

# KGPChamps at SemEval-2019 Task 3: A deep learning approach to detect emotions in the dialog utterances.

<sup>1</sup>Jasabanta Patro, <sup>2</sup>Nitin Choudhary, <sup>3</sup>Kalpiti Chittora, and <sup>4</sup>Animesh Mukherjee  
IIT Kharagpur, India  
{<sup>1</sup>jasabantapatro, <sup>2</sup>nitch13jan, <sup>3</sup>chittora.kalpiti }@iitkgp.ac.in, <sup>4</sup>animeshm@cse.iitkgp.ac.in

## Abstract

This paper describes our approach to solve *SemEval task 3: EmoContext*; where, given a textual dialogue, i.e., a user utterance along with two turns of context, we have to classify the emotion associated with the utterance as one of the following emotion classes: *Happy, Sad, Angry* or *Others*. To solve this problem, we experiment with different deep learning models ranging from simple LSTM to relatively more complex attention with Bi-LSTM model. We also experiment with word embeddings such as ConceptNet along with word embeddings generated from bi-directional LSTM taking input characters. We fine tune different parameters and hyper-parameters associated with each of our model and report the micro precision, micro recall and micro F1-score for each model. We identify the Bi-LSTM model, along with the input word embedding taken as the concatenation of the embeddings generated from the bidirectional character LSTM and ConceptNet embedding, as the best performing model with a highest micro-F1 score over the test set as 0.7261.

## 1 Introduction

In recent years, with the increase in the popularity of social media platforms, a significant amount of unstructured social media content (posts, tweets, messages etc.) has become available to the research community. People use social media as a platform to share their opinions, emotions, thoughts etc. This information has a huge potential to serve as a commercial catalyst to the business of companies and organizations, e.g., knowing the opinion of people about a product or a service could help the company to do betterment of their product or service according to the desire of the online consumers. In similar lines, emotions from the peoples' comments/opinion can help us to model the future popularity of the product or the

service. Further, knowing public emotions about different events can help political parties to set their agenda for elections. Thus mining of opinions and emotions has a lot of practical relevance. Even prior to the social media era, emotion detection had achieved significant attention of psychologists and linguistics. An elaborate discussion of emotion as a research topic is presented in the next section.

In this paper, we describe our system and the models, with which, we achieved significant performance improvement over the SemEval baseline for task 3. The task is described in (Chatterjee et al., 2019), where, given a textual dialogue, i.e., a user utterance along with two turns of context, we have to classify the emotion associated with the utterance into one of the following emotion classes: *Happy, Sad, Angry* or *Others*. To solve this problem, we experiment with different deep learning models ranging from simple LSTMs to more complex attention based Bi-LSTM models. We also experiment with different word embeddings such as ConceptNet along with word embeddings generated from bi-directional character LSTMs. Our best model gives a micro F1 of 0.7261 on the test set released by the organizers.

## 2 Related works:

From the last decades of the previous century, emotion as a topic of research has captured the attention of many scientists and researches from different sub-fields of computer science and psychology. While prior to the current century, researches tried to capture emotions from acoustic signals (Murray and Arnott, 1993; Banse and Scherer, 1996) and facial expressions (Ekman and Friesen, 1971; Ekman, 1993; Ekman et al., 1987), in the current century, due to the emergence of

Internet and social media, expression and detection of emotion through/from texts and social media, has grabbed significant attention (Alm et al., 2005; Fragopanagos and Taylor, 2005; Binali et al., 2010; Dini and Bittar, 2016; Canales and Martínez-Barco, 2014; Seyeditabari et al., 2018) of researchers. The whole literature around emotion can be broadly divided into two categories (i) theoretical studies and (ii) computational studies.

**Theoretical studies:** The theoretical studies include searching answers for whether facial expressions of emotion are universal (Ekman and Friesen, 1971), searching for cross-cultural agreement in the judgment of facial expression (Ekman et al., 1987), studying the acoustic profile of vocal emotion expression (Banse and Scherer, 1996) etc. An exploratory discussion of the literature detailing human vocal emotion and its principal findings are presented in (Murray and Arnott, 1993). Application of the literature to the construction of a system capable of producing synthetic speech with emotion has also been discussed. A brief description of how emotion is processed in our brain is discussed in (LeDoux, 2000).

