**Using neural networks to analyze tweets and predict the emoji they evoke**

By: Andrew Virts

**Abstract**

With the increase in development of social media, emojis are becoming more and more widely used by users when posting their messages. Therefore, it is important to study and understand the relationships between emojis and plain text messages. In “Tweet Emoji Prediction Using Hierarchical Model with Attention”, the authors present a neural approach to predicting multiple emojis evoked by plain text tweets (Wu, Wu, Wu, Huang, & Xie, 2018). The researchers use a model that contains three modules. The first is a character encoder to learn different representations of words based off their original characters using a convolutional neural network(CNN). The second is a sentence encoder that is used to learn representations of sentences using a combination of long and short-term memory, known as LSTM. The final model is a multi-label classification model that uses a convolutional neural network to predict the emojis that will be evoked by a tweet. Finally, the researchers apply attention mechanism at the word-level to select and determine important contexts.

**Introduction**

As social media rapidly grows and expands, so does the use of emojis. New emojis, each with their own unique meanings and interpretations are added with almost every update. Emojis are often combined with plain-text messages to convey different expressions and emotions (Wu et al., 2018). This presents a challenging task for researchers; they must be able to parse a text message and determine the meaning of the emoji or emojis within the text based off the context that surrounds it. This means that the researchers must develop a model that can recognize and understand context and the internet slang that is frequently used on social media and apply this to determine the correct emoji(s) to go along with the plain-text emojis. Being able to successfully analyze the relationship between emojis and text messages has a plethora of benefits, including emoji-based text generation, online information retrieval, and improved social media analysis (Wu et al., 2018). This research project is unique because most research projects out there focus on analyzing the usage and semantics of specific emojis, not the emoji and the message as a whole. These researchers want to be able to predict multiple emojis based off the context of tweets. The researchers propose a hierarchical neural model that utilizes attention mechanism to accomplish this difficult task. This hierarchical neural model is made up of three modules; the first is a character encoder designed to learn the hidden representations of words from their original characters using a convolutional neural network. The second is a word encoder that is used to learn sentence representations; it captures all different kinds of contextual information using a combination of Long short-term memory and convolutional neural nets (Wu et al., 2018). Since there are a large amount of words in tweets that are considered “uninformative”, attention mechanism is applied in this module to capture the “informative” words. Lastly, the third module is an emoji classification module that is used to predict the emojis. Combining these modules gives you an extremely intelligent model that is capable of combing through tweets and predicting the emojis that best match the tweet.

**Materials and methods**

The researchers for this project use an extremely intelligent model that is capable of parsing plain-text tweets, separating the “informative” and “uninformative” words, and determining the meanings of the emojis in these tweets based off the language around them. The model is extremely detailed and was highlighted briefly in the introduction.

A close up of a map

Description automatically generated

The first module of this project is a *Character Encoder*. It consists of two parts, the first is a character embedding layer that breaks down the character sequence of a word into a sequence of vectors (Wu et al., 2018). The resulting output of this layer is a vector sequence. Secondly, a convolutional neural network is used to break down words at the character level. It captures local information about these characters and uses this information to build hidden representations of each word. The output for this first module is the hidden representation of each word. There is often a lot of strange vocabulary, unique internet slang, and creative spelling used in tweets. Due to the unorthodox nature of social media lingo, it can difficult to understand and predict the meanings behind some of these phrases and they can present challenges for the model, leading to weak representations of these words. This is where the character encoder comes in; it learns from the original characters used in this “social media lingo” to find the patterns that are often present within these phrases and uses these patterns to enhance the representation of these words (Wu et al., 2018). Character-level analysis can also be used to provide emotion clues; the use of all-caps, an exclamation mark or a question mark can be indicative of the emotion presented by the tweet.

The second module consists of the *Word Encoder*, it is used to analyze and learn the representation of sentences at the word-level (Wu et al., 2018). The Word Encoder contains four components, the first is a word embedding layer. This layer is used to convert the sequences of words into a sequence of vectors (Wu et al., 2018). The resulting vectors will then be concatenated with the output from the character encoder, giving us the final word representations. The second part is the Bi-directional LSTM (Bi-LSTM) layer. LSTM is a proven technique used to capture long-distance information. For this project, both past and future contexts of words are used to inform the model of how it should build its representations of words, so a Bi-directional Long Short-Term Memory Network is employed to scan the input sequences in both directions (Wu et al., 2018). It takes the word sequence as input, and outputs the hidden representations of the words. The third component of module two is a *word-level convolutional neural network.* A convolutional neural network is a deep learning algorithm that takes an input and assigns varying degrees of importance to various objects of the input so it can differentiate them from one another. Local context is extremely important to help infer the meanings of sentences, which is then used to predict emojis. The CNN in this component is employed to capture local context and build representations of words based off of these contexts. It takes the hidden representations of words as input, and outputs feature maps. The fourth and final component of this model is the attention layer. Given different contexts, words can have varying degrees of meaning determined by how they are used in the sentence. For example, in the sentence “Cold weather makes me depressed”, the word “depressed” is much more informative to emoji prediction than the word “me”. Attention mechanism is used to highlight and select important contexts and separate them from the unimportant contexts.

