International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 06 | June 2018 www.irjet.net p-ISSN: 2395-0072

Twitter User Social Circle Detection Using Multi-View Network Structure

Sneha Rani[1], Prem Ranjan[2], Sagar Kumar[3],Ciba Tembhare[4], Prof.Smita Khot[5]

[1,2,3,4]Sneha Rani ,DYPIT, Sant Tukaram Nagar
[5]Prof. Dept. of Computer Engineering, DYPIT, India

Abstract – In our system, we can learn social circles in ego-networks which are based on multi-view network structure. We can classify information about the similar data or similar information. Here we can detect ego-network based on social circle. Automatically social circle detection in ego-networks is a very fundamentally important task for social network analysis. In this paper, we know, how to detect circles by leveraging multiple views of the network structure. For detection of this leveraging multiple views of the network structure, we crawl ego networks from Twitter and model them by six views, including user relationships, user interactions and user content. Friendship is the one view which is used in social circle detection. In this system characterizes the friend relation between alters by a similarity matrix where alters follow each other on Twitter. It is a most common view for social circle detection. Its only check the twitter users follow each other or not but it don't check the tweets of user. In our system we use Sentiment Classification of tweets using NLP (Natural Language Processing). It helps to find the accurate friend relation between alerts. We apply multi-view spectral clustering techniques to detect circles on these ego-networks. In this paper we can used a modified multi-view spectral clustering techniques over a single-view clustering methods. We integrate this how the bound may be affected by several network characteristics. How the different network characteristics affected on a social network.

Keyword - *Social Circle Detection, Data Crawling, Sentiment Analysis, Multi-View Spectral Clustering*

I. INTRODUCTION

In a proposed system, we can solve the problem of single value clustering technique we can used a multi-value clustering technique is used. First, we propose to effectively leverage multiple views of the network structure for better automatic social circle detection in ego-nets. To that end, we introduce multi-view spectral clustering techniques and demonstrate they superior circle detection performance, as compared with common single-view clustering techniques. Second, we propose to interpret the sparseness of ego-net structure as incompleteness, and conjecture the ignorance of such hidden incompleteness may result in performance bias. To that end, we first derive an upper bound for the performance bias, with implications supported in simulations; we then propose a modified multi-view clustering technique which selectively transfers information from sparse views, and demonstrate its superior circle detection performance as compared with the standard multi-view clustering technique which fully transfers information across views. Finally, extensive experimental evaluations are done based on the ego-nets we crawled from Twitter. Structural Views, Interaction Views, Content View are the three type of a view which can be applied for the view. In this system characterizes the friend relation between alters by a similarity matrix where alters follow each other on Twitter. It is a most common view for social circle detection. Its only check the twitter users follow each other or not but it don't check the tweets of user. In our system we use Sentiment Classification of tweets using NLP (Natural Language Processing). It helps to find the accurate friend relation between alerts.

Objective and Goal:

- Automatic social circle detection in ego-networks.
- The main goal of our system is finding out a multiple view clustering Information we can used a multi view clustering and a Selective Co-trained Spectral Clustering. Which give us appropriate information and correct information rather than the single view clustering information.

II. LITERATURE SURVEY

J. Yang, J. McAuley, and J. Leskovec present "Community Detection in Networks with Node Attributes". Communities from Edge Structure and Node Attributes (CESNA), an accurate and scalable algorithm for detecting overlapping communities in networks with node

attributes. It is find semantic relations between the terms to the general expression relation. If one source of information is missing or noisy, the other can make up for it. However, taking both node attributes and network topology for community detection is also difficult task. This is used to get community detection in networks with node attribute.

P. Shi, H. Xu, and Y. Chen proposed "Using Contextual Integrity to Examine Interpersonal Information Boundary on Social Network Sites". Identifies users' interpersonal privacy concerns that are rooted from informational norms outlined in the theory of contextual integrity. The tensions that occur within and cross these informational norms. It is too difficult to identify information. It is used to examine information boundary on social network sites.

P. Ferragina and U. Scaiella proposed "Fast and accurate annotation of short texts with wikipedia pages". The sophisticated graph of topics produced by Tag me for input text might lead to the design of innovative. It is difficult to implement rather than other techniques. It can use for topic information using a tag me.

C. Lan and J. Huan present "Reducing the unlabeled sample complexity of semi-supervised multi-view learning". We improve the state-of-art u.s.c. from $O(1/\epsilon)$ to $O(\log 1/\epsilon)$ for small error ϵ, under mild condition To obtain the improved result, as a primary step we prove a connection between the generalization error of a classifier and its contradictory, which measures the similarity between classifier and sample distribution. It is costly, time consuming, and often unnecessary to find communities for an entire network. In this paper we reduce a sample complexity of semi-supervised multi-view learning.

