

Improving Arabic Cognitive Distortion Classification in Twitter using BERTopic

Fatima Alhaj¹
The University of Jordan,
Amman, Jordan

Ali Al-Haj²
University of Suffolk
Ipswich, United Kingdom

Ahmad Sharieh³, Riad Jabri⁴
The University of Jordan,
Amman, Jordan

Abstract—Social media platforms allow users to share thoughts, experiences, and beliefs. These platforms represent a rich resource for natural language processing techniques to make inferences in the context of cognitive psychology. Some inaccurate and biased thinking patterns are defined as cognitive distortions. Detecting these distortions helps users restructure how to perceive thoughts in a healthier way. This paper proposed a machine learning-based approach to improve cognitive distortions' classification of the Arabic content over Twitter. One of the challenges that face this task is the text shortness, which results in a sparsity of co-occurrence patterns and a lack of context information (semantic features). The proposed approach enriches text representation by defining the latent topics within tweets. Although classification is a supervised learning concept, the enrichment step uses unsupervised learning. The proposed algorithm utilizes a transformer-based topic modeling (BERTopic). It employs two types of document representations and performs averaging and concatenation to produce contextual topic embeddings. A comparative analysis of F1-score, precision, recall, and accuracy is presented. The experimental results demonstrate that our enriched representation outperformed the baseline models by different rates. These encouraging results suggest that using latent topic distribution, obtained from the BERTopic technique, can improve the classifier's ability to distinguish between different CD categories.

Keywords—Arabic tweets; cognitive distortions' classification; machine learning; social media; supervised learning; unsupervised learning; transformers; BERTopic; topic modeling

I. INTRODUCTION

Researchers attempt to understand the psychological well-being of users by analyzing the published social media content [1], [2]. This content may hold negative, and inaccurate conclusions called cognitive distortions (CDs). Research showed that CDs could be detected in social media [3]. In cognitive psychology, CDs are biased perspectives, patterns, and beliefs that have been reinforced over the years [4]. Usually, individuals diagnosed with depression express higher levels of distorted thinking than others [5], which affect the content they share.

Moreover, CDs can lead to poor behavior and chronic anxiety and are related to symptoms of depression [6]. Examples of these categories are overgeneralization, emotional reasoning, and catastrophizing. Until being diagnosed, a CD can be so problematic to change. Machine learning (ML) and natural language processing (NLP) can improve mental health care by defining CDs within a text. Classifying CDs can be beneficial to the therapy process in two ways. It helps in replacing these inaccurate thoughts and evaluating the improvement over time.

While many previous works with CDs showed promising results with detecting CD, they faced disappointing results with the multi-class CD classification [7], [4], [8]. This work aims to improve the classification task by dealing with the text shortness problem. In general, many classification tasks working with short text fails to achieve high performance according to the sparse representation of the textual data [9], [10], [11]. Extra contextual information can be deployed to overcome the sparse representation and make the data more related, comprehensively expressed, and expand the efficiency of classifiers to handle unseen data.

Contextual topic modeling (TM) can provide extra latent information and identify semantically similar groups in the corpus. The topic-based embeddings can learn global semantics from the entire corpus [12], which provides the features with an opportunity to become more representative.

Also, while analyzing textual cognitive distortions, we noticed that individuals mostly use a common CD when expressing their thoughts about a specific topic or domain. For example, catastrophizing is highly related to relations, self-image terms, and academic achievement. General terms like life, humans, country and government are mostly related to over-generalizing and labeling. This observation encouraged us to utilize TM as extra contextual information in CD classification.

The TM algorithm used in this work is a transformer-based algorithm that utilizes the pre-trained knowledge in the modeling process to provide a topic distribution. So, our primary contribution in this paper is using the state-of-the-art transformer-based TM algorithm to enrich the CD representations and improve its classification task by exploring potential semantically meaningful categories in the data.

The remainder of the paper is structured as follows. In Section II, a summary of related work is presented. Section III overviews some background. Section IV presents the design and methodology used in the research, while the experimental results are analyzed in Section V. Finally, Section VI discuss conclusions.

II. RELATED WORK

The proposed approach utilizes the TM algorithm to overcome the shortness of social media text to improve CD classifications. Accordingly, this research is related to three main topics; CD classification, short text classification, and TM in text classification.

