International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 2, Issue 2(February 2013), PP.35-39
www.irjes.com

Current Trends in Reality Mining

Jyoti More[1], Chelpa Lingam[2]
[1]*Lokmanya Tilak College of Engg., Koparkhairane*
(affiliated to Mumbai University), INDIA
[2]*Pillai's HOC College of Engineering & Technology, Rasayani*
(affiliated to Mumbai University), INDIA

Abstract: *We live in a technology driven society where, each one of us continuously leaves digital traces behind. Our mobile phones, for example, continuously sense our movements and interactions. This socio-geographic data could be continuously captured by millions of people around the world and promises to reveal important behavioral clues about humans. Mining patterns of human behavior from large-scale mobile phone data has deep potential impact on society. Reality Mining, pioneered by Nathan Eagle and Alex Pentland, (Massachusetts Institute of Technology (MIT)) is defined as the study of human social behavior based on wireless mobile phone sensed data. Reality mining is based on data collected by sensors in mobile phones, cars, security cameras, RFID ('smart card') readers, and others, all allow for the measurement of human physical and social activity. Applications of reality mining are in diverse fields like epidemiology, psychology, urban planning, security, marketing and even analysis of poverty. This paper attempts to overview and analyzes the current trends in reality mining. It also presents the current challenges in this field.*

***Keywords**: Reality Mining, Social Network Mining, Context aware computing*

I. INTRODUCTION

Reality Mining is defined as t*he study of human social behavior based on wireless mobile phone sensed data* by Nathan Eagle and Alex Pentland, (Massachusetts Institute of Technology (MIT). It is the collection and analysis of machine-sensed environmental data pertaining to human social behavior, with the goal of identifying predictable patterns of behavior.

Mobile phones are promising electronic devices as sensors due to their vast usage over the world on a daily continuous basis, and also due to the numerous types of sensors embedded in the device. The mobile phone has developed, due to its paramount nature, from a simple communication device to include many other tools such as a cameras, browsers, games, calendars, alarm clocks, and will surely continue to develop in the future. All of these forms of data can be analyzed to reveal details about human behavior.

Sensors are everywhere, continuously gathering information as we live our daily lives. Whether using email, the telephone, a bank machine, or even simpler activities such as driving, using a photocopy machine, and a camera, all of these activities leave traces of our behavior. Recently, the communication devices have been viewed from an engineering perspective as sensors, capturing data which scientists in many disciplines are very excited about. This data potentially impacts every one of us as researchers begin to study the possibilities of their use. Research using mobile phone data has mostly focused on location-driven data analysis, more specifically, using Global Positioning System (GPS) data to predict transportation modes to predict user destinations or paths, and to predict daily step count. Other location-driven tasks have made use of Global System for Mobile Communications (GSM) data for indoor localization or WiFi for large-scale localization . There are several works related to activity modeling from location-driven phone sensor data. CitySense is a mobile application which uses GPS and WiFi data to summarize “hotspots” of activity a city, which can then be used to make recommendations to people regarding, for example, preferred restaurants. Applications to society as a whole are being investigated in terms of epidemiology and psychology, urban planning, security, and even in the analysis of poverty.

This paper focuses on the possibilities, scope and challenges related to reality mining. A brief review has been carried out for this purpose.

II. SOCIAL NETWORKS

A Social network is defined as a set of actors (individuals) and the ties (relationships) among them. Important research problems include the study of social networks’ structural properties (such as community detection and evolution), user properties (such as reputation and trustworthiness), and user social relations (including influence and trust). Social networks are either explicitly specified, such as a Facebook friends list, or implicitly inferred from social interactions such as email or mobile phone communications. Important research problems include the study of social networks' structural properties (such as community detection and evolution), user properties (such as reputation and trustworthiness), and user social relations (including influence and trust).

2.1 Social Networks as Graphs

Social networks are naturally modeled as undirected graphs(fig 1). The entities are the nodes, and an edge connects two nodes if the nodes are related by the relationship that characterizes the network. If there is a degree associated with the relationship, this degree is represented by labeling the edges.

