Data & Knowledge Engineering 148 (2023) 102220

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

Mixed emotion extraction analysis and visualisation of social media text

Yuming Li, Johnny Chan *, Gabrielle Peko, David Sundaram

University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

A R T I C L E I N F O

Keywords:
Sentiment analysis
Emotion analysis
Multi-label classification
Knowledge graph

A B S T R A C T

With the widespread use of social media and accelerated development of artificial intelligence, sentiment analysis is regarded as an important way to help enterprises understand user needs and conduct brand monitoring. It can also assist businesses in making data-driven decisions about product development, marketing strategies, and customer service. However, as social media information continues to grow exponentially, and industry demands increase, sentiment analysis should no longer be limited to fundamental polarity classification of positive, neutral, and negative. Instead, it should move to more precise classification of emotions. Therefore, in this paper, we expand sentiment analysis to analysis of eight different emotions based on Plutchik's wheel of emotions, and define it as a multi-label classification task to identify complex and mixed emotions in text. We achieved an overall precision of 0.7958 for the eight emotions multi-label classification based on the attention-based bidirectional long short-term memory with convolution layer (AC-BiLSTM) model on the SemEval-2018 dataset. In addition, we proposed the introduction of the NRC emotion lexicon and emotion correlation constraints to optimise the emotion classification results. This ultimately increased the overall precision to 0.8228 demonstrating the effectiveness of our approach. Finally, we store and visualise the emotion analysis results in a graph structure, in order to achieve deductibility and traceability of emotions.

1. Introduction

1.1. Background

With the ever-growing popularisation of the web, social networks have gradually become the main platform for people to express themselves or exchange ideas with others. Along with the huge number of users, vast quantities of text and other forms of user-generated contents follow. People are accustomed to post short texts on the Internet to express their opinions and feelings, which provides abundant data for natural language processing (NLP) researchers, and channels for institutions to glean public opinion. Amongst all the different types of information generated by the expanding social media user base, we choose to focus on the analysis of emotion related information in this paper. As a crucial task in computational linguistics and artificial intelligence, sentiment analysis has a long history which can be dated back to the 1960s [1]. Sentiment refers to the attitude (positive, neutral, and negative) conveyed by a piece of text, while sentiment analysis is the task of identifying and extracting this information from textual data [2].

In social media platforms such as Twitter and Facebook, sentiment analysis is important for monitoring brand reputation, tracking

* Corresponding author
E-mail address: jh.chan@auckland.ac.nz (J. Chan).

https://doi.org/10.1016/j.datak.2023.102220
Received 19 March 2022; Received in revised form 18 July 2023; Accepted 5 September 2023
Available online 11 September 2023
0169-023X/

customer feedback, and identifying emerging trends. By analysing sentiment, companies can make data-driven decisions about marketing strategies, customer service, and overall business operations. However, with the explosive increase in the amount of information and the rapidly increasing needs of industry, the recognition of positive, neutral, and negative polarities of text by traditional sentiment analysis is unable to meet the needs of users. Rather, it should incorporate more precise classification of emotions [3]. For example, when people express negative sentiment, distinguishing whether such sentiment is caused by anger, fear or sadness is of greater significance for further analysis and opinion mining. This has led to the emergence of emotion analysis as a field of research from sentiment analysis. In NLP, sentiment and emotion are related but distinct concepts, sentiment refers to the emotional tone or attitude conveyed by a piece of text, while emotion refers to specific affective states. Thus, sentiment analysis focuses on identifying the polarity (positive, neutral, and negative) of text, while emotion analysis aims to identify specific emotions such as happy, sad, angry etc. [4].

In this paper, we extend the scope of analysis of social media from the traditional sentiment level to the emotion level. Based on Plutchik‘s wheel of emotions [5] shown in Fig. 1, this paper divides the target emotions to be recognised into eight emotion categories: trust, fear, surprise, sadness, disgust, anger, anticipation, and joy.

Furthermore, the expressions from users of social network via short texts are more casual than formally written texts from newspapers or reports that have recognised structure and format. A post or a tweet could contain multiple emotions - "I am surprised that I will be so angry" has embedded both anger and surprise. Therefore, the traditional emotion analysis which belongs to single-label classification that one piece of text corresponds to one label of emotion may not be sufficient for this type of context [6]. In order to avoid omission, we need a better approach to extract all the emotions that could be mixed in the same piece of text. In this paper, we define social media emotion classification as a multi-label classification task. It involves identifying the presence of multiple emotions (one or more specific emotion labels) in a piece of text, which can provide more nuanced insights into the affective content of the text. After defining emotion analysis as a multi-label classification task, the paper uses two mechanisms to optimise the emotion analysis results based on neural network models. The first mechanism involves adding the emotion lexicon to training data, while the second mechanism involves leveraging the co-occurrence and correlation patterns of emotions for further optimisation.

The correlation between the emotion labels is an important factor influencing the accuracy of multi-label emotion classification. While emotions in Plutchik’s wheel are correlated to each other based on their spatial displacement [7], the relationships between the basic emotions are not statistically clear. Therefore, we visualise the correlation between emotion labels in the SemEval-2018 Task 1 dataset [8], which is also the dataset we use for evaluation of our proposed model (the more detailed data set descriptions and statistics are illustrated in Section 3.1). This dataset includes not only the eight emotions based on Plutchik’s wheel, but also three other labels. In view of the scope of this research on emotion, we combined the mainstream classic emotion theories and emotion classification models [9,10]. We finally selected Plutchik’s wheel, which has 8 emotions that highly overlap with the emotion classification theory model, as the range of emotions for our research. Fig. 2 shows the correlation of the selected emotion labels from the SemEval-2018 Twitter sentiment dataset. The cell element represents the dependency measurement between the corresponding two emotion labels, and the size of the measurement value is distinguished by colour. The brighter (or darker) the colour of the cell, the stronger (or weaker) the dependency relationship between the two emotion labels.

From Fig. 2, we can find the co-occurrence between emotions. For example, anger is often accompanied by disgust; and the chance for anger and trust to co-occur is extremely low. Accordingly, the emotions of fear, sadness, disgust, and anger have higher chances to co-occur as they are associated with the negative polarity; and the emotions of trust, anticipation and joy have higher chances to co-

Fig. 1. Wheel of emotions adopted from Plutchik [5].

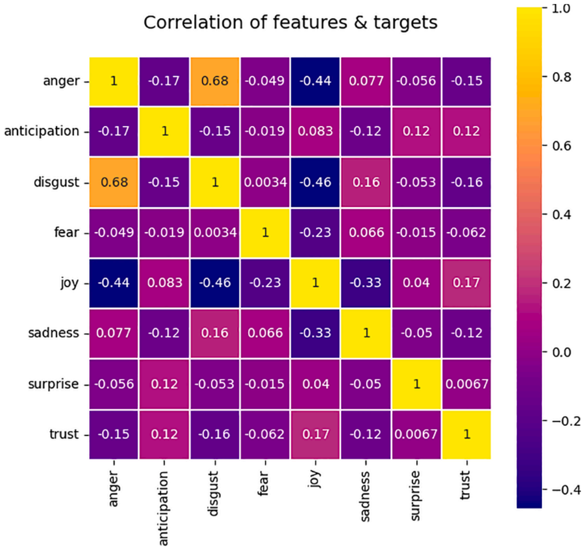

Fig. 2. Correlation of emotion labels in SemEval-2018 dataset.

occur as they are associated with the positive polarity. Therefore, we use them as theoretical support for our optimisation mechanism. We introduce emotion correspondence for optimising our multi-label emotion classification.

In addition, sentiment analysis and emotion analysis are usually followed by downstream activities like user behaviour prediction, brand analysis or public opinion mining. These tasks not only require the classified emotions as output, but they also need to know the internal reason for each emotion in order to conduct a more comprehensive analysis for decision support. The traditional table-based structure for data storage is relatively insufficient in inference and reasoning [11,12]. We introduce a graph-based structure to organise and store the tags, users, emotion categories and polarities classified from Twitter, and to construct an "emotion graph". As a derivative of the traditional knowledge graph, our emotion graph uses tags as nodes, and the associated users, polarities, and emotion categories as relationships. The emotion graph could describe the emotions of users among different tags more intuitively, and support subsequent inference, reasoning, and analysis.

At the algorithm level, most of the existing studies often choose to use a state-of-the-art neural network models to achieve more than 90% accuracy on multi-emotion classification tasks [3]. But these models ignore the recognition of complex emotions and mixed emotions, and only considers the recognition of a single emotion category. At the data storage and data visualisation level, there is sparse research that stores and visualises emotion classification results in a graph structure. The key contributions of this paper are focused on addressing these deficiencies.

1.2. Contribution

From an application perspective, this paper develops an emotion classification method that improves the accuracy and richness of sentiment analysis in social media compares to existing research. It considers the complexity of emotion attributes and mixed emotional expressions from social media texts, and expands emotional information classification from three basic sentiment polarities to eight additional emotion categories.

