Global Journal of Engineering and Technology Review

Journal homepage: http://gjetr.org/

Global Journal of Engineering and Technology Review 1 (1) 34 – 40 (2016)

A Trusted Behaviour Learning for Interest Prediction in Social Ontology Based on Weighted Graph

J. Kalavathi[1]*, S. Balamurali[2] and M. Venkatesulu[3]

[1,2,3] Department of Computer Applications, Kalasalingam University, Krishnankoil, 626126, TN, India

---

ABSTRACT

Objective – In this paper, an attempt is made to suggest a social ontology dependent behavior learning approach to anticipate the interest of group of users utilizing social graphs and density measure.
Methodology/Technique – Weighted clustering is implemented using trustiness to ensure better linking among the social domain. The social ontology includes a set of classes in which related terms and properties are referred regarding a category or interest. The suggested approach clusters the user groups into N number of weighted clusters of Ontology set S, and there would be a weight cluster for each one of the interests or subjects or classes defined in the social ontology. From the social data set, a social graph can be built where every user may be regarded as a node and there would be a connection between user solely if they have general interest. The suggested technique clusters the user groups based on the trust value which is mostly dependent on density measures.
Findings – The evaluation results have shown that the proposed technique has produced less time and space complexity values.
Novelty – The proposed technique identifies the user group interest in exact way and acquires effective results

Type of Paper: Review

Keywords: Social Ontology; User Interest; Density Measure; Behavior Learning; Weighted Clustering; Trusted Graphs

---

1. Introduction

The latest internet users share their views and thoughts by using social networks which have grown as the most modern tool. The set of nodes interlinked unitedly to build a network, which will be named as social networks. Node represents every user of network and each user can have a number of conversations among set of users. For instance, users may share data among them by using a social network, such as YouTube. The users in the group post comments or upload other video as a response for the comments or a video that was posted or uploaded by a few users in the same group [6]. This is because of the synonymous interest, the users publish comments or upload videos.

---

* Paper Info: Revised: July, 2016
Accepted: November, 2016

* Corresponding author:
E-mail: kalajeyaraj@gmail.com
Affiliation: Department of Computer Applications, Kalasalingam Universit, India

Plenty of users share data concerning a subject or their private on other modern social networks [1] such as face book and twitter. Likewise there have many groups with several names preserved by several social network users and an individual user may be a role of any number of groups. A user could share any data either to the public or a particular group of users [4]. At present, the social network can be utilized like a ballot machine, which means, if anyone needs to know regarding the effectiveness of any politician, then he/she just writes a comment and requires others to post their comments on that. The positive and negative votes are calculated, depend on these comments. In this manner, the social networks encompass higher effect on the practical world.

From the social data set and from the social graphs, the user interest or synonymous interested user groups could be distinguished. In a social graph that may be made for the social data, every user contains numerous edges and holding relations with other users. This paper offers a method to determine synonymous user groups by learning the users behaviors and anticipate the users interest. Social dimension, FacetNet and suggested techniques. It distinctly demonstrates that the suggested technique has acquired higher prediction accuracy than the other techniques [2].

For interest prediction utilizing social ontology and social data set, a behavior learning technique has been suggested in this paper. Data mining, image processing, networking, etc. are the different classes of the social ontology, and each class contains a number of sub classes. The relational synonyms which belong to a class or category can be demonstrated by the ontology. To cluster the synonymous interested users, such an ontology has been employed by the suggested technique.

2. Proposed Methodology

The suggested technique involves four steps: (i) Social graph construction phase does construction of social graphs from the social data set. (ii) Density measure calculation process calculates density measure for every social ontology class towards every user. (iii) User groups with synonymous interest will be distinguished utilizing the density measure to make the cluster at clustering phase. (iv) In final step, the user group interest can be distinguished from the clustered users and their behavior model. Figure 1 shows the architecture model of proposed system.

2.1. Social Weighted Graph Clustering and Construction

The social graph was built by utilizing the blog catalog dataset; the set of edges and nodes are extracted from the data set. The data regarding the edges between users as 1-3 that demonstrates that there's a connection between the nodes 1 and 3, were included in the data set. Similarly it includes 4,52,784 entries in the data set and contains 37 thousand distinct node numbers present.It establishes a node that has the value of node named as number for every distinct user and the connections are distinguished to produce link between nodes when all the nodes were established. Due to the memory constraints, it's restricted itself up to 2500 users. In the given technique, the constructed social graph will be used at the consecutive stage.

