Distant Supervision for Emotion Classification Task using emoji2emotion

Aisulu Rakhmetullina
Informatics Dept.
Garching, 85748
aisulu.rakhmetullina@tum.de

Dietrich Trautmann
Informatics Dept.
Garching, 85748
dietrich.trautmann@cs.tum.edu

Georg Groh
Informatics Dept.
Garching, 85748
grohg@in.tum.de

Technical University of Munich

Abstract

Increasing number of research in the area of distant supervision for emotion detection task requires a reliable mapping between noisy labels and emotion classes. We propose a method for an experimental creation of such a reliable mapping based on manually annotated data and quantitative relations between labels and classes on example of emoji-emotion pair in a form of emoji2emotion mapping.

1 Introduction

The japanese word *emoji* means "picture + character", and has no semantical connection to english *emotion* as you might have thought. However, emojis indeed very often carry the emotional state of the writer. That is why, no surprises that as a part of the digital text emojis were exploited in various NLP researches related to sentiment analysis or emotion classification.

In later works based on machine learning approaches, most of the time emojis are used as a noisy label for a distant supervision task. However, the matching between emoji and sentiment or emotion class is often done manually [WR16]. That approach implies subjectivity and could lead to mismatching. The goal of this work is to propose a method for experimental matching between emoji and classes that should be more reliable. To evaluate our method we apply it to emoji to emotion mapping. Since to our knowledge there is no such experimentally created mapping between them, we introduce the name for it - emoji2emotion.

There exist different emotion classification models, either discrete or dimensional. In this work we have chosen Plutchik's wheel of emotions [Plu91] that combine characteristics of both discrete and dimensional models. We use main 8 emotions out of it called Plutchik's eight (anger, anticipation, joy, trust, fear, surprise, sadness and disgust) that shown in Figure 1.
.

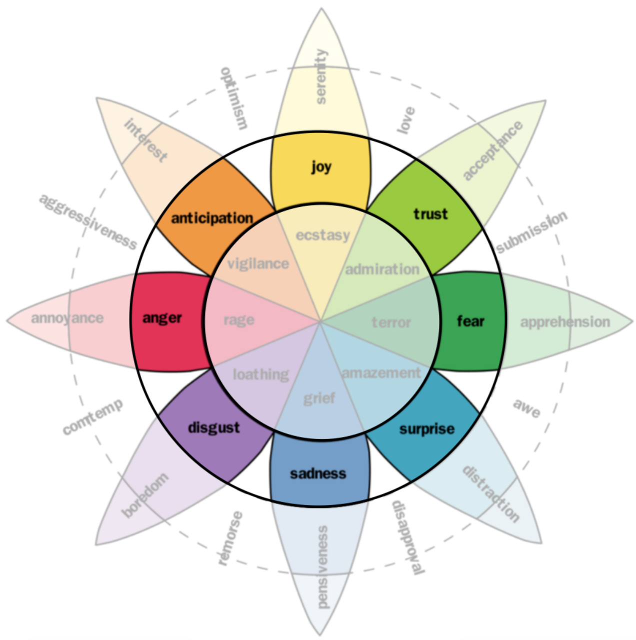

Figure 1: Plutchiks Wheel of Emotions with Plutchik's Eight highlighted [Plu91]

In: S. Wijeratne, E. Kiciman, H. Saggion, A. Sheth (eds.): Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018), Stanford, CA, USA, 25-JUN-2018, published at http://ceur-ws.org

2 Related Work

One of the first attempts to characterize emoji from its sentimental load perspective was a project called *Emoji Sentiment Ranking* - the first emoji sentiment lexicon (Figure 2). It was created by [NSSM15] and provides a map between 751 most frequently used emojis and sentiments. The valuable insights from it that we use: the majority of emojis are positive, especially the top popular ones; among tweets with emojis, the inter-annotator agreement tend to be higher.

In [ERA+16] authors release *emoji2vec*, set of pre-trained embeddings for all emojis in Unicode learned from emojisdescription taken from Unicode emoji standard. That is one of the examples of mapping emojis to another forms that are compatible to incorporate into machine learning tasks. And in general, representation learning and usage of pre-trained word embeddings is popular among natural language processing applications focused on social media.

