# Enhancing contextualised language models with static character and word embeddings for emotional intensity and sentiment strength detection in Arabic tweets

Alharbi, Abdullah I.; Smith, Phillip; Lee, Mark

*DOI:*
10.1016/j.procs.2021.05.089

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 189 (2021) 258–265

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on AI in Computational Linguistics

# Enhancing Contextualised Language Models with Static Character and Word Embeddings for Emotional Intensity and Sentiment Strength Detection in Arabic Tweets

Abdullah I. Alharbi[a,b,*], Phillip Smith[a], Mark Lee[a]

[a] *School of Computer Science, University of Birmingham, Birmingham, UK*

[b] *Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, KSA*

**Abstract**

Many studies have focused on Arabic sentiment or emotion classification tasks. However, research on alternative aspects of affect, such as emotional intensity and sentiment strength tasks, has been somewhat limited. In this paper, we propose a method for enriching a contextualised language model that incorporates static character and word embeddings for emotional intensity and sentiment strength in Arabic tweets. We examine the assumption that models using static embeddings that are trained specifically on a corpus containing extensive Arabic affect-related words can boost the performance of language models. Through the development of character-level embeddings, we have found that our method is able to overcome the limitations associated with out-of-vocabulary words, which is a common problem when dealing with Arabic informal text. Given this, the method that we have developed achieves state-of-the-art results for the detection of the intensity of emotion and sentiment strength in Arabic social media.



*Keywords:* Contextualised Language Models; Character and Word Embeddings; Arabic Emotional Intensity

## 1. Introduction

Social media channels produce a significant amount of user-generated content that can form the basis for models developed to predict people's emotions, sentiments and opinions. In addition to employing language to share their emotions and sentiments, people also use what they write to communicate the intensity of their emotions. The term 'affect' is adopted as a reference to various categories of emotion. These categories typically range from classification of sentiment (from positive to negative) to classification of finer-grained sentiment strength (e.g. high positive to low positive) and emotional intensity (e.g. high anger to low anger). A number of studies have analysed Arabic sentiment

* Corresponding author.
*E-mail address:* aia784@bham.ac.uk

1877-0509 
10.1016/j.procs.2021.05.089

classification; however, studies on alternative aspects of affect (sentiment strength and emotional intensity) are limited [30].

Detecting affect from text can be challenging, particularly in the context of social media microblogs. This is primarily because of the challenges associated with the use of informal language, character limitations and use of symbols and slang. This task becomes even more complicated when morphology-rich languages, such as Arabic, are involved [6]. Social media communications typically consist of a range of dialects and sub-dialects that are not ruled by consistent standards. As such, there is a need for effective methods and resources that can be adopted to better comprehend and treat a variety of linguistic forms when seeking to understand affect in Arabic tweets.

Traditional or static word embedding has been used effectively for a range of natural language processing (NLP) tasks [22, 41, 27, 19]. It uses dense vectors to represent words projecting into a continuous vector space, thereby decreasing the number of dimensions. Most works on affect tasks were derived from static word embedding models, such as word2vec [28], GloVe [33] and fastText [18]. However, although these early frameworks have achieved some significant advances, they lack contextualised information. Recently, Bidirectional Encoder Representations from Transformer (BERT) [21] have been shown to be able to generate effective representation for NLP tasks in various contexts. These dynamic language models provide words with different representations based on the contexts of these words. Transformer-based language models are particularly useful for Arabic sentiment and emotion classification tasks [5, 7]. However, to the best of our knowledge, no work has employed contextualised word embeddings for Arabic emotional intensity and sentiment strength.

Recent research shows that the combination of static and dynamic word embeddings can benefit downstream tasks [37, 34, 42, 8]. In our work, we propose an approach to enhancing a contextualised language model [2] with the integration of static character and word embeddings [9] for affect in Arabic tweets. We expect that static embedding models trained specifically on a corpus that is rich in Arabic-affect-related words can boost the performance of language models. Furthermore, character-level embedding (CE) has been approved to overcome out-of-vocabulary (OOV) words [9]. Our proposed method achieves state-of-the-art results for Arabic emotional intensity and sentiment strength tasks.