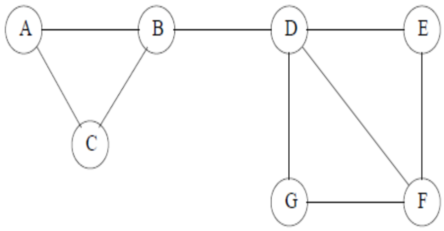

Fig 1.: Example of a small social network represented as graph

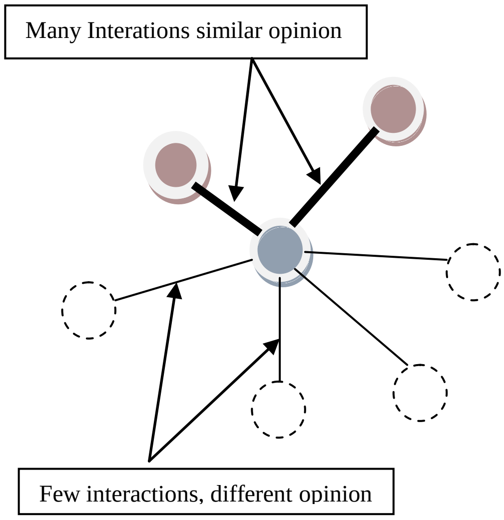

Fig 2.: Using Social Network Mining to estimate strengths of social relations

2.2 Social network mining

In social network mining, we apply data mining algorithms to study large-scale social networks. Social network mining has attracted a lot of attention for many reasons. For example, studying large social networks allows us to understand social behaviors in different contexts. In addition, by analyzing the roles of the people involved in the network, we can understand how information and opinions spread within the network, and who are the most influential people . In addition, since social network users may receive too much information from time to time, social network mining can be used to support them by providing recommendations and filtering information on their behalf.

III. CONTEXT AWARE COMPUTING

Context is a combination of any information that can be sensed or received by an entity which is useful to catch events and situations.Context-aware computing uses information about an end user's or object's environment, activities, connections and preferences to improve the quality of interaction with that end user or object. A contextually aware system anticipates the user's needs and proactively serves up the most appropriate and customized content, product or service. Applications that use context, whether on a desktop or in a mobile or ubiquitous computing environment, are called context-aware.

There are four catagories of context aware applications:

- Proximate Selection: *Presents information, which is selected considering some context to ease a choice.*

- Automatic Contextual Reconfiguration: *Current context automatically leads to new information.The entity creates new bindings to context resources.*
- Contextual Information and Commands: *Information and commands are shown / executed manually and adapted to the current situation.*
- Context-Triggered Actions: *The current context leads an application to start a process automatically [2]*

IV. CURRENT TRENDS

The Pioneers of reality Mining, the researchers of MIT, Nathan Eagle & Alex (Sandy) Pentland *et.al [4]* introduced a system for sensing complex social systems with data collected from 100 mobile phones over the course of 9 months. They demonstrated the ability to use standard Bluetooth-enabled mobile telephones to measure information access and use in different contexts, recognize social patterns in daily user activity, infer relationships, identify socially significant locations, and model organizational rhythms. The major findings included Behavior Prediction, Relationship Inference and Computational Social Science. Further, N. Eagle, A. Pentland, and D. Lazer *et. al[5]* analyzed 330,000 hours of continuous behavioral data logged by the mobile phones of 94 subjects, and compared these observations with self report relational data. The information from these two data sources is overlapping but distinct, and the accuracy of self report data is considerably affected by such factors as the recency and salience of particular interactions. A new method is proposed for precise measurements of large-scale human behavior based on contextualized proximity and communication data alone, and identify characteristic behavioral signatures of relationships that allowed to accurately predict 95% of the reciprocated friendships in the study. Using these behavioral signatures it could be possible to predict, individual-level outcomes such as job satisfaction. The dataset built during these experiments is still being used by many researchers for their experimentation.

Further the field was explored by many researchers for various applications. Huiqi Zhang; Dantu, R.; Cangussu, J.W. *et.al [6]* proposed in their paper, a socioscope model for social-network and human-behavior analysis based on mobile-phone call-detail records. Because of the diversity and complexity of human social behavior, no one technique will detect every attribute that arises when humans engage in social behaviors. They used multiple probability and statistical methods for quantifying social groups, relationships, and communication patterns and for detecting human-behavior changes. They proposed a new index to measure the level of reciprocity between users and their communication partners. This reciprocity index has application in homeland security, detection of unwanted calls (e.g., spam), telecommunication presence, and product marketing.