D. M. Boyd and N. B. Ellison proposed "Social network sites: definition, history and scholarship". This paper, which gives the information related to social network sites. We describe features of SNSs and propose a comprehensive definition and we get all the information related to social network. If the large amount of information to get it is too difficult.

W. Zhou, H. Jin, and Y. Liu proposed "Community discovery and profiling with social messages". This is use for get information related to Community discovery and profiling with social messages. The community's labels are existing but not yet developed or manifest, and each social document corresponds to an information sharing activity among the most probable community members regarding the most relevant community it is difficult to understand and implements.

T. Yang, R. Jin, Y. Chi, and S. Zhu proposed "Combining link and content for community detection: a discriminative approach". It is used to combining link and the content for a community detection. To alleviate the impact of irrelevant content attributes, we develop a discriminative model for content analysis. Discriminative LDA is that it is a supervised learning algorithm and cannot be applied directly to an unsupervised learning setup, which is the case of our problem.

Sneha Rani , Prem Ranjan, Sagar Kumar, Ciba Tembhare , Smita Khot "Twitter User Social Circle Detection Using Multi-View Network Structure". Vol. 5, Issue 12, December 2017 Website: www.ijircce.com. We can learn social circles in ego-networks which are based on multi-view network structure.

III SOFTWARE REQUIREMENT SPECIFICATION

User Classes and Characteristics

To design products that satisfy their target users, a deeper understanding is needed of their user characteristics and product properties in development related to unexpected problems that the user's faces every now and then while developing a project. The study will lead to an interaction model that provides an overview of the interaction between user characters and the classes. It discovers both positive and negative patterns in text documents as higher level features and deploys them over low-level features (terms). In proposed work is designed to implement above software requirement. To implement this design following software requirements and hardware requirements are used.

Software Requirements

- Operating System - Windows XP/7
- Programming Language - Java/J2EE
- Software Version - JDK 1.7 or above
- Tools - Eclipse
- Front End - JSP
- Database - Mysql

Hardware Requirements

- Processor - Pentium IV/Intel I3 core
- Speed - 1.1 GHz
- RAM - 512 MB (min)
- Hard Disk - 20GB

- Keyboard - Standard
- Mouse - Two or Three Button
- Monitor - LED Monitor

III. COMPARISON BETWEEN EXISTING SYSTEM AND PROPOSED SYSTEM

In an existing system, social learning circles in ego networks based on information about the structure of the network of multiple views can be classified according to the grouping methods of single view and six forms of friendship prediction based in the follower and following them they do not use NLP analysis.

In the proposed system, we can learn social circles in ego-networks which are based on multi-view network structure. We can classify information about the similar data or similar information. Here we can detect ego-network based on social circle. In an automatic social circle detection in ego-networks is a fundamentally important task for social network analysis. In this system, we know how to detect circles by leveraging multiple views of the network structure. For detection of this leveraging multiple views of the network structure, we crawl ego networks from Twitter and model them by six views, including user relationships, user interactions and user content. Friendship is the one view which is used in social circle detection. In this system characterizes the friend relation between alters by a similarity matrix where alters follow each other on Twitter. It is a most common view for social circle detection. Its only check the twitter users follow each other or not but it don't check the tweets of user. In our system we use Sentiment Classification of tweets using NLP (Natural Language Processing). It helps to find the accurate friend relation between alerts. We apply multi-view spectral clustering techniques to detect circles on these ego-networks. In this system we can used a modified multi-view spectral clustering techniques over a single-view clustering methods.

IV. ALGORITHM FOR RELEVANT FEATURE DISCOVERY

- **View Modeling**

  In this study, we crawl data from Twitter and employ classic techniques to model six views of its ego-net structures which are given below:

  1) Friendship
  2) Common Friend
  3) Reply
  4) Co-Reply
  5) Re-tweet
  6) Topic

- **K-means Algorithm**

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The main goal of the algorithm to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity. When start with *n* clusters, each containing one object and we will numbering the clusters 1 through *n*.

Pseudo code for k-means algorithm:

1) Compute the between-cluster distance $D(r, s)$ as the between-object distance of the two objects in r and s respectively, $r, s = 1, 2, ..., n$. Let the square matrix $D = (D(r, s))$. If the objects are represented by quantitative vectors we can use Euclidean distance.
2) Next, find the most similar pair of clusters r and s, such that the distance, $D(r, s)$, is minimum among all the pairwise distances.
3) Merge r and s to a new cluster t and compute the between-cluster distance $D(t, k)$ for any existing cluster $k \neq r, s$. Once the distances are obtained, delete the rows and columns corresponding to the old cluster r and s in the D matrix, because r and s do not exist anymore. Adding a new row and column in D similar to cluster t.
4) Repeat Step 3 a total of $n - 1$ times until there is only one cluster left.

