Innovative Semi-Automatic Methodology to Annotate Emotional Corpora

Lea Canales
University of Alicante
Alicante (Spain)
lcanales@dlsi.ua.es

Carlo Strapparava
Fondazione Bruno Kessler
Trento (Italy)
strappa@fbk.eu

Ester Boldrini
University of Alicante
Alicante (Spain)
eboldrini@dlsi.ua.es

Patricio Martínez-Barco
University of Alicante
Alicante (Spain)
patricio@dlsi.ua.es

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, in which the automatic detection of emotion is a crucial element. Emotion detection has the potential of high impact by contributing the benefit of business, society, politics or education. Given this context, the main objective of our research is to contribute to the resolution of one of the most important challenges in textual emotion detection task: the problems of emotional corpora annotation. This will be tackled by proposing a new semi-automatic methodology. Our innovative methodology consists in two main phases: (1) an automatic process to pre-annotate the unlabelled sentences with a reduced number of emotional categories; and (2) a refinement manual process where human annotators will determine which is the predominant emotion between the emotional categories selected in phase 1. Our proposal in this paper is to show and evaluate the pre-annotation process to analyse the feasibility and the benefits by the methodology proposed. The results obtained are promising and allow obtaining a substantial improvement of annotation time and cost and confirm the usefulness of our pre-annotation process to improve the annotation task.

1 Introduction

Automatic detection of affective states in text has wide range of applications for business, society, politics or education. This is because detecting emotions is becoming more and more important due to the fact that it has the potential of bringing substantial benefits for different sectors: example of this can be for instance detecting depression (Cherry et al., 2012), identifying cases of cyber-bullying (Dadvar et al., 2013), tracking well-being (Schwartz et al., 2013), or contributing to improve the student motivation and performance (Montero and Suhonen, 2014).

So far, many of the existing machine learning techniques for automatic detection of emotions are supervised; systems first infer a function from a set of examples labeled with the correct sentiment (this set of examples is called the training data or labelled corpus). After this, the model is able to predict the emotion of new examples. Hence, the training dataset employed in supervised machine learning algorithms is crucial to build accurate emotion detection systems that can generate reliable results.

The creation of a labelled corpus is not trivial, since detecting emotion in text can be difficult even for humans due to the influence of each own background that can influence emotion interpretation. Most relevant research carried out so far has shown that the amount of agreement between annotations when associating emotion to instances is significantly lower compared to other tasks such as Part-Of-Speech (POS) or Named Entity (NE) detection. This is due to the fact that manual annotations can be significantly influenced by a set of different factors such as clarity of instructions, difficulty of task, training of the annotators, and even by the annotation scheme (Mohammad, 2016). For this reason, in this paper an innovative semi-automatic methodology is proposed to resolve one of the most important challenges in textual emotion detection task: the problems of the annotation of an emotional corpus.

The methodology proposed in our research consists of two main phases: (1) an automatic process to pre-annotate the unlabelled sentences with a reduced number of emotional categories; and (2) a refine-

Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pages 91–100, Osaka, Japan, December 12 2016.

ment manual process where human annotators will determine which is the predominant emotion between the emotional categories selected in phase 1.

By means of proposing innovation in terms of annotation methodology, our aim is to reduce the complexity of emotion annotation task through reducing the number of emotional categories automatically, since the influence of the number of coding categories on reliability estimation is really important. As Antoine et al. (2014) concluded, the agreement values increase significantly when the number of classes decreases. Hence, our hypothesis is that the decrease of complexity of emotion annotation task through the reduction of the number of emotional categories will allow us to improve the reliability on the task. This methodology will allow us annotating large amount of emotional data in any genre efficiently and with guarantee of high standards of reliability. Our proposal in this paper is to show and evaluate the pre-annotation process to analyse the feasibility and the benefits by the methodology proposed.

The rest of the paper is organised as follows. Section 2 presents the related work and a reflection on the pending issues. After this, the proposed method is described in detail in the Section 3. Then, 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 summaries the most relevant emotional corpora developed for emotion detection purposes, their features and how they have been developed. Our analysis on Emotion Detection is focused on detecting areas of improvement that we aim to contribute to tackle with our research.

