Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/130606

Binomial Regressive Influence Behavior Ranking for Virtual Community Formation in Social Network

R. Gnanakumari[1,*] *and P. Vijayalakshmi*[2]

[1] Department of CSE, Nehru Institute of Engineering and Technology, Coimbatore, India.
[2] Dept. of ECE, Hindusthan College of Engineering and Technology, Coimbatore, India

Received: 2 Mar. 2019, Revised: 10 May 2019, Accepted: 13 May 2019
Published online: 1 Nov. 2019

Abstract: A Social Network (SN) is a website which permits people to share the data about their personal or business endeavor to form the virtual community. Due to communication between the users in the SN, a similar user's behavior identification causes a fundamental issue. The existing techniques still encounter the problem of identifying similar behavior accurately. Therefore, Statistic Dice Similarity Based Probabilistic Binomial Regression and Ranking (SDS-PBRR) method introduced. First, the similarity value between the user's behaviors is calculated using Statistic Dice Similarity Coefficient (SDSC). Second, Probabilistic Binomial Regression Analysis is carried out to evaluate the similarity value and to classify the users as Influencing Behavior (IB) or Non-Influencing Behavior (N-IB) minimum error rate in the SN. Last, firefly algorithm is applied to perform a ranking process for discovering the level of IB users in the SN. The simulation results show that SDS-PBRR method increases the True Positive Rate (TPR) and minimizes the False Positive Rate (FPR) as well as execution time.

Keywords: Social Network, Virtual Community, Statistic Dice Similarity Coefficient, Probabilistic Binomial Regression, User Behavior

1 Introduction

SN offers the web-based applications to enable users to create a profile and communicate with others on the network. There are different SN sites such as Facebook, Twitter, and Instagram, etc. In SN, influencer identification has become an issue for virtual community formation. Influencers in a SN are the members that have a more significant effect on the online SN than the average member. A Parallel Self-organizing Overlapping Community Detection (PSOCD) was introduced [1] using swarm intelligence in the extensive online SN. However, the error rate, as well as the time consumption, was not minimized. A Social Action-Based Influence Maximization (SAIM) model was introduced [2] for influence maximization in SNs. The accurate computation of influential users was not performed with less error.

A novel and effective centralized key management protocol was introduced [3] to provide a secure communication service between the users in the online SN. The mechanism was efficient for identifying malicious users in the online SN. However, the accurate detection of malicious users was not achieved. A new trust-based community detection algorithm was presented in [4] to broadly influence the trust among the users and mine the communities within the networks—the error rate was not minimized in the trust-based community detection.

A network representation learning technique was developed in [5] to discover the influential users for controlling the rumors. The method does not accurately discover influential users in the SN. Social influence modeling using information entropy was presented in [6] to detect the most influential users among mobile users. The approach failed to rapidly and efficiently indentify influential users in the mobile SN.

A Degree-Descending Search Evolution (DDSE) algorithm introduced for influence maximization with minimum time consumption [7]. The DDSE algorithm failed to enhance the performance of maximization influence.

The Dynamically Socialized User Networking (DSUN) model is presented in [8] for discovering influence-based communities based on implicit and

* Corresponding author e-mail: mailtokumari81@gmail.com

© 2019 NSP
Natural Sciences Publishing Cor.

explicit user correlations. It failed to perform in-depth analyses for discovering the influence-based communities. Susceptible View Forward Removed (SVFR) model was developed to describe the dynamic user activities in [9]. It did not perform the dynamic user behaviors ranking to detect the most IB of the users. A community-related influence evaluation model with a dynamic perspective (D-CIEM) developed in [10]. It optimizes a community ranking in SNs, but the execution time not minimized.

The above-mentioned pieces of literature reveal major issues, including high error rate, long execution time, failure to discover the influential users in the mobile SN etc. To resolve these issues, an efficient Statistic Dice Similarity-based Probabilistic Binomial Regression and Ranking (SDS-PBRR) Method introduced.

1.1 Significant Contribution of the SDS-PBRR Method

SDS-PBRR method improves the TPR and minimizes the execution time in the user classification. This contribution is achieved using similarity-based Probabilistic Binomial Regression Function (PBRF). It also computes the similarity between the behaviors of the users based on mutual dependence and independence.

PBRF classifies the users as IB users or N-IB users based on the similarity value. If the estimated regression value is higher than the threshold value, they are classified as IB users. Otherwise, they are classified as N-IB users. It helps to improve the TPR and reduce the FPR.

Firefly algorithm is applied to rank the IB user and thereby identify the first level of IB user based on the objective function. Initially, firefly optimization contains several IB users to create an initial population. The light intensity of each firefly is measured based on objective functions. The firefly with higher light intensity moves towards the other. Finally, the fireflies are ranked to identify the level of IB users with minimum execution time.

