THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations

OPEN ACCESS

Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations

Bushra Algotiml
Department of Computer Science
Umm Al-Qura University
Mecca, Saudi Arabia
bagotiml@uqu.edu.sa

AbdelRahim A. Elmadany
College of Computer and Information
Jazan University
Jazan, Saudi Arabia
aelmadany@jazanu.edu.sa

Walid Magdy
School of Informatics
University of Edinburgh
Edinburgh, UK
wmagdy@inf.ed.ac.uk

Abstract

Speech acts are the actions that a speaker intends when performing an utterance within conversations. In this paper, we proposed speech act classification for asynchronous conversations on Twitter using multiple machine learning methods including SVM and deep neural networks. We applied the proposed methods on the ArSAS tweets dataset. The obtained results show that superiority of deep learning methods compared to SVMs, where Bi-LSTM managed to achieve an accuracy of 87.5% and a macro-averaged F1 score 61.5%. We believe that our results are the first to be reported on the task of speech-act recognition for asynchronous conversations on Arabic Twitter.

1 Introduction

Speech act in linguistics is defined as the action that a speaker intends when performing an utterance such as asking question, recommending something, greeting or thanking, expressing a thought or making suggestion. Knowing speakers intention within a conversation is considered the one of the recent active research in Natural Language Understanding (NLU); which is called speech act recognition/classification. Speech act classification has been utilized in different Natural Language Processing (NLP) applications, such as summarization (Zhang et al., 2013; Bhatia et al., 2014), rumors verification (Vosoughi, 2015; Vosoughi and Roy, 2016a), hate speech or cyberbullying detection (Gambäck and Sikdar, 2017; Saravanaraj et al., 2016), and in the educational forum (Bayat et al., 2016).

Speech act classification task is usually treated as a multi-class classification problem. Most of researchers tend to use machine learning (ML) paradigm for the task in order to analyze and utilize the massive amount of data that found in online conversations. They usually apply their experiments to two type of conversations: 1) synchronous conversations, where the conversation is one-to-one, such as dialogues, chatting, meetings and transcribed phone conversations; and 2) asynchronous conversations, where the conversation is one-to-many such as emails, discussion forums and social media. Existing work on speech act classification mostly focuses on English language, with some focus on other languages such as German, French and Korean. Moreover, these studies have been conducted for both synchronous and asynchronous conversations. Limited studies have tackled this task for Arabic, and all focusing on synchronous conversations. To the best of our knowledge, there is no work so far for Arabic speech act classification for asynchronous conversations, such as that on social media. Twitter has become a communication medium containing a massive amount of data suitable for social and behavioural studies. Communication between users in Twitter can be considered as asynchronous conversations, within which people post questions, express feelings, recommend, request, report, or claim; all of which can be considered speech acts. Classifying the speech act of tweets can aid in understanding the intentions behind users posts, analyzing Twitter content, and understanding how users interact on social media (Vosoughi and Roy, 2016b).

Recently, an Arabic speech-act and sentiment corpus (ArSAS) of tweets corpus was released (El-madany et al., 2018). It contains more than 21K tweets, annotated with six speech acts. In this paper, this corpus is used to evaluate the effectiveness of different supervised ML approaches for speech act classification for Arabic tweets. In our work, we proposed two approaches based on SVM and multiple deep learning models to classify Arabic tweets into speech act labels. Our results show that Bi-LSTM models achieves the highest performance overall and over each of the individual speech-act classes, where it achieves an accuracy of 87.5% and a macro-F1 of 0.615.

2 Problem Definition

Linguistically, speech act theory studies the ways in which the words can be used to carry out actions rather than transmitting information. The speech act can be defined as the actions that may be performed by speakers to carry out their intentions when performing utterances. According to Searle (Searle, 1975), which is based on Austins work (Austin, 1975), speech acts can be classified by their intent of usage and he categorized it into five categories: Assertive, Directive, Commissive, Expressive, and Declaration.

