

RESEARCH

Open Access



TreeNet analysis of human stress behavior using socio-mobile data

B. Padmaja^{1*}, V. V. Rama Prasad² and K. V. N. Sunitha³

*Correspondence:

b.padmaja@gmail.com

¹ Department of CSE,
Institute of Aeronautical
Engineering, Hyderabad,
India

Full list of author information
is available at the end of the
article

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 111,444 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 immunity 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

Introduction

The Internet and mobile phone technologies have become part of our daily lives and have transformed the way of working and social interactions of modern societies. Modern day smart phone technologies contain 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 decade, 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.

Related work

Social networks play a key role in diffusion of ideas, opinions and recommendations. It is the ability of a node in a social network to influence other nodes to accept or reject the information [13]. The Social Evolution experiment conducted by MIT closely tracks the day-to-day life of humans with mobile phones and it studies their adaptation of diet, exercise, political affiliation, depression and stress [14]. The frequency of interaction patterns extracted from communication logs are used to infer the strength of social ties and identify relationships.

Stress is a state of mental condition that everyone experiences in their daily life. In modern era, stress is one of the major causes for various health problems like heart disease, stroke, depression and cancer. 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. A recent study at Harvard and Stanford Business Schools shows high mortality rates relating to stress which causes hypertension, cardiovascular and other mental health diseases. And this leads to 120,000 deaths in America each year [15].

Alison Dillon et al. studied the effectiveness of smartphone games which utilize bio-feedback in reducing stress [16]. Biofeedback method detects various physiological signals such as heart rate, respiration, muscle activity, or skin temperature etc. from the user's body, and it helps them to gain control over them [17].

Amir Muaremi et al. worked on stress prediction using smartphones and wearable devices during working days and sleep and found an accuracy rate of 61 % at three different stress levels [18].

Rosalind W. Picard et al. studied on stress recognition using wearable biosensors which automatically measure stress in real-life environments [19].

Matteo Ciman et al. conducted various human-smartphone interaction analysis to evaluate the stress state of various smartphone users [20].

Akane Sano et al. studied physiological or behavioral marker for stress by the using mobile phones, a wrist sensor and social surveys to classify whether the participants were stressed or not [21].

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, 22]. Smart phones in particular have been used to study human mobility patterns both at a macro and an individual level [23].

Human social interactions data collected from mobile phones have been identified as potent stressors which often impact his/her social behavior [24]. Previous research on human stress comes largely from neurobiology and few methods are based on physiological signals like blood pressure, heart rate [25], heart rate variability (HRV) [26], skin conductance [27, 28] and cortisol [29, 30]. Many stress assessing methods are based on surveys. Questions related to perceived stress scale (PSS) were used as an objective stress marker which 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 method and the efficiency of the results were verified after comparing it with other popular machine learning algorithms like Random Forest classifier.

Evaluation of individual well-being

Few behavior signals produced by smart phones have 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 [31].

Mapping social networks

Reality mining's capability for automatic mapping of social network is one of the important areas of research. A smart phone can sense and continuously monitor user's call, SMS patterns, location information; and by using statistical analysis of this data, we can show different behavioral patterns based on user's social relationship.

Evaluation of population well-being

Reality mining techniques are used to assess health conditions within a community. It shows that people tend to be least deprived in the regions where there is greater diversity of communication [32].

Infectious disease

GPS and other sensing technologies provided by smart phones are used to easily track people's 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 in improving the control over them [32].

Mental health

Reality mining techniques assist in the early detection of psychiatric disorders such as depression, attention deficit hyperactive disorder (ADHD) etc. Reality mining data stream approaches allow direct, continuous, and long term assessment of health patterns 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 [33].

Behavioral health analytics

Behavioral patterns collected from smart phone social sensors are used to improve the quality and to reduce the healthcare cost. Emerging mobile apps provide data of patients based on user’s location, call records, SMS records, app usage which is then used for big data analytics to find deviations in an individual’s daily activity to predict something wrong or suspicious even before an event occurs [34].