More importantly, the proposed emotion graph presents new opportunities for novel analysis techniques and future research. Since the emotions of users on specific topics expressed via social media texts often have some degree of transitivity and causality, they could influence each other and shape the direction of public opinion. Traditional sentiment analysis is not only limited to basic polarity classification from each post, but it also ignores the cause and effect among posts on the same topic. The emotion graph could go beyond such limitations to support reasoning and representation. For example, when the anger emotion on a topic exceeds a pre-defined threshold from a social network, the emotion graph could help track the related topic and users involved for follow-up monitoring and predicting in real-time. Using the out-degree and in-degree of the graph structure, it could quickly locate opinion leaders and relevant information within a specific range. This vastly improves the effectiveness of monitoring and management of emotional communication in social networks.

From a technical perspective, our contribution could be divided into the following:

- We transform social media sentiment analysis from a single-label classification task with three sentiment labels to a multi-label classification task with eight emotion labels, which is conducive to identifying complex and mixed emotions among social media texts.
- We propose a multi-label emotion classification hybrid model based on a dictionary and neural network, and we further optimise the classification results by utilising the dependence and co-occurrence rules between emotion labels.

- We use a graph-based structure instead of a table-based structure to organise, store and visualise the results of emotion analysis to show the potential internal connections and causalities among social media texts.

2. Related work

The recognition and judgment of human emotions through intelligent methods has a certain research history and research foundation that can be traced back to early efforts in computational linguistics to classify texts according to their emotional tone [13]. Over the years, researchers have developed a range of approaches to automatically identify and extract emotional information from text data, including rule-based, machine learning, and deep learning models [14].

First, we provide an overview of the algorithms used for emotion analysis tasks. As emotion analysis is an emerging topic that builds upon sentiment analysis, we combine the literature review of both topics to explore a systematic framework for identifying emotional information in texts.

The algorithms for processing sentiment and emotion analysis are roughly divided into dictionary based methods and deep learning based methods. Emotion analysis through the dictionary or lexicon is a classic method, which has achieved good results in theory and practice [15–18]. However, the dictionary based method comes with some problems: the scarcity of corpus resources, the low update frequency of emotion words, and insensitivity to new words, derived words, and different tenses of the same word. In addition, it is highly dependent on domain, time, language and other conditions, and subsequent expansion is relatively difficult. With the emergence and development of machine learning technologies, deep learning based methods are gradually overtaking the dictionary based method in textual emotion analysis [19]. This is due to the strong text representation ability of deep learning and its superiority to extract the semantic, syntax, grammar, and other deep abstract features. The three types of widely used deep learning based methods are convolutional neural network (CNN) [20,21], recurrent neural network (RNN) [22–26], long short term memory (LSTM) [3] and hybrid [27–29]. Recently, researchers have discovered a large number of mixed emotions among social network texts. The disadvantage of a deep learning based method is that it relies too much on the manually annotated corpus, and it cannot achieve good results if the sample set is small. Therefore, in this paper we choose to combine a dictionary based method with a deep learning based method to improve the overall effectiveness of emotion classification.

However, the above-mentioned widely cited literature mostly focus on the recognition of a single emotion status contained in a text, and only a few studies focus on the recognition of multiple mixed emotions [30]. The essence of sentiment and emotion analysis is a classification task, while recognising one or more emotions contained in the text is usually transformed to a multi-label classification problem at the technical level. At present the methods of constructing multi-label classification models are divided into two categories: problem conversion and algorithm adaptation [31]. Problem conversion aims to convert a multi-label classification problem into multi single-label classification or multi-class classification problems. The representative methods include binary relevance (BR) [32,33], classifier chain (CC) [34,35], random k-label set (RAkEL) [36], and calibrated label ranking methods (CLR) [37,38]. Algorithm adaptation directly changes the single-label classification task to multi-label classification scenarios. The representative methods include Support Vector Machine (SVM) based approach [39,40] and k-Nearest Neighbor (KNN) based approach [41,42].

Considering the construction of knowledge graph, they could be divided into general knowledge graph and domain knowledge graph. A general knowledge graph is also known as an encyclopaedia knowledge graph, such as Freebase [43], YAGO [44] and DBpedia [45], which use Wikipedia as the data source. A domain knowledge graph often uses a certain domain dataset to define personalised knowledge nodes and relationships. Specific applications include movie knowledge graphs [46,47], medical knowledge graphs [48–50], character knowledge graphs [51], biological knowledge graphs [52,53], geology knowledge graphs [54] and educational knowledge graphs [55].

A general knowledge graph adopts a bottom-up construction method. Specifically, it uses certain technologies to obtain content that may be the target entity or relationship, verifies whether its confidence is up to the standard through review, and then adds it to the knowledge graph. A domain knowledge graph adopts a top-down construction method. It pre-defines the category set of entity category attributes and relationships, follows the defined categories to extract the data, and then adds it to the knowledge graph. The emotion graph is a form of a domain knowledge graph, and most of the existing studies use a knowledge graph as a background tool for emotion analysis [56–58], rather than using it to structure the result of the emotion analysis.

From the above literature, we can summarise three key gaps in existing research. The first research gap is the classification of emotional information. Most of the existing research roughly divide them into three polarities of positive, neutral, and negative. Few studies focus on specific emotion attributes, which is obviously insufficient in the context of the increasing abundance of social media text. The second research gap is the definition of emotion analysis. Most studies define emotion analysis as a single-label classification task, that is, to identify only one final emotion contained in the text. But in real-world situations, emotions contained in social media texts are usually complex and mixed, so it is not enough to identify only one emotion. We need to redefine emotion analysis as a multi-label classification task, i.e., identifying one or more emotions contained in the text, to be more realistic. The third research gap is the visualisation and analysis of the emotion extraction result. Existing research usually stores emotion in a single table-based structure, while ignoring the inner connection and reasoning of emotion and emotion subjects. Knowledge graph shows outstanding advantages in reasoning and representation, but the existing knowledge graph application scenarios have almost no emotional relevance. Therefore, combining emotions and knowledge graph would significantly enhance classification.

3. Methodology

To address the research gaps, we first redefine the emotion analysis as a multi-classification task of eight emotions. We also propose a hybrid multi-label emotion classification model which is based on the attention-based bidirectional long short-term memory with convolution layer (AC-BiLSTM) model [59], external emotion lexicon and correlation processing. Since accuracy is much more difficult to achieve with the multi-label classification task than the single-label classification task, we use a hybrid model to synthesise the advantages of each module to ultimately improve the accuracy of the model. The BiLSTM module can process both single-label and multi-label classification tasks as well as the information that follows at the same time. Thus, the model can capture the context information to make more global judgment. Convolution layer could reduce the amount of data processing while retaining useful information. Attention mechanism is beneficial to capture important parts of long texts to increase the weight of important information. Finally, the introduction of external emotion lexicon and correlation processing allows us to add constraints on the basis of the original model to further improve the accuracy of the emotion classification task. In addition, we choose to store and visualise the emotion analysis results based on a knowledge graph. The purpose is to clearly express the relationship between emotion related subjects and emotion categories in the text through the advantages of reasoning and representation of the knowledge graph. At the same time, it is beneficial to explain and predict in combination with their internal connections during analysis. This helps us to address the research gap in visualisation and analysis of the emotion extraction result.

The overall model is presented in Fig. 3, with the related algorithm shown in Algorithm 1.

3.1. Training data with NRC emotion lexicon

In this paper, we use SemEval-2018 Task 1 dataset combined with NRC emotion lexicon as the overall training data for model training and evaluation. SemEval-2018 Task 1 is a dataset consisting of tweets labelled with one or more of eight emotion categories: anger, anticipation, disgust, fear, joy, love, sadness, and surprise. The dataset was created for the SemEval-2018 shared task on "Affect in Tweets," which aimed to advance research in NLP for emotion analysis. The dataset contains a total of 9,595 tweets, which were manually annotated by annotators through a crowdsourcing platform. The tweets were collected using several keyword queries related to the eight emotions, and were preprocessed to remove URLs, emojis etc. [8].

The NRC emotion lexicon was created by experts from the National Research Council of Canada. It is a widely used resource as training data for emotion analysis. Its use in training data can improve the accuracy and reliability of emotion analysis models by providing a standardised set of labelled data. It contains 105 languages with 14,182 vocabulary emotion labels. The classification criteria is based on Plutchik's wheel of emotions [5], including positive and negative polarities and multi-label classification of eight emotion categories: trust, fear, surprise, sadness, disgust, anger, anticipation, and joy. Since this paper only focuses on emotions, we focus on the data statistics of eight emotions. The distribution of labels in the SemEval-2018 Task 1 dataset and NRC emotion lexicon is shown in Table 1. Since the data set is multi-labelled annotation, we count the total number of samples containing each emotion label and concatenate the NRC Emotion Lexicon with the original training set as input training data to train the model.