The concept of a speech act in Arabic can be defined in the same manner as in English. However, a speech act is more profound in the Arabic rhetoric discipline, which is concerned with the semantics of stylistic. Rhetoric in Arabic lies between syntax and semantics and aims to enable the Arabic speaker to relay his or her intended communicative meaning to the listener through the application of rhetorical means and eloquent criteria. The speech act as part of Arabic rhetoric is concerned with the ways of delivering some relevant information with utterance in order to attract the listeners attention (Abdul-Raof, 2006). Interestingly, researchers (Al-Hindawi et al., 2014) have shown the existence of speech acts in Arabic prior to the work of Austin and Searle on English speech acts. They refer to the speech act that is mentioned in the holy Quran and pointed out that Arab scholars have seeded the Arabic Speech act theory even before Austin theory. All the aforementioned examples are written using classical Arabic; however, our work targets the speech act in Arabic dialects. For that, we use a domain-specific taxonomy of six speech act categories

that are commonly seen on Twitter (Vosoughi and Roy, 2016b), including assertion, recommendation, expression, question, request, and miscellaneous. These categories are all derived from Searles taxonomy but modified to make it suitable for speech acts in Twitter.

3 Literature Review

Extensive research has been conducted on speech acts classification in different languages using various techniques. The vast majority of literature in this field uses either supervised ML techniques with a variety of lexical, syntactic, contextual, and semantics features, or adopts deep learning techniques to automatically identify the speech act.

3.1 Trends in Speech Act Recognition

(Cohen et al., 2004) classified English email according to the intent of the sender using SVM classifiers with the use of bi-gram and Part-of-Speech (POS) tags as features. (OShea et al., 2010) proposed a novel technique based on function words (e.g. articles, prepositions, determiners etc.) that appear in the utterances using decision tree classifiers. Their results strongly demonstrate the ability of function words to discriminate between speech act classes. Bhatia et al. (Bhatia et al., 2014) used speech act classification to aid in the summarization of online forum discussion threads using Nave Bayes and they demonstrated the effectiveness of speech act labels in summarizing discussion threads. Additionally, (Tavafi et al., 2013) aimed to determine a domain-independent classifier that could achieve good results across all types of conversations (synchronous and asynchronous). They used different ML techniques such as SVM, Conditional Random Fields (CRF) and Hidden Markov Model (HMM) with a set of domain-independent features, including lexical features and the length of utterances as another feature. They demonstrated that HMM classifiers achieved the best performance for speech act classification on different synchronous and asynchronous English conversations.

Similar techniques was applied for speech-act recgonition for other languages. For example, (Kim et al., 2011) used a Decision Tree (DT), SVM and the Maximum Entropy Model (MEM) with a set of lexical, grammatical and contextual features for speech act classification tasks using a Korean dialogue corpus in a schedule management

domain. They showed that SVM outperformed other classifiers by achieving accuracy equal to 93%. Similarly, (Ko, 2015) used Korean corpus to conduct an experiment using SVM and k-NN classifiers with POS and lexical words as features. Ko also pointed out the inadequacy of the tf.idf feature weighting scheme for the task due to the short length of utterances in comparison to the documents length. He proposed a new weighting scheme and showed the effectively of SVM with the new weighting scheme in speech act classification using Korean corpus. Additionally, Bayat et al. (Bayat et al., 2016) used SVM and a set of lexical, contextual, and syntactic features to classify German messages posted on an online forum. They showed the effect of adding features to SVM classifiers in order to increase the performance for the task.

Recently, many researchers have recently applied deep learning techniques using word embeddings to capture speech act of the utterance in synchronous conversations. This is because deep learning techniques afford a significant advantage in capturing the semantics of lexical features (Yoo et al., 2017). The effectiveness of deep learning techniques for speech act recognition is evident in the work of (Cerisara et al., 2018), who proposed a novel deep neural network model based on Long Short Term Memory (LSTM) Recurrent Neural Networks (RNN) and validated this model in three different languages using synchronous conversation corpora. They demonstrated that the performance of this approach is consistent across these three languages (with an accuracy of 72%, 98% and 92% for English, Czech and French, respectively). Similarly, (Khanpour et al., 2016) used LSTM to classify dialogue acts in open-domain conversations. They reported that the word embedding parameters, dropout regularization, decay rate and number of layers have a significant impact on the final systems accuracy. (Yoo et al., 2017) applied a CNN to capture speech acts on Korean dialogues corpus. Their model has obtained a high accuracy (89%) in the speech act recognition task. (Kim and Kim, 2018) proposed an integrated neural network model based on CNN for identifying speech acts, predictors, and sentiments of dialogue utterances. They concluded that the integrated model can help in increasing the performance of intention identification. (Lee and Dernoncourt, 2016) applied both RNN and CNN on three different synchronous conversations datasets. They stated that the CNN model outperformed the LSTM model for all datasets by a very small margin.