A divide-and-conquer strategy with parallel computing mechanism is developed [11] to discover the top-K influential users with the community-based Greedy algorithm. It failed to discover influential users over time. Influence Diffusion Model (IDM) is introduced in [12] to identify the influencers based on the context relationship. The IDM used Page Rank model to rank the influence users. However, the similarity between the user's different activities is not computed.

A preference analysis and random walk framework were introduced [13] to identify the user Influences in an SN with high precision. The framework failed to detect the influence of users in a dynamic SN. Influence users detection in SN using quick sensor placement method is introduced in [14]. Though the method detects the influence users, the TPR was not improved.

Analysis of the dynamic influence of the users has been presented [15] in a SN based on network topology or statistical properties—the error rate is not minimized during the dynamic influence user identification. A Fuzzy Concept Analysis (FCA) and Latent Dirichlet Analysis (LDA) are performed in [16] to identify the right influencers. The time analysis in the influence detection remained unsolved.

An influence propagation model is presented in [17] to define the user's interaction and the influence on neighboring users. The model failed to consider the time parameter in the influence of users' identification. A structural diversity approach was introduced [18] to detect the top-k influential users in SNs. The model failed to use other optimization techniques for minimizing the time complexity.

An Evidential Influence Label Propagation Algorithm (EILPA) designed in [19] to efficiently detect the distributed community using influential users with minimum time complexity. The error rate not minimized. Identification of influence users performed in [20] based on multiple features. The PageRank concept is used to rank the influence of users in the SN. The accurate detection is not performed with lesser time complexity.

The above-mentioned issues identified from the existing techniques resolved by introducing a novel method called SDS-PBRR. The brief description of the proposed SDS-PBRR method presented in the next section.

2 Proposed Statistic Dice Similarity Based Probabilistic Binomial Regression and Ranking for Influence Behavior Detection

SNing services are an internet-based application at various online communities where people communicate with each other every day. In the SNing services, as well as quantifying user behavior is an essential task by analyzing the effective communications among the users throughout different periods. The successful prediction of user behavior is used to enhance the detection of influencers in the SN for access control in online groups. Influencers are the members of the SN who share positive/negative information and opinions about the services, brands/businesses among the other members of their online networks. Therefore, are detected to preserve the main features of the online SNs. Based on this motivation, Statistic Dice Similarity-based Probabilistic Binomial Regression and Ranking (SDS-PBRR) method are introduced in this paper. The proposed SDS-PBRR method is also designed. The architecture of the SDS-PBRR method is illustrated in Fig. 1.

Fig. 1 illustrates the architecture of the SDS-PBRR method to detect IB users in the SN. First, the similarities between the users based on their behaviors are calculated to find the influencer in the SN. After calculating the

© 2019 NSP
Natural Sciences Publishing Cor.

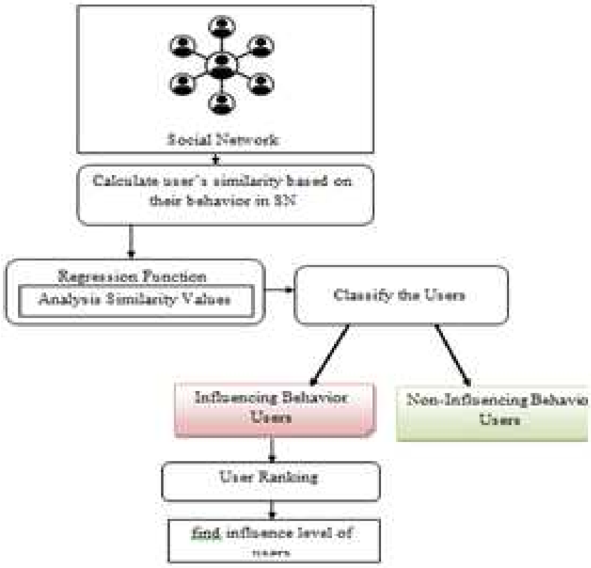

Fig. 1: Architecture of the SDS-PBRR method

similarity value, the SDS-PBRR method uses the probabilistic binomial regression function to analyze the similarity value with the threshold—the regression results used for identifying is applied to IB users and N-IB users. Finally, the firefly algorithm is applied to rank the IB users to find their first level in the SN—the brief description of the SDS-PBRR method is illustrated in the following subsection.

