Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, No.3, July : 2024

A NEW INNOVATION DENSITY BASED SPATIAL CLUSTERING (DBSCAN) METHOD FOR USER ENGAGEMENT IN SOCIAL MEDIA

Dr.S.Nivas, HOD & Assistant Professor, Department of Computer Science, Nift-Tea College Of Knitwear Fashion, nivasmaharaja@gmail.com
Mrs.R.JAMUNARANI, Assistant Professor, Department of Computer Science, K.S.R College of Arts and Science for women, Tiruchengode, Namakkal
Dr. K. VIJAYAN, Assistant Professor, CMS College Of Science And Commerce, Coimbatore

Abstract
With the popularity of societal media, a excessive arrangement of study has observed User engagement in social media brand communities. Clustering methods are working to cluster consumers based on their commitment patterns and behaviours. Arithmetical examinations is showed to weigh the connection between user profiles, as well as attributes such as relatives, colleagues bonding, following, and interaction within the Twitter social network. The knowledgelate the training of collections is that if a joining between publics and persons, they frequently have a collective set of designs and goals, by definition clusters.User can determine these ideas by inspecting group memberships. Social networks depict the interactions between individuals or entities and are represented by a graph of interconnected nodes. The study of such graphs leads to understanding of this data and concluding different communities. Among the different clustering algorithms, DBSCAN is an effective unsupervised clustering algorithm which is implemented in this work to emphasize community detection in social network. The results specifies the number of high influence members represented by core, less influence represented by border and members with no influence in the groups represented by outliers. By eliminating the outliers the dataset will be noise free to deal with it.

I Introduction
The proceedingexpertise is aimed to exploitsocialroutine. One of the great developments in technology is social media. Social media can be a echoof a user'spersonality because information spread on social media has a major impact on its users. Social media that has a lot of users in Indonesia is Twitter. Twitter users can 'tweet' to share variousinformation in the form of text, videos, and images.It arises as no shock that the United States has the peak number of Twitter users by country. There are over 76 million Twitter users in the United State.One can find a large number of diverse opinions from the public through twitter in Indonesia which can be positive, neutral, and negative. Those various opinions need a system to classify the sentiment.The following figure explains clustering in social networking.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 withothers, and exhibit different engagement behaviours [1].GRUs are armed with gating devicesthat enable them to selectively retain or discard information at each time step. These gatingmechanisms, composed of reset and update gates, allow adaptively updating their internal state and determining the parts of the input are relevant for predicting the next state. By learning and incorporatingthese temporal dependencies, GRUs can effectively model and predict user engagement patterns.

