SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I10P112

Volume 11 Issue 10, 114-121, October 2024
© 2024 Seventh Sense Research Group®

Original Article

Artificial Intelligence Enhanced Hybrid Fuzzy Technique in Social Media Mining for Anomaly Detection

Madhan .N[1], Dheva Rajan .S[2], Madhuri Jain[3]

[1, 2]College of Computing and Information Sciences, Mathematics Section,
University of Technology and Applied Sciences Al Musannah, Oman
[3]Department of Mathematics, Banasthali Vidyapith, Rajasthan, India.

[1]Corresponding Author : madhan.narayanan@utas.edu.om

Received: 08 August 2024 Revised: 08 September 2024 Accepted: 10 October 2024 Published: 30 October 2024

Abstract *- The investigation discusses Horizontal Anomalies in social media, emphasizing their multidimensionality and associated hazards. Traditional anomaly detection methods may fail to detect Horizontal Anomalies, demanding more advanced algorithms. Fuzzy logic, which can simulate uncertainty, is used alongside the K-means clustering technique to partition data. The anomaly identification process uses high membership numbers to determine activity level and engagement rate. Line plots illustrate membership values, whereas cluster centroids indicate the clusters. This multidisciplinary method, which includes fuzzy logic, clustering, anomaly detection, and visualization approaches, manages the complexities of Horizontal Anomaly detection, hence improving social media integrity and user security. Flawless identification of horizontal abnormalities in social media mining is crucial for maintaining platform integrity, safeguarding user privacy and security, and countering fraudulent activity. Despite its relevance, horizontal anomaly detection is one of the least researched aspects of social media mining. This study proposes a fuzzy logic-based technique supplemented with K-means clustering to detect horizontal abnormalities accurately and efficiently. The goal is to create a robust system capable of detecting unusual user activity across various dimensions, helping to progress anomaly detection techniques in social media mining.*

Keywords *- Artificial intelligence, Fuzzy, Membership, Anomaly, Engagement, Social media, Clustering.*

1. Introduction

Horizontal Anomaly (HA), or horizontal deviation, is unusual behavior displayed by people across many dimensions or qualities on a social networking site. HAs, instead of vertical anomalies, arise when there are inconsistencies or deviations over many dimensions simultaneously. HA occurs when a user's behavior deviates from expected patterns across many dimensions or features. This work is especially useful for discovering Social Media (SM) anomaly users. Changes in the sentiment conveyed in postings or comments. Changes in the devices used to access the platform and the geographical locations from which the user interacts. HAs are difficult to detect due to their multidimensional character. Traditional anomaly detection algorithms built for univariate or vertical anomalies may not be successful for HA. Analyzing user behavior across numerous dimensions at once necessitates sophisticated algorithms capable of detecting complicated patterns and deviations. HA provides a risk in SM mining and analysis and may signal a variety of concerns.

HA may indicate hostile actors' concerted efforts to distort debates, distribute disinformation, or sway public opinion. Sudden changes in user behavior across several parameters may indicate that unauthorised users have compromised or hijacked an account. Abnormal trends in user interactions or content sharing may raise worries about privacy violations or unauthorized access to sensitive data. Despite its importance, HA detection is one of the least investigated topics in SM mining. The complexities of assessing multidimensional deviations and the dynamic nature of SM platforms present significant difficulties to scholars and practitioners. HA in SM is an unusual behavior that individuals across many dimensions or qualities show simultaneously. Detecting and analyzing these abnormalities is critical for detecting attacks, ensuring platform integrity, and protecting user privacy and security.

1.1. Research Objective

Dealing with HA is the biggest issue in SM platforms. This work proposed a dynamic model instead of the available static models. Hence, it proposes adopting fuzzy logic-based membership functions to model the uncertainty and imprecision in the data. It applies the K-Means Clustering Algorithm (KMCA) to partition the data into clusters based on the combined membership values.

2. Literature

The use of social networks is a fundamental aspect of modern living. With the proliferation of online SM, the availability and use of information have become increasingly vulnerable to various irregularities. Anomalies are the leading source of internet fraud, as the anomalies allow unauthorized individuals to access and forge information. The HA is one type of anomaly that acts as a quiet attacker. These anomalies are generated by a user's inconsistent behavior toward diverse sources. HAs are difficult to identify and harmful to any network. The Web of Science search of the keyword titles 754 documents with “fuzzy” and “social media.”

Including the word “Anomaly” with the above keywords gives only four papers: SM survey techniques and flaws addressed by Batrinca and Treleaven (2015). An insight into various SM analytics platforms is given by Batrinca and Treleaven (2015) [1]. Alsayat and El-Sayed (2016) [2] proposed a comprehensive analysis of SM and proposed the optimal usage of KMCA. Hence, in this work, it is suggested that KMCA be used to detect HA. The major challenge in the SM is to collect the data.

