Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 5, July : 2023

HYBRID GATED RECURRENT UNITS AND WEIGHTED AVERAGE CLUSTERING METHODS FOR USER ENGAGEMENT IN SOCIAL MEDIA

A. Raja Research Scholar, Periyar University, Salem, Tamil Nadu, India.
Dr. S. Prema HOD Department ofComputer Application, ArulmiguArthanareeswarar Arts and Science College, Tiruchengodu, Namakkal, Tamil Nadu, India

Abstract
Hybrid gated recurrent units with weighted average clustering methods for user engagement in online social media is a sophisticated approach aimed at enhancing user engagement on social media platforms. User engagement refers to the level of interaction and interest displayed by users towards the content and activities available on these platforms.The hybrid approach combines two powerful techniques: Gated Recurrent Units (GRUs) and Weighted Average Clustering. GRUs are a type of recurrent neural network that can effectively model sequential data, making them suitable for analyzing user behavior patterns over time. They capture the temporal dynamics of user engagement by considering the sequential nature of user activities on social media.Weighted Average Clustering, on the other hand, is a method that groups users based on their similarities and assigns different weights to these groups. This allows for the identification of user segments with distinct engagement patterns and preferences.This approach provides a comprehensive understanding of user engagement in online social media by capturing both temporal dynamics and user segmentation. It can assist platform developers and content creators in tailoring their strategies to different user segments, thus optimizing user engagement and improving the overall social media experience.

Keywords: Content consumption, Clustering Methods, Gated Recurrent Units, Temporal dynamics, User Engagement, Online social media.

1. INTRODUCTION

User engagement is crucial for social media platforms has direct impacts on their success, growth, and user satisfaction. Higher levels of user engagement indicate the vibrant and active user base, increased content consumption, and a stronger sense of community.By analysing user engagement patterns, they can tailor their strategies, content, and features to cater to user preferences, increase user satisfaction, and drive continued engagementinvolve personalized recommendations.

GRUs can effectively analyse the sequential nature of user activities and interactions over time. They can capture how users navigate through social media platforms, consume content, interact with others, and exhibit different engagement behaviours [1].GRUs are equipped with gating mechanisms that enable them to selectively retain or discard information at each time step. These gating mechanisms, composed of reset and update gates, allow to adaptively update their internal state and determine the parts of the input are relevant for predicting the next state. By learning and incorporating these temporal dependencies, GRUs can effectively model and predict user engagement patterns.

Gated Recurrent Units (GRUs) are a type of recurrent neural network (RNN) architecture has been widely used in user engagement analysis in online social media. GRUs offer essential features that enable them to effectively model sequential data and capture temporal dependencies. These features include gating mechanisms and hidden states. The gating mechanisms in GRUs consist of two gates: the reset gate (r) and the update gate (z). These gates control the flow of information within the network and determine which information to retain and update. The equations for the reset gate and the update gate in a GRU are as follows:

$$gate\ r(t) = \sum (W_r[h(t-1), x(t)]) \quad (1)$$

$$gate\ z(t) = \sum (W_z[h(t-1), x(t)]) \quad (2)$$

Here, r(t) represents the reset gate at time step t, z(t) represents the update gate at time step t, h(t-1) represents the hidden state at the previous time step, x(t) represents the input at time step t, and W_r and W_z are weight matrices. The weight matrices, W_r and W_z, are multiplied with the concatenation of the previous hidden state (h(t-1)) and the current input (x(t)). This linear transformation allows the model to learn the importance and influence of different elements in the input for the gate calculations. The sigmoid function is applied to the transformed values to squash them between 0 and 1, representing the gate's activation or openness [2]. The reset gate helps the GRU decide how much of the previous hidden state to forget and incorporates new input information. The update gate controls how much of the previous hidden state is passed to the current time step. These gates allow the GRU to adaptively update its internal state based on the input and previous hidden state.

