EmoPars: A Collection of 30K Emotion-Annotated Persian Social Media Texts

Conference Paper · September 2021

DOI: 10.26615/issn.2603-2821.2021_023

Nazanin Sabri
University of California, San Diego

EmoPars: A Collection of 30K Emotion-Annotated Persian Social Media Texts

Nazanin Sabri, Reyhane Akhavan, Behnam Bahrak
School of Electrical and Computer Engineering, College of Engineering
University of Tehran, Tehran, Iran
{nazanin.sabri,reyhane.akhavan,bahrak}@ut.ac.ir

Abstract

The wide reach of social media platforms, such as Twitter, have enabled many users to share their thoughts, opinions and emotions on various topics online. The ability to detect these emotions automatically would allow social scientists, as well as, businesses to better understand responses from nations and costumers. In this study we introduce a dataset of 30,000 Persian Tweets labeled with Ekman's six basic emotions (Anger, Fear, Happiness, Sadness, Hatred, and Wonder). This is the first publicly available emotion dataset in the Persian language. In this paper, we explain the data collection and labeling scheme used for the creation of this dataset. We also analyze the created dataset, showing the different features and characteristics of the data. Among other things, we investigate co-occurrence of different emotions in the dataset, and the relationship between sentiment and emotion of textual instances. The dataset is publicly available at https://github.com/nazaninsbr/Persian-Emotion-Detection.

1 Introduction

As humans communicate emotions through text (Alm et al., 2005), the creation of text-based emotion detection models are a necessity for the analysis and understanding of the content posted online, more-natural speech generation, and better human-computer interaction tasks which are categorized as affective computing [1]. The lack of any other signals for emotions within text (i.e., facial expressions or body language) and the subjective nature of emotions, makes this task a challenging one.
Despite the availability of a number of studies on the topic of emotion recognition in NLP, they have been predominantly conducted in the English language. The unavailability of annotated data in the Persian language hinders the study of emotions in this language. As a result, in this study, we introduce a dataset of 30K Persian Tweets labeled with Ekman's six basic emotions (Ekman, 1999). To the best of our knowledge this is the first publicly available emotion-labeled dataset in the Persian language, which we believe will allow Persian NLP to progress and start the study of emotions.
The rest of this paper is structured as follows: in Section 2 we provide a brief review of studies on emotions and the automatic detection of emotions. We additionally point to other datasets available on the topic in other languages. Next, we discuss our data collection and labeling methods in Section 3. The statistics and properties of our datasets are presented in Section 4. Finally, we conclude the study in Section 5.

2 Related Work

Even though a significant body of work is available on the topic of emotions, a commonly agreed-upon definition is still lacking (Mulligan and Scherer, 2012). The American Psychology Association (APA) defines emotion as "a complex reaction pattern, involving experiential, behavioral, and physiological elements, by which an individual attempts to deal with a personally significant matter or event" (apa, (accessed December 1, 2020). The three key components of emotions, stated in the aforementioned definition, are subjective experience, physiological response, and behavioral response (Hockenbury and Hockenbury, 2010). Subjective experience has been explored in studies looking at the effects of culture, age, and gender on the emotions people feel (Barrett et al., 2007; Shaver et al., 1992; Fischer et al., 2004; Ekman et al., 1987; Kunzmann and Grühn, 2005). The physiological response to emotions could range from sweaty palms to a churning stomach. For instance, (Fernández

[1] "computing that relates to, arises from, or deliberately influences emotion" (Picard, 1999)

Proceedings of the Student Research Workshop associated with RANLP-2021, pages 167–173,
held online, Sep 1–3, 2021.
https://doi.org/10.26615/issn.2603-2821.2021_023

et al., 2012) shows that heart rate is significantly increased when people watch movies conveying fear or anger, (Waldstein et al., 2000) also explores frontal EEG activation in response to anger and happiness, and (LeDoux, 1995) reviews studies on the neurological aspects of emotion, focusing mostly on fear. In part one of (Philippot et al., 2004), the physiological processes during emotion regulation are discussed. Additionally, (Dalgleish, 2004) offers a comprehensive historical overview of studies on the neural bases of emotions. Lastly, behavioral responses could include smiling, sighing, and crying. These responses are found to often depend on societal norms and individual differences (Krys et al., 2016; Gross and Levenson, 1997; Van Hemert et al., 2011).

