Emotion Identification in FIFA World Cup Tweets using Convolutional Neural Network

Conference Paper · November 2015

DOI: 10.1109/INNOVATIONS.2015.7381514

4 authors:

Dario Stojanovski
Ludwig-Maximilians-University of Munich

Gjorgji Strezoski
University of Amsterdam

Gjorgji Madjarov
Ss. Cyril and Methodius University in Skopje

Ivica Dimitrovski
Ss. Cyril and Methodius University in Skopje

Emotion Identification in FIFA World Cup Tweets using Convolutional Neural Network

Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, Ivica Dimitrovski
Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Rugjer Boshkovikj 16 1000 Skopje, Republic of Macedonia
Email: stojanovski.dario@gmail.com, strezoski.g@gmail.com,
gjorgji.madjarov@finki.ukim.mk, ivica.dimitrovski@finki.ukim.mk

***Abstract*—Twitter has gained increasing popularity over the recent years with users generating an enormous amount of data on a variety of topics every day. Many of these posts contain real-time updates and opinions on ongoing sports games. In this paper, we present a convolutional neural network architecture for emotion identification in Twitter messages related to sporting events. The network leverages pre-trained word embeddings obtained by unsupervised learning on large text corpora. Training of the network is performed on automatically annotated tweets with 7 emotions where messages are labeled based on the presence of emotion-related hashtags on which our approach achieves 55.77% accuracy. The model is applied on Twitter messages for emotion identification during sports events on the 2014 FIFA World Cup. We also present the results of our analysis on three games that had significant impact on Twitter users.**

I. Introduction

Social media have gained increasing popularity over the recent years with the number of users on social networks and microblogging platforms growing rapidly. As of 2015, Twitter has over 300 million monthly active users and generates over 500 million messages per day[1]. Twitter messages are also known as tweets and have 140 character limitation. Tweets contain informal language, users use a lot of abbreviations, URLs, emoticons and Twitter specific symbols, such as hashtags and targets (user mentions).

Global sporting events, such as the FIFA World Cup or the Olympics, attract the attention of millions all around the world. The FIFA World Cup is arguably, the most globally followed event. During the latest 2014 FIFA World Cup, Twitter witnessed a staggering 672 million tweets related to the competition. The sheer amount of data generated in real-time along with the informal language, pose new challenges to an effective and efficient analysis of the content posted on Twitter. Twitter enables access to publicly available messages in real-time through the use of the Twitter Streaming API, opening up possibilities for many use cases. Chakrabarti et al. [1] and Nichols et al. [2] have explored Twitter activity in relation to sporting events in order to create a summary of sports games, while Corney et al. [3] developed a method for identifying events from Twitter streams and for associating messages related to sporting events to different teams' fans.

Sports games are arguably events that have tremendous effect on viewers and evoke a variety of emotional reactions. As users are inclined to share their feelings and opinions on social networks, popular sports games are accompanied with a number of Twitter posts where users express their views on the events of the game. Emotion identification and sentiment analysis have recently spiked the interest of both, academia and industry with the exponential grow of social media. Detecting users' reaction towards certain products and services can provide valuable insight for companies offering them. Additionally, it can be used to get information of the public opinion against different topics and events.

Emotion identification in sports tweets on the other hand can be utilized for different applications. There have been violent incidents between fans on many occasions during sports games. An emotion identification system can detect if a game has caused a lot of anger or other negative emotions amongst the fans, so the official organizers can be warned to take extra security measures after and during the match. Additionally, results from such system can be used to identify future potentially critical matches in terms of security.