The rest of this paper is organised in the following manner. Section 2 provides an overview of the related literature. The methodology for our proposed approach is summarised in Section 3. Section 4 explains the experimental setup (through which the proposed system was tested), the experimental results and a discussion of the outcomes. Section 5 ends the paper by presenting recommendations for future studies.

## 2. Related Works

### 2.1. Sentiment and Emotion Analysis Tasks

Over the last decade, Arabic Sentiment Analysis (SA) has attracted considerable attention due to the prevalence of opinions and sentiments in social media posts. The primary SA studies on Arabic have typically focused on the production of expensive resources that required human input to achieve acceptable levels of accuracy [24, 31, 16]. More recently, the standard of SA in Arabic has been enhanced by using different pre-trained language models [5]. The task of Arabic emotions classification followed the efforts spent in SA, trying to discover basic emotions conveyed within a given text; for instance, surprise, disgust, happiness, and anger [12, 25]. Such attempts have evolved in ways similar to the models and resources of SA.

While many studies have focused on sentiment and emotions classification tasks, research that aims to detect emotional intensity or sentiment strength is somewhat less common [29]. One notable study that focused on emotional intensity was the SemEval 2018 Task 1 (Affect in Tweets) [30]. The majority of the teams who performed effectively on the task used a combination of features originating from the existing lexicons associated with sentiments and emotions and different static word representations. The most effective method for Arabic emotional intensity proposed by AffecThor [1]. The researchers developed a system that employed various handcrafted lexicons in combination with a pre-trained word-level embedding framework that was based on a dataset consisting of 4 million tweets. These feature representations were used as an input features. They entered these features in a supervised learning framework accomplished with ensemble deep network architectures (Bidirectional Long-Short Term Memory (BiLSTM) with attention and convolutional neural network (CNN).

The second-best performing model was the EiTAKA system [26]. They proposed an ensemble model which combined N-Channels ConvNet, a deep learning approach, and XGBoost regressor, which focuses on a compilation of features based on embedding and lexicons. These integrated models assisted in delivering improved performance for the affect downstream tasks. The third best performing emotional intensity regression was the system proposed by the UNCC [3], while the third-best performing emotional intensity classification system was the EMA system [17]. These proposed methods used static pre-trained word-level embeddings within supervised learning models. The EMA system used applied stems as their processing techniques as opposed to lemmas because Arabic Twitter users typically use dialectical Arabic, and the majority of Arabic morphological analysers are trained using MSA data.

### 2.2. *Pre-trained Arabic Language Models*

The majority of the existing studies on Arabic static word embedding have focused on the use of word-level models [40, 39, 4, 14] and, to a lesser degree, character-level models [13, 9]. AraVec [39] is considered as one of the most popular static word embeddings that span six distinct word embedding frameworks for use with the Arabic language. The training data used to develop AraVec was extracted from three main sources: Wikipedia, Twitter, Common Crawl. They utilised skip-gram and CBOW to learn word representations for general Arabic natural language processing purposes. Mazajak [4] used over 250 million Arabic tweets to produce the largest word-level embedding model. Although they utilised a significant number of words to train the models, they were unable to identify the same words in alternative forms as employed within everyday speech due to the limitations of these word-level models. Generally speaking, the embedding effectiveness is largely dependent on the task [35] and is greatly affected by the range of words used to perform a given task [20].

Recently, [9] has created two static language models at different levels (character and word) to specifically target affect tasks. They collected 10 million tweets using a strategy to ensure the corpus was enriched with a variety of affect-related words from different Arabic variations. They employed Word2vec [28] for generating word-level embedding, and FastText [18] for the character-level embedding model. Finally, they proposed an approach to combine these models, that they refer to as Affect Character Word Embeddings (ACWE).

