10.1016/j.chb.2013.09.001>
- Kiritchenko, S., Zhu, X., Mohammad, S.M., 2014. Sentiment Analysis of Short Informal Texts. *J. Artif. Intell. Res.* 50, 723–762. <https://doi.org/https://doi.org/10.1613/jair.4272>
- Kline, T.J.B., 2005. *Psychological testing: A practical approach to design and evaluation*, 1st ed. SAGE Publications, Inc, Thousand Oaks, CA, US.
- Kotelnikov, E. V, Pletneva, M. V, 2016. Text Sentiment Classification Based on a Genetic Algorithm and Word and Document Co clustering 1. *J. Comput. Syst. Sci. Int.* 55, 115–123. <https://doi.org/10.1134/S1064230715060106>
- Ledbetter, A., Larson, K., 2008. Nonverbal cues in e-mail supportive communication: Associations with sender sex, recipient sex, and support satisfaction. *Information, Commun. Soc.* 11, 1089–1110.
- Li, W., Liu, Z., 2011. A method of SVM with Normalization in Intrusion Detection. *Procedia Environ. Sci.* 11, 256–262. <https://doi.org/https://doi.org/10.1016/j.proenv.2011.12.040>
- Liang, P.-W., Dai, B.-R., 2013. Opinion Mining on Social Media Data, in: 2013 IEEE 14th International Conference on Mobile Data Management. IEEE, Milan, Italy, pp. 91–96.
- Lin, S., Shen, Y., 2011. Taobao Data on Women’s Apparel Online Sales of Consumer Behavior Analysis, in: 2011 International Conference on Internet Technology and Applications. IEEE, Wuhan, China, pp. 1–4.
- Lin Su, 2018. “Say it! No ideas but in things—”: Punctuation Marks and American Locality in William Carlos Williams’s *Paterson*. *William Carlos Williams Rev.* 35, 128. <https://doi.org/10.5325/willcarlwillrevi.35.2.0128>
- Lochter, J. V, Zanetti, R.F., Reller, D., Almeida, T.A., 2016. Short text opinion detection using ensemble of classifiers and semantic indexing. *Expert Syst. with Appl. An Int. J.* 62, 243–249. <https://doi.org/10.1016/j.eswa.2016.06.025>
- Medagoda, N., Shanmuganathan, S., Whalley, J., 2016. Sentiment lexicon construction using SentiWordNet 3.0, in: *Proceedings - International Conference on Natural Computation*. IEEE, Zhangjiajie, China, pp. 802–807. <https://doi.org/10.1109/ICNC.2015.7378094>
- Mohammad, S.M., Kiritchenko, S., Zhu, X., 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, in: *The 7th International Workshop on Semantic Evaluation Exercises*. Association for Computational Linguistics, Atlanta, USA, pp. 321–327.
- Mohey, D., Hussein, E.M., 2016. A survey on sentiment analysis challenges. *J. King Saud Univ. - Eng. Sci.* 30, 330–338. <https://doi.org/10.1016/j.jksues.2016.04.002>
- Mori, H., Nagaoka, A., Arimoto, Y., 2016. Accuracy of Automatic Cross-Corpus Emotion Labeling for Conversational Speech Corpus Commonization, in: *10th International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), Portorož, Slovenia, pp. 4019–4023.
- Naradhupa, A.R., Purwarianti, A., 2011. Sentiment Classification for Indonesian Message in Social Media, in: *Cloud Computing and Social Networking (ICCCSN), 2011 International Conference on Electrical Engineering and Informatics*. IEEE, Bandung, Indonesia, pp. 1–5.
- Neviarouskaya, A., Prendinger, H., Ishizuka, M., 2011. Affect Analysis Model: novel rule-based approach to affect sensing from text. *Nat. Lang. Eng.* 17, 95–135. <https://doi.org/10.1017/S1351324910000239>
- Novak, P.K., Smailović, J., Sluban, B., Mozetič, I., 2015. Sentiment of emojis. *PLoS One* 10,

- 1–22. <https://doi.org/10.1371/journal.pone.0144296>
- Nunnally, J.C., Bernstein, I.H., 1979. This week's citation classic, Psychometric theory. McGraw-Hill.
