

Text classification based on gated recurrent unit combines with support vector machine

Muhammad Zulqarnain, Rozaida Ghazali, Yana Mazwin Mohmad Hassim, Muhammad Rehan

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia

Article Info

Article history:

Received Jan 22, 2019

Revised Jan 18, 2020

Accepted Feb 1, 2020

Keywords:

Deep learning

GRU

Support vector machine

Text classification

ABSTRACT

As the amount of unstructured text data that humanity produce largely and a lot of texts are grows on the Internet, so the one of the intelligent technique is require processing it and extracting different types of knowledge from it. Gated recurrent unit (GRU) and support vector machine (SVM) have been successfully used to Natural Language Processing (NLP) systems with comparative, remarkable results. GRU networks perform well in sequential learning tasks and overcome the issues of vanishing and explosion of gradients in standard recurrent neural networks (RNNs) when capturing long-term dependencies. In this paper, we proposed a text classification model based on improved approaches to this norm by presenting a linear support vector machine (SVM) as the replacement of Softmax in the final output layer of a GRU model. Furthermore, the cross-entropy function shall be replaced with a margin-based function. Empirical results present that the proposed GRU-SVM model achieved comparatively better results than the baseline approaches BLSTM-C, DABN.

Copyright © 2020 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Rozaida Ghazali,

Faculty of Computer Science and Information Technology,

Universiti Tun Hussein Onn Malaysia,

86400, Parit Raja, Batu Pahat, Johor, Malaysia.

Email: rozaida@uthm.edu.my

1. INTRODUCTION

With the quickly and continuous production of informations technologies in digital format, the rapidly increasing in the number of electronic text message, on the world wide web (www), such as important news sites, forums, blogs and soon text classification has become very serious problem for the big organization and companies to manage and organize the textual data. Text classification is a technique of classifying unorganized text automatically according to the predefined classes or categories based on the text contents through a given classification system. Nowadays, Text classification becomes an important task for automatically classifying the documents to their respective categories. Furthermore, text classification technique have to use in many applications of NLP, such as spam filtering [1], user intention analysis [2], text classification [3, 4], personalized news recommendation [5], email categorization [6], sentiment analysis [7], and sentences classification [8], in which require to allocate pre-defined categories to a sequential of text. A most of general and basic technique of representing texts is bag-of-words. While, in bag-of-words approach has an issue to lose the words sequence and ignore the semantic of words [5]. Therefore, in this context, how to organize and use these large amounts of text information becomes particularly important. Furthermore, at present most of the common and important classifier has been adopted in machine learning-based algorithms to initiated in traditional text classification are: naive Bayes, support vector machine, neural network, k-nearest neighbors, fast text, and decision trees algorithm [9]. Comparatively, SVM has greater implementation requirements to fulfill the theory, but it has been used to produce better results in many fields [10, 11]. However, their performance depends generally on the quality

of hand-crafted features [12]. Bin [13] proposed the latest type of RNN as known GRU, which utilizes the advantages of GRU and CNN jointly, to recognize medically relations in clinical records, with only words embedding features.

Over the years, various deep learning approaches have been presented to address this issue and have extended to implement of RNNs to real-world issues. However, based on the literature reviews and our research experiment, we found that some deficiencies of standard RNN are the gradient vanishing and exploding issues. It makes the training of RNN difficult, in two ways: (i) it cannot process very long sequences if using hyperbolic tanh activation function; (ii) it is very unstable if using the rectified linear unit (ReLU) as an activation function. In this research, for manage to overcome these issues we have to use the latest type of recurrent networks such as GRU network, initially introduced by Kyunghyun Cho 2014. GRU model indicates a strong performance on real-world sequential tasks. In fact, the influence of this research we have to combine two most of the strong state-of-the-art approaches, gated recurrent unit (GRU) and support vector machine (SVM), both of which can be traced back [14]. GRU has the capability of a dynamical system where the network output will be decided not individually by the current input but also by decide current network state. The output is base on the previous computations. It is characterized by the good model of sequential data and completely utilizations of sequence information.

Furthermore, in recent year gated recurrent unit (GRU) have used because of the short-comings of standard RNNs are addressing with long text. The advantage of the GRU model is that they determine in the process of synthesis to control how much information should be received in the current synthesis steps, how much is forgotten, and how much information is passed back. Through these gate controls, GRU has a proficient learning ability for long text. Given the existing literature, these models have been implemented all memory free. Consequently, the RNN networks with memory operation are obviously more appropriate for this type of task. In this research paper, we mainly perform a threefold contribution.

