Indonesian Journal of Electrical Engineering and Computer Science
Vol. 4, No. 1, October 2016, pp. 148 ~ 154
DOI: 10.11591/ijeecs.v4.i1.pp148-154

TreeNet Analysis of Human Stress Behavior using Socio-Mobile Data

B Padmaja*, V.V Rama Prasad, K.V.N Sunitha
Department of CSE, JNTUH, Hyderabad, Telangana, India
Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India
Department of CSE, BVRITH College of Engineering for Women, Hyderabad, Telangana, India
*Corresponding author, e-mail: b.padmaja@gmail.com

Abstract

Human behavior is essentially social and humans start their daily routines by interacting with others. There are many forms of social interactions and we have used mobile phone based social interaction features and social surveys for finding human stress behavior. For this, we gathered mobile phone call logs data set containing 111444 voice calls of 131 adult members of a living community for a period of more than 5 months. And we identified that top 5 social network measures like hierarchy, density, farness, reachability and eigenvector of individuals have profound influence on individuals stress levels in a social network. If an ego lies in the shortest path of all other alters then the ego receives more information and hence is more stressed. In this paper, we have used TreeNet machine learning algorithm for its speed and immune to outliers. We have tested our results with another Random Forest classifier as well and yet, we found TreeNet to be more efficient. This research can be of vital importance to economists, professionals, analysts, and policy makers.

Keywords: *Reality Mining, Social Network Analysis, Sensor data, Human Stress*

1. Introduction

The Internet and mobile phone technologies have become part our daily lives and it trans-formed the way of working and social interactions of modern societies. Modern day smart phone technologies contains a class of mobile applications, which supports social interaction among individuals, exploiting the growing power of smart phones to offer a variety of services.

Nathan Eagle and Alex Pentland from Media Laboratory, MIT coined the term ”Reality Mining” which is the collection and analysis of machine-sensed sensor data pertaining to human social behavior, with the goal of identifying predictable patterns of human behavior [1]. Human interactions are studied based on the usage of Smart phones and GPS systems and assemble a more complete picture of what individuals’ do, with whom they communicate and where they go.

Reality mining research shows the pattern of movement, known as behavior pattern, between the places where a person works, lives, eats and hangs out [2]. Social behavior of people has been shown to affect their obesity levels [3], reproductive fitness [4], productivity [5], software adoption [6], college choices, substance abuse, political affiliations [7], health characteristics [3, 8], spending behavior [9], happiness [10] and financial status [11]. Few reality mining experiments also focus on sleep and mood as they have significant public health impact with societal and financial effects [12].

In the last decades, many researchers in sociology have described stress behavior as a social construct by pointing out humans social influences which play a major role. And call log based social interaction patterns provide more predictive power on human stress.

The organization of the paper is as follows. We survey the related work followed by listing out the social network measures used in this paper. We present a TreeNet Gradient Boosting technique for characterizing stress behavior, and discuss the socio-mobile and stress features used to study the interconnections. Next, Social network features are listed out according to their priority and the influence of top predictor on the target class is shown through visualizations.

Received May 3, 2016; Revised August 26, 2016; Accepted September 11, 2016

2. Related Work

In modern era, stress is considered to be one of the major problems. Many times individuals are under stress due to deadline of projects and work and in long run high stress can be chronic. Several stress detection technologies may help people to better understand and relieve stress by increasing their awareness.

To understand multiple aspects of human behavior, in recent years many reality mining approaches have been used starting from individual to group to understand measures like personality, stress, interest level, spread of diseases and product adoption [3, 13]. Smart phones in particular have been used to study human mobility patterns both at a macro and an individual level [14].

Human social interactions data collected from mobile phones have been identified as potent stressors which often impact his/her social behavior [15, 16]. Previous research on human stress comes largely from neurobiology and few methods are based on physiological signals like blood pressure [17], heart rate [17], heart rate variability (HRV) [18], skin conductance [19, 20] and cortisol [21, 22]. Many stress assessing methods are based on surveys. Questions related to perceived stress scale (PSS) was used as an objective stress marker and it assessed to what degree a subject feels stressful in different situation.

Today we have many wearable devices and mobile phones containing various sensors to measure behavioral data in our day to day lives. This paper aims to use socio-mobile data to find human stress behavior using call log data collected from social-interactions through smart phones. At the same time, classification of human stress behavior using TreeNet analysis is a unique of its kind and the efficiency of results were verified after comparing with other popular machine learning algorithms like Random Forest classifier.