According to research in psychology, there is a number of theories about how to represent the emotions that humans can perceive and express. Among these theories, some of them are focused on defining the set of the basic emotions (Ekman, 1992; Plutchik, 1980), although there is not an universal consensus about which set of emotions are the most basic. Nevertheless, most of the work 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 that are generally more accurate. Among these resources, we can find corpora labelled with the six basic emotions categories proposed by Ekman such as: (Alm et al., 2005) annotated a sentence-level corpus of approximately 185 children stories with emotion categories; (Aman and Szpakowicz, 2007) annotated blog posts collected directly from Web with emotion categories and intensity; or (Strapparava and Mihalcea, 2007) 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: (Neviarouskaya et al., 2009) 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 (Boldrini and Martínez-Barco, 2012); or EmoTweet-28 corpus that consists of a collection of tweets annotated with 28 emotion categories (Liew et al., 2016).

The common feature of these emotional corpora is that have been annotated manually, 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 and with the aim of overcoming the cost and time consuming shortcoming of manual annotation, several emotional resources have recently been developed employing emotion word hashtags to create automatic emotional corpus on Twitter. (Mohammad, 2012a) 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 (Choudhury et al., 2012; Wang et al., 2012).

Thus, in Sentiment Analysis 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 cost; however, it can be applied on social networks and microblogging services exclusively because they are only used in these genres. For this reason, our objective is to develop a semi-automatic methodology for large-scale annotation of emotional corpora in any genre and with high standards of reliability.

3 Pre-annotation process

After a reflection on the pending issues, this section describes the pre-annotation process developed for improving the emotion annotation task. The section is divided into four subsections where the dataset employed and the main tasks carried out by the process are explained.

The process receives as input data a collection of unlabelled sentences/phrases and a set of emotions. The approach presented in this paper works with the Ekman‘s six basic emotions (Ekman, 1992), although the process can also be adapted for other set of group of emotions.

The overall pre-annotation process is described in Figure 1, which shows the two main steps the process: selecting emotional seed words and the association between emotions and sentences, explained in subsection 3.2 and subsection 3.3, respectively.

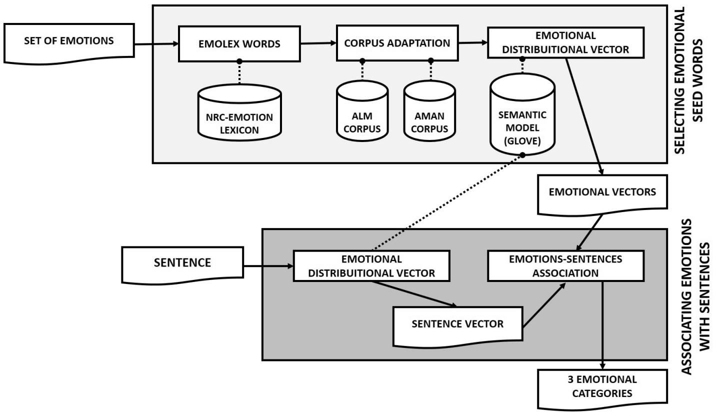

Figure 1: Overall pre-annotation process

3.1 Data

Regarding the corpora employed for the evaluation, this approach is assessed on two emotional corpora with sentence-level annotations: (i) Alm et al. (2005) corpus; and (ii) Aman and Szpakowicz (2007) corpus.

Alm corpus. This dataset consists in 1,580 annotated sentences from tales by the Grimm brothers, H.C. Andersen, and B. Potter. This corpus was annotated manually with an extended set of the Ekman’s basic emotions (angry, disgusted, fearful, happy, sad, positively surprised and negatively surprised). For our evaluation, we employ the version of the corpus where the merged label set was used: anger-disgust, fear, joy, sadness, and surprise.

Aman corpus. This dataset 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).

These corpora are selected because of several reasons: (i) both corpora are manually annotated allowing us to compare automatic annotation to manual annotation; (ii) they are relevant to emotion detection task since they have been employed in many works to detect emotions (Keshtkar and Inkpen, 2010; Chaffar and Inkpen, 2011; Mohammad, 2012b); and (iii) these corpora allow us to test our approach about corpora with different sources of information: tales and blogs from Web. Thus, the usability and effectiveness of our approach can be checked.