Static word embeddings take into consideration a full array of sentences in which a word is used as a means of generating a universal vector representation of it. However, the meaning of a given word can vary according to the context within which it is used. This led to generate contextualised word embeddings specifically for Arabic such as AraBERT [15] and MARBERT [2]. Both models have been applied to varied sentiment and emotion classification task, achieving the most effectiveness performance cross a number of benchmark datasets [11, 32, 36, 10]. However, we are not aware of any existing work that employs contextualised word embeddings for Arabic emotional intensity and/or sentiment strength.

## 3. Methodology

Our proposed methodology integrates static character and word embeddings (CE and WE) and contextualised embeddings (MARBERT). We combine these models at the sentence level, following a similar approach proposed by [8]. However, instead of using static word-level embeddings (WE) only, we incorporated CE to overcome the OOV problem. An overview of our proposed approach is illustrated in Figure 1. The following subsections explain each one of these models and how our proposed approach incorporates them.

### 3.1. *Pre-processing*

We used pre-processing techniques that have been employed in a range of studies [4, 23]. We started by removing unrecognisable symbols and any character that is not useful or used in Arabic, such as diacritics and punctuation marks. We did not remove the emojis because they are often of value in sentiment and emotion analysis tasks. Moreover, we normalised letters that appeared in different forms and re-rendered them in a single expression. For instance, the ‘hamza’on characters ( إ, أ ) was replaced with the ( ا ), and the ’t marbouta’ ( ة ) was replaced with ( ه ).

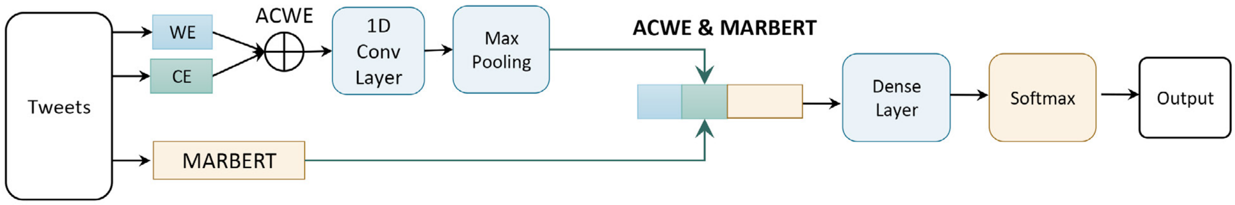


Fig. 1. The proposed system architecture.

### 3.2. Static Word Embeddings

Word embedding is a process by which text is represented in the form of a dense vector that represents how sets of words are semantically similar by positioning them near each other within the vector. Every word is encoded in a real-valued vector that consists of a large number of dimensions. Two word embedding levels were used in this study: WE and CE.

- **WE:** we employed a pre-trained WE model generated by [9]. This model was trained on a massive number of Arabic tweets that are rich in affect-related words from a variety of Arabic dialects. It was built by using the word2vec algorithm [28] to learn individual words and their representations. We used a large number of words to train the ara2vec model; however, a model of this nature cannot feasibly process every single word that is used in an everyday setting. Therefore, the OOV problem, in which WE cannot identify uncommon words, is expected to arise. This is one of the disadvantages of using a WE model.
- **CE:** Since a significant number of different dialects are spoken in Arabic, more significant OOV issues can develop with this language than with other languages. Alharbi and Lee [9] stated that WE can deliver high value in terms of semantic similarities; however, CE can more efficiently encode variations in word morphology and align them better within vector representations. Therefore, we employed their pre-trained character representation model (CE), which has been proven to perform well in Arabic affect tasks. Their CE was produced by training fastText [18] on a dataset containing 10 million tweets. This corpus included multiple affect words (i.e. words that express emotions and sentiments of varying intensity levels) from numerous Arabic dialects.