1. **Evaluation of Individual well-being**: Few behavior signals produced by smart phones had been correlated to the function of some major brain systems. Recent reality mining research on data streams offers direct assessment of cognitive and emotional states of individuals, perception of events, and information on their behaviors [23].

2. **Mapping Social Networks**: Reality minings capability for automatic mapping of social net-work is one of the important areas of research. A smart phone can sense and continuously monitor users call, SMS patterns, location information; and by using statistical analysis of these data, we can show different behavioral patterns based on users social relationship.

3. **Evaluation of Population Well-being**: Reality mining techniques are used to assess health conditions within a community. It shows people tend to be least deprived in regions where there is greater diversity of communication [24].

4. **Infectious Disease**: GPS and other sensing technologies provided by smart phones are used to easily track peoples movement to find out any disease spread by population or by physical proximity such as bird flu. And human behavior plays an important role in the spread of these diseases and improving the control over them [24].

5. **Mental Health**: Reality mining techniques assist in the early detection of psychiatric disor-ders such as depression, attention deficit hyperactive disorder (ADHD) etc. Reality mining data stream approaches allow direct, continuous, and long term assessment of health pat-terns and behaviors. The communication patterns and the frequency of communication with others of individuals, and their content and manner of speech are also the key signs of several psychiatric disorders [25].

6. **Behavioral Health Analytics**: Behavioral patterns collected from smart phone social sen-sors are used to improve the quality and to reduce the healthcare cost. Emerging mobile apps provide data of patients based on users location, call records, SMS records, app us-age which is then used for big data analytics to find deviations in an individuals daily activity to predict something wrong or suspicious even before an event occurs [26].

3. Social Network Analysis

A group of collaborating and/or competing individuals who are related to each other by one or more types of relations and are formally defined as a set of social actors, or nodes in a social network [27]. Social Network Analysis (SNA) is a technique which deals with the analysis of social networks to trace and understands the social relationships, and apply the inferred information among the members of the network. Many concepts from graph theory are adopted in

SNA. The reason is representation of Social Network through graph is the best way to analyze the relationships and interaction strengths among the actors or nodes [27-29]. Many graph tools have been developed to help researchers to visualize Social Networks.

In this paper, we have used the UCINET 6, a software package useful for Social Network analysis. In the literature of SNA, there are many metrics proposed to discover the characteristics of a Social Networks, like degree/size, density, different types of centralities, clustering coefficient, path analysis, flow, cohesion and influence, and other essential information which is obtained by various types of analysis [30]. In this section, for our analysis, we use the following metrics:

Degree: It is defined as the number of actors (alters) that an ego is directly connected to.

Farness: It is an aggregate of the weights of the shortest paths from ego / to ego to/ from all other nodes. If the social network is directed, then farness can be computed for sending and receiving information from alters, and the sum of geodesic distances from alters is called in-Farness and to other alters is called out-Farness.

Closeness Centrality: It is the reciprocal of farness. This metric is based on the notion of the average shortest path between a node and all the nodes in the graph. It is defined as the mean geodesic distance between an actor and all alters reachable from it. Closeness is an important measure which tells how long it will take information to spread from a given node to other nodes in the network. For a directed graph, in-closeness and out-closeness is calculated separately.

Structural Holes: These are the gaps / weaker connections between non redundant con-tacts or groups in the social structure. Individuals on either side of a structural hole circulate different flow of information. Structural holes are an opportunity to the broker to pass the information between people from opposite sides of the hole [31].

Ego Betweenness: It is the sum of ego's proportion of times that ego lies on the shortest path between each part of alters. If the alters are connected to each other not through ego, then the contribution of that pair is 0, for alters connected to each other only through ego, the contribution is 1. Similarly, alters which are linked to ego and one or more other alters, make the contribution 1/n, where n is the total number of nodes which connect the pair of those alters. Ego Betweenness is normalized by a function of the number of nodes in the ego network [32].

Proximal Betweenness: It measures the number of times a node occurs in a penultimate position on a geodesic. Let a_{jk} be the proportion of all geodesics linking vertex j and vertex k passing through vertex i, where i is the penultimate node on the geodesic, that is (i, k) is the last edge of the geodesic path. The proximal betweenness of a node i is the sum of all ajk where i, j and k are distinct.