Our research is based on smart phone call logs and it is the simplest way of maintaining a personal network diary of users on their phones with very little or almost no effort for long duration. This implements passive collection of data from smart phones and it helps us to gather complete information about user networks with no interruption to user’s daily activity. This dataset doesn’t suffer from cognitive/social biases in drawing network information.

Table 1 show the uniqueness of this research and also it explains the related work done by other researchers in terms of social granularity. Earlier most of the stress measurement work was done by collecting data from surveys, smart phone sensors and wearable devices. But in this paper, we have used call log data set of smart phone users to study

Table 1 Realted work done by other researchers on human stress

Work	Social granularity	Measure stress	Type of network data
Putnam	Communities	Survey	None
Lin	Individual, organization and communities	Surveys	Community
Kalimeri	Communities	Surveys	Sociometric badges and bluetooth
Burt	Communities	Surveys	Interview
Granovetter	Both on individuals and social groups	Surveys	Interview
Ciman	Group	Smart phone interactions	Human–smart phone interaction data
Dillon	Group	Smart phone games	Biofeedback
Gimpel	Group	Hardware, software sensors and surveys	Sensors
Anmol Madan	Group	Sensors and surveys	Smart phone sensors and socio-metric surveys
Amir	Group	Surveys and wearable devices	Sensors and wearable chest belts
Akane Sano	Group	Surveys and wearable devices	SMS, sensors and surveys
This work	Individual	Surveys	Call logs

human stress on daily basis over a period of time. Such method of finding stress behavior has never been explored before.

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 [35]. Social network analysis (SNA) is a technique which deals with the analysis of social networks to trace and understand 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 [35–37]. 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 [38]. 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 contacts 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 [39].
- *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 connecting the pair of those alters. N Ego Betweenness is normalized by a function of the number of nodes in the ego network [40].

- *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 a_{jk} 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 direct graph, $G = (V, E)$, Density is defined as: $\text{Density} = \frac{|E|}{|V| * (|V| - 1)}$, where $|V|$ is the total number of vertices and $|E|$ 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*: It can be defined as a sub-set of nodes where all probable pairs of nodes are directly linked to each other.

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. Table 2 shows sample data set containing two types of community users, SP and FA, who also participated in the survey. The data set contains call logs of both community and non community users, but we have preprocessed our data and concentrated only on community users.

Table 3 a sample of survey dataset containing 12,658 records on human stress from the same users and it is given in the range of 1–7. After preprocessing the data we computed Avg_Stress of each community user.

Table 2 Data set containing call logs of community users

Source	Destination	Date and time of call	Type of call
FA10-01-12	FA10-01-11	11/11/2010 0:19	Incoming
FA10-01-11	FA10-01-12	11/11/2010 0:19	Outgoing
FA10-01-11	FA10-01-12	11/11/2010 0:20	Outgoing
FA10-01-12	FA10-01-11	11/11/2010 0:20	Incoming
FA10-01-11	SP10-01-21	11/11/2010 0:20	Outgoing
SP10-01-21	FA10-01-11	11/11/2010 0:20	Incoming
FA10-01-07	2013346763	11/11/2010 0:47	Outgoing
FA10-01-07	2013346763	11/11/2010 0:48	Outgoing
FA10-01-07	2013346763	11/11/2010 0:48	Outgoing
FA10-01-07	2013346763	11/11/2010 0:49	Outgoing
FA10-01-07	2013346763	11/11/2010 0:50	Outgoing
FA10-01-07	2013346763	11/11/2010 0:51	Outgoing
FA10-01-07	2013346763	11/11/2010 0:52	Outgoing
FA10-01-07	2013346763	11/11/2010 0:53	Outgoing
SP10-01-25	6173195257	11/11/2010 2:50	Outgoing
SP10-01-52	8883002305	11/11/2010 4:42	Outgoing

