ArSEL: A Large Scale Arabic Sentiment and Emotion Lexicon

Gilbert Badaro, Hussein Jundi, Hazem Hajj, Wassim El-Hajj, Nizar Habash[†]
American University of Beirut, Beirut, Lebanon
[†]New York University Abu Dhabi, Abu Dhabi, UAE
ggb05@aub.edu.lb, haj14@mail.aub.edu, {hh63,we07}@aub.edu.lb, nizar.habash@nyu.edu

Abstract

With the advancement of Web 2.0, social networks experienced a great increase in the number of active users reaching 2 billion active users on Facebook at the end of 2017. Consequently, the size of text data on the Internet increased tremendously. This textual data is rich in knowledge, which attracted many data scientists as well as computational linguists to develop resources and models to automatically process the data and extract useful information. One major interest is sentiment and emotion classification from text. In fact, learning the opinion and emotions of people is important for businesses, marketers, government, politicians, etc. While focus had been given to sentiment analysis, recently emotion analysis has captured great interest as well. Several resources were developed for emotion analysis from text for English, however, very few targeted Arabic text. We present in this paper, ArSEL, the first large scale **Ar**abic **S**entiment and **E**motion **L**exicon. ArSEL is built in a way to augment the publicly available Arabic Sentiment Lexicon, ArSenL, and to generate a large scale lexicon that includes emotion and sentiment labels for almost every lemma in ArSenL. We also show the efficiency of using ArSEL in emotion regression and classification tasks using an Arabic translated version of annotated data from SemEval 2007 "Affective Task" as well as SemEval 2018 Task1 "Affect in Tweets" Arabic dataset. Coverages of 91% and 84% are achieved on the two datasets respectively. An improvement of 30% compared to majority baseline is achieved in terms of average F1 measure for emotion classification on SemEval 2018 Arabic dataset. ArSEL is publicly available on `http://oma-project.com`.

Keywords: Emotion Lexicon, Arabic Natural Language Processing, Emotion Classification, Regression

1. Introduction

The task of emotion recognition has been extensively studied from different modalities. For instance, several researchers tried to predict users' emotion by looking at their interaction with computers (Cowie et al., 2001; Pantic and Rothkrantz, 2003; Brave and Nass, 2003; Fragopanagos and Taylor, 2005; Jaimes and Sebe, 2007; Hibbeln et al., 2017; Patwardhan and Knapp, 2017; Constantine et al., 2016). Others have tried to assign to facial expressions emotion labels (Busso et al., 2004; Goldman and Sripada, 2005; Gunes and Piccardi, 2007; Trad et al., 2012; Wegrzyn et al., 2017). Recently, with the increase of textual data on the Web, computational linguists and data scientists started looking at emotion analysis from text. In fact, recognizing emotions of users is critical for different applications: first, it helps businesses and companies sense the feedback of its clients expressed on the Internet and consequently adapt their marketing strategies (Bougie et al., 2003); second, it allows providing customers with better personalized recommendations whether for advertisements or products (Mohammad and Yang, 2011) on top of collaborative filtering based recommender systems (Badaro et al., 2013; Badaro et al., 2014c; Badaro et al., 2014d); third, it can help in tracking emotions of users towards politicians, movies, music, products, etc, (Pang et al., 2008); fourth, it allows developing complex search algorithms that provide advanced search features filtered by emotions (Knautz et al., 2010) and last but not least, it allows a more accurate prediction of stock market prices (Bollen et al., 2011).

Some efforts have already been placed in developing emotion classification models from text (Shaheen et al., 2014; Houjeij et al., 2012; Abdul-Mageed and Ungar, 2017; Felbo et al., 2017). Since sentiment lexicons helped in improving the accuracy of sentiment classification models (Liu and Zhang, 2012; Taboada et al., 2011), several researchers are working on developing emotion lexicons for different languages such as English, French and Chinese (Mohammad, 2017; Bandhakavi et al., 2017; Yang et al., 2007; Poria et al., 2012; Mohammad and Turney, 2013; Das et al., 2012; Mohammad et al., 2013; Abdaoui et al., 2017; Staiano and Guerini, 2014). While sentiment is usually represented by three labels namely positive, negative or neutral, several representation models exist for emotions such as Ekman representation (Ekman, 1992) or Plutchik model (Plutchik, 1980; Plutchik, 1994) that includes Ekman's six emotions in addition to two labels: trust and anticipation. Despite the efforts for creating large scale emotion lexicons for English, the size of existing lexicons remain much smaller compared to sentiment lexicons. For example, DepecheMood (Staiano and Guerini, 2014), one of the largest publicly available emotion lexicon for English, includes around 37K while SentiWordNet (SWN) (Esuli and Sebastiani, 2007; Baccianella et al., 2010), a large scale English sentiment lexicon semi-automatically generated using English WordNet (EWN) (Fellbaum, 1998), includes around 150K terms annotated with three sentiment scores: positive, negative and objective. While some efforts have already been placed for developing emotion lexicons for English, we were only able to find two attempts for Arabic where the first emotion lexicon is a Google translation of an English Emotion lexicon, Emolex (Mohammad and Turney, 2013; Mohammad et al., 2013) and the second one is extracted from manually annotated Arabic documents for emotions (El Gohary et al., 2013). In fact, more work can be found related to sentiment analysis classification models for Arabic such as the work in (Badaro et al., 2014b; Badaro et al., 2015; Al Sallab et al., 2015;

Al-Sallab et al., 2017; Baly et al., 2017b; Abdul-Mageed, 2017) and to Arabic sentiment lexicon developments such as ArSenL (Badaro et al., 2014a), SIFAAT (Abdul-Mageed and Diab, 2012) and SANA (Abdul-Mageed and Diab, 2014). Developing emotion and sentiment classification models for Arabic is important given the tremendous increase of Arabic speaking users of Web 2.0. For instance, more than 11 million users are active on Twitter within the 22 Arab countries and more than 27 million tweets are generated daily.[1] Moreover, analyzing Arabic Twitter is a more complex task than MSA given that it includes different dialects with different characteristics (Baly et al., 2017a). Since the usage of sentiment lexicons in sentiment classification models showed significant improvement in the accuracy of such models (Al-Sallab et al., 2017), it is necessary to develop Arabic emotion lexicons for improved emotion classification models.