[3] An investigation on Neuro-fuzzy-oriented anomaly detection was conducted by Sharma et al.. However, only five states have been defined in this. The proposed methodology is defined as 10 states. (2018) [5] published an article on traffic information that used social data, not SM data. Oludare Isaac Abiodun et al. (2018) [6] perform mathematical modelling of individual profiling, but it deviates from the current objective.

Yuan et al. (2023) [7] used weighted fuzzy for anomaly detection but not exactly for HA. A wonderful review of the integration of big data analytics and SM has been done by Bazzaz Abkenar et al. (2021) [8]. The Twitter analytics of SM was done by Zachlod et al.. (2022) [10], who proposed their ideas for using SM marketing, in which the researchers insist that buyers be careful when buying. Dahish and Miah determined the data sentiment on SM. Cohen (2022) [12] proposed an algorithmic model using AI, but it was for financial forecasting. Dai et al. (2023) [13] have authored an article on data analytics techniques in SM. In that, he addressed the fact that KMCA is used least.

2.1. Research Gap

Unfortunately, adding the word “horizontal” to the word “anomaly” is obtained only in one article by Sharma et al. (2018) [4]. After changing the word SM to social networking, one can obtain the same results as 1 March 2024. HA is a novel problem and potential threat. Characterized by inconsistent behavior, HA is one of the least researched areas in SM mining. Apart from Sharma et al. (2018) [4], the study concerned finds no study that deals with this problem. Hence, HA is the least researched area in SM. This work addresses such a research gap in addressing HA.

3. Methodology

Unlike Sharma et al. (2018) [4], This research proposes using fuzzy numbers in the following way. In this case, the presence of any node in any network is either true or deceitful; in other words, it must be either ‘true’ or ‘false.’ Here, a Trapezoidal Fuzzy Number (TFN) is proposed. In this way, the membership function becomes trapezoidal, as given in Figure 1.

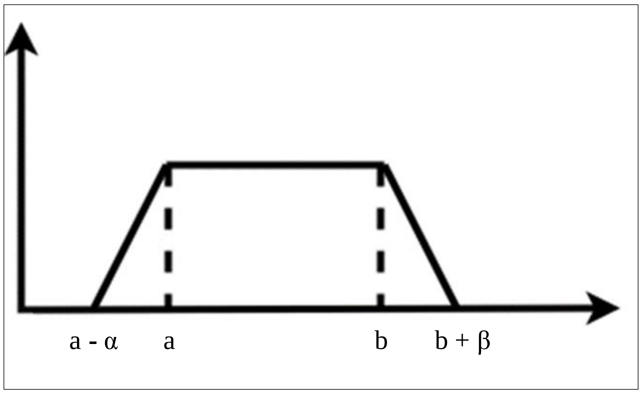

Fig. 1 Trapezoidal fuzzy membership function

Based on Figure 1, the following can be deduced.

$$A(t) = \begin{cases} L\left(\frac{\alpha-t}{\alpha}\right), \; if\ t \in [a-\alpha, a] \\ 1, \; if\ t \in [a, b] \\ R\left(\frac{t-b}{\beta}\right), \; if\ t \in [b, b+\beta] \\ 0, \; Otherwise \end{cases} \qquad (1)$$

Figure 2 shows the Fuzzy Membership Function (FMF) categories for the levels considered.

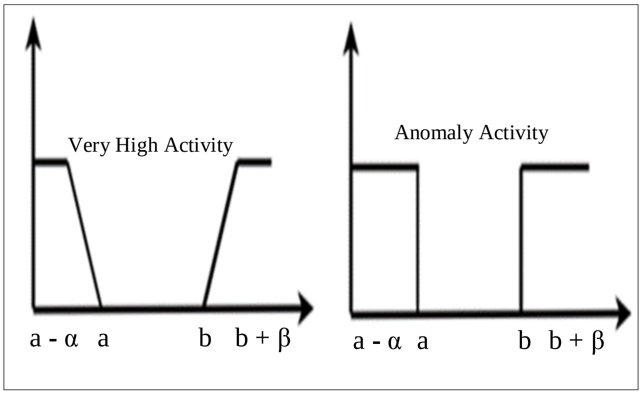

Fig. 2 Fuzzy membership function

By further usage of Zadeh (1975) [14], (1978a) [15], (1996) [17], and Carlsson and Fullér (2002) [18], approximate reasoning is also advocated to determine the membership function in the face of uncertain and imprecise data. Network data is immensely ‘big,’ and any information on them consists of large sets; hence, the analysis must be precise. The advantages of fuzzy in anomaly detection are Overseeing Uncertainty, Expressiveness, Interpretability, Robustness to Noise, Combining Multiple Sources of Information, Adaptability and Learning.

