

MULTIMEDIA RETRIEVAL USING EMOJI PREDICITON, ANTICIPATION AND RETRIEVAL

¹ Dinesh Kumar S, ² Sarath P, ³ Reena R, ⁴ Kapilavani R K

^{1,2}. Student, ³ Supervisor, ⁴ Coordinator, Department of Computer Science

Prince Shri Venkateshwara Padmavathy Engineering College, Ponmar, Chennai.

pavansdinesh@gmail.com

ABSTRACT

Over the past decade, emoji have emerged as a new and widespread form of digital communication, spanning diverse social networks and spoken languages. We propose treating these ideograms as a new modality in their own right, distinct in their semantic structure from both the text in which they are often embedded as well as the images which they resemble. As a new modality, emoji present rich novel possibilities for representation and interaction.

In this paper, we explore the challenges that arise naturally from considering the emoji modality through the lens of multimedia research, specifically the ways in which emoji can be related to other common modalities such as text and images. This study proposes predicting the emojis from the given input such as text, image and emoji using machine learning algorithm (Support Vector Machine) and deep learning algorithm (Deep convolutional Neural Network) with greater accuracy.

Keywords- Emoji, Text, Image, Support Vector Machine, Deep Convolutional Neural Networks, Machine Learning, Deep Learning.

I. INTRODUCTION

EMOJI, small ideograms depicting objects, people, and scenes, have exploded in popularity. They are now available on all major mobile phone platforms and social media websites, as well as many other places. According to the Oxford English Dictionary, the term emoji is a Japanese coinage meaning 'pictogram', created by combining e (picture) with moji (letter or character). Emoji as we know them were first introduced as a set of 176 pictogram available to users of Japanese mobile phones. The available range of ideograms has expanded greatly over the previous years, with 1,144 single emoji characters defined in Unicode 10.0 and many more defined through combinations of two or more emoji characters.

Emojis on smartphones, in chat, and email applications have become extremely popular worldwide. For example, Instagram, an online

mobile photo-sharing, video-sharing and social networking platform, reported in March 2015 that nearly half of the texts on Instagram contained emojis, In this study we have three modules for predicting emoji from text, image and emoji respectively.

➤ **Text to Emoji**

For a given input text message, the corresponding emojis can be predicted with a greater accuracy with the use of machine learning algorithm like SVM (Support Vector Machine) algorithm.

➤ **Emoji to Emoji**

For a given emoji as an input, it's relevant emojis can be predicted or retrieved.

➤ **Image to Emoji**

For a given image as an input, the corresponding images can be retrieved.

II. RELATED WORK

We start with introducing the background and literature related to our research. Our study is inspired by three streams of literature: nonverbal elements in communication, sentiment analysis, and information diffusion. Emoticons and Emojis People have been long using emoticons to provide non-verbal cues in online communications (Walther and Daddario 2003; Park et al. 2013). Such nonverbal cues help

people better interpret the nuance of meaning and the level of emotion not captured by language elements alone (Gajadhar and Green 2005; Lo 2008). There has also been research on the sentiment of emoticons (Boia et al. 2013), which reported that the sentiment of an emoticon is in substantial agreement with the sentiment of the entire tweet. Since the debut on Twitter and Instagram, emojis quickly expanded their territory from emotions to various objects (sports, foods, etc.). Researchers have been trying to understand the interpretation of emojis. and how emojis facilitate communications (Kelly and Watts 2015; Vidal, Ares, and Jaeger 2016), with a particular interest in their sentiments (Kralj et al. 2015). A key question is to find a good representation of emojis. Some researchers use the official description on the Unicode Website (Lu et al. 2016), while others utilize the word embedding tools to represent emojis with high-dimensional vectors (Eisner et al. 2016; Barbieri, Ronzano, and Saggion 2016).

Indeed, several word embedding tools have been developed to find distributed representations of words, and have shown great potentials in various tasks, such as classification and visualization (Tang et al. 2015; Mikolov et al. 2013b; 2013a). Similar to Barbieri, Ronzano, and Saggion (2016), we also applied a state-of-art embedding model to project words and emojis onto the same high-dimensional

vector space, from where we conduct extensive analysis on the popularity of emojis. Sentiment Analysis Both emoticons and emojis have been widely used to express the emotions, which relate our work to the sentiment analysis literature. Sentiment analysis has long been a core problem of natural language processing (Pang and Lee 2008; Mei et al. 2007; Liu 2012). Although various advanced sentiment analysis techniques have been proposed, accurately identifying sentiments and emotions from free text is still very challenging. The emergence of emoticons and emojis provide new opportunities to analyze the sentiment expressions in textual context. Many researches have attempted to model text sentiments with emoticons and emojis used in the context (Zhao et al. 2012; Kralj et al. 2015).