3.2 Selecting Emotional Seed Words

In this section, the process of creation the emotional seed words employing an emotional resource is presented. This approach employs NRC Word-Emotion Association Lexicon (Emolex) (Version 0.92) (Mohammad and Turney, 2013) as emotional lexicon, although the process can be adapted to another resource annotated with emotions.

Emolex is a lexicon of general domain consisting of 14,000 English unigrams (words) associated with the Plutchik‘s eight basic emotions (Plutchik, 1980) (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive) compiled by manual annotation. We adopted them because: (i) it is general domain and it can be apply in different corpora; (ii) it is annotated a superset of Ekman’s six basic emotions; and (iii) the most relevant feature of this resource is that the terms in this lexicon are carefully chosen to include some of the most frequent nouns, verbs, adjectives and adverbs.

The algorithm for the creation of the seed consists of:

- **Step 1 - Emolex words**: the process selects the Emolex words associated with only one of the Ekman’s basic emotions to create an accurate seed without ambiguous words. Thus, each emotional category is represented by a bag of words. Figure 2 shows an example for `ANGER`, `DISGUST` and `SADNESS` emotions.

- **Step 2 - Corpus adaptation**: These bags of words are adapted to each corpus removing those words that not appear in the corpus. In this manner, the seed contains only the emotional words employed in the corpus to annotate. Figure 2 shows an example of the adaptation process for Alm corpus.

- **Step 3 - Emotional distributional vector**: Each seed is transformed into a distributional vector adding up the distributional vectors of each word contained in the seed. To achieve that, a GloVe model (Pennington et al., 2014) built from the lemmas and POS of the British National Corpus (BNC)[1] is employed. This model is explained in detail in Section 3.3.

Once the process is completed, each emotion is represented by a distributional vector, a real-valued vector that stores its semantic features. Moreover, the process also creates a vector for a `NEUTRAL` category with the Emolex words not associated with the Ekman’s basic emotions.

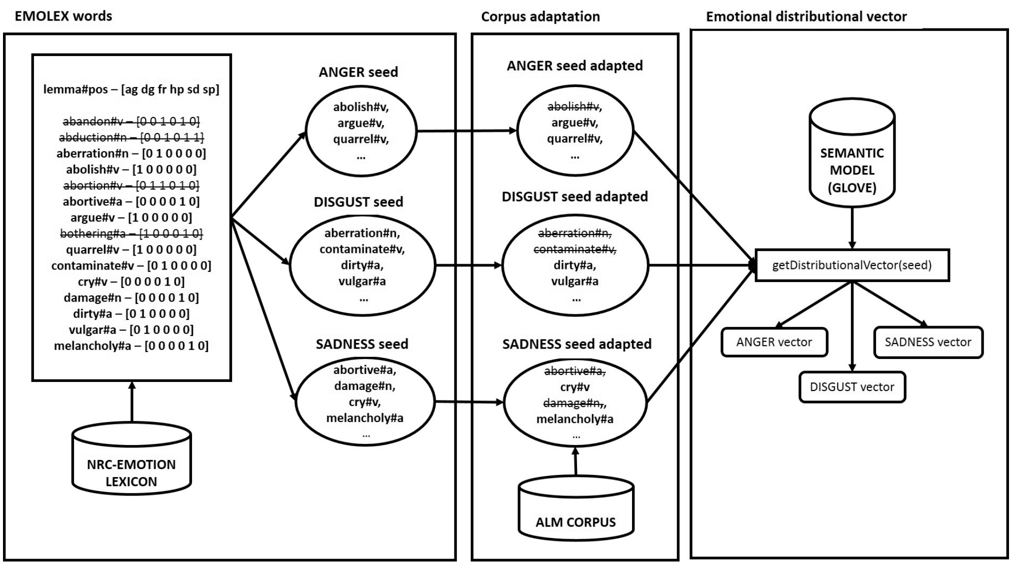

Figure 2: Creation of the emotional seed words for `ANGER`, `DISGUST` and `SADNESS` emotions (sample).