- Pak, I., Teh, P.L., 2018. Value of expressions behind the letter capitalization in product reviews, in: ACM International Conference Proceeding Series. <https://doi.org/10.1145/3185089.3185150>
- Pak, I., Teh, P.L., 2016. Machine Learning Classifiers : Evaluation of the Performance in Online Reviews. Indian J. Sci. Technol. 9, 1–9. <https://doi.org/10.17485/ijst/2016/v9i45/100703>
- Pak, I., Teh, P.L.P.L., Cheah, Y.-N.Y.-N.Y.-N., 2018. Hidden Sentiment Behind Letter Repetition in Online Reviews. J. Telecommun. Electron. Comput. Eng. 10, In Press.
- Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. Proc. Conf. Empir. Methods Nat. Lang. Process. July 6-7, 2002, Philadelphia, Pennsylvania, USA 79–86.
- Pfeifer, V.A., Armstrong, E.L., Lai, V.T., 2022. Do all facial emojis communicate emotion? The impact of facial emojis on perceived sender emotion and text processing. Comput. Human Behav. 126, 107016. <https://doi.org/https://doi.org/10.1016/j.chb.2021.107016>
- Podsakoff, P.M., MacKenzie, S.B., Podsakoff, N.P., Lee, J.-Y., 2003. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. J. Appl. Psychol. 88, 879–903.
- Prabowo, R., Thelwall, M., 2009. Sentiment analysis: A combined approach. J. Informetr. 3, 143–157. <https://doi.org/https://doi.org/10.1016/j.joi.2009.01.003>
- Rayson, P., 2008. From key words to key semantic domains. Int. J. Corpus Linguist. 13, 519–549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Riordan, M. a., Kreuz, R.J., 2010. Emotion encoding and interpretation in computer-mediated communication: Reasons for use. Comput. Human Behav. 26, 1667–1673. <https://doi.org/10.1016/j.chb.2010.06.015>
- Riordan, M.A., Kreuz, R.J., 2010. Cues in computer-mediated communication: A corpus analysis. Comput. Human Behav. 26, 1806–1817.
- Rong, Y., Zhang, X., Feng, X., Ho, T., Wei, W., Xu, D., 2015. Comparative analysis for traffic flow forecasting models with real-life data in Beijing. Adv. Mech. Eng. 7, 1–9. <https://doi.org/10.1177/1687814015620324>
- Schomaker, L., Bulacu, M., 2004. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. IEEE Trans. Pattern Anal. Mach. Intell. 26, 787–798. <https://doi.org/10.1109/TPAMI.2004.18>
- Sesa-Nogueras, E., Faundez-Zanuy, M., 2012. Biometric recognition using online uppercase handwritten text. Pattern Recognit. 45, 128–144. <https://doi.org/10.1016/j.patcog.2011.06.002>
- Sidi, Y., Glikson, E., Cheshin, A., 2021. Do You Get What I Mean?!? The Undesirable Outcomes of (Ab)Using Paralinguistic Cues in Computer-Mediated Communication. Front. Psychol. 12, 1337. <https://doi.org/10.3389/fpsyg.2021.658844>
- Smetanin, S., 2020. The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives. IEEE Access 8, 110693–110719. <https://doi.org/10.1109/ACCESS.2020.3002215>
- Soderlund, M., Oikarinen, E.-L., Tan, T.M., 2021. The happy virtual agent and its impact on the human customer in the service encounter. J. Retail. Consum. Serv. 59, 102401. <https://doi.org/https://doi.org/10.1016/j.jretconser.2020.102401>
- Subramanian, S.M., Vijayalakshmi, S., Venkataraman, B., 2016. CCCa Framework - Classification System in Big Data Environment with Clustering and Cache Concepts, in: Eighth International Conference on Soft Computing and Pattern Recognition (SoCPaR

- 2016). pp. 45–53. <https://doi.org/10.1007/978-3-319-60618-7>
- Tang, T., Tang, X., Yuan, T., 2020. Fine-Tuning BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text. *IEEE Access* 8, 193248–193256. <https://doi.org/10.1109/ACCESS.2020.3030468>
- Ted, T., Wu, L., Lu, H., Tao, Y., 2010. Computers in Human Behavior The effect of emoticons in simplex and complex task-oriented communication : An empirical study of instant messaging. *Comput. Human Behav.* 26, 889–895. <https://doi.org/10.1016/j.chb.2010.02.003>
- Teh, P.L., Pak, I., Rayson, P., Piao, S., 2015a. Exploring fine-grained sentiment values in online product reviews, in: 2015 IEEE Confernece on Open Systems (ICOS). IEEE, Melaka, pp. 114–118.