In this paper, we present ArSEL, the first publicly available large scale Arabic sentiment and emotion lexicon. ArSEL is an extension of ArSenL, where almost each lemma[2] in ArSenL is amended by eight emotion scores corresponding to: *afraid, amused, angry, annoyed, don't care, happy, inspired* and *sad*. The emotion scores are automatically obtained from DepecheMood (Staiano and Guerini, 2014), one of the largest publicly available English emotion lexicon. We first align DepecheMood with English WordNet (Fellbaum, 1998) and then, using synonymy semantic relation, we expand the coverage of DepecheMood and obtain EWN synsets annotated with emotion scores. Since ArSenL is linked to EWN 3.0, we can automatically assign the synsets' emotion scores to ArSenL lemmas. ArSEL can be used for several NLP tasks such as sentiment analysis, emotion analysis, or other semantic extraction tasks. It would be in particular useful for cases where it is desired to simultaneously extract the sentiment and emotion scores for words. In order to test the efficiency of ArSEL, we utilize ArSEL in emotion regression and classification tasks using unsupervised techniques similar to the way the efficiency of DepecheMood was tested with SemEval 2007 Affective Task dataset (Strapparava and Mihalcea, 2007). We also test the usefulness of ArSEL on a native Arabic dataset from SemEval 2018 Task1 "Affect in Tweets".[3]

The paper is organized as follows: in section 2, we present a literature review about emotion lexicon development. In section 3, we describe the approach followed for constructing ArSEL. In section 4, we evaluate ArSEL in emotion regression and classification tasks using first, SemEval 2007 news headlines data translated from English to Arabic using Google translate and second, SemEval 2018 Arabic Affect Tweets. We conclude the results of the paper in section 5 and present some ideas for future work.

[1] https://weedoo.tech/twitter-arab-world-statistics-feb-2017/

[2] For more information on issues of Arabic morphology in natural language processing, see (Habash, 2010).

[3] https://competitions.codalab.org/competitions/17751

2. Literature Review

We conduct a literature review on existing emotion lexicons for multiple languages. We present the techniques used to build the lexicons and the methods employed for evaluating their efficiency in emotion recognition tasks.

Strapparava et al. (2004) developed WordNet Affect by tagging specific synsets with affective meanings in EWN. They identified first a core number of synsets that represent emotions of a lexical database. They expanded then the coverage of the lexicon by checking semantically related synsets compared to the core set. They were able to annotate 2,874 synsets and 4,787 words. WordNet Affect was also tested in different applications such as affective text sensing systems and computational humor. WordNet Affect is of good quality given that it was manually created and validated, however, it is of limited size.

Mohammad and Turney (2013) presented challenges that researchers face for developing emotion lexicons and devised an annotation strategy to create a good quality and inexpensive emotion lexicon, EmoLex, by utilizing crowdsourcing. To create EmoLex, the authors first identified target terms for annotation extracted from Macquarie Thesaurus (Bernard and Bernard, 1986), WordNet Affect and the General Inquirer (Stone et al., 1966). Then, they launched the annotation task on Amazon's Mechanical Turk. EmoLex has around 10K terms annotated for emotions as well as for sentiment polarities. They evaluated the annotation quality using different techniques such as computing inter-annotator agreement and comparing a subsample of EmoLex with existing gold data. Moreover, they utilized Google translate to perform word translations into multiple languages including Arabic (Mohammad et al., 2013). However, the translation may include several errors: first, the translation may be incorrect since it is a word to word translation and second, the translation may be a transliteration instead in case the word is seen for the first time by the machine translator. Furthermore, the terms in the lexicon are not in their lemma form which make the lexicon harder to be utilized in an emotion classification task.

AffectNet (Cambria et al., 2012), part of the SenticNet project, includes also around 10K terms extracted from ConceptNet (Liu and Singh, 2004) and aligned with WordNet Affect. They extended WordNet Affect using the concepts in ConceptNet. While WordNet Affect, EmoLex and AffectNet include terms with emotion labels, Affect database (Neviarouskaya et al., 2007) and DepecheMood (Staiano and Guerini, 2014) include words that have emotion scores instead. Affect database extends SentiFul (Neviarouskaya et al., 2011) and covers around 2.5K words presented in their lemma form along with the corresponding part of speech tag.

DepecheMood is automatically built by harvesting social media data that were implicitly annotated with emotions. They utilize news articles from rappler.com. The articles are accompanied by Rappler's Mood Meter, which allows readers to express their emotions about the article they are reading. DepecheMood includes around 37K lemmas along with their part of speech (POS) tags and the lemmas are aligned with EWN. Staiano and Guerini also evaluated DepecheMood in emotion regression and classification tasks in unsupervised settings. They claim that, although they utilized a naïve unsupervised model, they were able to outperform existing lexicons when tested on SemEval 2007 dataset (Strapparava and Mihalcea, 2007).

Bandhakavi et al. worked on constructing emotion lexicons using Tweets annotated with emotion labels (Bandhakavi et al., 2014; Bandhakavi et al., 2017). They experiment different techniques for lexicon generation: term frequency models and iterative models including generative and expectation maximization algorithms. Bandhakavi et al. evaluated the different lexicons on a Twitter dataset (Wang et al., 2012) and utilized a feature based supervised approach for classifying emotion.

While the above emotion lexicons were mainly developed for English, Yang et al. (2007) constructed an emotion lexicon for Chinese language. The authors used web blog corpora in order to extract the lexicon terms and assigned emotion scores using point wise mutual information measure. They created two different lexicons by varying the number of documents downloaded from the Web. They also evaluated the lexicons in an emotion classification task using different prediction methods.

Xu et al. (2010) also worked on constructing emotion lexicon for Chinese using graph-based algorithm which ranks words according to a few seed emotion words. The graph algorithm utilizes different similarity measures derived from dictionaries, unlabeled corpora and heuristic rules. In order to improve the quality of the lexicon, they mixed manual verification with the automatic assignment of emotions.

Abdaoui et al. (2017) presented Feel, an emotion and sentiment lexicon for French. Abdaoui et al. utilized NRC emotion lexicon (Mohammad et al., 2013) and translated its terms to French using multiple online translators. Then, a professional human translator validated the translation along with their emotion labels. Abdaoui et al. also claimed that FEEL outperformed other French emotion lexicons in emotion classification from texts.