[1] http://www.natcorp.ox.ac.uk/

3.3 Associating Emotions with Sentences

After having the emotional seeds, the next step will consist in to associate the emotions represented by vectors with sentences, with the help of Distributional Semantics.

Distributional Semantic Models (DSM) 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[2]. Finally, each word is represented with a real-valued vector called word vector or *word embedding*.

Two are the 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 association between the emotional seeds and the sentences of our proposal is based on the estimation the similarity among them. For this reason, in this paper we test a model based on global matrix factorization methods: GloVe (Pennington et al., 2014).

This model is run with the default settings, 300 dimensions and on the lemmas of the British National Corpus (BNC)[3] that can be considered as a balanced resource since it includes texts from different genres and domains

The process of the association consists of:

- **Step 1 - Emotional distributional vector**: each sentence is pre-processed (tokenization, lemmatization and Part-Of-Speech Tagger) using Stanford Core NLP (Manning et al., 2014) and then is represented by a distributional vector adding up the vectors of their words (noum, verbs, adjectives and adverbs). Figure 3 shows an example for the sentence *'The bear in great fury ran after the carriage'*.

- **Step 2 - Emotions-Sentences Association**: the process measure the similarity between the vector of the sentence and the vectors of each emotional category and associates the three emotions whose semantic similarity is higher. Figure 3 shows the pre-annotated emotions for the example sentence, among which is the emotion of the gold standard of Alm corpus: `ANGER-DISGUST`.

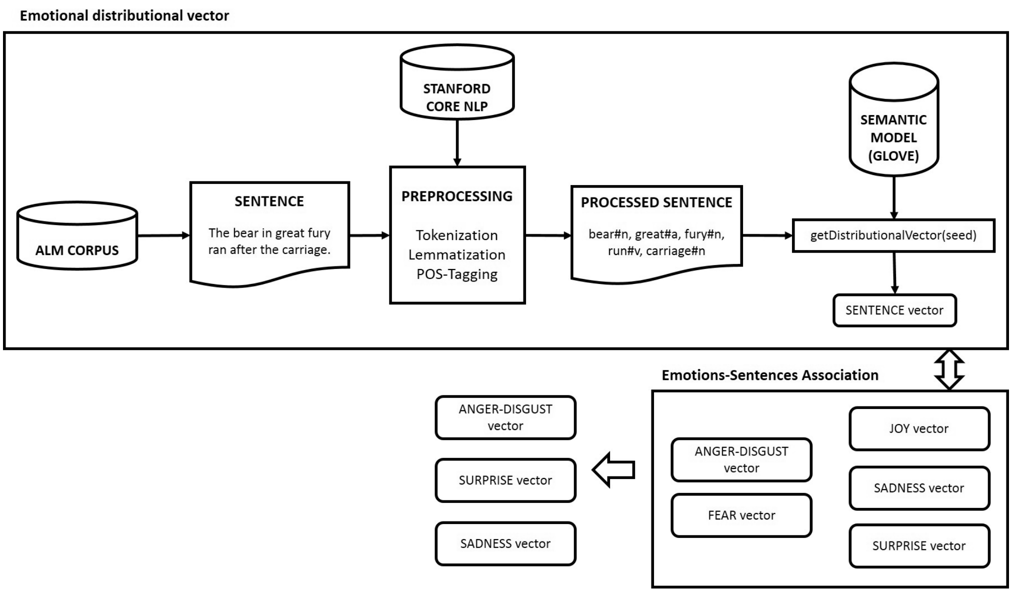

Figure 3: Association process between emotions and sentences with an example from Alm corpus.

[2] http://wordspace.collocations.de/doku.php/course:acl2010:start

[3] http://www.natcorp.ox.ac.uk/

At the end of the process, each sentence is annotated with the three emotional categories that are the ones more related to this sentence. The first phase of the methodology, the pre-annotation process, is finished with this step. Then, the second phase that consisting of a refinement manual process where human annotators will determine which is the predominant emotion would be developed. Although, this paper is focused on the evaluation of the pre-annotation process to evaluate the feasibility and the benefits by the methodology proposed before that the second phase be developed.