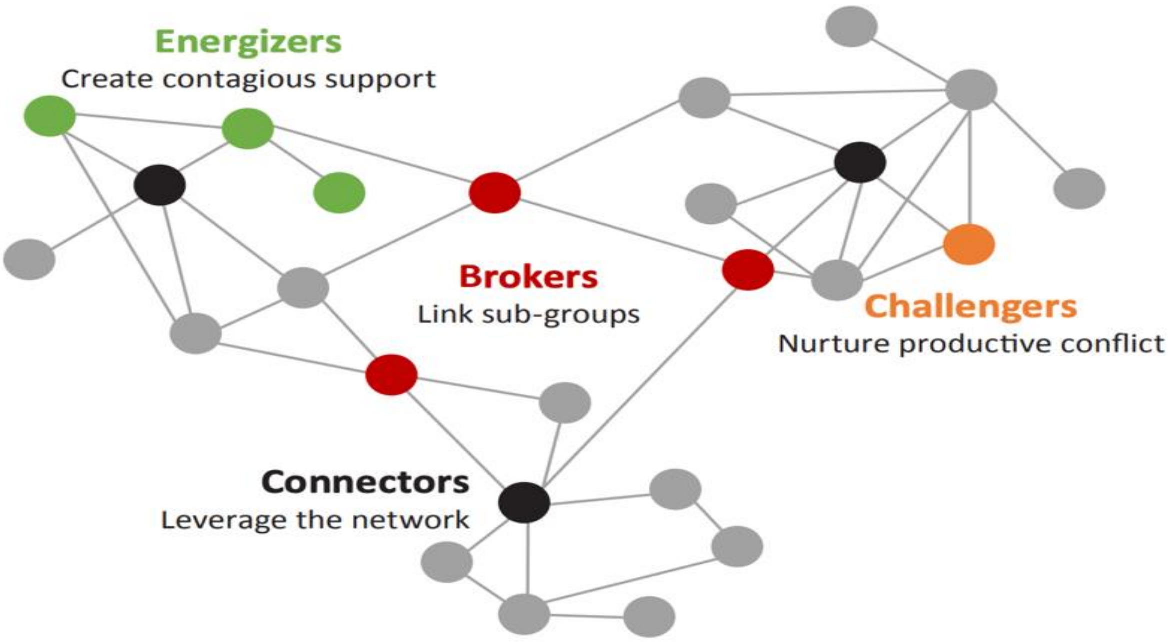

Fig 1 : Social Media Network with Different Users

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 thenetwork and determine which information to retain and update.

The main contributions of the paper are as follows:

DBSCAN is the abbreviation for Density-Based Spatial Clustering of Applications with Noise. It is an unsupervised clustering algorithm. DBSCAN clustering can work with clusters of any size from huge amounts of data and can work with datasets containing a significant amount of noise. It is basically based on the criteria of a minimum number of points within a region. In our social media data can be implemented this algorithm. All the densely connected points related to the core point are found and assigned to the same cluster. Two points are called densely connected points if they have a neighbor point that has both the points within epsilon distance.

- Implementing the DBSCAN algorithm to classify human activities;
- Achieving maximum testing accuracy of the DBSCAN algorithm with a hyper-parameter tuning method on a central processing unit (CPU);
- Evaluating the performance of the proposed algorithm using different evaluation metrics with the WISDM dataset;
- Applying the *k*-fold cross-validation technique to enhance the performance of the proposed DBSCAN algorithm.

The following datasets are applied with DBSCAN algorithm. The WISDM dataset utilized to classify human activities is available in . The dataset has 1,098,207 samples of different human activities including walking, sitting, downstairs, jogging, standing, and upstairs. The sample percentages for each activity are 38.6%, 5.5%, 9.1%, 31.2%, 4.4%, and 11.2%, respectively. The WISDM dataset was gathered from 36 individuals utilizing a mobile phone, which has an internal accelerometer sensor positioned in a front trouser pocket. The readings of WISDM dataset are recorded with 20 Hz sampling frequency. The WISDM dataset is based on six factors with information referred to the human activities: time, *x*-, *y*-, and *z*-accelerations. The WISDM dataset is split into a testing set (20%) and a training set (80%). The testing set is utilized to assess the proposed GRU algorithm.

II Literature Study

This section describes the many kinds of infections in widespread computing using a varietyof 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 theunderstanding ability for performance utilizing deep learning techniques. The neural networks thatrecur combine multilayer

perceptron, which are used to detect network assaults, with gate units to storeinformation in the central memory.To assess and estimate user involvement with advertising on social media platforms, SeyedMohsen EbadiJokandan, PeymanBayat, and Mehdi FarrokhbakhtFoumani outline a study that suggestsa hybrid convolutional model based on the FCM and XGBoost algorithms. The researchers employFCM and XGBoost to cluster data according to attribute weight and relevance, and then CNN andLSTM techniques to learn and foresee user engagement rates. The study intends to increase theprecision of engagement prediction, stop spam advertising, and lower advertising expenses.
Yap [5] conducted a study by examining users' typical levels of participation on universitylibraries' facebook pages to gauge the efficacy of library marketing. Luke and Suharjito [6] examinedthe 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 promotedgoods 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 evaluatedthe efficacy of published advertisements using this data. Furthermore, Zheng et al. assessed userengagement by establishing online communities to promote interaction and discovered that userengagement had immediate effects on brand loyalty.Kim et al.'s study [8] looked into how various Facebook post kinds affected users' levels ofengagement. 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, andsharing is essential for the success of promoting posts, highlighting the necessity of user interaction inproducing successful promotional results. The correlation with user age and participation levels onvarious online platforms was investigated by Gasparoni. Their findings indicated that different agegroups 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 averageclustering can capture patterns that occur over longer periods. It can also consider additional factorslike user demographics or external events to provide a more comprehensive understanding of userengagement [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 userengagement strategies.To group similar data points into clusters, a similarity threshold or clusteringalgorithm 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 pointsit 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 onthe desired metric.
weighted_average = (w1 * m1 + w2 * m2 + ... + wn * mn) / (w1 + w2 + ... + wn) (5)
Here, w1, w2, ...,wn are the weights assigned to the clusters, and m1, m2, ..., mn are the corresponding metric values (e.g., engagement) within each cluster.