Firstly, we concentrate on gated recurrent unit (GRU) network, a type of gated RNNs, which greatly minimizes the vanishing gradient and exploding issues of RNNs over gating mechanism that constructs the simple architecture while preserving the impact of LSTM (is a another type of RNN). Secondly, we proposed to replace the hyperbolic tangent activation function (tanh) with leaky rectified linear unit (LRelu) activation in the state to the modified equation. LRelu units have been demonstrated to be better performance than sigmoid non-linearities for feed-forward DNNs. Thirdly, we present a revising to this standard by implementing linear support vector machine (SVM) as the replacement of traditional Softmax in the final output layer of a GRU model.

2. SUPPORT VECTOR MACHINE

The support vector machine was introduced by Vapnik in 1990s and its popularity has increased quickly. It is essentially due to its achieved excellent state-of-the-art performances in various real-world applications and generalized better on invisible datasets. Furthermore, the alternate significant component is that, dissimilar neural networks, SVM generate consistent results. SVM was originally designed for problems of two classes and contains both positive and negative objects. The basic idea of SVMs is assign the input vector to an entity space with a complex dimensional feature and find the best possible linear hyperplane that splits into two target classes at the maximally the margin [15]. Furthermore, many support vectors are used to estimate the generalization performance, the number decreases as the number of support vectors increases. The features space mapping from the input space is achieved through the kernel tricks, which allows mapping to a higher space dimensionality without the essential to explicitly find this space. The kernel can be determined as any functionality that satisfies Mercer's theorem [16]. The radials base kernel and linear kernel are between the most extensively applied kernels with SVM.

However, development of the latest kernels that shows the similarities with various applications is an active exploration field and several kernels have been developed in recent years to handle particular application. SVM was initially designed as binary classifiers so several methods have used to increase SVMs to the multiple classes issue [17]. The principle methods are "one against all" (OAA) and "one against one" (OAO) where in the 1st technique n SVM classifiers will be construct, one for each class, and $\frac{(n-1)n}{2}$ binary SVM classifier is applied; the mainstream voting between these classifiers have used in prediction to new points cite-multiclass support vector machines. The basic equations of the SVM that are applied to estimation the decision function from a training dataset [18] are as follows:

$$f(x) = \text{sign}(\sum_{n=1}^l y_n a_n \cdot k(x, x_n) + b) \quad (1)$$

where b is the term of bias and l is presenting the number of the support vectors, $y_n \in \{-1,+1\}$ is the class signs to which the support vector belong and α is got as the solution of the subsequent quadratic optimization issues:

$$\begin{aligned} \min & \frac{1}{2} w^T w + C \sum_{i=1}^p \varepsilon_i (2) \\ \text{s.t } & y_i (w^T \phi(x_i) + b) \geq 1 - \varepsilon_i \\ & \varepsilon_i \geq 0, i = 1, \dots, N \end{aligned}$$

in the dataset, the number of data points cannot increase the number of support vectors; preferably it should be a comparatively small portion of the dataset [19].

3. GRU-BASED MODEL

The GRU was recently introduced by [20]. GRU is one of the newer generations of gated RNNs that have applied to solve the basic issues of vanishing gradient and exploding in standard RNNs when capturing long-term dependencies. This GRU approach is based on dynamically analysis that suffers from gradients explosion due to their non-linear dynamic. It was designed to properly update or reset its memory contents and is a lightly more simplify variation of LSTM [21]. It is combination of the input and forget gates into a single “update gate” and has an additional “reset gate”. The GRU model is simple and has fewer parameters as compare to traditional LSTM models and are gain progressively popularity.

Figure 1 illustrating, the multiple neurons are composed in a single input layer, the numbers of a neuron are decided by the scope of the features space. Equivalently, the output layer neurons are resembles to the output space. However, distinct the LSTM, the GRU completely exposed it’s memory content each time step and balances among the new memory contents and the previous memory contents severely use of the leaky combination, although with its adjustable time continuously control by update gate.