In the work presented in this paper, we showcase a deep learning system for emotion identification in Twitter messages from the 2014 FIFA World Cup. For this purpose, we use a convolutional neural network with multiple filters and varying window sizes which has not been adequately studied for the task of emotion detection. The approach is founded on the work of Kim [4] that reported state-of-the-art results in 4 out of 7 sentence classification tasks. We leverage pre-trained word embeddings, obtained by unsupervised learning on a large set of Twitter messages. The network is trained on automatically labeled tweets in respect to 7 emotions. We trained a model on a set with a balanced distribution of emotions to study emotional reactions on Twitter during games from the 2014 FIFA World Cup. Results are presented for three games: Belgium vs. USA, Brazil vs. Germany and Brazil vs. Netherlands.

The rest of the paper is organized as follows. Section 2 outlines existing approaches to analysis of Twitter data related to sporting events and emotion identification. In Section 3, we give an overview of the architecture that we employed for the task of detecting emotion in tweets. In Section 4, we present the dataset used for training and elaborate on the achieved performance with our approach. We apply our model and study the observations on messages collected during the 2014 FIFA World Cup in Section 5. Finally, we conclude our work and give future directions in Section 6.

[1]https://about.twitter.com/company

II. Related Work

There has been a lot of work done in the field of emotion identification. Current approaches mainly are based on unigrams, bigrams, Part-of-Speech tags and other hand-crafted features and machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes and maximum entropy.

Appropriate labeling of tweets with corresponding emotions still poses a challenge in the field. Manual annotation provides for best quality of the training and test sets. However, it is expensive and labor-intensive and it can be too domain specific. In addition, the size of the corpus is likely to be significantly smaller. Roberts et al. [5] used a manually annotated set of tweets with 7 labels for emotion. Their approach for classification is consisted of hand-crafted features such as unigrams, bigrams, indicators of exclamation and question marks, WordNet hypernyms and several other features and a SVM classifier.

Balabantaray et al. [6] also used manually labeled tweets and developed a system for emotion classification using hand-crafted features such as Part-of-Speech tags, unigrams, bigrams and others and the Word-net Affect emotion lexicon. The work presents an extensive analysis of the effect different combinations of features have on performance. Ghazi et al. [7] on the other hand, consider the problem of emotion recognition as a hierarchical problem. In this work, a classifier was build that arranges neutrality, polarity and emotions hierarchically.

Purver et al. [8] and Wang et al. [9] on the other hand, use distant supervision for automatic emotion annotation. By using well-known indicators of emotional content they were able to create noisily labeled datasets. Tweets were annotated by the presence of emotion-related hashtags. Wang et al. [9] applied additional heuristics to improve on the quality of the acquired dataset. They only considered messages where the emotion-related hashtag is at the end of the tweet, removed messages containing URLs and quotations and discarded tweets with more than 3 hashtags, non-English messages and retweets. Furthermore, the quality of the dataset was evaluated by randomly sampling a small number of tweets for manual inspection by two annotators. They received 95.08% precision on the development and 93.16% on the test set, where a tweet was manually checked whether the assigned label is relevant to the conveyed emotion. The dataset is publicly available and we build our work on portion of the data.

Sintsova et al. [10] used the Amazon Mechanical Turk (AMT) to build a human-based lexicon. Using the annotators' emotionally labeled tweets, they constructed a linguistic resource for emotion classification. Their approach is able to capture up to 20 distinct fine-grained emotions. Yu et al. [11] used tweets to examine US soccer fans' emotional response to sports games during the 2014 FIFA World Cup. They have conducted a detailed analysis of the expressed emotions during several games and causes for such user behavior.

Unlike the previously depicted approaches that use extensive feature engineering which can both, be time-consuming and produce over-specified and incomplete features, our approach is based on a deep learning technique. Deep learning approaches handle the feature extraction task automatically, potentially providing for more robust and adaptable models. Most of the work focuses on utilizing convolutional neural networks. Collobert et al. [12] proposed a unified neural network architecture that can be applied to various Natural Language Processing (NLP) tasks including sentiment analysis. Santos et al. [13] proposed a CNN for identifying sentiment analysis by exploiting character-level, word-level and sentence-level information. Kim [4] on the other hand, proposed an approach for sentence classification including sentiment analysis using a CNN with multiple filters and feature maps. This work also showed that continuously updating pre-trained word embeddings provides for better performance. In our work, we build on the aforementioned deep learning techniques. However, these approaches have been used for sentiment analysis, which is limited to classifying tweets with three labels, positive, negative and neutral. In this paper, we use convolutional neural network for a finer grained classification into 7 emotions.

