International Journal of Engineering & Technology, 7 (4.36) (2018) 766-772

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Friendship Identification on Location Based Social Networks Using Ensemble Learning Technique

Shaik Mastan Vali[1*], P. Sujatha[2]

[1]Research Scholar, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies, Chennai.
[2]Associate Professor, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies, Chennai.
E-mail:suja.research@gmail.com
**Corresponding author E-mail:Shaikmastan77@gmail.com*

Abstract

The brisk development of client information and geographic area information in the area built long range interpersonal communication applications, it is logically troublesome for clients to quick and absolutely discover the data they need. With the expedient development and generally abuse of cell phone, area based informal organization (LBSN) has turned out to be one critical stage for some novel applications. The area data will help to find companion relationship, companion suggestion, network identification, and manual for excursion, notice merchandise et cetera. We separated client social relationship, registration separation and registration compose are the three most huge key highlights. After the component extraction, we connected Adaboost troupe classifier with different base classifiers to order. In view of the trial results, Adaboost with Rehashed Incremental Pruning to Deliver Mistake Decrease (RIPPER) gives the best outcome contrasted with other base classifiers.

Index terms: *Location-based social network, friendship prediction, behavioral analysis, ensemble classifier.*

1. Introduction

In the prior quite a long while, Online Informal organizations (OSNs, for example, Facebook, Myspace, Twitter, LinkedIn, Snapchat and Instagram have turned out to be to a great degree prevalent with the greater part billion clients around the world. As of late, Area based Informal communities (LBSNs) have likewise developed to enable clients to see where their companions are, to look area labeled substance inside their social diagram, and to meet others close-by.

The area based online informal organizations have seen taking off prominence, pulling in a huge number of clients [1]. Numerous area based online person to person communication applications, for example, Foursquare, Blend, Gowalla, and so forth [2], have offered stunning, novel and significant administrations for clients dependent on the sharing of area data by monitoring those sites. Contrasted and conventional online informal organizations, an unmistakable qualities of LBSNs is the concurrence of both on the web and disconnected social relations, as appeared in the figure 1. It bolster run of the mill online person to person communication offices, for example, making companions, sharing remarks and photographs. Likewise LBSN bolster disconnected social communications, for example, checking in spots [3, 4]. The LBSNs are heterogeneous informal communities which contains of both on the web and disconnected social connections. Meanwhile, vertices in LBSNs typically have numerous properties, for example, properties of a client which incorporate number of devotees, number of followings and number of registration and a scene may have the properties of class, number of registration and number of guests [5].

Whatever remains of this exploration paper is sorted out as pursues. In Area II, we examine the related work which are concentrating on the fellowship forecast on area based interpersonal organization. Segment III displays the proposed technique and Area IV gives the trial results and dialogs. At long last, we finish up in Area V.

2. Related works

With the fast improvement of area based informal organizations, look into on examining the social relationship has pulled in a great deal of consideration among the exploration network. In which companionship forecast has turned out to be one of the significant investigations on area based informal organization. A portion of the current fellowship expectation strategies which are utilizing of the registration dataset are portrayed in this segment.

Crandalla et al [8] proposed to surmise the social ties from the geographic occurrences. They have utilized probabilistic model to derive the companionship from area information shared on Flickr. The model considers both spatial and transient data. The proposed model found few co-events that can give the outcome in a high observational probability of a social tie. Nonetheless, they make a solid theory that every client has just a single companion which isn't on account of continuous applications. Cranshaw et al [10] creators offered to make utilization of machine learning classifier to derive the fellowship among two clients. The highlights incorporate the ones identified with areas and the interpersonal organization structures. Likewise, they proposed area entropy, which estimates the assorted variety of special guests of area. They have demonstrated positive connection between the entropy of the areas the clients visits and the quantity of social ties that the

client has in the system. Mama et.al [6] recommended companionship expectation calculation for recommender framework was proposed to foresee the relationship among various clients. They have utilized topological and recorded association data as the highlights to pass judgment on the presence and the kind of connection relationship. The straight relapse calculation and strategic relapse calculation was actualized for highlight mix. Quercia et.al [7] pondered an administered theme arrangement and connection expectation in twitter datasets. They utilized coordinated informal organization connect forecast approach dependent on subject model. They broke down hub semantic data orchestrates arrange hub qualities and auxiliary attributes for connection expectation issue. Xu-Rui and Wei-Li [9] proposed a calculation for fellowship forecast utilizing area based informal organizations. They have utilized help vector machine classifier for fellowship forecast on LBSN. They have separated the social relationship, registration separation and registration compose to build up the expectation show. They demonstrated that the chose characteristics assume imperative job in the companionship forecast issue. Consequently, we have additionally utilized these highlights in our proposed work.