As discussed in Section 2, deep learning approaches for text classification have been widely used and proved to have superior performance when utilising recurrent neural network (RNN) based models. Long short-term memory (LSTM) and gated recurrent units (GRU) are variants of RNNs which are superior to RNNs because they address the vanishing gradient problem and can learn long-term dependencies more effectively [60]. Therefore, we conducted a comparative experiment using the SemEval-2018 Task 1 dataset, comparing several existing LSTM and GRU based models, to determine the optimal backbone model. We have compared 5 different

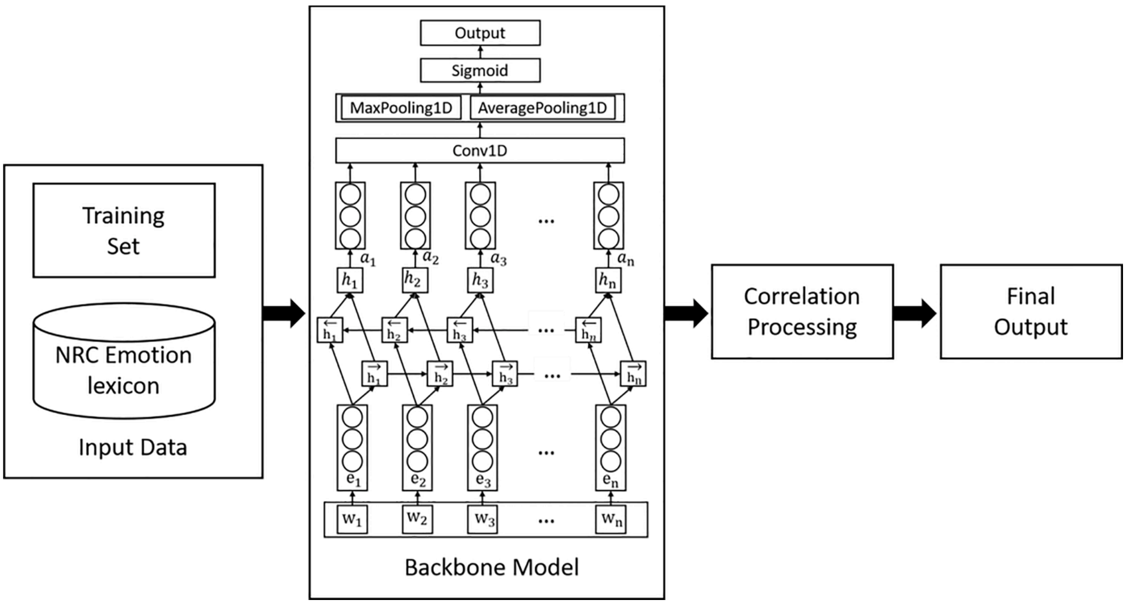

Fig. 3. Architecture of the overall model.

Algorithm 1
Framework of multi-label emotion classification model.

Require: D_{train}: training data, *dic*: Emotion Dictionary, D_{set}:training corpus, C: target class, cw: class weight, dp: dropout rate;
Ensure: model with the best weight, accuracy, precision, recall, F1-score;

$D_{train} = dic + D_{set}$;
for i=1 to *epoch* **do**
for data in D_{train} **do**
$S = tokenise(data)$
for w in S **do**
$e = embedding(w)$;
$E.append(e)$;
end for
$X = Dropout(E, dp)$;
$H = BiLSTM(X)$;
$A = attention(H)$;
$X_{out} = Conv1D(A)$;
$avg_{p}oolGlobalAveragePooling1D(X_{out})$;
$max_{p}ool = GlobalMaxPooling1D(X_{out})$;
$X_{out}concatenate[avg_{p}ool, max_{p}ool]$;
$output = Sigmoid(X_{out})$;
$loss = compute_{l}oss(output, C, cw)$;
do $backpropagation(loss)$;
end for
Evaluate Loss, Accuracy;
end for
return *output*;

Table 1
The data statistics of training data.

Emotion	Number of Samples in SemEval-2018 Task 1 dataset	Number of Samples in NRC emotion lexicon
Anger	1160	1247
Anticipation	1086	839
Disgust	1065	1058
Fear	1072	1476
Joy	1152	689
Sadness	1144	1191
Surprise	1741	534
Trust	1075	1231

combinations of the model, which are gated recurrent units with convolution layer (C-GRU), long short-term memory with convolution layer(C-LSTM), bidirectional gated recurrent units with convolution layer (C-BiGRU), bidirectional long short-term memory with convolution layer (C-BiLSTM) and attention-based bidirectional long short-term memory with convolution layer (AC-BiLSTM) Through this preliminary experiment, we found that the AC-BiLSTM model achieved the highest accuracy and selected it as the backbone model in our methodology (the results of the comparative experiment are shown in Table 2). Subsequently, in Sections 3.2-3.5, we provide a detailed description of each layer of this model.

Furthermore, the parameter setting of the AC-BiLSTM model, selected as the backbone model, is shown in Table 3:

The structure of the backbone model is shown in Fig. 4, which is divided into the input layer, embedding layer, BiLSTM layer, self-attention layer and convolution layer. These are discussed in the following sections.

3.2. Embedding layer

The embedding layer takes as its input, text T, which is composed of a series of words w_1,$w2$,...,w_n, where n is the length of the input text. Words are the abstract summary of human thought, and words need to be converted into numerical forms for subsequent calculations during text processing. In order to convert the text into a number that can be recognised by the computer, we used the global

Table 2
The comparative experiment results of backbone model.

Model	Loss	Accuracy
C-GRU	0.3263	0.3790
C-LSTM	0.3370	0.3649
C-BiGRU	0.9087	0.4046
C-BiLSTM	0.3108	0.3942
AC-BiLSTM	0.2096	**0.4296**

Table 3
Parameter setting of the AC-BiLSTM model.

Parameter	Value
max_features	16803
maxlen	150
embed_size	300
batch_size	32
epochs	10

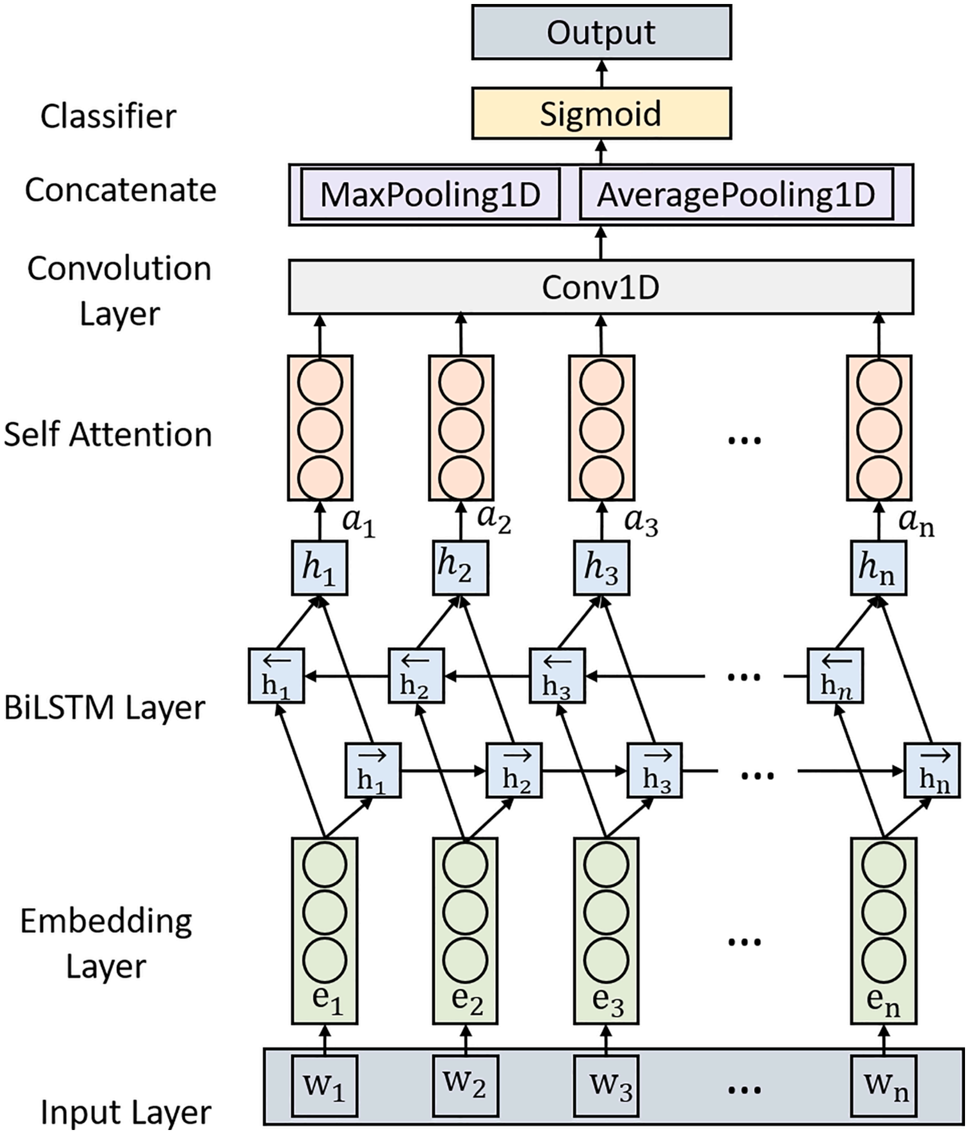

Fig. 4. Architecture of the backbone AC-BiLSTM model.

vectors for word representation (GloVe) pre-training model as word embedding to initialise the embedding layer of the neural network. GloVe is a word representation tool based on global word frequency statistics, which can express a word as a vector of real numbers. These vectors capture some semantic features between words, such as similarity and analogy. The implementation of GloVe is divided into the steps described in the following paragraphs.