3 Statistic Dice Similarity Measure

The first process in SDIBS-PBRR Method compute the similarity between the users (i.e., IB and the user's behaviors in an SN) for classification in SNs with higher TPR and lesser time consumption. Dice similarity is a statistical method used to compute the similarity between two users. Let us consider the numbers of users are $u_1, u_2, \ldots, u_n$ and their behaviors are represented as $B_1, B_2, \ldots, B_n$ in the SN. The statistic dice similarity is computed using the following mathematical equations:

$$\gamma(B_1, B_2) = 2 \times \left(\frac{B_1 \cap B_2}{B_1 \cup B_2}\right) \ldots\ldots\ldots \quad (1)$$

In (1), γ denotes an SDSC. B_1 represents the IB of user 1, B_2 represents the behavior of the user 2 in the SN. γ is the ratio function: the ratio of the mutual independence between the two users' behaviors and the dependence between the two users' behaviors. The intersection symbol $\cap$ denotes mutual independence which indicates that two users' behaviors are statistically independent. The union symbol $\cup$ represents a mutual dependence which denotes two users' behaviors are statistically dependent. The SDSC (γ) provides the similarity value between 0 and 1. Likewise, the similarity value of all users' behaviors is computed using SDSC. The estimated similarity values are analyzed using Probabilistic Binomial Regression Function (PBRF).

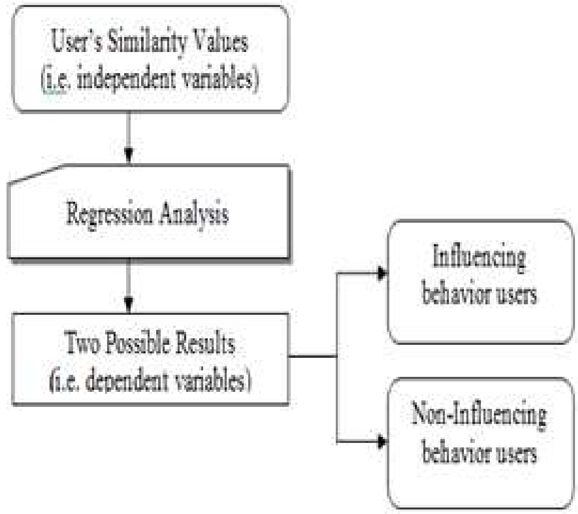

Fig. 2: Flow process of probabilistic binomial regression analysis

3.1 Probabilistic Binomial Regression Analysis based User's Behavior Classification

The second process in the SDS-PBRR method is the classification using Probabilistic Binomial Regression. The SDS-PBRR method identifies the IB in the SN through the regression analysis for virtual community formation. The IB users share the information to a massive number of individuals through an SN. Similarly, it is possible to reduce the unwanted content (*e.g., 'false news'*) shared to the other users. The primary issue in both cases is to identify the influential behavior users in the SN. Therefore, SDS-PBRR method identifies the influential behavior users using PBRF.

Binomial regression is a machine learning technique that offers the two possible results, such as IB users or N-IB users. Regression analysis is the statistical processes that compute the relationships between a dependent variable (i.e., two possible results) and one/more independent variables (i.e., users' similarity values). The process of probabilistic binomial regression is illustrated in Fig. 2.

Fig. 2 shows the flow process of PBRF to identify IB users or N-IB users. The input of the binomial regression function is a similarity value of the users in the SN. Let us consider the number of users' similarity values $v_1, v_2, v_3, \ldots, v_n$. The regression function provides the probabilities of occurrence of two possible results. The regression analysis is performed using the following

© 2019 NSP
Natural Sciences Publishing Cor.

mathematical equations:

$$\omega = \rho \cdot v_i + \varepsilon_r \tag{2}$$

In (2), ω denotes a regression function, ρ denotes a regression coefficient, v_i is a set of independent variables (i.e., users similarity values). ε_r represents the random variable as specifying error in the prediction. The ω provides the distributed results between 0 and 1. Based on the results obtained from the regression analysis, the threshold value is set to identify the N-IB users and IB users in the SN. The output of the binomial regression function expressed as follows,

$$y_r = \begin{cases} 1, & \omega > 0.5 \\ 0, & \omega < 0.5 \end{cases} \tag{3}$$

In (3), y_r denotes an output of the regression function. The threshold value set as 0.5. If the regression function (ω) value is higher than the threshold value 0.5, the users are classified as IB users. If the regression analysis function (ω) value is less than the threshold value 0.5, the users are classified as N-IB users. To explain, the SDS-PBRR method effectively analyzes the users' behaviors based on the social influence of virtual community formation.

3.2 Firefly Algorithm for Ranking the IB Users

After detecting the IB users from the SN, the ranking is carried out to identify the first level of the IB users $(iu_1, iu_2, \ldots, iu_3)$. It is performed by using the firefly algorithm. The Firefly is a meta-heuristic algorithm which is processed depending on light intensity behavior. In this case, the fireflies are considered as IB users in the SN. Application of the firefly algorithm, provides two significant variables, i.e., light intensity and attractiveness. The brighter one attracts the firefly with less light intensity. Therefore, the objective function of this algorithm represented as the light intensity (i.e., a ω value higher than the threshold value ($\omega > 0.5$).

