

Sentence Level Sentiment Analysis using Deep Learning Method



K.K. Uma, K. Meenakshisundaram

ABSTRACT: An Emoji is a small image representing facial expression, entity or a concept that can be either static or animated. In this paper, Emojis are used to study both cross-language and language based sentiment patterns. All the languages do not come with fair amount of labels. Emojis are useful signs of sentiment analysis in cross-lingual tweets. In this paper, an approach is proposed to extend the existing binary sentiment classification using multi-way classification. A novel Long Short-Term Memory (LSTM) - Convolutional Neural Networks (CNN) based model is proposed to obtain sentiments from emojis. The sentiments are classified using Deep Learning method like CNN. The proposed system outperforms the existing system in terms of Accuracy, Precision, Recall, F-Measure and Time Period. Finally the researcher manifests the fact that the CNN and LSTM combination as model shows an immense improvement to detecting the sentiment targets.

KEYWORDS: Emoji, Sentiment Analysis, Long Short-Term Memory (LSTM), Cross-language

I. INTRODUCTION

Classifying sentiments mainly depends on labeled data. It is seen that English has more labeled texts than other languages. This leads to an imbalance in the quality of associated information services for people of different languages.

Machine translation tools are used to connect the source and target languages. Cross-language sentiment patterns are thus learned in English and transferred to the target language. This method does not yield the sentiment knowledge of a target language.

Micro-blogging websites like Twitter are an unlimited source of different kinds of information. People put their opinions on diverse topics, argue about current issues, criticize and express their emotions. The spectators of these platforms and social networks increase day by day. Data obtained from these sources are used in opinion mining, sentiment and emotion analysis. Opinions, sentiments and emotions are involved in sentiment analysis and opinion mining. Data in social media are always noisy with many mistakes in spelling, grammar and punctuation with inconsistent data errors. Sentiment analysis plays an important role in data mining, and deals with identifying and analyzing sentiments available in social media.

Social networking websites generate huge volume of unstructured data. Twitter is mainly used, wherein tweets portray the users' viewpoints and sentiments. It is a rich source of data that supports opinion mining, sentiment and emotion analysis. At present, sentiment analysis methodologies deal with the polarity of positive and negative emotions. Polarity analysis yields only inadequate information. Defining positive/negative emotions is simple, but identifying the whole set is tedious.

To deal with this issue, a cross lingual sentiment classification scheme that shares information gained from a language with more labels to another language with fewer labels is designed.

II. RELATED WORK

Bi et al (2007) have used a amalgamation of diverse Machine Learning (ML) algorithms like Support Vector Machine (SVM), kNN (Nearest Neighbour), kNN model-based approach (kNNM) and Rocchio for text categorization using Dempster combination rule. The classifiers are trained separately using two datasets: 20-newsgroup and Reuters-21578 benchmark.

Xia et al (2011) have applied many linguistic features of text documents, consisting of part of speech and Term Frequency - Inverse Document Frequency (TF-IDF), along with ensemble methods for sentiment classification. They have also applied a combination of classifiers namely, SVM, Naïve Bayes (NB) and Maximum Entropy (ME) by using meta-classifier and weighted combination rules.

Da Silva et al (2014) have applied feature hashing and bag-of-words on Twitter dataset for sentiment analysis. They have ensemble SVM, NB, RF and Logistic Regression (LR) classifiers to improve performance.

Hussien et al (2016) have used a combination of classifiers such as SVM, Multinomial Naive Bayes (MNB) and Random Forest (RF) to get better classification accuracy. They aim of the experiment on issues pre-processing of single-emoji tweets proving the best part in the analysed dataset.

Wolny (2016) has dealt with analysing emotion tokens, emoticons and emoji ideograms. Emotion tokens are normally used in tweets to show emotions irrespective of the language. They support sentiment analysis in multilingual tweets. Binary sentiment classification approaches are extended by using a multi-way emotion classification.

