EmoWordNet: Automatic Expansion of Emotion Lexicon Using English WordNet

Gilbert Badaro, Hussein Jundi[†], Hazem Hajj, Wassim El-Hajj[†]
Department of Electrical & Computer Engineering
[†]Department of Computer Science
American University of Beirut
Beirut, Lebanon
{ggb05;haj14;hh63;we07}@aub.edu.lb

Abstract

Nowadays, social media have become a platform where people can easily express their opinions and emotions about any topic such as politics, movies, music, electronic products and many others. On the other hand, politicians, companies, and businesses are interested in analyzing automatically people's opinions and emotions. In the last decade, a lot of efforts has been put into extracting sentiment polarity from texts. Recently, the focus has expanded to also cover emotion recognition from texts. In this work, we expand an existing emotion lexicon, DepecheMood, by leveraging semantic knowledge from English WordNet (EWN). We create an expanded lexicon, EmoWordNet, consisting of 67K terms aligned with EWN, almost 1.8 times the size of DepecheMood. We also evaluate EmoWord-Net in an emotion recognition task using SemEval 2007 news headlines dataset and we achieve an improvement compared to the use of DepecheMood. EmoWordNet is publicly available to speed up research in the field on http://oma-project.com.

1 Introduction

Emotion recognition models have been extensively explored based on different modalities such as human computer interaction (Cowie et al., 2001; Pantic and Rothkrantz, 2003; Fragopanagos and Taylor, 2005; Jaimes and Sebe, 2007; Hibbeln et al., 2017; Patwardhan and Knapp, 2017; Constantine et al., 2016) and facial images and expressions (Goldman and Sripada, 2005; Gunes and Piccardi, 2007; Trad et al., 2012; Wegrzyn et al., 2017). Recently, special attention has been given to emotion recognition from text (Wu et al., 2006; Alm et al., 2005; Shaheen et al., 2014; Abdul-Mageed and Ungar, 2017; Badaro et al., 2018b,a). In fact, a tremendous amount of opinionated and emotionally charged text data is nowadays avail-

able on the Internet due to the increase of number of users of social networks such as Twitter and Facebook. For instance, Facebook reached more than 2 billion users on September 2017.[1] Recognizing emotions from text has several applications: first, it helps companies and businesses in shaping their marketing strategies based on consumers' emotions (Bougie et al., 2003); second, it allows improving typical collaborative filtering based recommender systems (Badaro et al., 2013, 2014c,d) in terms of products or advertisements recommendations (Mohammad and Yang, 2011); third, politicians can learn how to adapt their political speech based on people emotions (Pang et al., 2008) and last but not least emotion classification helps in stock market predictions (Bollen et al., 2011).

While plenty of works exist for sentiment analysis for different languages including analysis of social media data for sentiment characteristics (Al Sallab et al., 2015; Baly et al., 2014, 2017b,a), few works focused on emotion recognition from text. Since sentiment lexicons helped in improving the accuracy of sentiment classification models (Liu and Zhang, 2012; Al-Sallab et al., 2017; Badaro et al., 2014a,b, 2015), several researchers are working on developing emotion lexicons for different languages such as English, French, Polish 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; Maziarz et al., 2016; Janz et al., 2017). 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) (happiness, sadness, fear,

---

[1]https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

anger, surprise and disgust) or Plutchik model (Plutchik, 1994) that includes trust and anticipation in addition to Ekman's six emotions. Despite the efforts for creating large scale emotion lexicons for English, the size of existing emotion 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 terms 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.

In this paper, we focus on expanding coverage of existing emotion lexicon, namely DepecheMood, using the synonymy semantic relation available in English WordNet. We decide to expand DepecheMood since it is one of the largest emotion lexicon publicly available, and since its terms are aligned with EWN, thus allowing us to benefit from powerful semantic relations in EWN.

The paper is organized as follows. In section 2, we conduct a brief literature survey on existing emotion lexicons. In section 3, we describe the expansion approach to build EmoWordNet. In section 4, we compare the performance of EmoWordNet against DepecheMood using SemEval 2007 dataset and in section 5, we present a conclusion of our results and future work.