In summary, several techniques are employed for building emotion lexicons and can be mainly grouped into two categories: the first one is based on manual annotation provided by professional individuals or through crowdsourcing, the second technique is rather automatic and lexicons are derived from annotated corpora. Only couple of papers worked on developing emotion lexicon for Arabic, thus, we focus on developing a large-scale Arabic emotion lexicon. We present next the methodology followed to construct automatically ArSEL by utilizing DepecheMood, EWN and ArSenL.

3. ArSEL

We describe in this section the process followed to construct ArSEL. We first briefly describe the harvested resources. Then, we present the expansion technique of DepecheMood and how we link it to ArSenL.

3.1. Resources

We make use of three resources: DepecheMood, English WordNet and ArSenL.

DepecheMood: (Staiano and Guerini, 2014) an emotion lexicon for English consisting of 37,771 words aligned with English WordNet. Each word along with its corresponding part of speech tag is annotated with 8 emotion scores (afraid, amused, angry, annoyed, don't care, happy, inspired and sad) derived automatically from annotated corpora collected from Rappler.com news website. Three variations of the lexicon were presented where the differences are related to the method of normalizing the emotion scores.

English WordNet 3.0: (Fellbaum, 1998; Fellbaum, 2010) is a hierarchical dictionary including more than 117,000 synsets and around 150,000 terms distributed among four part of speech tags: noun, verb, adjective and adverb. EWN has been used extensively in multiple natural language processing tasks and also for developing sentiment lexicons such as SentiWordNet (Esuli and Sebastiani, 2007; Baccianella et al., 2010) and emotion lexicons such as WordNet Affect (Strapparava et al., 2004).

ArSenL: (Badaro et al., 2014a) is a free publicly available large-scale Arabic Sentiment lexicon. ArSenL consists of Arabic lemmas assigned to EWN synsets along with three sentiment scores derived from English SentiWordNet. ArSenL was automatically developed by taking the union of two sentiment lexicons: the first one maps Arabic WordNet 2.0 (Black et al., 2006) to English SentiWordNet by using WordNet sense map files across WordNet versions 2.0, 2.1 and 3.0. The second lexicon is the result of performing gloss matching between English gloss terms of an Arabic lexical resource, SAMA (Standard Arabic Morphological Analyzer) (Graff et al., 2009), and EWN synset terms. In both sub-lexicons, the sentiment scores are obtained from English SentiWordNet. ArSenL includes 153,638 Arabic lemma-EWN synset pairs corresponding to 33,995 Arabic lemmas/POS tags annotated with three sentiment scores: positive, negative and objective.

We choose DepecheMood since it is the largest publicly available emotion lexicon in English and its terms are aligned with English WordNet. We benefit from the available alignment with English WordNet to expand the coverage of DepecheMood and obtain emotion scores for EWN synsets, in addition to emotion scores for an expanded list of EWN terms compared to those already in DepecheMood. We also utilize the advantage that ArSenL is connected to EWN synsets and hence, we automatically assign emotion scores of EmoWordNet to corresponding ArSenL entries.

3.2. Expansion of DepecheMood and Link to ArSenL

In Figure 1, we show an overview of the steps followed to expand DepecheMood into EmoWordNet (steps grouped under DepecheMood Expansion) and then linking EmoWordNet to ArSenL to obtain ArSEL.

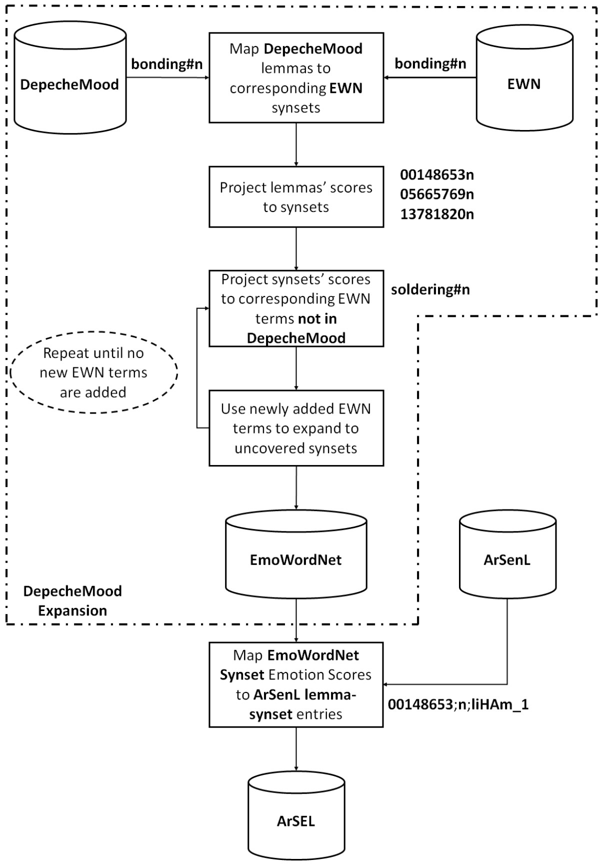

Figure 1: Overview of ArSEL Construction Methodology.

We detail first the steps utilized for expanding DepecheMood iteratively into what we name EmoWordNet.

Step 1: EWN synsets that include lemmas of DepecheMood are retrieved. A score is then computed for each retrieved synset, s. Let S denotes the set of all such synsets. Two cases may appear: either the retrieved synset includes only one lemma from DepecheMood, in this case the synset gets the same score of the lemma, or, the synset includes multiple lemmas, in this case the score is the average of the scores of the corresponding lemmas. A synset, s, includes two sets of terms, T, terms that are in DepecheMood, and $\bar{T}$, terms **not in** DepecheMood.

Step 2: using the synonymy semantic relation in EWN, and based on the concept that synonym words will likely share the same emotion scores, we assign the synset scores to its corresponding terms $\bar{T}$. Again, a term t in $\bar{T}$ may appear in one or multiple synsets from S. Hence, the score assigned to t will be either the one of its corresponding synset or the average of the scores of its corresponding synsets that belong to S.

Step 3: terms in $\bar{T}$ may also appear in synsets $\bar{s}$ that do not belong to S. $\bar{s}$ will get the score of its corresponding terms.