There is some debate on what basic emotions are and what we really mean by “basic”. (Ortony and Turner, 1990) discusses this issue at length. The paper refutes the claim that there exist basic emotions out of which all other emotions are built. (Ekman and Cordaro, 2011) disagrees and offers a concrete list of characteristics that basic emotions have. Nevertheless, multiple classifications of emotions have been introduced in the literature. One categorization, offered by Paul Ekman, suggests basic emotions to be: anger, disgust, fear, happiness, sadness and surprise (Ekman, 1999, 1992). Later on, however, Ekman does name other emotions that could potentially be proven to be basic (Ekman and Cordaro, 2011). Another well-known classification, offered by Plutchik, introduces 8 basic emotions composed of anger, fear, disgust, sadness, surprise, anticipation, trust, and joy (Plutchik, 1980).

With the growth of social media platforms, the automatic detection of emotions through text has come into focus in recent years. Several surveys of studies on emotion detection from text have been conducted, one in 2014 (Canales and Martínez-Barco, 2014), and two others in 2018 (Seyeditabari et al., 2018; Sailunaz et al., 2018). As a result, in our review, we will mainly focus on newer studies.

In 2019, a shared task on the detection of emotions in textual dialogue was organized (emo, (accessed December 1, 2020; Chatterjee et al., 2019), which resulted in a wave of studies on the topic. The best model on the task achieves an F1-score of 0.79, however the best two models on the task did not submit system description papers (Chatterjee et al., 2019). The third ranking model on the task (Agrawal and Suri, 2019), uses lexical features such as emotional intensity, valence-arousal-dominance scores (Warriner et al., 2013), and sentiment classifiers' scores to train a Light-GBM tree (Ke et al., 2017) model which achieves a micro-averaged F1 score of 0.77. (Basile et al., 2019) which was ranked fourth, uses a neural ensemble system made up of 4 neural models. The most used embedding-model among the top systems, was reported to be GloVe (Pennington et al., 2014). It is important to keep in mind that achieving the same level of accuracy on our dataset will be harder for two reasons: (1) our dataset does not include context for each Tweet which was available in the aforementioned task, and (2) Persian language is a low-resource language, thus some of the features that the participants used are not in our disposal.

Another shared-task in WASSA-2021 Workshop (Tafreshi et al., 2021) required participants to predict emotional tags (Ekman's six basic emotions) and empathy of news stories. The highest accuracy on the task was reported to be 0.62 with a corresponding F1 of 0.55 (Mundra et al., 2021). This result was obtained through data augmentation and fine-tuning of the ELECTRA (Clark et al., 2020) model. (Butala et al., 2021), another participant of the task, compares different conditional generation models (T5 (Raffel et al., 2019) and pegasus (Jingqing et al., 2019)) and pre-trained contextual embeddings (BERT (Devlin et al., 2019) and AlBERT (Lan et al., 2019)) and reports AlBERT to outperform BERT (Macro-F1 of 0.47 vs. 0.37). However, T5 is reported to have the best performance with a Macro-F1 of 0.57.

There are a number of studies unrelated to the shared-tasks as well. (Hasan et al., 2019) explores using a soft-classification model that assigns probabilities to each emotion. (Shoeb and de Melo, 2020) creates a dataset and introduces a method to find the correspondence between emojis and particular emotions. (Ishiwatari et al., 2020) tries to detect emotions in conversation. The authors incorporate speaker dependency into the model using graph attention networks (Veličković et al., 2017) as well as introducing a novel relational position encoding which is shown to improve the accuracy of the model. (Polignano et al., 2019) introduces a model made up of CNN, BiLSTM, and self-attention components, and compare different word-embeddings, finding that FastText vector spaces (Bojanowski et al., 2016) better capture the information they want. (Gollapalli et al., 2020) introduces an unsupervised emotion detection method which is built upon word co-occurrences and word associations. Some emotion datasets include: (Sosea and Caragea, 2020) from an English online health community with a focus on cancer, (Demszky et al., 2020) from English Reddit comments, (Liu et al., 2019) from long-form narratives in English, (Kumar et al., 2019) from Hindi stories, and (Almahdawi and Teahan, 2019) an Arabic dataset from Facebook posts written in the Iraqi dialect. While (Khosravi et al., 2019) uses machine learning models to detect emotions of Persian news texts, the dataset has not been published. Additionally, to the best of our knowledge, the dataset presented as part of this study is the first emotion dataset on the Persian social media texts. As social media content and formally-written news articles are structured differently, we believe that this dataset is of great value for the study of user-created content in social networks.