TABLE I. A SUMMARIZED SAMPLE OF CD-RELATED STUDIES

Ref.	Strength	Weakness
[4]	Using a representative dataset collected from multiple resources; crowdsourcing volunteers, mental health patients, and online therapy programs.	Authors reported F-score equal to 0.45, but the trained model cannot predict 7 out of 15 distortions.
[7]	<ul style="list-style-type: none">• Detecting and classifying CD from patient-therapist interactions.• Used a publicly available dataset called: Therapist Q&A, obtained from the Kaggle data science repository.• Represent text using different types of linguistic features.	<ul style="list-style-type: none">• Poor inter-annotator agreement.• Unclear distinction between different distortions.• Misclassification according to the presence of multiple types of distortions in a single text.• Detection of CD type fails to yield good results. Weighted F1-score did not score more than 0.30.
[13]	The proposed technology determines whether or not the input statement includes cognitive distortions (binary classification). When detecting distortion in a statement, the approach uses "cognitive restructuring" to reframe the statement to repair the distortion.	<ul style="list-style-type: none">• Authors did not utilize ML methods. Participants were asked to classify statements as distorted or undistorted.• The study only evaluates the empathy represented in structured and unstructured responses.
[14]	<ul style="list-style-type: none">• Uses a large scale of real-world materials from daily narration and diaries from books and web pages in the cognitive-behavioral therapy domain.• Utilizing deep learning techniques like word2vec and CNN.	<ul style="list-style-type: none">• The dataset is not available publicly.• The approach was described without experimental results or evaluation techniques.
[8]	Proved the possibility of detecting cognitive distortions (binary classification) automatically from personal blogs with relatively good accuracy (73.0%).	Experiments showed a poor performance in multi-class classification, which was justified by the limited size of individual posts and the overall dataset.

Researchers applied ML and NLP to look for mental health and cognitive psychology inferences. Many attempts were made to detect and classify a defined set of cognitive distortions in the English language. Table I summarizes a sample of this studies and describes the related advantages and limitations. Most of the proposed work in Table I couldn't classify CD efficiently. Besides, datasets were collected from different resources, and most of them are not publicly available. The proposed approach creates a dataset by crawling social media to overcome this constraint. Also, we utilized the pre-trained bidirectional encoder representations from transformers (BERT) model to improve classification performance.

The limited number of words in tweets leads to sparse co-occurrence patterns, making the classification task more challenging. To tackle the feature sparseness, works in the literature choose between two main approaches; either represent texts in a lower-dimensional space [9] or add external, implicit, and valuable information to enhance the data representation on the feature space [10], [11]. The second method works on enriching the representation of a short text using additional knowledge or semantics [15]. These semantics could be derived internally [16], or from an external resource, such as a collection of longer documents in a similar domain, or a huge resource such as Wikipedia and WordNet [16], [17], [18]. While the mentioned approaches require additional datasets to enrich the text representation, the proposed approach analyzes the topic distribution within the same dataset.

TM methods have been used to improve text classification in the literature. Anantharaman et al. [19] surveyed a set of TM algorithms for classifying short text and large text (document). Results suggested that the latent semantic analysis method (LSA) is the best for short text classification tasks, while latent dirichlet allocation (LDA) performs better on document classification. In addition, Albalawi et al. [20] investigated TM methods by comparing five frequently used methods. Methods were applied to short textual social data to show their capabilities in defining key and meaningful topics. Their study indicates that TM can overcome the noisy data contained in a short text. Also, an efficient application by Phan et al. [21] classified short medical text by exploiting the

hidden topics in large-scale data collections (Wikipedia and MEDLINE datasets).

Furthermore, Dai et al. [22] extended the bag-of-words representation and introduced a new feature space by using a hierarchical clustering algorithm. Yajian et al. [23] worked on extending the short text to a reasonably long one. In their approach, weighted synonyms and related words are generated for every word by the Word2vec and LDA model.

Most proposed approaches use traditional TM algorithms, such as LDA, a word co-occurrence-based method that only works efficiently with long text. However, the previous methods improved the classification results to some extent but still leave considerable space for improvement.