Table 3 Survey data set on human stress from community users

Nodes	Date	Stress
FA10-01-16	1/13/2011	4
FA10-01-16	1/14/2011	4
FA10-01-16	1/15/2011	1
SP10-01-08	4/20/2011	5
SP10-01-21	3/5/2011	4
SP10-01-19	11/26/2010	4
SP10-01-21	3/7/2011	5
SP10-01-21	3/1/2011	4
SP10-01-19	11/27/2010	4
FA10-01-27	4/17/2011	5
SP10-01-21	3/9/2011	4
FA10-01-48	2/27/2011	5
SP10-01-13	3/6/2011	2
SP10-01-42	2/27/2011	7

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 “[Social network analysis](#)” section. As already mentioned, we have used the UCINET 6 software package to compute the social network metrics. Figure 1 shows the social network metrics computed using UCINET 6.0 package.

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).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U																
	Nodes	Avg_	Stress	Between	Degree	Closeness	Betweenness	Size	ER	Efficiency	Constraint	Hierarchy	o	Between	Indirects	Density	n	Closeness	Aut	Closenes	Eigenvec	n	Eigenvec	Flow	n	Flow	Bet	Let	ret	ve	me	s					
1	DUMMY35c	NS	98.11164	0.04724	0.4667488	0.019636063	6	4	0.666667	0.443603	0.093419	13.666667	0.363889	0.4	2.072709	0.812798	2.072709	0.018278	2.584846	26.3556	0.167337	3.325E-09	0	0	0	0	0	0	0	0	0	0	0				
2	DUMMY35c	NS	98.11164	0.00787	47.375	0	1	1	1	1	1	0	0	0.3333	0.812798	0.812798	2.35E-11	3.325E-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
3	DUMMY35c	NS	83	0.01575	0.1976017	0.005102041	2	2	1	0.5	0	2	0	0	1.910828	1.910828	5.5E-09	7.78E-07	160	1.015873	160	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	FA10-01-01	NS	83	0.00787	0.1822115	0	1	1	1	1	1	0	0	0	1.887923	1.887923	-3.8E-08	-5.38E-06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
5	FA10-01-02	S	1	0.00787	0.3542056	0	1	1	1	1	1	0	0	0	2.030293	2.030293	0.000325	0.0459363	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
6	FA10-01-03	NS	1	0.00787	0.3542056	0	1	1	1	1	1	0	0	0	2.030293	2.030293	0.000325	0.0459363	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	FA10-01-04	NS	42.75	0.0315	0.4633252	0.036530662	4	1.5	0.375	0.740451	0.021037	0.6666667	0.708333	0.8333	2.071346	2.071346	0.008371	1.183885	185.5	1.177778	1006.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	FA10-01-05	S	83	0.01575	0.2155859	0.010141093	2	2	1	0.5	0	2	0	0	1.933702	1.933702	5.29E-09	7.487E-07	316	2.006349	316	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	FA10-01-06	NS	83	0.01575	0.236875	0.015117158	2	2	1	0.5	0	2	0	0	1.956522	1.956522	1.19E-07	1.682E-05	468	2.971429	468	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	FA10-01-07	S	11	0.02362	75.800003	0.00015747	3	3	1	0.333333	0	6	0	0	0.812956	0.812956	1.86E-10	2.633E-08	10	0.063492	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	FA10-01-08	NS	1	0.00787	47.375	0	1	1	1	1	1	0	0	0	0.812798	0.812798	2.35E-11	3.325E-09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	FA10-01-11	S	4.666667	0.0315	0.3575472	0.000845549	4	1.5	0.375	0.740451	0.021037	0.6666667	0.708333	0.8333	2.030948	2.030948	0.000776	0.1097534	4.66667	0.02963	0.666667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	FA10-01-12	S	18.91667	0.02362	0.4031915	0.003084736	3	1.667	0.555556	0.839506	0.073738	1	0.555556	0.6667	2.051115	2.051115	0.0014	0.1980167	3.16667	0.020106	3.16667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	FA10-01-13	S	30.23907	0.04724	0.4596752	0.003439267	6	4	0.666667	0.4674	0.029502	11	0.652778	0.4	2.06422	2.06422	0.175913	24.877831	17.1786	0.10907	42.46936	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	FA10-01-14	NS	151.0122	0.07874	0.4761307	0.021696197	10	6.7	0.67	0.295553	0.026778	39.5	0.690303	0.3667	2.076125	2.076125	0.264108	37.350445	75.3933	0.478687	564.8315	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	FA10-01-15	S	1	0.00787	379	0	1	1	1	1	1	0	0	0	0.4306	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	FA10-01-16	S	1	0.00787	379	0	1	1	1	1	1	0	0	0	0.4306	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	FA10-01-17	S	1	0.00787	0.3522305	0	1	1	1	1	1	0	0	0	0.4306	2.029639	2.029639	0.005528	0.7817327	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	FA10-01-18	S	88.05357	0.02362	0.4146608	0.003333869	3	3	1	0.333333	0	6	0	0	2.056136	2.056136	0.043812	6.1960173	170	1.079365	164.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	FA10-01-19	S	1	0.00787	379	0	1	1	1	1	1	0	0	0	0.6	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	FA10-01-20	S	1	0.00787	379	0	1	1	1	1	1	0	0	0	0.6	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	FA10-01-21	S	241	0.0315	0.2617403	0.015047334	4	2.5	0.625	0.583333	0.046809	6	0.5	0.5	1.971522	1.971522	2.31E-05	0.0032652	470	2.984127	468	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	FA10-01-22	NS	3	0.02362	0.2380653	0.000871066	3	1	0.333333	0.925926	0	0	0.666667	1	1.947751	1.947751	3.9E-06	0.000551	2	0.012698	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	FA10-01-23	NS	2	0.01575	94.75	6.29882E-05	2	1	0.5	1.125	0	0	0.5	1	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	FA10-01-24	NS	1	0.00787	63.166668	0	1	1	1	1	1	0	0	0	1	0.793651	0.793651	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 1 Shows popular centrality measures like degree, closeness, betweenness, reach centrality, structural holes, and density for each ego and avg_stress is taken as a binary variable for classification

