Bootstrapping Technique + Embeddings = Emotional Corpus Annotated Automatically

Lea Canales[1], Carlo Strapparava[2], Ester Boldrini[1], and Patricio Martínez-Barco[1]

[1] University of Alicante, Alicante, Spain
{lcanales,eboldrini,patricio}@dlsi.ua.es
[2] Fondazione Bruno Kessler, Trento, Italy
strappa@fbk.eu

Abstract. Detecting depression or personality traits, tutoring and student behaviour systems, or identifying cases of cyber-bulling are a few of the wide range of the applications applications, in which the automatic detection of emotion is crucial. This task can contribute to the benefit of business, society, politics or education. The main objective of our research is focused on the improvement of the supervised emotion detection systems developed so far, through the definition and implementation of a technique to annotate large scale English emotional corpora automatically and with high standards of reliability. Our proposal is based on a bootstrapping process made up two main steps: the creation of the seed using NRC Emotion Lexicon and its extension employing the distributional semantic similarity through *words embeddings*. The results obtained are promising and allow us to confirm the soundness of the bootstrapping technique combined with the *word embedding* to label emotional corpora automatically.

Keywords: Sentiment Analysis, Emotion Detection, Emotional Corpus, Bootstrapping, Word Embedding

1 Introduction

Emotion detection has been widely explored in neuroscience, psychology and behavior science, being an important element of human nature. In computer science, this task has also attracted the attention of many researchers, despite the challenges of dealing computationally with emotions such as the complexity of working exclusively with text as input, due to the lack of para-linguistic information like tone, emphasis and facial expressions.

Automatic detection of affective states in text has wide range of applications for business, society, politics or education. Detecting emotions is becoming more and more important due to the fact that it can bring substantial benefits for different sectors (detecting depression [5], identifying cases of cyber-bullying [10], tracking well-being [26], or contributing to improve the student motivation and performance [21]) that are demanding effective automatic detection systems for multiple purposes.

Many of the machine learning techniques for automatic detection of emotions are supervised, that is, systems first infer a function from a set of examples labeled with the correct emotion (this set of examples is called the training data or labelled corpus). Then the model is able to predict the emotion of new examples. Hence, the training data employed in supervised machine learning algorithms are crucial to build accurate emotion detection systems.

The creation of a labelled corpus is not trivial, since detecting emotion in text can be difficult even for humans. Much of works carried out so far have shown that the amount of agreement between annotations when associating emotion to instances is significantly lower compared to other tasks such as identifying part of speech or detecting named entities. This is due to the fact that manual annotations can be significantly influenced by clarity of instructions, difficulty of task, training of the annotators, and even by the annotation scheme [19]. For this reason, in this paper an innovative automatic technique is proposed to resolve the most important challenge in emotion detection task in text: the problems of the annotation of emotional data.

To do this, our proposal is to exploit a bootstrapping approach for automatic annotations with two main steps: 1) the creation of the seed where NRC Emotion Lexicon [20] is employed to annotate the sentences by its emotional words; and 2) the extension of the seed based on the distributional similarity calculated through *low-dimensional continuous representations* of instances and words (word vectors or *embeddings*). This technique will allow us the annotation of large amount of emotional data in any genre with efficiently and high standards of reliability.

The rest of the paper is organised as follows. Section 2 deals with the related works and a comparative analysis of our approach. In section 3, the proposed method is described in detail. Section 4 is aimed at showing the approaches proposed, the evaluation methodology, the results obtained and a discussion about these results. Finally, Section 5 details our conclusions and future works.

2 Related work

This section summarises the most relevant emotional corpora developed for emotion detection purposes, their features and how they have been developed. In addition, some of works where bootstrapping technique was applied for annotation and the weaknesses that motivate and justify the direction of our research are also analyzed.

Since there are hundreds of emotions that humans can perceive and express, much of the work in the community has been restricted to a reduced of emotions and valence categories.