Instead of analyzing the predicting power of emojis on the sentiment of the message, we are curious if the sentiments of messages would predict the usage of the emojis. Since the messages in the corpus are usually short, we chose lexicon-based approaches for sentiment analysis (Wilson, Wiebe, and Hoffmann 2005; Hu and Liu 2004; Kiritchenko, Zhu, and Mohammad 2014). Researchers have discovered various factors that explain the adoption and virality of certain messages, such as social network structure (Romero, Tan, and Ugander 2013), wording (Tan, Lee,

and Pang 2014), serendipity (Sun, Zhang, and Mei 2013), and other lexical characteristics (Danescu-Niculescu-Mizil et al. 2012). The object in this study is slightly different, as the emoji is brand-new non-verbal language units, and is phenomenally adopted by Internet users. Although there have also been re-researches on the adoption of both emoticons and emojis (Park et al. 2013; Lu et al. 2016), these researches focus on the culture-level difference in adoption, while our work tackles directly at the popularity of individual emoji.

A. MACHINE LEARNING

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning is closely related to computational statistics, which focuses on

making predictions using computers. Thus automatic accident detection systems are the need of time, which can reduce the accidents.

TYPES OF MACHINE LEARNING

- Supervised Learning Model
- Unsupervised Learning Model
- Semi Supervised Learning

i) SUPERVISED LEARNING MODEL.

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs. The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and the desired output, also known as a supervisory signal. In the mathematical model, each training example is represented by an array or vector, sometimes called a feature vector, and the training data is represented by a matrix.

Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs. An optimal function will allow the algorithm to correctly determine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its outputs or predictions over time is said to have learned to perform that task. Supervised learning algorithms classification and regression.

Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. Similarity learning is an area of supervised machine learning closely related to regression and classification, but the goal is to learn from examples using a similarity function that measures how similar or related two objects are.

ii) UNSUPERVISED LEARNING

Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data.

Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within the same cluster are similar according to one or more pre designated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some similarity metric and evaluated, for example, by internal

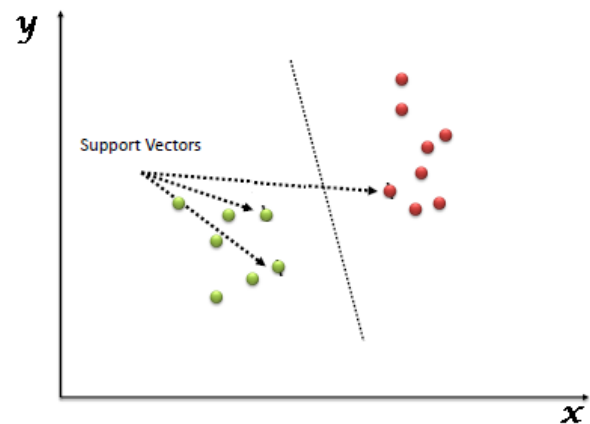
compactness, or the similarity between members of the same cluster, and separation, the difference between clusters. Other methods are based on estimated density and graph connectivity.

iii) SEMISUPERVISED LEARNING

Semi-supervised learning falls between unsupervised learning (without any labelled training data) and supervised learning (with completely labelled training data). Many machine-learning researchers have found that unlabelled data, when used in conjunction with a small amount of labelled data, can produce a considerable improvement in learning accuracy.

iv) SUPPORT VECTOR MACHINE

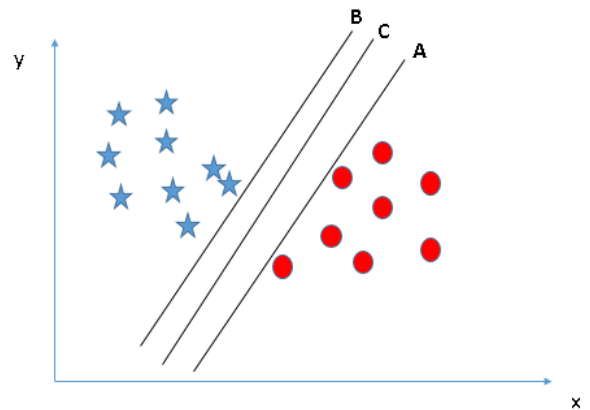
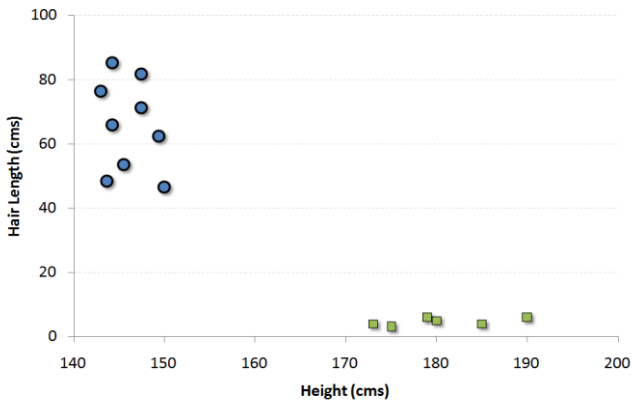
Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for either classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.