Stochastic Gradient TreeBoost algorithm

Stochastic Gradient_TreeBoost algorithm is a minor modification made to gradient boosting technique, where randomness is incorporated at each iteration of the algorithm [41]. At every iteration, a subsample of the training data set is drawn randomly (without replacement) and this subsample is used to fit the base learner and compute the model for the current iteration. Subsample size is a constant fraction f of the size of the training data set. Smaller values of f initiate randomness into the algorithm and it helps in preventing overfitting which acts as a kind of regularization.

Let $\{y^i, X^i\}_1^N$ be the entire training data sample and $\{r(i)\}_1^N$ be a random permutation of integers $\{1, 2, 3, \dots, N\}$. Then a random subsample $\tilde{N} < N$ is given by $\{y^{r(i)}, X^{r(i)}\}_1^{\tilde{N}}$.

Algorithm: Stochastic Gradient_TreeBoost

1. $F^0(x) = \arg \min_c \int_{i=1}^N \{y^i, c\}$
2. For $m = 1$ to M do:
3. $\{r(i)\}_1^{\tilde{N}} = \text{rand-perm}\{i\}_1^{\tilde{N}}$
4. $y_{r(i)m} = - \left[\frac{2 \{y_{r(i)}, F(x_{r(i)})\}}{2F(x_{r(i)})} \right]$
5. $\{R^m\}_1^{\tilde{N}} = L - \text{terminal nodetree}(\{y_{r(i)m}, x_{r(i)}\}_1^{\tilde{N}})$
6. $c^{im} = \arg \min_c \int_{x_{r(i)} f R^m} \{y_{r(i)}, F^{m-1}(x_{r(i)}) + c\}$
7. $F^m(x) = F^{m-1}(x) + o.c^{im} 1(x f R^m)$
8. endFor

The value $\tilde{N} = N$ introduces no randomness. Therefore the fraction $f = \tilde{N}/N$ causes more randomness in successive iterations thus improving overall randomness.