Zhang, Huiqi; Dantu, Ram *et. al [7]* proposed that the social-tie strengths of person-to-person are different one another even though they are in the same group. In this paper the researchers investigated the evolution of person-to-person social relationships, quantify and predict social tie strengths based on call-detail records of mobile phones. They proposed an affinity model for quantifying social-tie strengths in which a reciprocity index is integrated to measure the level of reciprocity between users and their communication partners. Since human social relationships change over time, they map the call-log data to time series of the social-tie strengths by the affinity model. Then they used ARIMA model to predict social-tie strengths. Farrahi, K.; Gatica-Perez, D. *et. al.[8]* suggested that human interaction data, or human proximity, obtained by mobile phone Bluetooth sensor data, can be integrated with human location data, obtained by mobile cell tower connections, to mine meaningful details about human activities from large and noisy datasets. They propose a model, called bag of multimodal behavior, that integrates the modeling of variations of location over multiple time-scales, and the modeling of interaction types from proximity. They further demonstrate the feasibility of the topic modeling framework for human routine discovery by predicting missing multimodal phone data at specific times of the day.

Simoes J., Magedanz, T. *et. al. [9]* proposed a work by combining social network analysis, reality mining techniques and context-aware systems. This work provides an architecture and ground steps for understanding and predicting human behavior and preferences within one of the most promising business models of the future: “Advertising”. Furthermore, it shows how user related data (context) can be securely managed and exposed to 3rd party providers, taking into account user context-aware privacy settings. The presented concepts are then realized in a prototype, which evaluates the basic functionalities previously described. Xu Yang, Yapeng Wang, *et. al.[10]* studied data mining for social network analysis purpose, which aims at finding people's social network patterns by analyzing the information about their mobile phone usage. In this research, the real database of MIT's Reality Mining project is employed. The classification model presented in this project provides a new approach to find the proximity between users - based on their registration frequencies to specific cellular towers associated their working places. K-means Algorithm is applied for clustering.

Huiqi Zhang; Dantu, R. *et. al.[11]* proposed the Bayesian inference model to calculate the willingness level of the callee to accept calls. Before making a call, the caller may use the willingness calculator to find out whether the callee is available. Based on this level the user can make a decision whether to make a call. They used time of the day, day of the week, talk-time and location for calculating the willingness level.

Michal Ficek, Lukas Kencl [12] proposed that the data captured from a live cellular network with the real users during their common daily routine help to understand how the users move within the network. Unlike the simulations with limited potential or expensive experimental studies, the research in user-mobility or spatio-temporal user behavior can be conducted on publicly available datasets such as the Reality Mining Dataset. These data have been for many years a source of valuable information about social interconnection between users and user-network associations. However, an important, spatial dimension is missing in this dataset. In this paper, the researchers present a methodology for retrieving geographical locations matching the GSM cell identifiers in the Reality Mining Dataset, an approach based on querying the Google Location API. A statistical analysis of the measure of success of locations retrieval is provided. Further, they presented the LAC-clustering method for detecting and removing outliers, a heuristic extension of general agglomerative hierarchical clustering. This methodology enables further, previously impossible analysis of the Reality Mining Dataset, such as studying user mobility patterns, describing spatial trajectories and mining the spatio-temporal data.

Zhenhui Li, Cindy Xide Lin *et. al.[13]* emphasized that Spatio-temporal data collected from GPS have become an important resource to study the relationships of moving objects. While previous studies focus on mining objects being together for a long time, discovering real-world relationships, such as friends or colleagues in human trajectory data, is a fundamentally different challenge. For example, it is possible that two individuals are friends but do not spend a lot of time being together every day. However, spending just one or two hours together at a location away from work on a Saturday night could be a strong indicator of friend relationship. Based on the above observations, in this paper the researchers aim to analyze and detect semantically meaningful relationships in a supervised way. That is, with an interested relationship in mind, a user can label some object pairs with and without such relationship. From labeled pairs, it is learnt that time intervals are the most important ones in order to characterize this relationship. These significant time intervals, namely T-Motifs, are then used to discover relationships hidden in the unlabeled moving object pairs.

Xiaowen Dong, Pascal Frossard *et. al.[14]* proposed that mobile phone data provides rich dynamic information on human activities in social network analysis. In this paper, the researchers represent data from two different modalities as a graph and functions defined on the vertex set of the graph. They propose a regularization framework for the joint utilization of these two modalities of data, which enables them to model evolution of social network information and efficiently classify relationships among mobile phone users.