3 Data Collection and Labeling

In this section, we begin by going over the data collection method (3.1), continuing on to explaining the labeling and validation process (3.2). The statistics of our dataset are explained in Section 4. The data is publicly available on GitHub[2].

3.1 Collection

The data presented as part of this study was collected using Twitter's official developer API (twi, (accessed December 2, 2020). To make sure our data is not biased by any topic or the discussions of a particular time we take two measures:

(1) we collect tweets using different keywords including articles and prepositions which are not specific to any particular topic and thus make sure no bias is introduced.

(2) the data is sampled from Tweets posted since the mid-2019 up until mid-2020, making sure to include at least some Tweets from each time period. As a result no time-specific issue dominates our dataset.

After the data has been collected we randomly sample 30,000 instances of the Tweets with a minimum character length of 20 (the minimum length is placed to ensure the Tweet is long enough to potentially reflect emotions). To preserve the anonymity of the Tweets, we replace any user mentions with *@USERNAME* and any link with *www.LINK.com* before the data has been inputted for labeling. In this manner we make sure to indicate there was a link or mention but that it was removed for anonymity purposes.

3.2 Labeling

We use iToll (ito, (accessed January 2, 2021), a Persian crowd-sourcing website to label the texts in our dataset. The labeling is done through Yes/No questions, where the user is shown the text and asked if emotion X is present in the text. This question is then repeated for all 6 basic emotions. Each question is shown to 5 different individuals and the final result is calculated through max voting[3].

To make certain the labels are valid, after the first round of labeling has been completed, any user whom more than 50% of their votes contradict the majority vote are considered to be unreliable user and his/her votes are removed from the dataset. These votes are then replaced with new votes by asking the questions from other individuals. A similar method of labeling was used by (Sosea and Caragea, 2020) and was shown to obtain satisfactory results. However, it is important to keep in mind that the perception of emotions can be quite subjective. Additionally since no context is available for each text, each user might read and perceive a text with a different attitude/tone. These differences could all lead to difficulties when it comes to labeling emotions.

4 Dataset Statistics

4.1 Instances of Each Emotion

As previously mentioned, our dataset is made up of 30,000 Tweets, labeled with six different emotion labels. Figure 1 shows the number of tweets in our dataset that are simultaneously labeled with N different emotions. We can see that 80% of the data does not have any emotion label at all, and less than 5% have 3 or more emotion labels. The large proportion of Tweets that do not have any emotion, make the task of automatic emotion detection,

[2] https://github.com/nazaninsbr/Persian-Emotion-Detection

[3] In the published dataset the vote counts have been provided. These numbers are in the range of $[0, 5]$. As a result, to ascertain whether or not an emotion is present (a binary classification), the user must check if the number is more than 2 (in other words if more than 2 people out of 5 have voted YES for that emotion). Should you want more certain labels (with less room for annotation errors), you could select a higher threshold for each emotion.

difficult. Figure 2 displays the number of Tweets with each emotion. We can see sadness is the most observed emotion, followed by anger and hatred. Fear is the least observed emotion, with only 690 instances among our data (closely followed by Happiness with only 692 positive instances).

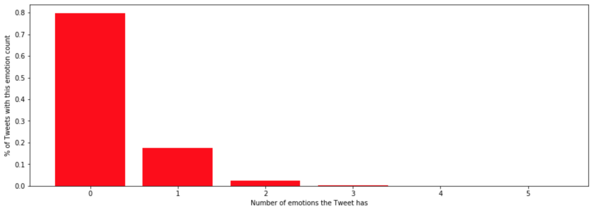

Figure 1: Number of Tweets in our dataset with zero to at most five labeled emotions

Figure 2: Number of Tweets with each emotion

4.2 Co-occurrence of Emotions

Co-occurrence of emotions is another issue we investigate in Figure 3. The value in each cell represents the proportion $\frac{e_1 \cap e_2}{e_1 \cup e_2}$ (where e_i represent the emotions on the row or column). We can see that positive and negative emotions are very unlikely to be labeled in the same text, further proving the validity of our labeling scheme.