Treenet analysis 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.

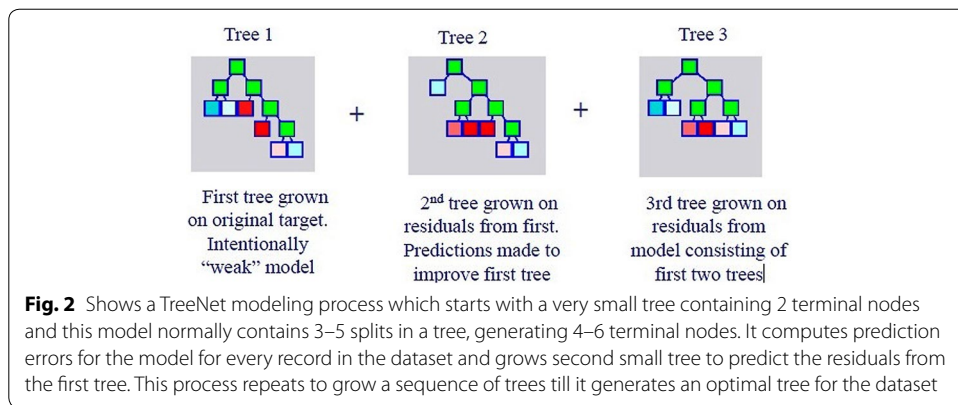
This model builds trees from several hundred to several thousand small trees and each tree depicts a small portion of the overall model. Finally, the model prediction is done by adding up all individual contributions. This model is similar to a long series expansion containing a sum of factors which becomes more accurate as the series expands. Figure 2 shows tree building process using gradient boosting technique.

$$F(X) = F_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \beta_3 T_3(X) + \dots + \beta_M T_M(X)$$

where each T_i is a small tree.

Procedure of TreeNet:

1. It begins with a very small tree with 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.

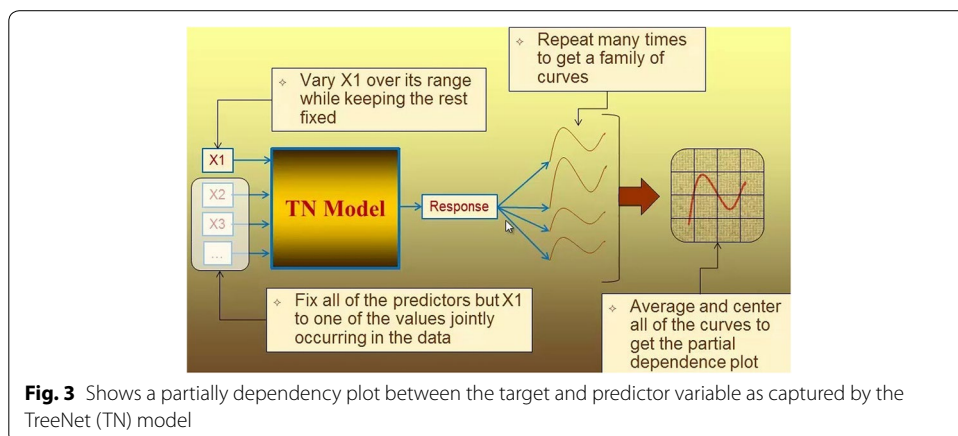


4. Then it computes residuals from this new two-tree model and grow a third tree to predict revised residuals.
5. Further it repeats this process to grow a sequence of tree.

Every tree yields minimum one positive and one negative node. Red shows a relatively large positive node 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.

TreeNet (TN) model is generally summarized in different ways like partial dependency plots, variable importance ranking, ROC curves and confusion matrix. Figure 3 shows a partially dependency plot between the target and predictor variable as captured by the model.

In 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 4 shows the average log likelihood (Negative) value of 0.653 for trees optimal $N = 171$ in our model.

TN model produces ROC curves which are unique for each record. This model allows records to be ranked from best to worst. Table 4 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).