Hence, it is proposed that FMF be used for outlier detection. Proving the effectiveness of a fuzzy inference system for HA detection mathematically is inherently challenging due to the subjective and ambiguous nature of FMF. Figure 3 gives the flow of the fuzzy analysis flow chart in anomaly detection.

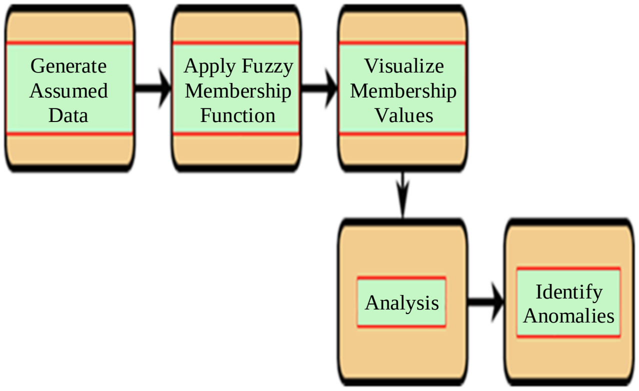

Fig. 3 Fuzzy analysis

In this investigation, activity and Engagement Rates (ER) play a crucial role in the computation of anomalies. The matrix definition of engagement involves the number of likes, comments, shares, clicks and impressions obtained for a particular post. The total engagement of a user is the sum of all those activities. Hence, a user post with 500 likes, 100 comments, 50 shares, and 25 clicks yields a total engagement 675.

The ER Sehl (2023) [19] is the contribution of every impression to the total engagement. The average total ER of a user can be obtained by taking the average across all the posts. Such computations drive us towards content performance, professional corrections, content improvement, visibility and teaching, and data-driven decisions. Of the references quoted, none have done a comparative analysis of all the parameters mentioned below. This paper addresses the gap in analyzing all the below-mentioned parameters.

3.1. ER by Reach (ERR)

This formula is the predominant method for calculating interaction with SM. ERR quantifies the proportion of those who opted to engage with the content after its exposure. Utilize the initial process for an individual post, then employ the subsequent one to determine the average rate across multiple posts.

$$ERR = \frac{Total\ number\ of\ engagements\ per\ post}{Total\ reach\ per\ post} * 100 \quad (2)$$

To calculate the average, sum all the ERRs from the selected posts and divide by the number of posts.

$$Average\ ERR = \frac{Total\ ERR}{Total\ posts} \quad (3)$$

Reach may serve as a more precise metric than follower count, as not all followers will engage with every piece of content. Individuals who do not follow may have encountered the posts via shares, hashtags, and other methods. Reach can vary for numerous reasons, rendering it a challenging variable to regulate. A minimal reach may result in an excessively high ER; conversely, it is essential to consider this factor.

3.2. ER by Posts (ER Post)

This formula measures engagements by followers on a specific post. In other words, it is like ERR, except instead of reaching, it tells the rate at which followers engage with the content. Most SM influencers calculate their average ER this way.

$$ER\ post = \frac{Total\ engagements\ on\ a\ post}{Total\ followers} * 100 \quad (4)$$

To get the average, sum all the ER posts intended for averaging and divide by the total number of posts.

$$Average\ ER\ by\ post = \frac{Total\ ER\ by\ post}{Total\ posts} \quad (5)$$

Although ERR provides a superior method for assessing interactions based on post visibility, this formula substitutes reach with followers, which is typically a more consistent indicator. If an individual's post reach varies frequently, employ this strategy for a more precise engagement assessment post-by-post basis.

Although this method offers a more consistent means of monitoring post engagements, it fails to capture the complete picture, as it neglects viral reach; also, an increase in follower count may lead to a small decline in ERs.

3.3. ER by Impressions (ER Impressions)

$$ER\ impressions = \frac{Total\ engagements\ on\ a\ post}{Total\ impressions} * 100 \quad (6)$$

$$Average\ ER\ impressions = \frac{Total\ ER\ impressions}{Total\ posts} \quad (7)$$

This strategy is advantageous for people managing paid content who require evaluation of effectiveness based on impressions. An ER formula based on impressions will invariably produce lower results than the ER Ratio and post-ERs (ER). Like reach, impression measurements may also demonstrate variability. Employing this method alongside reach may prove beneficial.

3.4. Daily ER

$$Daily\ ER\ = \frac{Total\ engagements\ in\ a\ day}{Total\ followers} * 100 \quad (8)$$

$$Average\ Daily\ ER\ = \frac{Total\ engagements\ for\ X\ days}{(X\ days * followers)} * 100 \quad (9)$$

This method effectively measures the daily interaction frequency of followers with the account rather than their engagement with individual posts. This is suitable for long-term analysis. Consequently, it incorporates interactions on both new and existing posts. This recipe can be customized for particular applications. If the brand's objective is to quantify daily comments, one might modify "total engagements" accordingly.