Step 2 and 3 are repeated until no new terms or synsets are added and scores of added terms converged. It is important to note that we decided to consider only synonyms for expansion since synonymy is the only semantic relation that preserves the emotion orientation and does not require manual validation (Strapparava et al., 2004).

Using the described automatic expansion approach, we were able to extend the size of DepecheMood by a factor of 1.8. We obtained emotion scores for an additional 29,967 EWN terms and for 59,952 EWN synsets. Overall, we construct EmoWordNet, an emotion lexicon consisting of 67,738 EWN terms and of 59,952 EWN synsets annotated with emotion scores.

Next, we match ArSenL entries to EmoWordNet synsets. Each entry in ArSenL consists mainly of an Arabic SAMA lemma, a corresponding POS tag, a corresponding EWN synset and three sentiment scores extracted from SentiWordNet. For each entry in ArSenL, if its assigned synset is found in EmoWordNet, emotion scores of the synset are automatically added to ArSenL entry. We were able to assign emotion scores to 149,634 ArSenL entries corresponding to 32,196 Arabic lemmas, i.e., 94.71% of ArSenL lemmas. We summarize the lexicon sizes per lemma in Table 1. We also show some sample lemmas of ArSEL along their corresponding 8 emotion scores in Table 3. We have picked samples that should be emotionally charged to check if the emotions represented by the lemma have the highest scores.

As a walking example of the steps described above, we added to the steps shown in Fig. 1 an example corresponding to each step. For instance, the DepecheMood term "bonding" having noun as POS tag is mapped to EWN term "bonding" with the same POS tag. "bonding" appears in three different noun synsets in EWN with the following offset IDs: "00148653; 05665769; 13781820". Since "bonding" is the only term having a DepecheMood representation in the three synsets, the three synsets will have the same emotion scores as "bonding". While synsets "05665769; 13781820" have only the term "bonding", "00148653" includes as well the lemma "soldering" which is not in DepecheMood. Thus, from step 2, "soldering" will have the same scores as "bonding". "soldering" does not appear in any other synset so there are no more iterations. The next step is to check if the retrieved synsets appear in ArSenL. For example, "00148653" corresponds to the lemma "liHAm" and hence the Arabic lemma will be assigned the emotion scores of the synset.

To test the efficiency of our emotion lexicon ArSEL, we evaluate in the next section the performance of ArSEL when employed in emotion regression and classification tasks.

Lexicon	Lemma Count
DepecheMood	37,771
EmoWordNet	67,738
ArSenL	33,995
ArSEL	32,196

Table 1: Lexicons Coverage.

SemEval	ArSEL
Fear	Afraid
Anger	Angry
Joy	Happy
Sadness	Sad
Surprise	Inspired
Disgust	-
-	Annoyed, Amused, Don't Care

Table 2: Mapping between SemEval and ArSEL Emotion Labels.

4. ArSEL Evaluation

Since ArSEL is generated based on ArSenL, the intrinsic evaluation results of ArSenL described in (Badaro et al., 2014a) are automatically inherited by ArSEL. Therefore, we focus in this section on performing extrinsic evaluation of ArSEL. We describe next the dataset used, the experiment setup, the regression and the classification results for the two datasets: SemEval 2007 and 2018 datasets.

4.1. Using SemEval 2007 Dataset

4.1.1. About the Dataset

We utilize SemEval 2007 Affective Task dataset (Strapparava and Mihalcea, 2007). The dataset consists of one thousand news headlines annotated with six emotion scores: anger, disgust, fear, joy, sadness and surprise. For the regression task, a score between 0 and 1 is provided for each emotion. For the classification task, a threshold is applied on the emotion scores to get a binary representation of the emotions: if the score of a certain emotion is greater than 0.5, the corresponding emotion label is set to 1, otherwise it is 0. The emotion labels used in the dataset correspond to the six emotions of the Ekman model (Ekman, 1992) while those in ArSEL, EmoWordNet and DepecheMood follow the ones provided by Rappler Mood Meter. We consider the same assumptions of emotion mapping presented in the work of (Staiano and Guerini, 2014) and summarized in Table 2. Disgust emotion label in SemEval is not aligned with any emotion in EmoWordNet and hence is discarded as also assumed in (Staiano and Guerini, 2014). The dataset is in English, thus, we use Google translate to translate it automatically to Arabic. Some examples of the news' headlines along with their Google and Human translations are shown in Table 4.

4.1.2. Experiment Setup

We perform the following preprocessing steps in order to proceed with the evaluation. We utilize MADAMIRA (Pasha et al., 2014) in order to perform lemmatization for the translated dataset. The output of MADMIRA is a list of lemmas in Buckwalter transliteration (Buckwalter, 2002) along with the corresponding POS tag. We exclude lemmas that do not belong to the main four POS tags: noun, verb, adjective and adverb. It is important to note that MADAMIRA generates many fine-grained POS tags that can be grouped into the above mentioned four POS tags. On ArSEL side, we compute the average of emotion scores per lemma since an Arabic lemma can be mapped to multiple EWN synsets. Next, we compute for each news' headline the sum and the average of emotion scores. The average turned out to give better results. For the regression task, we compute Pearson correlation coefficient between the computed headline emotion scores and the scores provided in SemEval taking into consideration the mapping of emotion labels as represented in Table 2. For the classification task, we first perform min-max normalization on the computed scores and then we apply thresholding with a threshold equals to 0.5. Thus, an emotion label will be set to 1 if its corresponding emotion score is greater than 0.5, otherwise it will be set to 0. The same thresholding is applied on SemEval scores. F1 measure is then computed to evaluate classification of emotions. The experiment process is summarized in Figure 2.

4.1.3. Regression and Classification Results

We present first the coverage results of ArSEL for the translated SemEval dataset. Only one headline ("Toshiba Portege R400", "توشيبا برتجي ر٤٠٠") did not include a lemma that matched to ArSEL. In terms of lemma counts, 2,688 unique lemmas represent the dataset. 301 lemmas were not identified by MADAMIRA, 121 lemmas had POS tags different than the four main ones and 2,266 lemmas were within the four POS tags: N, V, Adj and Adv. To evaluate the coverage of ArSEL, we compare ArSEL lemmas to the 2,266 lemmas that are within the main four POS tags. 91.41% of the 2,266 lemmas were found in ArSEL. Thus, we can conclude that ArSEL includes commonly used Arabic lemmas with a high coverage.

