A Paralinguistic Analysis of Emojis and Emoticons

Siddhesh Suresh Bangar
bangars@tcd.ie

Shaunak Pedgaonkar
pedgaoks@tcd.ie

Eamon Phelan
phelanec@tcd.ie

Parthiban Thandapani
thandapp@tcd.ie

Jiaxuan Xie
xiej1@tcd.ie

Pragati Aboti
abotip@tcd.ie

Abstract

In contemporary communication, the conveyance of affect extends beyond mere linguistic content, encompassing paralinguistic cues embedded within written discourse. This research investigates the multifaceted role of emoticons and emojis as vehicles for affective expression, examining their syntactic and semantic dimensions in diverse communicative contexts. Building upon prior studies that underscore the fluidity of emoticon and emoji usage, this paper addresses pertinent questions regarding their placement within textual structures and the semantic associations they evoke. By employing computational text analysis techniques, we probe the interplay between affective signals and textual similarity, discerning patterns that elucidate the differential impact of positive and negative emoticons and emojis on semantic coherence. Furthermore, we explore the phenomenon of 'confusions' in affect categories, elucidating instances where independent judges perceive irony or sarcasm in written expressions. Through this interdisciplinary inquiry, our study contributes nuanced insights into the evolving dynamics of affect signaling in written communication, shedding light on the intricate interplay between linguistic and paralinguistic modalities in shaping meaning and interpretation.

1 Introduction

Affect, or the feeling an individual has towards a topic, is a crucial aspect of communication. In spoken language, intonation and bodily movements convey affective information in addition to explicit content. Similarly, written language is also accompanied by paralinguistic content that contains affect signals, such as emoticons and emojis. These forms of expression are relatively recent and their use remains fluid, which raises questions about their syntax and semantics.

Emotive elements like emoticons and emojis have gained prominence due to their ability to enhance the emotional depth of digital communication, particularly in text-based exchanges. Although previously studied (Steinert and Dennis, 2022; Quintes and Ullrich, 2019), the consistency of their effects in new datasets and their syntactic and semantic properties remain under-explored. This research aims to address these gaps by examining the constraints governing their usage patterns, such as their placement within sentences and phrases, and exploring the extent to which their inclusion influences the interpretation of associated texts.

Previous research by Li and Yang (2018); Bai et al. (2019) has highlighted the transformative power of emoticons and emojis in shaping digital communication, revealing their capacity to convey complex emotional states and facilitate intimate connection among users. Studies (Cherbonnier and Michinov, 2022; Neel et al., 2023) have also noted variations in the effectiveness of certain emoticons and emojis depending on their design characteristics, such as facial expressions, color schemes, and cultural associations. Building upon this foundation, the proposed research endeavors to expand our understanding of these vital components of contemporary discourse and their implications for text analysis and interpretation.

Therefore, this research aims to investigate the syntax and semantics of emoticons and emojis in written language and their impact on text-based similarity measures. Specifically, we will examine the constraints that determine their placement and the extent to which their presence affects the interpretation of associated texts. By addressing these questions, we hope to contribute to a better understanding of the role of emoticons and emojis in written communication.

2 Related Work

2.1 Emojis

With their widespread use across social media platforms, emojis have evolved into indispensable tools for digital communication, especially when it comes to conveying emotions. A comprehensive analysis of emoji research by Bai et al. (2019) highlights their creation, use, purpose, and range of uses. It draws attention to how important emojis are in making up for the absence of nonverbal clues in computer-mediated communication (CMC), which improves emotional connection, the caliber of relationships, and the flow of one-on-one contact. The resulting body of work also highlights the variety of applications of emojis in network communication, emphasizing its emotive and semantic qualities as well as their applicability to industries other than social interaction, such as advertising and healthcare.