The basic idea behind the GloVe model is to create a co-occurrence matrix X, where each element X_{ij} represents the number of times word i and word j co-occur in a context window of a fixed size. This co-occurrence matrix can then be used to calculate a set of word vectors that capture the relationships between words in the corpus. The first step in the GloVe model is to construct a vocabulary of all the unique words in the input corpus. Each word is assigned a unique index in the vocabulary, which is used to identify it in the co-occurrence matrix. Next, a context window of size d is defined around each word in the corpus. The context window is used to capture the co-occurrence statistics of words. For each word w in the corpus, all the words that appear within a distance of d to the left and right of w are considered to be in its context.

Once the context windows are defined, the co-occurrence matrix M is constructed by counting the number of times each word appears in the context of every other word. This matrix is then used to calculate a set of word vectors using a weighted least squares regression model. The objective of the GloVe model is to learn word vectors such that the dot product between any two vectors reflects the co-occurrence statistics of the corresponding words. Specifically, the model minimises the following cost function:

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(e_i^T e_j + b_i + b_j - log\left(X_{ij}\right)\right)^2$$

Where e_i and e_j are the word vectors for words i and j, b_i and b_j are the bias term of the two word vectors respectively. $f\left(X_{ij}\right)$ is a weighting function that downweighs the contribution of rare co-occurrences, and $log(X_{ij})$ is the logarithm of the co-occurrence count.

The model is trained using stochastic gradient descent to minimise the cost function J. The final output of the GloVe model is a set of dense vector representations $e_1, e_2, \ldots, e_n$ for each word in the vocabulary, where n is the size of the vocabulary. These vectors can be used for a variety of natural language processing tasks, such as text classification, sentiment analysis, and machine translation. In this paper, the vectors would be passed to BiLSTM layer, which will be further illustrated in section 3.3.

3.3. BiLSTM layer

BiLSTM is a variant of RNN, which mainly includes forward LSTM and backward LSTM. The RNN possesses the ability to acquire global feature information, but has the drawbacks of gradient vanishing and gradient explosion [61]. However, the LSTM neural network can avoid these issues. At a specific time t, the basic neural unit of the LSTM model is composed of a memory cell c_t, and forget gate f_t, input gate i_t, and output gate o_t. The memory cell is a self-connected unit that can retain long-range contextual information, while these three gates jointly determine how to update the current memory cell c_t and current hidden state h_t.

Its state transition formula at time step t is described in the following paragraphs.

Based on the description in section 3.2, the input to the BiLSTM layer is the word vector e_t obtained using GloVe. The input gate i_t controls how much new information should be stored in the current storage unit, which is calculated using the previous output h_{t-1} and the current input e_t.

$$i_t = \sigma(W_i[h_{t-1}, e_t] + b_i)$$

Among them σ is the activation function Sigmoid, the output value is between 0 and 1, W_i is the weight matrix of the input gate, and b_i is the bias.

The forget gate f_t can be regarded as a function that controls the degree to which information from the old memory cell is discarded. The forget gate in BiLSTM is also calculated using the previous output h_{t-1} and the current input e_t.

$$f_t = \sigma\left(W_f[h_{t-1}, e_t] + b_f\right)$$

Where W_f and b_f is the weight matrix and bias of the forget gate.

The memory cell is updated as follows, $\widetilde{c}_t$ is the candidate memory cell, $tanh$ represents the hyperbolic tangent function whose output is [–1, 1], and $\odot$ represents the element corresponding multiplication.

$$\widetilde{c}_t = tanh(W_{\widetilde{c}}[h_{t-1}, e_t] + b_{\widetilde{c}})$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t$$

The output gate controls what to output based on the storage unit. The calculation of the output gate is as follows:

$$o_t = \sigma(W_o[h_{t-1}, e_t] + b_o)$$

Finally, the hidden state h_t is updated according to following formula and used for time step $t+1$

$$h_t = o_t \odot tanh(c_t)$$

In $W_i, W_f, W_{\widetilde{c}}, W_o$ are the weight matrices, $b_i, b_f, b_{\widetilde{c}}, b_o$ are the corresponding biases. In this layer, BiLSTM is applied to the word vector to capture the word level global structure information. Assuming the forward output is $\rightarrow_{h_i}$ and the reverse output is $\leftarrow_{h_i}$, the vector of the i^{th} word is: $h_i = [\rightarrow_{h_i} \oplus \leftarrow_{h_i}]$. Among them, the $\oplus$ represents the addition of the corresponding element.

3.4. Attention layer

In text with a high degree of freedom, such as Tweets, many words are considered inconsequential, whereas the emotions conveyed by the text typically centre around certain words. Moreover, BiLSTM models are unable to highlight the words that are crucial for emotion analysis in tweet texts, and their hidden layers would also lose some contextual information. However, the attention mechanism can effectively address this problem [62]. To calculate the attention allocation values, given the output of the BiLSTM layer $H = \{h_1, h_2, \ldots, h_n\}$, we calculate the importance score sc_i for the i^{th} output vector of BiLSTM layer h_i in the sequence as follows:

$$sc_i = tanh(W_a h_i + b_a)$$

Next, we obtain the normalised importance weight a_i for each word by applying the softmax function as follows:

$$a_i = \frac{expexp\ (sc_i)}{\sum_{j=1}^{n} expexp\ (sc_j)}$$

Where W_a is the weight matric learned during training process, b_a is the corresponding bias.

Finally, the hidden layer output value and the attention distribution value are dot-multiplied and accumulated to obtain the feature vector matrix T of the attention layer:

$$T = \sum_{0}^{n} a_i h_i$$

Where $A = a_1, a_2, \ldots, a_n$ is the final attention value, n is the data length, Q, K, V stands for key, and value respectively.

3.5. Convolution layer

After obtaining the global contextual features using the BiLSTM with attention mechanism, we pass the resulting global feature matrix through a convolutional layer for local feature extraction. By adding the convolution layer, we address the limitation of the BiLSTM model with attention mechanism in capturing local text features, while also avoiding the issue of contextual information neglect when using a convolutional neural network (CNN) directly. The specific computational process is described as follows:

$$J_i = f\left(\omega \times M_{i:i+g-1} + b\right)$$

The input to the convolutional layer is the text feature vector obtained from the BiLSTM and attention layers, where d represents the dimensionality of the word embeddings. ω is a convolution kernel of size $g * d$, and J_i is the feature map generated by the i^{th} convolution kernel. $M_{i:i+g-1}$ represents the local features extracted from rows i to $i + g - 1$ of the matrix M. f is a nonlinear transformation function, also known as an activation function, and b represents the bias term. The pooling function is used to downsample the feature maps generated by the convolution layer, effectively reducing the spatial dimension of the data, and controlling overfitting to some extent. In this paper we use the MaxPooling concatenated with AveragePooling to obtain the optimal solution of the local value.

3.6. Correlation processing

The correlation processing is based on the principle of polarity priority and the probability shown in the correlation matrix in Fig. 2 (i.e., negative emotion labels do not co-occur with positive polarity, and positive emotion labels do not co-occur with negative polarity). The final output of the model is one or more emotion labels of the input text.

3.7. Emotion graph construction

After utilising the aforementioned model to identify and analyse emotions contained in the text, we constructed an Emotion Graph (EG) based on the fundamental principles of Knowledge Graph (KG) for the visualisation of emotion analysis results. This approach is pioneering in its use of graphical and inferable structures to store and visualise emotion information extracted from text. The KG is a semantic network that employs nodes to represent entities and edges to represent relationships between entity pairs. The most basic unit of a knowledge graph is a triple construct consisting of Subject, Predicate, and Object (SPO). As an instance, the information presented in the statement "Spock is a character in Star Trek", which is constituted by the pairing of the two nodes, Spock and Star Trek, and the relationship "characterIn", resulting in the creation of the SPO triple (Spock, characterIn, Star Trek).

