

Citation:

M. R. Raza, W. Hussain and J. M. Merigó, "Cloud Sentiment Accuracy Comparison using RNN, LSTM and GRU," *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2021, pp. 1-5, doi: 10.1109/ASYU52992.2021.9599044.

Cloud Sentiment Accuracy Comparison using RNN, LSTM and GRU

Muhammad Raheel Raza
Department of Software Engineering
Firat University
Elazig, Turkey
191137125@firat.edu.tr
ORCID: 0000-0002-6305-2583

Walayat Hussain
Faculty of Engineering & Information
Technology
University of Technology Sydney
Sydney, NSW 2007, Australia
[walayat.hussain@uts.edu.au](mailto:walayathussain@uts.edu.au)
ORCID: 0000-0003-0610-4006

José Maria Merigó
Faculty of Engineering & Information
Technology
University of Technology Sydney
Sydney, NSW 2007, Australia
jose.merigo@uts.edu.au
ORCID: 0000-0002-4672-6961

Abstract— Cloud computing has become a de facto choice of many individuals and enterprises for computing solutions. In the last few years, many cloud providers appear in the market that offers the same services. It is a trivial job to choose an optimal service best suited for organisations in such a massive arms race of service providers. Existing consumer experience could help significantly build a holistic perception of their experiences that ultimately influence service adoption decisions. Sentiment analysis is an effective tool to understand consumer experience about the product or service. The sophisticated sentiment analysis could help businesses to gain a better insight and respond proactively to consumer issues. There are various methods for sentiment analysis that produces ideal results under different conditions. Therefore, it is very important to choose the right method to predict consumer's sentiment for a greatest result. In this paper we analyse the sentiment prediction accuracy of widely used neural network methods – recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent network (GRU). We use software as a service (SaaS) dataset having 6258 reviews. From analysis results we find that GRU outperforms the LSTM and RNN methods.

Keywords—Sentiment prediction; RNN; LSTM; GRU; Cloud reviews; Forecasting; Social influence.

I. INTRODUCTION

Cloud marketplace is growing every next day that provide convenience, dynamicity, better accessibility, affordability and affordance. Cloud services ranges from hardware to services in the form of infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS). In a myriad of cloud providers that offer similar type of services, it is a great challenge for a business and individuals to select the right service provider. Existing literature proposed numerous approaches to select the right service provider. Authors [1] used domain-specific ontology to extract related information from consumer's reviews for service selection. Different optimisation algorithms such as [2] simulated annealing, genetic algorithm and particle swarm optimisation algorithm are used for service selection. Other approaches [3-7] used various QoS prediction methods to assist the decision maker in a service selection. Sentiment analysis is an effective tool to analyse the user

experience of existing consumers that help a new consumer to determine the service decision. These feedbacks significantly impact consumer decision for service selection.

There are a variety of factors that directly or indirectly influence the behaviour of a consumer. Among them, word of mouth or social influence plays a key role in decision-making [8]. Social influence refers to the change in an individual's attitude or behaviour due to formal communication such as advertisement or informal communication such as word of mouth or feedback. Research [9] demonstrate that consumers are highly influenced by informal social influence such as product reviews or feedback. Such information acts as a foundation that ultimately impacts consumers' decision to buy or use any service [10]. Consumer reviews or online word of mouth have a great impact on the business process. Consumers reviews have a high impact on new consumers, and the written text enables consumers to process information based on their own pace and get greater details. According to one of the latest reviews by Local Consumer Reviews [11], 87% of consumers read reviews for local businesses, with 64% of consumers who spend at least 10 minutes before buying any product. Keeping in mind the importance of consumer's feedback it is essential to analyse and predict the sentiment for the optimal strategic decision-making process [12].