In several works [BFMP13], [HBF+15], [JLL+14], [KZM14] emoticons were used to create a lexicon for a later use in a knowledge-based approach for sentiment analysis or emotion detection. The common thing between these works is a utilization of a high number of emoticon types, usually hundreds. Later works based on machine learning approach in contrast to works in the previous paragraph use emoticons and emojis as noisy labels for distant supervision tasks. Such works are [Rea05], [GBH09], [DTR10], [ZDWX12].

The recent paper [FMS+17] presents a project called *DeepMoji* and shows that diversification of noisy labels set for the distant supervision allows models to learn richer representations. They obtained state-of-the-art performance on the 8 benchmark datasets according to sentiment, emotion and sarcasm detection, which proves the effectiveness of the noisy level approach. Furthermore, their analyses confirm the assumption that diversity of emotional labels results in a performance improvement comparing to previous distant supervision methods.

3 Data Acquisition and Annotation

In this section, the process of manually annotated corpus creation is described in detail. First, an acquisition of data for further annotation is explained in three steps: emoji list creation, tweets crawling and tweets preprocessing. Second, the annotation course is presented in another three steps: tweets filtering, annotation and averaging of vectors, analysis of resulting corpus.

3.1 Data Acquisition

The first step of an emoji containing tweets corpus creation is to choose the list of emojis. To select most popular emojis in the twitter and in general in text online, we looked into the Emojitracker [etr13] project as well as into Emoji Sentiment Ranking table [NSSM15]. By application of threshold for each ranking (>100 000 000 for Emojitracker and >100 for Emoji Sentiment Ranking) 31 emojis from the first list and 50 from the second was picked. We selected emojis that were the intersection of both lists and additionally handpicked some emojis that were in the top lists but not in the intersection one. That is how the set of 43 tweets was created. After that, we calculated the distribution percentages for each source and found the average. That average percentage was used to create the same natural balance in our corpus.

The second step of corpus creation is a collection of data using the results of the previous step. In this paper, we use easily accessible Twitter data that we crawl with help of tweepy library. As a result, 84777 tweets containing emojis were crawled. Turned out, the vast majority of them (92.3%) contains only one type of emoji and most of the time its quantity is equal to 1 (average emoji count per tweet is 1.2). That is why we decided to focus on single emoji type tweets and after filtering out tweets with multiple emoji types or with emoji types that are not in our emoji list, the 74670 tweets left for training purposes.

The last step in the creation of corpus for labelling is a tweets preprocessing. On this stage, the raw tweets downloaded in the previous step are processed to the ready tweets . To do so, the number of emoji types in a tweet is counted, as well as the number of occurrences per each emoji type present. The replacement of tags, hashtags and URLs by the placeholders is done.

3.2 Data Annotation

In order to start annotation process, we picked 500 tweets with additional requirement in order to enhance the quality of tweets to be annotated. The requirements were:

- Tweet does not contain **URL-s, TAG-s**. That is a common practice in NLP that allows to exclude meaningless parts of the text.
- Tweet does not contain **HASHTAGS**. Even though [DTR10] found hashtags useful for automated sentiment analysis, in our case we decided to eliminate them in order to increase the readability for annotators.
- Tweet contains **from 5 to 15 words**. That way we have not so short and not so long tweets.

- Tweet contains **no more than 2 uppercase words**. That is also for readability reasons.
- Tweet contains **no unlemmatizeable words** (using spacys lemmatizer). Here it serves the data purity purposes as well as the understandability of the text for annotators.
- Tweet contains **no certain keywords** (the list was manually generated after revision of corpus) in order to eliminate spam tweets.

After choosing these 500 tweets, 3 annotators were asked to go through the procedure of tweets evaluation using an Web Interface created by us. For each tweet they could choose arbitrary number of emotions (including none) out of Plutchik's Eight and set the intensity value for it from 1 to 3. The resulting labels were averaged according to rule of where more than half of annotators should agree on label.

The resulting corpus consists of 500 labeled tweets, where labels are vectors of size 8 containing emotion intensities for 8 emotions. In the annotated set nearly half of tweets has only one emotion type, and the other half the combination of them (up to 4 out of 8 at once), resulting 1.1 emotion per tweet in average. The most prevalent emotion was *joy* that appeared in 57% of tweets to some extent. Other emotions were not that spread, and appeared in a quarter or less of tweets each.