2   Literature Review

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 for emotions. 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. 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, which can be useful for compositional computations of emotion scores. Affect database extends SentiFul and covers around 2.5K words presented in their lemma form along with the corresponding part of speech (POS) tag. DepecheMood was automatically built by harvesting social media data that were implicitly annotated with emotions. Staiano and Guerini (2014) utilized 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 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 applied on SemEval 2007 dataset (Strapparava and Mihalcea, 2007). Since DepecheMood is aligned with EWN, is publicly available and has a better coverage and claimed performance compared to existing emotion lexicons, we decide to expand it using EWN semantic relations as described below in section 3.

To summarize, there are mainly two approaches that have been followed for building emotion lexicons for English. The first set of methods relies on manual annotation either done by specific indi-

viduals or through crowdsourcing, where the list of words is extracted from lexical resources. The second approach is automatic or semi-automatic and is based on annotated corpora for emotion. The first approach tends to produce limited size and highly accurate emotion lexicons but it is relatively expensive. On the other hand, the second approach is cheap and results in large scale emotion lexicons but with lower accuracy compared to manually developed emotion lexicons in terms of accurately representing the emotion of the term.

3 EmoWordNet

In this section, we describe the approach we followed in order to expand DepecheMood and build EmoWordNet. DepecheMood consists of 37,771 lemmas along with their corresponding POS tags where each entry is appended with scores for 8 emotion labels: afraid, amused, angry, annoyed, don't_care, happy, inspired and sad. Three variations of score representations exist for DepecheMood. We select to expand the DepecheMood variation with normalized scores since this variation performed best according to the presented results in (Staiano and Guerini, 2014).

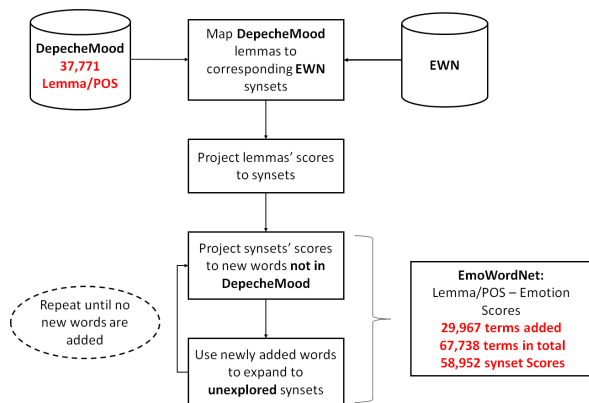In Fig. 1, we show an overview of the steps followed to expand DepecheMood.

Figure 1: Overview of DepecheMood Expansion Approach.

Step 1: EWN synsets that include lemmas of DepecheMood were retrieved. A score was then computed for each retrieved synset, s. Let S denotes the set of all such synsets. Two cases might appear: either the retrieved synset included only one lemma from DepecheMood, in this case the synset was assigned the same score of the lemma, or, the synset included multiple lemmas that exist

in DepecheMood, in this case the synset's score was the average of the scores of its corresponding lemmas. **Step 2:** A synset, s, includes two set of terms: T, terms that are **in** DepecheMood, and $\bar{T}$, terms **not in** DepecheMood. Using the synonymy semantic relation in EWN, and based on the concept that synonym words would likely share the same emotion scores, we assigned the synset's scores to its corresponding terms $\bar{T}$. Again, a term t in $\bar{T}$ might appear in one or multiple synsets from S. Hence, the score assigned to t would be either the one of its corresponding synset or the average of the scores of its corresponding synsets that belong to S. **Step 3:** after performing step 2, new synsets might be explored. Terms in $\bar{T}$ might also appear in synsets $\bar{s}$ that do not belong to S. $\bar{s}$ would get the score of its corresponding terms. Step 2 and 3 were repeated until no new terms or synsets were 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 mostly preserves the emotion orientation and does not require manual validation as described by Strapparava et al. (2004).

As a walking example of the steps described above, let us consider the DepecheMood term "bonding" having noun as POS tag. "bonding" can be found in three different EWN noun synsets 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 EWN synset so there are no more iterations.

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 present a simple extrinsic evaluation of EmoWordNet similar to the one performed for DepecheMood.

