CLASSIFICATION OF ENCOURAGEMENT AND WARNING ON CLASSICAL ARABIC USING DEEP LEARNING

LI ZHILONG, Saidah Binti Saad

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Malaysia.

p119470@siswa.ukm.edu.my, saidah@ukm.edu.my

ABSTRACT

The Qur'an, as a guidebook for all Muslims, plays an important role in the life of Islamic society. The Quran contains 6236 verses, with most verses conveying messages of warning and encouragement. As a guidebook, the Qur'an plays an important role in many fields, such as finance and medicine. So, it is necessary to classify the verses into encouraging and warning. Text classification techniques can be used to classify warning and encouraging verses in the Quran. Arabic is divided into three categories, which include Classical Arabic, Modern Standard Arabic (MSA) and Dialect Arabic. The language used in the Quran is classical Arabic, which is very rich in morphology. Text classification is divided into supervised learning method and unsupervised learning method. This research uses word embedding technology and uses different Machine Learning models to classify the Qur'an text into warning verses and encouraging verses. The result shows that the LSTM (85.84%) model performs better than Convolutional Neural Networks (80.54%) and Recurrent Neural Networks (73.48%). The Support Vector Machine (SVM) and Naive Bayes (NB) models had similar accuracy rates, 78.50% and 79.31%, respectively. In addition, the influence of epoch, batch size, and dropout parameters on the prediction effect of the model is also explored. Parameter experiments show that setting the parameter value too large or too small will reduce the accuracy of the model.

Keywords: **Text Classification, Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory, Classic Arabic.**

I. INTRODUCTION

Text classification is a basic but critical technology in the fields of natural language processing, text mining, information retrieval, and information security. Its main purpose is to solve the problem of messy information and provide more accurate information quickly. With the development of science and technology, various tools are becoming more and more intelligent, the amount of data is also increasing, and the text format is becoming more and more complex, which brings great challenges to text classification. In the early research, most of the methods used in text classification were traditional machine learning methods, which used traditional text representation methods and manually selected features, such as support vector machines, naive Bayesian, and logistic regression. These traditional machine learning methods can solve the text classification problem to a certain

extent, but there are also some shortcomings and deficiencies, such as the curse of dimensionality, low classification accuracy, and excessive manual intervention. The development of deep learning provides new ways to solve these problems. In recent years, deep learning has received extensive attention, and has made outstanding contributions in the fields of image processing and machine translation. With the development of word vectors, deep learning has also shown strong performance in text classification. The convolutional neural network and recurrent neural network memory attention mechanism of deep learning have achieved good results in text classification research, and their performance on various data sets is relatively good. Compared with the traditional machine learning method, the feature selection of the deep learning method does not require manual intervention, and the feature information in the text can be extracted from the input information, and the deep learning method can also retain the context information in the text. Therefore, with the continuous increase of information, text classification based on deep learning methods is particularly important.

Many researchers have been working on text categorization in English and other European languages, however few researchers work on text categorization for Arabic language. This is because Arabic is a language with rich forms and there are different methods in different places, so there is not much research on Arabic text classification. Despite the difficulties, considering the Qur'an as a guidebook for Muslims. There are many people who use the contents of the Qur'an as a guideline in their lives. Therefore, there is still a very important application value for the text classification of the Qur'an text.

Arabic is a language with many speakers, more than 420 million people in 27 countries and regions speak In Arabic. There are three Arabic languages, Modern Standard Arabic, Classical Arabic and Dialect Arabic. The Quran was writhed by classic Arabic. For readability, diacritics are used in Arabic. Each letter will have a different shape in different positions in the text. In addition, Arabic is written differently from our normal way. Arabic is written from right to left. Classical Arabic is very rich in morphology, and there are singular and plural nouns in Arabic. Verbs appear in different tenses in the text. Arabic has many challenges in sentiment classification.