There are various methods to implement sentiment analysis of consumer's feedback. Neural network methods such as Convolutional Neural Network (CNN), Deep Neural Network (DNN), Recurrent Neural Network (RNN) have set new standards to achieve high accuracy [13]. CNN is commonly used in different computer vision models that can use convolutions in the embedded features of the words in a sentence to extract sentiment information. RNN is commonly used for various natural language process (NLP) models to predict sentiment from text data. Long Short-term Memory (LSTM) is a type of RNN that handles long term dependencies of data and overcome the vanishing gradient issue of the RNN model. LSTM has three gates – input, output and forgets gate. Gated recurrent unit (GRU) is another method to handle short-term memory of the RNN

model with two gates – reset and update. Both the LSTM and GRU give the best results at various conditions [14]. To gain new insight and an optimal strategic decision-making process, the service provider always needs a sophisticated sentiment prediction system.

In this paper we analyse the prediction accuracy of RNN, LSTM and GRU on cloud reviews. The approaches are compared using SaaS reviews extracted by harvesting as a service (HaaS) crawler [15]. The approaches are compared based on precision, recall, F1-score and support. The rest of the paper is organised as follows: Section 2 present background study of sentiment classification and the proposed approach. Section 3 contains the implementation of analysis using the approaches. Results are displayed in Section 4.

II. SENTIMENT ANALYSIS APPROACHES

Sentiment Analysis is a technique of analyzing text and identifying opinions behind the text using various text classification methods of Machine Learning and NLP. Sentiment Analysis is used to recognize user reviews and classify them. It is applicable on social media, ecommerce website comments, and subjective political opinions to identify user comments and rank them to predict priority. Lappeman, et al. [12] has used human validation techniques to identify the sentiments of bigdata on social media and calculated the sentiment scores. The technique is quite useful for large organization and marketing purposes where the big data is complex to identify and the aim is to indicate absurd or misleading texts. Moreover, for performance improvement, customers' comments play a vital role and sentiment classification is the sole effective approach to achieve the purpose. Contextual based of HKT and TSI are used for sentiment identification purpose by Aziz & Starkey [13] following a tree structure to relate the words and sources of texts.

In cloud computing, a cloud consumer gets benefitted from the services provided by a cloud provider on demand. Palomba, et al. [14] inspect the mobile app reviews on crowdsourced websites that would help developers to bring innovations as desired. Services w.r.t storage, infrastructure and platform satisfies the service consumer and a trustful provider-consumer relationship is established [24]. In the cases, where the provider fails to allow sufficient resources, service consumers are distracted and tend to avail services from another available service provider. Service consumers' feedbacks regarding a specific service provider assists to identify an ideal service provider for a situation.

Our study performs sentiment analysis of cloud consumer reviews dataset extracted from web portals. Deep Learning approaches of RNN, LSTM and GRU clearly identifies which model brings the highest accuracy results. None of the literature, yet performed a combined approach to compare the conclusions.

A. Recurrnt Neural Network (RNN)

The Recurrent Neural Network (RNN), that is unfolded in time, constitutes a sort of Feedforward Neural Network that recognizes patterns of data and predicts them using sequential input. It follows the principle in such a way that output, from one computation in a cell, is reutilized and fed as an input to another cell. In this way, it is different from traditional feedforward neural network.

'al' stands for input layer, 'hdl' for the hidden layer, and 'cl' for the output layer. X, Y, and Z are network parameters that improve outputs. The present state would have a mixture of inputs, al(t) and hdl(t-1), for time 't'. As a result, the present state of the network can be expressed using the formula below:

$$hdl_t = f(hdl_{t-1}, al_t) \quad (1)$$

Here, hdl(t) is the current hidden state. Activation function tanh() is applied and the formula becomes:

$$hdl_t = \tanh(W_{hdl} * hdl_{t-1} + W_{al} * al_t) \quad (2)$$

Here, W(hdl) is the weight at the previous hidden state, W(al) is the weight at the current input state and tanh() is the activation function that implements non-linearity. The output formula is:

$$cl_t = W_l * hdl_t \quad (3)$$

cl(t) is the output of the current state, hdl(t) is the current hidden state and W(cl) is the weight at output state.