4.3 Agreement on the Existence of Emotions

Next we explored whether there is more agreement on the availability of some emotions compared to others. To answer this question, we begin by defining the following metric for each emotion:

$$\text{Average Agreement} = \frac{1}{M} * \sum_{i=1}^{M} \frac{|n_{yes}^{(i)} - n_{no}^{(i)}|}{5} \quad (1)$$

In Equation 1, M represents the number of texts in our dataset ($M = 30,000$) and $n_{yes}^{(i)}$ is the number of positive votes for Tweet i. The closer the metric is to 1, there is more agreement on existence of the emotion. The results are presented in Table 1. We observe that the most agreement is on happiness and the least agreement is on the presence of sadness.

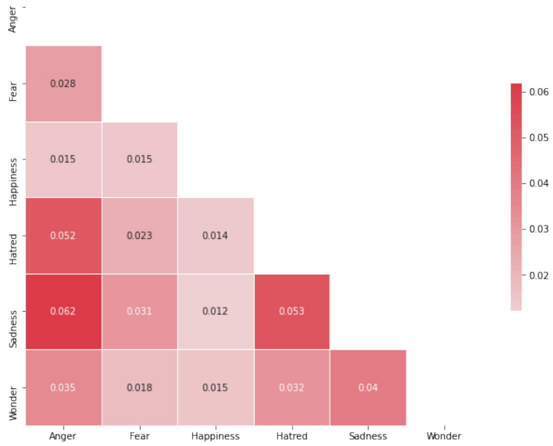

Figure 3: Emotion Co-occurrences, normalized by the total number of instances in the dataset

Emotion	Value of Equation 1
Anger	0.67
Sadness	0.65
Hatred	0.69
Happiness	0.75
Wonder	0.71
Fear	0.73

Table 1: Average agreement for different emotions on the sentences available in our dataset

4.4 Tweet Length

Next, we look into the average character length of instances of each emotion. The results are shown in Table 2. As Twitter allows 280 characters for each Tweet (twi), the average lengths are small. However this is not a phenomena limited to our dataset as (twi) also reports that only 1% of Tweets reach the limit.

4.5 Emojis, and Hashtags

Looking at the emojis used in the dataset, we see that 90% of sentences in the dataset do not have any emojis. Hashtags are also only observed in 19.3% of the data.

Emotion	Avg. Char. Len.	# Instances
Anger	133.3	1,632
Sadness	129.6	1,770
Hatred	132.6	1,256
Happiness	110.1	692
Wonder	124.1	986
Fear	140.5	690

Table 2: Average character length of sentences in our dataset

4.6 Sentiment vs. Emotion

To understand how the sentiment of each sentence relates to its emotion, we use Quecst (Jung et al., 2020), an online tool which labels texts with their sentiment score. We were able to detect the sentiment of 21,485 texts (71%) in our dataset. Figure 4 depicts these sentiment values. *0* means a negative sentiment and *1* refers to a positive one. We can see there are high number of tweets for both negative and positive extremes. Additionally, the majority of the data (45%) is shown to have a positive (> 0.5) sentiment. While neutral sentiment ($= 0.5$) makes up 17% and negative sentiment (< 0.5) makes up 37% of the data.

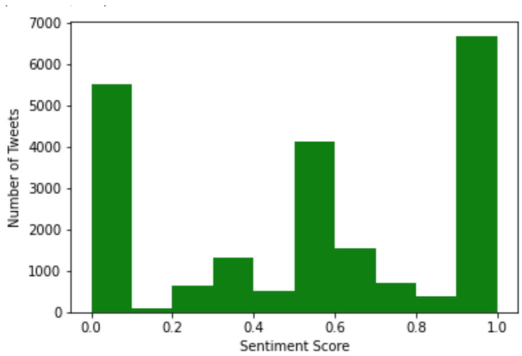

Figure 4: Histogram of sentiment scores of the tweets in our dataset

We further explore the distribution of sentiment scores for each emotion as shown in Figure 5. While some plausible relationships can be seen (for instance the availability of more positive tweets with the "happiness" label), no clear relationship can be observed between most other emotions (such as "anger") and the sentiment values.