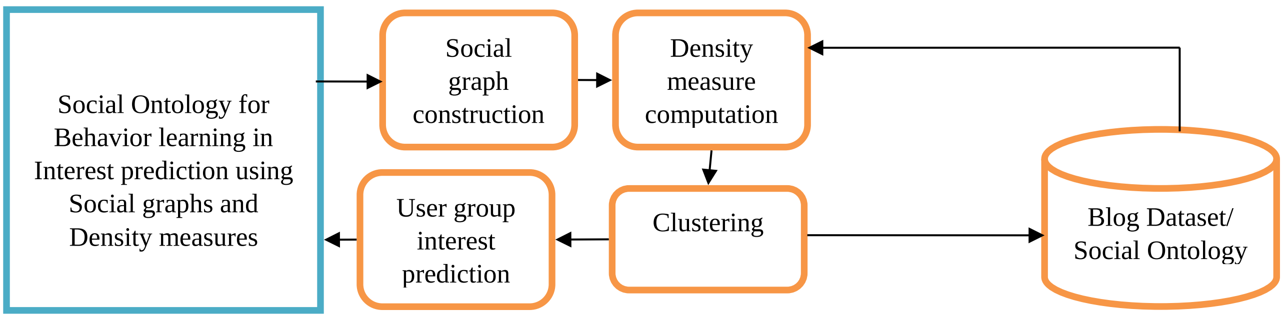

Figure 1.Proposed method of Architecture.

2.1.1. Problem Statement

Community structure finding is regarded as a graph clustering difficulty and an optimization difficulty [7]. An assumption is made that every person or a community has a density value and every pair of persons or communities contain a threshold value. A social network represents a graph, every person stands for a node, edges denote the relationship between people. For a presented sparse graph $G(V, E, W_V, S_E)$ that comprises of the node set V, the edge set E, the weight of node set W_V and also the weight of edge set S_E, it may be involve in finding the clusters of G like communities.

The outcome of graph clustering must divide a graph into different sub-graph (clusters), weight value is included in every portion, and in addition, there're attractiveness values between clusters that are identical with the edge weights. Equation (1) gives the optimization objective function where ρ represents the partitions of a graph.

$$\arg\max_{\rho}\left\{\sum_{k\in\rho}W(k)-\sum_{i,j\in\rho}S(i,j)\right\} \tag{1}$$

2.1.2. Preliminaries of Community in Weighted Graph

Definition 1: Density of cluster: A set of edges linked with each other is included in every node and the density of edges of each of its neighbors was calculated. The boundary of the cluster was demonstrated by the density measure and also it presents whether an individual node may be added to the cluster or not [5].

A novel node must have additional edges towards the nodes that already existed in the cluster. This scheme produces the cluster as additional exact one. Cluster density is defined as the average of all the weights of nodes in the cluster.

Definition 2: Attractiveness between clusters: The portion of the summation of all the edge's weights between two clusters and also the product of the node number of the two clusters is known as attractiveness between clusters. For the number of edges between clusters i and j be q, the edge weight can be S_e, $e \in 1,2,\ldots,q$, community i contains $Q_i$nodes and community j comprises of Q_j nodes, after that the attractiveness between cluster i and j is given as follows:

$$S_{ij}=\frac{\sum_{e=1}^{q}S_e}{Q_i\times Q_j} \tag{2}$$

Definition 3: Inter-interested clusters: The succeeding considerations that are given below must be fulfilled if cluster i and cluster j are inter-interested clusters.

$$q\geq Q_i, q\geq Q_j \tag{3}$$

Definition 4: Community: If a cluster i could be a community, then it should fulfill that:

$$S_{ij}<W_i+W_j, \forall j \tag{4}$$

The inter-interested cluster of cluster i implies cluster j.

2.1.3. Clustering Algorithm

Initially, every individual cluster refers to each node and the cluster algorithm will be an agglomerative algorithm.

First of all, it is necessary to determine which cluster is all its inter-interested clusters will obtain the highest attractiveness along with them, that is indicated by j, before performing the merger for cluster$i$.

The merger can't be performed straight, later finding the two clusters, since the attractiveness between them might be extremely smaller, implying that they might not be of the similar community, thus let have to do other judgment. Solely S_{ij}fulfill the following condition:

$$S_{ij} \geq W_i + W_j \tag{5}$$

The cluster i and j would be combined.

There occur two specific events during the merger, additionally. First, cluster i might not have inter-interested clusters, next cluster i won't combine with any other clusters, it would be a community; secondly, there may be more than one clusters contain the highest attractiveness with cluster i, and fulfills the eqn. (5) at the identical time, next it combines cluster i with anyone of them.