This strategy allows for a considerable margin of error. The calculation fails to consider that a single follower may engage ten times in one day instead of ten followers participating once each. Daily engagements may fluctuate for assorted reasons, including the number of posts shared. Consequently, it may be beneficial to graph daily participation against many posts.

3.5. ER by Views (ER views)

This ER views technique works well for video reach computation.

$$ER\ view\ = \frac{Total\ engagements\ on\ video\ post}{Total\ video\ views} * 100 \quad (10)$$

$$Average\ ER\ view\ = \frac{Total\ ER\ view}{Total\ posts} \quad (11)$$

ER views are an effective tracking method for creating engagement data for the uploaded video. Views frequently encompass several views from an individual user (non-unique views). Although the viewer may watch the movie repeatedly, this does not guarantee further engagement.

3.6. Cost per Engagement (CPE)

This CPE technique works well, especially for influence marketers.

$$CPE\ = \frac{Total\ amount\ spent}{Total\ engagements} \quad (12)$$

Most SM advertising platforms perform this computation to evaluate the engagement of various influencers alongside other object-oriented metrics, such as cost-per-click. Verify which interactions qualify as engagements to ensure precise comparisons. Furthermore, shares, saves, video views, video reach, and link clicks can all be considered metrics. The user’s profile matrices include follower growth over time, negative feedback rate, profile visits, reactions, and total ER inside profiles. There are a few SM analytics calculators available, such as Hootsuit [20].

A hybrid method that combines multiple engagement metrics, such as ERR, ER posts, ER impressions, Daily ER, and ER views, can provide a comprehensive understanding of user engagement across different dimensions. One hybrid method could involve weighted averaging of these metrics to assign importance based on their relevance and impact.

3.6.1. Weighted Averaging Method (WAM)

$$WAM\ = w_1 \times ERR\ + w_2 \times ER_{post} + w_3 \times ER_{impressions} + w_4 \times Daily\ ER + w_5 \times ER_{views} \quad (13)$$

Where, w_1, w_2, w_3, w_4, w_5 are weights assigned to each metric, representing their relative importance. ERR, ER posts, ER impressions, Daily ER, and ER views are the respective engagement metrics. The weights w_i can be determined based on factors such as the analysis's objectives, the nature of the content, and the platform's characteristics. For example, if maximizing reach is a priority, ERR and ER impressions may be given higher weights.

4. Analysis

The steps outlined in Table 1 in the proposed approach provide evidence that the Weighted Averaging Method (WAM) performs well compared to individual metrics and other existing methods. We will assume hypothetical data, assign weights, perform calculations, and evaluate the results for this demonstration.

Table 1. Anomaly detection score based on activity and ERs

S. No.	Activity Level	Engagement Rate	Anomaly Score
1	Very High	Very Low	High
2	High	Low	Moderate
3	Moderate	Moderate	Low
4	Low	High	Suspicious

It is obtained from a dataset with 10 users, where each user has corresponding values for activity_level and engagement_rate. The 10 chosen users have their accounts on Instagram, Facebook, and X Corp (formerly Twitter). Out of these 10 users, 9 are real, and 1 bot user (named user 9) is created and mixed with the existing user.

The investigation is to find the anomaly user perfectly. FMFs were defined for activity_level, partitioning it into various fuzzy sets like Very Low Activity (VLA), Low Activity (LA), Moderate Activity (MA), etc. Each user's activity level was evaluated using these membership functions, resulting in membership values for each fuzzy set. The membership values for each user were printed for different fuzzy sets (VLA, LA, MA, etc.).

4.1. K-Means Clustering (KMCA)

KMCA belongs to the unsupervised category. Alsayat and El-Sayed (2016) [2] This is used to find the cluster within the obtained data. For instance, for an obtained set of observations $y_1, y_2, y_3, \dots, y_n$, and each observation is a vector in the same dimension. The objective of KMCA is to partition the given 'n' number of observations into K-clusters, such as $U = U_1, U_2, U_3, \dots, U_K$, so that the within-cluster sum of squares can be minimized. Let c_i be the centroid points of S_i. The objective function of K-Means is given by,

$$\min_{S} \sum_{i=1}^{k} \sum_{x\in S_i} \left|\left|x_i - c_i\right|\right|^2 \qquad (14)$$

Here, KMCA is performed on the combined membership values obtained from all fuzzy sets. The K parameter was set to 5, indicating that the objective is to identify 5 clusters. The cluster centers represent each cluster's mean value of combined membership values. The output provides the cluster centers, which are the average combined membership values for each cluster:

Cluster 1: 1.000, Cluster 2: 1.500, Cluster 3: 0.950, Cluster 4: 0.820, and Cluster 5: 2.065. The activity levels and ERs of 10 users have been considered below.

activity_level = 5.62, 2.36, 7.89, 3.45, 8.21, 1.78, 6.54, 4.32, 9.87, 2.10 and engagement_rate = 3.21, 7.65, 2.98, 8.76, 1.23, 6.54, 4.32, 9.87, 2.10, 7.89, respectively. Figure 4 shows the membership values of various activities of selected users.

