Exploiting a Bootstrapping Approach for Automatic Annotation of Emotions in Texts

Lea Canales*, Carlo Strapparava†, Ester Boldrini* and Patricio Martnez-Barco*
*Department of Software and Computing Systems
Unviversity of Alicante, Alicante, Spain
Email: {lcanales, eboldrini, patricio}@dlsi.ua.es
†FBK-irst, Trento, Italy
Email: strappa@fbk.eu

***Abstract*—The objective of this research is to develop a technique to automatically annotate emotional corpora. The complexity of automatic annotation of emotional corpora still presents numerous challenges and thus there is a need to develop a technique that allow us to tackle the annotation task. The relevance of this research is demonstrated by the fact that people's emotions and the patterns of these emotions provide a great value for business, individuals, society or politics. Hence, the creation of a robust emotion detection system becomes crucial. Due to the subjectivity of the emotions, the main challenge for the creation of emotional resources is the annotation process. Thus, with this staring point in mind, the objective of our paper is to illustrate an innovative and effective bootstrapping process for automatic annotations of emotional corpora. The evaluations carried out confirm the soundness of the proposed approach and allow us to consider the bootstrapping process as an appropriate approach to create resources such as an emotional corpus that can be employed on supervised machine learning towards the improvement of emotion detection systems.**

***Keywords*-Sentiment Analysis; Emotion Detection; Emotional Corpora; Bootstrapping Technique**

I. 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. The goal of this research is to create resources and tools which evaluate and represent people‘s emotions through analyzing automatically on-line content such as the comments on the Social Web.

The relevance of this research is demonstrated by the large number of applications to detect people's emotional states and the great value provided by them for business, individuals, society or politics, and for applications in e-learning environment [1] or suicide prevention [2].

The main challenge for the creation of emotional resources is that they are complex to annotate. Traditionally, these resources have been annotated manually, where several annotators had to associate different sentences or phrases with emotional categories. However, the manual annotation is a hard and time-consuming task. Moreover, in emotion detection obtaining a good inter-annotator agreement is a challenge due to the subjectivity of the task.

For this reason, our proposal is to exploit a bootstrapping approach for automatic annotations of emotions. The process consists of two main steps; 1) the creation of the seed where NRC Emotion Lexicon [3] is employed to annotate the sentences by its emotional words, and 2) the extension of the seed based on generalised similarity measures (LSA and Word2vec models).

Achieving the same quality of manual annotations is difficult; although from a functional point of view, our approach demonstrates that obtaining automatic emotional annotations is possible.

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.

II. Related works

This section summarises the most relevant emotional corpora developed for emotion detection purposes, their features and how they have been developed, as well as, some of works where bootstrapping technique was applied for annotation are analysed.

An emotional corpus is a large and structured set of sentences where each sentence is tagged with one or more emotional tags. These corpora are a fundamental part of supervised-learning approaches, as they rely on a labelled training data, a set of examples. The supervised learning algorithm analyses the training data and infers a function, which it used for mapping new examples [4].

Concretely in the emotion detection area, supervised-learning technique is applied in different approaches, and hence the development of emotional corpora becomes crucial.

Generally, emotional corpora have been annotated manually, thus allowing machine learning algorithms learn from human annotations. Regarding corpora annotated manually with six basic emotion categories proposed by Ekman, there

978-1-5090-5206-6/16 $31.00 © 2016 IEEE
DOI 10.1109/DSAA.2016.78

are several works, such as: [5] annotated a sentence-level corpus of approximately 185 children stories with emotion categories; [6] annotated blog posts collected directly from Web with emotion categories and intensity; [7] annotated news headlines with emotion categories and valence; or [8] annotated 8,150 tweets collected from Web with emotion categories.

Moreover, manually annotated corpora with other group of emotions can be found in the literature: [9] corpus extracted 700 sentences from BuzzMetrics blog posts annotated with one emotion from the subset defined by Izard [10]; [11] corpus extracted 1,000 sentences from various stories annotated with one of 14 categories of their annotation scheme; [12] present 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 with a group of 15 emotions; [13] compile a dataset of 2012 US presidential election tweets which includes multi-layer annotation for emotion, polarity valence, style and purpose; or [14] present EmoTweet-28 corpus that consists of a collection of tweets annotated with 28 emotion categories.