4 Evaluation

Once the pre-annotation process has been detailed, this section shows its evaluation.

Given the importance of the creation of an accurate seed in the pre-annotation process and the size of Emolex when it works with Ekman's basic emotions, three approaches have been evaluated employing different versions of Emolex (original, WordNet (WN) synonyms and Oxford synonyms). The extension process of Emolex is completely automatic and is explained in detail Section 4.1.

4.1 Enriched approaches by WordNet and Oxford synonyms

The enriched approaches employed consist in the extension of Emolex employing the synonyms of WordNet (Version 3.0) (Miller, 1995) and the Oxford American Writer Thesaurus (Aubur et al., 2004).

In this process, each word contained in Emolex was looked up in WordNet/Oxford and the synonyms of all of senses were obtained and were added to the seed associated with the Emolex word. Figure 4 shows an example of the process employing WordNet. The word *alive* is contained in Emolex and has the emotion `JOY` associated. The process looks up *alive* in WordNet and obtains the synonyms of all of senses: *live*, *animated*, *active*, *alert* and *awake*. These synonyms are added to the seed of `JOY` emotion.

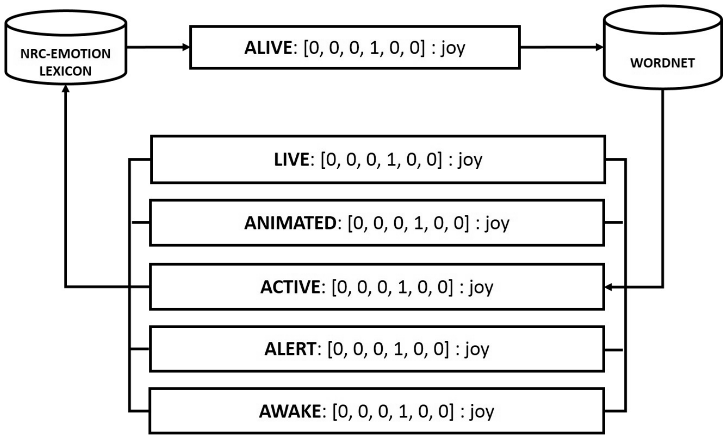

Figure 4: Process of the extension of Emolex by WordNet synonyms.

The enriched approaches run the same process than the original one, but employing the new versions of Emolex.

4.2 Evaluation Methodology

The pre-annotation process is assessed applying the measure of agreement between the gold standard of each corpus and our annotation. Since the pre-annotation process annotates the three emotional categories more related to each sentence, the evaluation process considers that there is an agreement if the correct emotion (the gold standard) is one of the three pre-annotated emotions. To achieve that, Cohen (1960) kappa and Krippendorff (2004) alpha are calculated. With both measures, we calculate the agreement based on a formula expressed in term of agreement (Cohen's kappa) and in terms of disagreement (Krippendorff's alpha). Since in metrics based on coefficients k, all disagreements are treated equally and disagreements are not all alike for semantic and pragmatic features (Artstein and Poesio, 2008), as the emotion detection task.

4.3 Results

The results obtained with both corpora are shown in Tables 1 and 2 below. Each table shows the Cohen's kappa and Krippendorff's alpha values obtained for each emotion employing the original and enriched approaches.

	Aman corpus					
	Cohen's Kappa			Krippendorff's Alpha		
	Original Appr.	WN Appr.	Oxford Appr.	Original Appr.	WN Appr.	Oxford Appr.
Anger	**0.50193**	0.39984	0.38544	**0.50199**	0.39991	0.38002
Disgust	0.35432	**0.56397**	0.53061	0.35033	**0.56386**	0.53061
Fear	**0.43424**	0.39342	0.26755	**0.43371**	0.39001	0.26252
Joy	**0.85897**	0.76931	0.71264	**0.85899**	0.76930	0.71251
Sadness	**0.51275**	0.45187	0.47945	**0.51280**	0.45188	0.47851
Surprise	**0.49255**	0.42098	0.38801	**0.48843**	0.41214	0.37709

Table 1: Cohen's kappa and Krippendorff's alpha values obtained by the Original Approach and the Enriched Approaches in the Comparison between their Annotations and the Gold of Aman Corpus.