Ljubešić & Fišer (2016) have investigated the spatial distribution of emojis on 17 million tweets. Cluster analysis is done over countries and correlation analysis of emoji distributions and World Development Indicators (WDI). The results were amazing paving the way to young researchers to get into it.

Manuscript published on 30 September 2019

* Correspondence Author

K.K. Uma*, Research Scholar, Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu, India.

Dr. K. Meenakshisundaram, Associate professor, Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Many focus on emoticons but less on emojis. [Guibon et al \(2016\)](#) have dealt with use of emoji in sentiment analysis. Emoji usage typology with linguistic and socio-linguistic studies on Emoji interpretation is done. The evolution of Emoji usages is considered.

[Lu et al \(2016\)](#) have analysed on the use of Emojis by smartphone users based on a data set composed from a popular emoji keyboard. They have shown that the categories and frequencies of Emojis aid in identifying and understanding cultural differences.

[Miller et al \(2016\)](#) have explored emoji versions across platforms leading to varied understandings of emoji. People's interpretations of the frequently used emoji characters on multiple platforms are analysed. Emoji that is likely to be misinterpreted is analysed in terms of sentiment and semantics. Miscommunications for both individual and different emoji renderings across platforms are found.

[Eisner et al \(2016\)](#) have used pre-trained models based on embedding for Unicode emojis which are known from the description in the Unicode emoji standard. These embedding are usable in downstream social Natural Language Processing (NLP) applications which is visibly seen.

For downstream sentiment analysis, the embedding from short descriptions yields better results in contrast to skip-gram model while circumventing the necessity for frequent patterns for estimating a representation.

III. PROPOSED SYSTEM

Messages on the online web contain text and Emoji. Both are separated and the scores for the same are computed. Long Short-Term Memory (LSTM) is used for sentence level prediction in tweets to discriminate emoji and text, classify them into positive, negative or neutral. Long Short-Term Memory (LSTM) ([Hochreiter & Schmidhuber 1997](#)) is a distinct type of Recurrent Neural Network (RNN). It is applied in this paper because it is capable of modelling the sequential property of text data.

It takes the input in the current layer and knowledge from the preceding steps to determine the hidden layer states. To deal with the gradient vanishing problem seen in RNNs, a gating mechanism is included to find the situations when and how the states can be restructured.

The LSTM units come with a memory cell and three gates namely, input, forget and output gates. The input gate takes control of the input activations into the memory cell, while the output gate deals with cell activation flows into the remaining portions of the network. The memory cells have self-loops, wherein the forget gates control the weights.

3.1 Long Short-Term Memory (LSTM)

A learning method that includes language based knowledge in cross-lingual sentiment classification is proposed. It uses Long Short-Term Memory (LSTM) model to find sentiments from images.

The labelled data in a source language is taken to learn a model and classify sentiments in a target language. Let the labelled English documents be 'l_{src}', the un-labelled data be 'u_{src}'. Let the target language be 'u_{tar}'. The

unlabelled data containing Emojis are often taken from social media like Twitter. The sentiment polarity of document written in the target language is computed based on the labelled data in the source language (l_{src}) and diverse forms of unlabelled data.

Sentence representation models are built for the source and the target languages. From word embedding, sentence representations are learnt by envisaging the emoji. This is a supervised learning process, wherein emojis are taken as sentiment labels. The labelled document in English is interpreted into the target language using Google Translate taking sentence by sentence.

The language representations are obtained by feeding the sentences and their translations into the representation models. The sentences are aggregated to reduced representations of the training document both in English and target language. The representations are used as features for forecasting the sentiment label of the document and classifying it. During testing, the document in the target language is translated into English and the above steps are followed to get the representation. Based on the representation, the label is predicted using the classifier.