These works demonstrate the existence of a large set of emotional corpora in different genres: children stories, blog posts, news headlines or Twitter, and with different group of emotions. However, all of them have been manually annotated, being a hard and time-consuming task.

Consequently, some emotional corpora have recently been developed automatically. For instance, [15] describe how a corpus from Twitter posts (Twitter Emotional Corpus) is created by using emotion word hashtags. This approach collects tweets with hashtags corresponding to Ekman's basic emotions: #anger, #disgust, #fear, #happy, #sadness, and #surprise. TEC has about 21,000 tweets from about 19,000 different people. In literature, it is possible to find several works that use emotion word hashtags to create emotional corpora from Twitter: (i) [16] dataset consists of 6.8 million affect-labelled posts, where each post is associated with one of 172 moods (classified on 11 affects); (ii) [17] contains about 2.5 million tweets annotated by harnessing emotion-related hashtags and employed 131 emotion hashtags as keywords aggregated by 7 emotion categories (joy, sadness, anger, love, fear, thankfulness, surprise); or (iii) [18] use Twitter hashtags to automatically label messages and choose Circumplex model [19], as model of emotional states which characterises affective experience along two dimensions: valence and arousal.

As previously mentioned, there is a considerable interest in developing emotional corpora for applying supervised-learning techniques. Thus, in scientific community, research on developing an automatic process to annotate has increased. Nevertheless, the techniques developed so far have been focused on Twitter. For this reason, our objective is to develop a technique to annotate an emotional corpus automatically from any genre.

In this work, a bootstrapping technique has been developed to contribute to resolve this problem. It is a semi-supervised technique that allows us to develop a process automatically or semi-automatically whose effectiveness has been demonstrated by the results obtained in a wide range of computational linguistics problems such as the word sense disambiguation [20] or the named entity classification [21]. More concretely, [22] and [23] demonstrate the adequacy of the bootstrapping technique for annotation task.

As a result of the conclusions drawn from the analysis of the previous work and a reflection on the pending issues, in the next section the bootstrapping process is described in detail.

III. Bootstrapping process

This section shows the bootstrapping process developed for automatic annotation. It is divided into four subsections where the dataset employed and the main tasks carried out by bootstrapping process are explained.

The process receives as input data a collection of unlabelled sentences/phrases and a set of emotions, concretely the Ekman‘s six basic emotions [24]. The objective of this task is to annotate unlabelled sentences with the emotions expressed in each sentence.

The overall bootstrapping process is described in Figure 1.

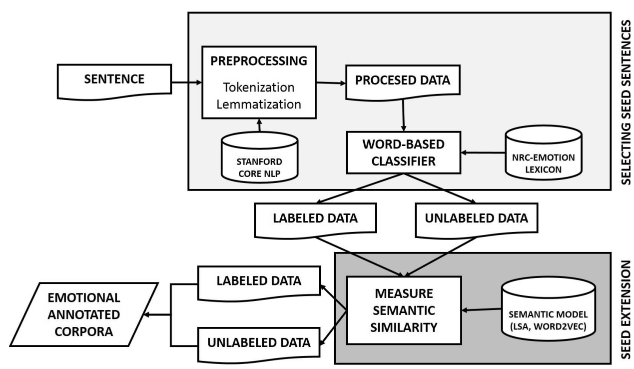

Figure 1. Overall bootstrapping process.

A. Dataset

The dataset employed to test our approaches is Aman corpus [6] that contains sentence-level annotation of 4,000 sentences from blogs posts collected directly from Web. This corpus was annotated manually with the six emotion categories proposed by Ekman and the emotion intensity (high, medium, or low). The distribution of each emotion in the gold corpus is shown in Table I.

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

Table I
THE DISTRIBUTION OF EMOTIONS IN THE GOLD AMAN CORPUS

	Aman Corpus # of instances
Anger	179
Disgust	172
Fear	115
Joy	536
Sadness	173
Surprise	115
Neutral	2,800
Total	4,090

is relevant to emotion detection task, since it has been employed in many works to detect emotions [25] [26] [27]; 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.

B. Selecting seed sentences

In this section, the process of creating the initial seed by exploring NRC Word-Emotion Association Lexicon (Emolex) [3] is presented. In this approach, Emolex is applied to annotate each sentence of the Aman corpus which contains emotional words.

Emolex is a lexicon of general domain consisting of 14,000 English unigrams (words) associated with an emotional vector of the Plutchik‘s eight basic emotions [28] (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive), 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. The coverage of this reduced version of Emolex is shown in Table II.