4 Evaluation of EmoWordNet

In this section, we evaluate the effectiveness of EmoWordNet in emotion recognition task from text. We evaluate regression as well as classification of emotions in unsupervised settings using similar techniques used for evaluating DepecheMood.

4.1 Dataset & Coverage

We utilized the dataset provided publicly by SemEval 2007 task on Affective text (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 EmoWordNet, as well as DepecheMood, follow the ones provided by Rappler Mood Meter. We considered the same emotion mapping assumptions presented in the work of (Staiano and Guerini, 2014): Fear → Afraid, Anger → Angry, Joy → Happy, Sadness → Sad and Surprise → Inspired. Disgust was not aligned with any emotion in EmoWordNet and hence was discarded as also assumed in (Staiano and Guerini, 2014). One important aspect of the extrinsic evaluation was checking the coverage of EmoWordNet against SemEval dataset. In order to compute coverage, we performed lemmatization of the news headlines using WordNet lemmatizer available through Python NLTK package. We excluded all words with POS tags different than noun, verb, adjective and adverb. EmoWordNet achieved a coverage of 68.6% while DepecheMood had a coverage of 67.1%. An increase in coverage was expected but since the size of the dataset is relatively small, the increase was only around 1.5%. In terms of headline coverage, only one headline ("Toshiba Portege R400") was left without any emotion scores when using both EmoWordNet and DepecheMood since none of its terms were found in any of the two lexicons.

4.2 Regression and Classification Results

We followed an approach similar to the one presented for evaluating DepecheMood. For preprocessing, we first lemmatized the headlines using WordNet lemmatizer available in Python NLTK package. We also accounted for multi-word terms that were solely available in EmoWordNet by looking at n-grams (up to n=3) after lemmatization. We then removed all terms that did not belong to any of the four POS tags: noun, verb, adjective and adverbs. For features computation, we considered two variations: the sum and the average of the emotion scores for the five emotion labels that overlapped between EmoWordNet and SemEval dataset. Using average turned out to perform better than when using sum for both lexicons. As stated in (Staiano and Guerini, 2014) paper, 'Disgust' emotion was excluded since there was no corresponding mapping in EmoWordNet/DepecheMood. The first evaluation consisted of measuring Pearson Correlation between the scores computed using the lexicons and those provided in SemEval. The results are reported in Table 1. We could see that the results are relatively close to each other: EmoWordNet slightly outperformed DepecheMood for the five different emotions. It was expected to have close results given that the coverage of EmoWordNet is very close to DepecheMood. Given the slight improvement, we expect EmoWordNet to perform much better on larger datasets.

For the classification task, we first transformed the numerical emotion scores of the headlines to a binary representation. We applied min-max normalization on the computed emotion scores per headline, and then assigned a '1' for the emotion label with score greater than '0.5', and a '0' otherwise. We used F1 measure for evaluation. Results are shown in Table 2. More significant improvement was observed in classification task compared to regression task when using EmoWordNet.

4.3 Results Analysis

In this section, we present some quantitative and qualitative analyses of the results. For quantitative analysis, we checked first whether the count of terms in a headline is correlated with having a correct emotion classification. Overall, the length of headlines was varying between 2 and 15 terms. Headlines with length between 5 and 10 terms were mostly correctly classified. Hence, one can

Emotion	EmoWordNet	DepecheMood
Fear	**0.59**	0.54
Anger	**0.42**	0.38
Joy	**0.33**	0.21
Sadness	**0.43**	0.40
Surprise	**0.51**	0.47
Average	0.46	0.40

Table 1: Pearson Correlation values between predicted and golden scores.

Emotion	EmoWordNet	DepecheMood
Fear	**0.45**	0.32
Anger	**0.17**	0.00
Joy	**0.48**	0.16
Sadness	**0.46**	0.30
Surprise	**0.43**	0.40
Average	0.40	0.24

Table 2: F1-Measure results for emotion classification.

Headline	Emotions
Hackers attack root servers	Anger; Fear
Subway collapse caught on camera	Fear; Sadness
Action games improve eyesight	Joy; Surprise
Study finds gritty air raises heart disease risk in older women	Fear; Sadness; Surprise
Wizardry at Harvard: physicists move light	Surprise

Table 3: Examples of correctly classified headlines.