According to research in psychology, there is a number of theories about how to represent emotions. Among these theories, some of them are focused on defining the set of the basic emotions [11, 25]. Although, there is not an universal consensus about which set of emotions are the most basic. Nevertheless, most of works in automatic detection of emotions in text has focused on the limited set of proposed basic emotions, since this allows reducing the cost in terms of time and money. Even though there also are approaches based on non-basic emotions.

Most of the emotional resources developed so far have been annotated manually, since, in this way, machine learning systems learn from human annotations. Among these resources, there are corpora labelled with the six basic emotions categories proporsed by Ekman such as: [1] annotated a sentence-level corpus of approximately 185 children stories with emotion categories; [2] annotated blog posts collected directly from Web with emotion categories and intensity; or [27] annotated news headlines with emotion categories and valence.

As mentioned previously, there are corpora labelled with other small set of emotions by manually annotation like: [22] corpus extracted 1,000 sentences from various stories; Emotiblog-corpus that consists of a collection of blog posts manually extracted from the Web and annotated with three annotation levels: document, sentence and element [4]; or EmoTweet-28 corpus that consists of a collection of tweets annotated with 28 emotion categories [15].

The common feature of these emotional corpora is that have been annotated manually, being a hard and time-consuming task where the obtaining an agreement between annotations is a challenge, due to the subjectivity of the task and the need to invest in many resources to annotate large scale emotional corpora.

Consequently, several emotional resources have recently been developed employing emotion word hashtags to create automatic emotional corpus on Twitter. [18] describe how they created a corpus from Twitter post (Twitter Emotional Corpus - TEC) using this technique. In literature, several works can be found with the use emotion word hashtags to create emotional corpora from Twitter [6, 28].

Thus, in research community, the interest of developing amounts of emotional corpora has increased because that would allow us to obtain better supervised machine learning systems. The use of emotion word hashtags as technique to label data is really simple and efficient in terms of time and money, but it can be applied on Twitter exclusively. For this reason, our objective is to develop a technique for large-scale annotation of emotional corpora automatically in any domain and with high standards of reliability.

Our proposal consists of a bootstrapping technique because it is a semi-supervised technique whose effectiveness has been demonstrated in a wide range of computational linguistic problems [29, 9] and more concretely for annotations task [14, 7]. Thus, our hypothesis is that the use of this technique for automatic annotation of emotions will provide us improvements on the emotion labelled task.

3 Bootstrapping process

This section describes the bootstrapping process developed for automatic annotation.

The process receives as input data a collection of unlabelled sentences/phrases and a set of emotions, concretely the Ekman‘s six basic emotions [11]. The objec-

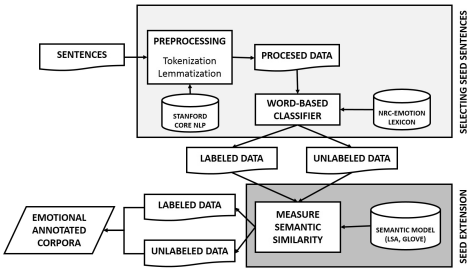

Fig. 1. Overall bootstrapping process.

tive of this task is to annotate unlabelled sentences with the emotions expressed in each sentence.

The overall bootstrapping process is described in Figure 1.

3.1 Dataset

The dataset employed to test our approaches is Aman corpus that contains sentence-level annotation of 4,000 sentences from blogs posts collected directly from Web. This resource was annotated manually with the six emotion categories proposed by Ekman and the emotion intensity (high, medium, or low).

The reasons to choose this corpus for testing the approach are: (i) it is manually annotated allowing us to compare automatic annotation to manual annotation; (ii) this corpus is relevant to emotion detection task, since it has been employed in many works to detect emotions; and (iii) it is possible to check the usability and effectiveness of our approach in Social Web domain, because this corpus contains sentences from blogs posts.

3.2 Selecting seed sentences

In this section, the process of creating the initial seed by exploring NRC Word-Emotion Association Lexicon (Emolex) [20] is presented.

Emolex is a lexicon of general domain consisting of 14,000 English unigrams (words) associated with the Plutchik‘s eight basic emotions [25], compiled by manual annotation. Our approach only employs the Ekman‘s basic emotions and for this reason the lexicon is reduced to 3,462 English unigrams.