	Alm corpus					
	Cohen's Kappa			Krippendorff's Alpha		
	Original Appr.	WN Appr.	Oxford Appr.	Original Appr.	WN Appr.	Oxford Appr.
Anger-Disgust	0.36762	0.53641	**0.56084**	0.34931	0.53655	**0.56043**
Fear	0.48990	0.58467	**0.59671**	0.48677	0.58481	**0.59667**
Joy	0.77948	0.75616	**0.79838**	0.77949	0.75523	**0.79823**
Sadness	0.59576	**0.72433**	0.57264	0.59566	**0.72424**	0.56721
Surprise	0.43095	0.38240	**0.44159**	0.42869	0.38251	**0.43351**

Table 2: Cohen's kappa and Krippendorff's alpha values obtained by the Original Approach and the Enriched Approaches in the Comparison between their Annotations and the Gold of Alm Corpus.

Several conclusions can be drown from Table 1. The results show the soundness of the original seed since they obtain the best results for most of the emotions except for `DISGUST` emotion. This can be due to the difficulty to distinguish between `ANGER` and `DISGUST` emotion in text even for humans since the results of these emotions are inverted in the enriched approaches. Thus, if we consider both emotions as an unique category, like on Alm corpus, these values would improve.

Regarding the rest of emotions, the values obtained by `FEAR` emotion are low, especially in the enriched approaches. This indicates that the seed of `FEAR` for Aman corpus not contains the words employed on blog posts to express `FEAR` emotion. But this could be improve it, including the words used in the blog posts in the seed. And about `SURPRISE` emotion, the results also are low although in this case it is coherent with many studies (Alm et al., 2005; Strapparava and Mihalcea, 2007).

About the conclusions on Alm corpus (Table 2), the enriched approach by Oxford synonyms demonstrates the improvements obtained by these synonyms since obtains the best results for most of emotions.

Taking into account that the pre-annotation have been carried out with a totally automatic process, the results on this corpus are considerably promising since the best approach obtains values higher 56% for the entire set of emotions except `SURPRISE`. Although, as we mentioned, these results are coherent with many studies. The lack of agreement in this emotion is due to the lack of para-linguistic information like tone, emphasis and facial expressions, relevant features for `SUPRISE` emotion.

4.4 Discussion

These results are interpreted taking into account that the gold standard of these corpora annotated manually (Aman and Alm corpus) achieved values of agreement less than 80%, the value needed to get a good reliability. Since there are cases in which the annotations of the gold standard seem questionable under a new review by humans. In these cases, our annotations can disagree with the gold standard but are considered errors of annotations and hence the agreement is worse.

Comparing both corpora, the results show that the pre-annotation process obtains better values on Alm corpus than on Aman data. This can be due to the genre of each corpus since the sentences on Aman corpus are from blog posts and the vocabulary employed is not formal and is not included in Emolex. Thus, the seed is less accurate than on Alm corpus, a corpus about children tales. Although, this is not a problem for our methodology since the pre-annotation process can be improved employing emotional lexicon adapted to different genres to create the seed. Hence, if the process employs a lexicon with the vocabulary employed on social media, the seed will be more accurate and the results will be improved.

Concerning the enriched approaches, the results show improvements on Alm corpus whereas on Aman corpus the best approach is the original one. This is related to the genre of the text because Oxford and Wordnet synonyms introduce noise on Aman corpus, since the vocabulary included in these resources is formal whereas the vocabulary employed in blog posts is informal.

5 Conclusion

As presented in the introductory section of this paper, the rationale beyond our research is the need to develop a methodology that allow us to tackle the annotation task of emotions with views on improving supervised learning techniques.

The paper presents an innovative semi-automatic methodology to annotate emotional corpora consisting of two main phases: (1) an automatic process to pre-annotate the unlabelled sentences with a reduced number of emotional categories; and (2) a refinement manual process where human annotators will determine which is the predominant emotion between the emotional categories selected in phase 1. A methodology adaptable to the genre of text and the set of emotions employed that will allow us the annotation of large amount of emotional data in any genre with efficiently and high standards of reliability.