In Table 1 the statistics of emotion and emotion combination distributions over the dataset is presented. For clarity emotions and emotion combinations are grouped into the positive, negative and neutral groups. Here the grouping was made under the assumption that emotions joy and trust are positive; sadness, anger, disgust, and fear are negative; and no emotions(neutral), anticipation and surprise are neutral. The combinations were determined by the prevailing sentiment, and in case of equality of positive and negative emotions, it was grouped into the neutral category.

The macro distribution shows that tweets with positive emotions are prevailing with about 60%, while the negative and neutral emotion tweets are only the rest. That is predicted that positive tweets will appear more (as stated in [NSSM15]), however, a distribution of classes is quite imbalanced.

4 Mapping emoji2emotion

Using annotated dataset from the previous step the percentage of emoji occurrences per emotion and vice verse was done. In order to create a mapping, we checked each possible pair of emotions and emojis for the following two conditions. First, emojis percentage

emotion [combination]	count in corpus	% in corpus
joy	206	41,20
joy/trust	43	8,60
joy/surpr	20	4,00
trust	12	2,40
joy/antcp	6	1,20
joy/surpr/trust	2	0,40
antcp/trust	1	0,20
joy/antcp/trust	1	0,20
joy/sadns/trust	1	0,20
joy/surpr/antcp	1	0,20
total:	293	58,60

emotion [combination]	count in corpus	% in corpus
no_emotion	94	18,80
surpr	8	1,60
joy/sadns	3	0,60
antcp	2	0,40
fear/joy	1	0,20
joy/surpr/antcp/trust	1	0,20
total:	109	21,80

emotion [combination]	count in corpus	% in corpus
sadns	49	9,80
anger/sadns	19	3,80
anger	14	2,80
sadns/surpr	2	0,40
fear/surpr	2	0,40
fear	2	0,40
disgt/sadns	2	0,40
anger/disgt	2	0,40
sadns/antcp	1	0,20
fear/sadns/surpr	1	0,20
fear/sadns	1	0,20
anger/sadns/trust	1	0,20
anger/sadns/surpr	1	0,20
anger/disgt/sadns	1	0,20
total:	98	19,60

Table 1: Distribution of positive, neutral and negative emotions across the resulting corpus

of appearing in the tweets subset of certain emotion should be at least equal to the median value of possible percentages. Second, an emotion should appear in certain emoji tweets at least half of the time. As a result, the following mapping was done as shown in Table 3.

emotion	***emojis***
anger	😒
joy	😂😍❤️😊💕😁☺️😄😘🙏
sadness	😭😩😢
surprise	😱

Table 2: Results of emoji2emotion mapping

To evaluate the quality of mapping, we use them as noisy labels in emotion annotation subtask of SemEval 2007 task 14 - Affective text [sem07]. That task aims to explore the connection between emotions and lexical semantics. Since the task is carried out in an unsupervised setting, only testing data is provided. It consists of 1000 short texts (news headlines) annotated according to 6 emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise) which are Ekman's Six, and their intensity. Due to the fact that 6 emotions of Plutchik's Eight compose Ekman's Six, this data could be compatible with ours. For that, we reduce the number of classes from 8 to 6 and labelled 74670 tweets from Data Acquisition step using emoji2emotion mapping to use as training data. We used coarse version of SemEval's test set as well as labelled our training set with binary vectors.

To train our model we turned the news headlines in test set as well as tweet texts in training set into word embeddings using the word2vec methodology and open source code of *emoji2vec*. Then we fed these word embeddings as well as noisy labels to 4 classifiers (SGD, Naive Bayes, Random Forest and k-NN) from the scikit library. Using the trained model we predicted emotion categories per headline for the 1000 test set mentioned before. The resulting precision, recall and f1 scores are presented in the Table 3. The bold values represent maximum values, while green values are those that outperform the SemEval's best scores.

That is evident that the training data has a imbalance towards certain emotion categories which we link to the number of emojis picked per emotion. That is why the results of training also translate that kind of bias. To avoid that bias we need a more balanced set, and for that, in turn, we need more balanced mapping. To achieve that, more training data will be needed in the next run of the experiment and we leave that for future development of the work.