Further research into their effects on emotional communication, social attributions, and information processing in online social media (OSM) by Boutet et al. (2021) underlines the importance of emojis. Emojis enhance the tone of digital messages by acting as tools for emotional expression and interpersonal engagement, leading to a more positive emotional tone, enhancing user trust, and facilitating deeper digital connections. In the context of romantic and sexual interactions initiated through CMC, emojis serve as vital tools for fostering relationships, as discussed by Gesselman et al. (2019). They imbue digital communication with necessary socio-emotional components, enabling affective self-disclosure and deeper levels of intimacy.

2.2 Text Similarity Analysis

In natural language processing applications, text similarity measures play a key role. A framework that combines semantic and syntactic information was introduced by Yang et al. (2021). It is evaluated on 24 datasets and the results show that it outperforms baseline methods in semantic text similarity tasks.

A text similarity measurement method for sentiment analysis was proposed by Tohti et al. (2021). It utilizes word vectors, a Chinese thesaurus for semantic similarity, and a sentiment dictionary for sentence and paragraph level sentiment analysis. Compared with traditional methods, this method shows higher precision, recall and time efficiency. Natural language processing (NLP) tools, Jaccard and cosine similarity measures were used by Qurashi et al. (2020). The study concluded that cosine similarity (which measures the angle between vectorized sentences) provides more accurate results compared to Jaccard similarity. A method for calculating text similarity was introduced by Mhatre et al. (2023). The approach blends statistical techniques and semantics, employing algorithms, libraries and semantic models. The method is suitable for use in different fields, emphasizing its flexibility and effectiveness for various text analysis applications.

These studies demonstrate the diverse approaches in the field of text similarity analysis. Each method has its unique application scenarios and advantages, demonstrating that combining multiple technologies and perspectives can improve the accuracy and efficiency of analysis in different text analysis tasks.

2.3 Parts of Speech Analysis

The use of incredibly sophisticated machine learning and deep learning algorithms has significantly changed sentiment analysis and Emoji recognition in textual data. Text sentiment identifications have shown a great deal of promise because to methods like CNN (Convolutional Neural Networks) by Liao et al. (2017), LSTM (Long Short Term Memory Networks) by Wang et al. (2016), and Bi-LSTM (Bidirectional LSTM) by Minaee et al. (2019), which can incorporate these algorithms and comprehend the context. By integrating these algorithms with word embeddings by Deho et al. (2018) and POS-tagging by Srividya and Sowjanya (2019), we have been able to evaluate emotions on a range of datasets, demonstrating their adaptability in managing the complexities of natural language, including the complex, free-form writing that appears on social media sites. We can also see by using transfer learning which is one of the key and important strategies for expanding the capabilities and efficacy of sentiment analysis models by Liu et al. (2019).

Here, it has been shown that hybrid models which mix machine learning and deep learning techniques have the ability to overcome the drawbacks of standalone approaches. Because these hybrid models combine techniques such Support Vector Machines (SVM) by Zainuddin and Selamat (2014) and random forests, they have achieved significant accuracy in applications like Amazon product review research. Furthermore, emoji sequences can also be examined more closely for structural regularities by treating them as beat gestures instead of speech segments. Emojis sequences can be highly repetitive, similar to beat gestures in speech that are repetitive in a rhythmic manner. Emoji are more like motions that go along with printed text than words because they don't have hierarchical grammar structures. According to the study by McCulloch and Gawne (2018), a emoji can emphasize their function as digital gestures rather than as formal language, emphasizing their similarity to beat gestures in spoken conversation.

3 Research Questions

3.1 Constraints on Emoji Placement

A gap in the literature exists for a large scale algorithmic analysis of the constraints on emoji placement within text that we will attempt to address in this paper. Previous studies were hand annotated with smaller sample sizes, limiting their conclusions. We aim to provide an up to date result detailing what emojis are most likely to occur and where e.g. within a phrase, at the end of a sentence, or as an entire message.

4 Research Methods

4.1 Datasets

Emojis and emoticons have become an increasingly popular way to express emotions and convey meaning in digital communication. As a result, several datasets have been developed to analyze the use of emojis and emoticons in various contexts, including social media Miller et al. (2016), marketing (Walther and D'addario, 2001) , and psychology (Kralj Novak et al., 2015; Mollahosseini et al., 2017).

