Emotion detection for written texts: techniques, their limitations and general challenges

Finn Alberts
Student pre-Master's programme Artificial Intelligence
Open University of The Netherlands
Email: finnalberts@gmail.com

Abstract—**Emotion detection is a subfield of sentiment analysis and focuses on identifying specific emotions such as happiness, sadness, or anger. This paper researches techniques for detecting emotions in written texts by studying literature from 2020 onwards. We zoom in on these different techniques, how they work and what their limitations are. We discuss the lexicon based approach, the rule based approach, machine learning, deep learning, transfer learning, transformer models, hybrid approaches, and the Multi-label Emotion Detection Architecture (MEDA). Apart from discussing limitations of each technique specifically, we discuss general challenges faced across all techniques. We especially look into challenges caused by a lack of datasets, both quantitatively and qualitatively. We also discuss challenges with detecting implicit emotions and sarcasm, as well as looking at the detection of multiple emotions for a single input. Future research should focus on enhancing dataset quality and quantity, exploring methods to better capture implicit emotions, and developing better models capable of detecting multiple emotions and their impact on each other. Addressing these challenges could significantly improve the accuracy and applicability of emotion detection technologies.**

I. Introduction

Everyone experiences emotions on a daily basis. Every day we face multiple emotions and express them in a number of ways, reaching from facial expressions to the things we say or write. By looking at emotions, we create more context and depth than by simply looking at the things explicitly stated.

Emotion detection is the branch of artificial intelligence (AI) that is focused on detecting emotions based on speech/voice, image, or textual data. More specifically, it is a branch of sentiment analysis [1]. There is, however, a key difference that makes emotion detection harder than sentiment analysis. Sentiment analysis only analyses if someone is expressing themselves positively, negatively, or neutral, whereas emotion detection aims to detect a specific emotion such as happiness, sadness, anger, or a combination of them [2].

Emotions can be detected by looking at multiple aspects, or by looking into one aspect, like written text, specifically. The latter does not take other elements into account such as facial expressions, and therefore makes it more difficult to correctly detect an emotion or multiple emotions [1].

Natural language processing (NLP) is the area of artificial intelligence that works on generating and understanding human language and is therefore closely related to detecting emotions in written text. Understanding human language does not solely involve understanding the explicit things stated, but also the feelings behind those things said.

Recognizing the emotions behind written texts can have many useful applications. For example, people nowadays use a lot of social media platforms, such as Facebook, Instagram, and YouTube. For businesses on these platforms it is useful to collect feedback about their products based on how people feel about them [3].

Another example use-case can be seen in the education sector. The quality of a teacher is not only determined by how well they know their subject, but also by how well they can teach it. To evaluate this, it is important to keep gathering feedback from students. However, when gathering feedback it can be difficult to derive conclusions on how students feel about their teacher manually. Here, emotion detection can help teachers evaluate themselves automatically [3].

There are many different approaches to detecting emotions in written text. This paper shows these different approaches, gives an overview of how they work, and highlights their limitations. Furthermore, this paper aims to show the general challenges for emotion detection for written text, applicable to all discussed approaches. The explicit research question this paper aims to answer is *"What are the limitations of current techniques for emotion recognition in written text?"*

To answer this research question, research has been conducted on existing literature solely.

The rest of this paper is built up as follows. Section 2 describes the research method used to gather relevant literature. Section 3 describes how emotion models work. The different techniques for detecting emotion and their limitations are discussed in section 4, and in section 5 general challenges for all techniques are presented. Section 6 describes related works, and the paper concludes with recommendations for future work in section 7.

II. Research method

A. Search method

To find the works used in this paper, a comprehensive search method was used. First of all, a selection of search queries was used in Google Scholar. These queries were combined with the criterion that all works must be published from 2020 onwards. This ensured that no works present outdated techniques for emotion recognition. An overview showing which query led to which source can be found in Table I.

The results of the queries were evaluated based on relevance for the topic of this paper. Here, we specifically looked for articles about emotion recognition in written text. Articles about emotion recognition in general, combining written text with other aspects such as audio and video, or articles about sentiment analysis were not deemed relevant.

Lastly, from articles that were deemed relevant for the topic of this article, information was extracted. We specifically looked at the techniques presented, their limitations, and limitations for emotion detection in written text in general.

TABLE I
OVERVIEW OF QUERIES AND RESULTING ARTICLES

Query	Article
emotion detection written text limitations	[1] [2] [3] [4]
emotion detection written text review	[1] [2] [3] [4]
emotion detection written text survey	[1] [2] [3] [4]
emotion recognition written text survey	[1] [2] [3] [4]
emotion recognition written text review	[1] [2] [3]
emotion recognition written text limitations	[1] [2]
emotion detection text limitations	[1] [2] [3] [4] [5] [6]

B. *Source evaluation*

The reliability of all relevant sources was evaluated based on a number of aspects. First of all, we looked at the number of citations for each article. Furthermore, the impact factor of the journal was taken into consideration.