	Prec.	**Rec.**	**F1**
	ANGER		
SGD	0,00	0,00	-
Nearest neighbors	**7,81**	**23,81**	**11,76**
Naive Bayes	0,00	0,00	-
Random Forest	0,00	0,00	-
	JOY		
SGD	**16,43**	41,59	**23,56**
Nearest neighbors	11,21	**86,73**	19,86
Naive Bayes	12,34	25,66	16,67
Random Forest	11,36	31,86	16,74
	SADNESS		
SGD	0,00	0,00	-
Nearest neighbors	4,76	1,92	2,74
Naive Bayes	0,00	0,00	-
Random Forest	**14,29**	**1,92**	**3,39**
	SURPRISE		
SGD	0,00	0,00	-
Nearest neighbors	**3,78**	**83,33**	**7,22**
Naive Bayes	3,36	59,52	6,37
Random Forest	0,00	0,00	-

Table 3: Results of applying emoji2emotion to task 14 of SemEval 2017 [sem07]

5 Findings and Contribution

We propose a method of experimental mapping between emoji and sentiment or emotion classes based on a special processing of manually annotated data. The processing includes the finding quantitate relation between emoji and emotion in form of cooccurrence percentage and further thresholding. To implement the method we annotated the corpus of 500 tweets containing emojis with help of 3 human judges. From the average annotation labels we constructed mapping as described above. Due to significant imbalance in emotions distribution across the dataset mapping was done only for 4 emotion categories and evaluated by exploiting as noisy labels for emotion detection task on those 4 emotions. The results on emotion detection task show that it is feasible to continue in that direction by increasing the size of the annotated corpus and further tuning the training parameters.

The resulting corpus of manually labeled emoji containing tweets is shared open source online (https://github.com/Aisulu/emoji2emotion) for the benefits of scientific society.

6 Challenges and Limitations

After the annotation process that was evident to us, that the labelling for 8 classes and 3 intensity levels for each of them require the high cognitive load from the annotators and in average takes 18 seconds per tweet. Even though we knew that the increase in the class numbers leads to the slower labelling [BKT+13], it was higher than we expected and lead to the decrease of the final corpus size. As a result, not all the emotions were presented in large enough size in the dataset which leads to the convergence of the classes to fewer classes.

7 Future Work

We aim to find less time-consuming form of annotation process for users to increase the size of the manually annotated corpus. After that we plan to repeat experimental procedures.

References

[BFMP13] Marina Boia, Boi Faltings, Claudiu-Cristian Musat, and Pearl Pu. A:) is worth a thousand words: How people attach sentiment to emoticons and words in tweets. In *Proceedings of the 2013 International Conference on Social Computing*, SOCIALCOM '13, pages 345–350, Washington, DC, USA, 2013. IEEE Computer Society.

[BKT+13] Michael Brooks, Katie Kuksenok, Megan K. Torkildson, Daniel Perry, John J. Robinson, Taylor J. Scott, Ona Anicello, Ariana Zukowski, Paul Harris, and Cecilia R. Aragon. Statistical affect detection in collaborative chat. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 317–328, New York, NY, USA, 2013. ACM.

[DTR10] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 241–249, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[ERA+16] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. emoji2vec: Learning emoji representations from their description. *CoRR*, abs/1609.08359, 2016.

[etr13] Emojitracker, 2013.

[FMS+17] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

[GBH09] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. 150, 01 2009.

[HBF+15] Alexander Hogenboom, Danella Bal, Flavius Frasincar, Malissa Bal, Franciska De Jong, and Uzay Kaymak. Exploiting emoticons in polarity classification of text. *J. Web Eng.*, 14(1-2):22–40, March 2015.

[JLL+14] Fei Jiang, Yiqun Liu, Huanbo Luan, Min Zhang, and Shaoping Ma. *Microblog Sentiment Analysis with Emoticon Space Model*, pages 76–87. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[KZM14] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M. Mohammad. Sentiment analysis of short informal texts. *J. Artif. Int. Res.*, 50(1):723–762, May 2014.

[NSSM15] Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. Sentiment of emojis. 2015.

[Plu91] R. Plutchik. *The Emotions*. University Press of America, 1991.

[Rea05] Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, ACLstudent '05, pages 43–48, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[sem07] Affective text. semeval task 14, 2007.

[WR16] I. D. Wood and S. Ruder. Emoji as emotion tags for tweets. *Proceedings of the Emotion and Sentiment Analysis Workshop LREC2016, Portoro, Slovenia*, pages 76–79, 2016.

[ZDWX12] Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. Moodlens: An emoticon-based sentiment analysis system for chinese tweets. 08 2012.