In this approach, Emolex is applied to annotate each sentence of the Aman corpus which contains emotional words. Each sentence has an emotional vector associated with a value to each emotion ([anger, disgust, fear, joy, sadness, surprise]) initialised to zero (Figure 2). In Emolex, each word has also an emotional vector associated.

The process starts tokenising and lemmatising each sentence using Stanford Core NLP [16]. Then, each word of the sentence is looked up in Emolex. If a word

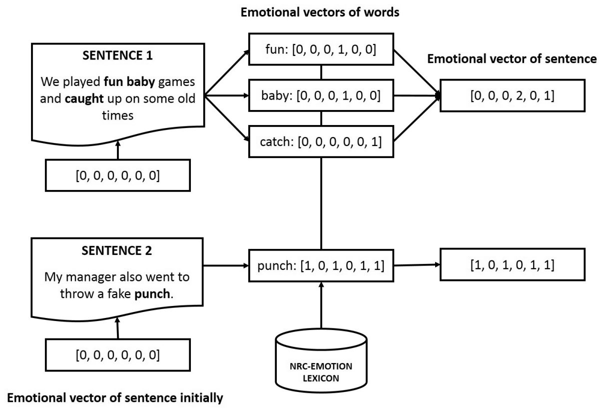

Fig. 2. Examples of the process of selecting seed sentences when a sentence is annotated.

of the sentence is in Emolex, its emotional values are added to the emotional vector of the sentence. Finally, the emotional vector of the sentence shows the emotions related to the sentence. The sentences are annotated with the emotion whose has the highest value (Figure 2 - Sentence 1). In this process, a sentence could have an emotional vector associated with several emotions in the same proportion. In this case, the process does not label any emotion because there is not a predominant emotion (Figure 2 - Sentence 2).

Figure 2 shows two examples of the creation of the seed. Sentence 1: *"We played fun baby games and caught up on some old time"*, whose emotional vector is initialised to zero, contains three emotional word: fun, baby and catch. The values of these three words are added and the sentence has finally associated this vector: [0, 0, 0, 2, 0, 1], this sentence has `JOY` emotion associated because this emotion has the highest value associated. Sentence 2: *"My manager also went to throw a fake punch."*, whose emotional vector is initialised to zero, contains one emotional word: punch. The sentence has finally associated this vector: [1, 0, 1, 0, 1, 1], hence this sentence has not associated any emotion.

Once the process is completed, there are non-annotated sentences because the sentences do not contain emotional words or do not contain a predominant emotion, and annotated sentences (seed sentences) with one of the emotions.

3.3 Seed extension

In this step, the purpose is to extend the seed sentences that have been obtained from the process explained in the previous subsection, with the help of a bootstrapping approach.

To achieve that, two Distributional Semantic Models (DSM) are adopted in this step. These models are based on the assumption that the meaning of a word can be inferred from its usage. Therefore, these models dynamically build semantic representations (high-dimensional semantic vector spaces) through a statistical analysis of the contexts in which words occur[1]. Finally, each word is represented with a real-valued vector called word vector or *word embedding*.

There are two main global families for learning word vectors: 1) global matrix factorization methods, and 2) local context windows methods. The methods based on local context windows poorly utilize the statistics of the corpus since they train on separate local context windows instead of on global co-occurrence counts and thus they are not as convenient as global matrix methods on word similarity task.

The extension of the seed of our proposal is based on estimating the similarity among non-annotated sentences and annotated sentences. For this reason, in this paper we test two models based on global matrix factorization methods: LSA and GloVe [23].

Both models are run on the lemmas of the British National Corpus (BNC)[2] that can be considered as a balanced resource since it includes texts from different genres and domains. Concretely, the LSA model employed is the one applied in [12] and GloVe model has been built by us.

The process of extension of the seed consists of measuring the similarity among non-annotated sentences and annotated sentences using the models listed and measuring the standard cosine similarity. When the similarity between a non-annotated sentence and an annotated sentence is higher than 80%, the non-annotated sentence is annotated with the emotions of the annotated one.