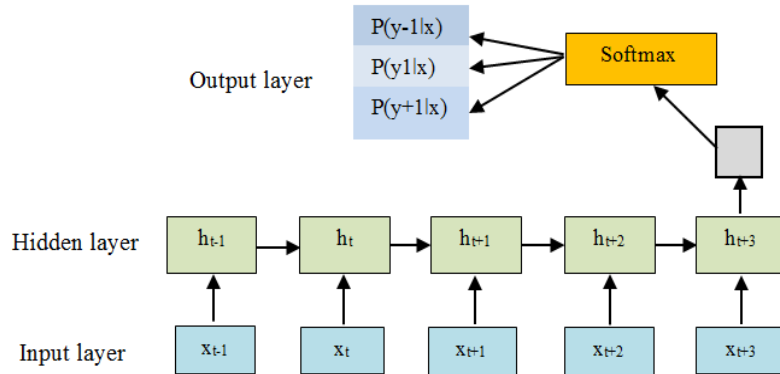


Figure 1. Structure of the GRU-based model

The activation h_t^i of the GRU at time t is a linear interpolating among the previous activations h_{t-1}^i and the candidate activation \hat{h}_t^i :

$$h_t^i = (1 - z_t^i) h_{t-1}^i + z_t^i \hat{h}_t^i \quad (3)$$

The update gate z_t^i determine how much the unit updates its activation:

$$z_t^i = \text{sigm} (W_z x_t + U_z h_{t-1}^i) \quad (4)$$

The candidate activation \hat{h}_t^i is calculated similar to the update gate:

$$\hat{h}_t^i = \text{tanh} (W x_t + U (r_t * h_{t-1}^i)) \quad (5)$$

where x_t is the input vector, h_t is the hidden state vector, r_t^i are reset gates and * denoted by element-wise product. When the reset gate is off ($r_t^i == 0$) it permits the unit to forget the previous information. This is similar to allow learning of input sequential symbol. The reset gate is evaluated using the subsequent equation.

$$r_t^i = \text{sigm}(W_r x_t + U_r h_{t-1}) \tag{6}$$

Update gate z decides how much the previous information (from the previous time step t) should throw away and what new state to add. Units of reset gate r_t^i are properly active on short-term dependencies. Units with long-term dependencies have an active update gate.

4. THE PROPOSED GRU-SVM ARCHITECTURE

In the exiting system mostly researchers have used machine learning and data mining algorithms to classify the text. But in our proposed model we concentrate on the deep learning algorithms to classify the text. In several studies and researches verified that deep learning algorithms produced better accuracy than machine learning and traditional algorithms. In this paper, we discussed the RNN algorithm to acquire better accuracy. The architecture of our proposed system is shown in Figure 2.

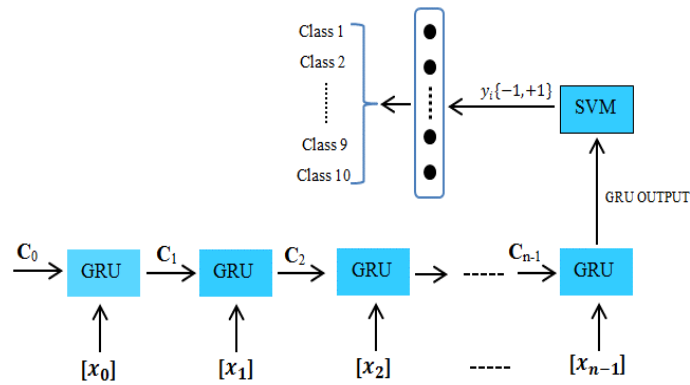


Figure 2. The proposed GRU-SVM architecture model, with timestep, GRU input unit and SVM as its classifier

Similarly, to the work of Alalshakmubarak [22] and shows that the propose to apply SVM as the classifier in a deep learning architecture especially GRU. Then, the parameters are trained across the gating mechanism of GRU architecture [23]. However, gated recurrent unit GRU overcomes such deficiencies of existing RNNs by adding the RNN with an update and reset gates that take as an input x_t , h_{t-1} , and generates \hat{h}_t , by the following:

- Update gate:

$$z_t = \sigma(W_z \cdot [U h_{t-1}, x_t]) \tag{7}$$

- Reset gate:

$$r_t = \sigma(W_r \cdot [U h_{t-1}, x_t]) \tag{8}$$

- Candidate hidden state:

$$\hat{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{9}$$

The second modification consists of replacing the standard hyperbolic tangent with Leaky-ReLU (LRelu) activation function. In particular, we modify the calculation of candidate state \hat{h}_t in (9) as follows:

$$\hat{h}_t = LRelu(W \cdot [r_t * h_{t-1}, x_t]) \tag{10}$$

- Hidden state:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \hat{h}_t \quad (11)$$

where W, U are weight matrices, σ is a logistic sigmoid activation function, and h_{t-1} is the hidden state (from previous time step t). But as a final layer introduced by SVM and its replacement of traditional softmax layer in a GRU the optimum separating hyper-plane parameters are also solved the subsequent regularization by improving the principal function of SVM as shown in equation 12:

$$J(\tilde{\omega}, \varepsilon) = \min \frac{1}{2} \tilde{\omega}^T w + C \sum_{i=1}^N \varepsilon_i \quad (12)$$

$$y_i(\tilde{\omega}^T \phi(i_{-t}, + b) \geq 1 - \varepsilon_i \text{ and } \varepsilon_i \geq 0, i = 1, \dots, N \quad (13)$$

Consider a training data set $G = \{[\tilde{x}_i, y_i]\}_{i=1}^n$, $\tilde{x}_i \in R^m$ is the i^{th} input design and y_i is its equivalent observed result $y_i \in \{+1, -1\}$. In test classification model, x_i denoted by attributes of text vector, y_i is class label, where c is a constant representing a tradeoff among the edge and the calculation of total errors. $\phi(\tilde{x})$ is a non-linear function that plots the input space into a superior dimensional feature space. The margin between the two parts is $2/w$.

The L2-SVM has to applied for the proposed GRU-SVM structure. As for the prediction, the decision function $f(x) = \text{sign}(\mathbf{w}\mathbf{x} + b)$ generates a score vector for each classes. So, to develop the predict class label y of a data x , the argmax function is employed:

$$\text{Predict-class} = \text{argmax}(\text{sign}(\mathbf{w}\mathbf{x} + b)) \quad (14)$$

The proposed GRU-SVM approach is summarized as follows:

- Input the dataset features $\{x_i | x_i \in R^m\}$ to the GRU network.
- Initialize arbitrary values with the learning parameters weights and biases (they will be adjusted through training).
- The GRU network cell state is calculated based on the input features x_i , and its learning parameters values.
- At the last time step, the prediction of the model is computed using the decision function of SVM: $f(x) = \text{sign}(\mathbf{w}\mathbf{x} + b)$.
- An optimization method is employed for loss minimization (Adam optimizer was applied in this research). Optimization adjusts the weights and biases based on the computed loss.
- This process is repeated until the recurrent neural network achieves the appropriate accuracy or achieves the highest possible accuracy.

5. EXPERIMENT SETUP

In this section, we explain the experimental settings and empirical results of the proposed model.

5.1. Collection of datasets

In this section, we adopted the ordinary experimental datasets that contain Chinese corpus and collected by Fudan University Dr. Li Rongla. In the automatically text classification system, the experimental dataset is typically divided into 2 parts: the training set and testing set. We randomly selected 10 categories from the corpus and deleted some error documents. Finally, the corpus consists of 10 categories and 1885 documents in the training sets and testing sets contain 940 documents. According to the text categorization system setting, each class should include a particular amount of training text. Each one of these texts was concluded using the classifier, and then the classification results distinction to the accurate determination. Therefore, we can determine the model impact with the SVM classifier. Detail description of dataset is presented in Table 1.

In addition, there are three evaluation index measures the efficiency of text classification:

$$\text{Recall} = \frac{a}{a + c} \quad (15)$$

$$\text{Precision} = \frac{a}{a + b} \quad (16)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (17)$$

where a represent the positive instance test documentations that are used accurately classified the class as belong to the number of this category. b are showing the negative instance of test documentations that are error classified for belongs to the number of this category. c are the positive instances of test documentations that are be error classified for does not belong to the number of this category. Furthermore, higher precision means that a model return dramatically more related results than irrelevant, also high recall means that an algorithm has returned best of the relevant results. So, we aspect to obtain higher values of precision and recall.