TN model gives stable variable importance rankings after assessing the relative importance of predictors. Table 5 shows the importance of top 5 predictor variables for the target class.

Figure 5 shows the top two predictor dependence variables for our target class STRESS (S).

We have used another modeling engine for our dataset like RandomForest classifier and found the results as follow:

TN model generates the confusion matrix using an adjustable score threshold and this matrix is an indicators of model's false positive and false negative rates. Table 6 shows the sensitivity (true positive rate) or specificity (true negative rate) more in TreeNet than

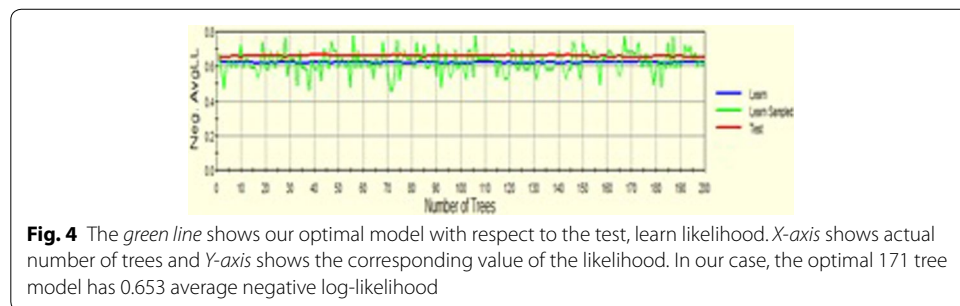


Fig. 4 The green line shows our optimal model with respect to the test, learn likelihood. X-axis shows actual number of trees and Y-axis shows the corresponding value of the likelihood. In our case, the optimal 171 tree model has 0.653 average negative log-likelihood

Table 4 Shows overall performance of our model

By	Neg. AvgLL	ROC	Misclass	Lift
Measure	0.65369	0.50000	0.36047	1.00000
NTrees	171	1	1	1

Table 5 Top 5 predictor variables for the target class Avg_stress

Variable	Score
Closeness	100.00
Hierarchy	93.27
Neigenvec	89.61
Density	80.97
Noutdwreach	72.86

Hypothesis testing

We have analyzed call log data set on daily basis over a period of time and we have formulated a set of hypothesis which is given in Table 7. These hypotheses we have tested using Person correlation analysis at 95 % confidence level, 0.05 significance level.

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. In a social network, when an ego is well-located on the communication paths linking pairs of others nodes, then that person is in central location. A centrally positioned ego can influence the group by restraining information in transmission. Such individuals are responsible for the maintenance of communications and they act as potential coordinators of group. If a node is more central in a network then the total distance is lower from all other

Table 7 Shows a set of formulated hypothesis for testing our results

H No	Hypothesis granularity	Hypothesis: description	Pearson correlation ρ value	Result: significant	Direction
H1	Overall period	People with more closeness centrality will have more stress than people with less closeness centrality	-0.0568	Yes	Negative
H2	Overall period	An individual's hierarchical position enables him/her to play the role of intermediary between other nodes, hence obtains more opportunities which makes his/her life more stressful	0.1214	No	Positive
H3	Overall period	The eigenvalue of an individual positions him/her as central nodes in a network and captures more information, hence feels more stressed	-0.0611	Yes	Negative
H4	Overall period	An individual in a denser network will have more potential connections/relationships with other nodes and promote more sharing of resources, hence individuals are more stressed	-0.0957	Yes	Negative
H5	Overall period	An individual who are well positioned in a network and reach to many people in few number steps are more stressful	-0.0542	Yes	Negative

nodes. Eigenvector centrality assigns high scores to the nodes based on the influence of a node in a network. 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 acts like a hub in the network.

From our analysis, we identified that the closeness, eigenvector and reach centrality are the major predictors for identifying stress behavior. Hierarchy describes the nature of constraint on ego which is a crucial property for shaping the nature of social interactions. Yet another important criterion is density because it shows the extent to which information diffusion takes place among the nodes. Therefore, an individual is more stressed based on its social hierarchy and density. We hope this study can be taken as benchmark for the further research in the field of social science.