Table II
THE COVERAGE OF EMOTIONS IN THE REDUCED VERSION OF EMOLEX

	Emolex (Ekman‘s emotions) # of Words
Anger	1,247
Disgust	1,058
Fear	1,476
Joy	689
Sadness	1,191
Surprise	534

The algorithm of the creation of the seed consists of:

- Step 1: each sentence has an emotional vector associated with a value to each emotion ([anger, disgust, fear, joy, sadness, surprise]) initialised to zero (Figure 2).
- Step 2: each sentence is tokenized and lemmatized using Stanford Core NLP [29].
- Step 3: each word of the sentence is looked up in Emolex. If a word is in Emolex, its emotional values are added to the emotional vector of the sentence.
- Step 4: each sentence is annotated with the emotion whose has the highest value in the emotional vector of the sentence.

Figure 2 shows an example of the creation of the seed. The sentence *“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.

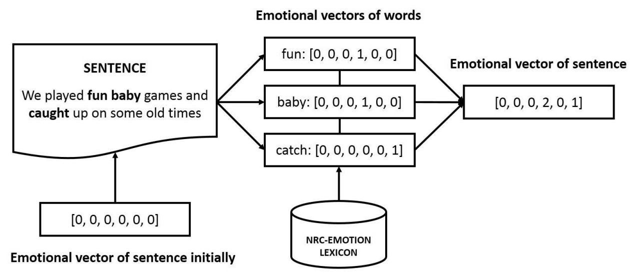

Figure 2. Example of the process of selecting seed sentences when a sentence is annotated.

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 3 shows an example about a sentence without predominant emotion. The sentence *“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.

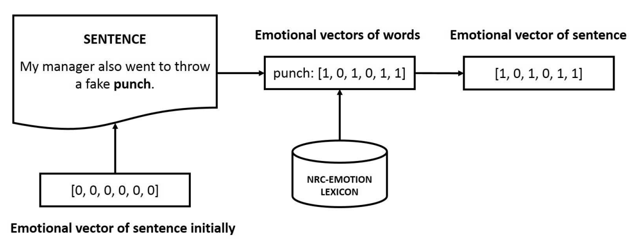

Figure 3. Example of the process of selecting seed sentences when a sentence is not annotated.

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.

C. 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, a similar approach to [30] is adopted, who use latent semantic spaces to estimate the similarity between documents and words. In our case, the similarity among non-annotated sentences and annotated sentences is estimated using Latent Semantic Analysis (LSA) and Word2Vec models (W2V).

As far as the LSA model is concerned, the one applied in this work is [31]. They run the SVD on the lemmas of the British National Corpus (BNC)[1] that can be considered as a balanced resource since it includes texts from different genres and domains.

Concerning Word2Vec models, the new models for learning distributed representation of words (CBOW and SKIP-gram) are applied. In particular, the word2vec operation is run on the lemmas of one of the source of Annotated English Gigaword[2]: New York Times Newswire Service to build CBOW and SKIP-gram models, with the default settings. Moreover, the English vectors learned with word2vec on the words of the BNC and WackyPedia/ukWaC [32] are also applied to test another approach.

Hence, a LSA model and three Word2vec models: (i) a CBOW model built from English GigaWord; (ii) a SKIP-gram model built from English Gigaword; (iii) a CBOW model built from BNC and WackyPedia/ukWaC are applied in the extension of seed. The process of extension of the seed consists of measuring the similarity among non-annotated sentences and annotated sentences using the models listed. 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.

D. 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 [33] is applied, representing the sentences as a vector of words weighted by their counts using Weka [34].

[1]http://www.natcorp.ox.ac.uk/
[2]https://catalog.ldc.upenn.edu/LDC2012T21

IV. Evaluation

Given the importance of the creation of an accurate seed in bootstrapping process and the size of Emolex when it works with Ekman‘s basic emotions, it has been considered to develop several approaches employing different versions of Emolex.

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 Plutchik‘s emotions. Therefore, the improvement of Emolex with synonyms can be considered interesting to test if the creation of the seed improves. The extension process of Emolex is completely automatic and is explained in detail in the next sections.

A. Enriched approach by WordNet synonyms

One of the enriched approach employed consists of the extension of Emolex employing the synonyms of WordNet [35].

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 4 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‘.