Table 1. Experimental dataset

Text class	Training sets	Test sets
Computer	133	67
Art	165	83
Entertainment	134	67
Education	149	74
Economy	214	105
Policy	332	164
Sport	300	150
Medicine	136	68
Business	181	90
Transportation	141	70

5.2. Hyperparameters and software

Data preprocessing and manipulate have performed in Python 3.6, basis on the sklearn, numpy and pandas packages. Deep learning GRU network and traditional DNNs have to executed with TensorFlow, an open source software library for numerically computations through data flows graph. Performance of all methods was based on pre-defined assessment measures. Completely simulations were implemented on Intel Core i7-3770XPU @3.40 GHz, and 4GB of RAM machine. The detail of hyperparameters of the proposed model is presented in Table 2.

Table 2. Detail of hyperparameters used in proposed GRU-SVM model

Hyperparameters	GRU-SVM
Batch Size	128
Cell Size	128
Dropout Rate	0.85
Epochs	15
Learning Rate	1e-5
SVM_C	0.5
No. of classes	10

6. RESULTS AND DISCUSSION

To evaluates and compared the performance of GRU-SVM text classification algorithm with baseline DABN and BLSTM-C classification algorithms, we hav experimented on well-known Chinese corpus datasets. we have concluded that the GRU-SVM approach has best text classifications capability in the term of the recall, precision and the F1 compare with the existing approaches of deep autoencoder belief networks (DABN) [24] and Bidirectional Long Short Term Memory with convolutional layer (BLSTM-C) [25]. The proposed technique performed better text classification on the datasets such as computer, art, entertainment, politics, education, sports, and transportation. While proposed improved GRU-SVM model performance is not good to classify accurate class in the kind of business, economy, and medicine. This may be because in eliminating related features of a test result, and lost several information. Therefore, that the recall rate criteria have affected. Consequently, it is also require for further improvement. Therefore, the proposed GRU-SVM model can provide a powerful algorithm in performing text classification tasks. Figure 3 to Figure 5 shows the performance of the proposed GRU-SVM model with the traditional baseline approaches.

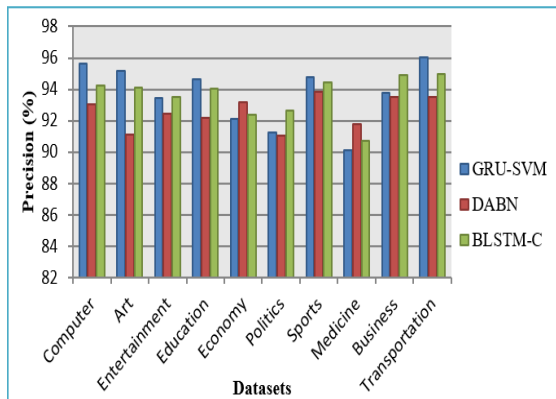


Figure 3. Comparison performance of the proposed model with baseline approaches in the term of precision

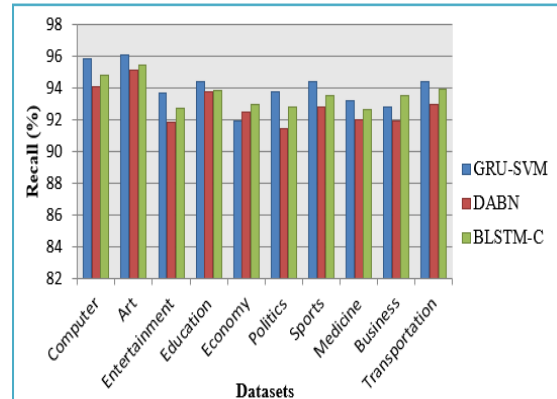


Figure 4. Comparison performance of the proposed model with baseline approaches in the term of recall

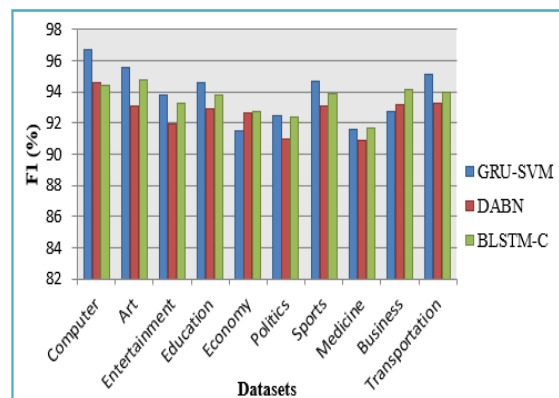


Figure 5. Comparison performance of the proposed model with baseline approaches in the term of F1

7. CONCLUSION

In this paper, we have proposed a gated recurrent unit network and replace its traditional softmax output layer with SVM for text classification. We have performed an experiment on the well-known Chinese corpus datasets to evaluate the performance of the GRU-SVM. The dataset is publicly available and included 1885 documents for training and 940 documents for testing. The results have been compared with two state-of-the-art DABN and BLSTM-C models. Through the practically data experiment, the proposed GRU-SVM model achieved the best text classification accuracy/performance rate of 94.75%, which demonstrated the potential of implementing GRU-SVM on multi-class classification issues with a small number of classes. Furthermore, our proposed method achieved much better performance in the term of precision, recall and F1 than DABN and BLSTM-C, particularly when the size of the storage limited.