Let the sentences in 'emoji_{src}' or 'emoji_{tar}' be (x, e), where,

$$x = [v_1, v_2 \dots v_L] \quad (1)$$

a sequence of word vectors indicating plain text without emojis

e - Emoji in the text

At step 't', the unit states are found as follows:

$$ip^{(t)} = \sigma(u_{ip}v_t + w_{ip}h^{(t-1)}b_{ip}) \quad (2)$$

$$fg^{(t)} = \sigma(u_{fg}v_t + w_{fg}h^{(t-1)}b_{fg}) \quad (3)$$

$$op^{(t)} = \sigma(u_{op}v_t + w_{op}h^{(t-1)}b_{op}) \quad (4)$$

$$cell^{(t)} = fg_t \cdot cell^{(t-1)} + ip^{(t)} \cdot \tanh(u_{cell}v_t + w_{cell}h^{(t-1)}b_{cell}) \quad (5)$$

$$h^{(t)} = op^{(t)} \cdot \tanh(cell^{(t)}) \quad (6)$$

At step 't',

ip^(t) - Input gate

fg^(t) - Forget gate

op^(t) - Output gate

cell^(t) - Memory cell

h^(t) - Hidden layer

w - Weights

u - Input weights

b - Biases

. - Element-wise product

Latent vectors are extracted from LSTM. To obtain information from the given context, before and after a word, bi-directional LSTM is used. In LSTM the gates play an vital role to outsmart the difficulties faced in Recurrent Neural Network. Latent vectors are joined from both directions and a bi-directional encoded vector is built.

$$\vec{H}_1 = \overrightarrow{\text{LSTM}}(v_i), i \in [1, L] \quad (7)$$

$$\overleftarrow{H}_1 = \overleftarrow{\text{LSTM}}(v_i), i \in [L, 1] \quad (8)$$

$$H_i = [\vec{H}_i, \overleftarrow{H}_i] \quad (9)$$

Based on the pre-trained sentence representations, document representations are learnt and cross-lingual sentiment classification is performed.

For each English document, the pre-trained representation models are used to embed the sentences in it. Secondly, the representations are combined into a compressed form.

As diverse portions contribute in different ways, the sentiment is obtained using attention mechanism. The next chapters discusses about comparison between various classifications techniques and their performances.

IV. CLASSIFICATION USING KNN, SVM AND CNN

In this paper, Emojis are classified into positive, negative and neutral using K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

This section throws light on each of these methods. It is seen that LSTM yields best results with CNN.

4.1 K-Nearest Neighbour (KNN)

K-Nearest Neighbour (K-NN) is parameter independent method involved in classification and regression (Altman 1992). K-closest training samples are taken in the feature space. K-NN is a kind of instance-based learning also known as lazy learning. The functions are locally approximated and the computations are postponed until classification. It is the simplest of all machine learning algorithms.

Output depends on the usage, classification or regression. It is widely used in classification. The aspects include ease to interpret output, calculation time and predictive capability.

When K-NN is used for classification, the output gives the class membership. Objects are classified by taking the vote obtained from neighbours. The objects are assigned to a class that is highly common among its K-NNs.

In case of regression, the output gives the property of the object which is the average of the values of KNNs.

4.2 Support Vector Machine (SVM)

Support Vector Machines (SVMs) is one of the new techniques for pattern classification, and it has been widely used in many application areas. It is a supervised learning model associated with learning algorithms to analyse data and to recognize patterns. The kernel parameters setting for SVM in training process impacts on the classification accuracy. Feature selection is another factor that impacts classification accuracy. Support vector machines (SVM) were first suggested by Vapnik (1995) and have recently been used in a range of problems including pattern recognition (Pontil & Verri 1998), bioinformatics (Yu et al. 2008) and text categorization (Joachims 1998).

SVM classifies data with different class labels by determining a set of support vectors that are members of the set of training inputs that outline a hyperplane in the feature space. SVM provides a generic mechanism that fits the hyperplane surface to the training data using a kernel function. The user may select a kernel function (e.g. linear, polynomial, or sigmoid) for the SVM during the training

process that selects support vectors along the surface of this function.