Matrix S represents the cluster attractiveness matrix, which is a k-order matrix, where $S_{ij} = S_{ji}$ refers the attractiveness between the cluster i and j, and in each iteration k will be modified. Considering that the total number of nodes be$n$, then the attractiveness matrix S can be a n-order matrix at beginning.

It is essential to update the matrix S, because after merging, the number of clusters will be minimized, the density of clusters would modify and the attractiveness between clusters may additionally alter consequently. It may utilize the previous one to cut down the amount of computation, once updating the novel attractiveness matrix. The group of clusters can be CM_{pre}, the similar attractiveness matrix will be S', the number of clusters be k', earlier the merge function; then the group of clusters be CM_{cur} and also the similar attractiveness matrix will be S, the number of clusters be k, later the merger.

$$S = \begin{pmatrix}
0 & 5 & 0 & 0 & 16 & 45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 12 & 44 & 0 & 43 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 12 & 0 & 0 & 22 & 38 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\
0 & 44 & 0 & 0 & 66 & 69 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
16 & 0 & 22 & 66 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\
45 & 43 & 38 & 69 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 18 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 & 41 & 0 & 58 & 64 \\
0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 & 23 & 0 & 0 & 0 & 37 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 & 35 & 48 \\
0 & 0 & 0 & 0 & 8 & 0 & 41 & 0 & 0 & 0 & 9 & 0 & 45 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 \\
0 & 0 & 8 & 0 & 0 & 18 & 58 & 0 & 35 & 0 & 0 & 0 & 68 \\
0 & 0 & 0 & 0 & 0 & 0 & 64 & 37 & 48 & 45 & 0 & 68 & 0
\end{pmatrix} \tag{6}$$

CM_{pre} is represented by using the succeeding mathematical expression.

$$CM_{pre} = \left\{ cm_l \setminus l = 1,2,\ldots,k' \right\} \tag{7}$$

Where cm_l refers the l^{th} cluster in CM_{pre}.

The mathematical formula can be given as follows, when the cluster p in CM_{cur}comprises of a number of clusters in CM_{pre}, and also the number is m, specifically, cluster p inCM_{cur} may be organized by the merger of m clusters in CM_{pre}.

$$CM_{cur}^{p} = \left\{ cm_t \middle| cm_t \in CM_{pre}, t = 1,2,\ldots,m \right\} \tag{8}$$

Where CM_{cur}^{p} stands for the cluster p in CM_{cur}. The attractiveness between community i and community j, will be updated by eqn. (9) afterwards, which is the element S_{ij} in matrix S.

$$S_{ij} = \frac{\sum_{cm_r \in CM_{cur}^{i}, cm_t \in CM_{cur}^{j}} S'_{cm_r, cm_t} \times Q_{cm_r} \times Q_{cm_t}}{Q_i \times Q_j} \quad (9)$$

A novel attractiveness matrix can be obtained after updating the elements in *S* one after another with eqn. (9).

2.2. BTN: Building the Trust Network

Two matters required to be made to form a trust network from SOURCE to SINK.

- Determine as numerous short paths for the two presented nodes as feasible, that may be a normal breadth-first search. For finding trusted acquaintance chains, PSN [8] is used for every step.
- Include trust data between directly linked nodes.

For the breadth-first search two methods are offered: (i) *Centralized* technique, specific for smaller online social networks, like a social network for a university, a company, and so on; (ii) D*istributed* technique, more suitable for larger online social networks that may ease the weight on the servers.

Assume *G* denotes the social network later the *PSN* procedure, *L* refers the maximum length of paths and *c* is utilized to moderate the path length, for Algorithm 1 and Algorithm 2. The set of trusted acquaintances of *u* can be given by $L^+(u)$, through the *PSN* procedure they are chosen and they are arranged in descending order by means of their precedencies in every class of longest contacts, longer contacts and local neighbors. The unvisited nodes can be represented by R_{source}.

Algorithm 1

Step 1: Start

Step 2: Read Social Data set SD.

Step 3: Initialize Social Graph set SG, Node List NL.

Step 4: for every entry from Nodes File NF

Create Node N_i = NF(Node.ID).

Add Node to the Node List NL.

$NL = N_i + \sum N_i \{NL\}$

End.

Step 5: for every entry from Edge File EF

Identify Source Node $SN = \int NL \times EF(SAddr)$

Identify Destination Node $DN == \int NL \times EF(DAddr)$

Create connection between SN and DN.