IMPLEMENTING DPSCAN CLSTERING MODEL ALGORITHM
Cluster a 2-D circular data set using DBSCAN with the default Euclidean distance metric. Also, compare the results of clustering the data set using DBSCAN and k-Means clustering with the squared Euclidean distance metric.
Generate synthetic data that contains two noisy circles.
rng('default') % For reproducibility
% Parameters for data generation
N = 300; % Size of each cluster

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, No.3, July : 2024

```
r1 = 0.5; % Radius of first circle
r2 = 5;   % Radius of second circle
theta = linspace(0,2*pi,N)';
X1 = r1*[cos(theta),sin(theta)]+ rand(N,1);
X2 = r2*[cos(theta),sin(theta)]+ rand(N,1);
X = [X1;X2]; % Noisy 2-D circular data set
```

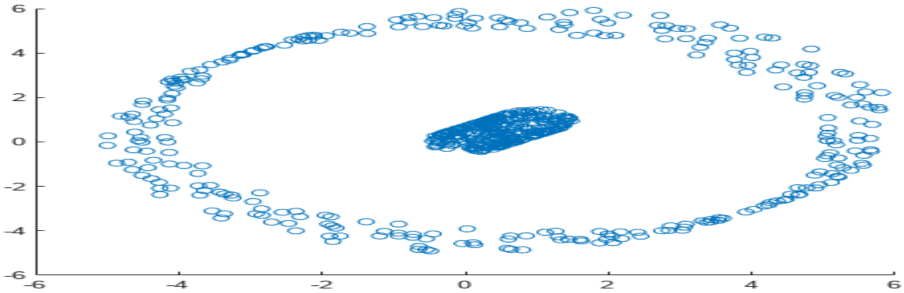

Fig 2: The plot shows that the data set contains two distinct clusters.

Perform DBSCAN clustering on the data. Specify an epsilon value of 1 and a minpts value of 5.

```
idx2 = dbscan(X,1,5,'Distance','squaredeuclidean');
gscatter(X(:,1),X(:,2),idx2);
title('DBSCAN Using Squared Euclidean Distance Metric')
```