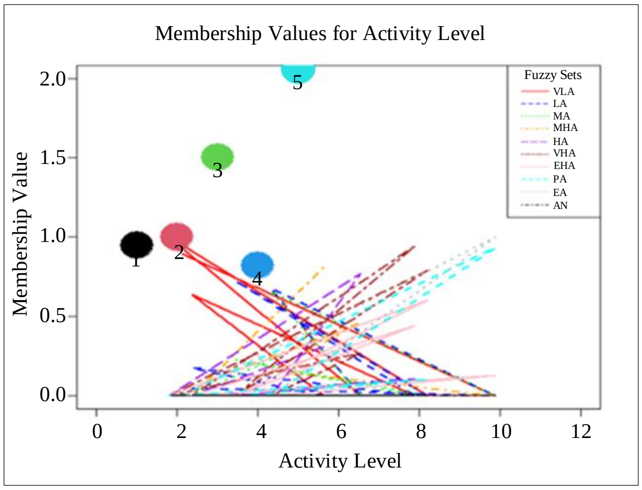

Fig. 4 Membership values of various activities

The success of this approach relies heavily on the chosen features, the defined fuzzy sets and rules, and the chosen defuzzification method. Experimentation, validation, and comparison with other methods are crucial for optimizing the fuzzy inference system for practical use in HA detection. Figure 5 compares Row Sums (RS), Weighted Sum (WS), product, and minimum and maximum methods.

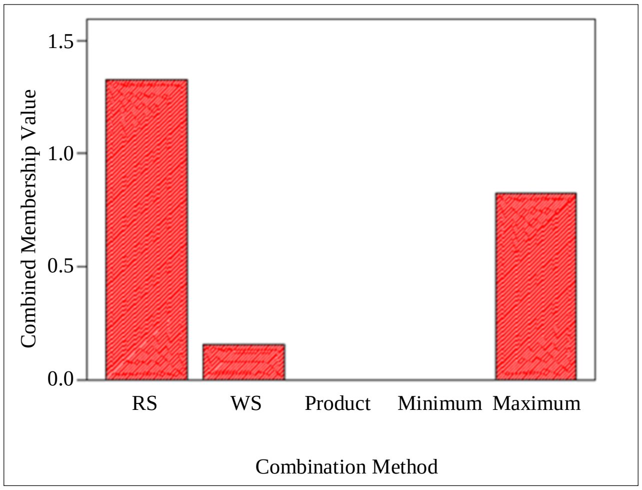

Fig. 5 Comparison of various methods

Row sums are the most suitable for the proposed technique when comparing various methods like row sums, weighted sums, and product, minimum, and maximum methods. Weighted sums have the least value, and if such a value is chosen, it becomes redundant as it was already assigned fuzzy values. Since the zero fuzzy values were transferred to the least bothered, the minimum value always becomes 0. A similar argument applied for maximum also. If the product method is used, then due to zero values, the product always gives zero values. Hence, the most appropriate method is row sums. Figure 6 shows a plot of histograms for ERR and ER_post, followed by ER_impressions and Daily_ER comparison in Figure 7 and ER_Views and WAM in Figure 8.

Figures 7 and 8 show the histogram distribution of values of ER_Impressions, Daily_ER, ER_Views and WAM, respectively, which gives us an equal contribution of the said parameters in the assessment. ER_Impressions and ER_View histogram are filled with bars, whereas in daily_ER, the frequencies are less. These deviations have a significant impact, particularly the deviation in WAM.

The linear regression test p-value summary is greater than 0.05 for Daily_ER, and there is no significant difference between Daily ER. The linear regression test p-value summary is less than 0.05 for ER_views, with a substantial difference in the ER_views. On the other hand, the p-value is 0.0469, equal to 0.05. Hence, it is on the borderline of non-rejection of the null hypothesis. The entire ANOVA analysis of the content provides a p-value of 0.477, indicating no significant differences between the observed values. Figure 9 shows the engagement of a user's activity. This graph shows the highest activeness on Jan 09 and Jan 31. However, I found nothing in seeking the posts and shares; only the activity went to the peak. That means the user is so active in commenting on other posts and other usual activities. This leads to suspect that the user.