**Computational studies:** From last two decades detecting and analysis of emotion in texts and social media content has grabbed significant attention of computational linguists and social scientists. (Litman and Forbes-Riley, 2004) determine the utility of speech and lexical features for predicting student emotions in computer-human spoken tutoring dialogues. They develop an annotated corpora that mark each student dialogue for negative, neutral, positive and mixed emotions. Then they extract acoustic-prosodic features from the speech signal, and lexical items from the transcribed or recognized speech and apply machine learning approaches to detect the emotions. In the same year, (Busso et al., 2004) came up with an analysis of emotion recognition techniques, using facial expressions, speech and multimodal information etc. They conclude that the system based on facial expression gives better performance than the system based on just acoustic information for the emotions considered. Sentiment classification seeks to identify a piece of text according to its authors general feeling toward their subject, be it positive or negative. Traditional machine learning techniques have been

applied to this problem with reasonable success, but they have been shown to work well only when there is a good match between the training and test data with respect to the topic. (Read, 2005) use emoticons to reduce dependency in machine learning techniques for sentiment classification. (Wiebe et al., 2005) came up with a corpus having an annotation of opinions, emotions, sentiments, speculations, evaluations and other private states in the language of 10000 lines. In the second half of the last decade several studies came up that analyze and detect (Fragopanagos and Taylor, 2005; Binali et al., 2010; Hancock et al., 2007; Strapparava and Mihalcea, 2008) emotion from the text using machine learning techniques of the text context. Detection of emotion over social media content (Yassine and Hajj, 2010; Pak and Paroubek, 2010; Gupta et al., 2010) and electronic media content (Neviarouskaya et al., 2007; Yang et al., 2007) started to become popular during this period. Emotion cause detection (Chen et al., 2010) introduce another interesting problem in this period. In the current decade many problems in this domain have been introduced like emotion detection in code-switching texts (Wang et al., 2015), metaphor detection with topic transition, emotion and cognition in context (Jang et al., 2016), sentence and clause level emotion annotation and detection (Tafreshi and Diab, 2018), detecting emotion in social media contents (Roberts et al., 2012; Liew, 2014), detecting emotion in multilingual contexts (Das, 2011) etc. to name a few. Several corpora have been introduced having an annotation of emotions and other associated things such as emotion over multi-genre corpus (Tafreshi and Diab, 2018), emotion corpus of multi-party conversations (Hsu et al., 2018), a fine-grained emotion corpus for sentiment analysis (Liew et al., 2016), a dataset of emotion annotated tweets to understand the interaction between affect categories (Mohammad and Kiritchenko, 2018) etc. to name few. Simultaneously, methodological novelty in emotion detection is also an important contribution by researchers in the recent times; works like emotion detection by GRUs (Abdul-Mageed and Ungar, 2017), representation mapping (Buechel and Hahn, 2018), hybrid neural networks (Li et al., 2016) etc. are a few such latest techniques. A detail description of different hidden challenges

present in emotion detection over social media content is present in (Dini and Bittar, 2016). Few survey papers (Canales and Martínez-Barco, 2014; Seyeditabari et al., 2018) describing different emotion analysis and detection methods adopted in past years also came up during this period.

### 3 Dataset and preprocessing

#### 3.1 Dataset

The dataset consists of three parts, (i) training data, (ii) development data (dev set), and (iii) test data. The training dataset consists of 30k conversations, where each conversation contains three turns of user utterances. The dev set and the test set contains 2754 and 5508 conversations respectively. These have been collected and annotated by the organisers. All of the conversations are classified into four classes, 'angry', 'sad', 'happy' and 'others'. Training data consists of about 5k samples each from 'angry', 'sad', 'happy' class, and 15k samples from 'others' class, whereas, both dev and test sets have a real-life distribution, which is about 4% each of 'angry', 'sad', 'happy' class and the rest is 'others' class.

#### 3.2 Preprocessing

Before feeding the conversations to our model, we perform the following operations on the text:

- The three turns of the conversation are joined to form a single sentence; also if there are multiple instances of punctuation, then we keep only a single instance. The joined utterance contains the conversations in the same order as that is given in the data set.
- Each emoji is replaced by its respective English translation. Example: ':-)' is replaced by 'happy'.
- All the possible English contractions are replaced by their expanded forms. for example: 'don't' is converted 'do not'.
- We use Ekphrasis toolkit (Baziotis et al., 2017) to normalize the occurrence of the URL, e-mail, percent, money, phone, user, time, date, and number in the comments. For example, URLs are replaced by 'url', and all occurrences of @someone are replaced by 'user'.

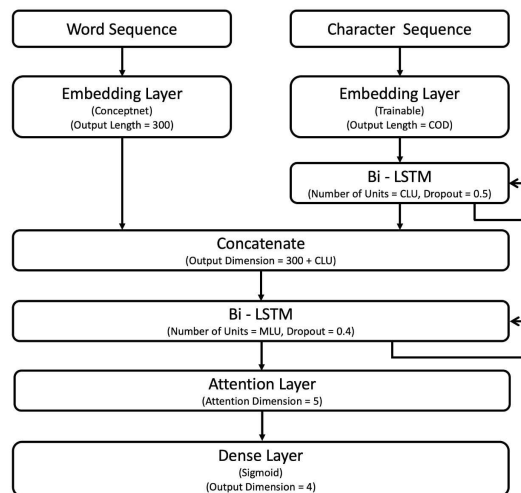


Figure 1: Overall schematic architecture of our system.