We also looked at the reliability of the main author of the articles. This is done by looking at the h-index of the author, the amount of articles published by them, and the number of citations for all published articles combined.

It should be noted that some of the authors had published a small amount of works. In these cases, we focused on the h-index and amount of citations for the article to assess its reliability.

III. Emotion models

To be able to detect emotions, we first have to define what emotions are. There are two main models used to represent emotions: discrete emotion models and dimensional emotion models [2], [3], [4], [5].

Discrete emotion models place emotions into distinct categories, like happiness, sadness, disgust, surprise, and fear. There are multiple discrete emotion models, each defining their own categories of emotions [2], [3], [4], [5].

Dimensional emotion models, on the other hand, acknowledge that emotions are not independent of each other. This implies that there is a relation between emotions and that they therefore need to be placed in a dimensional space. Like discrete emotion models, there are also multiple variations of dimensional emotion models. They differ in the amount of dimensions (e.g. 1-D, 2-D or 3-D). The emotions are usually represented by vectors in these models [2], [3], [4], [5].

Because of its simplicity compared to dimensional based models, discrete based models have been widely adopted for emotion classification. The dimensional based models are more suitable for works where similarities in emotions are represented [2].

IV. Emotion detection techniques and their limitations

A. *Lexicon based approach*

In a lexicon based approach, also known as a keyword based approach, emotions are detected by looking for emotional keywords. These keywords are assigned to a specific emotion. The relations between the keywords and the associated emotions are captured within so called lexicons. There are some popular lexicons, like Word-Net-Affect, and NRC Word-emotion lexicon, which are openly available [3], [4].

The lexicon based approach relies heavily on keywords in the given text that imply a certain emotion. This means that this approach is very limited for detecting implicit emotions. Furthermore, the lexicon based approach is reliant on the keyword being present in the lexicon. Lastly, a challenge for the lexicon based approach is context. The reason for this is that the meaning of a word can drastically change by the context of the sentence (e.g. ”Does it seem like I am happy?”) [4].

B. *Rule based approach*

The rule based approach encompasses the lexicon based approach and expands upon it. The key difference is that the rule based approach does not just look at specific keywords, but extracts rules based on linguistics, statistics, and computational concepts [2], [4].

The same problem as seen by the lexicon based approach is seen here. The context of the sentence makes it possible that a sentence is misclassified [2]. Furthermore, rule based approaches are affected by the quality of the text. For example, grammatical mistakes may cause the model to be unable to classify an emotion correctly. Additionally, detecting implicit emotions is only possible if a rule in the rule set represents it [4].

C. *Machine learning based approach*

Another possible approach is the machine learning based approach. Both supervised learning and unsupervised learning are options for detecting emotions [2]. The machine learning based approach can be implemented using one of the more traditional machine learning approaches, such as Naïve Bayes, decision trees, support vector machines (SVM), logistic regression or conditional random fields (CRF) [2], [3]. According to Alswaidan and Menai, SVM is the most used machine learning based approach [4].

In recent years, it has been seen that deep learning has been more robust compared to traditional machine learning based approaches. Deep learning also shows to be outperforming these traditional approaches [2]. Deep learning is explored further in the next section.

D. Deep learning based approach

Deep learning is a part of machine learning that tries to process information in the same way the human brain does [3]. Here, the program learns from experience and breaks down complicated concepts into simpler ones [4].

When using deep learning for emotion detection, there are a few steps involved. First, an embedding layer is built. This layer converts the input into a vector representation. This embedding layer is then fed into a neural network. This neural network finally gives an output [4], [5].

For both the embedding layer and the neural network there are multiple options. According to Deng and Ren, there are two types of embeddings: typical word embeddings, which are more general and word embeddings specifically designed for emotion detection, called emotional word embeddings [5]. For the neural network options include recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent neural nets (GRNNs) [5]. LSTMs are the most used for emotion detection for written text [4].

E. Transfer learning based approach

Transfer learning is not an approach by itself, but helps to improve machine learning (and especially deep learning). The transfer learning approach allows one to reuse existing models. This is done by using a model from a source domain and transferring it to a target domain. [3], [5].

Transfer learning can be done in multiple ways. The most frequently used method in NLP is using sequential transfer learning. This method consists of two stages. The first stage is the pre-training phase, where we try to train a model with universal knowledge of the natural language. In the second stage, the transfer phase, the model is finalized to be used for the target domain [5].