The hidden state of the GRU (h(t)) is computed as a combination of the previous hidden state and the current input, controlled by the reset and update gates given in equation (1) and (2) as in (3),

$$h(t) = [1 - z(t)] \times h(t-1) + z(t) \times h'(t) \quad (3)$$

Here, h'(t) represents the candidate hidden state that combines the reset gate and the current input as given in equation (4),

$$h'(t) = \tanh(W \times [r(t) \times h(t-1), x(t)]) \quad (4)$$

The candidate hidden state represents a potential update to the hidden state, considering the influence of the reset gate and the current input.GRUs may struggle to capture long-term dependencies, meaning they might have difficulty recognizing patterns that occur over extended periods [3]. Additionally, GRUs typically focus on the input sequence and may not consider important contextual information or external factors that can influence user engagement.

They may also face challenges in predicting engagement for new or recently joined users who have limited activity history. Moreover, the complexity of GRUs can make them less interpretable compared to simpler models, making it harder to understand the specific factors driving user engagement. To overcome the limitations the weighted average clustering method, incorporate with GRU are discussed in the literature background [4].

2.LITERATURE BACKGROUND

This section describes the many kinds of infections in widespread computing using a variety of strategies used in current research. The suggested strategy infers the advantages of each technique. The restrictions are removed in order to raise each study's evaluation metric.

This work used gated recurrent units and weighted average clustering to project the understanding ability for performance utilizing deep learning techniques. The neural networks that recur combine multilayer perceptron, which are used to detect network assaults, with gate units to store information in the central memory.

To assess and estimate user involvement with advertising on social media platforms, Seyed Mohsen EbadiJokandan, PeymanBayat, and Mehdi FarrokhbakhtFoumani outline a study that suggests a hybrid convolutional model based on the FCM and XGBoost algorithms. The researchers employ FCM and XGBoost to cluster data according to attribute weight and relevance, and then CNN and LSTM techniques to learn and foresee user engagement rates. The study intends to increase the precision of engagement prediction, stop spam advertising, and lower advertising expenses.

Yap[5]conducted a study by examining users' typical levels of participation on university libraries' Facebook pages to gauge the efficacy of library marketing. Luke and Suharjito [6] examined the efficiency of Twitter product marketing by examining user interaction with promotional tweets. To categorize and calculate the usage percentage of Twitter followers depending on the promoted goods or services, they used the Nave Bayes algorithm.

By examining user engagement rates via comments and likes on Instagram, Bonilla-Quijadaetal. [7] performed research on the effectiveness of urban tourist advertising. They evaluated the efficacy of published advertisements using this data. Furthermore, Zheng et al. assessed user engagement by establishing online communities to promote interaction and discovered that user engagement had immediate effects on brand loyalty.

Kim et al.'s study [8] looked into how various Facebook post kinds affected users' levels of engagement. They discovered that informative articles had the greatest impact on online users, underscoring the significance of participation in the online dissemination of knowledge. Additionally, Stefko et al. [9] Noted that enhancing user involvement through actions like likes, comments, and sharing is essential for the success of promoting posts, highlighting the necessity of user interaction in producing successful promotional results. The correlation with user age and participation levels on various online platforms was investigated by Gasparoni. Their findings indicated that different age groups exhibit distinct preferences and behaviours on different platforms, with older individuals favouring Facebook and younger individuals being more active on Instagram.

By aggregating historical engagement data over specific time intervals, weighted average clustering can capture patterns that occur over longer periods. It can also consider additional factors like user demographics or external events to provide a more comprehensive understanding of user engagement [10]. Furthermore, by leveraging similarities between user clusters, weighted average clustering can provide insights for new users with limited activity history.

The interpretability of clustering results allows for actionable insights to optimize user engagement strategies.To group similar data points into clusters, a similarity threshold or clustering algorithm is used.The assignment of weights to clusters can be based on various criteria as like

- Equal weights: Each cluster is assigned an equal weight, assuming equal importance.
- Proportional weights: The weight of a cluster is determined based on the number of data points it contains or its relative size compared to other clusters.
- Attribute-based weights: The weight of a cluster can be assigned based on a specific attribute, such as the average engagement level of users within the cluster.