Headline	Gold	Predicted
A film star in Kampala, conjuring aminos ghost	Fear; Surprise	Anger; Joy; Sadness
Damaged Japanese whaling ship may resume hunting off Antarctica	Joy; Sadness	Anger; Fear; Surprise
Apple revs up Mac attacks on Vista	Surprise	Anger; Fear; Joy; Sadness
Serbia rejects United Nation's Kosovo plan	Anger; Sadness; Surprise	Fear; Joy
Taliban leader killed in airstrike	Joy; Sadness	Anger; Fear; Surprise

Table 4: Examples of misclassified headlines.

conclude that having a headline with couple of terms only may not allow the system to clearly decide on the emotion label and having headlines with many terms may cause the system to over predict emotions. In addition to headline length, we checked whether POS tags are correlated with correct or erroneous emotion predictions. Given that the dataset consists of news headlines, the "noun" POS tag was the most frequent in both correctly classified headlines and misclassified ones.

For qualitative analysis, we analyze few correctly classified headlines and few other misclassified ones. We show in Table 3 few examples of correctly classified headlines and in table 4 other examples of misclassified headlines. By looking at the misclassified examples, we observe that the golden annotation tend to be sometimes conflicting such as the second and the fifth examples in Table 4 where we have joy and sadness as assigned emotions for the two headlines. An explanation for having conflicting emotions for the same headline is that the annotators reflected their personal point of view of the information conveyed by the headline. Hence, some people were happy to read the headline others were sad. In order to incorporate such challenging aspect of emotion recognition from text, more sophisticated emotion recognition models need to be considered and tested.

5 Conclusion and Future Work

We presented EmoWordNet, a large scale emotion lexicon, consisting of around 67K EWN words and 58K EWN synsets annotated with 8 emotion scores. EmoWordNet is automatically constructed by applying a semantic expansion approach using EWN and DepecheMood. When utilized for emotion recognition, EmoWordNet outperformed existing emotion lexicons and had a better lexical coverage. For future work, we would like to evaluate the performance of EmoWordNet on larger datasets and we would like to improve the accuracy of the recognition model. EmoWordNet is publicly available on http://oma-project.com.

References

Amine Abdaoui, Jérôme Azé, Sandra Bringay, and Pascal Poncelet. 2017. Feel: a french expanded emotion lexicon. *Language Resources and Evaluation*, 51(3):833–855.

Muhammad Abdul-Mageed and Lyle Ungar. 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.

Ahmad Al-Sallab, Ramy Baly, Hazem Hajj, Khaled Bashir Shaban, Wassim El-Hajj, and Gilbert Badaro. 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.

Ahmad A Al Sallab, Ramy Baly, Gilbert Badaro, Hazem Hajj, Wassim El Hajj, and Khaled B Shaban. 2015. Deep learning models for sentiment analysis in arabic. In *ANLP Workshop*, volume 9.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 579–586. Association for Computational Linguistics.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.

Gilbert Badaro, Ramy Baly, Rana Akel, Linda Fayad, Jeffrey Khairallah, Hazem Hajj, Khaled Shaban, and Wassim El-Hajj. 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.

Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj. 2014a. A large scale Arabic sentiment lexicon for Arabic opinion mining. *ANLP 2014*, 165.

Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, Wassim El-hajj, and Khaled Shaban. 2014b. An efficient model for sentiment classification of Arabic tweets on mobiles. In *Qatar Foundation Annual Research Conference*, 1, page ITPP0631.

Gilbert Badaro, Obeida El Jundi, Alaa Khaddaj, Alaa Maarouf, Raslan Kain, Hazem Hajj, and Wassim El-Hajj. 2018a. Ema at semeval-2018 task 1: Emotion mining for arabic. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*.

Gilbert Badaro, Hazem Hajj, Wassim El-Hajj, and Lama Nachman. 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.

Gilbert Badaro, Hazem Hajj, Ali Haddad, Wassim El-Hajj, and Khaled Bashir Shaban. 2014c. A multiresolution approach to recommender systems. In *Proceedings of the 8th Workshop on Social Network Mining and Analysis*, page 9. ACM.

Gilbert Badaro, Hazem Hajj, Ali Haddad, Wassim El-Hajj, and Khaled Bashir Shaban. 2014d. Recommender systems using harmonic analysis. In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, pages 1004–1011. IEEE.