Another approach for transfer learning is multi-task learning (MTL). Here, target and source tasks are related and trained simultaneously. Compared to sequential transfer learning this method is less vulnerable to overfitting [5].

F. Transformer models based approach

Transformer models are a branch of deep learning where the importance of the input data is weighed differently, based on so-called attention vectors in an attention layer. In NLP tasks, it has created many state of the art solutions, since its birth in 2017 [1].

There are many different transformer models, including Transformer-XL, Generative Pre-Training (GPT), Bidirectional Encoder Representations from Transforms (BERT), Cross-Lingual Language Model (XLM), and XLNet. As of 2021, BERT is the most researched transformer based model for emotion detection in written text [1].

When specifically looking at BERT, we can see that it is performing well on extracting contextual information. However, it can only detect emotions in a monolingual text. Furthermore, the size of the input sentence is limited. Lastly, it suffers from pragmatic inference, which means it can make assumptions based on previous knowledge, but is not necessarily true for the given input [1].

G. Hybrid approaches

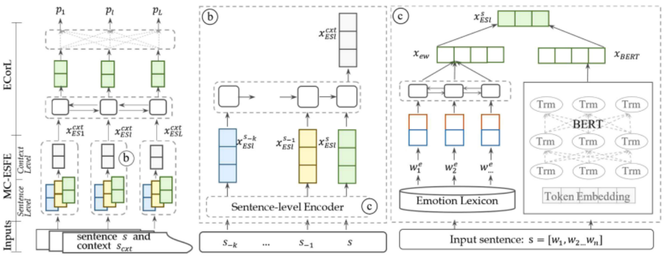
In the hybrid approach, the rule based approach and machine learning based approach are combined into a unified model. It is also possible to combine the rule based approach with the deep learning based approach or transformer models based approach, as they are both machine learning techniques.

By combining the two techniques, the strengths of both techniques can be utilised, while also making the limitations of the individual techniques less of an issue. However, the hybrid based approach relies on the type of deep learning technique used and therefore requires a well-performing deep learning model to have satisfactory results [2].

H. Multi-label Emotion Detection Architecture (MEDA)

In most of the techniques mentioned before, emotions are seen as independent of each other and detected as such. However, emotions are closely related and impact each other. This means that emotions have correlations, which can be used to increase performance (e.g. it is less likely that sadness and happiness are both present in a given sentence) [6].

The Multi-Label Emotion Detection Architecture (MEDA) extracts both emotion-specified features and emotion correlations. The architecture consists of two parts: the Multi-Channel Emotion-Specified Feature Extractor (MC-ESFE) and the Emotion Correlation Learner (ECorL). In the MC-ESFE, each emotion is detected seperately in a dedicated channel. Then, in the ECorL, the correlations between emotions are taken into consideration to give the final prediction [6]. MEDA is shown in Figure 1.



(a) MEDA

(b) *lth* ESFE channel of MC-ESFE

(c) Sentence-level encoder of *lth* ESFE channel

Fig. 1. The Multi-Label Emotion Detection Architecture (MEDA) [6]

Although MEDA shows satisfactory results, one of its limitations is that weaker emotions can be suppressed by stronger emotions. This can cause the architecture to not detect these emotions [6].

V. General challenges

Apart from the limitations of each of the techniques mentioned in section 4, there are also some challenges for emotion detection for written text in general. One of the main issues is the quality and quantity of (labeled) datasets.

One of these dataset related issues is spelling mistakes or incorrect grammar. This is an issue especially for datasets based on data from social media sites (e.g. ”Y have u been soooo late?” has many mistakes, but is common on these sites). These spelling and grammar errors can cause the detection to be more difficult. This same issue applies for new terms and words that are not yet represented in the datasets. Emoji are not well-represented in these datasets as well [3].

Another issue with the available datasets is that most of the datasets are in English and few other languages are available. This makes it harder to detect emotions in these other languages [2], [3], [4].

The same issue of language goes for the domains the datasets are applicable to. Many domains are yet to be represented well in datasets, making it harder to detect emotions in these domains [2], [3].

Furthermore, many of the datasets show an imbalance in the representation of emotions in them. For example, a dataset may have many ”happy” labeled data, but a lot less ”sad” labeled data [4], [5].

It is possible to create more datasets, not only to increase the amount of available data, but also have more data for currently underrepresented emotions or domains. It should be noted that labelling data however is a time intensive task. Furthermore, the labelling of the data can be subjective, based on the interpretation of the person doing the labelling [3], [5].

This last point is not only applicable when labelling the data. Boundaries between emotions are fuzzy, making it not only hard for people to properly classify an emotion, but for emotion detection algorithms as well [5].