Betweenness Centrality: It measures the position of a node and is defined as the number of times a node connects pairs of other nodes who otherwise would not be able to reach one another and plays the role of intermediary in the interaction between the other nodes.

Flow Betweenness: Let a_{jk} be the amount of flow between node j and node k which must pass through i for any maximum flow. The flow betweenness of node i is the sum of all a_{jk} where i, j and k are distinct and j < k. The flow betweenness is, therefore, a measure of the contribution of a node to all possible maximum flows.

Reach Centrality: It counts the number of nodes where each node can reach in k or less steps. For k = 1, this is equivalent to degree centrality. For directed networks, it calculates separate measures for out-Reach and in-Reach. In a social network, when we find the key individuals who are positioned well in the network, via them we can reach many people in just a few steps. This measure gives us a natural metric for evaluating each node.

Density: It is defined as the total number of ties divided by the total number of possible ties. Given a directed graph, Density = jEj jVj * (jVj 1), where jVj is the total number of vertices and jEj is the total number of the edges of the graph.

Degree Centrality: It is defined as the number of direct connections a node has with other actors or alters. A node with a high degree centrality acts as a hub in the network and for a directed network; degree centrality is the sum of in-degree and out-degree. It signifies activity or popularity of that node in the network due to large number of interactions with other nodes.

Clique: Clique can be defined as a sub-set of nodes where all probable pairs of nodes are directly linked to each other.

4. A Social Analysis of Human Stress Behavior

Our study is based on smart phone call logs dataset. This dataset contains continuous collection of call logs including the date, time and duration of call of individuals residing in a community. Here the call types are incoming, outgoing or missed call between individuals. Using this dataset, we build a phone communication network for the community where each node is an actor and each link is the type of calls made by them.

In this section, we present study of human stress behavior with the aim to elicit some useful information by using social network analysis and in particular the metrics shown in Sect. 3. As already mentioned in Sect. 3, we have used the UCINET 6 software package to compute the social network metrics. Here we propose the study of the human stress behavior from a social point of view, with the goal of extracting information from the dynamics of the relationships among the members of the network. We have used TreeNet algorithm, which typically generates thousands of small decision trees built in a sequential error-correcting process to converge to an accurate model given by Salford Systems (SPM 8.0).

4.1. Treenet Ananlysis of Call Log Dataset

TreeNet is a stochastic Gradient Boosting technique, a new machine learning approach which is good for classification and regression problems. It is built on CART trees and thus is fast, efficient, data driven, immune to outliers and invariant to monotone transformation of variables.

Procedure of TreeNet:

1. It begins with a very small tree as initial model containing as small as one split generating 2 terminal nodes.
2. But generally a model has 3-5 splits in a tree, rendering 4-6 terminal nodes.
3. After Tree, it computes residuals for this simple model for every record in data. Then it grows a second small tree to predict the residuals from the first tree.
4. Then it computes residuals from this new two-tree model and grow a third tree to predict revised residuals.
5. Further it rrepeats this process to grow a sequence of tree.

Every tree yields minimum one positive and one negative node. Red shows a relatively large positive and deep blue indicates a relatively negative node. Total score is obtained by identifying a relevant terminal node in every tree and summing the score across all trees in the model.