In this process, non-annotated sentences could be matched to two or more annotated sentences. The process selects the annotated sentence whose similarity with non-annotated one is higher and annotates it.

3.4 Training a supervised classifier

In the second step of the bootstrapping technique, the annotated and the non-annotated sentences from the previous step are exploited to train a supervised classifier. Concretely, a multi-classifier Support Vector Machines (SVM) with Sequential Minimal Optimization [24] is applied, representing the sentences as a vector of words weighted by their counts using Weka [13].

4 Evaluation

The objective of this research is to assess the viability of the use of bootstrapping technique to built emotional corpora. To achieve that, in this paper two evaluation, explained in Section 4.3, are carried out.

Furthermore, as mentioned previously, Emolex contains 3,462 words when it works with Ekman‘s emotions compared to the 14,000 words when it works with

[1] http://wordspace.collocations.de/doku.php/course:acl2010:start
[2] http://www.natcorp.ox.ac.uk/

Plutchik's emotions. Therefore, the improvement of Emolex with synonyms can be considered interesting to test if the creation of the seed improves. For this reason, three approaches have been evaluated employing different versions of Emolex (original, WN synonyms and Oxford synonyms). The extension process of Emolex is completely automatic and is explained in detail in the next sections.

4.1 Enriched approach by WordNet synonyms

One of the enriched approach employed consists of the extension of Emolex employing the synonyms of WordNet [17].

In this process, each word contained in Emolex was looked up in WordNet, the synonyms of its more frequent sense were obtained and were annotated with the emotions of the Emolex word. Figure 3 shows an example of the process. The word 'alarm' is contained in Emolex and has the emotions `FEAR` and `SURPRISE` associated. The process looks up 'alarm' in WordNet and obtains the synonyms of its more frequent sense: 'dismay' and 'consternation'. These synonyms are added to Emolex and annotated with the same emotions of 'alarm'.

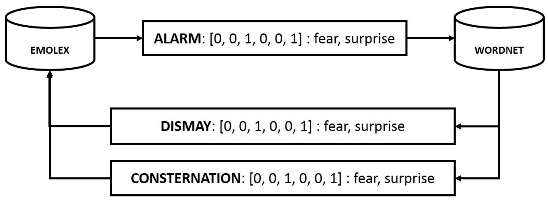

Fig. 3. Process of the extension of Emolex by WordNet synonyms.

After the process, Emolex has been extended with 4,029 words more, resulting a lexicon with 7,491 words.

The enriched approach by WordNet synonyms runs the same process than the original approach, but employing the new version of Emolex.

4.2 Enriched approach by Oxford synonyms

The enriched approach by Oxford synonyms was carried out with the aim of analysing the relevance of selecting a set of synonyms or other.

First, each word contained in Emolex was looked up in the Oxford American Writer Thesaurus [3] and all of the synonyms for all of its senses were collected. Then, each synonym of a word was associated with the emotions of the Emolex word and was added in Emolex. If a synonym was already in Emolex, their emotions associated will be the result of matching the emotional vector stored in Emolex and the new emotional vector.

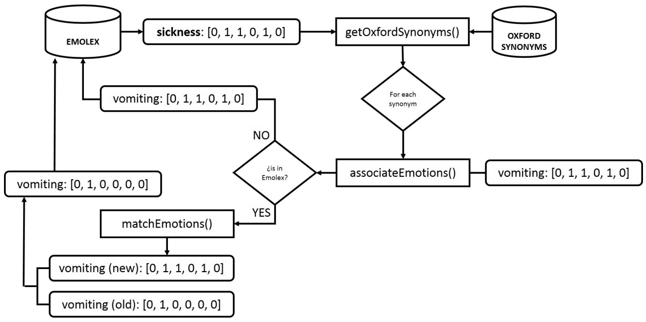

Fig. 4. Process of the extension of Emolex by Oxford synonyms.