The Qur'an serves as the code of conduct for many Muslims in their daily lives, and it gives corresponding guidance in many aspects. Such as legal, financial, and marriage. Therefore, the study of the Qur'an is very necessary. More importantly, the Qur'an contains information in different fields such as astronomy, geography, and biology. This is also a way for relevant researchers to acquire knowledge. Alike sentiment sentences in general text, In Quran, there is also some categorization which are encouragement verses and warning verses. Therefore, it is of great value for the study of the Qur'an. So, this research uses several machine learning methods to classify Quranic texts into

encouragement and warning. In addition, the effect of model parameters on model performance is tested.

II. LITERATURE REVIEW

Alshammeri, M.& Atwell, E (2021) used NLP combined with ML to classify the verses of the Quran into 15 predefined classes. The semantics of the verses were captured using Doc2vec embeddings that were used to group similar documents. Their model achieved a classification accuracy of 68% and 56% F1 score. The results confirmed that the classifier scored higher accuracy with the distributed bags of words architecture of the Doc2vec model. Next, they evaluated our model by examining the semantic similarity of the Quranic verses. Derived classes showed relatively high average similarity for some classes using the distributed bags of words architecture. The three top classes that achieved higher average same-category similarity and mean-difference are Faith, Pillars of Islam, and Religions.

Al-Kabi, Mohammed et al (2013) measured the effectiveness of four classification techniques to classify different Quranic verses according to their topics. They performed a preprocessing step on the Qur'an text, removing diacritics and Qur'anic symbols in the Qur'an. Their evaluation results showed that the NB classifier scored the highest accuracy results with lowest error rate whereas the J48 Decision Tree classifier scored the lowest accuracy results with highest error rate.

In another text classification study of Arabic, Hadi et al (2010) use a dataset consisting of 2,244 Arabic document. These documents are unequally partitioned into 5 categories. They aim to evaluate two classification algorithms Support Vector Machine and Naïve Bayes. It concludes that SVM perform better than NB.

Boukil & Samir, et al (2018) classify Arabic text based on Convolutional Neural Networks (CNN). They first use text pre-processing technology to remove some symbols and numbers in the text, remove stop words, and find the stem of the words. Then, they used Bag of Words method for text representation. After that, they applied Term Frequency-Inverse Document Frequency (TF-IDF) as a dimension reduction method to find the most important words. The dataset used in this research was 111,728 documents categorized into five classes. Finally, the results achieved an accuracy rate 92.94% based on trained CNN model.

Almost little research work is based on classical Arabic. Alhasani Hatem & Saidah Saad et al (2018) used sentiment analysis techniques to classify classical Arabic texts and achieved good results by using the selected top 5000 features to perform experiments on Naive Bayesian and Support

Vector Machines, corresponding to the accuracy rates are 95.6% and 91.5%, respectively. In addition, the data in many research work mainly comes from social media. And most are in Modern Standard Arabic or dialect Arabic, with very little work based on Classical Arabic. Alhasani Hatem & Saidah Saad et al (2018) used sentiment analysis techniques to classify classical Arabic texts and achieved good results by using the selected top 5000 features to perform experiments on Naive Bayesian and Support Vector Machines, corresponding to the accuracy rates are 95.6% and 91.5%, respectively.

III.   RESEARCH MODEL AND RESEARCH QUESTIONS

The objective of this study is to use several state-of-the-art based on Machine Learning model and deep learning model. The machine learning method can quickly and efficiently finish the classification task, and the text is divided into positive, negative, and neutral. So, machine learning methods can be seen in many fields. This work, different machine learning models were used to classify Qur'anic encouragement and warning verses.

A.   Convolution Neural Networks

CNN model can well detect the important features and related features contained in the text, which is beneficial to the text classification task. The word embedding generated by the model Skip-gram is sent to the CNN model layer, and the filter is used on the convolution layer to extract the features of the text and generate a feature matrix. Then there is the pooling layer, the pooling layer will continue to filter out the most important features in the feature matrix, and then the flattening layer will merge all the features together. The last layer is a fully connected layer, using Relu as the activation function, and finally using the SoftMax function to classify the text into different categories. In this study we select CNN model to predict the encouragement verse and warning verse of the Qur'an.