RNN models need to be trained for backward propagation. For that purpose, an extension of backpropagation is used, called as Backpropagation through Time (BPTT) algorithm. For too long sequences, the backpropagation process may be too expensive w.r.t operational time and computational cost. Thus, due to gradient vanishing and exploding gradient problems, a Truncated BPTT is used that truncates feedback connection and forces the gradient to flow back, disabling the network to learn full dependencies, thus solves the exploding gradient problem. For the vanishing gradient problem, choosing a suitable activation function can be a solution. However, as an alternative, it is preferred to shift to another ANN architecture, known as Long Short-Term Memory (LSTM), to deal with gradient problems efficiently.

B. Long Short-Term Memory (LSTM)

LSTMs are a sort of RNN that has been modified to deal with the complexity of RNN design [27]. Unlike RNN, LSTM is meant to solve the vanishing gradient problem by remembering long-term historical data as a default behaviour. As a result, it learns long-term dependencies and does backpropagation model training.

The architecture of LSTM comprises of a combination of cells, memory blocks that receive and transfer information to each other. A cell decides which data to store,

write, read or erase. Within the network, two states are transferred from one cell to the other, the cell state and the hidden state. Each cell contains three gates to process data: forget gate, input gate and output gate. Forget gate removes unnecessary information from the cell by erasing it so that LSTM will not use it. Input gate is responsible for adding new information for long-term to a cell state. Output of cell is expressed through the output gate. Sigmoid layer gives output as 0 or 1 where 0 means 'to remove' and 1 means 'to keep'.

A sigmoid layer accepts a mix of current data 'al(t)' and hidden state from previous cell 'hdl(t-1)' whenever data 'al' is supplied into a cell for time 't'. The sigmoid activation function's output of 0 or 1 determines how much of the previous data the LSTM should remember. As a result, it provides forget gate f(t). The formula is as follows:

$$f_t = a(W_f * [hdl_{t-1}, al_t] + bias_f) \quad (4)$$

It is calculated in two parts to determine how much data should be saved in cell state. To return 0 or 1, the combination of al(t) and hdl(t-1) is first sent via a sigmoid layer. Second, the same data is passed through the tanh function, which returns C(t), which is then added to the cell state. The input gate formulas are as follows:

$$i_t = a(W_i * [hdl_{t-1}, al_t] + bias_i) \quad (5)$$

$$c_t = \tanh(W_c * [hdl_{t-1}, al_t] + bias_c) \quad (6)$$

For updating the previous cell state C(t), multiply C(t-1) with f(t) to remove all extraneous data, multiply C'(t) with i(t) to obtain the new data, then sum both resultants as:

$$c_t = (f_t * c_{t-1}) + (i_t * c_t) \quad (7)$$

At the end, the output gate displays information about what the input and cell decide to output. Its formulas are:

$$o_t = a(W_o * [hdl_{t-1}, al_t] + bias_o) \quad (8)$$

$$h_t = o_t * \tanh(c_t) \quad (9)$$

LSTM method is suitable for huge datasets in which the model is trained using huge amount of training data. LSTM performs best for large organizations dealing with Bigdata.

C. Gated Recurrent Unit (GRU)

GRU are a kind of LSTM with gating structure for the information flow within and between cells. GRU contains two gates, the update gate and the reset gate. These gates help in controlling the vanishing gradient problem. Unlike LSTM, GRU doesnot maintain the cell state, instead a hidden state is preserved. As a result of which, GRU is easy to easier and faster for training small datasets.

The Reset gate handles the short term memory or hidden state of GRU. Its equation is given as:

$$r_t = a(W_r * H_{t-1} + a_t) \quad (10)$$

Update gate maintains the long term memory of GRU. Its equation is given as:

$$u_t = a(W_u * H_{t-1} + u * a_t) \quad (11)$$

To get the hidden state value H(t), a candidate hidden state is generated, as presented by the equation:

$$H'_t = \tanh(W * (H_{t-1} * r_t) + a_t) \quad (12)$$

From the candidate hidden state H'(), we can find out the current hidden state, as given in the equation:

$$H_t = u_t * H_{t-1} + (1 - u_t) * H'_t \quad (13)$$

III. IMPLEMENTATION

This section contains the implementation procedures carried out for sentiment analysis using various deep learning techniques.