By applying the firefly algorithm (population of the firefly), IB users are generated. After the initialization, the objective function of each firefly is expressed as $\omega_1, \omega_2, \ldots, \omega_n$. For each firefly, the light intensity is formulated in connection with the objective function.

$$L(F) = f(x) \tag{4}$$

In (4), $L(F)$ denotes a light intensity of the firefly, and $f(x)$ denotes an objective function. Estimating new solutions and updating any pair of two fireflies positions are expressed as follows:

$$F_i^{t+1} = F_i^t + \rho \exp[-\tau] \left(F_j^t - F_i^t\right) + \beta_t \delta_t \tag{5}$$

In (5), F_i^{t+1} denotes an updated position of the firefly F_i. F^T represents the current location of the firefly F_i, ρ denotes an attractiveness of the firefly, τ is the light absorption coefficient, F_j^t denotes a current location of the firefly F_j, β_t is a randomization parameter between 0 and 1, δ_t is a vector drawn from a Gaussian/ other distribution. The updated position of the firefly is a new location of the firefly. Based on the updated results, the fireflies are ranked to determine the current best solution.

$$R \rightarrow F_1, F_2, \ldots, F_n \tag{6}$$

In (6), R denotes a rank assigned to all the fireflies based on their objective function. Based on ranking, the topmost levels of the IB users identified. In an SN, the highly-ranked users influence SNs more than the others. This process takes minimum time for ranking the users. The algorithmic process of statistic dice similarity based probabilistic binomial regression and ranking described as follows.

Algorithm 1: A Statistic Dice Similarity Based Probabilistic Binomial Regression and Ranking

Input : Number of users $u_1 u_2 u_3 \ldots u_n$, behaviors of the users $B_1 B_2 \ldots B_n$
Output: Identify the IB users in an SN

begin
  Compute similarity between the two users based on behaviors $\gamma(B_1, B_2)$
  Perform regression analysis $\omega = \rho . v_i + \varepsilon_r$
  if $(\omega > 0.5)$ **then**
    regression output $y_r = 1$ IB users
    else if $(\omega < 0.5)$ **then**
      regression output $y_r = 0$ N-IB users
    end
  end
  Generate an initial population of IB users
  Define the objective function $f(x)$
  Formulate light intensity L based on objective function $f(x)$
  for $i = 1 : n$ *(all n fireflies)* **do**
    for $j = 1 : n$ *(all n fireflies)* **do**
      if $\left(L\left(F_j\right) > L\left(F_i\right)\right)$ **then**
        Firefly F_i moves towards F_j
        Evaluate new solutions and update light intensity
      end
    end
  end
  Rank the F_i based on their updated location
  Identify top rank F_i has high IB users
end

Algorithm 1 describes a process of identifying the IB users in the SN. First, the similarity between the users' behaviors is measured using the statistical dice similarity coefficient value. The estimated similarity values are given to the regression function to identify IB users and N-IB users. If the estimated regression value is below the

© 2019 NSP
Natural Sciences Publishing Cor.

Table 1: Columns and description

Columns	Description
Target	The polarity of the tweet 0 = Negative, 2 = Neutral 4 = Positive
IDS	The ID of the tweet
Date	The Date of the tweet
Flag	The QUERY (lyx). If there is no query, then value is NO_QUERY.
User	The user that tweeted
Text	The text of the tweet

threshold value, the users are classified be N-IB users. Otherwise, the users are classified as an IB user. Finally, the firefly algorithm is applied to identify the level of the N-IB users among the regular users—the populations of IB users initialized with the objective function. If the light intensity of the firefly is higher than the other, then the low-intensity firefly moves towards the higher intensity firefly. Then, the fireflies light intensity updated. Finally, fireflies are ranked based on the new location of the firefly. The top-ranked firefly is a high IB user.

4 Experimental Evaluation

The proposed SDS-PBRR method and existing methods Self-organizing Overlapping Community Detection (PSOCD) algorithm, Social Action-Based Influence Maximization (SAIM) Model are implemented using Java language. The Sentiment140 dataset with 1.6 million tweets is derived from thehttps://www.kaggle.com/kazanova/sentiment140. The dataset contains tweets with negative emotions and others with positive emotions. The sentiment1 40 dataset comprises 16, 00,000 tweets gathered using the Twitter API. The dataset comprises the six columns listed in Table 1

The proposed work is to conduct an experimental and analytical evaluation of the users' classification based on behavior in a SN with the dataset extracted from kaggle.com. The experimental evaluation is conducted based on various factors such as TPR, FPR, and execution time. Overall 250 tweet users are taken from the dataset as input for finding the IB of the users in the SN.

5 Results and Discussion

The simulation results of SDS-PBRR method, the existing PSOCD algorithm, and SAIM Model are discussed with the specific parameters, such as TPR, FPR and execution time. The simulation parameters results, explained with the help of tables and graphical representation. For each sub-section, the mathematical calculation is presented using three different methods.