The proposed approach considers the shortness of social media text in CD classification, which (to the best of our knowledge) is not explored in the literature. The novelty of this approach is based on employing a BERT-based model (BERTopic) for the TM, which provides a better contextual representation. The BERTopic model was employed previously for Arabic text modeling in [24]. It showed a better performance than non-negative matrix factorization (NMF) and LDA.

III. BACKGROUND

A. Cognitive Distortions

Social media textual content represents a rich resource for natural language processing techniques to make inferences in cognitive psychology. Machine learning approaches can be applied to learn the distorted thinking patterns to perform CD detection and classification. Table II shows the five considered CDs used in this work.

B. Term Frequency-Inverse Document Frequency

One of the weighting methods for feature-based representation is term frequency-inverse document frequency (TF-IDF). It provides a weight for each word. A TF-IDF value is determined by the relative frequency of a word in a specific text and the inverse proportion of the word over the entire corpus, which reflects how relevant a word is to a particular

TABLE II. COGNITIVE DISTORTION AND THEIR DEFINITIONS

Cognitive distortion	Definition
Inflexibility	"Having rigid beliefs about how things or people must or ought to be [25].
Over-generalizing	Creation negative conclusions without full evidence [5].
Labeling	To describe self or others negatively without giving credit to counterevidence [25].
Emotional reasoning	Depend on feelings as a source of facts [5].
Catastrophizing	Enlarging negative statements or events into disasters [25].

text. The TF-IDF is given by the equation 1. Where the weight of the word i in the text (document) j is w_{ij} . N is the number of documents, and tf_{ij} is the frequency of the word i in the document j , and df_i is the number of documents that contain the word i [26].

$$W_{ij} = tf_{ij} * \log \frac{N}{df_i} \quad (1)$$

C. Topic Modeling with BERT

Topic modeling (TM) is an analytical unsupervised learning model that discovers topics distribution in a corpus. In this context, a topic can be defined as a repeated pattern of terms [27]. Mainly, the TM algorithm provides two distributions; document-topic and topic-term.

Recently, a transformer-based algorithm was deployed for TM called BERTopic [28]. The algorithm uses three primary phases to produce a topic's distribution for a set of documents. First, it creates sentence embedding. Second, it creates clusters of semantically similar sentences. The last step includes creating topic representation with c-TF-IDF. Each step will be elaborated on in the methodology section.

The usual TF-IDF equation defines the importance of a word between different documents. Differently, the class-based TF-IDF (c-TF-IDF) [28] defines the importance of a word within a class (topic). It treats all documents in a single class as a single document. Equation (2) finds the c-TF-IDF of each word, where W_{ic} is the weight of word i in class c . The frequency of word i in class c is tf_{ic} . A is the average number of a word per class, and f_i is the frequency of word i across all classes.

$$W_{ic} = tf_{ic} \times \log\left(1 + \frac{A}{f_i}\right) \quad (2)$$

Essentially, the BERTopic technique utilized two Algorithms; UMAP and HDBSCAN. Uniform manifold approximation and projection (UMAP) [29] is an algorithm for non-linear dimension reduction that combines manifold learning and topological data analysis. The algorithm optimizes a low-dimensional graph to be structurally similar to the original graph. While HDBSCAN algorithm [30] is a hierarchical clustering algorithm that stands for hierarchical density-based spatial clustering of applications with noise. The algorithm first transforms the space according to the density or sparsity of the data to provide a distance-weighted graph, and then it creates the minimum spanning tree for this graph. The algorithm also represents the connected components by a cluster hierarchy

then condenses them based on the minimum cluster size parameter. Finally, it extracts clusters from the condensed tree.

IV. METHODOLOGY

In general, usual text mining methods have some limitations in classifying short texts. The limitations are related to the sparsity of co-occurrence patterns in short texts, and the lack of context information (semantic features) [31]. A common method to overcome these limitations is to enrich the texts with additional information to make data more related and extend the classifiers' coverage to perform better in future data [21]. The additional information can be directly attached to the textual representation. After that, the standard classification is performed.

The proposed approach in this paper contains two phases; the enrichment phase and the classification phase. Two main steps are performed in the enrichment phase; the first step is acquiring the additional information, while the second step combines them with the original data representation.