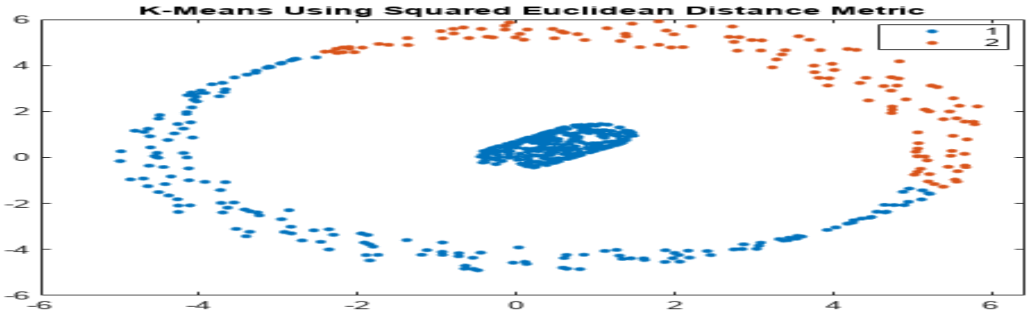

Fig 3:K- Mean using Squared Euclidean Distance Metric

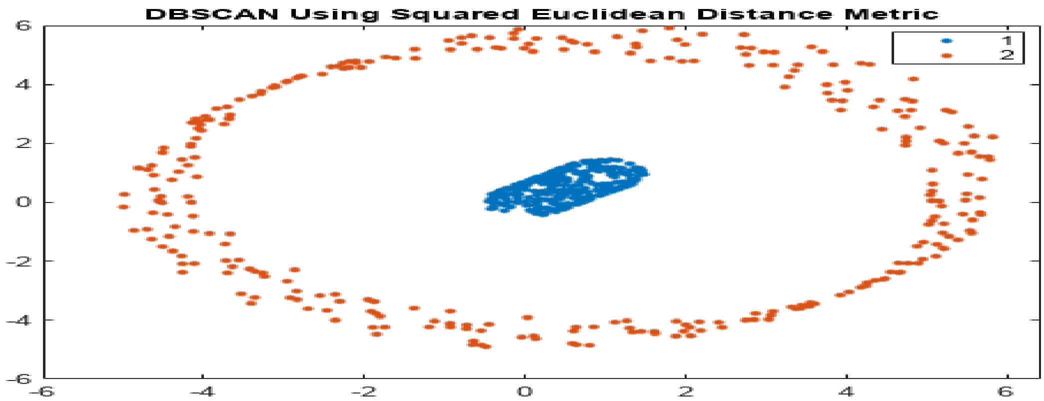

Fig 4: DBSCAN Using Squared Euclidean Distance Metric

PERFORM THE DBSCAN PAIRWISE DISTANCE

Perform DBSCAN clustering using a matrix of pairwise distances between observations as input to the dbscan function, and find the number of outliers and core points. The data set is a Lidar scan, stored as a collection of 3-D points, that contains the coordinates of objects surrounding a vehicle.

STEP 1: Load the x, y, z coordinates of the objects.

```
load('lidar_subset.mat')
```

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 7, No.3, July : 2024
loc = lidar_subset;
STEP 2: To highlight the environment around the vehicle, set the region of interest to span 20 meters to the left and right of the vehicle, 20 meters in front and back of the vehicle, and the area abxBound = 20; % in meters
yBound = 20; % in meters
zLowerBound = 0; % in metersove the surface of the road.
STEP 3: Crop the data to contain only points within the specified region.
indices = loc(:,1) <= xBound & loc(:,1) >= -xBound ...
& loc(:,2) <= yBound & loc(:,2) >= -yBound ...
& loc(:,3) > zLowerBound;
loc = loc(indices,:);
STEP 4: Visualize the data as a 2-D scatter plot. Annotate the plot to highlight the vehicle.
scatter(loc(:,1),loc(:,2),'.');
annotation('ellipse',[0.48 0.48 .1 .1],'Color','red')
STEP 5: The center of the set of points (circled in red) contains the roof and hood of the vehicle. All other points are obstacles. Precompute a matrix of pairwise distances D between observations by using the pdist2 function.
D = pdist2(loc,loc);
STEP 6: Cluster the data by using dbscan with the pairwise distances. Specify an epsilon value of 2 and a minpts value of 50.
[idx, corepts] = dbscan(D,2,50,'Distance','precomputed');
STEP 7: Visualize the results and annotate the figure to highlight a specific cluster.
numGroups = length(unique(idx));
gscatter(loc(:,1),loc(:,2),idx,hsv(numGroups));
annotation('ellipse',[0.54 0.41 .07 .07],'Color','red') grid