Gilbert Badaro, Hussein Jundi, Hazem Hajj, Wassim El-Hajj, and Nizar Habash. 2018b. Arsel: A large scale arabic sentiment and emotion lexicon. In *Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT3) co-located with LREC2018*.

Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash, and Khaled Shaban. 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.

Ramy Baly, Gilbert Badaro, Hazem Hajj, Nizar Habash, Wassim El Hajj, and Khaled Shaban.

2014. Semantic model representation for human's pre-conceived notions in arabic text with applications to sentiment mining. In *Qatar Foundation Annual Research Conference*, 1, page ITPP1075.

Ramy Baly, Gilbert Badaro, Ali Hamdi, Rawan Moukalled, Rita Aoun, Georges El-Khoury, Ahmad Al Sallab, Hazem Hajj, Nizar Habash, Khaled Shaban, et al. 2017b. Omam at semeval-2017 task 4: Evaluation of english state-of-the-art sentiment analysis models for arabic and a new topic-based model. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 603–610.

Anil Bandhakavi, Nirmalie Wiratunga, Stewart Massie, and Deepak Padmanabhan. 2017. Lexicon generation for emotion detection from text. *IEEE intelligent systems*, 32(1):102–108.

John Rupert Lyon-Bowes Bernard and John Rupert Lyon-Bowes Bernard. 1986. *The Macquarie Thesaurus*. Macquarie.

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

Roger Bougie, Rik Pieters, and Marcel Zeelenberg. 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.

Erik Cambria, Catherine Havasi, and Amir Hussain. 2012. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS conference*, pages 202–207.

Layale Constantine, Gilbert Badaro, Hazem Hajj, Wassim El-Hajj, Lama Nachman, Mohamed BenSaleh, and Abdulfattah Obeid. 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)*.

Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias,

Winfried Fellenz, and John G Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80.

Dipankar Das, Soujanya Poria, and Sivaji Bandyopadhyay. 2012. A classifier based approach to emotion lexicon construction. In *NLDB*, pages 320–326. Springer.

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

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

Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

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

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

Hatice Gunes and Massimo Piccardi. 2007. Bimodal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4):1334–1345.

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

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

Arkadiusz Janz, Jan Kocoń, Maciej Piasecki, and Monika Zaśko-Zielińska. 2017. plwordnet as a basis for large emotive lexicons of polish.

Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.

Hugo Liu and Push Singh. 2004. Conceptneta practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kedzia. 2016. plwordnet 3.0–a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268.

Saif M Mohammad. 2017. Word affect intensities. *arXiv preprint arXiv:1704.08798*.

Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

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

Saif M Mohammad and Tony Wenda Yang. 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.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2007. Textual affect sensing for sociable and expressive online communication. *Affective Computing and Intelligent Interaction*, pages 218–229.

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

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

Amol S Patwardhan and Gerald M Knapp. 2017. Multimodal affect analysis for product feedback assessment. *arXiv preprint arXiv:1705.02694*.

Robert Plutchik. 1994. *The psychology and biology of emotion.* HarperCollins College Publishers.

Soujanya Poria, Alexander Gelbukh, Dipankar Das, and Sivaji Bandyopadhyay. 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.

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

Jacopo Staiano and Marco Guerini. 2014. Depechemood: A lexicon for emotion analysis from crowd-annotated news. *arXiv preprint arXiv:1405.1605*.

Philip J Stone, Dexter C Dunphy, and Marshall S Smith. 1966. The general inquirer: A computer approach to content analysis.

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

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

Chadi Trad, Hazem M Hajj, Wassim El-Hajj, and Fatima Al-Jamil. 2012. Facial action unit and emotion recognition with head pose variations. In *ADMA*, pages 383–394. Springer.

Martin Wegrzyn, Maria Vogt, Berna Kireclioglu, Julia Schneider, and Johanna Kissler. 2017. Mapping the emotional face. how individual face parts contribute to successful emotion recognition. *PloS one*, 12(5):e0177239.

Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. 2006. Emotion recognition from text using semantic labels and separable mixture models. *ACM transactions on Asian language information processing (TALIP)*, 5(2):165–183.

Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 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.