- Teh, P.L., Rayson, P., Pak, I., Piao, S., 2015b. Sentiment Analysis Tools Should Take Account of the Number of Exclamation Marks!!!, in: *IiWAS '15 Proceedings of the 17th International Conference on Information Integration and Web-Based Applications & Services*. ACM New York, NY, USA ©2015, p. Article No. 35. <https://doi.org/10.1145/2837185.2837216>
- Teh, P.L., Rayson, P., Pak, I., Piao, S., Yeng, S.M., 2016. Reversing Polarity With Emoticons, in: *Natural Language Processing and Information Systems*. Springer, Salford, pp. 453–458.
- Thelwall, M., Buckley, K., Paltoglou, G., 2012. Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.* 63, 163–173.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., 2010. Sentiment Strength Detection in Short Informal Text. *Am. Soc. Informational Sci. Technol.* 61, 2544–2558. <https://doi.org/10.1002/asi>
- Thelwall, M., Wilkinson, D., 2010. Public dialogs in social network sites: What is their purpose? *Am. Soc. Informational Sci. Technolo-gy* 61, 392–404.
- Tossell, C.C., Kortum, P., Shepard, C., Barg-Walkow, L.H., Rahmati, A., Zhong, L., 2012. A longitudinal study of emoticon use in text messaging from smartphones. *Comput. Human Behav.* 28, 659–663. <https://doi.org/10.1016/j.chb.2011.11.012>
- Urabe, Y., Rzepka, R., Araki, K., 2013. Emoticon recommendation system for effective communication, in: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). IEEE, Niagara Falls, ON, Canada, pp. 1460–1461.
- Vandergriff, I., 2013. Emotive communication online: A contextual analysis of computer-mediated communication (CMC) cues. *J. Pragmat.* 51, 1–12. <https://doi.org/10.1016/j.pragma.2013.02.008>
- Wu, Y., Zhang, Y., Luo, S.M., Wang, X.J., 2007. Comprehensive information based semantic orientation identification, in: *IEEE NLP-KE 2007 - Proceedings of International Conference on Natural Language Processing and Knowledge Engineering*. IEEE, Beijing, pp. 274–279.
- Xue, B., Fu, C., Shaobin, Z., 2014. A Study on Sentiment Computing and Classification of Sina Weibo with Word2vec, in: 2014 IEEE International Congress on Big Data. IEEE, Anchorage, AK, USA, pp. 358–363.
- Zavattaro, S.M., French, P.E., Mohanty, S.D., 2015. A sentiment analysis of U.S. local government tweets: The connection between tone and citizen involvement. *Gov. Inf. Q.* 32, 333–341. <https://doi.org/10.1016/j.giq.2015.03.003>
- Zin, T.T., Tin, P., Toriu, T., Hama, H., 2012. Visual behavior analysis tool for consumer video surveillance, in: *The 1st IEEE Global Conference on Consumer Electronics 2012*. IEEE, Tokyo, Japan, pp. 718–719.