Figure 4 shows an example of the process for the word 'sickness'. The first step is get their Oxford synonyms and for each synonym (in this example the synonym 'vomiting'): 1) associate the emotions of 'sickness', this is, DISGUST, FEAR and SADNESS; and 2) check if 'vomiting' is already in Emolex. If it is not, their emotions associated will be the same that 'sickness'. In another case, their emotional vector will contain the emotion in common between the vector saved in Emolex (old) and the new emotional vector (new). In this case, 'vomiting' will be associated with DISGUST emotion.

After the process, Emolex has been extended with 6,789 words more, resulting a lexicon with 10,251 words.

Once extended, the process of the enriched approach by Oxford synonyms is the same than the original approach, but employing the new version of Emolex.

4.3 Evaluation Methodology

As we mentioned, the evaluation methodology is divided into two steps. On the one hand, an emotional model is built from the corpus annotated automatically to evaluate the usability of this corpus. On the other hand, the quality of automatic annotations is assessed through the measure of agreement between the corpus developed with our approach (automatic annotation) and the gold standard of Aman corpus (manual annotation).

To evaluate the automatic emotion classification, the multi-classifier employed is performed with a 10-fold cross-validation on the corpus annotated automatically. Specifically, precision, recall and F1-score are calculated in each model.

Concerning our evaluation of agreement on Aman corpus, we employ the Cohen's kappa [8] to measure the inter-tagger agreement between automatic and manual annotations like the original work.

4.4 Results

The results obtained by each classifier in all of our approaches are shown in the tables below; Tables 1 and 2 detail results obtained with LSA and GloVe similarity models respectively. Moreover, Table 1 also shows the results obtained by the original Aman corpus, employing SMO classifier with the same features. Precision (P), recall (R) and F1-values (F1) valued are shown for each emotion employing the original approach and the enriched approaches.

Regarding the comparison between automatic and manual annotations, Cohen's kappa values obtained by each one of our approaches when they are compared to the gold standard of Aman corpus are shown in Table 3.

4.5 Discussion

As seen in the results section in the average F1-values of each approach, the results obtained are promising considering the state-of-art results (around 75% F1-value) and the fact that the features employed in the classifier are the sentences as a vector of words weighted by their counts. Thus, these results could be improved with selecting other features.

Regarding F1-values obtained by original and enriched approaches, the results show the improvements achieved by `ANGER`, `DISGUST`, `FEAR`, `JOY` and `SADNESS` emotions when the set of WN and Oxford synonyms are employed. These advances in the results are obtained by both DSMs: LSA and GloVe. Hence, the results confirm the benefit of extending Emolex with synonyms.

In terms of comparison between DSMs, the results demonstrate improvements when GloVe is employed, since this model obtains higher values for nearly all emotions and the global results with respect to LSA model. Hence, GloVe is validated as an effective and proper model to extend the seed in the bootstrapping process.

Table 1. Precision, Recall and F1-values obtained by the SMO Multi-Classifier on the Corpus Developed Applying LSA as Semantic Metric in the Extension of the seed.

	LSA model (Aman corpus)									Aman corpus		
	Original approach			WN approach			Oxford approach			Original		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Anger	0.198	0.137	0.162	0.444	0.348	**0.391**	0.338	0.330	0.334	0.538	0.274	0.363
Disgust	0.250	0.068	0.107	0.308	0.178	**0.225**	0.353	0.120	0.179	0.714	0.320	0.442
Fear	0.401	0.236	0.297	0.392	0.303	**0.342**	0.412	0.251	0.312	0.672	0.357	0.466
Joy	0.574	0.571	0.572	0.677	0.702	**0.689**	0.565	0.604	0.584	0.720	0.513	0.599
Sadness	0.247	0.107	0.149	0.467	0.269	0.341	0.591	0.462	**0.519**	0.577	0.260	0.359
Surprise	0.459	0.224	**0.301**	0.366	0.152	0.214	0.359	0.192	0.250	0.553	0.226	0.321
Neutral	0.706	0.846	**0.770**	0.559	0.676	0.612	0.551	0.668	0.604	0.798	0.955	0.869
Macro Avg.	0.595	0.633	**0.605**	0.571	0.586	0.573	0.525	0.533	0.523	0.753	0.774	0.745

Table 2. Precision, Recall and F1-values obtained by the SMO Multi-Classifier on the Corpus Developed Applying GloVe as Semantic Metric in the Extension of the seed.