Recently, more attention was directed to speech act classification for tweets as a kind of asynchronous conversations. (Zhang et al., 2011) proposed a set of word-based and character-based features to recognise the speech acts of tweets in order to analyze tweeters behavior collectively or individually. They suggested word-based features composed of a set of N-grams, abbreviations, acronyms, vulgar and opinion words, emoticon, and Twitter-specific features. SVM was used for this task, and showed a weighted-average F1 value of nearly 0.70. Another work by (Vosoughi and Roy, 2016b) explored speech act recognition on Twitter by training SVM, Naive Bayes, decision tree and logistic regression classifiers with a set of semantic and syntactic features. Their approach in features engineering is similar to (Zhang et al., 2011), though they added the dependency subtrees and POS tags to their syntactic features set. They achieved a performance with a weighted-average F1 score equal to 0.70, which is similar to the results of (Zhang et al., 2011). Furthermore, they also applied their speech act classifier to detect rumors on Twitter based on assertion speech act detection in tweets (Vosoughi, 2015). Another work by (Joty and Hoque, 2016) applied deep learning techniques by proposing a model that used LSTM and RNNs for speech acts modeling showed the effectiveness on asynchronous conversations, such as emails and forums.

3.2 Arabic Speech Act Recognition

Much less attention was directed to speech act classification for Arabic. Only limited amount of work exist in literature, but only for classifying speech act in Arabic synchronous conversations. One of the earliest works in this area is (Graja et al., 2013) who used CRF to perform a semantic labelling task for spontaneous speech in Tunisian dialects by using the TuDiCol corpus. Another work, (Elmadany, 2016; Elmadany et al., 2016) utilized the JANA corpus to solve the issue of automatic dialogue act classification for Egyptian Arabic dialect using SVM. Also, (Sherkawi et al., 2018) applied different machine learning techniques on a small Arabic corpus. Their corpus is relatively small and written in MSA.

Recently, a new corpus of Arabic tweets (Ar-SAS) annotated speech-act was released (El-madany et al., 2018), which is the first to model speech-act in asynchronous conversations for Arabic. Nonetheless, no work has been published on this data yet. In this paper, we utilise the Ar-SAS dataset for speech-act classification of Arabic tweets. We believe we are the first exploring speech act classification for Arabic asynchronous conversations.

4 Methodology

We propose two different approaches for classifying Arabic tweets into pre-defined speech act categories: 1) SVM with a set of syntactic and semantic features, and 2) Deep learning with word embedding using different neural network architectures.

4.1 Support Vector Machine Model

SVM has demonstrated significant performance in most of the related work for speech act classification (Cohen et al., 2004; Elmadany, 2016; Hemphill and Roback, 2014; Kim et al., 2011; Ravi and Kim, 2007; Zhang et al., 2013; Tavafi et al., 2013; Zhang et al., 2011; Vosoughi, 2015; Ko, 2015). In preliminary experimentation using simple bag-of-words as features, we found SVM to be superior to other basic ML methods including Naive Bayes, k-NN, RF and DT. Thus, it was chosen for further experiments with additional set of features.

4.1.1 Features Selection

Based on literature, we extracted multiple set of features from tweets to model speech act. We grouped the features into three groups: lexical, syntactic, and structural features.

Lexical features: This is simply the words n-grams in the tweet text. We used uni-gram, bi-gram and tri-gram phrases. These features were extracted after applying preprocessing to the text, including character normalisation to the different forms of the Arabic letters {ي ,ء ,ة ,أ} and diacritics removal (Darwish et al., 2014).

Syntactic features: These features represent the syntactic style of the text of the tweet, and it include four sets of features: punctuation marks, Twitter special characters, Emojis, and Links. Binary features representing specific punctuation such as question and exclamation marks are set to one when appear, since they can be indicative to specific speech act classes such as expressions, requests or question. In addition, the rest of punctuation marks are combined and added as additional binary features to indicate the appearance of any other punctuation marks in the tweet. Twitter special characters such as presence of hashtags (#) and mentions (@) were also used as binary features. Hashtags might be an indicative feature of some speech act classes such as assertion, where user might use hashtags to announce something. In addition, emojis are usually used in tweets to communicate specific feelings thus it was added as an additional feature. Finally a binary feature indicating the presence of links in the tweets was also used.