$SN = \sum NL_i + NL(DN)$.

End.

Step 6: Stop.

Algorithm 2

Distributed BFS (G, SOURCE, SINK) Algorithm

Step 1: **Input:** G. SOURCE, a trustor; SINK, a trustee.

Step 2: **Output:** D, a path set from SOURCE to SINK.

Step 3: $c \leftarrow L - 1$. Let SOURCE be the *current node*.

Step 4: **for** every unvisited neighbor*u* of *current node***do**

Step 5: **if***u* is SINK **then**

Step 6: *u* sends *response* backward to his *requester*, add *u* to *responder*; Iterate till a *response* is transmitted to SOURCE. Reverse *responder* to obtain a path *P*, then add *P* into *D*.

Step 7: **else**

Step 8: **if***c*> 0 **then**

Step 9: $c \leftarrow c - 1$, *u* transmits request to $L^+(u)$.

Step 10: Set *u* as visited.

Step 11: **end if**

Step 12: **end if**

Step 13: **end for**

2.2.1. Defining Trust between Directly Connected Nodes

The work given in this paper distinguishes between *referral trust* and *functional trust* that has been introduced in [3]. The trust network can be obtained by using *referral trust* to the entire edges excluding for the final hop and *functional trust* to the final hop. But, mining trust data for various social networks which regards data mining and other methods was a difficult job.

2.2.2.1 Referral Trust.

Referral trust can be delineated like the priority to be chosen as the consecutive hop that can be within the range of [0, 1]. The *referral trust* solely utilizes the data of *domain* that will be stable and objective and can’t be modified as per the wish.

Definition 5: referral trust from i to *j* is equivalent to the precedency of *j* to be chosen like the consecutive hop by *i*: $RT(i, j) = p(i, j)$.

2.2.2.2. Functional Trust.

It solely presents an idea for delineating functional trust, while to find short paths is the important function of this paper. Various types of features, like *social relations*, *explicit ratings*, *reputation*, and *similarity* are regarded.

3. Results and Discussion

The suggested technique was applied and measured utilizing the consecutive data sets as presented in Table 1. Every data set from this table has entry for more number of users and has more size that can't be processed because of the memory limitations. For instance, the Blog catalog data set has entry and relations for more than 35,000 users and has 4,52,000 edges. Building social graphs of like large size isn't possible and it have limited itself to 2500 for evaluation function.

Table 1: The Data Set used.

Data Set	Number of Users	Number of Messages
Enron	2359	32,789
DBLP	343, 103	491726
Blog Catalog	32,700	4,52,000

Figure 2a depicts the time complexity taken by several approaches, whereas utilizing Blog catalog data set, and it presents that the suggested technique has less time complexity than the others. The Figure 2b shows the time scatter plot of the minimal information centrality which is used to find the average distance among the nodes for communication. Figure 3 generaly the space taken by different algorithms that compared to evaluate the proposed technique. It depicts that the proposed method has utilized solely a less memory where as the others have taken more memory.

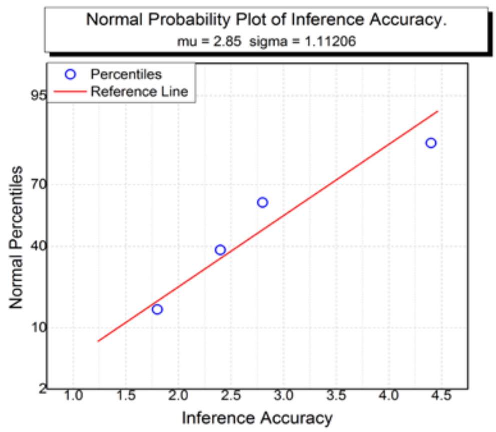

Figure.2a. The time complexity of different approaches with normal Probability Plot.

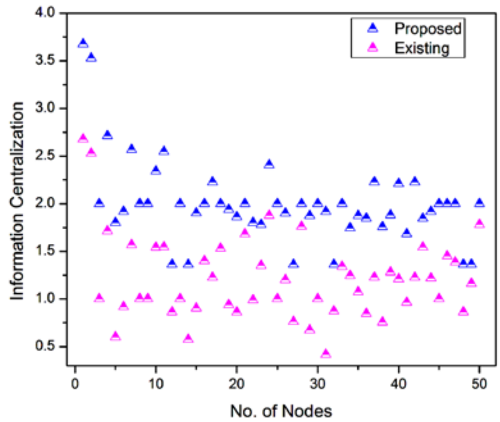

Figure.2b. The scatter plot of different approaches.