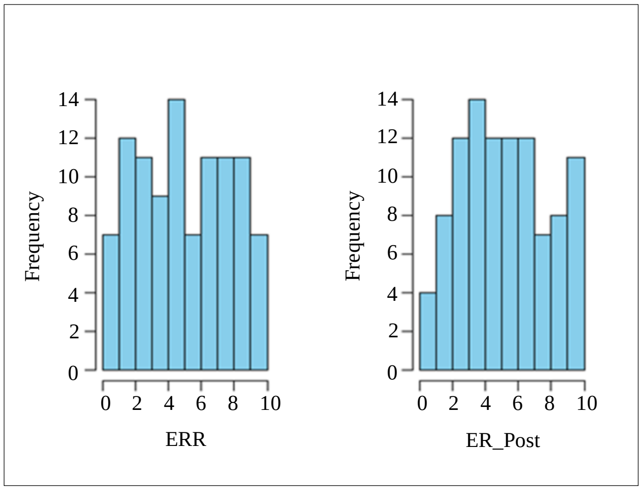

Fig. 6 Histogram of ERR and ER_post activities

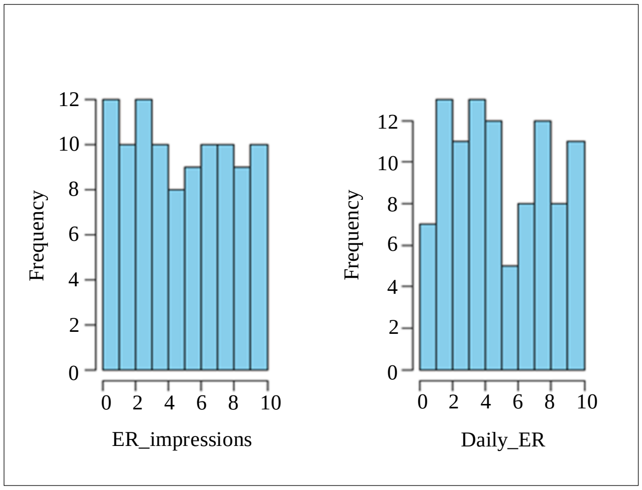

Fig. 7 Histogram of ER_impressions and daily_ER activities

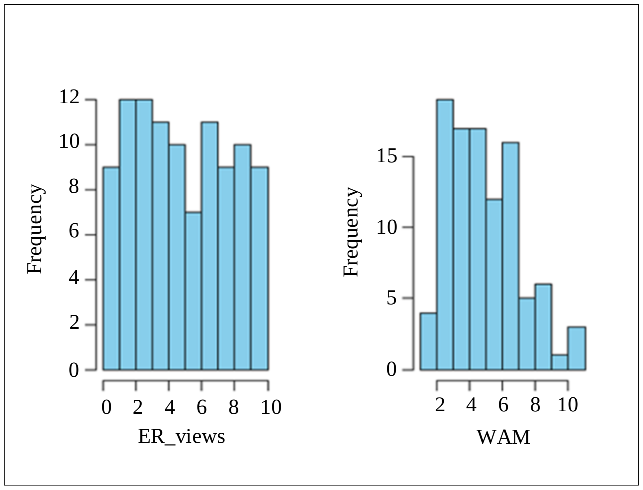

Fig. 8 Histogram of ER_views and WAM activities

After identifying with the user ID, it is a user with ID user9. When finding the correlation coefficients, it possesses a weak positive correlation, with correlation coefficients between 0.1 – 0.25 for all the parameters. Hence, these statistical measures alone cannot take us through anomaly detection; they are performed as outlier detection.

The outlier’s values of each parameter have been found, and the results are given below:

```
$ERR
[1] 20 62 87

$ER_post
[1]  7 93

$ER_impressions
[1]  2 22 72 77 87 92 97

$Daily_ER
[1]   2  27  40  47  52  80 100

$ER_views
[1] 17 92
```

These outlier value users are under investigation. However, with these values alone, one cannot determine the anomaly. From these values, outliers were chosen and moved to further analysis of identification of anomaly detection using the clustering technique.

The KMCA algorithm grouped the users into 5 clusters based on their combined membership values across different fuzzy sets. Figure 10 shows the fuzzy clustering of engagement matrices.

Each cluster center represents the average combined membership value for users in that cluster. Although Figure 4 shows the clustering positions of membership values, Figure 10 shows the clustering, particularly for ER_post. Similar analysis can be performed for other measures to understand anomalies better.

For instance, Cluster 1 has a center of 1.000, indicating that users in this cluster have an average combined membership value of 1.000 across all fuzzy sets. This clustering helps identify similar patterns among users based on their fuzzy membership values, providing insights into different user segments. Two users, user 5 and user 9, have membership values of activity level and ER.

The anomaly identification is only based on activity_level>8 and engagement_rate<3.