In this work, the additional information source is the output of a TM algorithm (described briefly in the background section). The algorithm generates probability distributions of the hidden topics in the corpus. BERTopic [28] is used as a TM technique, where it takes advantage of BERT embeddings. The BERTopic technique produces the distribution of topics within a document d . Later, this distribution (BERTopic (doc_d)) is combined with text embedding.

Fig. 1 shows an overview of the proposed approach that uses BERTopic processes to enrich the document representation with the hidden topic distribution. The following steps summarize the BERTopic processes:

- 1) Generate embeddings for tweets to provide numerical vector representations, where each tweet is considered a single document. This step is based on a pre-trained deep bidirectional transformer. This type of embedding is very powerful for language understanding and can capture the semantic relations between words.
- 2) Reduce the dimensionality of the embedding vectors to create a lower-dimensional embedding of tweets by using the UMAP algorithm [29]. The algorithm preserves the local and global structure in lower dimensionality. Cosine-similarity is used as a distance metric to measure distances between data points.
- 3) Perform clustering using HDBSCAN [30] to find highly similar text in the semantic space that produces a single topic. Defining the number of topics in this stage is a crucial decision. It is supposed to be reasonable and compatible with the dataset size and topics-related words. The generated topics can be too fine-grained, which results in a massive number of topics. Also, it can dismiss important details by defining a small number of clusters. In BERTopic, the algorithm infers the ideal number from the data itself. This step applies iterative merging for the most similar topics. Cosine

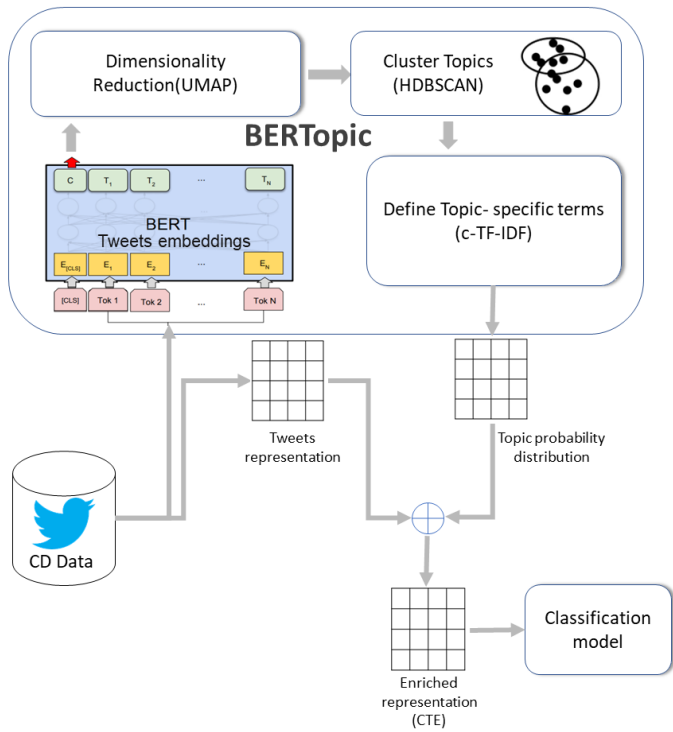


Fig. 1. A Graphical Representation of the TM Enrichment Approach for CD Classification.

similarity between c-TF-IDF vectors is utilized in this process.

- 4) Define the essential words in a topic according to the c-TF-IDF equation. Words with a high score are more representative of the topic. Accordingly, every tweet is related to a set of topics with different probabilities. In this step, BERTopic generates a topic distribution vector for each tweet, $Pr(topic | tweet)$. A graphical representation map for this process is shown in Fig. 2.

The previous processes produce topic distribution for each document as $BERTopic(doc_d)$. Then, the distribution is used in Algorithm 1 to produce the final contextual topic embedding (CTE) for each document. Mainly, the algorithm uses two types of document representations; topic distribution and word embedding. Then, it performs averaging and concatenation. The nested for (line 2-4) creates word embeddings for each word in a document d using the word2vec model, while line 5 constructs document embeddings. All word vectors in document d are averaged to produce a single vector in the same embedding space.