Once the clusters and their weights are determined, a weighted average can be calculated based on the desired metric.

$$\text{weighted\_average} = (w_1 * m_1 + w_2 * m_2 + ... + w_n * m_n) / (w_1 + w_2 + ... + w_n) \quad (5)$$

Here, w1, w2, ..., wn are the weights assigned to the clusters, and m_1, m_2, ..., m_n are the corresponding metric values (e.g., engagement) within each cluster.

3. PROPOSED HYBRID GRUWAC METHOD

The proposedhybrid nature of combining Gated Recurrent Units (GRUs) and weighted average clustering methods offers a promising approach by analysing user engagement in online social media. By integrating the strengths of both techniques, this hybrid approach aims to overcome the limitations of individual methods and provide enhanced insights into user engagement dynamics.weighted average clustering methods can be combined with GRUs. Clustering techniques, such as k-means or hierarchical clustering, can group users based on similar engagement patterns or other relevant features. This clustering step allows for the identification of distinct user segments with similar behaviour.

Steps involved in Proposed Algorithm

By combining GRUs and weighted average clustering, the hybrid approach enables the modelling of both short-term dependencies through the GRUs and long-term engagement patterns through the cluster-based aggregation. This integration allows for a more comprehensive analysis of user engagement in online social media, capturing both the sequential dynamics and the broader context of user behaviour.The target engagement metric refers to the specific measure or metric that is the focus of analysis or prediction when assessing user engagement in online social media.

Proposed Algorithm

Input: Prepare user engagement data, representing sequential patterns (e.g., engagement levels over time) and the target engagement metric (e.g., likes, comments).

Step 1: Define the architecture of the GRU model, including the number of hidden units and layers.

Step 2: Train the GRU model using the engagement data, optimizing it to learn the sequential dependencies and capture temporal patterns.

Step 3:Specify the desired number of clusters or use techniques such as silhouette analysis to determine the optimal number of clusters.

Step 4: Apply the clustering algorithm to group similar engagement patterns together, forming clusters.

Step 5: Define criteria for assigning weights to each cluster. This can include the number of users in each cluster, the average engagement levels within each cluster, or other relevant factors.

Step 6: Calculate the weights for each cluster based on the chosen criteria.

Step 7: Examine the weighted average engagement values for each cluster, representing the summarized engagement patterns.

Step 8: Analyse the differences between clusters to gain insights into distinct user engagement behaviours.

Step 9: Interpret the findings to understand the factors that contribute to higher or lower engagement levels in different user segments.

To encode the weighted average clustering data with the contextual information can compute the score uses the separate attention mechanism as given in equation (7). Now $\sigma$is an element-wise sigmoidlogistic function defined as (6),

$$\sigma(x) = 1/(1+e^{-x}); \sigma \in [-1,1] \quad (6)$$

To compute a score distribution over different clusters the proposed method learns the task-specific notion of headedness. The weighted embedding of each span is calculated by equations (7-9).

$$\alpha_t^c = w_\alpha .kNN_\alpha(x_t^*) + b_\alpha^C + g_i \quad (7)$$

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 5, July : 2023

$$a_{i,t}^{c} = \frac{\exp(\alpha_i . x_t^u)}{\sum_{t=i}^{n} \exp(\alpha_i . x_k^u)} \quad (8)$$

$$\breve{x}_i^c = \sum_{t=i}^{n} a_i^n, t(x_t^u + e_i) \quad (9)$$

Where $\breve{x}_i^c$ is a weighted cluster identified with GRU in spanned Cluster Band, g_i be the target with the previous cluster, b_α^C is the bias parameter and e_i is the embedder input cluster.Here, the weights $a_{i,t}^c$ are automatically learned by the proposed algorithm.The decision indicatorat this stage of sorting is considered by using the following equation (10)

$$b_i = p_{j+1}^i - p_j^i \quad (10)$$

Where$b_i$parameter isused to sort the alternatives in each cluster. The features and cluster weightage are ranked more closelyrelated and more similar to the intended mention, and the probabilityof their coreferention is higher. The block diagram of the proposed approach is shown in the Figure 1.0