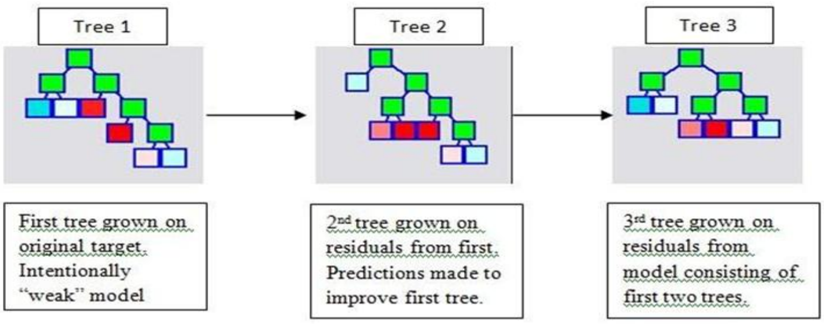

Figure 1. TreeNet modeling process

Figure 2. Number of Trees versus Average LogLikelihood (Negative) at N = 171

5. Result and Analysis

Our dataset, Avg Stress is our target variable and there are 17 predictors used to predict our target class STRESS(S). We have used train and test set ratio as 80:20, learn rate of 0.1, 200 as initial number of trees and cross entropy technique to determine number of trees optimal for our logistic model. Our model is based on a binary classification system, where a person is STRESSED (S) or NOT STRESSED (NS) is determined by whether the sign of the predicted outcome is positive or negative. This technique produces the smallest possible two-node tree in each stage. The default TreeNet uses a six-node tree, but in our model, the optimal likelihood and ROC models are attained when 171 trees are grown, that means at this level, the tree shows the best performance.

Figure 2 shows the average log likelihood (Negative) value of 0.653 for trees optimal N = 171 in our model.

Table 1 shows our optimal model, where area under the ROC curve shows a measure of overall model performance of 50% and an average Log-Likelihood of 65% to emphasize the probability interpretation of the model predictions for the target variable STRESS (S). Table 2 shows the importance of top 5 predictor variables for the target class. Figure 3 shows the top one predictor dependence variables for our target class STRESS (S).

Table 1. Optimal TreeNet Model for Call Log Dataset

By	Neg.	ROC	Misclass	Lift
Measure	0.65369	0.50000	0.36047	1.00000
N Trees	171	1	1	1

Table 2. Top 5 predictor variable for the target class STRESS(S)

Variable	Score
Hierarchy	100.00
Density	51.65
Infarness	45.37
Noutdwreach	44.19
Neigenvec	42.60

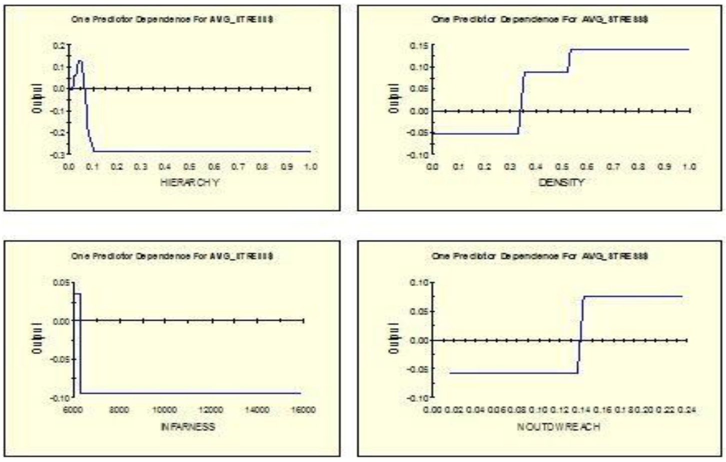

Figure 3. One predictor dependence for Avg Stress

6. Conclusion

In this paper, we presented the study of human stress behavior from a social point of view. The position of an individual/actor and the flow of information through that node in a social network preferably decide the stress behavior of that person. From our analysis, we identified that the hierarchy describes the nature of constraint on ego and it is an important property for shaping the nature of social interactions. Therefore, we conclude that an individual is more stressed based on its social hierarchy. Density is the second important criteria to decide the stress behavior of a node because it shows the extent to which information diffusion takes place among the nodes and the actors who have high level of social capital/social constraint. Farness is the third important criteria in deciding our target class because if a person lies in the shortest path from all other nodes then more information is passed through him/her and in the case of in-Farness, the ego receives far more information from other alters and hence experiences more stress. Similarly, the reach centrality of a node in a social network also decides the reachability of the node from other nodes. If the out-Reach of a node is high then the node is like a hub in the social network, and hence is more stressed.