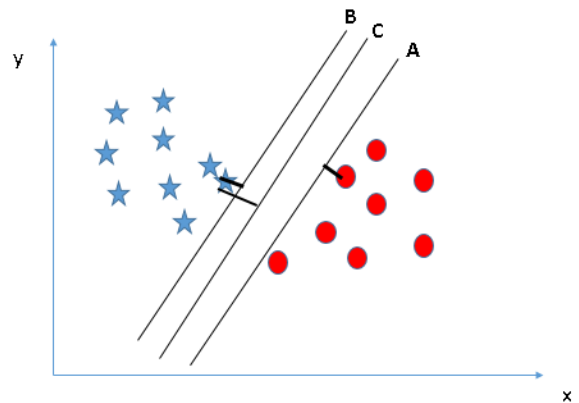
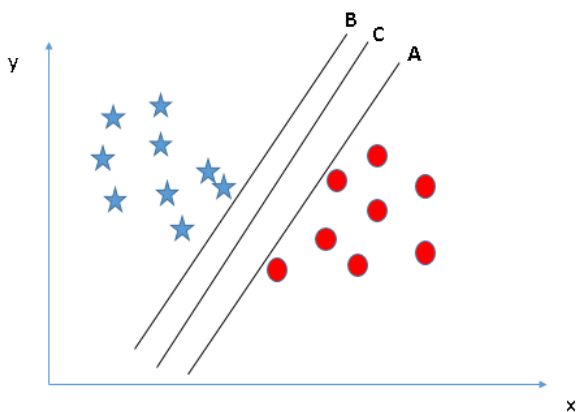
Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

v) CLASSIFICATION ANALYSIS

Let's consider an example to understand these concepts. We have a population composed of 50%-50% Males and Females. Using a sample of this population, you want to create some set of rules which will guide us the gender class for rest of the population. Using this algorithm, we intend to build a robot which can identify whether a person is a Male or a Female. This is a sample problem of classification analysis. Using some set of rules, we will try to classify the population into two possible segments. For simplicity, let's assume that the two differentiating factors identified are: Height of the individual and Hair Length. Following is a scatter plot of the sample.

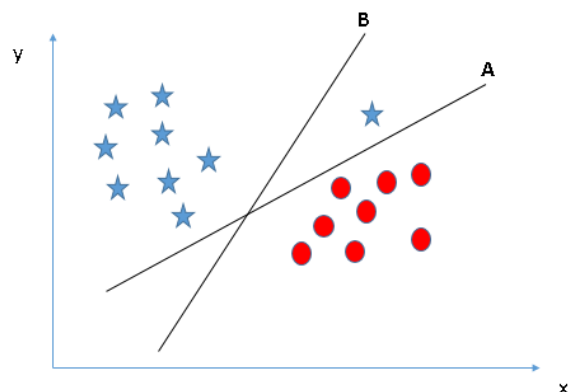


Identify the right hyper-plane Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.



Identify the right hyper-plane (Scenario-3): Use the rules as discussed in previous section to identify the right hyper-plane

Identify the right hyper-plane (Scenario-2): Here, we have three hyper-planes (A, B and C) and all are segregating the classes well. Now, how can we identify the right hyper-plane?



Can we classify two classes (Scenario-4) Below, I am unable to segregate the two

classes using a straight line, as one of star lies in the territory of other(circle) class as an outlier.



Pros and Cons associated with SVM

Pros

- It works really well with clear margin of separation
- It is effective in high dimensional spaces.
- It is effective in cases where number of dimensions is greater than the number of samples.
- It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

Cons

- It doesn't perform well, when we have large data set because the required training time is higher

- It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping

B. DEEP LEARNING

Deep Learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabelled. Also known as deep neural learning or deep neural network. Deep learning is an AI function that mimics the workings of the human brain in processing data for use in detecting objects, recognizing speech, translating languages, and making decisions. Deep learning AI is able to learn without human supervision, drawing from data that is both unstructured and unlabelled.

i) DEEP CONVOLUTIONAAL NEURAL NETWORK

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels

that scan the hidden layers and translation invariance characteristics.

Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. Typical ways of regularization include varying the weights as the loss function gets minimized while randomly trimming connectivity.


























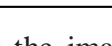
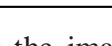

CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in the filters.

III. REGRESSION ANALYSIS MODEL

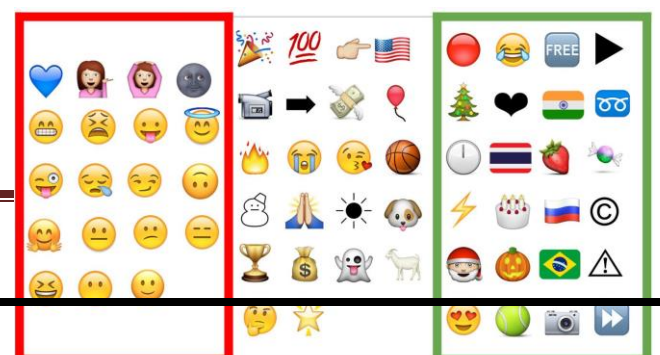
A. OVERVIEW OF DATASET

To facilitate research on these challenges, it is necessary to use a dataset with sufficient examples of the relationship between emoji and other modalities. Existing works on emoji have either forgone the use of an annotated emoji dataset or have used datasets comprised of only a small subset of available emoji. Both of these settings are artificial and fail to adequately represent the challenge and promise of emoji. Instead, we target the full range of potential emoji, including their very long tail, and seek to learn their real-world usage rather than place any prior assumptions on them.

The balanced subset is selected such that no single emoji annotation applies to more than 10 examples. To train toward this objective while still leveraging the breadth of the available data, we construct our mini batches so that each emoji has an equal chance of being selected.

	Image		Image-only	Text-only	True Emoji
A		rt U : nah this nymar x jordan collab is pure heat			
B		rt U : one of the short poetry i have done , #watercolor #art			
C		thank you			
D		turned my ghetto concrete workshop room into my own cool little space			
E		no one will ever understand what it's like to have a best friend like this so lucky i am U			
F		not food			
G		im that weird girl that likes to hold snakes			

A surprising result is that the image modality actually out- performs the text modality in most of the metrics. Because the semantic space is learned on textual data, one might expect the text modality to be the most reliably embedded modality within the shared space, but that does not seem to be the case. Perhaps this is a result of many distracting terms in the textual data, which supervised approaches learn to filter out. Meanwhile, the limited vocabulary of the CNN



concepts are likely to be a strong signal. Nonetheless, the fusion of the two modalities improves performance across all metrics.

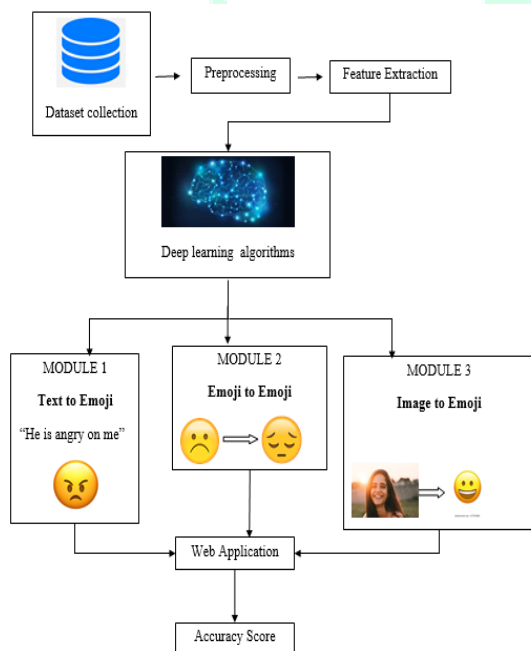
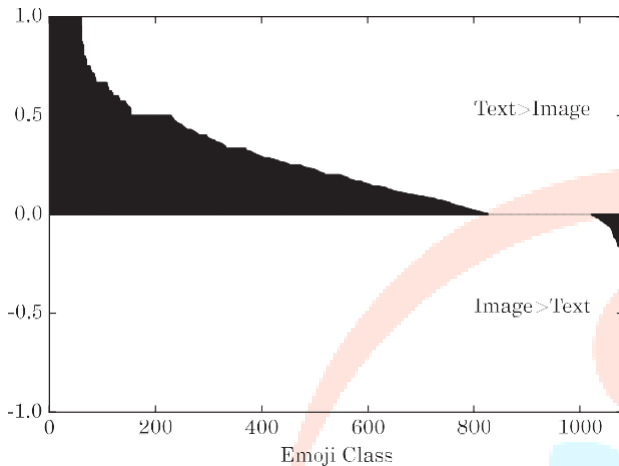


Fig: 3.1 Architecture Diagram

i) LIST OF MODULES

- Text to Emoji

- Emoji to Emoji

- Image to Emoji

❖ Text to Emoji:

For a given input text message, the corresponding emojis can be predicted with a greater accuracy with the use of machine learning algorithm like SVM (Support Vector Machine) algorithm.