Combining BERTopic distribution and word2vec representation keeps the semantic information and creates contextual topic representation. The topic representation and the word embedding are composed in line 6 to produce CTE for a given document (d), where \oplus represents the arithmetic concatenation operator. We use "featureUnion" function¹ in Python to

¹<https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.FeatureUnion.html>

perform the concatenation.

Algorithm 1: Generate Contextual Topic Embedding

input : M : Maximum sequence of words,
 D : Number of documents,
 $doc_d = \{w_1, \dots, w_M\}$.
 $BERTopic(doc_d)$

output: $CTE(doc_d)$

- 1 **for** $d \in \{1, \dots, D\}$ **do**
- 2 **for** $n \in \{1, \dots, M\}$ **do**
- 3 | Embedding(w_{dn})= word2vec(w_{dn})
- 4 **end**
- 5 Embedding(doc_d)= Avg(Embedding(w_{d1}),...,Embedding(w_{dm}))
- 6 $CTE(doc_d)$ =Embedding(doc_d) \oplus $BERTopic(doc_d)$
- 7 **end**

The next step is feeding the enriched representation $CTE(doc_d)$ to different classifiers and evaluating the results. Because this approach uses hidden topic knowledge of the corpus as the primary source of enrichment, its effectiveness is corpus-dependent.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The dataset was collected by crawling Twitter data. First, a translated version of the cognitive distortion scale [32] was utilized. Volunteers were asked to express distorted thoughts they had. The initial responses were used to define a keyword list for each category. Keywords were defined as the top frequent and influential words in the responses. For example, one of the CD patterns is inflexibility. The related seed words were: (لازم / يجب / لابد / واجب / مفروض) / (محتّم / المفترض), which mean: (obligatory / must / must / obligatory / obligatory / unavoidable / supposed) respectively. The defined seed words were used for crawling Twitter by streaming API in June 2021.

Tweets were labeled by two reliable annotators who work in the psychotherapy domain. The annotators labeled the data independently of one another. Inter annotator agreement (IAA) was calculated to evaluate the annotation quality. Cohen's kappa coefficient between annotators was 0.817, which indicates an almost perfect agreement according to Kappa's interpretations [33]. We only considered texts that both annotators labeled with the same label (9250).

We excluded the noise usually found in a social media text. The preprocessing operation included removing Arabic diacritics, Arabic and English punctuations, stop words, emojis, and non-Arabic characters. Also, The Arabic letters were normalized, where we replaced letters with different shapes with one of its defined shapes. "Alef" (آ, إ, أ) into ا, and "Ta Marbouta" (ة) into (o). In addition, we removed tweets-related noise like users' mentions, usernames, hashtags, and Twitter handles (@user), RT, <>. The last stage in the preprocessing is stemming, which removes any suffixes, prefixes, or infixes from words. This step reduces the derived or inflected words

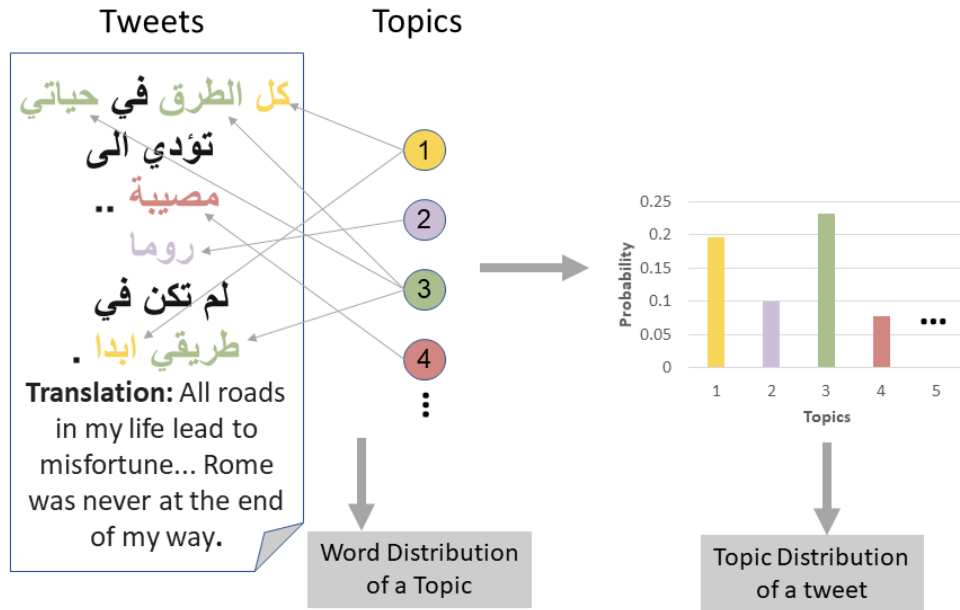


Fig. 2. A Graphical Representation for Generating Topic Distribution Vectors.