ACKNOWLEDGEMENTS

The authors would like to thanks Ministry of Education Malaysia, Universiti Tun Hussein Onn Malaysia and Research Management Center (RMC) for funding this research activity under the Fundamental Research Grant Scheme (FRGS), vote no.1641.

REFERENCES

- [1] G. Tzortzis and A. Likas, "Deep Belief Networks for Spam Filtering," *19th IEEE Int. Conf. Tools with Artif. Intell.* 2007), pp. 306–309, 2007.
- [2] Y. Ning *et al.*, "Multi-task deep learning for user intention understanding in speech interaction systems," *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 161–167, 2017.

- [3] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Rehan, "A comparative review on deep learning models for text classification," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 1856–1866, 2020.
- [4] R. Ghazali, N. A. Husaini, L. H. Ismail, T. Herawan, and Y. M. M. Hassim, "The performance of a Recurrent HONN for temperature time series prediction," *Proc. Int. Jt. Conf. Neural Networks*, no. July, pp. 518–524, 2014.
- [5] K. J. Oh, W. J. Lee, C. G. Lim, and H. J. Choi, "Personalized news recommendation using classified keywords to capture user preference," *Int. Conf. Adv. Commun. Technol. ICACT*, pp. 1283–1287, 2014.
- [6] G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email Classification Research Trends: Review and Open Issues," *IEEE Access*, vol. 5, pp. 9044–9064, 2017.
- [7] S. M. Saqib, F. M. Kundi, and S. Ahmad, "Unsupervised Learning Method for Sorting Positive and Negative Reviews Using LSI (Latent Semantic Indexing) with Automatic Generated Queries," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 1, pp. 56–62, 2018.
- [8] M. Zulqarnain, R. Ghazali, S. H. Khaleefah, and A. Rehan, "An Improved the Performance of GRU Model based on Batch Normalization for Sentence Classification," *Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 9, pp. 176–186, 2019.
- [9] V. Korde and C. N. Mahender, "Text Classification and Classifiers: A Survey," *Int. J. Artif. Intell. Appl.*, vol. 3, no. 2, pp. 85–99, 2012.
- [10] W. Sharif, N. A. Samsudin, M. M. Deris, and M. Aamir, "Improved relative discriminative criterion feature ranking technique for text classification," *Int. J. Artif. Intell.*, vol. 15, no. 2, pp. 61–78, 2017.
- [11] R. Ghazali, Z. A. Bakar, Y. Mazwin, and M. Hassim, "Functional Link Neural Network with Modified Cuckoo Search Training Algorithm," *Int. Conf. Intell. Comput.*, pp. 285–291, 2014.
- [12] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, "Deep Learning for Extreme Multi-label Text Classification," *Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '17*, pp. 115–124, 2017.
- [13] B. He, Y. Guan, and R. Dai, "Convolutional Gated Recurrent Units for Medical Relation Classification," *J. Mach. Learn. Res.*, no. July, pp. 118–127, 2018.
- [14] L. Wei, B. Wei, and B. Wang, "Text Classification Using Support Vector Machine with Mixture of Kernel," *A J. Softw. Eng. Appl.*, vol. 2012, no. December, pp. 55–58, 2012.
- [15] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A Novel Active Learning Method Using SVM for Text Classification," *Int. J. Autom. Comput.*, vol. 15, no. 3, pp. 290–298, 2018.
- [16] M. K. S. Varma, "Pixel-based Classification Using Support Vector Machine Classifier," *2016 IEEE 6th Int. Conf. Adv. Comput.*, pp. 51–55, 2016.
- [17] V. Hooshmand Moghaddam and J. Hamidzadeh, "New Hermite orthogonal polynomial kernel and combined kernels in Support Vector Machine classifier," *Pattern Recognit.*, vol. 60, pp. 921–935, 2016.
- [18] G. Indrawan, I. K. P. Sudiarsa, and K. Agustini, "Smooth Support Vector Machine for Suicide-Related Behaviours Prediction," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 5, pp. 3399–3406, 2018.
- [19] W. Sharif, U. Tun, H. Onn, I. Tri, and R. Yanto, "An Optimised Support Vector Machine with Ringed Seal Search Algorithm for Efficient Text Classification," *J. Eng. Sci. Technol.*, vol. 14, no. 3, pp. 1601–1613, 2019.
- [20] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *arXiv*, no. September, pp. 1–15, 2014.
- [21] M. Zulqarnain, S. A. Ishak, R. Ghazali, and N. M. Nawati, "An Improved Deep Learning Approach based on Variant Two-State Gated Recurrent Unit and Word Embeddings for Sentiment Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 594–603, 2020.
- [22] A. Alalshkumbarak and L. S. Smith, "A Novel Approach Combining Recurrent Neural Network and Support Vector Machines For Time Series Classification," *2013 9th Int. Conf. Innov. Inf. Technol.*, no. August, pp. 42–47, 2013.
- [23] M. Zulqarnain, R. Ghazali, M. G. Ghouse, and M. F. Mushtaq, "Efficient Processing of GRU Based on Word Embedding for Text Classification," *Int. J. Informatics Vis.*, vol. 3, no. 4, pp. 377–383, 2019.
- [24] Z. Yang and X. Pang, "Research and Implementation of Text Classification Model Based on Combination of DAE and DBN," *2017 10th Int. Symp. Comput. Intell. Des.*, pp. 190–193, 2017.
- [25] Y. Li, X. Wang, and P. Xu, "Chinese Text Classification Model Based on Deep Learning," *J. King Saud Univ. - Comput. Inf. Sci.*, no. November, pp. 20–32, 2018.