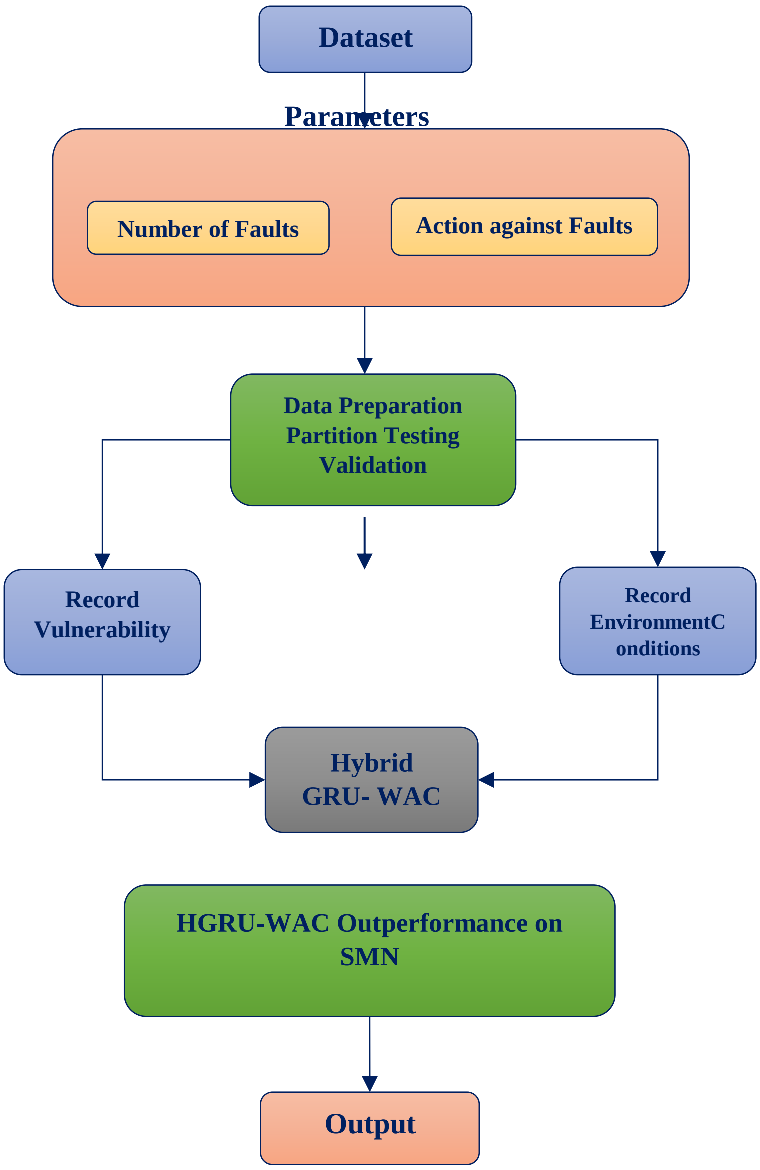


Figure 1.0 Block Diagram of the proposed HGRUWAC Method

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 5, July : 2023

In order to create a multi-trust reflexive aggregation model, two weighted confidence aggregation operators are designed. This prevents secondary uncertainty brought on by subjective weighting when considering multi-node, multi-path chains. A social network trust measurement approach based on uncertainty theory quantifies the confidence degrees of the nodes, the trust chains, and the entire network. In social networks, trust develops during the course of social interaction as users exchange information or resources. At the same time as trust is ingrained in the social network, it also emerges from it.