5.1 Impact of TPR

TPR defined as the ratios of No. of users is correctly identified as IB users or N-IB users to the total number of users in the SN. The mathematical formula for computing the TPR is expressed as follows:

$$TPR = \frac{\begin{array}{l}\text{no. of users correctly identified}\\ \text{as influencing behavior or}\\ \text{non-influencing behavior}\end{array}}{n} \times 100 \tag{7}$$

In (7), TPR represents the TPR, n denotes the number of users in the SN. The TPR is measured in percentage (%). The sample mathematical computations for TPR using three different methods is presented as follows:

5.2 Sample mathematical calculation for TPR

Proposed SDS-PBRR method: Number of users correctly identified as IB users/N-IB users is 22, and the total number of users is 25. Then the TPR is calculated as follows:

$$TPR = \frac{22}{25} \times 100 = 88\%$$

Existing PSOCD: Number of users correctly identified as IB users/N- IB users is 20, and the total number of users is 25. Then the TPR is calculated as follows:

$$TPR = \frac{20}{25} \times 100 = 80\%$$

Existing SAIM: Number of users correctly identified as IB users/N- IB users is 19, and the total number of users is 25. Then the TPR is calculated as follows:

$$TPR = \frac{19}{25} \times 100 = 76\%$$

The above-mentioned mathematical computation results are used to show how effective the proposed SDS-PBRR method in identifying the IB users. Let us consider the 25 users, the SDS-PBRR methods correctly identify as IB users or N-IB users in the Twitter dataset, are 22 but the other two methods correctly identify 20 and 19 users. Therefore, the TPR of SDS-PBRR method is 88% and the TPR rate of the other two existing methods PSOCD, SAIM is 80% and 76%, respectively. Similarly, the nine remaining runs are carried out with the various users. Overall, ten different results are reported in Table 1.

Table 1 describes the performance results of TPR Vs. No. of users in the SN using three different methods, i.e. SDS-PBRR, PSOCD, and SAIM. For the simulation purposes, the numbers of users in the SN are considered as inputs for computing the TPR. For the simulation purposes, the numbers of users are assigned from 25 to 250. The performance results of the TPR using SDS-PBRR method are higher compared to those when

© 2019 NSP
Natural Sciences Publishing Cor.

Table 2: TPR Vs. number of users

Number of users	TPR (%)		
	SDS-PBRR	PSOCD	SAIM
25	88	80	76
50	90	86	80
75	91	85	83
100	87	82	78
125	88	83	79
150	89	85	81
175	86	81	77
200	88	83	79
225	89	84	77
250	90	85	76

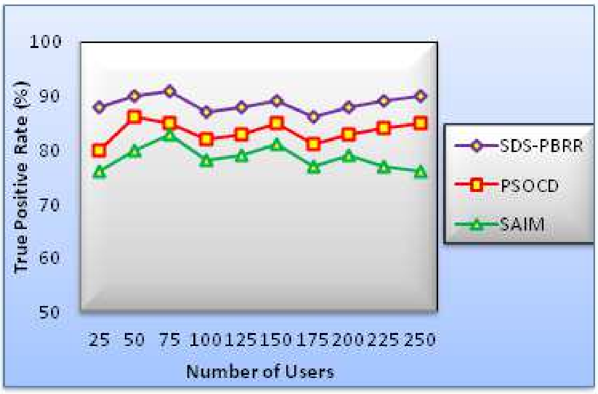

Fig. 3: Performance results of TPR vs. number of users

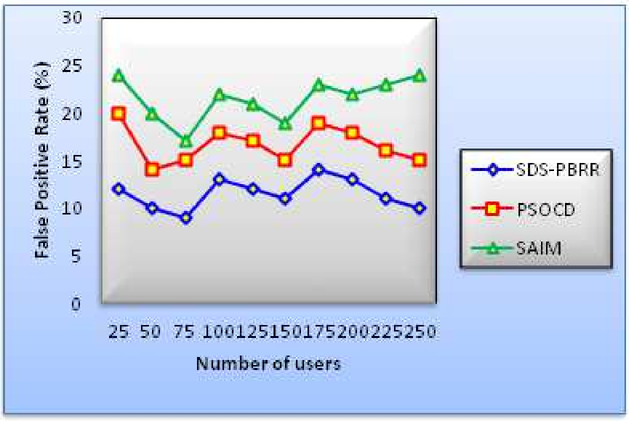

Fig. 4: Performance results of FPA vs. number of users

using PSOCD [1] and SAIM [2]. The results are plotted in the two-dimensional graph shown in Fig. 3.