In Table 7, Pearson correlation results are presented when using ArSEL and when using EmoWordNet on the translated SemEval Dataset and the original one respectively. We notice that the performance of ArSEL is very similar to EmoWordNet. The small difference in the scores obtained is expected since the automatic Online translation from English to Arabic cannot be guaranteed to be 100% accurate as can be seen in some of the examples shown in Table 4. Moreover, some English words may have an emotion score while their Arabic translation may not be present in ArSEL. In order to check if looking at both the English and Arabic data improves the accuracy of emotion prediction, we combine the two scores obtained from using EmoWordNet on English SemEval 2007 and

Lemma#POS	English Gloss	Afraid	Amused	Angry	Annoyed	Don’t Care	Happy	Inspired	Sad
xawof#n لحام خوف	fear	**0.16866352**	0.10374394	0.13578057	0.11578797	0.09626842	0.10521568	0.12802106	0.14651883
saEAdap#n سعادة	happiness	0.01080941	0.16735222	0.01801752	0.04023918	0.18246141	**0.38946541**	0.16637971	0.02527514
taEAsap#n تعاسة	misery	0.11482094	0.11724791	0.07061617	0.13834278	0.04755821	0.1362515	0.16859612	**0.20656636**
DaHik#v ضحك	laugh	0.04837066	**0.21422647**	0.07150008	0.11078673	0.13726831	0.11134006	**0.21054358**	0.09596412
Huzon#n حزن	grief	0.01551373	0.13148076	0.0687485	0.10431947	0.06042494	0.08809219	0.21824078	**0.31317963**
$ajan#n شجن	anxiety	0.159757	0.08634377	0.10675246	0.10506455	0.11995604	0.14099477	0.05896844	**0.22216298**
maqotal#n مقتل	assault; killing	0.15997316	0.0616973	**0.33435758**	0.10675574	0.06770851	0.07205292	0.03512961	0.16232519
<izoEAj#n ازعاج	disturbance	0.05707528	0.06349826	**0.34656472**	0.14284707	0.11914421	0.11311906	0.06151737	0.096234049
kuwayĩs#a كويس	well	0.0221555	**0.24858529**	0.03319092	0.11484484	0.23663404	0.1073459	0.22167375	0.01556974
$ayo’#n شيء	thing	0.08178512	0.14615643	0.13145998	0.14008017	0.14118469	0.11244018	**0.14626546**	0.10062796

Table 3: Sample of ArSEL Arabic Lemmas with Emotion Scores.

English News’ Headline	Google Translation	Human Translation
Women protest Pakistan demolition	المرأة تحتج على هدم باكستان	المرأة تحتج عل التفجير في باكستان
Dolphins, sea lions may report for duty soon	الدلافين، أسود البحر قد تقرير عن واجب قريبا	الدلافين، أسود البحر توضع في الخدمة قريباً
Woman fights to keep drunken driver in jail	امرأة تحارب للحفاظ على سائق سكران في السجن	إمرأة تحارب لإبقاء سائق سكران في السجن
Female astronaut sets record	سجل رائد فضاء أنثى	رائدة فضاء تسجل رقما قياسيا
Astronaut’s arrest tests NASA’s mettle	اختبارات اعتقال رائد الفضاء ناسا هزة	إعتقال رائد فضاء يضع ناسا تحت الاختبار

Table 4: News’ Headlines’ Examples to Show Differences between Google Translations and Human Translations.

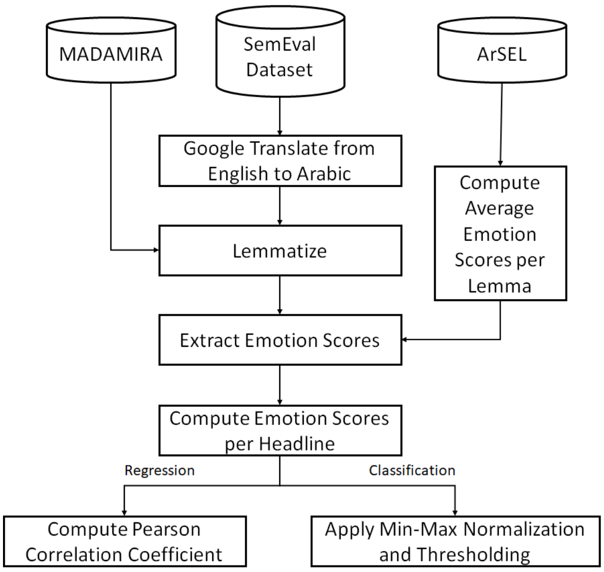

Figure 2: Overview of ArSEL Evaluation Steps.

from using ArSEL on the translated version of the same dataset. We compute the average of the two resulting scores and use it to perform regression and classification. We report the regression results in Table 7 under Combined column. As can be seen, combining the scores obtained through ArSEL and EmoWordNet improved Pearson correlation on average and consistently for all emotions except for Surprise. The discrepancy between the results achieved by EmoWordNet and ArSEL is due to the translation errors incurred by Google translate. The translation errors cause MADAMIRA to generate erroneous analysis of lemmas and hence the total emotion scores of the headline will be incorrect. The same error analysis can be inferred by looking at the other emotion classes as well.

In Table 8, we also compare F1 measure achieved by using ArSEL and EmoWordNet on translated SemEval and original one respectively. We also notice that the results for emotion classification are very close to each other. We also test the performance of combining the output of the two lexicons based on the parallel dataset shown under combined column in Table 8. Hence, we can conclude that the efficiency of EmoWordNet is preserved in ArSEL when used for emotion recognition from text. We can also deduce that emotion scores of EmoWordNet are correctly represented in ArSEL. In Table 5, we show some examples of news’ headlines that were correctly classified and in Table 6, examples of news’ headlines that were misclassified. By looking at the misclassified examples, we notice that misclassification is either due to predicting additional emotion labels to the actual ones (precision issue) or by predicting different emotion labels than the actual ones (recall issue). Similar to the regression task, translation errors incurred by Google translate have a negative impact on the analysis performed by MADAMIRA, thus, the translated headline is misrepresented and emotion scores assigned to the headline are incorrect.