Once more, using trust, we can spot powerful nodes throughout the network [26,49]. Nodes have the option to determine whether to propagate data to other nodes using the "trust measure" or not after receiving it. Contrarily, a lack of confidence will significantly impede interpersonal connections. Social network platforms can only offer users network growth, a solid platform for cooperation and exchange, more reliable information assistance, effective user attraction, better privacy protection for security, and the promotion of social network consensus with adequate and stable trust support [11].

4. Experimental Results

This section describes the considered courses for the online engagement of users along with the corresponding datasets.compare the LSTM unit, GRU and tanh unit in the task of sequence modelling.a rigorous exploratory data analysis,applied a variety of deep learning algorithms specialized for time series geospatial data andperform a comparative analysis of Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU),Convolutional Neural Network (CNN), and the proposed method.

Data for the experiments are received from the online data sources are obtained and integrated. This is a customer segmentation dataset with 51,000 records, created for beginners and professionals to practice data science and deep learning activities. It includes various customer data points such as name, gender, email, location, job title, and more.

Customer data is defined as the information your customers provide while interacting with your business via your website, mobile applications, surveys, social media, marketing campaigns, and other online and offline avenues. Customer data is a cornerstone of a successful business strategy. Data-driven organizations realize the importance of this and act to ensure that they collect the necessary customer data points that would enable them to improve customer experience and fine-tune business strategy over time.

4.1 Evaluation Metrics

The proposed hybrid approaches enhance the performance of given models and provide more accurate prediction based on the followingmetrics and also applying hybrid feature extraction on speed at a Specific Performance Time, Root Mean Square Error, Mean AbsolutePercentage Error and Accuracy.

$$\text{Specific Performance Time (SPT)} = (C_0 / C_m) * 100 \qquad (11)$$

Here C_0 be the current cluster segment and C_m be the average size of the segment and SPT also provides a normalized expected performance of user segment whichprevents having extreme values.

RMSE is also known asRMSD (root-mean-square deviation). It is the square root of the mean squared differencebetween desired output and predicted output. The same is explained in Equation (12)

$$RMSE = \sqrt{(\sum_i C_{di} - C_{pi})^2 / n} \qquad (12)$$

Where C_{di} be the density of the specific cluster, C_{pi} be the predicted clusters and n be the number of observations.

MAPE (mean-absolute-percentage-error) is robust to large outliers.It eliminates thescaling factor and explains the error in the form of percentage. The formula of MAPE isexplained in Equation (13).

$$MAPE = median(\left|C_{di} - C_{pi}\right| / C_{di}) \qquad (13)$$

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, No. 5, July : 2023
Accuracy was taken as the performancemeasure when counting the number of correct Cluster segment detections. The accuracy functionwas calculated using Equation (14).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

Where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

Dataset

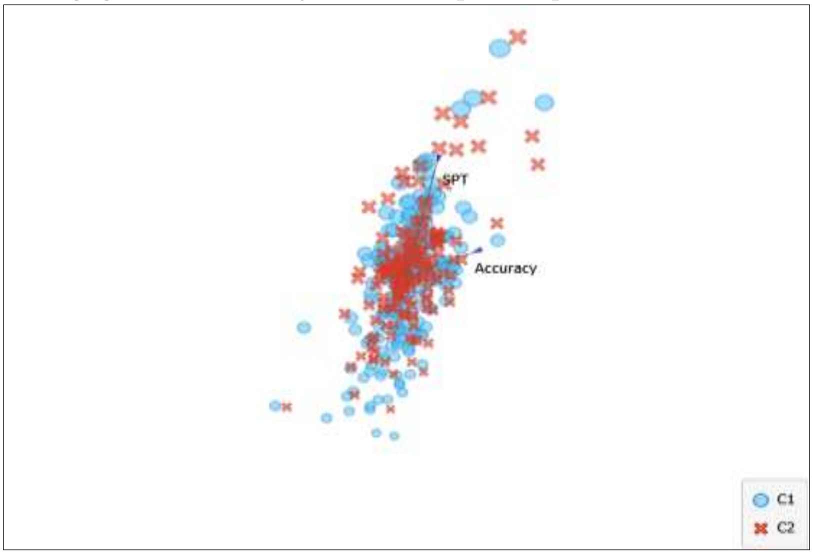
The social graph dataset is a collection network of twitter followers and followers is included in the social graph dataset. The individual who had the most followers at the moment (Lady Gaga) had 10,000 followers randomly selected. Furthermore, the dataset included and considered the social media profiles of these 10,000 individuals. The tweets were also pulled from this dataset.Between January 2010 and October 2010, the tweets were crawled [12].The categories that need to be predicted based on engagement aspects. The label is shown as 1 and the timestamp indicates that the user interacted with the tweet if it is included in those features. The designation is marked as 0, in every other scenario.