Figure 4 shows the clustered output obtained from the proposed approach based on the better connectivity. The minimal shortest path among the network is given by 0.004096 with weighted edge of 5. While the information centrality is given by 0.065972 with minimum clustered classifier as three. The standard deviation obtained from the proposed approach for information centrality is 0.010191.

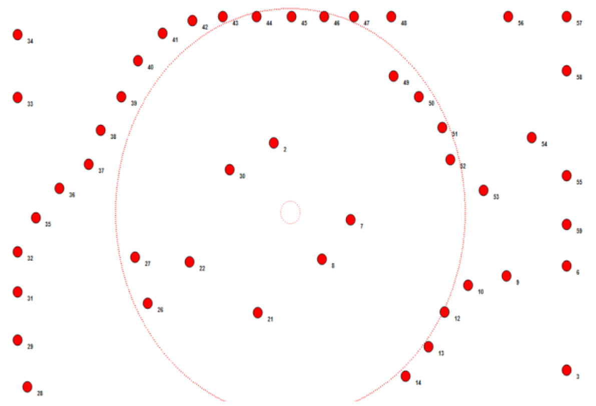

Figure. 3. The improvised clustered output obtained from the proposed system for better connection among the social network.

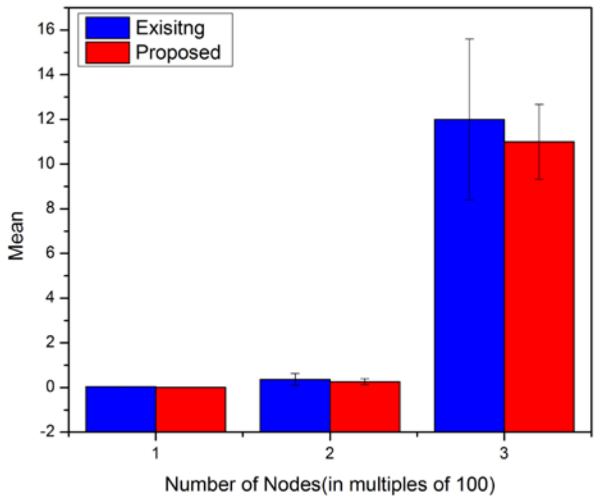

Figure. 4. The space occupied by different algorithms on blog catalog data set.

4. Conclusions

This paper has proposed a social ontology based interest prediction utilizing social graphs and density measures. The proposed technique clusters the users into groups based on density measures. The density measure stands for the relevancy of users around a semantic meaning or classes. The cluster groups interest is identified employing semantic weight that is calculated utilizing the semantic measure and interest persistence measure. The proposed technique identifies the user group interest in exact way and acquires effective results. The evaluation results have shown that the proposed technique has produced less time and space complexity values.

References

[1] Akshay P, Juan L,Jie G.Predicting group stability in online social networks. Inter conf on WWW, proceeding 2013; 1021-1030.

[2] Jingxuan L, Wei P, Tao L, Tong S, Qianmu L, Jian X. Social network user influence sense-making and dynamics prediction. Expt Sys with Appln 2014; 41:5115–5124.

[3] Jøsang A, Hayward R, Pope S. Trust network analysis with subjective logic. 29th Aust Comp SciConf 2006; 85–94.

[4] Kalavathi J, Balamurali S, Venkatesulu M. An efficient evolutionary approach for identifying evolving groups in dynamic social networks using genetic modelingProc Comp Sci 2015; 57:428-437.

[5] Lin Y.R, Chi Y, Zhu S, Sundaram H, Tseng B.L.Facetnet: A Framework for Analyzing Communities and Their Evolutions in Dynamic Networks. 17th IntConf on WWW 2008; 685-694.

[6] Tang L, Liu H. Scalable Learning of Collective Behavior Based on Sparse Social Dimensions. 18th ACM Conf on Inf and KnwMang 2009; 1107-1116.

[7] Yang T, Chi Y, Zhu S, Gao Y, Jin R. A Bayesian Approach Toward Finding Communities and Their Evolutions in Dynamic Social Networks. SIAM IntConf on Data Mining 2009; 990-1001.

[8] Wenjun J, Guojun W, Jie W. Generating trusted graphs for trust evaluation in online social networks. Adc in Comp SupptCollab Sys and Tech 2014; 31:48 – 58.