Fig. 3 illustrates the performance results of TPR Vs. No. of users in the SN. In the graph, the number of users is taken as input in the x axis, but the corresponding TPR results are obtained at the y axis. The graph shows that the actual positive results of three different methods SDS-PBRR, PSOCD, SAIM are distinguished by the three different colors of the line: blue, red and green, respectively. The above-mentioned figure clearly illustrates that the SDS-PBRR method increases TPR compared to the existing methods. This significant improvement of the SDS-PBRR method achieved by applying the PBRF. The regression function analyzes the users' behaviors based on their similarity value. First, the dice similarities between the users' behaviors are computed. After that, the regression function takes the users similarity values as input to be compared with the threshold value. Based on the similarity threshold value, the regression function correctly identifies the IB user/ N-IB users. As a result, the regression function in SDS-PBRR method increases the TPR more than the existing methods.

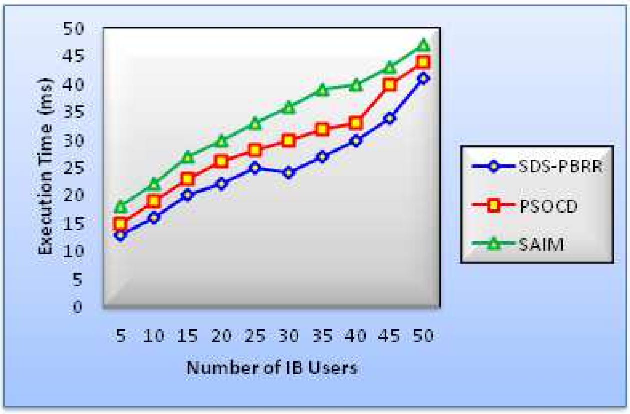

Fig. 5: Performance results of execution time Vs. number of IB users

After performing the ten runs with No. of users, the average results are computed to identify the percentage results of SDS-PBRR method. The results show that the SDS-PBRR method significantly increased the performance of correct positive rate by 6% and 13% when compared to PSOCD and SAIM.

5.3 Impact of FPR

The FPR defined as the ratio of No. of users incorrectly identified as IB user or N-IB users to the total number of users in the SN. The FPR is computed using the following mathematical equation,

$$FPR = \frac{\begin{array}{l}\text{no. of users incorrectly identified}\\ \text{as influencing behavior or}\\ \text{non-influencing behavior}\end{array}}{n} \times 100 \quad (8)$$

In (8), FPR represents the false positive rate, n denotes the number of users in the SN. The FPR is measured in terms of percentage (%). The sample mathematical calculations

© 2019 NSP
Natural Sciences Publishing Cor.

Table 3: FPR vs. number of users

Number of Users	FPR (%)		
	SDS-PBRR	PSOCD	SAIM
25	12	20	24
50	10	14	20
75	9	15	17
100	13	18	22
125	12	17	21
150	11	15	19
175	14	19	23
200	13	18	22
225	11	16	23
250	10	15	24

for the FPR using three different methods is presented as follows:

Sample mathematical calculation for an FPR

– Proposed SDS-PBRR: Number of users incorrectly identified as IB users/ N-IB users is 3, and the total number of users is 25. Then the FPR is mathematically computed as follows:

$$FPR = \frac{3}{25} * 100 = 12\%$$

– Existing PSOCD: Number of users incorrectly identified as IB users/N- IB users is 5, and the total number of users is 25. Then the FPR is computed as follows,

$$FPR = \frac{5}{25} * 100 = 20\%$$

– Existing SAIM: Number of users incorrectly identified as IB users/N- IB users is 6, and the total number of users is 25. Then the FPR is calculated as follows:

$$FPR = \frac{6}{25} * 100 = 24\%$$

Let us consider 25 users; the above-mentioned mathematical calculation shows that the FPR of SDS-PBRR is 12% and the FPR of PSOCD and SAIM is 20% as well as 24%, respectively. The result shows that the SDS-PBRR is better than the PSOCD and SAIM. The ten different results are computed and described in Table 2.

Table 2 indicates the FPR vs. several users. Despite varying the number of users as input, the different false positive results are achieved. The results report that the FPR using SDS-PBRR method minimized when compared to the existing PSOCD and SAIM. The results are plotted in the two-dimensional graph as shown in Fig. 4.

Fig. 4 depicts the performance results of the FPR concerning No. of users in the SN. The graphical result shows that the FPR considerably minimized using SDS-PBRR method. Because it effectively identifies the IB user and N-IB user through the regression analysis. The binomial regression function uses the regression coefficients to identify the behavior of the users in the SN. In the regression analysis, the error during the behavior analysis minimized. It helps to minimize the false identification of the IB/N-IB users. Moreover, the binomial regression function set the threshold value. They are classified as N-IB user. Otherwise, they are classified as IB user. This process minimizes false user identification in the SN.

Totally ten results are obtained for three methods, i.e. SDS-PBRR, PSOCD, and SAIM. The proposed false positive results compared with the actual results. Then the average value is taken for ten different results. Finally, the result shows that SDS-PBRR method decreases the FPR by 31% when compared to existing PSOCD. Similarly, the performance results of FPR minimized by 46% when compared to existing SAIM.