The final module of this model is the *Multilabel Emoji Classification* module. The sole purpose of this module is to jointly predict whether each emoji is evoked by the input tweet or not (Wu et al., 2018). This means that the module will try to predict multiple mutually dependent outcomes and use these outcomes to help predict the rest of the sequence.

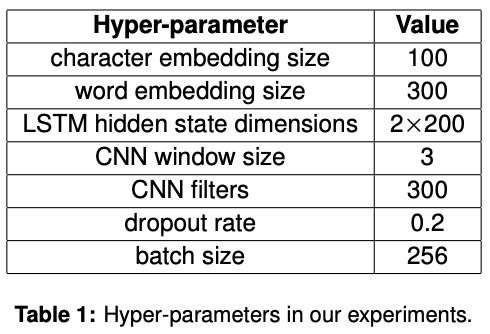
Combine all three modules together and you have a fully function neural network model capable of parsing at the character-level and word-level and using this information to predict the emojis evoked by tweets. The model analyzes context and uses its own representations of words and characters it makes to create its predictions. This model has been shown to outperform several baselines as well as humans at predicting emojis from tweets.

A screenshot of a cell phone

Description automatically generated

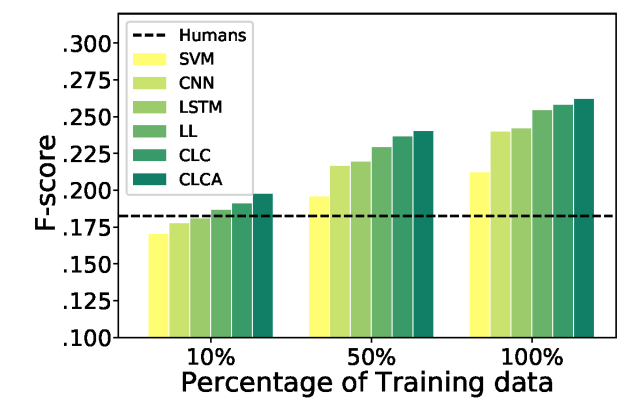
**Experiments**

The dataset used for this model was retrieved using twitter’s API. The total dataset contains 3,720,122 tweets. This project decided to use just the 30 most frequently used emojis for their model. Out of the total dataset, 896,441 tweets contained one of the 30 emojis (Wu et al., 2018). Out of the 896,441 tweets containing emojis, 36.8% of these contained more than one distinct emoji (Wu et al., 2018). The researchers randomly selected 500,000 tweets to use to train their model, 50,000 for validation, and another 50,000 for the actual test (Wu et al., 2018).

 The researchers maintain several hyper-parameter settings that were selected via cross validation on the training set (Wu et al., 2018). Hyper-parameters are parameters that are external to the module and cannot be estimated from the data, they are completely independent factors. These hyper parameters are used to help estimate the model parameters and find better fitting data. A dropout strategy is applied at each layer in the network and is implemented to mitigate overfitting (Wu et al., 2018). Overfitting can occur when the model doesn’t generalize well from the training data given to it, to unseen data. The researchers use the macro F-score over all the emojis as the task metric (Wu et al., 2018). Each experiment is repeated 10 times and the average result is reported.

The researchers for this project compared the performance of their model with several baselines. They compare their model to 5 different methods, the first is “SVM”, which uses a bag-of-words approach to predict emojis (Wu et al., 2018). In a bag-of-words model, text is represented as a bag of its words, completely disregarding grammar and word order when predicting its solution. The second method is “CNN” and uses a convolutional neural network as the word encoder only and uses the results to predict emojis jointly (Wu et al., 2018). For this method tweets are not being analyzed at the character level and therefore no character information is captured, only the words are analyzed. The next method is the “LSTM” method. Similar to the CNN method above it analyzes just words and not characters. However, it uses Long short-term memory to analyze tweets and predict the resulting emoji. The following method is the “LL” method. It is a hierarchical model that utilizes long short-term memory in both character and word encoders to predict emojis. This means the “LL” model should be able to more accurately predict emojis compared to the LSTM model because it analyzes both words and characters. The last method compared against is the “CLC” model. This model is a variant of the researchers model just without attention mechanism at the end. Let’s recall that attention mechanism is used to highlight and select important contexts and separate them from the unimportant contexts. Finally, CLCA is the model proposed by the researchers that is being compared against the others mentioned above. In addition to comparing their model’s performance against these baselines, the researchers also compare their model against humans by inviting a group of volunteers to the solve the same problem (Wu et al., 2018).