Fig. 3. Bar Charts for Top c-TF-IDF Scores in Six Topics.

TABLE III. A PERFORMANCE COMPARISON OF MULTIPLE CLASSIFIERS, USING BOTH WORD2VEC AND THE ENRICHED FEATURES (CTE)

Model	Word2vec				Enriched Representation (CTE)			
	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall
DT	0.6329	0.6329	1	0.6333	0.9667	0.9667	0.9667	0.9667
SVM	0.9273	0.9274	1	0.9268	0.9792	0.9792	0.9792	0.9792
RF	0.8372	0.8632	1	0.8644	0.9004	0.9004	0.9004	0.9004
KNN	0.7701	0.7714	1	0.7911	0.8640	0.8641	1	0.8641
XGBoost	0.8789	0.8782	1	0.8797	0.9823	0.9823	0.9823	0.9823
Bagging	0.7760	0.7798	1	0.7771	0.9710	0.9709	0.9709	0.9709
Stacking	0.9280	0.9297	1	0.9270	0.9710	0.9813	0.9710	0.9710

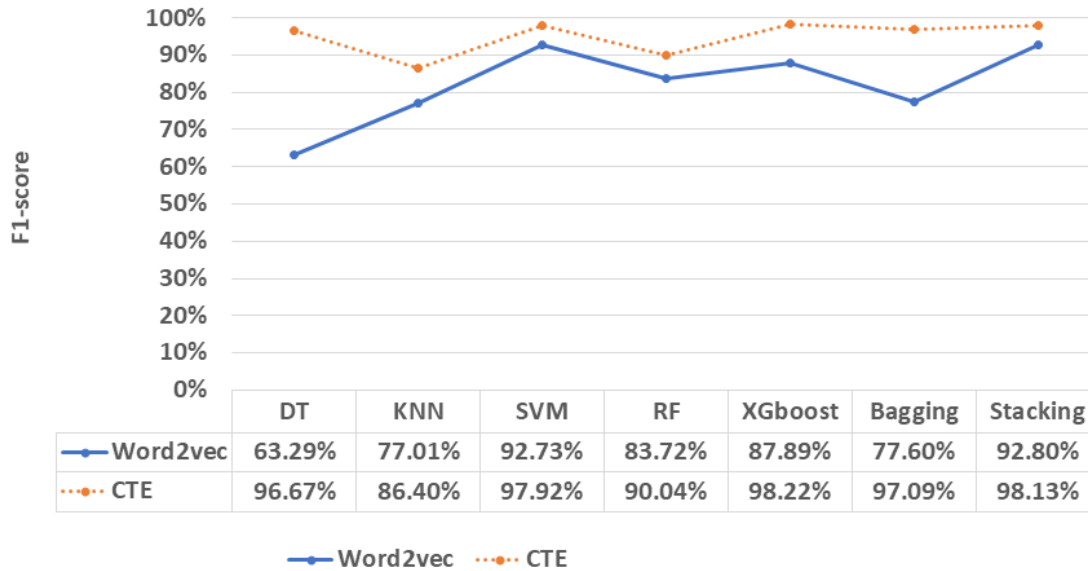


Fig. 4. F1-score Comparison between Classifiers that used Word2vec only and Classifiers that Utilized the Enriched Features (CTE).

into their related stems and relates all the variations into one stem. We used Tashaphyne², an Arabic light stemmer tool.

All experiments were applied by splitting data randomly into 75% (6,940 tweets) for training and 25% (2,310 tweets) for testing. Ten runs were performed for each experiment, and we reported the average results to overcome the performance variation problem. We used the NVIDIA Tesla K80 GPU provided on the Google colaboratory notebook.