5.4 Impact of Execution Time

Execution time is defined as the amount of time required to rank the users for identifying the first level of the IB user. The execution time mathematically is computed as follows:

$$ET = n \times \text{time (rank one users)} \tag{9}$$

In (9), ET represents the execution time, n denotes the number of users. Execution time is measured in milliseconds.

5.5 Sample Mathematical Calculation for Execution Time

Proposed SDS-PBRR: Number of users is 5 and the time for ranking one user is 2.6 ms, then the execution time is computed as follows:

$$ET = 5 \times 2.6 \text{ ms} = 13 \text{ ms}$$

Existing PSOCD: Number of users is 5 and the time for ranking one user is 3 ms. Then the execution time is calculated as follows:

$$ET = 5 \times 3 \text{ ms} = 15 \text{ ms}$$

Existing SAIM: Number of users is 5 and the time for ranking one user is 3.6 ms. Then the execution time is calculated as follows:

$$ET = 5 \times 3.6 \text{ ms} = 18 \text{ ms}$$

Performance results of execution time for several IB users are described in Table 3. For the simulation purpose, the number of IB users taken as input varied from 5 to 50. The different execution time results are attained using three different methods, i.e. SDS-PBRR, PSOCD, and

© 2019 NSP
Natural Sciences Publishing Cor.

Table 4: Execution time vs. number of IB users

Number of IB users	Execution time (ms)		
	SDS-PBRR	PSOCD	SAIM
5	13	15	18
10	16	19	22
15	20	23	27
20	22	26	30
25	25	28	33
30	24	30	36
35	27	32	39
40	30	33	40
45	34	40	43
50	41	44	47

SAIM. The table value shows that the execution time minimized using SDS-PBRR when compared to the existing because the proposed method efficiently identifies the level of the IB user through the ranking process. The results are plotted in the following graph.

As shown in Fig. 5, the performance results execution time Vs. No. of IB users. The above graphical results confirm that the SDS-PBRR method minimizes the execution time when compared to other methods. The proposed SDS-PBRR method identifies most IB users among the number of users through the ranking process. The SDS-PBRR method uses the firefly algorithm for ranking the IB users based on the objective function.

The regression function values of two users are derived from the firefly algorithm analysis. If a user has a high objective function than the other position changes. Similarly, all the users are arranged based on the regression function values. Finally, the fireflies are ranked based on light intensity. The top-ranked fireflies are chosen as most IB users on the SN with minimum time. Thus, the SDS-PBRR method significantly identifies the level of the IB users and minimizes the execution time by 13% and 25% when compared to the existing PSOCD and SAIM, respectively.

According to above-mentioned discussions, the performance results of SDS-PBRR method effectively identify the influencing behaviors users in the SN with high TPR and less FPR as well as minimum execution time when compared to the existing methods.

6 Conclusion

An efficient method called Statistic Dice Similarity-based Probabilistic Binomial Regression Ranking (SDS-PBRR) is developed to identify the IB users in the SN with less execution time. First, the numbers of users in the SN is taken as input for virtual community formation. The statistic dice similarities between the users are computed for classifying the users. The PBRF evaluates the similarity value of the users with the threshold value. Based on the threshold value, the regression function classifies the user as IB/N-IB users with minimum error rate. It helps to improve the TPR and minimize the FPR. Finally, IB users are ranked to identify the first level of IB users in the SN. The firefly algorithm is applied to rank the users with minimum execution time. The simulation is carried out using twitter dataset with three different parameters: the correct positive rate, FPR and execution time. The performance results show that SDS-PBRR method increases the TPR in the IB detection and minimizes the FPR as well as execution time compared to the state-of-the-art methods.

References

[1] H. Sun, W. Jie, J. Loo, L. Wang, S. Ma, G. Han, Z. Wang and Wei Xing, A parallel self-organizing overlapping community detection algorithm based on swarm intelligence for large scale complex networks, Future Generation Computer Systems, Elsevier, **89**, 265–285 (2018).

[2] M. Azaouzi and L. B. Romdhane, An Efficient Two-Phase Model for Computing Influential Users in Social Networks Using Social Actions, Journal of Computer Science and Technology, Springer, **33**(2), 286–304 (2018).

[3] M. Al-Qurishi, Sk. Md. Mizan Rahman, M. Shamim Hossain, A. Almogren, M. Alrubaian, A. Alamri, B. B. Gupta, An efficient key agreement protocol for Sybil-precaution in the online social network, Future Generation Computer Systems, Elsevier, **84**, 139–148 (2018).

[4] X. Chen, C. Xia, J. Wang, A novel trust-based community detection algorithm used in social networks, Chaos, Solitons & Fractals, Elsevier, **108**, 57–65 (2018).