3. Proposed friendship prediction system

The proposed kinship forecast framework begins with the gathering of open source area based interpersonal organization dataset. After the dataset accumulation, the framework plays out the conduct examination of area based informal community datasets to discover the importance of the dataset. From that point forward, in light of the essentialness, separated three noteworthy highlights. At that point the chose highlights are passed to outfit learning structure to anticipate the kinship. The general procedure of the proposed framework is given in the figure 1.

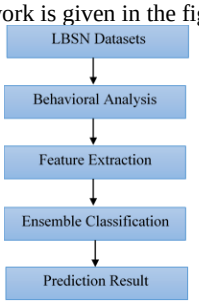


Fig. 1: Proposed friendship prediction system

A. Dataset description and behavioural analysis

Brightkite dataset

This undirected system contains user– client companionship relations from Brightkite, a previous area based interpersonal organization were client shared their areas. A hub speaks to a client and an edge shows that a companionship exists between the client spoken to by the left hub and the client spoken to by the correct hub. It was established in 2007 as a person to person communication site which enables clients to impart their area to their companions: it is accessible worldwide and it depends on making registration at spots, where clients can see who is adjacent and who has been there previously. Brightkite clients can build up shared companionship connections and they can drive their registration to their Twitter and Facebook accounts. This dataset speaks to a total preview of a well known area based administration in its underlying development stage. The factual portrayal of the brightkite dataset is given in table 1.

Table 1: Brightkite dataset statistical information

Characteristics	Descriptions
Duration	April 2008 to October 2010
Vertex type	User
Edge type	Friendship
Network Format	Undirected
Edge weights	Unweighted
Size (Number of nodes)	58,228 vertices (users)
Volume (Number of edges)	214,078 edges (friendships)
Average degree (Average no. of edges attached to a vertex)	7.3531 edges / vertex
Edges in largest weak connected component	212945
Nodes in largest weak connected component	56739
Clustering coefficient	11.1%
Check-ins	4,491,143

From the dataset, we have derived the following figures. The figure 2 shows the local clustering coefficient of the edges in the brightkite datasets. The figure 3 shows the frequency of the degree distribution. The connections are more dense in the range of 10^2 to 10^3.

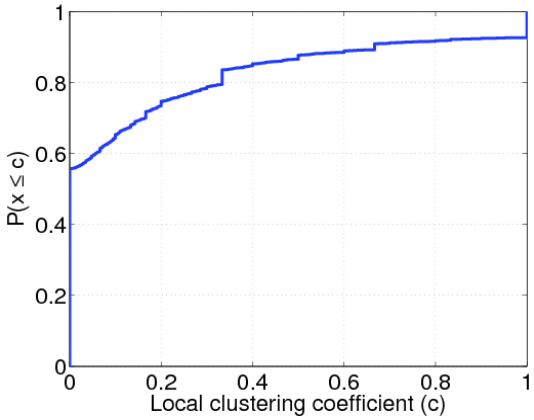
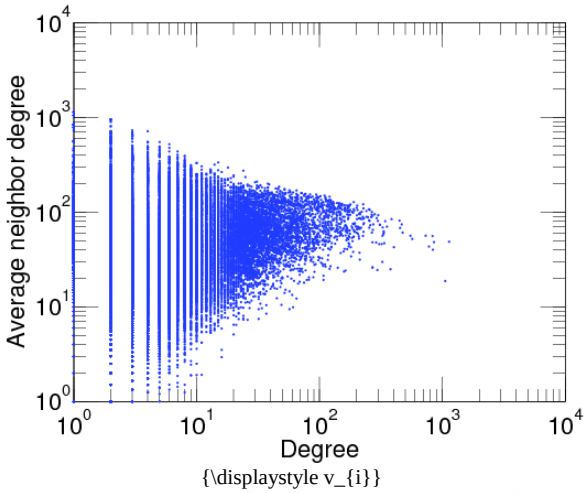


Fig. 2: Clustering coefficient distribution for Brightkite dataset

The local clustering coefficient of a node in a graph measures how close its neighbours are to being a complete graph. From this figure, one can infer that the local clustering coefficient of a node is calculated between the ranges of 0.5 to 1.0. The measures are close to 1, if every neighbour is connected to every other node within the neighborhood, and 0 if no node that is connected.