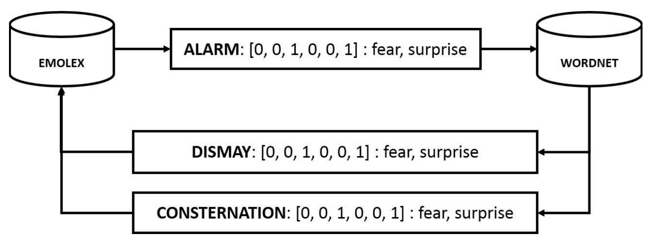

Figure 4. Process of the extension of Emolex by WordNet synonyms.

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.

B. 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 [36] 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. Figure 5 shows an example of the process, where the synonyms of ‘blackmail‘ are analyzed. In Oxford dictionary, ‘blackmail‘ has two synonyms associated: ‘extortion‘ and ‘exaction‘. Before adding them to Emolex, the process checks if the synonym is already in Emolex.

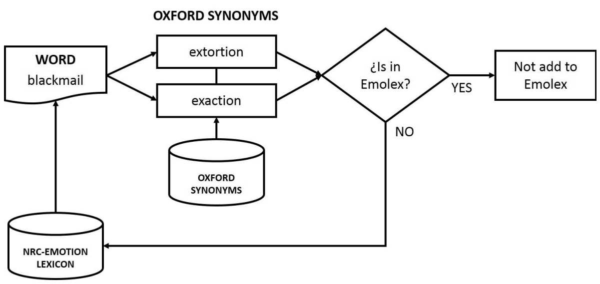

Figure 5. Process of the extension of Emolex by Oxford synonyms.

During this process, there were synonyms associated with different words of Emolex, and these words were associated with emotional vectors with different values. In these cases, these synonyms were associated with the emotions in common of all of emotional vectors. Figure 6 shows an example of the process, where the word ‘vomiting‘ is synonym of two Emolex words: ‘sickness‘ and ‘nausea‘. These words are annotated with different emotions. For this reason, the process annotates the word ‘vomiting‘ with shared emotions between two Emolex words. In this case, ‘vomiting‘ is annotated with `DISGUST` emotion.

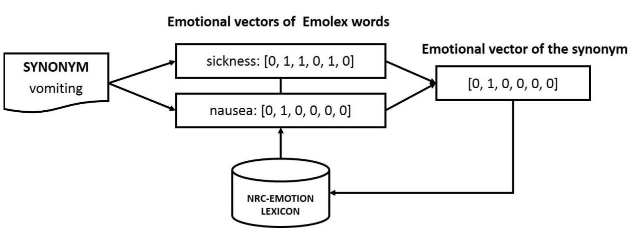

Figure 6. Process of the synonyms associated with two or more Emolex words.

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.

C. Evaluation Methodology

The evaluation methodology is divided into two steps. On the one hand, the usability of the corpus annotated automatically to built an emotional model is evaluated. On the other hand, the agreement between automatic and manual annotations is also assessed, employing an agreement measure.

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

For the comparative between manual and automatic annotations, a detailed knowledge of features of emotion annotation task developed on Aman Corpus is required. This task was manually developed by four annotators who received no training, though they were given samples of annotated sentences to illustrate the kind of annotations required. Concerning the emotion categories, the Ekman‘s basic emotions were selected and two further categories were added: (i) mixed emotions and (ii) no emotion, resulting in eight categories to which a sentence could be assigned. To measure how the annotators agree on classifying a sentence, Cohen’s kappa [37] was employed because popularly used to compare the extent of consensus between annotators.

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

D. Results

The results obtained by each classifier in all of our approaches and on Aman corpus are shown in the tables below (Tables III-VII). There is a table for Aman corpus and one table for each semantic similarity model: LSA, Word2Vec model built from Gigaword (CBOW and Skip-gram) and Word2Vec model built from BNC and WackyPedia/ukWaC. Each table shows the precision (P), recall (R) and F1-values (F1) obtained for each emotion employing the original approach and the enriched approaches.

Table III
PRECISION, RECALL AND F1-VALUES OBTAINED BY SMO MULTI-CLASSIFIER ON THE GOLD OF AMAN CORPUS

	Aman Corpus		
	P	R	F1
Anger	0.538	0.274	0.363
Disgust	0.714	0.32	0.442
Fear	0.672	0.357	0.466
Joy	0.720	0.513	0.599
Sadness	0.577	0.260	0.359
Surprise	0.553	0.226	0.321
Neutral	0.798	0.955	0.869
Macro Avg.	0.753	0.774	**0.745**

Regarding the results obtained in the comparison between automatic and manual annotations, these are shown in Tables VIII-IX. These tables show Cohen‘s kappa values obtained by each one of our approaches when they are compared to the gold standard of Aman corpus.