Based on the knowledge graph setting, we derive emotion information into (Emotion subject, relation, emotion object) according to the principle of SPO. We use Twitter tags as our starting point, as sentiment / emotion analysis often involves categorising emotion subjects based on different thematic events. These subjects can be used to search for posts related to a specific topic, to create trends or internet groups. We refer to emotion subjects as the tags contained in the Twitter text. These tags typically consist of a hash mark "#" followed by a specific word or phrase. As for the emotion object, we define it as the emotions conveyed by the tweet, the users involved, and the sentiment polarity of the tweet. Emotions are identified using the model described in sections 3.1-3.6, users are extracted based on features of the tweet using a rule-based method that looks for words starting with "@", and sentiment polarity is determined based on the rules outlined in the correlation matrix in section 3.6. For relation, we have defined three types of relationships: "Emotion" between the tag and its emotion, "Related_User" between the tag and its author, and "Sentiment_Polarity" between the tag and its sentiment polarity. We use Neo4j graph database [63] to store and visualise the nodes and relationships. Neo4j is a high-performance, NoSQL graph database. Compared with the traditional table containing rows and columns, the node space is a network composed of many nodes, relationships, and attributes (i.e., key-value pairs).

For instance, In the sentence, *"@Uber very disappointing that support has not responded to my email!! #bad #uber"*, we construct the emotion graph by following three steps:

1 Extract the emotion subjects, which refer to the tags #**bad** and #**uber**.
2 Conduct emotion analysis on the sentence using the multi-label emotion analysis model described in sections 3.2-3.6, resulting in **anger** and **sadness** as the final emotion analysis results, then define the relation with emotion subject extracted in step 1 as 'Emotion'.
3 Extract the emotion object, which refers to the user **@Uber**, then define the relation with emotion subject extracted in step 1 as 'Related_User'.

By constructing all potential combinations as triples (Emotion subject, relation, emotion object) of the three elements, we obtain the emotion graph as shown in Fig. 5.

4. Result

The result of the experiment is shown in Table 4. We use precision, recall, F1-score, and accuracy to evaluate the separate result and overall result of our model. As the most commonly used evaluation index for machine learning classification models, precision and recall can fully reflect the correct proportion of model prediction results and the correct proportion in all data. The precision rate is oriented to the prediction result. The precision rate represents the proportion of samples predicted to be in certain category that actually belong to that category. While the recall rate is for the original sample. The recall rate represents the proportion of items in a certain category in the original sample that were correctly predicted. And F1-score is a comprehensive index calculated based on precision rate and recall rate, the higher the F1 score, the better the model is. In addition, the paper also uses precise accuracy, which is different from the global calculation method of overall precision, it calculates the proportion of data that each label all predicts correctly.

The calculation criteria of these three indicators are shown below:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2 \ \times precision \times recall}{precision + recall}$$

True positive (TP) refers to samples that actually belong to the category they are predicted to be in. False positive (FP) refers to samples that do not belong to the category they are predicted to be in. False negative (FN) refers to samples that should belong to one category but they are predicted to be in another category.

We conducted a baseline and three comparative experiments based on the SemEval-2018 dataset, all of which applied the AC-BiLSTM algorithm described in sections 3.2-3.5 as the backbone model. The "Original Model (baseline)" employed only the SemEval-2018 training set as the training data and generated results through the backbone model. The "Model+ Emotion Dictionary" incorporated the NRC emotion lexicon into the training dataset (with the training data described in section 3.1), and generated results through the backbone model. The "Model+ Correlation" refers to the results obtained by using only the SemEval-2018 training set as the training data and through the backbone model, followed by the correlation constraint described in section 3.6 as the optimisation mechanism. The "Model+ Emotion Dictionary+ Correlation" incorporated the NRC emotion lexicon into the training dataset (with the

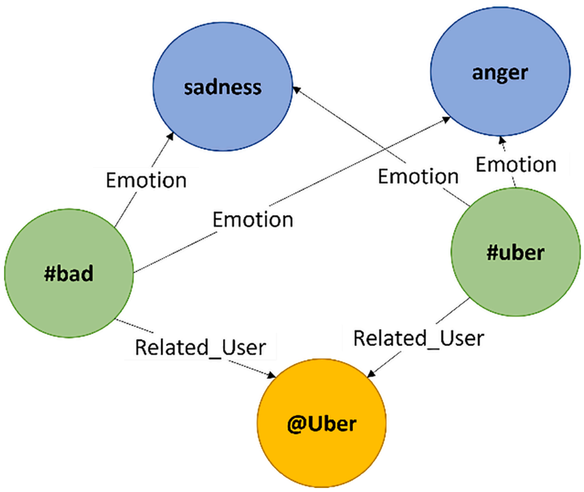

Fig. 5. The final output of the emotion graph for the example "*"@Uber very disappointing that support has not responded to my email!! #bad #uber*".

Table 4
Results of the experiment.

Model	Evaluation	Emotion Type							
		Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Original Model (baseline)	Precision	0.7945	0.6116	0.7335	0.8312	0.8135	0.7221	0.8398	0.7259
	Recall	0.7957	0.5318	0.7465	0.8678	0.7877	0.7161	0.5702	0.5110
	F1-score	0.7951	0.5308	0.7369	0.8480	0.7914	0.7189	0.6097	0.5096
	*Precision	0.7958							
	*Recall	0.7703							
	*F1-score	0.7817							
	*Accuracy	0.3181							
Model+ Emotion Dictionary	Precision	0.8000	0.5985	0.7321	0.8029	0.8127	0.7288	0.7195	0.5604
	Recall	0.7883	0.5992	0.7279	0.8563	0.7961	0.7042	0.6089	0.5345
	F1-score	0.7933	0.5989	0.7298	0.8260	0.7994	0.7135	0.6432	0.5422
	*Precision	0.7826							
	*Recall	0.7623							
	*F1-score	0.7715							
	*Accuracy	0.3101							
Model+ Correlation	Precision	0.8303	0.6240	0.7827	0.8334	0.8173	0.7425	0.8398	0.7259
	Recall	0.8099	0.5331	0.7835	0.8191	0.7866	0.7035	0.5702	0.5110
	F1-score	0.8180	0.5325	0.7831	0.8261	0.7904	0.7165	0.6097	0.5096
	*Precision	0.8003							
	*Recall	0.7565							
	*F1-score	0.7744							
	*Accuracy	**0.3455**							
Model+ Emotion Dictionary+ Correlation	Precision	0.8244	0.6028	0.7673	0.8075	0.8211	0.7449	0.7195	0.5640
	Recall	0.7943	0.5963	0.7477	0.8096	0.7993	0.6888	0.6089	0.5351
	F1-score	0.8051	0.5993	0.7548	0.8086	0.8031	0.7043	0.6432	0.5434
	*Precision	**0.8228**							
	*Recall	0.7663							
	*F1-score	**0.7885**							
	*Accuracy	0.3204							

*Precision = Overall Precision; *Recall = Overall Recall; *F1-score = Overall F1-score.

training data described in section 3.1), followed by the application of the correlation constraint described in section 3.6, and through the backbone model, followed by the correlation constraint described in section 3.6 as the optimisation mechanism. As shown in Table 4, the overall precision and precise accuracy of our proposed model was improved by 2.7% and 0.23% respectively compared with the original model with the addition of an emotion dictionary and correlation constraint, and the F1-score of anger, anticipation, disgust, joy, surprise and trust was also improved. Although the use of emotion dictionary alone makes the precise accuracy slightly lower than the original model, the F1-score of the classes (such as anticipate and trust) whose precision is significantly lower than that of other categories has been improved to a certain extent, which makes the individual F1-score of each category more average. To sum up, adding an emotion dictionary and correlation constraint to the original model improves the results of multi-label emotion analysis. Among them, the emotion dictionary is beneficial to optimise the recognition accuracy rate, which is significantly worse than other specific categories, so that the recognition results between different categories are more balanced, and the relationship constraint is beneficial to the improvement of the F1-score of each class and precise accuracy.

At the end of the multi-label emotion analysis, we store and visualise the recognition results of the SemEval-2018 test set in the form of a graph. All the results are shown in Fig. 6, in which the green node represents the tag, the blue node represents the emotion, the yellow node represents the user involved, and the red node represents the sentiment polarity. The arrow represents the relationship between nodes, and the direction of the arrow is the direction of the relationship.

Considering the density of Fig. 6, in order to improve the readability of the emotion graph in Fig. 6, we randomly selected 8 samples from the SemEval-2018 test set and conducted a small demo to focus on discussion. Table 5 is a small sample of the emotion analysis result using SemEval-2018 test set. After identifying the emotion and sentiment polarity contained in each Tweet (sentiment polarity is get through the result of label "optimism" and "pessimism" in SemEval-2018 dataset), and using the regular method to extract the users and tags involved in the text, the emotion graph is shown in Fig. 7. Following the emotion graph construction steps outlined in section 3.7, we extracted the emotion subject, emotion, emotion object, and sentiment of each of the tweet samples. Table 5 represents the table structured storage format of the aforementioned information, which is also the conventional storage format for existing emotion analysis tools. Fig. 7 represents the emotion graph corresponding to the data in Table 5. Through a comparison of the two, we found that the graph structure is more effective than the table structure in reflecting the intrinsic connections between the data. For example, the user '@CNN' and the user '@Uber' seem to be irrelevant, but it can be found from the graph that they both involve tags with negative sentiment, which may provide new correlation parameters in the specific application of public opinion analysis.