{\displaystyle v_{i}}

Fig. 3: Degree distribution of average neighbor degree of brightkite dataset

The normal neighbor degree availability is the normal closest neighbor level of hubs with degree d. It is broadly used to gauge the conditions between degrees of neighbor hubs in a system. From the figure, one can affirm that, there is a nearby degree dispersion between neighboring hubs in the scope of 101 to 102.

In view of the observational examination of the accessible datasets, we have inferred the three critical highlights for discovery of companionship in area based online informal organizations.

B. Feature extraction

These days, the principle route for fellowship forecast is to choose highlights that influence companionship and build up an expectation demonstrate. Distinguishing the critical highlights is a standout amongst the most imperative undertaking that chooses the forecast estimates, for example, accuracy and review. In this exploration, we have separated three noteworthy highlights, which are portrayed as underneath.

Social relationship

The social relationship is one of the vital variables for new fellowship arrangement [11]. To evaluate the social connection between the clients I and j, we utilized the accompanying prevalent hub likeness measures.

Common neighbors (CN) $Sim(i,j) = \left| adj_G(i) \cap adj_G(j) \right|$

Where $adj_{(i)}$ and $adj_{(j)}$ denote the set of neighbors of node i and j respectively.

Check-in Distance

The distance between two users i and j can be measured by the distance between their frequent movement areas. In this paper we used the following equation to characterize the feature.

$$a_{i,j} = \frac{\sum_{x=1}^{X} \frac{\sum_{y=1}^{Y} d(l_{ix}, l_{jy})}{Y}}{\sum_{x=1}^{X} \sum_{y=1}^{Y} d(l_{ix}, l_{jy})}$$

Where $l_{i1}, l_{i2}, \ldots, l_{iX}$ is check-in sequence for user i, $d(l_{ix}, l_{jy})$ is the distance between the x^{th} check-in location of user i and the y^{th} for user j. We define $a_{i,j}$ as the characteristic property of check-in distance among different users.

Check-in type

The check-in types of users reflect their own interest. Sometimes different users check in different places. We define $t_i$as the check-in type set of user i, t_j as the check-in type set of user j. Describing a_t as the feature of check-in type among different users, as described in the following equation.

$$a_t(i,j) = \frac{t_i \cap t_j}{t_i \cup t_j}$$

The extracted features are passed to the ensemble classification framework to predict the friendship.

C. Ensemble classification

Group learning technique prepare various classifiers rather than a solitary classifier and consolidate the aftereffect of various classifiers to yield a superior outcome than individual frail classifier [12]. It generally join countless classifiers trying to deliver a solid classifier. The term gathering is normally saved for strategies that make numerous speculations by utilizing a similar base classifier.

The frail classifiers are additionally called as the base classifiers which are normally created from base learning calculations that can be irregular woodland, JRip, bolster vector machine and innocent Bayesian multinomial. The real objective of outfit strategy is to consolidate the expectation of a few models that is worked with a grouping calculation to enhance the speculation or strength over a solitary classifier.

Determination of frail classifiers: The Adaboost calculation requires a gathering of powerless classifiers planned in advance. An individual powerless classifiers is chosen first. Its arrangement precision is moderately low.

We have chosen the base classifiers as arbitrary woodland, JRip, and bolster vector machine (SVM).

Irregular Woods

Irregular woods strategy is an uncommon sort of group technique which depends on the lead partition – and – vanquish plot utilized in the characterization mission [13].

As it is a troupe procedure, it amalgamates a gathering of delicate student to deliver well-manufactured less fatty which can arrange the information correctly.

It works by developing a large number of choice trees at preparing time and yielding the class that is the method of the classes yield by individual trees.