Authors' contributions

BP made substantial contributions in formulating the concept, design, and carried out experimentation, analysis and interpretation of data; WVRP participated in concept formulation and statistical analysis. KVNS participated in design and helped in literature survey and drafting the manuscript. All authors read and approved the final manuscript.

Authors' information

Ms. B Padmaja is a faculty member in the Department of Computer Science, Institute of Aeronautical Engineering, Hyderabad, Telangana, India. She has received her B.Tech from North Eastern Regional Institute of Science and Technology (NERIST), Arunachal Pradesh, India in May 2001. She completed her M.Tech from School of IT, JNTUH, Hyderabad, India. Currently she is pursuing her research in Human Behavior Analysis using Reality Mining from JNTUH, Hyderabad. She is a member of ISTE and CSI.

Dr. V V Rama Prasad received the M.Sc (Tech.) degree in Electronic Instrumentation from Sri Venkateswara University, Tirupati, India in 1986 and M.E degree in Information Systems from BITS, Pilani, India in 1991. He has vast teaching and research experience and presently, he is working as Professor of Computer Science and Engineering department at Sree Vidyanikethan Engineering College, Tirupati, India. He was awarded the Ph.D. degree in Computer Science by J.N.T. University, Hyderabad, during 2007 for the thesis in Fractal Image Compression. He has published about 10 papers in National and International journals and presented several papers in National and International conferences. He has edited books, and refereed conferences. He is also a reviewer for 08 International Journals. His current areas of research interest include Computer Graphics, Image Processing, Computer Networks, Computer Architecture and Neural Networks. He is a member of IEEE, IACSIT, IAENG, ISTE and CSI.

Dr. K V N Sunitha completed her B.Tech ECE from Nagarjuna University, M.Tech Computer Science from REC Warangal. She was awarded Ph.D from JNTUH, Hyderabad in 2006. She has 24 years of Teaching Experience. Currently she is working as Founder Principal, BVRIT Hyderabad College of Engineering for Women, Hyderabad, India since August 2012. She has guided five Ph.Ds and currently guiding 8 research scholars. She has published more than 90 papers in International and National Journals and conferences. She is a reviewer for many national and International Journals. She is fellow of Institute of Engineers, Senior member for IEEE and International association CSIT, and life member of many technical associations like CSI and ACM.

Author details

¹ Department of CSE, Institute of Aeronautical Engineering, Hyderabad, India. ² Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, India. ³ Department of CSE, BVRITH College of Engineering for Women, Hyderabad, India.

Acknowledgements

We would like to acknowledge Neha for her assistance in editing and proofreading the paper.

Competing interests

The authors declare that they have no competing interests.

Received: 10 June 2016 Accepted: 11 October 2016

Published online: 16 November 2016

References

1. Berka P. Reality Mining: data mining or something else?. Bratislava: IWKM; 2015.
2. Pentland A, et al. Using reality mining to improve public health and medicine. *Stud Health Technol Inform*. 2009;149:93–102.
3. Christakis NA, Fowler JH. *Connected: the surprising power of our social networks and how they shape our lives*. Brown: Little; 2009.
4. Reed SC, Palm JD. Social fitness versus reproductive fitness. *Science*. 1951;113:294–6.
5. Padmaja B, et al. Connecting productivity with social capital via daily mobile phone logs. *Soc Netw*. 2016;5:62–74.
6. Pan W, et al. Composite social network for predicting mobile apps installation. *Proceedings of the 25th Conference on Artificial Intelligence. AAAI-11: San Francisco*; 2011.