In addition to linear classification SVM is proved to be best suited for non-linear classification using kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

4.3 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN), a Deep Learning (DL) algorithm takes the input and assigns weights to various aspects and differentiates them. CNN involves less pre-processing in contrast to other classification algorithms. CNNs are capable of learning characteristics (Fukushima 2007).

The architecture of a CNN is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. CNNs are stimulated by biological processes. The connection between neurons is similar to animal visual cortex. Individual neurons react to stimuli in the receptive field.

V. RESULTS & DISCUSSION

Twitter dataset with weekend and vacation based tweets are taken for study. The dataset is divided into positive (happy emoticons) and negative (sad emoticon) samples. Neutral emotions can also be detected in between.

The performance of the system is measured in terms of Accuracy, Precision, Recall, F-Measure and Time Period.

✓ **Classification Accuracy:** Classification accuracy is the most commonly used measure for determining performance of classifiers. It is the rate of number of correct predictions made by a model over a data set.

✓ **Sensitivity and Specificity:** Sensitivity and Specificity: Sensitivity is the true positive rate, and specificity is the true negative rate.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{FP+TN}$$

The convolution neural networks (CNN) and long Short term Memory (LSTM) together works to do magical kind of results in the review dataset of social media big data.

Irrespective of voluminous dataset and dimensionality it works rationally to yield impeccable way of results in classification. It is seen that the proposed architecture CNN+LSTM scheme offers better results when compared to the existing classification algorithms like KNN, SVM, CNN, LSTM+KNN and LSTM+SVM

Sentence Level Sentiment Analysis using Deep Learning Method

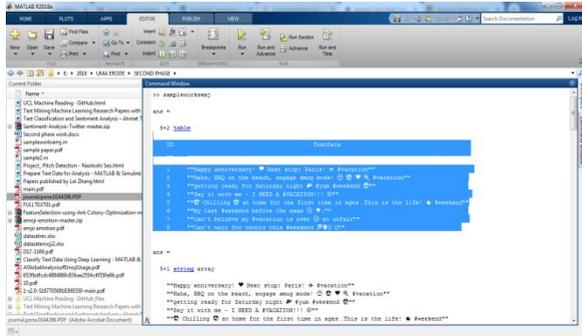


Figure 1: Input Review Data

Figure 1 shows the pre-processed twitter dataset where all the cleaning activities took place like steaming, tokenization, pruning, and removal of stop words, etc.

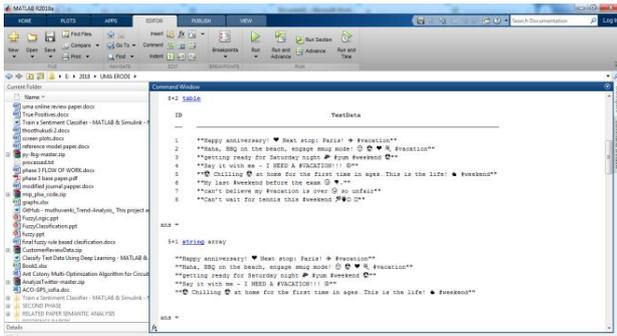


Figure 2: Pre-Processed Output

A complete set a clear review data set is obtained ready for the next action.

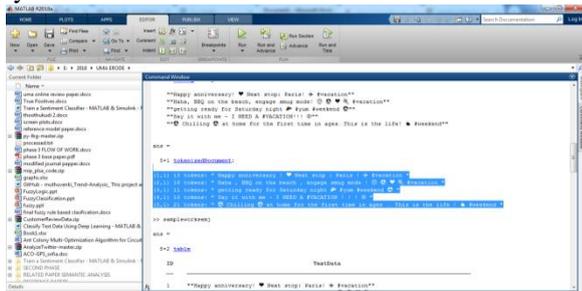


Figure 3: Emoji and Text classification

A set of review with text and emoji is taken together in the work to extract the key attribute using the proposed architecture.