Table 2 shows the user activity level, ER, and anomaly detection data. Figure 11 shows the users' comparison of various activity levels.

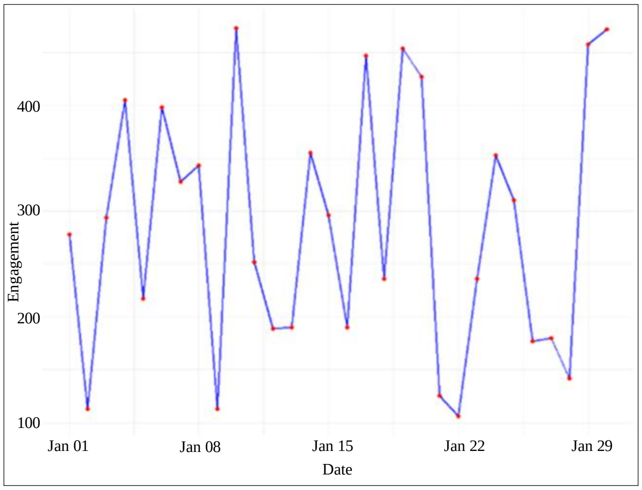

Fig. 9 Engagement graph by user

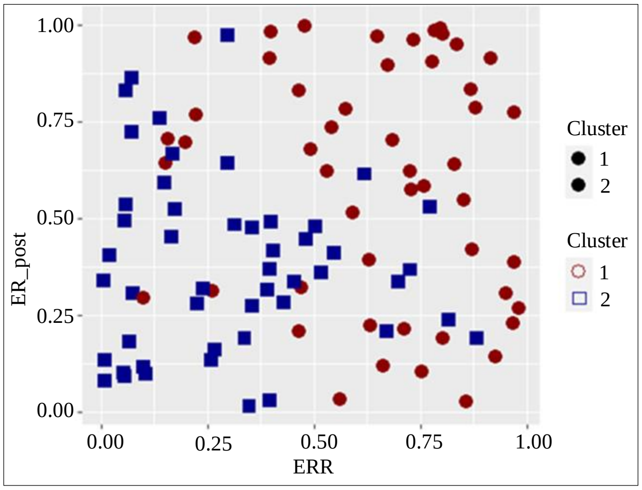

Fig. 10 Fuzzy clustering of engagement matrices

Table 2. Activity level data

User	Activity Level	Engagement Rate	Anomaly
User 1	5.62	3.21	No
User 2	2.36	7.65	No
User 3	7.89	2.98	No
User 4	3.45	8.76	No
User 5	8.21	1.23	No
User 6	1.78	6.54	No
User 7	6.54	4.32	No
User 8	4.32	9.87	No
User 9	9.87	2.1	Yes
User 10	2.1	9.99	No

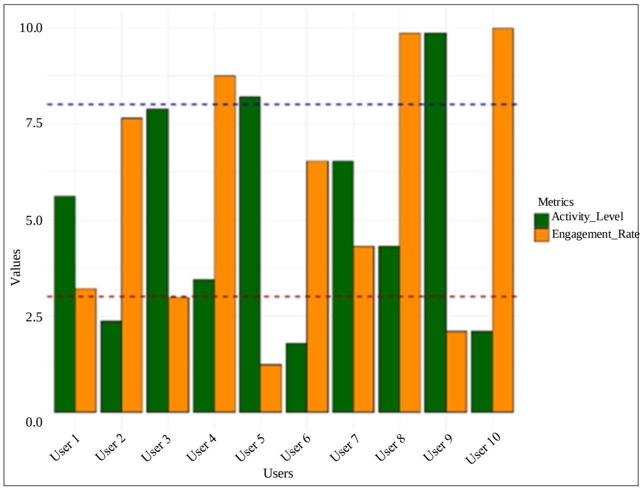

Fig. 11 User comparison of engagement matrices

```
> print(anomalies)
user activity_level engagement_rate combined_membership

5 user5       8.21        1.23          1.500
9 user9       9.87        2.10          2.065
```

The anomaly user is found when incorporating the third condition as a combined activity level, as given below.

```
> cat("Anomalies:\n")

Anomalies:
> print(anomalies)
user activity_level engagement_rate combined_membership

9 user9       9.87        2.1          2.065
```

Hence, user 9 can be considered an anomaly, and further investigation must be performed on this user.

5. Conclusion

The proposed investigation addresses the HA problem in SM, focusing on detecting anomaly user behavior across multiple dimensions. Because of the various dimensions of anomalies, it becomes challenging to detect anomalies in SM, making the detection methods less effective. Developing and proving the effectiveness of a fuzzy inference system for HA detection requires a comprehensive approach. While conclusive mathematical proof is not feasible due to the inherent non-determinism of fuzzy, one can analyze the system's components, simulate its behavior, and optimize its parameters to demonstrate its potential for accurate HA detection in the specific context.