4.2. Using SemEval 2018 Arabic Affective Tweets Dataset

While in the previous section we performed an extrinsic evaluation of ArSEL against a translated dataset from English, we present in this section an evaluation against a native Arabic dataset extracted from SemEval 2018 Task 1 “Affect in Tweets”. We describe first the dataset and the coverage achieved by ArSEL and then we present results of applying regression and classification using the same approach described in section 4.1.2.

English News' Headline	Google Translation	True Emotions
Ice storms kill 21 across nation	العواصف الثلجية تقتل ٢١ عبر الأمة	fear; sadness
Thailand attacks kill three, injure 70	هجمات تايلاند قتل ثلاثة، وإصابة ٧٠	fear; sadness
Heavy snow causes travel chaos and shuts schools	تسبب الثلوج الكثيفة فوضى السفر وتغلق المدارس	fear; sadness; surprise
Israeli, Lebanese clash on border	اشتباك إسرائيلي، لبناني على الحدود	anger; fear; sadness
Catania punished for fan violence	كاتانيا يعاقب على العنف مروحة	anger; sadness

Table 5: Examples of Correctly Classified News' Headlines from SemEval 2007.

English News' Headline	Google Translation	True Emotions	Predicted Emotions
Closings and cancellations top advice on flu outbreak	إغلاق وإلغاء المشورة العليا بشأن تفشي الأنفلونزا	joy	fear; surprise
Discovered boys bring shock, joy	اكتشاف الأولاد تجلب صدمة، والفرح	joy; surprise	sadness; surprise
Iraqi sunni lands show new oil and gas promise	وتظهر الأراضي السنية العراقية وعدا جديدا للنفط والغاز	joy	fear; surprise
Golden Globes on their way	غولدن غلوب في طريقهم	joy	joy; sadness; surprise
Bush adamant on troops to Iraq	بوش يصر على القوات إلى العراق	anger; sadness	fear

Table 6: Examples of Misclassified News' Headlines from SemEval 2007.

Emotion	EmoWordNet	ArSEL	Combined
Fear	0.51	0.44	**0.53**
Anger	0.31	0.34	**0.37**
Joy	0.33	0.26	**0.35**
Sadness	**0.41**	0.31	**0.41**
Surprise	**0.17**	0.1	0.14
Average	0.35	0.29	**0.36**

Table 7: Pearson Correlation Values.

Emotion	EmoWordNet	ArSEL	Combined
Fear	0.45	**0.57**	0.55
Anger	0.17	**0.36**	**0.36**
Joy	0.48	0.55	**0.59**
Sadness	0.46	0.50	**0.55**
Surprise	0.43	0.52	**0.53**
Average	0.40	0.50	**0.52**

Table 8: F1-Measure results for emotion classification using EmoWordNet on English SemEval 2007, using ArSEL on the Arabic translated version and when combining the two scores.

4.2.1. About the Data

In SemEval 2017, a task was created for Arabic Twitter sentiment analysis (Rosenthal et al., 2017). Several teams participated and the winning teams were NileTMRG (El-Beltagy et al., 2017) and OMAM (Baly et al., 2017b; Onyibe and Habash, 2017). In SemEval 2018, the focus was on Emotion classification from text. We utilize the provided competition dataset to evaluate ArSEL. SemEval 2018 dataset consists of Arabic tweets that are annotated with four emotions: anger, fear, joy and sadness along with the intensity present for each one. We have only access to the training and the development sets. In total, there are 2,871 tweets. In Table 9, we show the distribution of emotions across the tweets. The frequencies of the emotions are very close to each other with "Sadness" being the most frequent in the dataset. We follow the same experiment setup described in section 4.1.2, but we do not need the translation part since the data is already in

Emotion	Number of Occurrence
Fear	1028
Anger	1027
Joy	952
Sadness	**1030**

Table 9: Distribution of Emotion Labels across the Tweets.

Arabic. Instead, we perform additional preprocessing steps given that the dataset is extracted from Twitter. We clean the tweets from the hash tag and the underscore characters. We then feed the tweets to MADAMIRA to extract lemmas. In terms of ArSEL coverage, we were able to match 83.47% of the generated lemmas that belong to one of the four main POS tags. We were not able to generate any emotion scores for three tweets that mainly consisted of dialectal Arabic terms (عيونج , your eyes) elongations (خاااف, fear) and emoticons.

4.2.2. Regression and Classification Results

We follow the same approach described in section 4.1.2 to perform regression and classification with the modifications described in section 4.2.1. We use the average of the scores of the four emotions (joy, fear, anger and sadness), mutually present in ArSEL and in SemEval 2018 dataset. We have tried the sum of the emotions' scores as well, but, using average showed to be better. For the regression, we evaluate Pearson correlation coefficient against the intensity scores provided in the Twitter data. On average, we achieve an R score of 0.26. Table 12 shows the results per emotion.

For classification, we also apply min-max normalization and compare against the provided labels in the data. We use F1 measure as an evaluation metric. We also compare the results of our naïve unsupervised classifier to a majority baseline classifier where the predictor will always assign "Sadness" to the tweet since it is the most frequent emotion. The results are shown in Table 13. We outperform the baseline by an average of 30% as F1-score. Thus, we can confirm the efficiency of using ArSEL for emotion recognition tasks. We expect better results to be

Arabic Tweet	Translation	True Emotions
عيد سعيد جدا ما نعرف كلام كفار	A very happy holiday we don't know words of unbelievers	joy
!!اليش هالفتره عباره عن دموع وحرقت قلب؟	Why is this period full of tears and sorrow	sadness
يـارب هالفجر جايب معاه كل خير وسعاده وتوفيق	Oh God, this dawn is bringing all good, happiness and reconciliation	joy
ماعليك زود ياالجوهرة كلك لطف وحبابة يسعدك ربي	You are a diamond of niceness and loveliness, may God make happy	joy
صباح الخير و الفرح بيوم المرأة العالمي انشالله أيامك كلها أعياد مايا	Good morning and joy in international Women's day	anger; joy

Table 10: Examples of Correctly Classified Arabic Tweets from SemEval 2018.