V. SYSTEM ARCHITECTURE

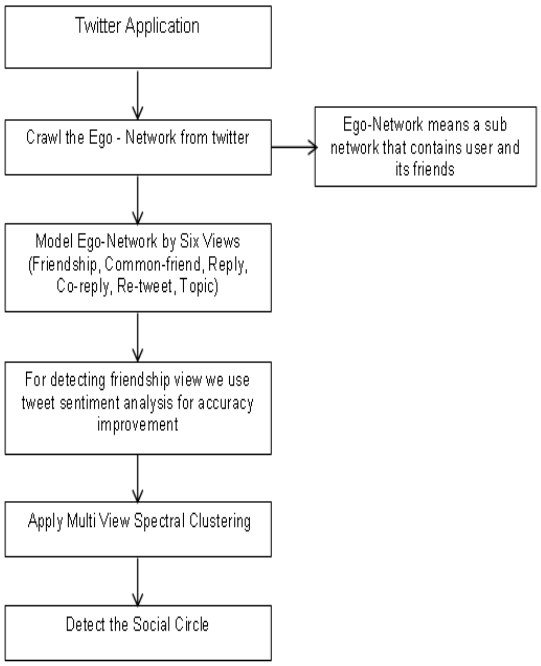

Figure 1: System Architecture

VI. MATHEMATICAL MODULE

The following terms shows in detail working of project.

Let us consider S as a system for Social Circle detection in Twitter Users.

S= {

INPUT:

Identify the inputs

F= {f1, f2, f3 ....., fn| F as set of functions to execute commands}

I = {i1, i2, i3|I sets of inputs to the function set}

O= {o1, o2, o3.|O Set of outputs from the function sets}

S= {I, F, O}

I = {Twitter user data extracted from Twitter Application, …}

O = {Output of Twitter user Social Circle,...}

F = {Functions implemented to get the twitter users data, model into six views and clustering}

- **Failures:**
  - Huge data can lead to more time consumption to get the information
  - Hardware failure
  - Software failure
- **Success:**

Successfully algorithm implementation and proper input

- **Space Complexity:**

The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

- **Time Complexity:**

  Check No. of patterns available in the datasets= n

  If (n>1) then retrieving of information can be time consuming. So the time complexity of this algorithm is $O(n^n)$.

Above mathematical model is NP hard because some time result is not accurate.

VII. EXPERIMENTAL SET UP AND RESULT TABLE

1. Result Table

User Id	Followings Count	Followers Count
1	20	13
2	9	17
3	15	8
4	21	25
5	25	18
6	23	21

Table 1: User's Followings and Followers Count

Above table shows that each users followings and followers count

2. Result Graph

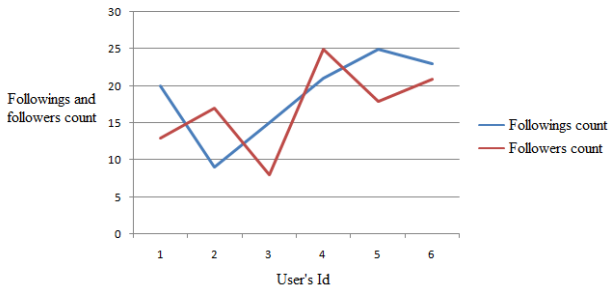

Figure 2: Result Graph of Followings and Followers Count

The above figure shows that the graph of each user's followings and followers count.

CONCLUSION

In this system, we can classified the information using a multi view clustering .In this technique we proposed to automatically detect social circles of an ego-net based on its multi-view network structure. We crawled and modeled Twitter ego-nets by six views, and showed multi-view spectral clustering outperformed the commonly adopted single-view clustering on these ego nets. We also showed, by treating sparse views as inherently incomplete ones and selectively transferring information across views, our modified multi-view clustering technique outperformed the standard multi-view clustering technique.

References

[1] HaiboHe,andEdwardo A. Garcia "Learning from Imbalanced Data".

[2] V. Roth "Probabilistic Discriminative Kernel Classifiers for Multi-class Problems".

[3] SHIGERU KATAGIRI, BIING-HWANG JUANG, FELLOW, IEEE,AND CHIN-HUI LEE "Pattern Recognition Using a Family of Design Algorithms Based Upon the Generalized Probabilistic Descent Method".

[4] Ji ZHU and Trevor HASTIE" Kernel Logistic Regression and the Import Vector Machine

[5] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning".

[6] Yuchun Tang, Member, IEEE, Yan-Qing Zhang, Member, IEEE, Nitesh V. Chawla "SVMs Modeling for Highly Imbalanced Classification".

[7] Gavin C. Cawley and Nicola L. C. Talbot, "Efficient Model Selection for Kernel Logistic Regression".

[8] Sneha Rani , Prem Ranjan, Sagar Kumar, Ciba Tembhare , Prof.Smita Khot "Twitter User Social Circle Detection Using Multi-View Network Structure".