Figure 4: key features of emoji and text from twitter dataset

A sample screenshot of received features involving top rated words with hashing in twitter dataset showing the result of

proposed architecture.

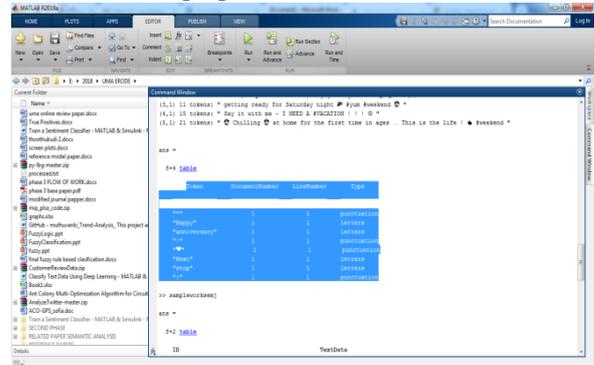


Figure 5: resultant of feature selection

The above figure explicitly indulges an action separating emoji from text for extracting key terms and the way to define scores.



Figure 6: Emoji Detection using LSTM+CNN

A comparative analysis of different life classification algorithm is discussed in this selection.

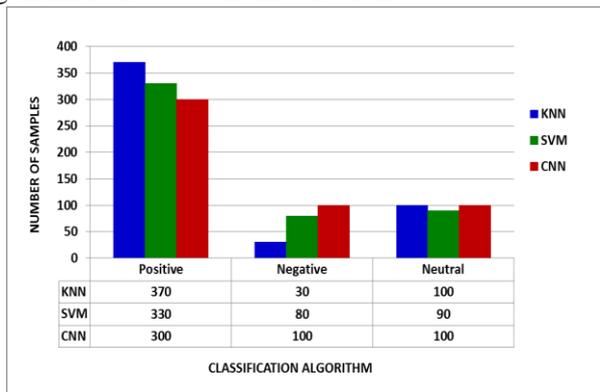


Figure 7: Number of polarity in the dataset

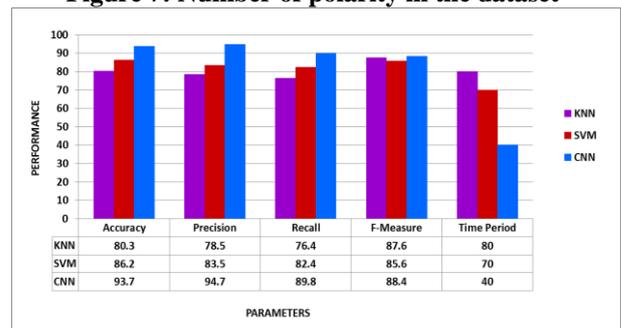


Figure 8: Performance of the different classification algorithm

The resultant graphs shows a legible idea of obtained positive, negative and neutral scores in the dataset. It is considerably proved that the Convolutional Neural Network works hefty working well in all the grounds of parameters.

Also when it is added up with long short term memory the promising results are alluring in the fixed parameters on the other hand the time period has also been reduced in predicting the labels. Therefore a picturesque architecture is designed to work for sentiment analysis.