**Results**

The researchers compared the 5 models listed above to their own model (CLCA), and the results are as follows. Neural methods consistently outperformed SVM in predicting emojis (Wu et al., 2018). Since the SVM method does not analyze at the character-level it fails to capture as much information as other methods that do analyze at the character-level. Furthermore, due to its bag-of-words approach it misses out on important contextual information that can be extremely helpful in predicting emojis.

The hierarchical models such as CLCA and LL outperformed the flatten models like CNN and LSTM (Wu et al., 2018). Due to their inability to capture character specific information these models miss out on important data, information and hints that would be helpful for predicting emojis. Since CLCA captures information at the word and character-level it has more contextual information to base its prediction off of, leading to more accurate predictions.

The researchers proposed model (CLCA), consistently outperformed the LL model. This demonstrated that using a combination of convolutional neural networks and long short-term memory can yield better performance than using only long short-term memory (Wu et al., 2018). The researchers hypothesize that it outperforms LSTM because when you combine CNN and LSTM you can capture both local and long-range contexts, which is evidently extremely useful in mining the relationships between plain-text and emojis (Wu et al., 2018). Combining both of these into one model allows the model to gather more data and context from the tweets, leading to better predictions. The CLCA model also applies attention mechanism, while LSTM does not, resulting in the CLCA model focusing more on the important words in tweets that are more useful in predicting the emojis evoked by a tweet (Wu et al., 2018).

Automatic methods were shown to outperform humans if the training data was sufficient. However, researchers suspect it might be because an automatic neural model can predict emojis more objectively than a human can. It is cheap for researchers to feed a model a very large amount of detailed data to analyze and base its predictions off of, while humans do not have this benefit.

There were several prominent findings after gathering the results from all the tests. The researchers’ model (CLCA) outperformed other baselines in most emoji categories and achieved great performance when predicting frequently used emojis such as a heart or the laughing emoji (Wu et al., 2018). The researchers believe their models success when predicting frequently used emojis could be due to the sufficient amount of training data that was supplied containing these A picture containing writing implement, stationary

Description automatically generatedemojis. Certain infrequent emojis such as the big red exclamation mark and the party popper achieved satisfactory performance. The researchers believe this is due to the specific use of these emojis. They are easier for the model to understand as they are commonly associated with certain emotions and contexts and are used to express specific meanings (Wu et al., 2018). The researchers found some interesting results in predicting sad emojis. Sad emojis such as the crying face were harder for the model to predict. This could be due to the fact these emojis are used in a wide variety of contexts, can often be used completely out of context, and are often implied in the context of the tweet, making it particularly difficult for the model to infer off a plain-text message (Wu et al., 2018).

**Conclusion**

A hierarchical neural model with attention mechanism stood tall against all baselines and achieved the greatest success in determining emoji prediction in tweets. The model contains three major modules, each with their own functionality. The character encoder uses a convolutional neural network to learn and understand hidden representations of words. The word encoder is designed to analyze the words of a tweet and learn sentence representations using a combination of convolutional neural nets and long short-term memory. Lastly, an emoji classifier is used to predict the emojis for different tweets. Attention Mechanism is incorporated into the model to analyze contexts and determine important from unimportant words to build high quality sentence representations. Combined together these modules create an intelligent neural network that is capable of parsing tweets and producing an emoji that it deems most accurately fits the tweet. The model examines individual characters, words, word sequences and long and short-term context to create hidden representations that it uses to further produce more intelligent predictions. The model exceeded expectations and was able to successfully predict frequently used emojis a majority of the time.

**References**

Wu C, Wu F, Wu S, Huang Y, Xie X. Tweet Emoji Prediction Using Hierarchical Model with Attention. October 2018. doi: 10.1145/3267305.3274181

Matsumoto, K., Yoshida, M., & Kita, K. (2018). Classification of Emoji Categories from Tweet Based on Deep Neural Networks. *Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval - NLPIR 2018*. doi: 10.1145/3278293.3278306

Tomihira, Toshiki, et al. “What Does Your Tweet Emotion Mean? Neural Emoji Prediction for Sentiment Analysis.” *Proceedings of the 20th International Conference on Information Integration and Web-Based Applications & Services - iiWAS2018*, 2018, doi:10.1145/3282373.3282406

Chen, J. (2019, October 13). Neural Network Definition. Retrieved from https://www.investopedia.com/terms/n/neuralnetwork.asp.

Understanding LSTM Networks. (2015, August 27). Retrieved from https://colah.github.io/posts/2015-08-Understanding-LSTMs/.