Structural features: These set of features represent the structural form of the sentence, and it includes: (1) The length of the tweet in characters and words, which was shown previously to be a useful feature for speech act recognition (Zhang et al., 2013; Tavafi et al., 2013; Elmadany, 2016). (2) POS tags of the words in the tweets, which has been shown to be an effective features in speech act recognition for English (Zhang et al., 2013; Vosoughi and Roy, 2016b). We used FARASA POS tagger [1] for extracting the POS tags. We modified the tagger to include hashtags, URLs and emojis as tags within the sequence. Then, we added uni-gram, bi-gram and tri-gram PoS sequences to the features vector for each tweet.

4.2 Deep Learning Approach

In this paper, we implemented different variants of deep learning neural network in order to determine the most effective type of neural network for the task of Arabic speech act recognition. Two variants of deep learning approaches were considered in this task using different architectures: RNN in particular LSTM and BiLSTM; and CNN. In addition, several combinations of neural networks variants have been applied for this task such as CNN on top of LSTM, CNN on top of BiLSTM, LSTM on top of CNN, and BiLSTM on top of CNN.

A skip-gram word2vec embedding have been used to represent the words in each tweet before inputting them to any neural network. An Arabic pretrained word embedding has been utilized for this work called "AraVec" (Soliman et al., 2017). AraVec is a skip-gram model trained on 70M Ara-

---

[1] http://qatsdemo.cloudapp.net/farasa/

Speech Acts	Examples
Assertion	‏#الشروق: السسي: كلي فخر واعتزاز بالنخبة المتميزة المشاركة في شباب العالم
	#Sunrise: El-Sisi: I am proud of all the elite who are contributed in the world cup forum
Expression	‏أشعر أن الربع العربي إشعاع من الحرة
	I feel that Arab revolutions are radiation of freedom
Recommendation	‏الكرة الاطالة تحتاج لتركي ال الشخ
	Italian football needs Tukey Al-Shaikh
Request	‏بعد منتدي شباب العالم أطالب بعمل منتدي للفساد لإظهار الحقائق
	After the world cup forum, I request to do a forum to reveal truths
Miscellaneous	‏ولد سلمان ممكن يلعب مكان محمد صلاح وممكن يلعب مكان عبد الشافي
	Walid Suliman can play instead of Muhammad Salah or possibly instead of Abdul-Shafy

Table 1: Example tweets of different speech acts categories in the ArSAS dataset

bic tweets containing 204K unique words and 300 dimensions.

5 Experimental Setup

5.1 Dataset

We utilized the recently published tweets corpus "Arabic Speech Act and Sentiment" (ArSAS) for our experimentation (Elmadany et al., 2018). ArSAS contains a large set of 21,081 Arabic tweets in different Arabic dialects and annotated by six speech act classes: Assertion, Recommendation, Expression, Question, Request, and Miscellaneous. The tweets in the corpus covers 20 topics including long-standing topics, events and entities (celebrities or organization). Table 1 shows few examples of tweets in the corpus with their corresponding speech act label.

The size of samples in each speech acts class varies a lot in ArSAS corpus, ranging between 60 samples to 11.7K samples per class. The smallest two classes are miscellaneous and recommendation classes that have only 60 and 109 tweets respectively. Therefore, we decided to merge these two classes into one and called it miscellaneous. The final distribution of the five classes in the corpus is: expression (11734), assertion (8233), question (752), request (183), and miscellaneous (169).

5.2 Classifiers Implementation and Setup

For the SVM classifier, SVM LinearSVC implemented in the SKlearn[2] was utilised for our experiments. We examined "One-vs-All" and "One-vs-One" strategies for SVM, and noticed better performance for the "One-vs-All" implementation, and thus it was conducted for our experiments.