To this end, we followed [9] in that we combined static character- and word-level models to take advantage of each of them and to semantically and morphologically represent words in our Arabic affect tasks. The model resulting from this integration is called Affect Character and Word Embeddings (ACWE). Hence, we employed the convolutional neural network (CNN) proposed by [23] and using ACWE to set the weights of the embedding layer. These weights were subsequently modified during the training process to ensure that they are suitable for the tasks before different kernels and filters were used to produce max pooled features. Finally, these features were concatenated with the output representations of the contextualised embedding model explained in the following subsection.

### 3.3. Contextualised Embedding Model

One popular model of contextualised or dynamic language models is BERT, which was created by the Google research team and has been proven to be highly effective in a myriad of NLP tasks [21]. Given that the dataset involved in our downstream tasks consisted of multiple Arabic dialects, we chose to use a pre-trained model that was specifically created for Arabic dialect tasks: MARBERT [2]. Pre-trained on a vast dataset containing 6 billion tweets, MARBERT produces state-of-the-art outcomes in numerous tasks involving Arabic-language NLP. This model employs 12 attention headers, 12 encoder blocks and 768 hidden dimensions. It can process sequences of up to 512 tokens. When input into a model, these tokens create a sequence representation. The first token in the sequence token is [CLS], which comprises targeted classification embedding. Following the authors, we implemented the model and fine-tuned it in accordance with each downstream task.

### 3.4. Combining MARBERT and ACWE

We integrated static ACWE models and contextualized MARBERT at the tweet level. Figure 1 presents the overview architecture of our proposed method. After fine-tuning MARBERT on the training dataset for each task, we retrieved the contextualised vectors (each having average of 12 hidden layers). For ACWE, we obtained the feature vector after the CNN model training (explained in section 3.2). We simply concatenated both obtained vectors which were linked with the remaining network layers. Finally, after applying a dropout, the final concatenated vectors were forwarded to a dense layer with Mean Square Error activation function for regression tasks and Softmax activation for classification tasks.

## 4. Experiments

### 4.1. Datasets

Our proposed method was evaluated using four affect datasets released by the organisers in the SemEval 2018 shared task (Affect in Tweets) [30]. The four datasets used in our experiments are described as follows:

- **Emotion Intensity Regression Task (EI-reg ):** This task involves the use of four sub-sets, one to represent each emotion: joy, sadness, fear, and anger. The aim is to identify the emotional intensity (EI) expressed within a given tweet. The data consists of 1800 Arabic tweets separated into three sets for each emotion: a training, development, and test set. The EI-reg task was annotated by a real-value score that varied from zero (lowest intensity) to one (highest intensity). The details of the EI-reg dataset can be observed in Table 1.
- **Emotion Intensity ordinal classification Task (EI-oc):** This task is similar to EI-reg. However, its main objective is to predict EI classes using a value between 0 and 3. 0 represents an unrelated emotion, 1 signifies the lowest EI, and 3 signifies the highest EI. The details of the EI-oc dataset can be observed in Table 1.
- **Valence Intensity regression Task (V-reg ):** The aim of the task is to use a real-value score to predict the valence (V) that is best aligned with the sentiment strength or valance represented within a given tweet. The V-reg task scores range from 0 (most negative) to 1 (most positive). The details of the V-reg dataset can be observed in Table 1.
- **Valence Intensity ordinal classification Task (V-oc):** The goal of this task is to classify a given tweet using one of seven class labels, which range from -3 (very negative) to +3 (very positive). Neutral class is represented by a 0 score. The details of the V-oc dataset can be observed in Table 1.