A. Features of Dataset

The dataset used in this study is composed of cloud service consumer reviews concerning the services granted by different service providers. These services can be in terms of application, infrastructure and storage services. The dataset of reviews is harvested from diverse web portals using a novel method of Harvesting-as-a-Service (HaaS) [15]. The method works in such a way that it extracts the service information available on different web sources. The extracted data is then structured into a cloud reviews dataset. A total of 6258 cloud reviews are present in our dataset and are sentimentally analyzed into positive, negative, and neutral classes. A total of 5855 reviews are categorized as positive, 264 negative and 139 as neutral natured reviews. The dataset is split into 80:20 in terms of training and testing data. Results of the study depend upon the performance of the training dataset, which reflects the results obtained for the testing section. Some of the properties of the SaaS dataset used to be evaluated using three different deep learning approaches are given in Table 1.

TABLE I. OVERVIEW OF SAAS DATASET

SaaS Reviews Dataset	Training Dataset	Validation Section	Testing Dataset	Total
No. of Reviews	4895	140	1224	6259
Percentage of Reviews	78.207%	2.236%	19.555%	100.00%

B. Hyperparameters Configuration

The model used is Sequential in nature, in which there is a linear stack of layers. Each model contains a single layer of its specified layer. The overfitting problem is avoided using the spatial dropout layer with 0.2 as its dropout value. Learning rate of 0.001 balances the learning ability along with the Sigmoid activation function for all three models. Table 2 displays various hyperparameters of the LSTM architecture used for the model.

TABLE II. HYPERPARAMETER CONFIGURATION OF MODELS

Hyperparameters	SaaS Review Dataset
Model	Sequential
Epochs	5
Dropout value	0.2
Hidden units	196
Learning rate	0.001
Activation function	Sigmoid
Layers	[1-LSTM / 1-RNN / 1-GRU] per model

IV. EVALUATION OUTCOMES

Deep Learning approaches of RNN, LSTM and GRU are implemented to train the model using training sample of dataset. Each model used in our experiment, runs upto 5 epochs to complete its training. Within an epoch, the model trains itself though the whole training dataset in one cycle. After the training stage, the test sample is tested on the model based on the results obtained. Table 3 shows a comparison of accuracy rates achieved in each epoch by the implementation of DL approaches.

TABLE III. EVALUATION OF ACCURACY RATES PER EPOCH

Deep Learning approaches	Epochs (%)				
	1 st	2 nd	3 rd	4 th	5 th
RNN	0.945	0.958	0.958	0.956	0.964
LSTM	0.951	0.958	0.958	0.958	0.958
GRU	0.954	0.958	0.962	0.970	0.977

Comparing the results of prediction accuracy of the deep learning models, GRU outperforms the LSTM and RNN approaches. With each progressing epoch, GRU model seems to improve the accuracy values which indicates the accurate training and improved performance of the model. A graphical representation observes the differences between the defined approaches for analyzing sentiments. In the graph, the y-axis denotes the accuracy rates while the x-axis represents the number of epochs.

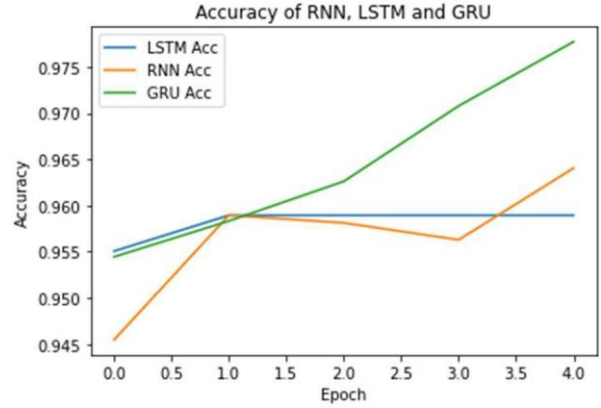


Fig. 1. Graphical Representation of Accuracy Rates of three approaches

The results are also classified with respect to some performance metrics. Precision, F1-score, Recall and Support values determine the success of approaches for a particular dataset. Following are the values of performance metrics that our experiment derives for RNN, LSTM and GRU models. Table 4, 5 and 6 divides the metric values about 0 for negative reviews and 1 for positive reviews.