B.   Recurrent Neural Networks

Due to the special structure of the cyclic neural network, RNN is better at processing some order-related texts, or some tasks that change over time. Due to the special structural characteristics of the RNN model, the input of the hidden layer consists of two parts, one part is the output of the input layer, and the other part is the output of the hidden layer at the previous moment. Therefore, RNNs are best suited for processing text or speech data. In this study, we use the RNNs model as our classifier, and then compare and analyze the experimental results of different models.

C.   Long Short-Term Memory

LSTM is composed of memory units, input units, and forget units. It is a special RNN that can avoid the problem of long-term dependence. The function of input unit is to control the flow of the information from the cell. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. The function of the storage unit is to remember data values at different times. The function of the forget unit is to decide which useless information to discard. This step is mainly realized through the forget gate. In this study, the word vectors trained by Skip-gram are sent to the LSTM layer for training. The trained words are connected to the fully connected layer, and the fully connected layer uses the Relu activation function. Finally, the predicted text is classified using the SoftMax function.

In this study, firstly, the corpus required for the experiment is constructed by selecting encouraging verses and warning verses in the Qur'an. Encouraging statements are grouped into one category, while warning statements are grouped into another. The focus of this research is to classify Quranic texts as warning verses or encouraging verses at the sentence level using different deep learning methods respectively. The study aims to answer the following research questions:

i) Q1: A comparative study using several state-of-art deep learning-based methods, and determine which model classifies best?

ii) Q2: What are the best parameters of the model on the experimental results?
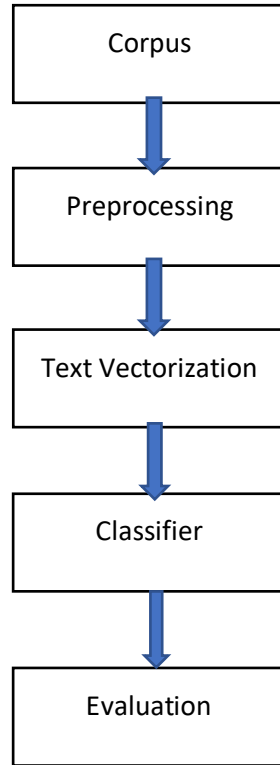
IV.   METHODS: DATA COLLECTION AND CLASSIFICATION PROCESS.

The data of this study come from Qur'an. The Qur'an contains many verses of encouragement and warning. These verses can be roughly divided into three types, The first type is that the verses only contain messages of encouragement or warning. The second is that a single verse contains both a message of encouragement and a message of warning. The third is that several verses together express a message of warning or encouragement. This study deals with the first type of verses. The dataset comes from the previous research work of (Alhasani Hatem & Saidah Saad et al 2018) who experimented on support vector machines and naive Bayesian classifiers by selecting different features and achieved good results.

The text classification experiment consists of five steps. First, obtain the data set from the Quran, and perform data pre-processing on the dataset. The pre-processing process include normalization, tokenization, lemmatization, stemming, and removal stop word. Then use the Word2Vec algorithm to extract text features from the cleaned data to obtain low-dimensional word

embedding. Then use three different neural network structures (CNN, RNN LSTM) as a text classification model, and finally use the built model to classify the Qur'an text into encouraging verses and warning verses. The flowchart is shown in Figure 1.

Figure 1 Research phases of classification of encouragement and warning

```
        ┌─────────────────┐
        │     Corpus      │
        └─────────────────┘
                 │
                 ▼
        ┌─────────────────┐
        │  Preprocessing  │
        └─────────────────┘
                 │
                 ▼
        ┌─────────────────┐
        │Text Vectorization│
        └─────────────────┘
                 │
                 ▼
        ┌─────────────────┐
        │    Classifier   │
        └─────────────────┘
                 │
                 ▼
        ┌─────────────────┐
        │    Evaluation   │
        └─────────────────┘
```

V.   RESULTS & DISCUSSION

This dataset is used the verses classification done by Hatem et.al (2018) which is on Arabic Quranic Text. The total number of verses contained in the Quranic corpus is 6,236 verses but the dataset used for this experiment consists of 1,114 verses of encouragement and 1,953 verses of warning. To improve the accuracy and to avoid predominance of the warning verses, we had to balance the sub-corpus. Therefore, 2,000 verses were randomly chosen, and duplicate verses were eliminated. The balance corpus consisted of 2,000 verses where 1,000 are encouragement verses and the other 1,000 are warning verses. For this experiment 80% of all text data is extracted as training data and 20% as test data. Training data is used to train the model, and the test data is used for performance evaluation. The table below show the main information.