Table 1. Number of tweets in *EI-reg*, *EI-oc*, *V-reg* and *V-oc* datasets and the statistics of the datasets splits.

| Task | Emotion | Labels | Train | Dev | Test | Total |
|---|---|---|---|---|---|---|
| | anger | o to 1 | 877 | 150 | 373 | 1,400 |
| EI-reg/ | fear | (real-value)/ | 882 | 146 | 372 | 1,400 |
| EI-oc | joy | 0,1,2,3 | 728 | 224 | 448 | 1,400 |
| | sadness | (classes) | 889 | 141 | 370 | 1,400 |
| V-reg/V-oc | - | real-value/ 7 classes | 932 | 138 | 730 | 1,800 |

### 4.2. Results

To evaluate the outcomes of our experiments we utilised Pearson's correlation coefficient, which measures a bivariate linear correlation between two given variables. Pearson's was used as an evaluation metric as it is the official measurement metric for the benchmark datasets. The coefficient showed a link between the score expected by our systems and the score provided by the test results in our experiments.

Table 2 summarises the performance of ACWE and MABERT when they were used individually and of our proposed approach when they were combined. The Pearson correlation result for the dynamic MARBERT model significantly outperformed that of the static ACWE model. This outcome confirms that contextualised word embedding models can also provide outstanding performance in emotional intensity and sentiment strength tasks, as demonstrated in previous studies on sentiment polarity or basic emotion classification tasks. Moreover, the proposed method revealed an improvement from about 1% to 3% cross all four tasks. This result demonstrates the performance of the proposed method and the importance of leveraging dynamic and static word embeddings in Arabic affect tasks within the context of social media microblogs.

Overall, the proposed approach improved the effectiveness across all four affect tasks. The enhanced results can be explained by the fact that ACWE trained on a large corpus that was built specifically for the domain of affect tasks, whereas MARBERT trained on an enormous dataset for a general domain. Therefore, the combination approach enhances the quality of the final representations with additional information. A recent study [38] also found that contextualised language models still struggle to understand rare words even though they trained with vast data. Thus, leveraging these models with a combination of character and word embeddings can enhance the performance across all affect tasks.

Table 2. Pearson correlation coefficient results using MARBERT alone and the proposed combination method cross four affect datasets.

| Model | EI-reg | | | | | EI-oc | | | | | V-reg | V-oc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | anger | fear | joy | sad | avg | anger | fear | joy | sad | avg | | |
| Previous SOTA | 64.7 | 64.2 | 75.6 | 69.4 | 68.5 | 55.1 | 55.1 | 63.1 | 61.8 | 58.7 | 82.8 | 80.9 |
| MARBERT | 69.0 | 70.6 | 78.4 | 70.6 | 72.1 | 56.0 | 64.4 | 68.9 | 65.6 | 63.7 | 85.4 | 81.2 |
| ACWE+MARBERT | **69.9** | **70.9** | **79.9** | **71.1** | **73.0** | **57.4** | **67.8** | **70.9** | **66.6** | **65.7** | **86.1** | **83.1** |

## 5. Conclusion

In this paper, we propose a method for enhancing a contextualised language model by incorporating static character and word embeddings for affect tasks, particularly emotional intensity and sentiment strength identification. We obtain outstanding results, significantly outperforming previous state-of-the-art cross four affect tasks. Our proposed method indicates the importance of leveraging static affect-specific representations and dynamic language models. Training BERT on large datasets from scratch is time-consuming and requires high-performance hardware that is not always available. Alternatively, statistic word embedding algorithms can be used at two levels (character and word embeddings) and then combined with pre-trained BERT models. In future work, we will explore more advanced approaches to combine static and dynamic language models. Moreover, given the fact that the dataset is a small size , we will investigate data augmentation strategies to enable deep neural network learning from more samples during the training stage.

## References


[1] Abdou, M., Kulmizev, A., Ginés i Ametllé, J., 2018. AffecThor at SemEval-2018 task 1: A cross-linguistic approach to sentiment intensity quantification in tweets, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana. pp. 210–217.

[2] Abdul-Mageed, M., Elmadany, A., Nagoudi, E.M.B., 2020. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. arXiv:2101.01785.