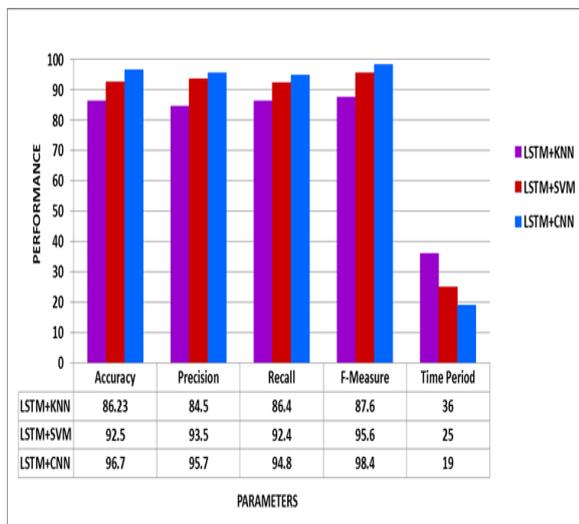


Figure 9: Performance of classification algorithm with LSTM

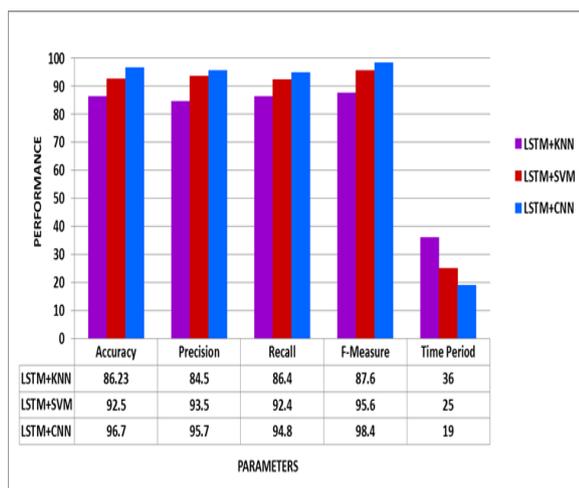


Figure 10: Performance of the Proposed LSTM+CNN

VI. CONCLUSION

Long Short-Term Memory (LSTM) - Convolutional Neural Networks (CNN) based model is proposed to segregate Emojis from text, calculate score, classify sentiments into positive, negative or neutral. This proposed methodology strives to complement the label deficient languages aiding in multi-lingual sentiment analysis. The main aim is to classify text and emojis using Deep Learning techniques. It is seen that Deep Learning based technique yields better performance compared to the existing K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) methods of classification.

REFERENCES

1. Ai, W., Lu, X., Liu, X., Wang, N., Huang, G., & Mei, Q. (2017, April). Untangling emoji popularity through semantic embeddings. In Eleventh International AAAI Conference on Web and Social Media.
2. Bi, Y., Bell, D., Wang, H., Guo, G., & Guan, J. (2007). Combining multiple classifiers using Dempster's rule for text categorization. *Applied Artificial Intelligence*, 21(3), 211-239
3. Barbieri, F., Kruszewski, G., Ronzano, F., & Saggion, H. (2016, October). How cosmopolitan are emojis?: Exploring emojis usage and meanings over different languages with distributional semantics. In Proceedings of the 24th ACM international conference on Multimedia (pp. 531-535). ACM.
4. Chen, Z., Lu, X., Ai, W., Li, H., Mei, Q., & Liu, X. (2018, April). Through a gender lens: learning usage patterns of emojis from large-scale android users. In Proceedings of the 2018 World Wide Web Conference on World Wide Web (pp. 763-772). International World Wide Web Conferences Steering Committee.
5. Cramer, H., de Juan, P., & Tetreault, J. (2016, September). Sender-intended functions of emojis in US messaging. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 504-509). ACM.
6. Choudhary, N., Singh, R., Rao, V. A., & Shrivastava, M. (2018, August). Twitter corpus of resource-scarce languages for sentiment analysis and multilingual emoji prediction. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1570-1577).
7. Cui, H., Lin, Y., & Utsuro, T. (2018). Sentiment Analysis of Tweets by CNN utilizing Tweets with Emoji as Training Data.
8. Da Silva, N. F., Hruschka, E. R., & Hruschka Jr, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66, 170-179.
9. Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M., & Riedel, S. (2016). emoji2vec: Learning emoji representations from their description.
10. Fukushima, K. (2007). Neocognitron. *Scholarpedia*, 2(1), 1717.
11. Fernández-Gavilanes, M., Juncal-Martínez, J., García Mendez, S., Costa-Montenegro, E., & González-Castaño, F. J. (2018). Creating emoji lexica from unsupervised learning of their descriptions. *Expert Systems with Applications*, 103, 74-91.
12. Guibon, G., Ochs, M., & Bellot, P. (2016, June). From emojis to sentiment analysis.
13. Ge, J., & Gretzel, U. (2018). A social media influencer perspective. *Journal of Marketing Management*, 34(15-16), 1272-1295.
14. Hern, Alex (February 6, 2015). "Don't know the difference between emoji and emoticons? Let me explain". *The Guardian*.
15. Hu, T., Guo, H., Sun, H., Nguyen, T. V. T., & Luo, J. (2017, May). Spice up your chat: the intentions and sentiment effects of using emojis. In Eleventh International AAAI Conference on Web and Social Media.
16. Hussien, W. A., Tashtoush, Y. M., Al-Ayyoub, M., & Al-Kabi, M. N. (2016, July). Are emoticons good enough to train emotion classifiers of Arabic tweets?. In 7th IEEE International Conference on Computer Science and Information Technology (CSIT) pp. 1-6
17. Hussien, W., Al-Ayyoub, M., Tashtoush, Y., & Al-Kabi, M. (2019). On the Use of Emojis to Train Emotion Classifiers.
18. Kopev, D., Atanasov, A., Zlatkova, D., Hardalov, M., Koychev, I., Nikolova, I., & Angelova, G. (2018, June). Tweety at semeval-2018 task 2: Predicting Emojis using hierarchical attention neural networks and support vector machine, In Proceedings of the 12th International Workshop on Semantic Evaluation (pp. 497-501).
19. Ljubešić, N., & Fišer, D. (2016, August). A global analysis of emoji usage. In Proceedings of the 10th Web as Corpus Workshop (pp. 82-89).
20. Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., & Mei, Q. (2016, September). Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 770-780). ACM.