The Performance metrics for the proposed methods are compared with the existing LSTM, GRU, CNN are given in the following Table 1.0

Table 1.0. The Performance Metrics Based on Deep Learning Models

Model	**RMSE**	**MAE**	**MAPE**
LSTM	4.86	2.13	6.95
GRU	5.05	2.29	7.7
CNN	30.3	25.96	64.10
Proposed Method	4.6	2.08	6.85

Figure 2.0. The Class engagement accuracy based on specific performance time



The Figure 2.0 mentioned the Specific performance Time for Classified users' engagement on the twitter social network shown.

Conclusion

This paper utilizes the specific performance time (SPT) as the network state evaluation indicator. The factors are act as influencing the engagement frequency in sentiments of the tweet. The proposed hybrid

GRUWAC method performance are better than the existing methods. The proposed approach properlymanages the problem of coreference resolution with the lowesterror rate.

References:

1. Belli, Luca, et al. “Privacy-Preserving Recommender Systems Challenge on Twitter’s Home Timeline.” arXiv preprint arXiv:2004.13715 (2020).
2.Bratsas, C.; Koupidis, K.; Salanova, J.M.; Giannakopoulos, K.; Kaloudis, A.; Aifadopoulou, G. A comparison of machine learning methods for the prediction of traffic speed in urban places. Sustainability 2020, 12, 142.
3. Chen, Y., Zhang, H., Li, P., Li, Q., & Li, B. (2020). Deep Learning for User Engagement Prediction in Social Media Platforms: A Survey. ACM Transactions on Multimedia Computing, Communications, and Applications, 16(1), 1-25.
4. Chung, J., Gulcehre, C., Cho, K., &Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555.
5. Du, J., Xie, X., & Zhang, Y. (2020). Improving User Engagement Prediction with Social Attention and Gated Recurrent Units. Neurocomputing, 410, 1-10.
6. Ghasemzadeh, H., Li, X., & Zhou, X. (2020). A Comprehensive Survey of User Engagement in Social Media. ACM Transactions on Social Computing, 3(1), 1-33.
7. Koshorek, S., & Beck, H. (2018). Identifying User Engagement in Online Social Networks with RNNs. Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM), 565-573.
8. Lan, Y., Wang, Y., & Liang, Y. (2019). Time-Aware User Engagement Prediction in Online Social Networks. Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM), 1365-1374.
9. Li, J., Lu, J., Wu, X., & Yang, H. (2018). User Engagement Modeling in Online Social Media with Recurrent Neural Networks. IEEE Transactions on Multimedia, 20(4), 918-929.
10. Yi, H.; Bui, K.H.N. An automated hyperparameter search-based deep learning model for highway traffic prediction. IEEE Trans. Intell. Transp. Syst. 2020, 22, 5486–5495.
11. Zarezade, A., Gao, J., Ram, S., &Zhai, C. X. (2018). Predicting User Engagement in Online News Comments: An Ensemble Approach. In Proceedings of the 12th ACM Conference on Recommender Systems (pp. 55-63).
12. X. Peng, J. Dai, H. Garg, Exponential operation and aggregation operator for q-rung orthopair fuzzy set and their decision-making method with a new score function, Int. J. Intell. Syst. 33 (2018), 2255–2282.