In our research endeavor, we direct our attention towards datasets enriched with emojis and/or emoticons that closely align with the focal points of our inquiry. Notably, Guibon et al. (2018) have curated a dataset comprising messages or posts interspersed with emojis, reflective of varied emotions, reactions, or intentions. Each datum within this corpus encompasses textual content alongside the accompanying emojis, thus facilitating nuanced investigations into the paralinguistic dimensions of emojis and emoticons.

Moreover, the comprehensive corpus fashioned by Shardlow et al. (2022) warrants attention, as it encompasses a vast array of messages sourced from three prominent social media platforms: Twitter, Reddit, and TikTok. This expansive compilation forms the bedrock for scrutinizing patterns of emoji utilization and their consequential ramifications for emotional expression within the realm of human-computer interaction. Leveraging this extensive repository, researchers are empowered to delve into the frequencies, diversities, and contextual intricacies surrounding emoji usage across divergent demographics and online platforms, thereby furnishing invaluable insights into the multifaceted roles assumed by emojis in contemporary online discourse.

4.2 Text Similarity Approach

Text similarity techniques are essential for improving numerous components of sentiment classification and emoji analysis jobs. This section of the study examines how these techniques have been used in recent research studies to enhance sentiment analysis accuracy and expand our knowledge of emoji semantics. Unicode characters are frequently used to represent emoticons and emojis. Emojis and emoticons can be used to compare Unicode character sequences in texts by first converting them into their Unicode equivalents. For instance, based on previous research by Wijeratne et al. (2017)"A Semantics-Based Measure of Emoji Similarity," text similarity methods are used to determine the semantic similarity between emojis. Generally, it is attempted to extract descriptions and use word embedding models trained on Twitter and Google News to transform meanings into vector representations by using machine-readable emoji meanings from EmojiNet. With this knowledge, we are able to understand that this method helps with sentiment analysis tasks and expands on our grasp of emoji semantics. Emojis also change how people perceive emotion in affectively neutral text messages. They use text similarity in an indirect way as part of the sentiment analysis process. In order to examine how emojis affect emotional tones, researchers have probably used machine learning algorithms. They might also use text similarity metrics to evaluate how similar new texts are to training texts. These algorithms are instrumental in classifying texts, with text similarity potentially aiding in assessing new texts' similarity to training data.Neel et al. (2023)

Emoji prediction mostly relies on text similarity. To improve the accuracy of emoji predictions, word2vec and emoji2vec models are used in addition to similarity modelling techniques like Jaccard similarity and skip-gramme neural networks. As mentioned by Raj and Balachandran (2020) Text similarity techniques have a substantial overall impact on improving sentiment analysis and emoji analysis, underscoring their significance in textual analysis.

4.3 Parts of Speech Approach

This portion of the paper aims to address the lack of large corpus research in the area of emoji and emoticon placement within text. Previous research has shown that rather than replacing words emojis tend to act as multi-modal indicators of affect or stance (Na'aman et al., 2017) and are overwhelming at the end of sentences (Cramer et al., 2016). This existing research however has focused on hand annotated datasets of smaller size.

In this paper we use automated text analysis to develop a richer understanding of both where in text emojis occur and what types of emojis appear at particular locations. As a source of textual emoticons we reuse a dataset created as part of previous work (Vogel and Janssen, 2009) while the full set of Unicode emojis are sourced from an open source repository of up to date Unicode characters and classified according to the Unicode consortium's groupings.

The locations of emoticons and emojis are first identified within the text using regular expressions. The text is then parsed using the spaCy natural language processing library (Montani et al., 2023) to produce sentence, chunk, and parts of speech data. The locations of the emojis are then compared to the spaCy object boundaries to determine where within sentences, phrases, and words emoji occur, and which emoji occur where. As an example the thumbs up emoji may occur far more often on its own as a full sentence while the Birthday Cake emoji may only occur within related sentences.