The word2vec representation and BERTopic [28] algorithm are employed in this approach. We applied the BERTopic algorithm on the preprocessed CD dataset using Python’s ”bertopic” library. Arabert model [34] was used as a pre-trained language model. Each term in the corpus contributes differently to each topic. Fig. 3 shows the bar charts of the first six topics with a subset of the most contributing terms according to the c-TF-IDF scores.

We reduced the number of topics by starting from the least frequent topic and merging topics as long as the similarity exceeds 0.915, the threshold suggested in the BERTopic documentation. The number of topics was reduced from 70 topics to 47. Afterward, the resulting topic distributions for tweets in the dataset were concatenated to the word2vec representation.

The resulting enriched features are fed into six classifiers.

Table III demonstrates the results of four performance metrics; macro-accuracy, macro-precision, macro-recall, and macro-F1. Also, Fig. 4 shows F1-scores of classification using enriched features approach compared with classification using word2vec features. According to Fig. 4, the performance of the proposed approach improved all the classification results. DT and bagging results boosted the most. These encouraging results suggest that using latent topic distribution, obtained from BERTopic, can improve the classifier’s ability to distinguish between different CD categories. The baseline classifiers could not fully capture the relations of a CD category and a subset of topics due to the shortness of tweets. On the contrary, the enriched features emphasized these relations.

VI. CONCLUSION

The constant evolution in natural language processing and machine learning has made it possible to address different cognitive distortions in a text. This paper suggests an approach for enriching short textual representations to improve cognitive distortions’ classification of the Arabic context over Twitter. The enrichment approach considers the shortness of social media text and the sparsity of word co-occurrence patterns, which (to the best of our knowledge) is not explored in the literature. It also utilizes a transformer-based topic modeling algorithm (BERTopic) that employs a pre-trained language model (AraBERT). For evaluation, various

²<https://pypi.org/project/Tashaphyne/>

classifiers were used; DT, k-NN, SVM, RF, XGBoost, stacking and bagging. Experimental results showed that the proposed approach improved the performance of the defined classifiers with different rates using an Arabic CD dataset. We will study other TM approaches with advanced feature representations and embeddings for future work. In addition, CD data from other sources and languages can be tested to examine the generality of the proposed approach. As a recommendation for other researchers, we encourage examining transformer-based topic modeling to empower different supervised tasks in NLP. Also, it is good practice to define an approximation for topics' granularity in different domains.