Through experimentation, we discovered that expanding the recognition of emotional information in the text to eight categories is feasible. In addition, the inclusion of the NRC lexicon and correlation constraints further enhanced the accuracy of the emotion analysis model. Most importantly, for the multi-label and multi-classification task set forth in this paper, storing the final emotion analysis results in a graph database and visualising them in the form of a knowledge graph is more advantageous for downstream applications that require reasoning. Graph databases are designed to handle highly interconnected data and are optimised for

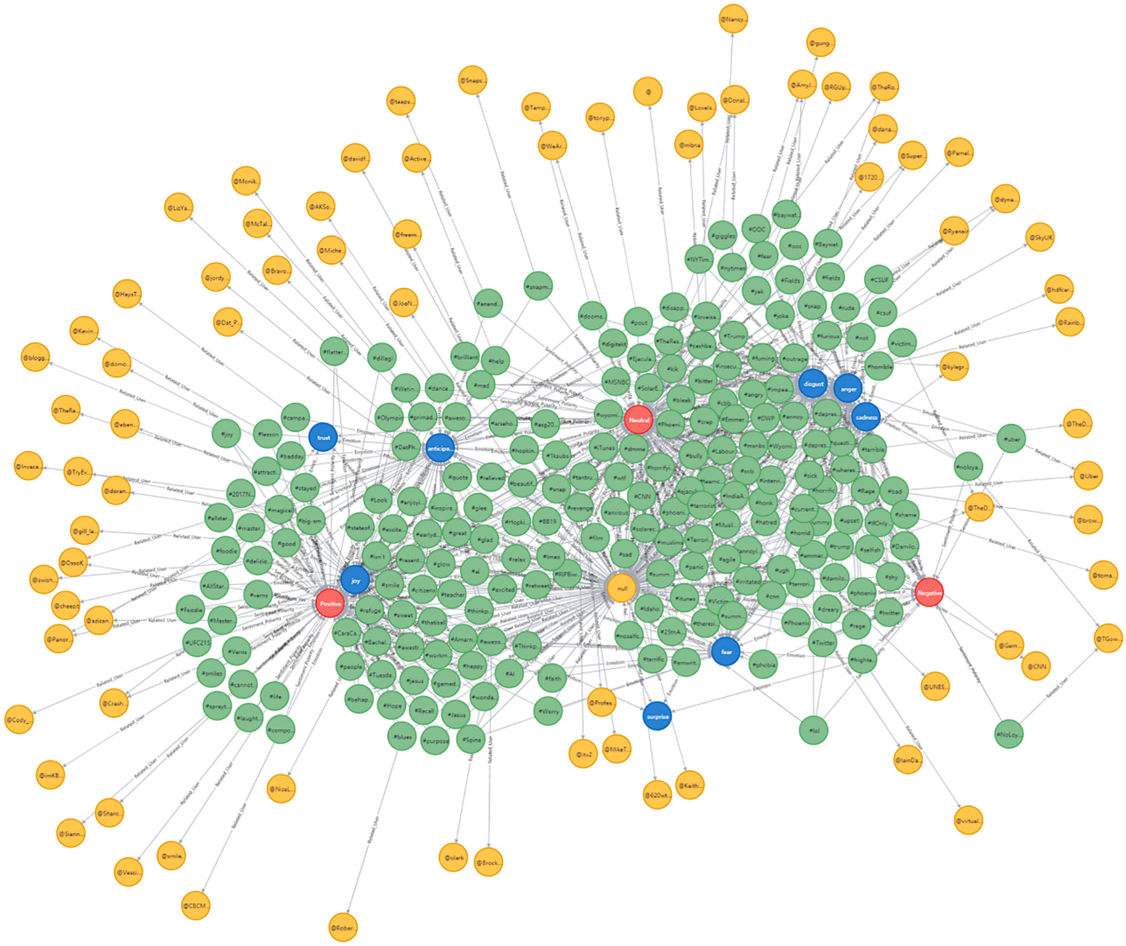

Fig. 6. The final output of the emotion graph.

Table 5
A small sample of the classification result using SemEval-2018 test set.

Tweet	Emotion subject	Emotion	Emotion object	Sentiment
@virtualalien there are more #frightening things in life	#frightening	anger fear surprise	@virtualalien	Negative
@danawhite you're a crook. #joke	#joke	anger	@danawhite	Neutral
Yes @itv2 putting that Angel back in the villa was the best move #revenge	#revenge	joy	@itv2 @MikeThalassitis	Positive
@Uber very disappointing that support has not responded to my email!! #bad #uber	#bad #uber	anger sadness	@Uber	Negative
Shame the cashback @mbna credit card comes to an end. I used to look forward to that end of year bonus. Sad really. #cashback	#cashback	sadness anticipation disgust	@mbna	Neutral
@LoveIsland #loveisand next time there's a dumping get Olivia out of there	#loveisand	anticipation disgust	@LoveIsland	Neutral
@CNN If #trump #whitehouse aren't held accountable for their actions,what precedent is being set for future presidencies. #nightmare	#trump #whitehouse	fear	@CNN	Negative
@TGowdySC No wonder why we're in this mess. #NoLoyalty	#NoLoyalty	anger disgust	@TGowdySC	Negative

managing complex relationships between data points. In contrast, traditional table-based databases, such as SQL databases, are designed for handling structured data in a tabular format. Compared to the traditional table-based databases, graph databases are highly flexible, agile, and scalable. Graph databases make it easy to uncover insights and patterns within highly connected data. This is particularly useful in applications such as social networks, recommendation engines, fraud detection, and network analysis. Therefore, this paper makes a strong contribution to the extension of traditional sentiment analysis, and has played a significant role in promoting the development of intelligence.

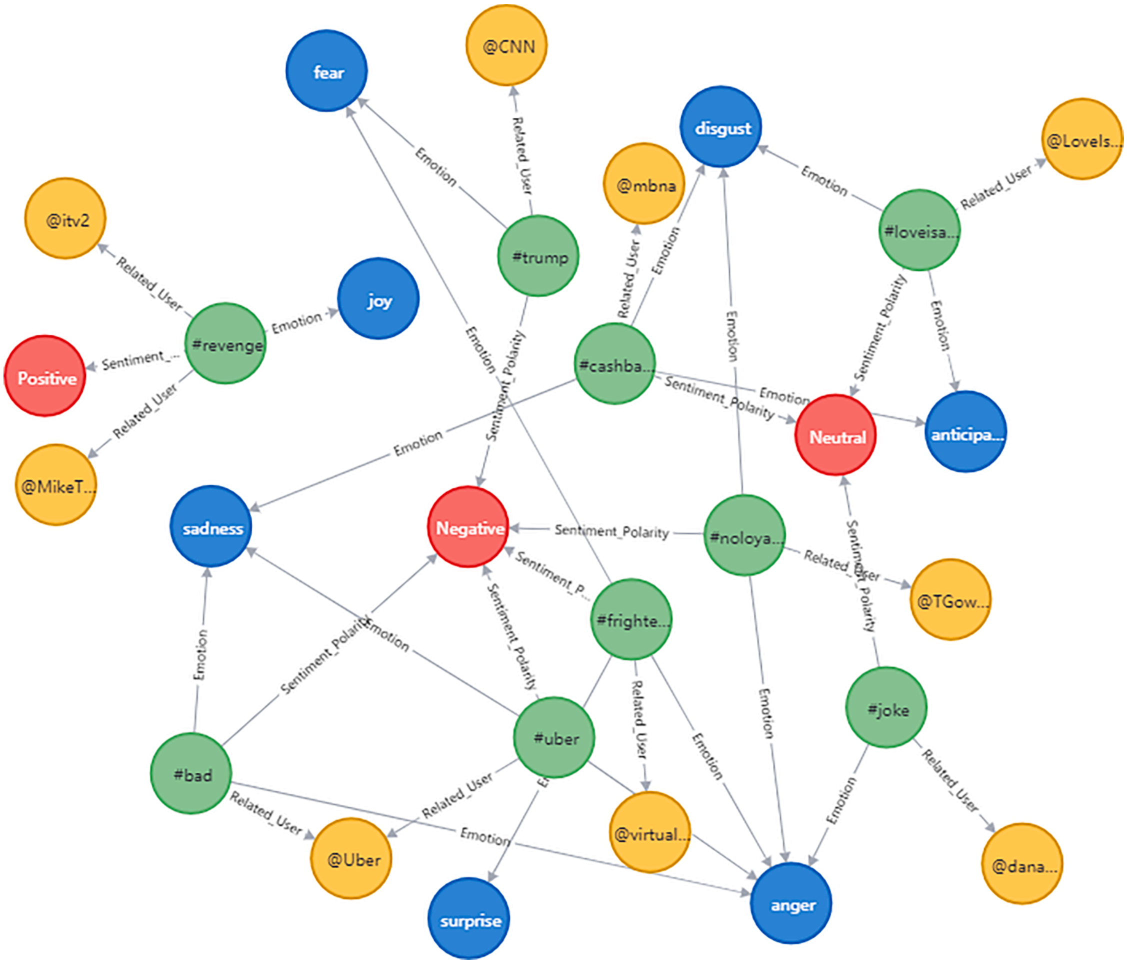

Fig. 7. Zooming in on Fig. 6.