III. System Architecture

The approach we used for emotion identification in this work is a convolutional neural network architecture which is depicted in detail in Fig 1. The model is consisted of a simple network with one convolutional layer and a softmax output layer. Each token in a tweet is represented by a word embedding or word representation generated by a neural language model [14], [15]. These features are fed to the convolutional layer of the network. The proposed system is not dependent on hand-crafted features and manually created lexicons. Consequently, the approach is more robust than traditional NLP techniques and more adaptable when applied to different tasks and domains.

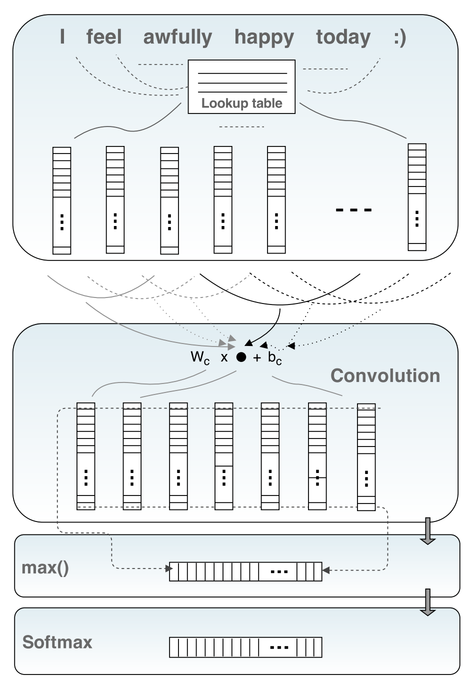

Fig. 1: Convolutional neural network architecture.

A. Pre-processing

Pre-processing is often applied to Twitter messages because of the language irregularities that are present. In order to clean noise from the tweets, we applied several pre-processing steps. We replaced each occurrence of a user mention with a generic token and lowercased all words. Additionally, we removed all HTML entities and punctuation except for exclamation and question marks and stripped hashtag symbols, but we kept emoticons as one of the strongest indicators of sentiment and emotions.

Moreover, all elongated words were normalized. Elongated words are common in social media posts when users want to emphasize some words. Although the number of repeated characters may differ (haaapy, haaaapy), users that posted such tweets most likely tried to convey the same emotion. As we do not want to differentiate between such words, we limited the number of repeated characters to three.

B. Pre-trained word embeddings

One of the most challenging parts of NLP is generating a suitable representation for the text that is being processed. Traditional NLP approaches utilize one-hot vectors which don't scale well to big corpora and are oblivious to the actual meaning of the word. As a result, we leverage word representations or embeddings, learned on large corpus of unlabeled textual data using neural language models. These embeddings capture syntactic and semantic regularities of words and have recently been used in many tasks in NLP.

The system works by first, constructing a lookup table, where each word is mapped to an appropriate feature vector or word embedding. One way of building the lookup table is by randomly initializing the word vectors. However, previous work in sentiment analysis [4], [13] showcased that initializing the word embeddings with pre-trained ones provides for better performance as opposed to random initialization. In this work, we do not do the pre-training ourselves as there are several already available word representations. One of the most popular are *word2vec*[2] [14] which are trained on Google News dataset with 100 billion tokens. Also, there are the Sentiment Specific Word Embeddings (SSWE) [16] and the global vectors for word representation or GloVe[3] word embeddings. There are several available pre-trained GloVe word embeddings which differ in the data used for learning. In order to train different word vectors they have used Wikipedia dumps, data obtained from a crawl of the Web and tweets.