The first evaluation performed for this innovative methodology confirms its feasibility and benefits since the agreements values are promising. Thus, our main conclusion is that the reduction of the number of categories could provide us benefits that will revert in positive impact in the emotion annotation task and therefore to improve the reliability on emotional corpora.

Taking into account the results obtained, our future work will be focused on developing a manual annotation task with the sentences pre-annotated by our automatic process to verify the benefits of the new methodology; analysis of the process to create a more accurate seed; and employing other emotional resources to create the seeds adapted to different genres and set of emotions.

Acknowledgements

This research has been supported by the FPI grant (BES-2013-065950) and the research stay grant (EEBB-I-16-11174) 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

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the conference on HLT-EMNLP*, pages 579–586.

Saima Aman and Stan Szpakowicz. 2007. Identifying Expressions of Emotion in Text. In *Text, Speech and Dialogue*, pages 196–205.

Jean-yves Antoine, Jeanne Villaneau, and Anaïs Lefeuvre. 2014. Weighted Krippendorff's alpha is a more reliable metrics for multi- coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, number 1, pages 550–559.

Ron Artstein and Massimo Poesio. 2008. Inter-coder Agreement for Computational Linguistics. *Comput. Linguist.*, 34(4):555–596.

David Aubur, Rae Armantrout, David Crystal, and Michael Dirda. 2004. *Oxford American Writer's Thesaurus*. Oxford University Press.

Ester Boldrini and Patricio Martínez-Barco. 2012. *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.

Soumaya Chaffar and Diana Inkpen. 2011. Using a Heterogeneous Dataset for Emotion Analysis in Text. In *Proceedings of the 24th Canadian Conference on Advances in Artificial Intelligence*, Canadian AI'11, pages 62–67, Berlin, Heidelberg. Springer-Verlag.

Colin Cherry, Saif M. Mohammad, and Berry De Bruijn. 2012. Binary Classifiers and Latent Sequence Models for Emotion Detection in Suicide Notes. *Biomedical informatics insights*, 5(Suppl 1):147–154.

Munmun De Choudhury, Michael Gamon, and Scott Counts. 2012. 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*.

J Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37.

Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong. 2013. Improving cyberbullying detection with user context. In *Advances in Information Retrieval*, pages 693–696.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, pages 169–200.

Fazel Keshtkar and Diana Inkpen. 2010. A Corpus-based Method for Extracting Paraphrases of Emotion Terms. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 35–44, Stroudsburg, PA, USA. Association for Computational Linguistics.

Klaus Krippendorff. 2004. Content Analysis: An Introduction to Its Methodology. In *Content Analysis: An Introduction to Its Methodology*, chapter 11. Sage Publications.

Jasy Suet Yan Liew, Howard R. Turtle, and Elizabeth D. Liddy. 2016. EmoTweet-28: A Fine-Grained Emotion Corpus for Sentiment Analysis. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David McClosky. 2014. The {Stanford} {CoreNLP} Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

George A Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computation and Language*, 29 (3):436–465.

Saif Mohammad. 2012a. #Emotional Tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*.

Saif Mohammad. 2012b. Portable Features for Classifying Emotional Text. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 587–591, Montréal, Canada. Association for Computational Linguistics.

Saif M Mohammad. 2016. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. In Herb Meiselman, editor, *Emotion Measurement*. Elsevier.

Calkin Suero Montero and Jarkko Suhonen. 2014. Emotion analysis meets learning analytics: online learner profiling beyond numerical data. In *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*, pages 165–169.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2009. Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In *Proceedings of the Third International ICWSM Conference*, pages 278–281.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of Emotion*, pages 3–33.

Hansen Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Richard E. Lucas, Megha Agrawal, Gregory J. Park, Shrinidhi K. Lakshmikanth, Sneha Jha, Martin E. P. Seligman, and Lyle Ungar. 2013. Characterizing Geographic Variation in Well-Being Using Tweets. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P. Sheth. 2012. Harnessing Twitter "Big Data" for Automatic Emotion Identification. In *International Confernece on Social Computing (SocialCom)*.