REFERENCES

- [1] D. D'Hotman and E. Loh, "Ai enabled suicide prediction tools: a qualitative narrative review," *BMJ Health & Care Informatics*, vol. 27, no. 3, 2020.
- [2] A. Wongkoblap, M. A. Vadillo, and V. Curcin, "Social media big data analysis for mental health research," in *Mental Health in a Digital World*. Elsevier, 2022, pp. 109–143.
- [3] D. J. Dozois, "Depressive cognition on twitter," *Nature Human Behaviour*, vol. 5, no. 4, pp. 414–415, 2021.
- [4] B. Shickel, S. Siegel, M. Heesacker, S. Benton, and P. Rashidi, "Automatic detection and classification of cognitive distortions in mental health text," in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2020, pp. 275–280.
- [5] K. C. Bathina, M. Ten Thij, L. Lorenzo-Luaces, L. A. Rutter, and J. Bollen, "Individuals with depression express more distorted thinking on social media," *Nature Human Behaviour*, vol. 5, no. 4, pp. 458–466, 2021.
- [6] A. Barak and J. M. Grohol, "Current and future trends in internet-supported mental health interventions," *Journal of Technology in Human Services*, vol. 29, no. 3, pp. 155–196, 2011.
- [7] S. Shreevastava and P. Foltz, "Detecting cognitive distortions from patient-therapist interactions," in *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, 2021, pp. 151–158.
- [8] T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez, and C. Giraud-Carrier, "Detecting cognitive distortions through machine learning text analytics," in *2017 IEEE international conference on healthcare informatics (ICHI)*. IEEE, 2017, pp. 508–512.
- [9] H. M. K. Kumar and B. S. Harish, "A new feature selection method for sentiment analysis in short text," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1122–1134, 2020. [Online]. Available: <https://doi.org/10.1515/jisys-2018-0171>
- [10] A. Bagheri, A. Sammani, P. G. van der Heijden, F. W. Asselbergs, and D. L. Oberski, "Etm: Enrichment by topic modeling for automated clinical sentence classification to detect patients' disease history," *Journal of Intelligent Information Systems*, vol. 55, no. 2, pp. 329–349, 2020.
- [11] D. Bollegala, V. Atanasov, T. Maehara, and K.-i. Kawarabayashi, "Classinet—predicting missing features for short-text classification," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 5, pp. 1–29, 2018.
- [12] S. Seifollahi, M. Piccardi, and A. Jolfaei, "An embedding-based topic model for document classification," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 3, pp. 1–13, 2021.
- [13] R. R. Morris and R. Picard, "Crowdsourcing collective emotional intelligence," *arXiv preprint arXiv:1204.3481*, 2012.
- [14] X. Zhao¹², C. Miao¹², and Z. Xing, "Identifying cognitive distortion by convolutional neural network based text classification," *International Journal of Information Technology*, vol. 23, no. 1, 2017.
- [15] A. Sun, "Short text classification using very few words," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 1145–1146.
- [16] X. Hu, N. Sun, C. Zhang, and T.-S. Chua, "Exploiting internal and external semantics for the clustering of short texts using world knowledge," in *Proceedings of the 18th ACM conference on Information and knowledge management*, 2009, pp. 919–928.
- [17] S. Zelikovitz, W. W. Cohen, and H. Hirsh, "Extending whirl with background knowledge for improved text classification," *Information Retrieval*, vol. 10, no. 1, pp. 35–67, 2007.
- [18] M. Chen, X. Jin, and D. Shen, "Short text classification improved by learning multi-granularity topics," in *Twenty-second international joint conference on artificial intelligence*, 2011.
- [19] A. Anantharaman, A. Jadya, C. T. S. Siri, B. N. Adikar, and B. Mohan, "Performance evaluation of topic modeling algorithms for text classification," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 704–708.
- [20] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Frontiers in Artificial Intelligence*, vol. 3, p. 42, 2020.
- [21] X.-H. Phan, L.-M. Nguyen, and S. Horiguchi, "Learning to classify short and sparse text & web with hidden topics from large-scale data collections," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 91–100.
- [22] Z. Dai, A. Sun, and X.-Y. Liu, "Crest: Cluster-based representation enrichment for short text classification," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2013, pp. 256–267.
- [23] Z. Yajian, D. Dingpeng, and C. Junhui, "A short text classification algorithm based on semantic extension," *Chinese Journal of Electronics*, vol. 30, no. 1, pp. 153–159, 2021.
- [24] A. Abuzayed and H. Al-Khalifa, "Bert for arabic topic modeling: An experimental study on bertopic technique," *Procedia Computer Science*, vol. 189, pp. 191–194, 2021, aI in Computational Linguistics. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921012199>
- [25] L. Rojas-Barahona, B.-H. Tseng, Y. Dai, C. Mansfield, O. Ramadan, S. Ultes, M. Crawford, and M. Gasic, "Deep learning for language understanding of mental health concepts derived from cognitive behavioural therapy," *arXiv preprint arXiv:1809.00640*, 2018.
- [26] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of tf*idf, lsi and multi-words for text classification," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2758–2765, 2011.
- [27] V. K. Gunjan, J. M. Zurada, B. Raman, and G. Gangadharan, *Modern Approaches in Machine Learning and Cognitive Science: A Walk-through*. Springer, 2020.
- [28] M. Grootendorst, "Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics." 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.4381785>
- [29] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.
- [30] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.
- [31] L. Yang, C. Li, Q. Ding, and L. Li, "Combining lexical and semantic features for short text classification," *Procedia Computer Science*, vol. 22, pp. 78–86, 2013.
- [32] R. Covin, D. J. Dozois, A. Ogniewicz, and P. M. Seeds, "Measuring cognitive errors: Initial development of the cognitive distortions scale (cds)," *International Journal of Cognitive Therapy*, vol. 4, no. 3, pp. 297–322, 2011.
- [33] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.
- [34] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.