BIOGRAPHIES OF AUTHORS



Muhammad Zulqarnain received his Bachelor and Master degree in Computer Science and Information Technology from The Islamia University of Bahawalpur (IUB), Pakistan. He received his M.Phil degree (Master of Philosophy) from National College of Business Administration and Economics, Lahore, Pakistan. He is currently pursuing Ph.D from University Tun Hussein Onn Malaysia. His research interest is Machine Learning and Deep learning for natural language processing and its application



Rozaida Ghazali is currently a Professor at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She graduated with a Ph.D. degree in Higher Order Neural Networks from the School of Computing and Mathematical Sciences at Liverpool John Moores University, United Kingdom in 2007. Earlier, in 2003 she completed her M.Sc. degree in Computer Science from Universiti Teknologi Malaysia (UTM). She received her B.Sc. (Hons) degree in Computer Science from Universiti Sains Malaysia (USM) in 1997. In 2001, Rozaida joined the academic staff in UTHM. Her research area includes neural networks, swarm intelligence, optimization, data mining, and time series prediction. She has successfully supervised a number of PhD and master students and published more than 100 articles in various international journals and conference proceedings. She acts as a reviewer for various journals and conferences, and as an editor in a few Springer conference proceedings. She has also served as a conference chair, and as a technical committee for numerous international conferences.



Yana Mazwin Mohamad Hassim is a senior lecturer at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She graduated with a PhD degree from Universiti Tun Hussein Onn Malaysia (UTHM) in 2016. Earlier, in 2006 she completed her Master's degree in Computer Science from Universiti of Malaya (UM). She received her Bachelor of Information Technology (Hons) degree majoring in Industrial Computing from Universiti Kebangsaan Malaysia (UKM) in 2001. In 2003, Yana Mazwin joined the academic staff in UTHM. Her research area includes neural networks, swarm intelligence, optimization and classification.



Muhammad Rehan is an Assistant professor at the Department of Computer Science and Information Technology, Baqai Medical University (BMU) Karachi, Pakistan. He received his M.Phil. Degree (Master of Philosophy) from the Faculty of Computer Science and Information Technology, Hamdard University core area is Computer Networks in 2016. Karachi, Pakistan. He received his Bachelor of Computer Science degree majoring in Software Engineering from Dadabhoy Institute of Higher Education (DIHE) Karachi, Pakistan in 2004. He is currently pursuing his Ph.D. from University Tun Hussein Onn Malaysia (UTHM). His research area is Data Mining and Machine Learning for Optimization and its application.