Table 1 dataset

Data	Training	Test	Total

Encourage data	800	200	1000
Warning data	800	200	1000
Total	1600	400	2000

In this research, Machine Learning models and Deep Learning models are used to conduct experiments respectively. The parameters used by the different models are shown in Table 2.

Table 2 Default model parameters configuration

Model	Hyper-parameter	Value
CNN	Batch sizes	32
	Number of epochs	20
	Number of units in fully connected layer	768
	Learning rate	0.003
	Dropout rate	0.5
	Number of filters	256
	Filter size	3
RNN	Batch size	32
	Activation	Relu
	Number of epochs	20
	Dropout rate	0.5
LSTM	Number of epochs	20
	Number of units in fully connected layer	100
	Dropout rate	0.5
	Learning rate	3e-5
	LSTM hidden state dimension	300
	Batch size	32

The results of text classification based on machine learning are shown in Table 3.

Table 3 Model Experiment Results

Method	Accuracy	precision	recall	F1-score
SVM + TF-IDF	78.50	78.66	78.47	78.45
NB + TF-IDF	79.31	79.31	79.31	79.31

The results of text classification based on deep learning are shown in Table 4.

Table 4 Result of deep learning model

Method	Accuracy	precision	recall	F1-score
RNN	69.00	69.27	69.00	68.89
RNN + Word2vec	73.48	73.49	73.38	73.40
CNN	79.36	79.34	79.33	79.36

CNN + Word2vec	80.53	80.53	80.49	80.49
LSTM	84.00	84.95	83.83	83.83
LSTM + Word2vec	85.84	85.88	85.87	85.84

From the above table, it can be found that the naive Bayesian model based on the traditional machine learning method is slightly better than the support vector machine model, and the corresponding accuracy rates are 79.31% and 78.50%, respectively.

It can be found from Table 4 that the deep learning model using word2vec word vector is better than the model without word2vec. the text classification method based on RNN performs the worst, and the corresponding accuracy rate was 73.48%. and the classification method based on LSTM and CNN performs better, and the corresponding accuracy rate was 85.84%, 80.53%.

It can be found by comparing traditional machine learning methods and deep learning methods. The results show that the sentiment analysis method based on LSTM and CNN have a better performance than SVM and NB. RNN have a worst performance. The main reason is that the RNN model can only obtain information for a short period of time, and the connection between different verses in the Qur'an is not very close. So, RNN performs poorly on the task of classifying the Qur'an text. Finally, the LSTM and CNN models achieved good results in Quran text classification, with corresponding accuracies of 85.85% and 80.53%, respectively.

VI.   CONCLUSION

The main research goal of this research is to divide the Qur'an text into encouraging verse and warning verse. The main work is summarized as follows:

1. In this research, two sets of experiments were set up. The first group is based on SVM and NB classifiers, using the TF-IDF algorithm to calculate the feature vectors. The second group is based on deep learning methods CNN, RNN and LSTM, using Word2ve technology to train feature vectors. The experimental results of the two groups of models were compared and analysed.

2. Based on the LSTM model, the influence of different parameters on the performance of the model is explored.

The results of the two groups of models were compared and analysed. The results show that the sentiment analysis method based on LSTM and CNN have a better performance than SVM and NB. RNN have a worst performance. The main reason is that the RNN model can only obtain

information for a short period of time, and the connection between different verses in the Qur'an is not very close. So, RNN performs poorly on the task of classifying the Qur'an text. Finally, the LSTM and CNN models achieved good results in Quran text classification, with corresponding accuracies of 85.85% and 80.53%, respectively.

This research also compares and analyzes the influence of different parameter values on the model results.

1. As for the dimension of word embedding, when the dimension of word embedding gradually increases from low to high, the accuracy of the model also increases, but when the dimension increases to a certain value, the accuracy of the model no longer increases, on the contrary there will be Downward trend.