- Finally, we use NLTK Wordnet lemmatizer (Loper and Bird, 2002) to lemmatize the words to their roots.

### 4 System description

Our overall system is illustrated in Figure 1. We run different variants of our system by changing associated parameters, hyper-parameters and layers. For input, we consider a variety of options, which include (i) creating word embeddings using a Bi-LSTM trained on the character sequence of the sentence/utterance, (ii) using a pre-trained word embedding, i.e., Conceptnet, and (iii) concatenating (i) and (ii). From the architecture point of view our systems can be categorized into three types – (i) simple LSTM model, where an LSTM layer is considered instead of a Bi-LSTM layer (see figure 1) with no attention, i.e., the final hidden vector of LSTM layer is fed to the dense layer bypassing the attention layer (ii) simple Bi-LSTM model, where no attention layer is present, i.e., the final hidden vector of Bi-LSTM layer is fed to the dense layer by-passing the attention layer in figure 1, and (iii) Bi-LSTM model + attention, where we keep the attention layer active as shown in figure 1. We use the python module keras for our implementation.

### 5 Models and results

As previously stated, we experiment with different variants of the model. In this section, we discuss some of the top performing models and their performance. The results for different sys-

Model Type	$Acc_\mu$	$Pre_\mu$	$Rec_\mu$	$F1_\mu$	$F1_{test}$
LSTM + Conceptnet	0.90	0.88	0.90	0.89	0.6825
Bi-LSTM + ConceptNet	0.90	0.86	0.91	0.88	0.6686
BiLSTM + (char embed. + Conceptnet)	0.90	0.88	0.89	0.89	<b>0.7261</b>
Bi-LSTM + (character embed. + Conceptnet) + no emojis	0.90	0.88	0.90	0.89	0.6418
Attentive Bi-LSTM + character embedding + Conceptnet	0.89	0.88	0.88	0.88	0.6900

Table 1: Results of different models; accuracy ( $Acc_\mu$ ), micro-precision ( $Pre_\mu$ ), micro-recall ( $Rec_\mu$ ) and micro-F1 score ( $F1_\mu$ ) over the training data for five-fold cross validation;  $F1_{test}$  is the micro-F1 score over the test set released by the organisers.

tems and their description are as shown in Table 1. We report two types of results (i) performance over training set which we obtain through five-fold cross-validation and (ii) performance over test data as reported by SemEval organizers. A short description of the model variants and their results are given below.

1. The first two models present in Table 1, as the name suggests, contains a layer of LSTM (1<sup>st</sup> model) and Bi-LSTM (2<sup>nd</sup> model). Sequence of words padded to a fixed length is given as input to this layer. The input sequence is then converted to an embedding vector with the help of pre-trained embedding matrices. We tried various pre-trained embedding matrices such as GloVe, fastText, ConceptNet and Google word2vec, out of which for Conceptnet we get best results. The outcome of LSTM/Bi-LSTM is given as input to the final dense layer which contains four nodes with sigmoid activation function for four emotions.
2. In the next model, we append character embeddings to the Conceptnet embeddings. This model produces the best performance over the test set, i.e., micro F1 score over the test set is 0.7261 as released by the organizers. The input to this model is a 2-D vector of words with characters in the second dimension.
3. The fourth model is the same as the previous model but emotion words (words which replaced emojis) are removed. As we can infer from the table this choice though did not affect the performance over the training set, the test set performance is significantly affected.
4. Finally, in the attentive Bi-LSTM model, we switch on the attention layer. Other param-

eters are kept the same as the third model model.

## 6 Conclusion

In this paper, we present a neural network based model to detect emotions from textual conversations. The usage of pre-trained embedding, Conceptnet gives a huge boost to the performance of our system. The performance reported in our paper could further be improved by implementing a better preprocessing pipeline and using more advanced RNN models. Furthermore, the dataset provided had a huge imbalance among different classes, therefore sampling among classes could result in increased performance. On the other hand, studying emotion in social media text can be linked further to the popularity of a product, service etc. which might be linked to financial interests of organizations. Further, how users expressing a particular predominant emotion interact with other users could be another line of future study.