JRip

JRip famously known as Rehashed Incremental Pruning to Create Blunder Decrease (RIPPER) is one of the fundamental and most mainstream calculation [14]. It depends on the Incremental Diminished Mistake Pruning (IREP) calculation. It delivers an arrangement of standards, each one in turn, through two stages: development and pruning. In the development stage, the guidelines are at first vacant and conditions are added to the standards with the end goal to augment their inclusion of the preparation informationSupport vector machine

Support vector machine is a classification technique that has been emerged for the analysis of data for the classification process as well as regression [15]. However, it is mostly used for classification problems.

D. Experimental results and discussions

The various weak classifiers are identified and used in our proposed system are Random Forest, JRip, Support vector machine and Naïve Bayesian Multinomial. We have used these weak classifiers along with the boosting algorithm to improve the classification accuracy. To evaluate the performance of the proposed system, we have used the five essential measures namely, Precision, Recall, F-measure, Accuracy and False Positive Rate. The measures are derived as follows.

True Positive (TP) – The friendship prediction correctly identified as a friendship prediction.

False Positive (FP) – The non-friendship prediction wrongly identified as a friendship prediction.

True Negative (TN) – The non-friendship prediction correctly identified as a non-friendship prediction.

False Negative (FN) – The friendship_prediction wrongly identified as a non-friendship prediction.

Precision is a measure of what fraction of test data is detected as friendship prediction actually from the friendship prediction classes.

$$\text{Precision} = TP / (TP + FP)$$

Recall measures the fraction of friendship prediction class that was correctly detected.

$$\text{Recall} = TP/(TP+FN)$$

F-Measure is a measure of tests accuracy, which measures the balance between precision and recall.

$$F-measure = 2 * \frac{precision * recall}{precision + recall}$$

False Positive Rate (FPR) is the ratio of false positives (FP) over the sum of false positives (FP) and true negatives (TN).

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP+TN}$$

Accuracy (ACC) is calculated as the number of all correct predictions of friendship_prediction divided by the total number of instances.

$$\text{Accuracy (ACC)} = \frac{TP + TN}{(TP + FP + TN + FN)}$$

With the end goal to demonstrate that the chose highlights are significant to grouping calculations, we utilized adaboost gathering learning with three base characterization models for fellowship surmising, in particular Arbitrary Woodlands, JRip and SVM. We utilize Weka [16] to execute the gathering learning system. The highlights are chosen as social connection, registration separation and registration compose for every client, and Adaboost troupe classifier with different feeble classifiers are chosen for fellowship expectation. Likewise, we gained relating grouping results for the area based datasets. The order results are appeared in the table 2. Every one of the strategies are approved utilizing 10-crease cross approval strategy. Qualities underlined in strong relate to the best outcomes accomplished.

Table 2: Experimental Results of the Adaboost Ensemble Classifier Combined with Various Weak Classifier for Brightkite Dataset

Methods	Precision	Recall	F-Measure	Accuracy	FPR
Adaboost with Random Forest	0.985	0.985	0.985	98.472	0.014
Adaboost with JRip	**0.985**	**0.985**	**0.985**	**98.539**	**0.015**
Adaboost with SVM	0.982	0.982	0.982	98.207	0.021

From the experimental results, one can notice that the proposed method can predict friendship effectively due to the significant feature selection and discriminant classification selection. For the sake of comparison and better visibility, the measures such as precision, recall and f-measure are plotted using bar chart as shown in the figure 4.

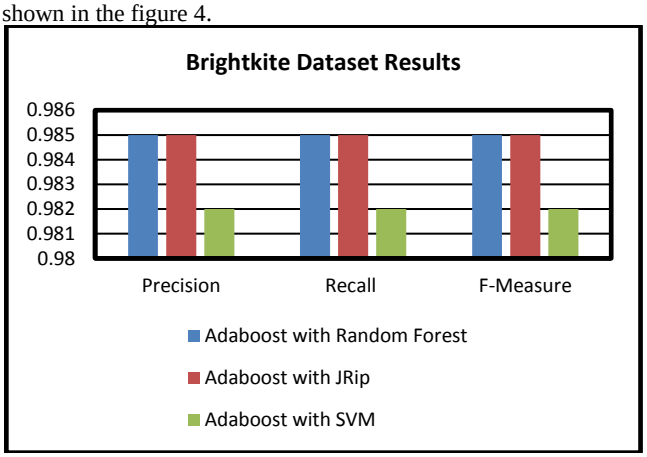


Fig. 4: Performance comparison

From the figure 4, one can affirm that the Adaboost with JRip strategy accomplishes higher accuracy, higher review and higher f-measure in both datasets contrasted with different strategies. One more deduction is that, very nearly three techniques accomplishes high review and low exactness, where high review identifies with a low false negative rate and high accuracy identifies with a low false positive rate. However, a high F-measure esteem can be accomplished if both accuracy and review esteems are high.