5. Conclusion

Social media emotion analysis is a growing field in NLP and data science. With the explosive growth of social media platforms, there is a wealth of data available for analysis, providing insights into public opinion, sentiment, and user emotional states. This has led to increased interest in developing techniques to automatically identify and extract emotions from social media data, which has many potential applications in areas such as marketing, politics, and mental health.

The focus of this paper is how to precisely capture emotion information from social media and how to inferably store and visualise the reasoning for downstream applications. Based on the Plutchik's wheel of emotions, this paper expands social media emotional information classification from the traditional sentiment-level positive, neutral, and negative polarities to eight emotion-level specific emotion categories. Our research refines the capture of social media users' emotions to a certain extent, and enables richer and more detailed information to be obtained from user-generated text. At the same time, the paper proposes to extend the emotion analysis from single-label classification task to multi-label classification task, so as to realise the recognition of complex emotions and mixed emotions contained in social media texts. This solution is closer to reality, and provides more realistic and detailed results for follow-up emotion information acquisition. Finally, this paper proposes to store and visualise emotional information in a graph structure instead of traditional table-based structure, which makes the entire emotion analysis traceable, and also enables users and related organisations to more clearly understand the source and internal connections of the emotional attributes contained in the social media text. Therefore, this paper makes up for the shortcomings of traditional sentiment-level analysis such as being too general and unable to trace the source of emotions, and enhances the task of capturing and recognising emotion information in social media.

6. Limitation and future work

This paper has some limitations. The first limitation is lack of data. The SemEval-2018 dataset we use is the only multi-label emotion-level analysis dataset we could find. The model training results rely too much on the original annotations, which leads to a slight lack of robustness and generality. Moreover, the quality of the dataset directly determines the effect of the model, and it is less convincing in measuring the specific accuracy of the model.

The second limitation is the lack of comparison with state-of-the-art models technically. The focus of this research is on expanding the form of traditional sentiment analysis (from sentiment analysis to emotion analysis, from single-label classification task to multi-label classification task) and introducing the feasibility of using the NRC Emotion Lexicon and adding correlation constraints, as well as the feasibility and superiority of using graph structures to store and visualise emotion analysis results. We assumed that the relative comparison was made within the same algorithm. Even if this method has theoretically explicable advantages compared to state-of-the-art methods, we defined the emotion analysis as a specific multi-label classification task based on Plutchik's wheel of emotions, which has an advantage in detail compared to existing sentiment polarity analysis and single-label emotion classification. At the same time, our proposed emotion graph has advantages in deducibility and traceability compared to state-of-the-art methods that only stop at the emotion recognition results.

The third limitation represents an extension of the first and second limitations, as the focus of this paper is on the framework rather than the technical aspect. Therefore, upon the documentation of SemEval-2018 dataset that the dataset has undergone preliminary preprocessing, we did not separately add a preprocessing module in the algorithmic section. However, as the dataset expands in the future, along with its migration and widespread application, the lack of a separate data preprocessing module can lead to a decrease in overall model robustness and transferability.

In future work with regard to the first limitation - the lack of data, we plan to expand our search area and attempt to collaborate with relevant organisations or individuals to annotate and propose datasets for multi-label emotion analysis. In addition, we plan to use data augmentation techniques or employ pre-labelling technology to expand existing datasets, thereby addressing the limitation of data volume. To address the second limitation - the lack of comparison with state-of-the-art models technically, we plan to optimise the algorithm at the technical level in future work. We intend to customise a pre-trained model based on emotion information and increase comparison with state-of-the-art algorithms to further optimise the entire emotion analysis framework technically. To tackle the third limitation, we plan to add a dedicated preprocessing module for emotion information in future work, which will include specific preprocessing steps similar to emoji processing, built upon the foundation of traditional text preprocessing techniques.

Author Statement

During the preparation of this work the authors did not use any generative AI or AI-assisted tools for in the writing process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] K. Ananthajothi, K. Karthikayani, R. Prabha, Explicit and implicit oriented Aspect-Based Sentiment Analysis with optimal feature selection and deep learning for demonetization in India, Data Knowledge Eng. 142 (2022), 102092.
[2] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in information retrieval, 2(1–2), 1-135.
[3] A.S. Imran, S.M. Daudpota, Z Kastrati, R. Batra, Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets, IEEE Access 8 (2020) 181074–181090. IEEE.
[4] C. Strapparava, R. Mihalcea, Semeval-2007 task 14: affective text, in: Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), 2007, pp. 70–74.
[5] Robert. Plutchik, The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice, Am. Sci. 89 (4) (2001) 344–350. JSTOR.
[6] M. Jabreel, A. Moreno, A deep learning-based approach for multi-label emotion classification in tweets, Appl. Sci. 9 (6) (2019) 1123.
[7] A. Semeraro, S. Vilella, G. Ruffo, PyPlutchik: visualising and comparing emotion-annotated corpora, PLoS One 16 (9) (2021), e0256503.
[8] Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, Svetlana Kiritchenko, SemEval-2018 task 1: affect in tweets, in: Proceedings of the International Workshop on Semantic Evaluation (Semeval-2018), New Orleans, LA, USA, 2018.
[9] P.E. Ekman, R.J. Davidson, The nature of emotion: fundamental questions, Oxford University Press, 1994.
[10] C.E. Izard, The psychology of emotions, Springer Science & Business Media, 1991.
[11] J. Pujara, H. Miao, L. Getoor, W. Cohen, Knowledge graph identification, in: Proceedings of the International Semantic Web Conference, Berlin, Heidelberg, Springer, 2013, pp. 542–557.
[12] J. You, J. Leskovec, K. He, S. Xie, Graph structure of neural networks, in: Proceedings of the International Conference on Machine Learning, PMLR, 2020, pp. 10881–10891.
[13] Lori Young, Soroka Stuart, Affective news: The automated coding of sentiment in political texts, Political Communication 29 (2012) 205–231.
[14] P. Ray, A. Chakrabarti, A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis, Appl. Comput. Inform. 18 (2022) 163–178.
[15] Keliang Jia, Zhinuo Li, Chinese micro-blog sentiment classification based on emotion dictionary and semantic rules, in: Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (Cibda), IEEE, 2020, pp. 309–312.
[16] Zhi Li, Rui Li, Guanghao Jin, Sentiment analysis of Danmaku videos based on naïve Bayes and sentiment dictionary, IEEE Access 8 (2020) 75073–75084. IEEE.