Arabic Tweet and English Translation	True Emotions	Predicted Emotions
كنت اكتر انسانه بتخاف بس صارلي كم سنه كتير جريئه حتى بحضر افلام رعب لحالي و عادي I used to be scared from watching scary movies but I have been watching them by myself since a while	joy	fear
هرفي لو اعطوني ببلاش ما اخذت من عندهم شي ، مرتين دخلت المستشفى بسببهم Even if they gave it for free I won't take it, I was admitted to the hospital twice because of them	fear	anger; sadness
عشان كده العرب كانو بيشوفو ان اللي عيونهم ملونه نذير شؤم تحسي عينه فيها غدر وشر That's why Arabs thought that people with colored eyes are evil	fear	joy; sadness
شيء ما يقوم بإشعال فتيل الرهبة في قلبي كلما تعلق الموضوع بالحب I have fear feelings whenever the subject is related to love	fear	sadness
طيب طالما هوا عتاب كيف صار فراق ؟؟ Since it was reproach why did it become separation?	sadness	joy

Table 11: Examples of Misclassified Arabic Tweets from SemEval 2018.

Emotion	R Value
Fear	0.26
Anger	0.25
Joy	0.31
Sadness	0.22
Average	0.26

Table 12: Pearson Correlation Results on SemEval 2018 Arabic Tweets Dataset.

Emotion	ArSEL	Majority Baseline
Fear	0.32	0
Anger	0.41	0
Joy	0.52	0
Sadness	0.46	0.5
Average	**0.43**	0.13

Table 13: Classification F1-score Results on SemEval 2018 Arabic Tweets Dataset.

achieved when utilizing more sophisticated regression and classification techniques.

We also show examples of correctly classified tweets in Table 10, whereas in Table 11, we present examples of misclassified tweets.

By analyzing some of the misclassification examples we can see that several tweets are in dialectal Arabic which may produce erroneous morphological analysis. Moreover, some words have different meanings and emotion significance especially when used in dialectal Arabic such as the word "طيب" which could mean good, ok, tasty or alright. Last but not least, it is important to have a comprehensive model that takes into consideration the whole tweet rather than only word components as for instance in the first example in Table 11: although the words "خاف" and "رعب", which relate to fear are present in the tweet, the overall emotion is joy since the writer is happy that she has overcome her fear and she has been able to watch scary movies without any problem.

5. Conclusion and Future Work

We presented in this paper ArSEL, a large scale Arabic Sentiment and Emotion Lexicon. ArSEL is constructed automatically by using three lexical resources: DepecheMood, English WordNet and ArSenL. First, DepecheMood is mapped to EWN. Then, it is expanded iteratively using EWN synonymy semantic relation. The resulting expanded version of DepecheMood, EmoWordNet, is then linked to ArSenL entries using EWN synset IDs that exist in both lexicons. ArSEL consists of 32,196 Arabic lemmas annotated simultaneously with sentiment and emotion scores. ArSEL will be made publicly available on `http://oma-project.com` to speed up research in the area of emotion recognition from text. Moreover, using ArSEL in emotion classification task proved to be efficient with comparable performance to when utilizing EmoWordNet on an English dataset. Using ArSEL in a simplistic classification model outperformed a majority baseline predictor by 30% in terms of F1 measure. As future work, we would like to investigate more complex and sophisticated emotion recognition models and test the proposed models on larger datasets.

6. Bibliographical References

Abdaoui, A., Azé, J., Bringay, S., and Poncelet, P. (2017). Feel: a french expanded emotion lexicon. *Language Resources and Evaluation*, 51(3):833–855.

Abdul-Mageed, M. and Diab, M. (2012). Toward building a large-scale Arabic sentiment lexicon. In *Proceedings of the 6th International Global WordNet Conference*, pages 18–22.

Abdul-Mageed, M. and Diab, M. T. (2014). Sana: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In *LREC*, pages 1162–1169.

Abdul-Mageed, M. and Ungar, L. (2017). Emonet: Fine–grained emotion detection with gated recurrent neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 718–728.

Abdul-Mageed, M. (2017). Modeling Arabic subjectivity and sentiment in lexical space. *Information Processing & Management*.

Al Sallab, A. A., Baly, R., Badaro, G., Hajj, H., El Hajj, W., and Shaban, K. B. (2015). Deep learning models for sentiment analysis in arabic. In *ANLP Workshop*, volume 9.

Al-Sallab, A., Baly, R., Hajj, H., Shaban, K. B., El-Hajj, W., and Badaro, G. (2017). Aroma: A recursive deep learning model for opinion mining in arabic as a low resource language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(4):25.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.

Badaro, G., Hajj, H., El-Hajj, W., and Nachman, L. (2013). A hybrid approach with collaborative filtering for recommender systems. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pages 349–354. IEEE.

Badaro, G., Baly, R., Hajj, H., Habash, N., and El-Hajj, W. (2014a). A large scale Arabic sentiment lexicon for Arabic opinion mining. *ANLP 2014*, 165.

Badaro, G., Baly, R., Hajj, H., Habash, N., El-hajj, W., and Shaban, K. (2014b). An efficient model for sentiment classification of Arabic tweets on mobiles. In *Qatar Foundation Annual Research Conference*, number 1, page ITPP0631.

Badaro, G., Hajj, H., Haddad, A., El-Hajj, W., and Shaban, K. B. (2014c). A multiresolution approach to recommender systems. In *Proceedings of the 8th Workshop on Social Network Mining and Analysis*, page 9. ACM.

Badaro, G., Hajj, H., Haddad, A., El-Hajj, W., and Shaban, K. B. (2014d). Recommender systems using harmonic analysis. In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, pages 1004–1011. IEEE.

Badaro, G., Baly, R., Akel, R., Fayad, L., Khairallah, J., Hajj, H., Shaban, K., and El-Hajj, W. (2015). A light lexicon-based mobile application for sentiment mining of arabic tweets. In *Proceedings of the Second Workshop on Arabic Natural Language Processing*, pages 18–25.

Baly, R., Badaro, G., El-Khoury, G., Moukalled, R., Aoun, R., Hajj, H., El-Hajj, W., Habash, N., and Shaban, K. (2017a). A characterization study of arabic twitter data with a benchmarking for state-of-the-art opinion mining models. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 110–118.

Baly, R., Hajj, H., Habash, N., Shaban, K. B., and El-Hajj, W. (2017b). A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in arabic. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(4):23.