21. Mathews, S., & Lee, S. E. (2018). Use of Emoji as a Marketing Tool: An Exploratory Content Analysis. *Fashion, Industry and Education*, 16(1), 46-55.
22. Miller, H. J., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016, March). "Blissfully Happy" or "Ready to Fight": Varying Interpretations of Emoji. In *Tenth International AAAI Conference on Web and Social Media*.
23. Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), 764-779.
24. Pontil, M & Verri, A 1998, 'Support vector machines for 3D object recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 637-646.
25. Seyednezhad, S. M. M., Fede, H., Herrera, I., & Menezes, R. (2018, May). Emoji-word network analysis: Sentiments and semantics. In *The Thirty-First International Flairs Conference*.
26. Shiha, M., & Ayzav, S. (2017). The effects of emoji in sentiment analysis. *Int. J. Comput. Electr. Eng.(IJCEE.)*, 9(1), 360-369.
27. Wolny, W. (2016). Emotion analysis of Twitter data that use emoticons and emoji ideograms.
28. Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181(6), 1138-1152.
29. Vapnik V (1995), 'Support-vector networks', *Machine Learning*, vol.20, pp.273-297.

AUTHORS PROFILE



K.K.UMA is currently pursuing Ph.D as a part time research scholar under the guidance of DR.K. Meenakshi Sundaram in the department of computer science, Erode Arts and science college[Autonomous] ,Erode, Tamilnadu, India. She has obtained her master degree and M.Phil degree in computer science under Bharatiyar University Coimbatore. Her area of interest includes Data Mining and Big Data. she has presented 4 papers in national and international

level conferences. She published 5 international journal



DR.K. Meenakshi Sundaram having 31 years of teaching experience. He produced 45 M.Phil research scholars in computer science. Also 6 Ph.D. public viva voice examination as examiner. He scored as a member in board of studies of various universities and Autonomous colleges. He published 45 papers in International ournals and presented 60 papers in National /International conferences. His area of interest is Data Mining

and Software Engineering.