7. Fowler JH, Kam C. Beyond the self: social identity, altruism, and political participation. *J Pol.* 2007;69(3):813–27.
8. Lane RE. *The loss of happiness in market democracies*. New Haven: Yale University Press; 2001.
9. Singh VK, et al. Classifying spending behavior using socio-mobile data. *ASE Hum J.* 2013.
10. Singh VK, et al. Predicting Happiness using socio-mobile data. symposium on success, Ignite talk, ISQSS, Harvard University; 2013.
11. Pan W, et al. Fortune monitor or fortune teller: understanding the connection between interaction patterns and financial status. *IEEE International Conference on social computing*; 2011.
12. Moturu ST, et al. Using social sensing to understand the links between sleep, mood, and sociability. *IEEE Third International Conference on social computing (SocialCom)*; 2011. p. 208–14.
13. Madan A, et al. Reality mining and personal privacy—will privacy disapper when social sensors learn our lives?. *Int Forum Manag Pers Electron Info.* 2009.
14. <http://realitycommons.media.mit.edu/socialevolution.html>.
15. Joel G, et al. The relationship between workplace stressors and mortality and health costs in the United States. *Manage Sci.* 2015;62(2):608–28.
16. Dillon Alison, et al. Smartphone applications utilizing biofeedback can aid stress reduction. *Front Psychol.* 2016;7:832.
17. Fond G, et al. Ketamine administration in depressive disorders: a systematic review and meta-analysis. *Psychopharmacology.* 2014;231(18):3663–76.
18. Muaremi Amir, et al. Towards measuring stress with smartphones and wearable devices during workday and sleep. *Springer J BionanoSci.* 2013;3(2):172–83.
19. <https://www.media.mit.edu/special/groups/advancing-wellbeing>.
20. Matteo C, et al. iSenseStress: accessing stress through human–smartphone interaction analysis. *9th International Conference on pervasive computing technologies for healthcare (PervasiveHealth)*; 2015.
21. Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. *Humaine Association Conference on affective computing and intelligent interaction*; 2013. p. 671–76.
22. Pentland A. *Honest signals: how they shape our world*. Cambridge: MIT Press; 2008.
23. Eagle N, et al. Network diversity and economic development. *Science.* 2010;328(5981):1029–31.
24. Beery AK, Kaufer D. Stress, social behavior, and resilience: insights from rodents. *Neurobiol Stress.* 2015;1:116–27.
25. Vrijkotte TG, et al. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension.* 2000;35(4):880–6.
26. Dishman RK, et al. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. *Int J Psychophysiol.* 2000;37(2):121–33.
27. Hernandez J, et al. Call center stress recognition with person-specific models. *Affect Comput Intell Interact.* 2011;6974:125–34.
28. Setz C, et al. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans Inf Technol Biomed.* 2010;14(2):410–7.
29. Dickerson SS, et al. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol Bull.* 2004;130(3):355–91.
30. Van Eck M, et al. The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosom Med.* 1996;58(5):447–58.
31. Pentland A, et al. Improving public health and medicine by use of reality mining. A whitepaper submitted for the Robert Wood Johnson Foundation; 2009.
32. Funk S, et al. Modelling the influence of human behavior on the spread of infectious diseases: a review. *J R Soc Interface.* 2010;7(50):1247–56.
33. Torous J, et al. Realizing the potential of mobile mental health: new methoids for new data in phychiatry. *Current Psychiatry Rep.* 2015;17(8):1–7.
34. Wlodarczak P, et al. Reality mining in eHealth. *Health Info Sci.* 2015;9085:1–6.
35. Wasserman S, Faust K. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press; 1994.
36. Marin A, Wellman B. *Social network analysis: an introduction*. Handbook of social network analysis. Thousand Oaks: Sage; 2010.
37. Zhang M. *Social network analysis: History, concepts, and research*. Handbook of social network technologies. New York: Springer; 2010.
38. Slaninova K, et al. Analysis of social networks extracted from log files. *Handbook of Social Network Technologies*. New York: Springer; 2010.
39. Burt RS, et al. Structural holes versus network closure as social capital. *New York: Social capital: theory and research, sociology and economics*; 2001.
40. Martin E, Borgatti SP. Ego network betweenness. *Soc Netw.* 2005;27(1):31–8.
41. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal.* 2002;38(4):367–78.