Bandhakavi, A., Wiratunga, N., Deepak, P., and Massie, S. (2014). Generating a word-emotion lexicon from# emotional tweets. In * *SEM@ COLING*, pages 12–21.

Bandhakavi, A., Wiratunga, N., Massie, S., and Padmanabhan, D. (2017). Lexicon generation for emotion detection from text. *IEEE intelligent systems*, 32(1):102–108.

Bernard, J. R. L.-B. and Bernard, J. R. L.-B. (1986). *The Macquarie Thesaurus*. Macquarie.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C. (2006). Introducing the Arabic wordnet project. In *Proceedings of the third international WordNet conference*, pages 295–300.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Bougie, R., Pieters, R., and Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, 31(4):377–393.

Brave, S. and Nass, C. (2003). Emotion in human–computer interaction. *Human-Computer Interaction*, page 53.

Buckwalter, T. (2002). Buckwalter {Arabic} morphological analyzer version 1.0.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211. ACM.

Cambria, E., Havasi, C., and Hussain, A. (2012). Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS conference*, pages 202–207.

Constantine, L., Badaro, G., Hajj, H., El-Hajj, W., Nachman, L., BenSaleh, M., and Obeid, A. (2016). A framework for emotion recognition from human computer interaction in natural setting. *22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2016), Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2016)*.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80.

Das, D., Poria, S., and Bandyopadhyay, S. (2012). A classifier based approach to emotion lexicon construction. In *NLDB*, pages 320–326. Springer.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

El-Beltagy, S. R., Kalamawy, M. E., and Soliman, A. B. (2017). Niletmrg at semeval-2017 task 4: Arabic sentiment analysis. *arXiv preprint arXiv:1710.08458*.

El Gohary, A. F., Sultan, T. I., Hana, M. A., and El Dosoky, M. (2013). A computational approach for analyzing and detecting emotions in Arabic text. *International Journal of Engineering Research and Applications (IJERA)*, 3:100–107.

Esuli, A. and Sebastiani, F. (2007). Sentiwordnet: A high-coverage lexical resource for opinion mining. *Evaluation*, pages 1–26.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017). Using millions of emoji occurrences to learn any–domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.

Fellbaum, C. (1998). *WordNet*. Wiley Online Library.

Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Fragopanagos, N. and Taylor, J. G. (2005). Emotion recognition in human–computer interaction. *Neural Networks*, 18(4):389–405.

Goldman, A. I. and Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, 94(3):193–213.

Graff, D., Maamouri, M., Bouziri, B., Krouna, S., Kulick, S., and Buckwalter, T. (2009). Standard Arabic morphological analyzer (sama) version 3.1. *Linguistic Data Consortium LDC2009E73*.

Gunes, H. and Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4):1334–1345.

Habash, N. Y. (2010). Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1):1–187.

Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., and Weinmann, M. (2017). How is your user feeling? inferring emotion through human–computer interaction devices. *MIS Quarterly*, 41(1).

Houjeij, A., Hamieh, L., Mehdi, N., and Hajj, H. (2012). A novel approach for emotion classification based on fusion of text and speech. In *Telecommunications (ICT), 2012 19th International Conference on*, pages 1–6. IEEE.

Jaimes, A. and Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer vision and image understanding*, 108(1):116–134.

Knautz, K., Siebenlist, T., and Stock, W. G. (2010). Memose: search engine for emotions in multimedia documents. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and development in information retrieval*, pages 791–792. ACM.

Liu, H. and Singh, P. (2004). Conceptnet: a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226.

Liu, B. and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Mohammad, S. M. and Yang, T. W. (2011). Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 70–79. Association for Computational Linguistics.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

Mohammad, S. M. (2017). Word affect intensities. *arXiv preprint arXiv:1704.08798*.

Neviarouskaya, A., Prendinger, H., and Ishizuka, M. (2007). Textual affect sensing for sociable and expressive online communication. *Affective Computing and Intelligent Interaction*, pages 218–229.

Neviarouskaya, A., Prendinger, H., and Ishizuka, M. (2011). Sentiful: A lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, 2(1):22–36.

Onyibe, C. and Habash, N. (2017). Omam at semeval-2017 task 4: English sentiment analysis with conditional random fields. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 670–674.

Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Pantic, M. and Rothkrantz, L. J. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390.

Pasha, A., Al-Badrashiny, M., Diab, M. T., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O., and Roth, R. (2014). Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. In *LREC*, volume 14, pages 1094–1101.

Patwardhan, A. S. and Knapp, G. M. (2017). Multimodal affect analysis for product feedback assessment. *arXiv preprint arXiv:1705.02694*.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Theories of emotion*, 1(3-31):4.

Plutchik, R. (1994). *The psychology and biology of emotion.* HarperCollins College Publishers.

Poria, S., Gelbukh, A., Das, D., and Bandyopadhyay, S. (2012). Fuzzy clustering for semi-supervised learning–case study: Construction of an emotion lexicon. In *Mexican International Conference on Artificial Intelligence*, pages 73–86. Springer.

Rosenthal, S., Farra, N., and Nakov, P. (2017). Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518.

Shaheen, S., El-Hajj, W., Hajj, H., and Elbassuoni, S. (2014). Emotion recognition from text based on automatically generated rules. In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, pages 383–392. IEEE.

Staiano, J. and Guerini, M. (2014). Depechemood: A lexicon for emotion analysis from crowd-annotated news. *arXiv preprint arXiv:1405.1605*.

Stone, P. J., Dunphy, D. C., and Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Strapparava, C. and Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.

Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Trad, C., Hajj, H. M., El-Hajj, W., and Al-Jamil, F. (2012). Facial action unit and emotion recognition with head pose variations. In *ADMA*, pages 383–394. Springer.

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

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., and Kissler, J. (2017). Mapping the emotional face. how individual face parts contribute to successful emotion recognition. *PloS one*, 12(5):e0177239.

Xu, G., Meng, X., and Wang, H. (2010). Build chinese emotion lexicons using a graph-based algorithm and multiple resources. In *Proceedings of the 23rd international conference on computational linguistics*, pages 1209–1217. Association for Computational Linguistics.

Yang, C., Lin, K. H.-Y., and Chen, H.-H. (2007). Building emotion lexicon from weblog corpora. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 133–136. Association for Computational Linguistics.