As emoji usage varies widely between languages and regions(Kejriwal et al., 2021) the analysis will be limited to English text.

References

Bai, Q., Q. Dan, Z. Mu, and M. Yang (2019). A systematic review of emoji: Current research and future perspectives. *Frontiers in psychology 10*, 2221.

Boutet, I., M. LeBlanc, J. Chamberland, and C. A. Collin (2021). Emojis influence emotional communication, social attributions, and information processing. *Journal of Social and Personal Relationships 38*(2), 577–598.

Cherbonnier, A. and N. Michinov (2022). The recognition of emotions conveyed by emoticons and emojis: A systematic literature review.

Cramer, H., P. De Juan, and J. Tetreault (2016). Sender-intended functions of emojis in us messaging. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*, pp. 504–509.

Deho, B. O., A. W. Agangiba, L. F. Aryeh, and A. J. Ansah (2018). Sentiment analysis with word embedding. In *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)*, pp. 1–4. IEEE.

Gesselman, A. N., V. P. Ta, and J. R. Garcia (2019). Worth a thousand interpersonal words: Emoji as affective signals for relationship-oriented digital communication. *PLOS ONE 14*(8), e0221297.

Guibon, G., M. Ochs, and P. Bellot (2018). From emoji usage to categorical emoji prediction. In *International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 329–338. Springer.

Kejriwal, M., Q. Wang, H. Li, and L. Wang (2021). An empirical study of emoji usage on twitter in linguistic and national contexts. *Online Social Networks and Media 24*, 100149.

Kralj Novak, P., J. Smailovič, B. Sluban, and I. Mozetič (2015). Sentiment of emojis. *PLOS ONE 10*(12), e0144296.

Li, L. and Y. Yang (2018). Pragmatic functions of emoji in internet-based communication—a corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education 3*(1), 1–12.

Liao, S., J. Wang, R. Yu, K. Sato, and Z. Cheng (2017). Cnn for situations understanding based on sentiment analysis of twitter data. *Procedia computer science 111*, 376–381.

Liu, R., Y. Shi, C. Ji, and M. Jia (2019). A survey of sentiment analysis based on transfer learning. *IEEE access 7*, 85401–85412.

McCulloch, G. and L. Gawne (2018). Emoji grammar as beat gestures. In *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media, Standford [en línea]. Disponible en http://knoesis. org/resources/Emoji2018/Emoji2018_Papers/Paper13_Emoji2018. pdf [Consulta 11/12/2019].*

Mhatre, S., S. Satre, M. Hajare, A. Hire, A. Itankar, and S. Patil (2023). Text comparison based on semantic similarity. In *2023 3rd International Conference on Intelligent Technologies (CONIT)*, pp. 1–5.

Miller, C. H., Y. Zhang, and S. A. Sivo (2016). An exploration of the impact of emoticons on service quality perceptions. *Journal of Service Theory and Practice 26*(6), 787–807.

Minaee, S., E. Azimi, and A. Abdolrashidi (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. *arXiv preprint arXiv:1904.04206*.

Mollahosseini, A., B. Hasani, and M. H. Mahoor (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing 10*(1), 18–31.

Montani, I., M. Honnibal, M. Honnibal, A. Boyd, S. V. Landeghem, and H. Peters (2023, October). spacy: Industrial-strength natural language processing in python.

Na'aman, N., H. Provenza, and O. Montoya (2017). Varying linguistic purposes of emoji in (twitter) context. In *Proceedings of ACL 2017, student research workshop*, pp. 136–141.

Neel, L. A., J. G. McKechnie, C. M. Robus, and C. J. Hand (2023). Emoji alter the perception of emotion in affectively neutral text messages. *Journal of Nonverbal Behavior 47*(1), 83–97.

Quintes, C. and D. Ullrich (2019). Omg i'm laughing so hard–alienation in digital communication and potential countermeasures. *i-com 18*(3), 301–307.