TABLE IV. CLASSIFICATION METRICS FOR LSTM MODEL

Sentiment	Precision	Recall	F1-score	Support
0	1.00	0.85	0.09	63
1	0.95	1.00	0.97	1161

TABLE V. CLASSIFICATION METRICS FOR RNN MODEL

Sentiment	Precision	Recall	F1-score	Support
0	0.29	0.08	0.12	63
1	0.95	0.99	0.97	1161

TABLE VI. CLASSIFICATION METRICS FOR GRU MODEL

Sentiment	Precision	Recall	F1-score	Support
0	0.30	0.22	0.25	63
1	0.96	0.97	0.96	1161

V. CONCLUSION

Consumers give feedback to express their satisfaction or dissatisfaction about the product that it has used. Such feedback significantly influences consumer behaviour especially potential new consumers. In a highly competitive market such as a cloud marketplace, the service providers must identify the sentiment of their consumers and address any discrepancies proactively. Many sentiment analysis methods perform best at different conditions. Noen

of the literature performed a comparative analysis of all three deep learning approaches on a cloud dataset. This paper has selected three neural network methods – RNN, LSTM, GRU- and analysed their prediction accuracy using a cloud dataset. From analysis results, we have found that GRU performs best than RNN and LSTM methods. Thus, it can be derived that for small datasets, GRU outperforms LSTM and RNN techniques. In our future work, we will use the approach to analyse the sentiment of social media users in a complex decision-making process.