❖ Emoji to Emoji:

Using a Deep convolutional Neural Network, for a given emoji as an input, it's relevant emojis can be predicted or retrieved

❖ Image to Emoji:

Using a Deep convolutional Neural Network, for a given image as an input, the corresponding images can be retrieved.

V. PERFORMANCE METRICS

We need to correlate two properties of the Emoji Sentiment Ranking with other data. In the first case we correlate the emojis ranked by occurrence to the Emoji tracker list—the property of the list elements is the number of occurrences. In the second case we correlate the emojis ranked by sentiment to subsets of emojis from the 13 different languages—the property of the list elements is the sentiment score. For any two lists x and y , of length n , we first compute the Pearson correlation coefficient [29]:

$R(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$, where \bar{x} and \bar{y} are the list's mean values, respectively. The Spearman's rank correlation coefficient [30] is computed in the same way, the property values of the x and y elements are just replaced with their ranks. In both cases we report the correlation coefficients at the 1% significance level.

$$\text{Alpha} = 1 - D_o/D_e,$$

where D_o is the observed disagreement between the annotators, and D_e is the disagreement expected by chance. When the annotators agree perfectly, $\text{Alpha} = 1$, and when the level of agreement equals the agreement by chance, $\text{Alpha} = 0$. The two disagreement measures are defined as follows:

$$D_o = \frac{1}{N} \sum_{c, c'} N(c, c') \cdot \delta^2(c, c'),$$

$$D_e = \frac{1}{N} \sum_{c, c'} (N - 1) X_{c, c'} N(c) \cdot N(c') \cdot \delta^2(c, c').$$

The arguments, N , $N(c, c')$, $N(c)$, and $N(c')$, refer to the frequencies in a coincidence matrix, defined below. $\delta(c, c')$ is a difference function between the values of c and c' , and depends on the metric properties of the variable. In our case, for the discrete sentiment variables c and c' , the difference function δ is defined as:

$$\delta(c, c') = |c - c'|, c, c' \in \{-1, 0, +1\}.$$

In [33], this is called the interval difference function. Note that the function attributes a disagreement of 1 between the negative (or positive) and the neutral sentiment, and a

disagreement of 2 between the negative and positive sentiments.

VI. CONCLUSION

In this paper, we have approached emoji as a modality distinct from text and images. There is sufficient motivation for doing so, and considerable future opportunities for research and applications with the emoji modality. Emoji are everywhere, and are becoming more pervasive. It is our hope that this work and the challenge tasks defined within further research and understanding of emoji within the multimedia community. In this paper we propose the prediction of emojis from a given text, image and emoji with a greater accuracy. Unlike the usual text message transmission between the users, the emoji usage will be very useful and the communication can be done effectively. The use of emoji has its wide range of applications starting from customer review, breaking language barriers and to many of its kind.

VII. FUTURE ENHANCEMENT

Currently English is the only language that is used for the text classification and for the emoji prediction. In future many languages other than English language can be used. Using Web scrapping method, the input text can be provided in the future.

VIII. REFERENCE

[1]W.Ai *et al.*, “Untangling emoji popularity through semantic embeddings,” in *Proc. 11th Int. AAAI Conf. Web Social Media*, 2017, pp. 2–11.

[2]L. M. Aiello *et al.*, “Sensing trending topics in twitter,” *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1268–1282, Oct. 2013.

[3] Z. Akata, F.Perronnin, Z. Harchaoui, and C. Schmid, “Label-embedding for image classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1425–1438, Jul. 2016.

[4]F. Barbieri, M. Ballesteros, and H. Saggion, “Are emojis predictable? In *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics: vol. 2, Short Papers*, 2017, pp. 105–111.

[5]F.Barbieri, G.Kruszewski, F. Ronzano, and H. Saggion, “How cosmopolitan are emojis: Exploring emojis usage and meaning over different languages with distributional semantics,” in *Proc. ACM Conf. Multimedia*, 2016, pp. 531–535.