Another issue making detecting emotions difficult, is implicit emotions. Implicit emotions are emotions which are not explicitly stated in the input, but can logically be derived from it. This, however, asks for the algorithm to understand the linguistics and context of the text, which is a difficult task [4].

Closely related to this previous issue, are the difficulties that arise with sarcasm and irony. These can point an algorithm in the completely opposite direction of the emotion actually present in the input [3].

Another difficulty in emotion detection for written texts is detecting emotions in dialogue, where one sentence is a reaction to another. Most techniques presented in this paper detect emotions on a sentence level and therefore ignore the context of the rest of the conversation, which may include crucial information for detecting the correct emotion [5].

Lastly, even though we presented MEDA as a possible architecture for detecting multiple emotions, this remains a challenge. Emotions in a sentence can contradict each other and weaker emotions might be surpressed by stronger emotions [3], [6].

VI. Related works

This paper is based on multiple different works. Nandwani and Verma presented a review on emotion detection for written text and gave insight into different emotion detection techniques. They discussed a lexicon based approach, a machine learning based approach, a deep learning based approach, and a transfer learning based approach. They also discussed challenges for emotion detection in written text in general [3].

Alswaidan and Menai dive deeper into some of the same approaches, as well as introducing some new ones. Techniques discussed in their paper are the keyword based approach, the rule based approach, classical learning (machine learning), deep learning, and hybrid approaches. They also present some general challenges for these techniques [4].

Acheampong, Wenyu and Nunoo-Mensah also discuss some of these techniques, specifically being the rule based approach, machine learning, and the hybrid approaches. They too discuss general challenges for emotion detection in written text [2]. In a later paper they zoom in on transformer models for detecting emotions. They specifically zoom in on BERT based approaches [1]

Deng and Ren as well explored emotion detection techniques focused especially on deep learning. They brought insight into the inner workings of these deep learning approaches and also discussed the different options for deep learning there are [5].

Deng and Ren also developed an architecture (MEDA) for detecting multiple emotions and the correlations between them. By looking at these correlations, MEDA aims to improve the detection of emotions [6].

This paper aimed to gather information from all these different sources, to create a single overview of all the techniques presented in these different papers. Furthermore, this overview explicitly states weakness and limitations for these techniques, as to create a complete overview of where there are open issues are for all emotion detection techniques.

VII. Conclusion

In summary, this paper discussed different techniques for detecting emotions in written text by researching existing works. These works were used to create an overview of techniques, their limitations, and challenges in general.

We discussed the lexicon based approach where emotions are detected by looking at specific keywords in the input. It is, however, limited in detecting implicit emotions.

We also discussed the rule based approach, which expands on the lexicon based approach by extracting rules based on linguistics, statistics, and computational concepts. It does however suffer the same issues as the lexicon based approach and is strongly reliant on the quality of the text.

Furthermore, we discussed machine learning approaches, also including deep learning techniques, transfer learning approaches, and transformer models based approaches. These techniques have less limitations on their own, but are still limited by the general challenges emotion detection faces.

Hybrid approaches were also discussed, in which the rule based approach and machine learning based approach are combined. This approached helped in making the issues for these techniques on their own smaller, but is reliant on a well-performing deep learning model.

Lastly, the MEDA architecture was discussed as a technique for detecting multiple emotions and their correlations. By detecting these correlations, the detection could be improved. Weaker emotions, however, could be suppressed by stronger emotions.

Apart from limitations of the individual techniques, we also discussed some general challenges. A main challenge is the lack of enough high-quality datasets. There are many issues related to this, including a lack of non-English datasets and an imbalance of emotions.

Another issue discussed is the issue of fuzzy boundaries between emotions, making it hard to draw a hard line between them. Not only can this affect the correct labelling of the datasets, but also for the algorithms this makes it harder to properly classify written text.

There also lay difficulties in implicit emotions, where emotions are not as visible as when explicitly stated. In these cases, the context is very important. The same goes for sarcasm and irony.

Lastly, even though MEDA aims to bring a solution to detecting multiple emotions, it remains difficult to do so.

Further research is needed to solve these issues. Especially improvements for the datasets could have a big impact, as many issues are related to them. Furthermore, research in how we can improve detecting implicit emotions and multiple emotions is needed. Research in these areas could open up new doors for emotion detection in written text. This could create many new use-cases for these techniques.

REFERENCES

[1] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: a review of bert-based approaches," *Artificial Intelligence Review*, pp. 1–41, 2021.

[2] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, p. e12189, 2020.

[3] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, 2021.

[4] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems*, vol. 62, pp. 2937–2987, 2020.

[5] J. Deng and F. Ren, "A survey of textual emotion recognition and its challenges," *IEEE Transactions on Affective Computing*, 2021.

[6] ——, "Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning," *IEEE Transactions on Affective Computing*, 2020.