	GloVe model (Aman corpus)								
	Original approach			WN approach			Oxford approach		
	P	R	F1	P	R	F1	P	R	F1
Anger	0.103	0.037	0.054	0.500	0.357	**0.417**	0.353	0.320	0.336
Disgust	0.000	0.000	0.000	0.194	0.085	**0.118**	0.273	0.075	**0.118**
Fear	0.527	0.298	0.381	0.475	0.320	**0.382**	0.341	0.162	0.220
Joy	0.822	0.565	0.670	0.708	0.678	**0.693**	0.594	0.541	0.566
Sadness	0.500	0.109	0.178	0.485	0.263	0.341	0.642	0.480	**0.549**
Surprise	0.787	0.349	**0.484**	0.438	0.097	0.159	0.478	0.180	0.262
Neutral	0.828	0.959	**0.889**	0.650	0.777	0.708	0.631	0.765	0.692
Macro Avg.	0.781	0.812	**0.782**	0.631	0.647	0.631	0.585	0.599	0.583

Table 3. Cohen's kappa values obtained by LSA and GloVe models (the Original Approach and the Enriched Approaches) in the Comparison of their Annotations to the Gold of Aman Corpus.

	Cohen's kappa values					
	LSA			GloVe		
	Original Appr.	WN Appr.	Oxford Appr.	Original Appr.	WN Appr.	Oxford Appr.
Anger	**0.9368**	0.9051	0.8882	**0.9470**	0.9133	0.9034
Disgust	0.9495	0.9417	**0.9537**	0.9534	0.9460	**0.9547**
Fear	0.9226	0.8919	**0.9323**	**0.9492**	0.9238	0.9417
Joy	**0.7719**	0.6041	0.7241	**0.8630**	0.6810	0.7814
Sadness	**0.9285**	0.9193	0.8033	**0.9425**	0.9223	0.8508
Surprise	0.9186	**0.9512**	0.9345	0.9504	**0.9574**	0.9470

Concerning agreement values, most of the best values (in bold) are higher than 80% showing the quality of the automatic annotation obtained with our approach. About the comparison of the two DSMs, the best results in LSA approach are improved by GloVe model. Thus, the benefits of GloVe are also endorsed by the agreement values.

5 Conclusion

The basis of our research is the need to develop a technique that allow us to tackle the annotation task of emotions and thus improving supervised learning techniques. Therefore, this paper is focused on exploiting an innovative bootstrapping approach to automatically annotate emotional corpora.

Afterwards the evaluation performed, we can confirm that our approach is appropriate and reliable. Thus, our main conclusions are that: 1) the results confirm the soundness of the proposed approach for automatic annotation of emotions; 2) the relevance of the extension of Emolex with a set of synonyms to improve the result; and 3) the effectiveness of GloVe model to extend the seed.

Our future research will deal with exploring this bootstrapping process in other corpora to verify the results in any genre; analysis of the process to create a more accurate seed; employing other corpora (e.g. social media corpus) to built the GloVe model as well as Word2Vec (W2V) models; and exhaustive manual review to detect potential improvement.

Acknowledgment

This research has been supported by the FPI grant (BES-2013-065950) and the research stay grant (EEBB-I-15-10108) from the Spanish Ministry of Science and Innovation. It has also funded by the Spanish Government (DIGITY ref. TIN2015-65136-C02-2-R) and the Valencian Government (grant no. PROMETEOII/ 2014/001).