[3] Abdullah, M., Shaikh, S., 2018. TeamUNCC at SemEval-2018 task 1: Emotion detection in English and Arabic tweets using deep learning, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana. pp. 350–357.

[4] Abu Farha, I., Magdy, W., 2019. Mazajak: An online Arabic sentiment analyser, in: Proceedings of the Fourth Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Florence, Italy. pp. 192–198.

[5] Abu Farha, I., Magdy, W., 2021. A comparative study of effective approaches for Arabic sentiment analysis. Information Processing Management 58, 102438.

[6] Al-Ayyoub, M., Khamaiseh, A.A., Jararweh, Y., Al-Kabi, M.N., 2019. A comprehensive survey of Arabic sentiment analysis. Information Processing & Management 56, 320–342.

[7] Al-Twairesh, N., 2021. The evolution of language models applied to emotion analysis of arabic tweets. Information 12, 84.

[8] Alghanmi, I., Espinosa Anke, L., Schockaert, S., 2020. Combining BERT with static word embeddings for categorizing social media, in: Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), Association for Computational Linguistics, Online. pp. 28–33.

[9] Alharbi, A.I., Lee, M., 2020. Combining character and word embeddings for affect in Arabic informal social media microblogs, in: Métais, E., Meziane, F., Horacek, H., Cimiano, P. (Eds.), Natural Language Processing and Information Systems, Springer International Publishing, Cham. pp. 213–224.

[10] Alhuzali, H., Abdul-Mageed, M., Ungar, L., 2018. Enabling deep learning of emotion with first-person seed expressions, in: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, Association for Computational Linguistics, New Orleans, Louisiana, USA. pp. 25–35.
[11] Alomari, K.M., ElSherif, H.M., Shaalan, K., 2017. Arabic tweets sentimental analysis using machine learning, in: Benferhat, S., Tabia, K., Ali, M. (Eds.), Advances in Artificial Intelligence: From Theory to Practice, Springer International Publishing, Cham. pp. 602–610.
[12] Alswaidan, N., Menai, M.E.B., 2020. Hybrid feature model for emotion recognition in Arabic text. IEEE Access 8, 37843–37854.
[13] Altowayan, A.A., Elnagar, A., 2017. Improving Arabic sentiment analysis with sentiment-specific embeddings, in: 2017 IEEE International Conference on Big Data (Big Data), IEEE, Boston. pp. 4314–4320.
[14] Altowayan, A.A., Tao, L., 2016. Word embeddings for Arabic sentiment analysis, in: 2016 IEEE International Conference on Big Data (Big Data), IEEE, Washington. pp. 3820–3825.
[15] Antoun, W., Baly, F., Hajj, H., 2020. AraBERT: Transformer-based model for Arabic language understanding, in: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection, European Language Resource Association, Marseille, France. pp. 9–15.
[16] Badaro, G., Baly, R., Hajj, H., Habash, N., El-Hajj, W., 2014. A large scale Arabic sentiment lexicon for Arabic opinion mining, in: Proceedings of the EMNLP 2014 workshop on arabic natural language processing (ANLP), pp. 165–173.
[17] Badaro, G., El Jundi, O., Khaddaj, A., Maarouf, A., Kain, R., Hajj, H., El-Hajj, W., 2018. EMA at SemEval-2018 task 1: Emotion mining for Arabic, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana. pp. 236–244.
[18] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics 5, 135–146.
[19] Bordes, A., Chopra, S., Weston, J., 2014. Question answering with subgraph embeddings, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar. pp. 615–620.
[20] Çano, E., Morisio, M., 2017. Quality of word embeddings on sentiment analysis tasks, in: Natural Language Processing and Information Systems, Springer International Publishing, Cham. pp. 332–338.
[21] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis. pp. 4171–4186.
[22] Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., Makhoul, J., 2014. Fast and robust neural network joint models for statistical machine translation, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, Maryland. pp. 1370–1380.
[23] Duwairi, R., El-Orfali, M., 2014. A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. Journal of Information Science 40, 501–513.
[24] Eskander, R., Rambow, O., 2015. SLSA: A sentiment lexicon for Standard Arabic, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal. pp. 2545–2550.
[25] Hifny, Y., Ali, A., 2019. Efficient Arabic emotion recognition using deep neural networks, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6710–6714.
[26] Jabreel, M., Moreno, A., 2018. EiTAKA at SemEval-2018 task 1: an ensemble of n-channels ConvNet and XGboost regressors for emotion analysis of tweets, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana. pp. 193–199.
[27] Lin, C.C., Ammar, W., Dyer, C., Levin, L., 2015. Unsupervised POS induction with word embeddings, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Denver, Colorado. pp. 1311–1316.
[28] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., 2013. Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, Curran Associates Inc., Red Hook, NY, USA. p. 3111–3119.
[29] Mohammad, S., Bravo-Marquez, F., 2017. Emotion intensities in tweets, in: Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017), Association for Computational Linguistics, Vancouver, Canada. pp. 65–77.
[30] Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S., 2018. SemEval-2018 task 1: Affect in Tweets, in: Proceedings of The 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana. pp. 1–17.
[31] Mourad, A., Darwish, K., 2013. Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs, in: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 55–64.
[32] Nabil, M., Aly, M., Atiya, A., 2015. ASTD: Arabic sentiment tweets dataset, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal. pp. 2515–2519.
[33] Pennington, J., Socher, R., Manning, C., 2014. GloVe: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar. pp. 1532–1543.
[34] Peters, M.E., Neumann, M., Logan, R., Schwartz, R., Joshi, V., Singh, S., Smith, N.A., 2019. Knowledge enhanced contextual word representations, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China. pp. 43–54.
[35] Qu, L., Ferraro, G., Zhou, L., Hou, W., Schneider, N., Baldwin, T., 2015. Big data small data, in domain out-of domain, known word unknown word: The impact of word representations on sequence labelling tasks, in: Proceedings of the Nineteenth Conference on Computational Natural Language Learning, Association for Computational Linguistics, Beijing, China. pp. 83–93.
[36] Rosenthal, S., Farra, N., Nakov, P., 2017. SemEval-2017 task 4: Sentiment analysis in Twitter, in: Proceedings of the 11th International