Figure 5: Distribution of sentiment scores for each emotion

5 Conclusion

In this study we presented a dataset of Emotion-labeled Persian Tweets and discussed the properties of the dataset. We believe this dataset is a valuable resource for future studies in Persian NLP. Future work could investigate models for the task of emotion detection and investigate emotions surrounding various topics on social media platforms.

References

Twitters doubling of character count from 140 to 280 had little impact on length of tweets. https://techcrunch.com/2018/10/30/twitters-doubling-of-character-count-from-140-to-280-had-little-impact-on-length-of-tweets/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAGtt9NmxoAHZ5quCVKsd9mzI7Qso9sy1WW3QjFWfLyoLuTuA0kETvOI3PnlMpVirGuqogD7pRxNSQEovhs-nyPmOYxQrDSzJd3vkhvQ-X3LvBk77KDYPhILYaski5hJ3BUr7inLE-r9laS7A-7XQRMetXJM1DXVVpUzNMvStbR7H.

(accessed December 1, 2020). *APA Dictionary of Psychology*. https://dictionary.apa.org/emotion.

(accessed December 1, 2020). *EMOCONTEXT : A SHARED TASK AT SEMEVAL 2019*. https://www.humanizing-ai.com/emocontext.html.

(accessed December 2, 2020). *Twitter Developer API*. https://developer.twitter.com/en/docs/api-reference-index.

(accessed January 2, 2021). *Itoll*. https://check.itoll.ir.

Parag Agrawal and Anshuman Suri. 2019. Nelec at semeval-2019 task 3: think twice before going deep. *arXiv preprint arXiv:1904.03223*.

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

Amer J Almahdawi and William J Teahan. 2019. A new arabic dataset for emotion recognition. In *Intelligent Computing-Proceedings of the Computing Conference*, pages 200–216. Springer.

Lisa Feldman Barrett, Batja Mesquita, Kevin N Ochsner, and James J Gross. 2007. The experience of emotion. *Annu. Rev. Psychol.*, 58:373–403.

Angelo Basile, Marc Franco-Salvador, Neha Pawar, Sanja Štajner, Mara Chinea Rios, and Yassine Benajiba. 2019. Symantoresearch at semeval-2019 task 3: combined neural models for emotion classification in human-chatbot conversations. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 330–334.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Yash Butala, Kanishk Singh, Adarsh Kumar, and Shrey Shrivastava. 2021. Team phoenix at wassa 2021: Emotion analysis on news stories with pre-trained language models. *arXiv preprint arXiv:2103.06057*.

Lea Canales and Patricio Martínez-Barco. 2014. Emotion detection from text: A survey. In *Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC)*, pages 37–43.

Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. Semeval-2019 task 3: Emocontext contextual emotion detection in text. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 39–48.

Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

Tim Dalgleish. 2004. The emotional brain. *Nature Reviews Neuroscience*, 5(7):583–589.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding.

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

Paul Ekman. 1999. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16.

Paul Ekman and Daniel Cordaro. 2011. What is meant by calling emotions basic. *Emotion review*, 3(4):364–370.

Paul Ekman, Wallace V Friesen, Maureen O'sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712.

Cristina Fernández, Juan C Pascual, Joaquim Soler, Matilde Elices, Maria J Portella, and Enrique Fernández-Abascal. 2012. Physiological responses induced by emotion-eliciting films. *Applied psychophysiology and biofeedback*, 37(2):73–79.

Agneta H Fischer, Patricia M Rodriguez Mosquera, Annelies EM Van Vianen, and Antony SR Manstead. 2004. Gender and culture differences in emotion. *Emotion*, 4(1):87.

Sujatha Das Gollapalli, Polina Rozenshtein, and See Kiong Ng. 2020. Ester: Combining word cooccurrences and word associations for unsupervised emotion detection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 1043–1056.

James J Gross and Robert W Levenson. 1997. Hiding feelings: the acute effects of inhibiting negative and positive emotion. *Journal of abnormal psychology*, 106(1):95.

Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. 2019. Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, 7(1):35–51.

Don H Hockenbury and Sandra E Hockenbury. 2010. *Discovering psychology*. Macmillan.

Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7360–7370.

Zhang Jingqing, Zhao Yao, Saleh Mohammad, et al. 2019. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv preprint arXiv: 1912.08777*.

S. Jung, J. Salminen, and B. J. Jansen. 2020. *Text2Sentiment (Version 1.0) [Computer software]*. Qatar Computing Research Institute. `https://qu ecst.qcri.org/tool/Text2Sentiment`.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, pages 3146–3154.

Ali Khosravi, Manoochehr Kelarestaghi, and Mehdi Purmohammad. 2019. Emotion detection in persian text; a machine learning model. *Contemporary Psychology, Biannual Journal of the Iranian Psychological Association*, 14(1):42–48.

Kuba Krys, C-Melanie Vauclair, Colin A Capaldi, Vivian Miu-Chi Lun, Michael Harris Bond, Alejandra Domínguez-Espinosa, Claudio Torres, Ottmar V Lipp, L Sam S Manickam, Cai Xing, et al. 2016. Be careful where you smile: Culture shapes judgments of intelligence and honesty of smiling individuals. *Journal of nonverbal behavior*, 40(2):101–116.

Yaman Kumar, Debanjan Mahata, Sagar Aggarwal, Anmol Chugh, Rajat Maheshwari, and Rajiv Ratn Shah. 2019. Bhaav-a text corpus for emotion analysis from hindi stories. *arXiv preprint arXiv:1910.04073*.

Ute Kunzmann and Daniel Grühn. 2005. Age differences in emotional reactivity: the sample case of sadness. *Psychology and aging*, 20(1):47.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Joseph E LeDoux. 1995. Emotion: Clues from the brain. *Annual review of psychology*, 46(1):209–235.

Chen Liu, Muhammad Osama, and Anderson De Andrade. 2019. Dens: a dataset for multi-class emotion analysis. *arXiv preprint arXiv:1910.11769*.

Kevin Mulligan and Klaus R Scherer. 2012. Toward a working definition of emotion. *Emotion Review*, 4(4):345–357.

Jay Mundra, Rohan Gupta, and Sagnik Mukherjee. 2021. Wassa@ iitk at wassa 2021: Multi-task learning and transformer finetuning for emotion classification and empathy prediction. *arXiv preprint arXiv:2104.09827*.

Andrew Ortony and Terence J Turner. 1990. What's basic about basic emotions? *Psychological review*, 97(3):315.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Pierre Philippot, Robert S Feldman, et al. 2004. *The regulation of emotion*. Psychology Press.

Rosalind W Picard. 1999. Affective computing for hci. In *HCI (1)*, pages 829–833. Citeseer.

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

Marco Polignano, Pierpaolo Basile, Marco de Gemmis, and Giovanni Semeraro. 2019. A comparison of word-embeddings in emotion detection from text using bilstm, cnn and self-attention. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pages 63–68.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. 2018. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, 8(1):28.

Armin Seyeditabari, Narges Tabari, and Wlodek Zadrozny. 2018. Emotion detection in text: a review. *arXiv preprint arXiv:1806.00674*.

Phillip R Shaver, Shelley Wu, and Judith C Schwartz. 1992. Cross-cultural similarities and differences in emotion and its representation.

Abu Awal Md Shoeb and Gerard de Melo. 2020. EmoTag1200: Understanding the association between emojis and emotions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8957–8967, Online. Association for Computational Linguistics.

Tiberiu Sosea and Cornelia Caragea. 2020. Canceremo: A dataset for fine-grained emotion detection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8892–8904.

Shabnam Tafreshi, Orphée De Clercq, Valentin Barriere, Sven Buechel, João Sedoc, and Alexandra Balahur. 2021. Wassa 2021 shared task: Predicting empathy and emotion in reaction to news stories. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 92–104.

Dianne A Van Hemert, Fons JR van de Vijver, and Ad JJM Vingerhoets. 2011. Culture and crying: Prevalences and gender differences. *Cross-Cultural Research*, 45(4):399–431.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Shari R Waldstein, Willem J Kop, Louis A Schmidt, Amy J Haufler, David S Krantz, and Nathan A Fox. 2000. Frontal electrocortical and cardiovascular reactivity during happiness and anger. *Biological psychology*, 55(1):3–23.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207.