However, since we are dealing with Twitter messages, which usually contain a lot of informal language, slang and abbreviations, using word embeddings trained on corpus where more formal language is used may not be suitable. Using word vectors trained on Twitter data is more fitting in our task as we assume there will be less missing tokens in the lookup table and the representations will be more meaningful in this context. Therefore, we leverage 200 dimensional GloVe embeddings [15] trained on 2 billion tweets (20 billion tokens). The SSWE are trained on distant supervised tweets, but with considerably less data (10 million tweets) and with a dimensionality of 50. For words that are not present in the vocabulary of word vectors, we use random initialization. Kim [4] suggests to use a range of $[-a, a]$ where a is set so that random initialized words have the same variance as the pre-trained ones. In our case, we set a to 0.25.

However, since the training is done in an unsupervised manner, there is no sentiment or emotion regularities encoded in the embeddings. As a result, words such as "bad" and "good", that likely appeared in similar context in the corpus are neighboring words based on cosine similarity. In contrast to the approach used in [16], we use available word embeddings and by back-propagation, during network training, update them in order to adapt to the specific task at hand. The intuition behind this approach is that by back-propagating the classification errors, emotion regularities are encoded into the word representations. In the work of Kim [4] it was showcased that non-static word vectors capture sentiment regularities in words. Upon finishing network training, "good" was no longer one of the most similar words to "bad" and vice versa. Additionally, this approach enables for building a more meaningful representation for words that are not present in the lookup table and for which random initialization is used.

C. Convolutional Neural Network

In this work, we utilize a convolutional neural network for classification of tweets into 7 emotion classes. Our approach is based on the work of Kim [4] which was used for different sentence classification tasks including sentiment analysis. CNNs with pooling operation deal naturally with variable length sentences and also take into account the ordering of the words and the context each word appears in. For simplicity, we consider that each tweet represent one sentence.

The network is trained by supplying tweets to the convolutional layer. Tweets can vary in length, but dealing with variable sized texts is inherently built into the network by employing a max-over-time pooling scheme. Each word or token of an input tweet, with the appropriate padding at the beginning and end of it, is mapped to an appropriate word representation. Padding length is defined as $h/2$ where h is the window size of the filter. Words are mapped from a lookup table $L \in R^{k \times |V|}$, where k is the dimension of the word vectors and V is a vocabulary of the words in the lookup table. Each word or token is projected to a vector $w_i \in R^k$. After the mapping, a tweet is represented as a concatenation of the word embeddings

$$x = \{w_1, w_2, \dots w_n\}. \tag{1}$$

The convolution operation is then applied to the mapped word embeddings and max-over-time pooling in order to get a fixed sized vector. In this step, we apply multiple filters with varying windows sizes h. We use rectified linear units in the convolutional layer and windows of size 3, 4 and 5. As tweets are short texts with limited character count, having such window sizes is adequate and using larger ones would not be beneficial. Filters are applied to every possible window of words in the tweet and a feature map is produced as a result. For each of the filters, a weight matrix W_c and a bias term b_c are learned. The weight matrix is used to extract local features

[2]https://code.google.com/p/word2vec/
[3]http://nlp.stanford.edu/projects/glove/

around each word window. The convolution operation can be formally expressed as

$$x'_i = f(W_c \cdot x_{i:i+h-1} + b_c), \quad (2)$$

where $f(\cdot)$ is the activation function and $x_{i:i+h-1}$ is the concatenation of word vectors from position i to position $i + h - 1$. The generated feature map x' is then passed through a max-over-time pooling layer

$$x' = max\{x'_1, x'_2 \dots x'_{n-h+1}\}. \quad (3)$$

which outputs a fixed sized vector where the size is a hyper-parameter to be determined by the user. In our case, we set the size of this vector to 100 and this hyper-parameter corresponds to the number of hidden units in the convolutional layer. By doing so, we extract the most important features for each feature map.