The f-measure estimation of a classifier is wanted to be as high as could reasonably be expected and it is near 1 shows that the calculation is promising.

4. Conclusion

Fellowship expectation in area based informal communities is valuable for different applications, for example, companion distinguishing proof, put suggestion, protection administration, and so on. In this examination paper, we proposed to anticipate the companionship by utilizing the most prominent open source area based informal organization datasets. In light of the exact investigation, recognized social relationship, registration separation and registration compose assumes a vital job in the distinguishing proof of companionship.

It is demonstrated that these highlights are effective for recognizing the kinship while applying to the Adaboost group classifier with JRip base student.

References

[1] Tan R, Gu J, Yang J, Lin X, Chen P & Qiao Z, "Designs of privacy protection in location-aware mobile social networking applications", *Journal of Software*, (2010), pp.298-309.

[2] Cho E, Myers SA & Leskovec J, "Friendship and mobility: User movement in location-based social networks", *KDD*, (2011), pp.1082-1090.

[3] Li N & Chen G, "Multi-Layered Friendship Modeling for Location-Based Mobile Social Networks", *Int'l. Conf. Mobile and Ubiquitous Systems: Computing, Networking and Services*, (2009), pp.1-10.

[4] Sadilek HK and Bigham JP, "Finding your friends and following them to where you are", *Fifth ACM International Conference on Web Search and Data Mining*, (2012), pp.723-732.

[5] Leskovec J, Huttenlocher D & Kleinberg J, "Predicting positive and negative links in online social networks", *Proceedings of the19th international conference on World Wide Web*, (2010), pp.641-650.

[6] Ma J, Xu H & Chen H, "Friendship prediction in recommender system", *Journal of National University of Defense Technology*, Vol.35, No.1, (2013), pp.163–168.

[7] Quercia D, Askham H & Crowcroft J, "Tweet LDA: supervised topic classification and link prediction in Twitter", *Proceedings of the 4th Annual ACM Web Science Conference*, (2012), pp.247–250.

[8] Crandalla DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D & Kleinberg J, "Inferring social ties from geographic coincidences", *National Academy of Sciences*, Vol.107, No.52, (2010), pp.22436–22441.

[9] Xu-Rui G, Li W & Wei-Li W, "An algorithm for friendship prediction on location-based social networks", *International Conference on Computational Social Networks*, (2015), pp.193-204.

[10] Cranshaw J, Toch E, Hone J, Kittur A & Sadeh N, "Bridging the gap between physical location and online social networks", *12th ACM International Conference on Ubiquitous Computing (UbiComp)*, (2010), pp.119–128.

[11] Wang H, Li Z & Lee WC, "PGT: Measuring mobility relationship using personal, global and temporal factors", *14th IEEE International Conference on Data Mining (ICDM)*, (2014), pp.570–579.

[12] Krawczyk B, Minku LL, Gama J, Stefanowski J & Woźniak M, "Ensemble learning for data stream analysis: A survey", *Information Fusion*, Vol.37, (2017), pp.132-156.

[13] Malekipirbazari M & Aksakalli V, "Risk assessment in social lending via random forests", *Expert Systems with Applications*, Vol.42, No.10, (2015), pp.4621-4631.

[14] D'Andrea E, Ducange P, Lazzerini B & Marcelloni F, "Real-time detection of traffic from twitter stream analysis", *IEEE transactions on intelligent transportation systems*, Vol.16, No.4, (2015), pp.2269-2283.

[15] Xiao C, Freeman DM & Hwa T, "Detecting clusters of fake accounts in online social networks", *8th ACM Workshop on Artificial Intelligence and Security*, (2015), pp.91-101.

[16] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P & Witten IH, "The WEKA data mining software: an update", *ACM SIGKDD explorations newsletter*, Vol.11, No.1, (2009), pp.10-18.