Hyper-parameters	Choice
Output layer activation function	softmax
Cost Function	Cross-entropy
Optimizer	ADAM
Learning Rate	0.0001
Batch size	50
Epoch size	30
Dropout rate	0.5
LSTM Units	100
CNN filters	2 and 3
CNN features map	32
Pool size	2
LSTM hidden layers	2
B-LSTM hidden layers	2
CNN hidden layers	4

Table 2: Neural network hyper-parameters

For the neural network classifiers, we used Keras 2.1.3[3] implementations of the multiple models we examined. For the training process of our deep learning models, Table 2 describes the hyper-parameters we used after multiple iterations for reaching the optimal performance.

5.3 Evaluation

For measuring the performance of our approaches, we split the data into five folds and applied 5-fold cross validation for training and testing. Data are split into folds over the class level, where we insure that 20% of the samples of each class exists in each fold. This was essential step to ensure the presence of samples from the small classes in each fold.

For evaluation, three scores are applied: accuracy, micro F-score, and macro F-score. Accuracy and micro-F1 should demonstrate the overall performance of the approaches, while macro-F1 would indicate the average performance of the approaches over each class individually.

---

[2]https://scikit-learn.org/

[3]https://github.com/keras-team/keras

	Macro-F1	Micro-F1	Accuracy
Lexical	0.510	0.840	0.840
Lexical+Syntactic	0.520	0.850	0.850
Lexical+Structural	0.520	0.850	0.860
All features	**0.532**	**0.862**	**0.865**

Table 3: The performance of SVM for Arabic speech act classification using different sets of features

	Macro-F1	Micro-F1	Accuracy
CNN	0.540	0.841	0.850
LSTM	0.570	0.850	0.860
BiLSTM	**0.615**	**0.86**	**0.875**
CNN on top of LSTM	0.535	0.850	0.865
CNN on top of BiLSTM	0.558	0.850	0.860
LSTM on top of CNN	0.585	0.860	0.870
BiLSTM on top of CNN	0.600	0.860	0.870

Table 4: Comparison between different deep learning architectures for Arabic speech act classification in term of macro-averaged F1, micro-F1 and accuracy

6 Results

6.1 SVM Results

Table 3 reports the results obtained when applying SVM classifier for our task using different sets of features on the ArSAS dataset with 5-fold cross-validation. As shown, the performance of different set of features is almost similar, and the performance when applying all the set of features achieves the best results of accuracy 86.5%, micro-F1 of 0.862, and macro-F1 of 0.532. While the overall performance is relatively high, the performance for some of the classes is considerably low. This could be explained as a reason to the high imbalance of our classes, where some of the classes (such as 'miscellaneous', 'requests', and 'questions') are tiny compared to the two major classes 'expression' and 'assertion'. Actually, nearly 90% of the samples in the classes miscellaneous and request were incorrectly classified.

These results are comparable to the state-of-the-art in other languages such as English. Comparing our work to the work by (Vosoughi and Roy, 2016b) and (Zhang et al., 2011) for speech act classification for English tweets, they report in term of micro F1 (0.69 and 0.70) respectively, and they also explain this due to the high imbalance of classes. Our achieved micro-F1 is even higher 0.86. This might indicate the suitability of using the same techniques —used for English speech act classification— for the Arabic task.

6.2 Deep Learning Results

Table 4 reports the results obtained when applying seven different architectures of RNN and CNN for Arabic speech act classification on the ArSAS dataset with 5-fold cross-validation. As shown, the performance of most of the models is close to those obtained by the SVM models in terms of accuracy, but consistently higher when measured using macro-F1. The BiLSTM and the BiLSTM on top of CNN architectures achieved significantly higher results in terms of macro-F1 compared to all the other models, which indicates better performance on the class level. The best performing model was the BiLSTM model with an accuracy of 87.5%, micro-F1 of 0.86, and macro-F1 of 0.615. This confirms the effectiveness of using the bidirectional LSTM to capture the context in the tweet, which the miscellaneous class actually need. Additionally, BiLSTM succeeded in recognizing both question and request classes better than any other model.

Table 5 shows the performance on the best performing model using BiLSTM on each of the classes individually. As shown, the performance over the two large classes 'assertion' and 'expression' is high (0.9 and 0.87 F1 respectively) compared to the other classes. The 'request' class achieved the lowest performance (0.2 F1). This shows the challenge of recognising the speech act in asynchronous conversations for some of the in-

Class	F1 Score
Assertion	0.90
Expression	0.87
Miscellaneous	0.53
Question	0.57
Request	0.20

Table 5: The performance of best performing BiLSTM model on each class

	Macro-F1	Micro-F1	Accuracy
SVM	0.532	**0.862**	0.865
Bi-LSTM	**0.615**	0.860	**0.875**

Table 6: The best performing SVM model and neural network architecture on the ArSAS dataset

frequent classes.