Table IV
PRECISION, RECALL AND F1-VALUES OBTAINED BY SMO MULTI-CLASSIFIER ON THE CORPUS DEVELOPED APPLYING LSA AS SEMANTIC METRIC IN THE EXTENSION OF THE SEED.

	LSA								
	Original approach			Enriched approach WN			Enriched approach Oxford		
	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
Disgust	0.250	0.068	0.107	0.308	0.178	**0.225**	0.353	0.120	0.179
Fear	0.401	0.236	0.297	0.392	0.303	**0.342**	0.412	0.251	0.312
Joy	0.574	0.571	0.572	0.677	0.702	**0.689**	0.565	0.604	0.584
Sadness	0.247	0.107	0.149	0.467	0.269	0.341	0.591	0.462	**0.519**
Surprise	0.459	0.224	0.301	0.366	0.152	0.214	0.359	0.192	**0.250**
Neutral	0.706	0.846	**0.770**	0.559	0.676	0.612	0.551	0.668	0.604
Macro Avg.	0.595	0.633	**0.605**	0.571	0.586	0.573	0.525	0.533	0.523

Table V
PRECISION, RECALL AND F1-VALUES OBTAINED BY SMO MULTI-CLASSIFIER ON THE CORPUS DEVELOPED APPLYING WORD2VEC MODEL (CBOW ARCHITECTURE) BUILT FROM BNC AND WACKYPEDIA/UKWAC AS SEMANTIC METRIC IN THE EXTENSION OF THE SEED.

	ukWak W2V (CBOW)								
	Original approach			Enriched approach WN			Enriched approach Oxford		
	P	R	F1	P	R	F1	P	R	F1
Anger	0.184	0.152	0.167	0.413	0.330	**0.367**	0.360	0.356	0.358
Disgust	0.121	0.047	0.067	0.350	0.286	**0.315**	0.200	0.098	0.132
Fear	0.289	0.179	0.221	0.409	0.282	**0.334**	0.336	0.219	0.265
Joy	0.507	0.586	0.544	0.680	0.796	**0.733**	0.520	0.600	0.557
Sadness	0.307	0.226	0.260	0.406	0.241	0.303	0.552	0.586	**0.568**
Surprise	0.345	0.185	0.241	0.294	0.103	0.153	0.376	0.229	**0.285**
Neutral	0.608	0.702	0.652	0.587	0.573	0.580	0.596	0.554	**0.574**
Macro Avg.	0.475	0.504	0.483	0.586	0.610	**0.592**	0.511	0.518	0.511

Table VI
PRECISION, RECALL AND F1-VALUES OBTAINED BY SMO MULTI-CLASSIFIER ON THE CORPUS DEVELOPED APPLYING WORD2VEC MODEL (CBOW ARCHITECTURE) BUILT FROM GIGAWORD AS SEMANTIC METRIC IN THE EXTENSION OF THE SEED.

	Gigaword W2V CBOW)								
	Original approach			Enriched approach WN			Enriched approach Oxford		
	P	R	F1	P	R	F1	P	R	F1
Anger	0.113	0.074	0.089	0.541	0.400	**0.460**	0.399	0.385	0.392
Disgust	0.250	0.052	0.086	0.262	0.129	**0.173**	0.235	0.080	0.119
Fear	0.419	0.233	0.300	0.387	0.287	**0.329**	0.300	0.172	0.219
Joy	0.554	0.423	0.480	0.674	0.706	**0.690**	0.550	0.556	0.553
Sadness	0.305	0.105	0.157	0.496	0.298	0.372	0.554	0.459	**0.502**
Surprise	0.407	0.222	**0.287**	0.406	0.160	0.230	0.338	0.150	0.208
Neutral	0.719	0.876	**0.790**	0.591	0.679	0.632	0.540	0.648	0.589
Macro Avg.	0.618	0.660	**0.627**	0.590	0.604	0.592	0.508	0.520	0.508

Table VII
PRECISION, RECALL AND F1-VALUES OBTAINED BY SMO MULTI-CLASSIFIER ON THE CORPUS DEVELOPED APPLYING WORD2VEC MODEL (SKIP ARCHITECTURE) BUILT FROM GIGAWORD AS SEMANTIC METRIC IN THE EXTENSION OF THE SEED.