REFERENCES

- [1] A. M. Alkalbani and W. Hussain, "Cloud service discovery method: A framework for automatic derivation of cloud marketplace and cloud intelligence to assist consumers in finding cloud services," *International Journal of Communication Systems*, vol. 34, no. 8, p. e4780, 2021.
- [2] W. Hussain, F. K. Hussain, O. K. Hussain, E. Damiani, and E. Chang, "Formulating and managing viable SLAs in cloud computing from a small to medium service provider's viewpoint: A state-of-the-art review," *Information Systems*, vol. 71, pp. 240-259, 2017.
- [3] A. Alghamdi, W. Hussain, A. Alharthi, and A. B. Almusheqah, "The need of an optimal QoS repository and assessment framework in forming a trusted relationship in cloud: A systematic review," in *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)*, 2017, pp. 301-306: IEEE.
- [4] A. M. Mohamed and H. M. Abdelsalam, "A multicriteria optimization model for cloud service provider selection in multicloud environments," *Software: Practice and Experience*, vol. 50, no. 6, pp. 925-947, 2020.
- [5] W. Hussain, O. Sohaib, M. Naderpour, and H. Gao, "Cloud Marginal Resource Allocation: A Decision Support Model," *Mobile Networks and Applications*, vol. 25, pp. 1418-1433, 2020.
- [6] M. R. Raza, W. Hussain, E. Tanyıldızı, and A. Varol, "Sentiment analysis using deep learning in cloud," in *2021 9th International Symposium on Digital Forensics and Security (ISDFS)*, 2021, pp. 1-5: IEEE.
- [7] W. Hussain, F. K. Hussain, M. Saberi, O. K. Hussain, and E. Chang, "Comparing time series with machine learning-based prediction approaches for violation management in cloud SLAs," *Future Generation Computer Systems*, vol. 89, pp. 464-477, 2018.
- [8] W. Hussain and O. Sohaib, "Analysing Cloud QoS Prediction Approaches and its Control Parameters: Considering Overall Accuracy and Freshness of a Dataset," *IEEE Access*, vol. 7, pp. 82649-82671, 2019.
- [9] R. Spears, "Social influence and group identity," *Annual Review of Psychology*, vol. 72, pp. 367-390, 2021.
- [10] M. R. Raza, A. Varol, and W. Hussain, "Blockchain-based IoT: An Overview," in *2021 9th International Symposium on Digital Forensics and Security (ISDFS)*, 2021, pp. 1-6: IEEE.
- [11] J. Hörisch, I. Wulfsberg, and S. Schaltegger, "The influence of feedback and awareness of consequences on the development of corporate sustainability action over time," *Business Strategy and the Environment*, vol. 29, no. 2, pp. 638-650, 2020.
- [12] W. Hussain, F. Hussain, and O. Hussain, "QoS prediction methods to avoid SLA violation in post-interaction time phase," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, 2016, pp. 32-37: IEEE.
- [13] H. Butt, M. R. Raza, M. J. Ramzan, M. J. Ali, and M. Haris, "Attention-Based CNN-RNN Arabic Text Recognition from Natural Scene Images," *Forecasting*, vol. 3, no. 3, pp. 520-540, 2021.
- [14] BrightLocal. (2020, 03 Jul 2021). *Local Consumer Review Survey 2020*. Available: <https://www.brightlocal.com/research/local-consumer-review-survey/>
- [15] A. M. Alkalbani, W. Hussain, and J. Y. Kim, "A Centralised Cloud Services Repository (CCSR) Framework for Optimal Cloud Service Advertisement Discovery From Heterogenous Web Portals," *IEEE Access*, vol. 7, pp. 128213-128223, 2019.
- [16] W. Hussain, F. K. Hussain, O. Hussain, and E. Chang, "Profile-based viable service level agreement (SLA) violation prediction model in the cloud," in *2015 10th international conference on P2P, parallel, grid, cloud and internet computing (3PGCIC)*, 2015, pp. 268-272: IEEE.
- [17] M. R. Raza and A. Varol, "QoS parameters for viable SLA in cloud," in *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, 2020, pp. 1-5: IEEE.
- [18] A. Aldahiri, B. Alrashed, and W. Hussain, "Trends in Using IoT with Machine Learning in Health Prediction System," *Forecasting*, vol. 3, no. 1, pp. 181-206, 2021.
- [19] W. Hussain, F. K. Hussain, and O. K. Hussain, "Comparative analysis of consumer profile-based methods to predict SLA violation," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2015, pp. 1-8: IEEE.
- [20] M. U. Salur and I. Aydin, "A novel hybrid deep learning model for sentiment classification," *IEEE Access*, vol. 8, pp. 58080-58093, 2020.
- [21] M. R. Raza, A. Varol, and N. Varol, "Cloud and fog computing: A survey to the concept and challenges," in *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, 2020, pp. 1-6: IEEE.
- [22] S. Gao *et al.*, "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *Journal of Hydrology*, vol. 589, p. 125188, 2020.
- [23] J. Lappeman, *et al.*, "Studying social media sentiment using human validated analysis." *MethodsX*, 2020. 7: p. 100867.
- [24] W. Hussain, F. K. Hussain, and O. K. Hussain, "Risk management framework to avoid SLA violation in cloud from a provider's perspective," in *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2016, pp. 233-241: Springer.
- [25] A.A. Aziz, and A. Starkey, "Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches." *IEEE Access*, 2019. 8: p. 17722-17733.
- [26] F Palomba., *et al.*, "Crowdsourcing user reviews to support the evolution of mobile apps." *Journal of Systems*, 2018. 137: p. 143-162
- [27] W. Hussain, F. K. Hussain, and O. K. Hussain, "Maintaining trust in cloud computing through SLA monitoring," in *International Conference on Neural Information Processing*, 2014, pp. 690-697: Springer.
- [28] W. Hussain, F. K. Hussain, O. K. Hussain, and E. Chang, "Provider-Based Optimized Personalized Viable SLA (OPV-SLA) Framework to Prevent SLA Violation," *The Computer Journal*, vol. 59, no. 12, pp. 1760-1783, 2016.
- [29] M. R. Raza, W. Hussain, and J. M. Merigó, "Long Short-Term Memory-based Sentiment Classification of Cloud dataset," in *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Elazig, Turkey, 2021: IEEE, pp. 1-6.