While our performance is comparable to performance in other languages, we believe there is still large room for improving the performance. We hope that our work would be considered as a baseline for future work on speech act classification for Arabic.

6.3 Discussion

We explored speech act classification in Twitter using SVM classifier with sets of lexical, syntactic, and structural features, and using several neural network architectures with pretrained word embedding for word representation. The best SVM model with all the extracted features has achieved 53.2% in the term of macro-averaged F1 using 5-folds cross validation on the ArSAS dataset. After applying deep learning for the task using variants of neural network architectures, our experiments showed that all the suggested architectures have outperformed the best SVM model with the sets of features. This highlights the superiority of deep learning models especially for a highly inflected language such as Arabic, and in particular Arabic dialect as present in our tweets dataset. BiLSTM has achieved 61.5% in the term of macro-averaged F1, which is 8.3% higher than the best SVM model as shown in Table 6. This confirms the fact that deep learning usually performing better even without any feature engineering.

Potentially, the performance of BiLSTM might improve if these extracted features get fed as an input to the network, especially the PoS tags. PoS features might give some structural characteristics to the neural network.

Moreover, the overall performance was highly affected by the imbalance distribution for the classes amongst the corpus. During our experimentation, we examined some solutions for this issue, such as data over sampling, but it did not lead to improved performance. We believe it might be useful if there are more new examples for the small classes, which would potentially enhance the performance of neural networks as well.

Furthermore, ArSAS corpus contains labels for the type of the topic for each tweet such as long-standing, entity, and event. We suggest considering this attribute in implementation by training different classifier for each type of topics separately. It might be better for the classifier to classify tweets in the same type than classifying tweets from mixed types.

7 Conclusion

In this paper, we have presented two ML approaches for speech act classification in Twitter platform using dialectical Arabic tweets. An SVM classifier with different sets of lexical, syntactic, and structural features was proposed. In addition, a set of different neural network architectures was examined for the task. For both approaches, we exploited the recent published Arabic corpus called "ArSAS" which has more than 21K tweets that annotated by six different speech acts, that we decided to squash to only five classes after merging the smallest two classes into one. Our results showed that deep learning is a more effective approach for speech act classification of Arabic tweets compared to SVM. In particular, the BiLSTM implementation achieved the highest performance especially with the macro-F1 score that was 61.5% compared to only 53.2% for SVM. The best achieved overall accuracy for BiLSTM and SVM were comparable with values of 87.5% and 86.5% respectively.

As a general observation, all the used techniques have showed an acceptable performance, especially when compared to the current state-of-the-art for English speech act classification for asynchronous conversations. Nevertheless, as we discussed, there is the potential of multiple directions for improvements that could be explored in future work.

References

Hussein Abdul-Raof. 2006. *Arabic rhetoric: A pragmatic analysis*. Routledge.

Fareed Hameed Al-Hindawi, Hameed Hasoon Al-Masudi, and Ramia Fuad Mirza. 2014. The speech act theory in english and arabic. *Open Journal of Modern Linguistics*, 4(01):27.

John Langshaw Austin. 1975. *How to do things with words*. Oxford university press.

Berken Bayat, Christopher Krauss, Agathe Merceron, and Stefan Arbanowski. 2016. Supervised speech act classification of messages in german online discussions. In *The Twenty-Ninth International Flairs Conference*.

Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. 2014. Summarizing online forum discussions–can dialog acts of individual messages help? In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 2127–2131.

Christophe Cerisara, Pavel Kral, and Ladislav Lenc. 2018. On the effects of using word2vec representations in neural networks for dialogue act recognition. *Computer Speech & Language*, 47:175–193.

William W Cohen, Vitor R Carvalho, and Tom M Mitchell. 2004. Learning to classify email into speech acts. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 309–316.

Kareem Darwish, Walid Magdy, et al. 2014. Arabic information retrieval. *Foundations and Trends® in Information Retrieval*, 7(4):239–342.

A Elmadany. 2016. *Automatic Act Classification for Arabic Dialogue Context*. Ph.D. thesis, Thesis.