## References

- Muhammad Abdul-Mageed and Lyle H. Ungar. 2017. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In *ACL*.
- Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *HLT/EMNLP*.
- Rainer Banse and Klaus R. Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 70 3:614–36.
- Christos Baziotis, Nikos Pelekis, and Christos Douk-eridis. 2017. Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 747–754.

- Haji Binali, Chen W Wu, and Vidyasagar Potdar. 2010. Computational approaches for emotion detection in text. *4th IEEE International Conference on Digital Ecosystems and Technologies*, pages 172–177.
- Sven Buechel and Udo Hahn. 2018. Representation mapping: A novel approach to generate high-quality multi-lingual emotion lexicons. *CoRR*, abs/1807.00775.
- Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *ICMI*.
- Lea Canales and Patricio Martínez-Barco. 2014. Emotion detection from text : A survey.
- Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. Semeval-2019 task 3: Emocontext: Contextual emotion detection in text. In *Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval-2019)*, Minneapolis, Minnesota.
- Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *COLING 2010*.
- Dipankar Das. 2011. Analysis and tracking of emotions in english and bengali texts: a computational approach. In *WWW*.
- Luca Dini and André Bittar. 2016. Emotion analysis on twitter: The hidden challenge. In *LREC*.
- Paul Ekman. 1993. Facial expression and emotion. *The American psychologist*, 48 4:384–92.
- Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17 2:124–9.
- Paul Ekman, Wallace V. Friesen, Maree O’Sullivan, Aryola Chan, I Diacoyanni-Tarlatzis, K G Heider, Rainer Krause, W A LeCompte, Tom K Pitcairn, and P. E. Ricci-Bitti. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53 4:712–7.
- Nickolaos F. Fragopanagos and John G. Taylor. 2005. Emotion recognition in human-computer interaction. *Neural networks : the official journal of the International Neural Network Society*, 18 4:389–405.
- Narendra K. Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2010. Emotion detection in email customer care. *Computational Intelligence*, 29:489–505.
- Jeffrey T. Hancock, Christopher Landrigan, and Courtney Silver. 2007. Expressing emotion in text-based communication. In *CHI*.
- Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao K. Huang, and Lun-Wei Ku. 2018. Emotionlines: An emotion corpus of multi-party conversations. *CoRR*, abs/1802.08379.
- Hyeju Jang, Yohan Jo, Qinlan Shen, Michael Z. Miller, Seungwhan Moon, and Carolyn Penstein Rosé. 2016. Metaphor detection with topic transition, emotion and cognition in context. In *ACL*.
- Joseph E. LeDoux. 2000. Emotion circuits in the brain. *Annual review of neuroscience*, 23:155–84.
- Xiangsheng Li, Jianhui Pang, Biyun Mo, and Yanghui Rao. 2016. Hybrid neural networks for social emotion detection over short text. *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 537–544.
- Jasy Suet Yan Liew. 2014. Expanding the range of automatic emotion detection in microblogging text. In *EACL*.
- Jasy Suet Yan Liew, Howard R. Turtle, and Elizabeth D. Liddy. 2016. Emotweet-28: A fine-grained emotion corpus for sentiment analysis. In *LREC*.
- Diane J. Litman and Katherine Forbes-Riley. 2004. Predicting student emotions in computer-human tutoring dialogues. In *ACL*.
- Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit in proceedings of the acl workshop on effective tools and methodologies for teaching natural language processing and computational linguistics. *Philadelphia, Association for Computational Linguistics*, pages 62–69.
- Saif Mohammad and Svetlana Kiritchenko. 2018. Understanding emotions: A dataset of tweets to study interactions between affect categories. In *LREC*.
- Iain R. Murray and John L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93 2:1097–108.
- Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2007. Narrowing the social gap among people involved in global dialog: Automatic emotion detection in blog posts. In *ICWSM*.
- Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*.
- Jonathon Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *ACL*.
- Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. 2012. Emopatweet: Annotating and detecting emotions on twitter. In *LREC*.

- Armin Seyeditabari, Narges Tabari, and Wlodek Zadrozny. 2018. Emotion detection in text: a review. *CoRR*, abs/1806.00674.
- Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *SAC*.
- Shabnam Tafreshi and Mona T. Diab. 2018. Sentence and clause level emotion annotation, detection, and classification in a multi-genre corpus. In *LREC*.
- Zhongqing Wang, Sophia Yat Mei Lee, Shoushan Li, and Guodong Zhou. 2015. Emotion detection in code-switching texts via bilingual and sentimental information. In *ACL*.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39:165–210.
- Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 2007. Emotion classification using web blog corpora. *IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*, pages 275–278.
- Mohamed Yassine and Hazem M. Hajj. 2010. A framework for emotion mining from text in online social networks. *2010 IEEE International Conference on Data Mining Workshops*, pages 1136–1142.