Qurashi, A. W., V. Holmes, and A. P. Johnson (2020). Document processing: Methods for semantic text similarity analysis. In *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pp. 1–6.

Raj, H. W. and S. Balachandran (2020). Future emoji entry prediction using neural networks. *Journal of Computer Science 16*(2), 150–157.

Shardlow, M., L. Gerber, and R. Nawaz (2022). One emoji, many meanings: A corpus for the prediction and disambiguation of emoji sense. *Expert Systems with Applications 198*, 116862.

Srividya, K. and A. M. Sowjanya (2019). Aspect based sentiment analysis using pos tagging and tfidf. *International Journal of Engineering and Advanced Technology (IJEAT) 8*(6).

Steinert, S. and M. J. Dennis (2022). Emotions and digital well-being: On social media's emotional affordances. *Philosophy & Technology 35*(2), 36.

Tohti, T., S. Li, and A. Hamdulla (2021). A text similarity measurement employs semantic dictionary-based sentiment analysis. In *2021 International Conference on Asian Language Processing (IALP)*, pp. 358–362.

Vogel, C. and J. F. Janssen (2009). Emoticonsciousness. In *Multimodal Signals: Cognitive and Algorithmic Issues: COST Action 2102 and euCognition International School Vietri sul Mare, Italy, April 21-26, 2008 Revised Selected and Invited Papers*, pp. 271–287. Springer.

Walther, J. B. and K. P. D'addario (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review 19*(3), 324–347.

Wang, J., L.-C. Yu, K. R. Lai, and X. Zhang (2016). Dimensional sentiment analysis using a regional cnn-lstm model. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, pp. 225–230.

Wijeratne, S., L. Balasuriya, A. Sheth, and D. Doran (2017). A semantics-based measure of emoji similarity. In *Proceedings of the international conference on web intelligence*, pp. 646–653.

Yang, J., Y. Li, C. Gao, and Y. Zhang (2021). Measuring the short text similarity based on semantic and syntactic information. *Future Generation Computer Systems 114*, 169–180.

Zainuddin, N. and A. Selamat (2014). Sentiment analysis using support vector machine. In *2014 international conference on computer, communications, and control technology (I4CT)*, pp. 333–337. IEEE.

<u>Statement of Contribution</u>

Statement of Contribution:

All authors contributed equally to every section of the project. Here's a summary of each member's contributions:

- **Siddhesh Suresh Bangar (Chair):** Led the project, involved in all aspects including data collection, cleaning, research question development. Provided unique insights and ensured effective communication and collaboration within the group.
- **Shaunak Pedgaonkar (Recorder):** Responsible for detailed meeting minutes and maintaining project documentation. Contributed to research and actively participated in discussions. Ensured the team met objectives and deadlines.
- **Eamon Phelan (Accountant):** Ensured accurate record-keeping. Actively participated in all project aspects including research question development, literature review and research methods writing.
- **Parthiban Thandapani (Monitor)**: Ensured that everyone, provided a concise summary of a scholarly journal or conference article relevant to the project. Monitored the completeness of bibliographic details of each article, ensuring they were ready for inclusion in the mid-term synthetic review within the project essay.
- **Pragati Aboti (Verifier):** Ensured research integrity, oversaw team responsibilities, conducted literature review, and contributed to data collection and literature review. Ensured adherence to high research standards.
- **Jiaxuan Xie (Ambassador):** Participated in other group meetings, exchanged ideas, and facilitated discussions as well as helped in writing literature review.

In summary, each member made significant contributions to various aspects of the project, including research, data analysis, documentation, integrity assurance, and collaboration. All contributions were essential in achieving the project's success.

Siddhesh Suresh Bangar, Shaunak Pedgaonkar, Eamon Phelan, Parthiban Thandapani, Jiaxuan Xie, Pragati Aboti

Aboti Jiaxuan Xie T. Parthiban Shaunak. P