AbdelRahim Elmadany, Sherif M Abdou, and Mervat Gheith. 2016. Jana: A human-human dialogues corpus for egyptian dialect ldc2016t24.

AbdelRahim A Elmadany, Hamdy Mubarak, and Walid Magdy. 2018. Arsas: An arabic speech-act and sentiment corpus of tweets. In *OSACT 3: The 3rd Workshop on Open-Source Arabic Corpora and Processing Tools*, page 20.

Björn Gambäck and Utpal Kumar Sikdar. 2017. Using convolutional neural networks to classify hatespeech. In *Proceedings of the first workshop on abusive language online*, pages 85–90.

Marwa Graja, Maher Jaoua, and Lamia Hadrich Belguith. 2013. Discriminative framework for spoken tunisian dialect understanding. In *International Conference on Statistical Language and Speech Processing*, pages 102–110. Springer.

Libby Hemphill and Andrew J Roback. 2014. Tweet acts: how constituents lobby congress via twitter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1200–1210. ACM.

Shafiq Joty and Enamul Hoque. 2016. Speech act modeling of written asynchronous conversations with task-specific embeddings and conditional structured models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1746–1756.

Hamed Khanpour, Nishitha Guntakandla, and Rodney Nielsen. 2016. Dialogue act classification in domain-independent conversations using a deep recurrent neural network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2012–2021.

Hark-Soo Kim, Choong-Nyoung Seon, and Jung-Yun Seo. 2011. Review of korean speech act classification: machine learning methods. *Journal of Computing Science and Engineering*, 5(4):288–293.

Minkyoung Kim and Harksoo Kim. 2018. Integrated neural network model for identifying speech acts, predicators, and sentiments of dialogue utterances. *Pattern Recognition Letters*, 101:1–5.

Youngjoong Ko. 2015. New feature weighting approaches for speech-act classification. *Pattern Recognition Letters*, 51:107–111.

Ji Young Lee and Franck Dernoncourt. 2016. Sequential short-text classification with recurrent and convolutional neural networks. *arXiv preprint arXiv:1603.03827*.

James OShea, Zuhair Bandar, and Keeley Crockett. 2010. A machine learning approach to speech act classification using function words. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, pages 82–91. Springer.

Sujith Ravi and Jihie Kim. 2007. Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications*, 158:357.

A Saravanaraj, JI Sheeba, and S Pradeep Devaneyan. 2016. Automatic detection of cyberbullying from twitter. *International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN*, pages 2249–9555.

John R Searle. 1975. A taxonomy of illocutionary acts.

Lina Sherkawi, Nada Ghneim, and Oumayma Al Dakkak. 2018. Arabic speech act recognition techniques. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(3):18.

Abu Bakr Soliman, Kareem Eissa, and Samhaa R El-Beltagy. 2017. Aravec: A set of arabic word embedding models for use in arabic nlp. *Procedia Computer Science*, 117:256–265.

Maryam Tavafi, Yashar Mehdad, Shafiq Joty, Giuseppe Carenini, and Raymond Ng. 2013. Dialogue act recognition in synchronous and asynchronous conversations. In *Proceedings of the SIGDIAL 2013 Conference*, pages 117–121.

Soroush Vosoughi. 2015. *Automatic detection and verification of rumors on Twitter*. Ph.D. thesis, Massachusetts Institute of Technology.

Soroush Vosoughi and Deb Roy. 2016a. A semi-automatic method for efficient detection of stories on social media. In *Tenth International AAAI Conference on Web and Social Media*.

Soroush Vosoughi and Deb Roy. 2016b. Tweet acts: A speech act classifier for twitter. In *Tenth International AAAI Conference on Web and Social Media*.

Donghyun Yoo, Youngjoong Ko, and Jungyun Seo. 2017. Speech-act classification using a convolutional neural network based on pos tag and dependency-relation bigram embedding. *IEICE Transactions on Information and Systems*, 100(12):3081–3084.

Renxian Zhang, Dehong Gao, and Wenjie Li. 2011. What are tweeters doing: Recognizing speech acts in twitter. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Renxian Zhang, Wenjie Li, Dehong Gao, and You Ouyang. 2013. Automatic twitter topic summarization with speech acts. *IEEE transactions on audio, speech, and language processing*, 21(3):649–658.