The output of the pooling operation for each of the multiple filters with varying window sizes is concatenated. Predictions are generated using a softmax regression classifier. The concatenated features from the max-over-time pooling layer are passed to a fully connected softmax layer whose output is the probability distribution over the labels.

Deep neural networks suffer from overfitting due to the high number of parameters that need to be learned. In order to counteract this issue, we use dropout regularization which essentially randomly drops a portion of hidden units (sets to zero) during training. As a result, the network prevents co-adaption between the hidden units. The proportion of units to be dropped is hyper-parameter to be determined by the user. The network is trained using stochastic gradient descent over shuffled mini-batches.

IV. Experiments

A. Dataset

In order to train our model, we utilize an already available annotated set provided in the work of Wang et al. [9]. They have based their work on existing psychology work [17] where human emotions are organized in a hierarchy. The first layer contains 6 basic emotions (*love*, *joy*, *surprise*, *anger*, *sadness* and *fear*) which expand to 25 subcategories. They have added another emotion, *thankfulness*, which is not included in [17]. Out of these categories they have extrapolated a set of keywords and their lexical variants to represent a single category of human emotions. They queried the Twitter API for tweets containing any of the keywords in the form of a hashtag.

However, due to Twitter privacy policy, only the IDs of the tweets were available for download. Using the Twitter API, we collected the messages, but because of changed privacy settings or deletion, a significant portion of the messages was not available. Out of 1991184 for training, 247798 for development and 250000 for test, we were able to retrieve 1347959, 168003 and 169114, respectively. Nonetheless, this is still a representative dataset. Moreover, the distribution of emotions in the tweets was similar to that of the original set. We applied the same heuristics that are pertaining to the removal of the hashtags indicative of the emotion from the Twitter message.

B. Experimental setup and results

We reused several hyper-parameters which were used in the work of Kim [4], mini-batch size of 50, l_2 constraint of 3, rectified linear units for the convolutional layer and filter windows of 3, 4 and 5. We set the learning rate to 0.02 and the dropout parameter to 0.7. These parameters, along with the decision to use rectified linear units over hyperbolic tangent was done by doing a grid search using the 1000 samples training set.

Due to technical limitations, we did not utilized the full capacity of the dataset. We tested our model with 2000 Twitter messages while for the development set we used 1000 messages. Both sets were generated by randomly sampling from the retrieved tweets. We trained the model with 1000 and 10000 training samples, in order to observe the gains from a larger training set. Employing the above mentioned parameters we were able to achieve 50.12% with 1000 training samples and 55.77% with 10000. Wang et al. [9] reported 43.41% and 52.92% respectively. Our model achieves higher accuracy in both cases on our reduced set.

For the three most popular emotions, *joy* (28%), *sadness* (24%) and *anger* (22%), we observe the highest F-measure of 64.95%, 55.48% and 58.71% respectively. Both precision and recall are relatively high in comparison to the other emotions. Precision and recall for *joy* are 59.59% and 71.38%, 51.06% and 60.74% for *sadness* and 59.28% and 58.16% for *anger*. For the less popular ones, *love* (12%), *fear* (5%), *thankfulness* (5%) precision is relatively high, 44.68%, 52.78% and 71.01% respectively, but recall is significantly lower compared with the top 3 emotions, 34.29%, 16.52% and 41.89% respectively. The F-measures for each of these emotions are 38.8%, 25.16% and 52.69% accordingly. The imbalance of class distribution in the dataset, leads to a classifier that will rarely classify a sample with uncommon labels. Since *surprise* accounts for only 1% of the training data, on the randomly sampled test set in our work, the classifier did not classify correctly any test example with *surprise*. For future work, we would like to explore the full capacity of the dataset and compare with the results reported by Wang et al. [9]. Additionally, further examination is required to improve on the results for rare emotions in the training set, such as *surprise*.