[5] H. Wei, Z. Pan, G. Hu, L. Zhang, H. Yang, X. Li, X. Zhou, Identifying influential users based on network representation learning in complex networks, PLoS ONE, **13**(7), 1–13 (2018).

[6] S. Peng, A. Yang, L. Cao, S. Yu, D. Xie, Social influence modeling using information theory in mobile social networks, Information Sciences, Elsevier, **379**, 146–159 (2017).

[7] L. Cui, H. Hu, S. Yu, Q. Yan, Z. Ming, Z. Wen, N. Lu, DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks, Journal of Network and Computer Applications, Elsevier, **103**, 119–130 (2018).

[8] X. Zhou, B. Wu, Q. Jin, Analysis of User Network and Correlation for Community Discovery Based on Topic-Aware Similarity and Behavioral Influence, IEEE Transactions on Human-Machine Systems, **48**(6), 559–571, (2018).

[9] J. Dong, B. Chen, L. Liu, C. Ai, F. Zhang, The Analysis of Influencing Factors of Information Dissemination on Cascade Size Distribution in Social Networks, IEEE Access, **6**, 54185–54194 (2018).

[10] M. Li, Y. Xiang, Q. Zhang, and B. Zhang, A Community-Related Influence Evaluation Model for Social Networks with a Dynamic Perspective, Mathematical Problems in Engineering, Hindawi, **9**, 1–17 (2016).

[11] Sudhakar Sengan and S. Chenthur Pandian, Hybrid Cluster based Geographical Routing Protocol to Mitigate Malicious

Nodes in Mobile Ad Hoc Network, International Journal of Ad Hoc and Ubiquitous Computing, **21**(4), 224–236 (2016).
[12] J. Wei, G. Mengdi, W. Xiaoxi, W. Xianda, A new evaluation algorithm for the influence of user in the social network, China Communications, **13**(2), 200–206 (2016).
[13] C. Mao and W. Xiao, A Comprehensive Algorithm for Evaluating User Influences in Social Networks Based on Preference Analysis and Random Walk, Complexity, Hindawi, 1–16 (2018).
[14] S. Pei, S. Tang, Z. Zheng, Detecting the Influence of Spreading in Social Networks with Excitable Sensor Networks, PLoS ONE, **10**(5), 1–19 (2014).
[15] H-J. Yin, H. Yu, Y-L. Zhao, Z-L. Zhu, and W. Zhang, Analysis of the Dynamic Influence of Social Network Users, Scientific Programming, Hindawi, pp. 1–6 (2017).
[16] S. A. Rí os, F. Aguilera, J. David, Nuñ ez-Gonzalez, Manuel Grañ a, Semantically enhanced network analysis for influencer identification in online social networks, Neurocomputing, Elsevier, **326–327**, 71–81 (2019).
[17] Sudhakar Sengan and S. Chenthur Pandian, Trustworthy Position Based Routing to Mitigate against the Malicious Attacks to Signifies Secured Data Packet using Geographic Routing Protocol in MANET, WSEAS Transactions on Communications, **12**(11), 584–2013 (2013).
[18] W. Xu, W. Liang, X. Lin, J. X. Yu, Finding top-*k* influential users in social networks under the structural diversity model, Information Sciences, Elsevier, **355–356**, 110–126 (2016).
[19] M. Azaouzi and L. B. Romdhane, An evidential influence-based label propagation algorithm for distributed community detection in social networks, Procedia Computer Science, Elsevier, **112**, 407–416 (2017).
[20] Q. Sun, N. Wang, Y. Zhou and Z. Luo, Identification of Influential Online Social Network Users Based on Multi-Features, International Journal of Pattern Recognition and Artificial Intelligence, **30**(6), 1–15 (2016).

R. Gnanakumari serves as an as Assistant Professor in the Dept. of CSE at Nehru Institute of Engg. & Tech. Coimbatore. She is pursuing Ph.D. Degree under Faculty of ICE, Anna University, and Chennai. She received B.E.-CSE from Sri Ramakrishna Engineering College, Coimbatore. Also, she completed M. E.-CSE from the Anna University of Technology, Coimbatore. She has published various research papers in National & International Journals & Conferences. Her research area is Data Mining, Machine Learning, and Social Networks.

P. Vijayalakshmi serves as Associate Professor in the Dept. of ECE at Hindusthan College of Engg. & Tech., Coimbatore. In 2014, she completed Ph.D., in Faculty of ICE, Anna University, Chennai. She did B. E.-ECE from PSG College of Technology. Also, she completed a Master's Degree from GCT, Coimbatore, Anna University, Chennai. She has more than 22 years of teaching experience. Moreover, she is supervising 05 Research Scholars at Anna University. She has published 8 papers in National and International Journals and 26 papers in National and International Conferences. Her research interest is Wireless Communications and VLSI Design.

© 2019 NSP
Natural Sciences Publishing Cor.