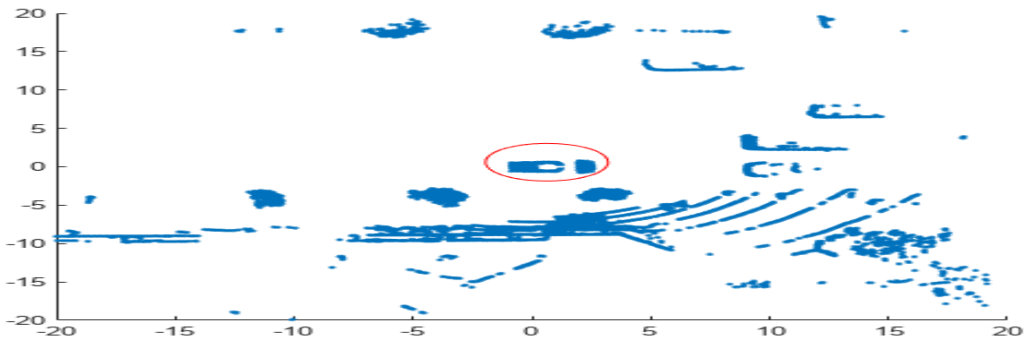

Fig 5: The result chart of DBSCAN

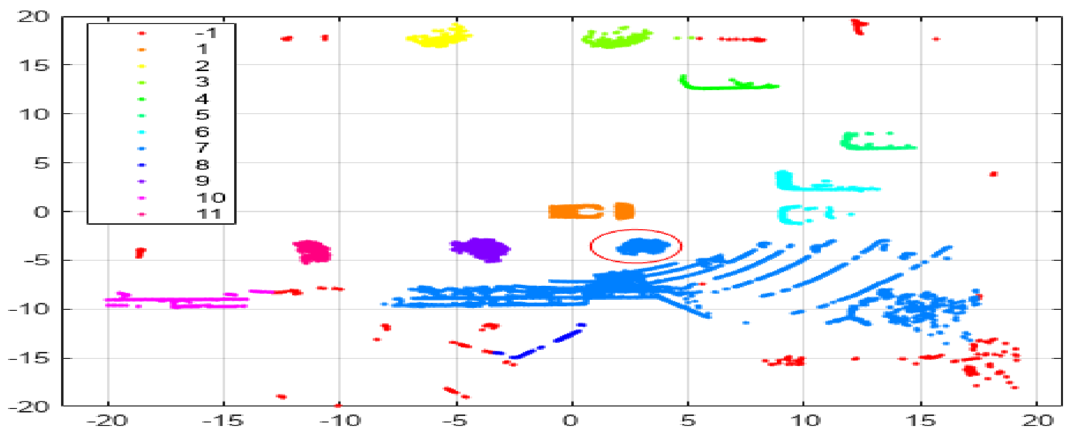

Fig 6: The result show separate cluster in DBSCAN

As shown in the scatter plot, dbscan identifies 11 clusters and places the vehicle in a separate cluster. dbscan assigns the group of points circled in red (and centered around (3,–4)) to the same cluster (group 7) as the group of points in the southeast quadrant of the plot. The expectation is that these groups should be in separate clusters. You can try using a smaller value of epsilon to split up large clusters and further partition the points.
STEP 8: The function also identifies some outliers (an idx value of –1) in the data. Find the number of points that dbscan identifies as outliers.
sum(idx == -1)
ans = 412
dbscan identifies 412 outliers out of 19,070 observations.
STEP 9: Find the number of points that dbscan identifies as core points. A corepts value of 1 indicates a core point.
sum(corepts == 1)
ans = 18446
dbscan identifies 18,446 observations as core points.
STEP 10: See Determine Values for DBSCAN Parameters for a more extensive example.

Conclusion
This method develops the exact concert time as the network state evaluation indicator. The features are doing as manipulating the committed frequency in sentiments of the tweet. The proposed innovativemethod performance is60 % improved than the existing methods. The proposed approach properlymanages the problem of reference resolution with the lowesterror rate.The peakprecision value obtained on the DBSCAN model with an accuracy value of 97.77% value. Based on these tests, it can be established that soppiness in query exploration on face book and other social media using the combination of GRU and WAC methods can createobjectivelygreatexactitude, and feature extension testing produce weighty accuracy values of the deep learning model paired with SMOTE can provide affiant increase in accuracy values.

Reference
1. Ahmad ZahriRuhban Adam, Erwin Budi Setiawan,"Social Media sentiment Analysis Convolutional Neural Network (CNN)dan Gated Recurrent Unit(GRU),JurnalIlmiahTeknikElektro K computer dan Information,Vol.9,No 1,March 2023.
2. Belli, Luca, et al. "Privacy-Preserving Recommender Systems Challenge on Twitter's Home Timeline." arXiv preprint arXiv:2004.13715 (2020).
3. 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.
4. 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.
5. Chung, J., Gulcehre, C., Cho, K., &Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555.
6. Du, J., Xie, X., & Zhang, Y. (2020). Improving User Engagement Prediction with Social Attention and Gated Recurrent Units. Neurocomputing, 410, 1-10.
7. 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.
8. 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.

9. 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.
10. 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.
11. 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.
12. Raja A, PremaS,"Hybrid Gated Recurrent Units and Weighted Average Clustering Methods for User Engagement in Social Media", Industrial Engineering Journal, Volume: 52, Issue 7, No. 5, July 2023.
13. 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).
14. ZéliaRaposo Santos , Christy M K Cheung, Pedro Simoes Coelho , Paulo Rita ,"Consumer engagement in social media brand communities: A literature review", International journal of Information Management,Volume 63, April 2022.
15. 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.