References

1. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: Machine learning for text-based emotion prediction. In: Proceedings of the conference on HLT-EMNLP. pp. 579–586 (2005)
2. Aman, S., Szpakowicz, S.: Identifying Expressions of Emotion in Text. In: Text, Speech and Dialogue, pp. 196–205 (2007)
3. Aubur, D., Armantrout, R., Crystal, D., Dirda, M.: Oxford American Writer's Thesaurus. Oxford University Press (2004)
4. Boldrini, E., Martínez-Barco, P.: EMOTIBLOG: A model to Learn Subjetive Information Detection in the New Textual Genres of the Web 2.0-Multilingual and Multi-Genre Approach-. Ph.D. thesis (2012)
5. Cherry, C., Mohammad, S.M., De Bruijn, B.: Binary Classifiers and Latent Sequence Models for Emotion Detection in Suicide Notes. Biomedical informatics insights 5(Suppl 1), 147–154 (2012)
6. Choudhury, M.D., Gamon, M., Counts, S.: Happy, Nervous or Surprised? Classification of Human Affective States in Social Media. In: Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (2012)
7. Chowdhury, S., Chowdhury, W.: Performing Sentiment Analysis in Bangla Microblog Posts. In: International Conference on Informatics, Electronics & Vision (ICIEV). IEEE (2014)
8. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37 (1960)
9. Collins, M., Singer, Y.: Unsupervised Models for Named Entity Classification. In: In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora. pp. 100–110 (1999)
10. Dadvar, M., Trieschnigg, D., Ordelman, R., de Jong, F.: Improving cyberbullying detection with user context. In: Advances in Information Retrieval. pp. 693–696 (2013)

11. Ekman, P.: An argument for basic emotions. Cognition and Emotion pp. 169–200 (1992)
12. Gliozzo, A., Strapparava, C.: Semantic Domains in Computational Linguistics. Springer-Verlag Berlin Heidelberg (2009)
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1), 10–18 (2009)
14. Lee, S., Lee, G.G.: A Bootstrapping Approach for Geographic Named Entity Annotation. In: Asia Information Retrieval Symposium (AIRS). pp. 178–189 (2004)
15. Liew, J.S.Y., Turtle, H.R., Liddy, E.D.: EmoTweet-28: A Fine-Grained Emotion Corpus for Sentiment Analysis. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016) (2016)
16. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The {Stanford} {CoreNLP} Natural Language Processing Toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations. pp. 55–60 (2014)
17. Miller, G.A.: WordNet: A Lexical Database for English. Communications of the ACM 38(11), 39–41 (1995)
18. Mohammad, S.: #Emotional Tweets. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics (2012)
19. Mohammad, S.M.: Sentiment analysis: Detecting valence, emotions, and other affectual states from text. Emotion Measurement (2015)
20. Mohammad, S.M., Turney, P.D.: Crowdsourcing a Word-Emotion Association Lexicon. Computation and Language 29 (3), 436–465 (2013)
21. Montero, C.S., Suhonen, J.: Emotion analysis meets learning analytics: online learner profiling beyond numerical data. In: Proceedings of the 14th Koli Calling International Conference on Computing Education Research. pp. 165–169 (2014)
22. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In: Proceedings of the Third International ICWSM Conference. pp. 278–281 (2009)
23. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global Vectors for Word Representation. In: Empirical Methods in Natural Language Processing (EMNLP) (2014)
24. Platt, J.: Using Analytic QP and Sparseness to Speed Training of Support Vector Machines. In: Proc. Advances in Neural Information Processing Systems. pp. 557–563 (1999)
25. Plutchik, R.: A general psychoevolutionary theory of emotion. In: Theories of Emotion, pp. 3–33 (1980)
26. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Lucas, R.E., Agrawal, M., Park, G.J., Lakshmikanth, S.K., Jha, S., Seligman, M.E.P., Ungar, L.: Characterizing Geographic Variation in Well-Being Using Tweets. In: Proceedings of the International AAAI Conference on Weblogs and Social Media (2013)
27. Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: Affective text. In: Proceedings of the 4th International Workshop on Semantic Evaluations. pp. 70–74 (2007)
28. Wang, W., Chen, L., Thirunarayan, K., Sheth, A.P.: Harnessing Twitter "Big Data" for Automatic Emotion Identification. In: International Confernece on Social Computing (SocialCom) (2012)
29. Yarowsky, D.: Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In: Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics. pp. 189–196. ACL '95, Association for Computational Linguistics, Stroudsburg, PA, USA (1995)