	Gigaword W2V (SKIP)								
	Original approach			Enriched approach WN			Enriched approach Oxford		
	P	R	F1	P	R	F1	P	R	F1
Anger	0.139	0.094	0.112	0.465	0.354	**0.402**	0.383	0.385	0.384
Disgust	0.176	0.037	0.061	0.365	0.256	**0.301**	0.160	0.073	0.100
Fear	0.336	0.223	0.268	0.388	0.286	**0.329**	0.257	0.167	0.203
Joy	0.528	0.597	0.560	0.688	0.748	**0.717**	0.557	0.609	0.582
Sadness	0.273	0.143	0.188	0.435	0.213	0.286	0.544	0.417	**0.472**
Surprise	0.353	0.156	0.217	0.250	0.092	0.134	0.359	0.168	**0.229**
Neutral	0.629	0.751	**0.685**	0.538	0.636	0.583	0.485	0.595	0.534
Macro Avg.	0.507	0.545	0.517	0.573	0.593	**0.577**	0.482	0.493	0.482

Table VIII
COHEN'S KAPPA VALUES OBTAINED BY LSA AND UKWAK W2V (CBOW) (THE ORIGINAL APPROACH AND THE ENRICHED APPROACHES) IN THE COMPARISON OF THEIR ANNOTATIONS TO THE GOLD OF AMAN CORPUS.

	Cohen's kappa values					
	LSA			ukWak W2V (CBOW)		
	Original	Enriched WN	Enriched Oxford	Original	Enriched WN	Enriched Oxford
Anger	**0.9368**	0.9051	0.8882	**0.9193**	0.9004	0.8713
Disgust	0.9495	0.9417	**0.9537**	0.9452	0.9392	**0.9529**
Fear	0.9226	0.8919	**0.9323**	0.9315	0.9099	**0.9328**
Joy	**0.7719**	0.6041	0.7241	0.6987	0.5359	**0.7219**
Sadness	**0.9285**	0.9193	0.8033	0.8750	**0.9119**	0.7425
Surprise	0.9186	**0.9512**	0.9345	0.9014	**0.9522**	0.9338

Table IX
COHEN'S KAPPA VALUES OBTAINED BY GIGAWORD W2V (CBOW) AND GIGAWORD W2V (SKIP) (THE ORIGINAL APPROACH AND THE ENRICHED APPROACHES) IN THE COMPARISON OF THEIR ANNOTATIONS TO THE GOLD OF AMAN CORPUS.

	Cohen's kappa values					
	Gigaword W2V (CBOW)			Gigaword W2V (SKIP)		
	Original	Enriched WN	Enriched Oxford	Original	Enriched WN	Enriched Oxford
Anger	**0.9430**	0.9089	0.8875	**0.9328**	0.9044	0.8675
Disgust	0.9507	0.9430	**0.9527**	0.9460	0.9412	**0.9514**
Fear	**0.9380**	0.9136	0.9343	0.9106	0.9009	**0.9223**
Joy	**0.8053**	0.6414	0.7443	**0.7281**	0.5752	0.6942
Sadness	**0.9340**	0.9173	0.8396	**0.9131**	0.9066	0.8230
Surprise	0.9368	**0.9557**	0.9325	0.9146	**0.9509**	0.9295

E. Discussion

Concerning agreement values (Tables VIII-IX), the results obtained demonstrate a high level of agreement between automatic annotations and the gold standard of Aman corpus for each emotion except `JOY` emotion. This is due to the fact that the process of the creation of the seed annotates a few sentences incorrectly and this error is expanded in the second part of bootstrapping process. For this reason, the number of sentences annotated with `JOY` are higher than they should be and the agreement is worse. The errors in the seed may be caused by the use of Emolex, a general domain resource, and not address the ambiguity problems.

Regarding F1-values obtained by the three approaches (original, enriched by WN and enriched by Oxford), these can be considered promising, since their best F1-values are near 60%, obtaining the best value of 62.7% with Gigaword W2V (CBOW). Although, these results do not improve the values obtained by Aman corpus (74.5%), the three approaches are considered encouraging because the corpora employed have been developed with a totally automatic process.