V. FIFA World Cup 2014 Emotion Analysis

The dataset containing Twitter messages from the 2014 FIFA World Cup was collected using the Twitter Streaming API. We retrieved English tweets only for the duration of a game. The API was queried for tweets containing the official game hashtag (i.e. #BRAGER) or general World Cup hashtags (i.e. #WorldCup2014). In the following section, we show the results of emotion identification on three games that have attracted a significant amount of attention on Twitter. As Brazil was the host of the competition, we chose their game in the semi-finals and for the third place. Additionally, the game between USA and Belgium was chosen because of the fact that the teams scored three goals in overtime, potentially causing a variety of emotions amongst the viewers.

TABLE I: Sample tweets from the World Cup set with the predicted emotion labels

Sadness	*That Robben booking was ridiculous. Another example of home field advantage for Brazil #BRAvsNED*
Joy	*Looks like Brazil will lose horribly again. NICE! #WorldCup2014*
Fear	*Please don't let this be another massacre. #BRAvsNED*

As mentioned before, the original training set has an imbalanced distribution of emotion labels. As a result, the classifier is bias towards emotions that are more present in the training set. In our experiments, we observed that a insignificant amount of the tweets related to the World Cup was labeled with *surprise* and *fear*. In order to overcome this problem, we train our model on a set with a balanced distribution of emotions. The model was trained on 10000 samples with selective sampling in order to get an even distribution. Table I shows sample tweets that were classified using our model. The model is being applied to the sports tweets in the same way as it was applied in the experiments explained above.

The following figures depict the distribution of emotions during the observed World Cup games. The time on the x-axis is in GMT. One could expect that the Twitter activity will experience a sudden increase at and immediately after important events during the game. From the figures we observe that scored goals are understandably the biggest actuators of emotions. We track the global sentiment towards the events of a game instead of reactions in separate regions or countries. Having such approach, we expect a mixture of emotions. However, a bias towards certain teams can be expected, as we only collect English tweets. This is apparent on Fig 2 where at the beginning of the game, *love* dominates over other emotions and *joy* is very close to *sadness*. We assume that a significant portion of English tweets will be originating from USA and as a result of early enthusiasm, users were more likely inclined to post positive tweets.

Twitter activity related to the game between Belgium and USA which ended as a draw in regular time, was at its peak in the additional time because of the three scored goals. The first spike in Twitter activity happened immediately after the end of regular time which is a proof that games that go in overtime or potentially are decided by a penalty shootout evoke reaction amongst the users. Belgium scored the first two goals in the overtime. The spike representing the first goal scored in the 93rd minute of the game (21:56 GMT) is somewhat smaller than the one representing the second goal, scored in the 105th minute of the game (22:08 GMT), but both spikes have an almost equal distribution of love and sadness. However, the third goal (107th minute - 22:10 GMT), the one scored by the USA team, caused a positive reaction amongst the viewers, a sentiment which continued to be present even though USA lost the quarter-final.

The bias towards a team, which was assumed for the Belgium vs. USA game, is likely to be less present for the remaining two games. This enables to get a better overview of the global sentiment, though we again assume that Brazil might be slightly more favored in respect to the Netherlands and Germany due to traditional fans' inclination towards host