References

[1] P Berka. *Reality Mining: Data Mining or Something else?*. 10th IWKM 2015. Bratislava, Slovakia. 2015.
[2] A Pentland, et al. Using Reality Mining to Improve Public Health and Medicine. *Strategy for the Future of Health.* 2009; 149: 93-102.
[3] NA Christakis, JH Fowler. Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. Little, Brown. 2009.
[4] SC Reed, JD Palm. Social Fitness versus Reproductive Fitness. *Science.*.1951; 113: 294-296.
[5] B Padmaja, et al. Connecting Productivity with Social Capital via Daily Mobile Phone Logs. *International Journal of Social Networking.* 2016; 5: 62-74.
[6] W Pan, et al. *Composite Social Network for Predicting Mobile Apps Installation.* Proceedings of the 25th Conference on Artificial Intelligence, AAAI-11. San Francisco, CA. 2011.
[7] JH Fowler, C Kam. Beyond the Self: Social Identity, Altruism, and Political Participation. *Journal of Politics.* 2007; 69(3): 813-827.
[8] Lane E Robert. The Loss of Happiness in Market Democracies. Yale University Press. 2001.
[9] VK Singh, et al. Classifying Spending Behavior using Socio-Mobile Data. *ASE Human Journal.* 2013.
[10] VK Singh, et al. *Predicting Happiness using Socio-Mobile Data.* Symposium on Success Ignite talk, ISQSS. Harvard University. 2013.
[11] W Pan, et al. *Fortune Monitor or Fortune Teller: Understanding the Connection between In-teraction Patterns and Financial Status.* IEEE International Conference on Social Computing. 2011.
[12] ST Moturu, et al. *Using Social Sensing to Understand the Links between Sleep, Mood, and Sociability.* IEEE Third International Conference on Social Computing (SocialCom). 2011: 208-214.
[13] A Pentland. Honest Signals: How They Shape Our World. MIT Press. 2008.
[14] N Eagle, et al. Network Diversity and Economic Development. *Science*. 2010; 328(5981): 1029-1031.
[15] A Sano, RW Picard. *Stress Recognition using Wearable Sensors and Mobile Phones.* 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013: 671-676.
[16] AK Beery, D Kaufer. Stress, Social Behavior, and Resilience: Insights from Rodents. *Neu-robiology of stress.* 2015; 1: 116-127.
[17] TG Vrijkotte, et al. Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. *Hypertension*. 2000; 35(4): 880-886.
[18] RK Dishman, et al. Heart Rate Variability, Trait Anxiety, and Perceived Stress among Physically Fit Men and Women. *International Journal of Psychophysiology.* 2000; 37(2): 121-133.
[19] J Hernandez, et al. Call Center Stress Recognition with Person-Specific Models. *Affective Computing and Intelligent Interaction.* 2011; 6974: 125-134.
[20] C Setz, et al. *Discriminating Stress from Cognitive Load using a Wearable EDA Device.* IEEE transactions on information technology in biomedicine: A publication of the IEEE Engi-neering in Medicine and Biology Society. 2010; 14(2): 410-417.
[21] SS Dickerson, ME Kemeny. Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychological bulletin.* 2004; 130(3): 355-391.
[22] M Van Eck, et al. The Effects of Perceived Stress, Traits, Mood States, and Stressful Daily Events on Salivary Cortisol. *Psychosomatic medicine.* 1996; 58(5): 447-458.
[23] A Pentland, et al. Improving Public Health and Medicine by use of Reality Mining. A whitepaper submitted for the Robert Wood Johnson Foundation. 2009.
[24] S Funk, et al. Modelling the Influence of Human Behavior on the Spread of Infectious Dis-eases: A Review. *Journal of the Royal Society Interface.* 2010; 7(50).
[25] J Torous, et al. Realizing the Potential of Mobile Mental Health: New Methods for New Data in Phychiatry. *Current Psychiatry Reports.* 2015; 17(8): 1-7.
[26] P Wlodarczak, et al. Reality Mining in eHealth. Lec-ture Notes in Computer Science. Springer, 2015; 9085: 1-6.
[27] S Wasserman, K Faust. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. 1994.

[28] A Marin, B Wellman. Social Network Analysis: An Introduction. Handbook of Social Net-work Analysis. Sage: Thousand Oaks. 2010.
[29] M Zhang. Social Network Analysis: History, Concepts, and Research. Handbook of Social Network Technologies. New York: Springer. 2010: 321.
[30] K Slaninova, et al. Analysis of Social Networks Extracted from Log Files. Handbook of Social Network Technologies. New York: Springer. 2010: 115-146.
[31] S Ronald Burt, et al. *Structural Holes versus Network Closure as Social Capital*. Social Capital: Theory and Research, Sociology and Economics. New York. 2001: 31-56.
[32] E Martin, SP Borgatti. Ego Network Betweenness. *Social Networks*. 2005; 27(1): 31-38.