

Friend Recommendation System using Ensemble Voting and Distance Heuristics

Viswanathan Appakkudal Ramani, Dhariya Harsh Parikh, and Yashwanth Varre

¹Computer Science Department, Pace University, New York, NY, USA

²Computer Science, Essex County College, Newark, NJ, 07102, USA
{va99912n,dp08607n,yv28667p}@pace.edu

Friend Recommendation System helps to forecast future friends by analyzing social network using an ensemble of classifiers and weighted voting strategy. Social network analysis is a mechanism through which graphs are analyzed to find patterns and deduce relationships between multiple nodes. Three-person P1, P2, P3 form a length-3 following chain if a person P1 follows person P2 and person P2 follows P3. Given a length-3 following chain, it may make sense to recommend to P1 to follow person P3. It may make even more sense doing that if there are several 3-following chains between P1 and P3.

We considered the Facebook dataset provided by Stanford which contains 4039 nodes and 88234 edges and makes a directed network. We cleaned the graph by eliminating isolated nodes, and nodes with fake edges. We extracted features such as Jaccard Similarity, Adamic-Adar Index, Preferential attachment, Resource Allocation Index, and Common Neighbors [1]. Common neighbors find the number of mutual friends between two persons. Jaccard similarity compares members for two sets to see which members are shared and which are distinct. Resource Allocation Index is a measure that calculates the amount of resource that a node can send to the destination node via its neighbors. Preferential attachment measures the influence of a node over the other. Adamic-Adar index measures the number of shared links between two nodes. We modeled the features using an ensemble of ANN, Logistic regression, and Nave Bayes classifiers using a weighted voting mechanism [2]. Classifiers are weighted based on their accuracy score; the classifier with maximum accuracy gets the highest weights, and weights reduce for a classifier as the accuracy reduces. And, if the accuracy is less than a threshold, or if the classifier overfits/underfits the data, the classifier is eliminated, thereby keeping only significant classifiers.

References

- [1] Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., and Elovici, Y., Link Prediction in Social Networks Using Computationally Efficient Topological Features, in *Proceedings of IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third International Conference on Social Computing*, 2011, pp 73-80
- [2] Chebrolu, S., Abraham A., and Thomas, J. P., Feature deduction and ensemble design of intrusion detection system, *International journal of computers and security*, vol. 24, no. 4, pp. 295-307, 2005