[17] Yanghui Rao, Jingsheng Lei, Liu Wenyin, Qing Li, Mingliang Chen, Building emotional dictionary for sentiment analysis of online news, World Wide Web 17 (4) (2014) 723–742. Springer.
[18] Q. Wang, G. Zhu, S. Zhang, K.C. Li, X. Chen, H. Xu, Extending emotional lexicon for improving the classification accuracy of Chinese film reviews, Connection Sci. 33 (2) (2021) 153–172.
[19] J. Guo, Deep learning approach to text analysis for human emotion detection from big data, J. Intell. Syst. 31 (1) (2022) 113–126.
[20] P. Bhuvaneshwari, A.N. Rao, Y.H. Robinson, M.N. Thippeswamy, Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model, Multimedia Tools Appl. 81 (9) (2022) 12405–12419.
[21] L. Khan, A. Amjad, K.M. Afaq, H.T. Chang, Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media, Appl. Sci. 12 (5) (2022) 2694.
[22] Huang, Minlie, Qiao Qian, and Xiaoyan Zhu. 2017. "Encoding Syntactic Knowledge in Neural Networks for Sentiment Classification." ACM Transactions on Information Systems (TOIS) 35 (3). ACM New York, NY, USA: 1–27.
[23] Qian, Qiao, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2016. "Linguistically Regularized Lstms for Sentiment Classification." arXiv Preprint arXiv:1611.03949.
[24] A.P. Rodrigues, R. Fernandes, A. Shetty, K. Lakshmanna, R.M. Shafi, Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques, Comput. Intell. Neurosci. 2022 (2022).
[25] Xin Wang, Yuanchao Liu, Cheng-Jie Sun, Baoxun Wang, Xiaolong Wang, Predicting polarities of tweets by composing word embeddings with long short-term memory, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 1343–1353.
[26] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, Yueting Zhuang, Microblog sentiment classification via recurrent random walk network learning, IJCAI 17 (2017) 3532–3538.
[27] Md Shad Akhtar, Abhishek Kumar, Deepanway Ghosal, Asif Ekbal, Pushpak Bhattacharyya, A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 540–546.
[28] S. Mai, Y. Zeng, S. Zheng, H. Hu, Hybrid contrastive learning of tri-modal representation for multimodal sentiment analysis, IEEE Trans. Affective Comput. (2022).
[29] Xingyou Wang, Weijie Jiang, Zhiyong Luo, Combination of convolutional and recurrent neural network for sentiment analysis of short texts, in: Proceedings of the Coling 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 2428–2437.
[30] I. Ameer, G. Sidorov, H. Gomez-Adorno, R.M.A. Nawab, Multi-label emotion classification on code-mixed text: data and methods, IEEE Access 10 (2022) 8779–8789.
[31] X. Zheng, P. Li, Z. Chu, X. Hu, A survey on multi-label data stream classification, IEEE Access 8 (2019) 1249–1275.
[32] Matthew R Boutell, Jiebo Luo, Xipeng Shen, Christopher M Brown, Learning multi-label scene classification, Pattern Recognit. 37 (9) (2004) 1757–1771. Elsevier.
[33] G. Wu, R. Zheng, Y. Tian, D. Liu, Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification, Neural Netw. 122 (2020) 24–39.
[34] W. Gerych, T. Hartvigsen, L. Buquicchio, E. Agu, E.A. Rundensteiner, Recurrent bayesian classifier chains for exact multi-label classification, Adv. Neural Info. Process. Syst. 34 (2021) 15981–15992.
[35] Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank, Classifier chains for multi-label classification, Machine Learn. 85 (3) (2011) 333. Springer.
[36] Grigorios Tsoumakas, Ioannis Vlahavas, Random K-Labelsets: an ensemble method for multilabel classification, in: Proceedings of the European Conference on Machine Learning, Springer, 2007, pp. 406–417.
[37] Johannes Fürnkranz, Eyke Hüllermeier, Eneldo Loza Mencía, Klaus Brinker, Multilabel classification via calibrated label ranking, Machine Learn. 73 (2) (2008) 133–153. Springer.
[38] S. Moral-García, C.J. Mantas, J.G. Castellano, J. Abellán, Using Credal C4. 5 for Calibrated Label Ranking in Multi-Label Classification, Int. J. Approximate Reasoning 147 (2022) 60–77.
[39] M. Azad-Manjiri, A. Amiri, A. Saleh Sedghpour, ML-SLSTSVM: a new structural least square twin support vector machine for multi-label learning, Pattern Anal. Appl. 23 (2020) 295–308.
[40] André Elisseeff, Jason Weston, A kernel method for multi-labelled classification. Advances in Neural Information Processing Systems, 2002, pp. 681–687.
[41] G. Alberghini, S.B. Junior, A. Cano, Adaptive ensemble of self-adjusting nearest neighbor subspaces for multi-label drifting data streams, Neurocomputing 481 (2022) 228–248.
[42] Min-Ling Zhang, Zhi-Hua Zhou, ML-Knn: a lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048. Elsevier.
[43] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, Jamie Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: Proceedings of the 2008 Acm Sigmod International Conference on Management of Data, 2008, pp. 1247–1250.
[44] Fabian M Suchanek, Gjergji Kasneci, Gerhard Weikum, Yago: a core of semantic knowledge, in: Proceedings of the 16th International Conference on World Wide Web, 2007, pp. 697–706.
[45] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, Zachary Ives, Dbpedia: a nucleus for a web of open data. The Semantic Web, Springer, 2007, pp. 722–735.
[46] J.A. Sacenti, R. Fileto, R. Willrich, Knowledge graph summarization impacts on movie recommendations, J. Intell. Info. Syst. 58 (1) (2022) 43–66.
[47] Weiwei Wang, Zhigang Wang, Liangming Pan, Yang Liu, Jiangtao Zhang, Research on the Construction of Bilingual Movie Knowledge Graph, Acta Scientiarum Naturalium Universitatis Pekinensis 52 (1) (2016) 25–34.
[48] Peipei Ping, Karol Watson, Jiawei Han, Alex Bui, Individualized knowledge graph: a viable informatics path to precision medicine, Circ. Res. 120 (7) (2017) 1078–1080. Am Heart Assoc.
[49] Longxiang Shi, Shijian Li, Xiaoran Yang, Jiaheng Qi, Gang Pan, Binbin Zhou, Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services, Biomed. Res. Int. 2017 (2017). Hindawi.
[50] Kou, Z., Shang, L., Zhang, Y., & Wang, D. (2022). Hc-covid: A hierarchical crowdsource knowledge graph approach to explainable covid-19 misinformation detection. Proceedings of the ACM on Human-Computer Interaction, 6(GROUP), 1-25.
[51] Wang, Zhouxia, Tianshui Chen, Jimmy Ren, Weihao Yu, Hui Cheng, and Liang Lin. 2018. "Deep Reasoning with Knowledge Graph for Social Relationship Understanding." arXiv Preprint arXiv:1807.00504.
[52] P. Chandak, K. Huang, M. Zitnik, Building a knowledge graph to enable precision medicine, Scientific Data 10 (1) (2023) 67.
[53] Nazar Zaki, Chandana Tennakoon, Hany Al Ashwal, Knowledge graph construction and search for biological databases, in: Proceedings of the 2017 International Conference on Research and Innovation in Information Systems (Icriis), 2017, pp. 1–6. IEEE.
[54] Chengbin Wang, Xiaogang Ma, Jianguo Chen, Jingwen Chen, Information extraction and knowledge graph construction from geoscience literature, Comput. Geosci. 112 (2018) 112–120. Elsevier.
[55] Penghe Chen, Yu Lu, Vincent W Zheng, Xiyang Chen, Boda Yang, KnowEdu: a system to construct knowledge graph for education, Ieee Access 6 (2018) 31553–31563. IEEE.
[56] Mireille Fares, Angela Moufarrej, Eliane Jreij, Joe Tekli, William Grosky, Unsupervised word-level affect analysis and propagation in a lexical knowledge graph, Knowledge-Based Syst. 165 (2019) 432–459. Elsevier.
[57] Xu, Peng, Zihan Liu, Genta Indra Winata, Zhaojiang Lin, and Pascale Fung. 2020. "Emograph: Capturing Emotion Correlations Using Graph Networks." arXiv Preprint arXiv:2008.09378.
[58] Wenying Yu, Shuai Ding, Zijie Yue, Shanlin Yang, Emotion recognition from facial expressions and contactless heart rate using knowledge graph, in: Proceedings of the 2020 IEEE International Conference on Knowledge Graph (Ickg), IEEE, 2020, pp. 64–69.
[59] G. Liu, J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, Neurocomputing 337 (2019) 325–338.

[60] I. Benchaji, S. Douzi, B. El Ouahidi, J. Jaafari, Enhanced credit card fraud detection based on attention mechanism and LSTM deep model, J. Big Data 8 (2021) 1–21.
[61] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, Independently recurrent neural network (indrnn): Building a longer and deeper rnn, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5457–5466.
[62] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
[63] Jim. Webber, A programmatic introduction to Neo4j, in: Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity, 2012, pp. 217–218.

Yuming Li: Yuming Li is a PhD Candidate of Information Systems and Operations Management at the University of Auckland Business School. She holds a Bachelor's degree in Management and a Master's degree in Computer Science. Her research interests are in information systems, natural language processing and deep learning.

Johnny Chan: Johnny Chan is a Lecturer of Information Systems and Operations Management at the University of Auckland Business School. His primary interests are in the research and development of emerging technologies including artificial intelligence, blockchain, internet of things, and mixed reality. His vision is to democratise applications of emerging technologies as means to contribute towards the sustainability and wellbeing of individual, organisation, and society, through the lenses of decision, behavioural, design and data sciences.

Gabrielle Peko: Gabrielle Peko is a Senior Lecturer of Information Systems and Operations Management at the University of Auckland Business School. She has two principal, intertwined streams of research that complement and enhance each other including (1) the transformation of traditional entities / organisations / communities with rigid structures, processes, and systems into intelligent, agile, adaptive, evolvable, and sustainable entities / organisations / communities and (2) decision support, collaborative, and social systems, networks and technologies that support decision making and transformation.

David Sundaram: David Sundaram is a Professor of Information Systems and Operations Management at the University of Auckland Business School. His primary research interests are in the design and implementation of adaptive systems that are flexible and evolvable and support the informational, decisional, knowledge and collaboration needs of organisations. Related research interests that support the above include (1) adaptive organisations and supporting them through the interweaving of deliberate and emergent strategies (2) process, information, and decision modelling (2) sustainability modelling and reporting (3) ubiquitous information systems (4) enterprise systems and their implementation.