The proposed method employs a mathematical model that uses the FMF and the KMCA algorithm to tackle this issue. The proposed method integrates anomaly detection, clustering, and data visualization techniques to analyze and derive insights from SM. The analysis identifies the anomaly user perfectly, even though it has not identified user 5 (with less score). The feedback obtained from the user is not considered a parameter in the proposed work, which becomes the limitation of the current study. The TFN function is used; future researchers may use more enhanced methods. The model can be enhanced with other combination functions. As a future development, advanced fuzzy logic models and membership functions can be utilized to improve the precision and interpretability of anomaly detection results.

The current method uses the static method, which may be used in future research on dynamic behavior in HA detection. The integration of deep learning with cross-platform for HA can be considered.

Acknowledgements

Sincere acknowledgements for all the co-authors.

References

[1] Bogdan Batrinca, and Philip C. Treleaven, “Social Media Analytics: A Survey of Techniques, Tools and Platforms,” *AI & Society*, vol. 30, pp. 89-116, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ahmed Alsayat, and Hoda El-Sayed, “Social Media Analysis Using Optimized K-Means Clustering,” *IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, USA, pp. 61-66, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Stefan Stieglitz et al., “Social Media Analytics - Challenges in Topic Discovery, Data Collection, and Data Preparation,” *International Journal of Information Management*, vol. 39, pp. 156-168, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Vishal Sharma et al., “NHAD: Neuro-Fuzzy Based Horizontal Anomaly Detection in Online Social Networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 11, pp. 2171-2184, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Zhihao Zheng et al., “Framework for Fusing Traffic Information from Social and Physical Transportation Data,” *PLOS ONE*, vol. 13, no. 8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Oludare Isaac Abiodun et al., “Terrorism Prevention: A Mathematical Model for Assessing Individuals with Profiling,” *International Journal of Computer Science and Network Security*, vol. 18, no. 7, pp. 117-127, 2018. [Google Scholar]

[7] Zhong Yuan et al., “Anomaly Detection Based on Weighted Fuzzy-Rough Density,” *Applied Soft Computing*, vol. 134, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sepideh Bazzaz Abkenar et al., “Big Data Analytics Meets Social Media: A Systematic Review of Techniques, Open Issues, and Future Directions,” *Telematics and Informatics*, vol. 57, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Cécile Zachlod et al., “Analytics of Social Media Data - State of Characteristics and Application,” *Journal of Business Research*, vol. 144, pp. 1064-1076, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Teissir Benslama, and Rim Jallouli, “Social Media Data Analytics for Marketing Strategies: The Path from Data to Value,” *Journal of Telecommunications and the Digital Economy*, vol. 10, no. 2, pp. 96-110, 2022. [Google Scholar] [Publisher Link]

[11] Zahra Dahish, and Shah J. Miah, “Exploring Sentiment Analysis Research: A Social Media Data Perspective,” *International Journal of Soft Computing*, vol. 14, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Gil Cohen, “Algorithmic Trading and Financial Forecasting Using Advanced Artificial Intelligence Methodologies,” *Mathematics*, vol. 10, no. 18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Lingxi Dai et al., “Data Analytics in the Social Media Industry,” *Lecture Notes in Education Psychology and Public Media*, vol. 3, no. 1, pp. 856-865, 2023. [Publisher Link]

[14] L.A. Zadeh, “The Concept of a Linguistic Variable and its Application to Approximate Reasoning-III,” *Information Sciences*, vol. 9, no. 1, pp. 43-80, 1975. [CrossRef] [Google Scholar] [Publisher Link]

[15] L.A. Zadeh, “Fuzzy Sets as a Basis for a Theory of Possibility,” *Fuzzy Sets Systems*, vol. 1, no. 1, pp. 3-28, 1978. [CrossRef] [Google Scholar] [Publisher Link]

[16] L.A. Zadeh, “PRUF-A Meaning Representation Language for Natural Languages,” *International Journal of Man-Machine Studies*, vol. 10, no. 4, pp. 395-460, 1978. [CrossRef] [Google Scholar] [Publisher Link]

[17] L.A. Zadeh, “Fuzzy Sets and Information Granularity,” *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems*, pp. 433-448, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[18] Christer Carlsson, and Robert Fullér, *Fuzzy Reasoning in Decision Making and Optimization*, Physica Heidelberg Publisher, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[19] Katie Sehl, and Shannon Tien, Engagement Rate Calculator + Guide for 2023, Strategy, Blog, 2023. [Online]. Available: https://www.indikit.net/document/371-engagement-rate-calculator

[20] Understand What's Working on Your Social with Hootsuite Analytics, Hootsuite, 2024. [Online]. Available: https://www.hootsuite.com/platform/analytics