More concretely, comparing the original approach to the enriched ones, the results show the need to improve the creation of the seed to increase the F1-values for each emotion. All of the emotions have improved their F1-values in the enriched approaches, except the `SURPRISE` emotion in Gigaword W2V (CBOW) which obtains the best F1-value in original approach. In this case, the Oxford and WN synonyms added to Emolex for `SURPRISE` emotion are not beneficial when the Gigaword W2V (CBOW) model is employed because these synonyms could be semantically similar to other words not related with the `SURPRISE` emotion. Hence, the training data employed in this case could contain incorrect sentence annotated with `SURPRISE` emotion.

Finally comparing the enriched approaches, the results show the improvements obtained by WN synonyms for `ANGER`, `DISGUST`, `FEAR` and `JOY` emotions, and the improvements of F1-values for `SADNESS` and `SURPRISE` emotions obtained by Oxford synonyms. These improvements are shown regardless the semantic similarity model employed. Thus, we consider an extension employing WN synonyms or Oxford synonyms depending on each emotion could be the best solution.

V. CONCLUSION

In this paper, we exploit a bootstrapping approach to automatically annotate emotional corpora which allows us to address the need to annotate emotional corpora and to improve supervised learning techniques. Moreover, we present two enriched approaches focused on analysing the creation of the seed, extending Emolex with two sets of synonyms.

The evaluation performed has demonstrated the contributions of our approach for emotional annotation task, since the results obtained by emotional models and agreement metrics are promising. That shows it is possible to create a good emotion model and there is an agreement between automatic and manual annotations. Thus, the results allow us to verify the integrity of the bootstrapping process to annotate emotional corpora automatically.

Our main conclusions are that the results confirm the soundness of the proposed approach for automatic annotation and the relevance of the extension of Emolex with a set of synonyms to improve the results. Hence, our proposal allows us to consider the bootstrapping process as a good approach to create resources such as emotional corpora useful to be employed on supervised machine learning, without developing a hard and time-consuming annotation task. Thus, the bootstrapping process could help us to improve the current emotion detection systems for the generation of emotional and personality profiles.

Our future research will deal with exploring this bootstrapping process in other corpora to verity the results; analysis of the process to create a more accurate seed to resolve the ambiguity problems.; testing new semantic similarity metric like GloVe: Global Vectors for Word Representation[3]; 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] P. Rodríguez, A. Ortigosa, and R. M. Carro, "Extracting Emotions from Texts in E-Learning Environments." in *Complex, Intelligent and Software Intensive Systems (CISIS)*, L. Barolli, F. Xhafa, S. Vitabile, and M. Uehara, Eds. IEEE Computer Society, 2012, pp. 887–892. [Online]. Available: http://dblp.uni-trier.de/db/conf/cisis/cisis2012.html#RodriguezOC12

[2] B. Desmet and V. Hoste, "Emotion Detection in Suicide Notes," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6351–6358, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2013.05.050

[3] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," in *Computational Intelligence*, vol. 29, no. 3, 2013, pp. 436–465.

[3] http://nlp.stanford.edu/projects/glove/

[4] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. MIT Press, 2012.

[5] C. O. Alm, "Emotions from text: Machine learning for text-based emotion prediction," in *In Proceedings of HLT/EMNLP*, 2005, pp. 347–354.

[6] S. Aman and S. Szpakowicz, "Identifying Expressions of Emotion in Text." in *Proceedings of the 10th International Conference on Text, Speech and Dialogue TSD 2007*, ser. Lecture Notes in Computer Science, V. Matousek and P. Mautner, Eds., vol. 4629. Springer, 2007, pp. 196–205. [Online]. Available: http://dblp.uni-trier.de/db/conf/tsd/tsd2007.html#AmanS07

[7] C. Strapparava and R. Mihalcea, "Learning to Identify Emotions in Text," in *Proceedings of the 2008 ACM Symposium on Applied Computing*, ser. SAC '08. New York, NY, USA: ACM, 2008, pp. 1556–1560. [Online]. Available: http://doi.acm.org/10.1145/1363686.1364052

[8] R. C. 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.

[9] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Affect Analysis Model: Novel Rule-based Approach to Affect Sensing from Text," *Natural Language Engineering*, vol. 17, no. 1, pp. 95–135, 2011. [Online]. Available: http://dx.doi.org/10.1017/S1351324910000239

[10] C. E. Izard, *The face of emotion*. Appleton-Century-Crofts, New York :, 1971.