Workshop on Semantic Evaluation (SemEval-2017), Association for Computational Linguistics, Vancouver, Canada. pp. 502–518.

[37] Roy, A., Pan, S., 2020. Incorporating extra knowledge to enhance word embedding, in: Bessiere, C. (Ed.), Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, International Joint Conferences on Artificial Intelligence Organization. pp. 4929–4935. Survey track.

[38] Schick, T., Schütze, H., 2020. Rare words: A major problem for contextualized embeddings and how to fix it by attentive mimicking, in: The Tenth EAAI Symposium on Educational Advances in Artificial Intelligence, New York, USA,, AAAI Press. pp. 8766–8774.

[39] Soliman, A.B., Eissa, K., El-Beltagy, S.R., 2017. AraVec: A set of Arabic word embedding models for use in Arabic NLP. Procedia Computer Science 117, 256–265.

[40] Zahran, M.A., Magooda, A., Mahgoub, A.Y., Raafat, H., Rashwan, M., Atyia, A., 2015. Word representations in vector space and their applications for Arabic, in: Gelbukh, A. (Ed.), Computational Linguistics and Intelligent Text Processing, Springer International Publishing, Cham. pp. 430–443.

[41] Zhang, J., Liu, S., Li, M., Zhou, M., Zong, C., 2014. Bilingually-constrained phrase embeddings for machine translation, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Baltimore, Maryland. pp. 111–121.

[42] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q., 2019. ERNIE: Enhanced language representation with informative entities, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy. pp. 1441–1451.