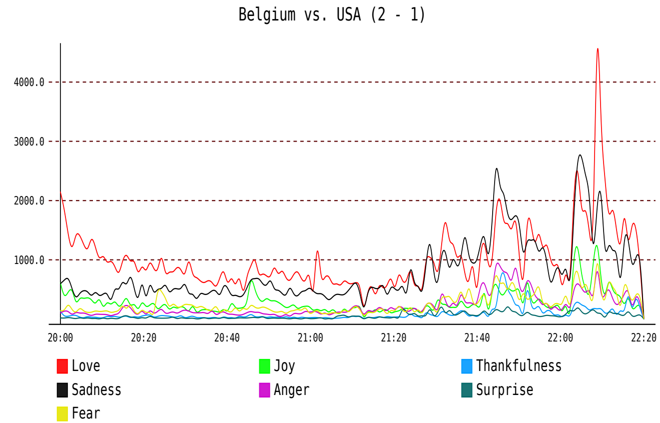

Fig. 2: Quarter-final: Belgium vs. USA

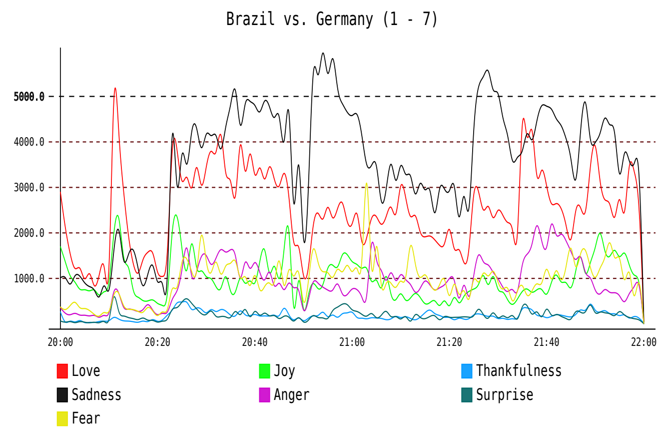

Fig. 3: Semi-final: Brazil vs. Germany

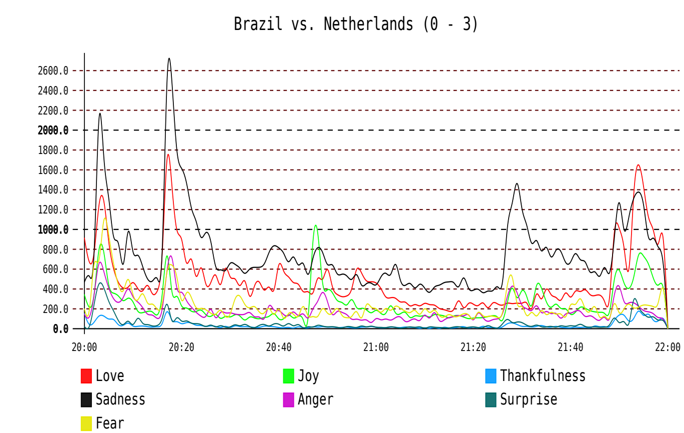

Fig. 4: 3rd place: Brazil vs. Netherlands

teams in World Cups.

The second game is evidently, one that has attracted an abundance of attention on Twitter. For the duration of the game we collected over 1.1 million tweets. Germany achieved a high 7-1 victory over hosts Brazil. As a result, we witness an almost constant high data stream with a variety of emotions. Having such a high number of scored goals and in addition to the fact that Brazil conceded 4 goals in just 6 minutes, it not only provoked regular soccer fans to post updates for the game, but those that rarely tweet about sports too. The first goal scored in the 11th minute (20:11 GMT) by the German team resulted in a dominantly positive reaction amongst Twitter users. However, as soon as Germany scored the second in the 23rd minute (20:23 GMT) and later on the next three consecutive goals (24th minute - 20:24 GMT, 26th minute - 20:26 GMT, 29th minute - 20:29 GMT), users' sentiment has shifted dramatically. *Sadness* has become the dominant emotion with *anger* and *fear* also having a significant appearance. Throughout the game, negative emotions were generally more present which confirms our assumption that fans were more supportive of Brazil. We can also observe that the activity has lowered dramatically since the beginning of the second half since this was a period of the game with significantly less excitement. However, a dramatic increase of the number of posted tweets can again be observed after the sixth goal in the 69th minute (21:24 GMT). It is interesting to note that the last goal in the 90th minute (21:53 GMT) of the game which was scored by Brazil, did not produce a noticeable increase in activity in relation to the previous goals.