2. As for the value of Dropout, If the Dropout value is too large and too few neural units are reserved, it will have a bad impact on the training results.

3. As for the value of batch size, the batch size value should not be too large or too small. When the batch size value is small, the time consumed by each iteration is greater, but the accuracy of the model is not significantly improved. When the batch size is too large, the processing speed is faster, but because the parameter correction is slower, more iterations are needed to achieve the same result. Therefore, an appropriate size needs to be selected.

ACKNOWLEDGEMENT

REFERENCE

Ahmad Al-Sallab, Ramy Baly, Hazem Hajj, Khaled Bashir Shaban, Wassim El-Hajj, and Gilbert Badaro. 2017. Aroma: a recursive deep learning model for opinion mining in arabic as a low resource language. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16(4):25.

Chu D. C., & Abdulla M. M. 2014. Self-efficacy beliefs and preferred gender role in policing: An examination of policewomen's perceptions in Dubai, the United Arab Emirates. British journal of criminology 54(3): 449-468.

Collobert R, Weston J. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. Proceedings of the 25th international conference on Machine

learning. ACM, 2008: 160-167.

Guido Zarrella and Amy Marsh. 2016. Mitre at semeval-2016 task 6: Transfer learning for sentance detection. arXiv preprint arXiv:1606.03784, 2016.

Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20 (1995): 273-297.

LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

Hochreiter S., Schmidhuber J. 1997. "Longshort-termmemory". Neuralcomputation9, 1735–1780.

Kreaa A., Ahmad A. and KabalanK. 2014. ARABIC WORDS STEMMING APPROACH USING ARABIC ORDNET. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.4, No.6, November.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 16:321–357.

Nizar Y Habash. 2010. Introduction to Arabic natural language pro- cessing. Synthesis Lectures on Human Language Technologies 3 1(2010):1–187.

S. Morsy. 2011. "Recognizing Contextual Valence Shifters in Document- Level Sentiment Classification". Department of Computer Science and Engineering, The American University in Cairo (AUC).

Sadam Al-Azani and El-Sayed El-Alfy. 2018. Emojis-based sentiment classification of arabic microblogs using deep recurrent neural networks. International Conference on Computing Sciences and Engineering (ICCSE) 1–6.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) 31–41.

Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengjiao Wang. 2016. pkudblab at semeval-2016 task 6: A specific convolutional neural network system for effective stance detection. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) 384–388.

Yoon Kim.2014. Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)\ 1746–1751.

Al-Azani, Sadam, and El-Sayed El-Alfy. "Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks." 2018 international conference on computing sciences and engineering (ICCSE). IEEE, 2018

AlHasani, H., Saad, S., & Kassim, J.M. (2018). Classification of Encouragement (Targhib) And Warning (Tarhib) Using Sentiment Analysis on Classical Arabic. International Journal on Advanced Science, Engineering and Information Technology.

Alshammeri, M., Atwell, E., & Alsalka, M. (2021, December). Classifying Verses of the Quran using Doc2vec. In Proceedings of the 18th International Conference on Natural Language Processing (ICON) (pp. 284-288).

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

Al-Kabi, Mohammed N., et al. "A topical classification of Quranic Arabic text." Proceedings of the 2013 Taibah University International Conference on Advances in Information Technology for

the Holy Quran and Its Sciences. 2013.

Hadi, Wa'el Musa, Ma'an Salam, and Jaber A. Al-Widian. "Performance of NB and SVM classifiers in Islamic Arabic data." Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications. 2010.

Boukil, Samir, et al. "Arabic text classification using deep learning technics." International Journal of Grid and Distributed Computing 11.9 (2018): 103-114.

Abdeen, Mohammad AR, et al. "A closer look at arabic text classification." International Journal of Advanced Computer Science and Applications 10.11 (2019).

Khreisat, Laila. "A machine learning approach for Arabic text classification using N-gram frequency statistics." Journal of Informetrics 3.1 (2009): 72-77.

Liu, Bing, et al. "Building text classifiers using positive and unlabeled examples." Third IEEE international conference on data mining. IEEE, 2003.

Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." The Journal of physiology 148.3 (1959): 574.