[11] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Recognition of Affect, Judgment, and Appreciation in Text," in *Proceedings of the 23rd International Conference on Computational Linguistics*, ser. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 806–814. [Online]. Available: http://dl.acm.org/citation.cfm?id=1873781.1873872

[12] E. Boldrini and P. Martínez-Barco, "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. dissertation, University of Alicante, 2012.

[13] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," *Information Processing and Management*, vol. 51, no. 4, pp. 480 – 499, 2015.

[14] J. S. Y. Liew, H. R. Turtle, and E. D. Liddy, "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.

[15] S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," *Computational Intelligence*, vol. 31, no. 2, pp. 301–326, 2015. [Online]. Available: http://dx.doi.org/10.1111/coin.12024

[16] M. D. Choudhury, M. Gamon, and S. Counts, "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 - ICWSM 2012*. Association for the Advancement of Artificial Intelligence, 2012. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=167851

[17] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing Twitter "Big Data" for Automatic Emotion Identification," in *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, ser. SOCIALCOM-PASSAT '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 587–592. [Online]. Available: http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.119

[18] M. Hasan, E. Rundensteiner, and E. Agu, "EMOTEX: Detecting Emotions in Twitter Messages," in *ASE BIG-DATA/SOCIALCOM/CYBERSECURITY Conference*, 2014, pp. 27–31.

[19] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39(6), pp. 1161–1178, 1980.

[20] D. Yarowsky, "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," in *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, ser. ACL '95. Stroudsburg, PA, USA: Association for Computational Linguistics, 1995, pp. 189–196. [Online]. Available: http://dx.doi.org/10.3115/981658.981684

[21] M. Collins and Y. Singer, "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*, 1999, pp. 100–110.

[22] S. Chowdhury and W. Chowdhury, "Performing Sentiment Analysis in Bangla Microblog Posts," in *International Conference on Informatics, Electronics & Vision (ICIEV), 2014*. IEEE, 2014.

[23] S. Lee and G. G. Lee, "A Bootstrapping Approach for Geographic Named Entity Annotation," in *Asia Information Retrieval Symposium (AIRS)*, 2004, pp. 178–189.

[24] P. Ekman, "Basic emotions," in *Handbook of cognition and emotion*, 1999, pp. 45–60.

[25] F. Keshtkar and D. Inkpen, "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*, ser. CAAGET '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 35–44. [Online]. Available: http://dl.acm.org/citation.cfm?id=1860631.1860636

[26] S. Chaffar and D. Inkpen, *Using a Heterogeneous Dataset for Emotion Analysis in Text*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 62–67. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21043-38

[27] S. Mohammad, "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*, ser. NAACL HLT '12. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012, pp. 587–591. [Online]. Available: http://dl.acm.org/citation.cfm?id=2382029.2382123

[28] R. Plutchik, "Emotion: Theory, Research and Experience," in *Theories of emotion*. Academic Press, 1980, vol. 11, no. 01, p. 399.

[29] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The {Stanford} {CoreNLP} Natural Language Processing Toolkit," in *Association for Computational Linguistics (ACL) System Demonstrations*, 2014, pp. 55–60. [Online]. Available: http://www.aclweb.org/anthology/P/P14/P14-5010

[30] A. Gliozzo, C. Strapparava, and I. D. O. Dagan, "Improving Text Categorization Bootstrapping via Unsupervised Learning," *ACM Transactions on Speech and Language Processing*, vol. 6, no. 1, 2009.

[31] A. Gliozzo and C. Strapparava, *Semantic Domains in Computational Linguistics*. Springer-Verlag Berlin Heidelberg, 2009.

[32] G. Dinu and M. Baroni, "Improving zero-shot learning by mitigating the hubness problem," *CoRR*, vol. abs/1412.6, 2014. [Online]. Available: http://arxiv.org/abs/1412.6568

[33] J. Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines," in *Proc. Advances in Neural Information Processing Systems 11*, 1999, pp. 557–563. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=68525

[34] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656278

[35] G. A. Miller, "WordNet: A Lexical Database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995. [Online]. Available: http://doi.acm.org/10.1145/219717.219748

[36] D. Aubur, R. Armantrout, D. Crystal, and M. Dirda, *Oxford American Writer's Thesaurus*. Oxford University Press, 2004.

[37] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, p. 37, 1960.