In the game between Brazil and Netherlands, the hosts received two early goals. As it can be discovered from Fig 4, the two spikes relate to the Netherlands goals in the 3rd (20:03 GMT) and 17th minute (20:17 GMT). After the defeat against Germany, fans were likely concerned of another high loss and *sadness* was the dominant expressed emotion. In the remainder of the game, the emotion distribution evened out, since there were not too many intriguing events apart from the last minute goal (91st minute - 21:54 GMT) which again raised the interest among Twitter users in the game.

VI. Conclusion

In this paper, we presented a convolutional neural network for emotion identification in Twitter messages and we applied it to monitoring emotions during sports events. The model was evaluated on a set of hashtag labeled tweets with 7 distinct emotions. Using the presented architecture, we achieved improvements over current state-of-the-art performance with the dataset on reduced training set. Our approach obtains an accuracy of 50.12% and 55.77% with a training set of 1000 and 10000 samples. We trained the model on a set with a balanced distribution of emotion labels and we applied it to tweets related and posted during games on the 2014 FIFA World Cup. We showcased our system on three games from the competition that attracted a lot of attention in the public and we examine the emotion response received during these games. For future work, we would like to explore the effect different word embeddings and a deeper network architecture would have on the performance. Additionally, we would like to evaluate our model on the whole set to fully compare to existing approaches.

VII. Acknowledgments

We would like to acknowledge the support of the European Commission through the project MAESTRA Learning from Massive, Incompletely annotated, and Structured Data (Grant number ICT-2013-612944).

References

[1] D. Chakrabarti and K. Punera, "Event summarization using tweets." *ICWSM*, vol. 11, pp. 66–73, 2011.

[2] J. Nichols, J. Mahmud, and C. Drews, "Summarizing sporting events using twitter," in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. ACM, 2012, pp. 189–198.

[3] D. Corney, C. Martin, and A. Göker, "Spot the ball: Detecting sports events on Twitter," in *Advances in Information Retrieval*. Springer, 2014, pp. 449–454.

[4] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1746–1751. [Online]. Available: http://www.aclweb.org/anthology/D14-1181

[5] K. Roberts, M. A. Roach, J. Johnson, J. Guthrie, and S. M. Harabagiu, "Empatweet: Annotating and detecting emotions on Twitter." in *LREC*, 2012, pp. 3806–3813.

[6] R. Balabantaray, M. Mohammad, and N. Sharma, "Multi-class twitter emotion classification: A new approach," *International Journal of Applied Information Systems*, vol. 4, no. 1, pp. 48–53, 2012.

[7] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Hierarchical approach to emotion recognition and classification in texts," in *Advances in Artificial Intelligence*. Springer, 2010, pp. 40–50.

[8] M. Purver and S. Battersby, "Experimenting with distant supervision for emotion classification," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2012, pp. 482–491.

[9] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing Twitter" big data" for automatic emotion identification," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 2012, pp. 587–592.

[10] V. Sintsova, C.-C. Musat, and P. Pu Faltings, "Fine-grained emotion recognition in olympic tweets based on human computation," in *4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, no. EPFL-CONF-197185, 2013.

[11] Y. Yu and X. Wang, "World cup 2014 in the Twitter world: A big data analysis of sentiments in us sports fans tweets," *Computers in Human Behavior*, vol. 48, pp. 392–400, 2015.

[12] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Nov. 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1953048.2078186

[13] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, 2014, pp. 69–78. [Online]. Available: http://aclweb.org/anthology/C14-1008

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[15] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*, vol. 12, pp. 1532–1543, 2014.

[16] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for Twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2014, pp. 1555–1565. [Online]. Available: http://aclweb.org/anthology/P14-1146

[17] P. Shaver, J. Schwartz, D. Kirson, and C. O'connor, "Emotion knowledge: further exploration of a prototype approach." *Journal of personality and social psychology*, vol. 52, no. 6, p. 1061, 1987.