From the above survey it is observed that the reality mining still has a large potential of getting explored and contribute to ubiquitous computing field. It is one aspect of digital footprint analysis. Much more experimentations could be carried out to analyze the social ties and predict human behavior that could be helpful in exploring parallel universes of opinion mining, emotional mining, social network mining, etc.

V. SIGNIFICANCE OF THE STUDY

- **Security:** Reality mining can be a great tool to track terrorists as mobile phone networks can identify unusual patterns of movement and communication. GPS-enabled mobile phones and tracking devices are installed on commercial vehicles to monitor traffic conditions. It facilitates in tracking of real-time traffic congestion data.
- **Business:** It can help companies to boost inter-office cooperation. Mining Task-Based Social Networks can be used to explore Collaboration in Software Teams. Event planners who manage multi-million dollar conventions and conferences can avail the data and make the best out of it. Telecom companies can analyze the service usage and can enhance customer service.
- **Healthcare:** Reality mining has the ability to contribute immensely towards healthcare. By gathering health related information through mobiles, they can predict disease outbreaks. With the aid of audio or motion sensors, changes in the nervous system can be deducted and this information could be used to screen depression.
- **Viral marketing**, **viral advertising**: these are the buzzwords referring to marketing techniques that use pre-existing social networks and other technologies to produce increases in brand awareness or to achieve other marketing objectives (such as product sales) through self-replicating viral processes, analogous to the spread of viruses or computer viruses. It can be delivered by word of mouth or enhanced by the network effects of the Internet and mobile networks. Viral marketing may take the form of video clips, interactive Flash games, advergames, ebooks, images, text messages,email messages, or web pages.
- **Digital Footprinting**: Uncovering or tracing the tourists with User-Generated Content[3]

VI. CURRENT CHALLENGES

The need for effective methods and mathematical models for analysis becomes crucial in order to make good use of the sources. In machine learning, algorithms have been developed to recognize complex patterns and make intelligent decisions based on data. Traditional machine learning models are recognized as useful tools for large- scale data analysis. They have been used in the domain of human behavior analysis, though their limitations with new types of data and human-centric questions become apparent. For example, many of the traditional machine learning models is supervised, requiring training data which is often impossible or illegal to collect on human subjects. Other specifications related to human-centric data include the multimodal aspect, the noise, the massive quantity, and the complex questions of interest.

More specifically, data collected by mobile phone sensors include many types, ranging from GPS, Bluetooth, accelerometer, to voice features. Each of these sensors may be sampled with varying frequencies, each has varying timescales and differing characteristics, and each has its own sources of noise.

Although many basic conceptual questions remain unresolved, the major roadblock in defining the fundamental predictability limits for technosocial systems is their sensitivity and dependence on social adaptive behavior.

Addressing these problems involves tackling three major scientific challenges. The first is the gathering of large-scale data on information spread and social reactions that occur during periods of crisis. This is not presently out of reach, via largescale mobile communication databases (such as mobile telephones, Twitter logs, and social Web tools) operating at the moment of specific disaster or crisis events. The second challenge is the formulation of formal models that make it possible to quantify the effect of risk perception and awareness phenomena of individuals on the technosocial network structure and dynamics. The third challenge is that of maintaining privacy i.e. to do privacy –preserving-mining.

REFERENCES

Books:

[1] Earl Cox , "*Fuzzy Modeling and Genetic Algorithms for Data mining and Exploration*", Morgan Kaufmann Publishers/ Elsevier

[2] B. Schilit, N. Adams, R. Want, "Context-Aware Computing Applications" Proceedings of Workshop on Mobile Computing Systems and Applications, 1994

Theses:

[3] Katayoun Farrahi, *"A Probabilistic Approach to Socio-Geographic Reality Mining"* THESIS No5018 (2011) *submitted to the* Faculty of Science and Technology Engineer École Polytechnique Fédérale de Lausanne *to obtain the degree of Doctor of Science*

Journal Papers:

[4] Nathan Eagle & Alex (Sandy) Pentland, "Reality mining: sensing complex social systems" *Pers Ubiquit Comput (2006) 10: 255–268, Springer-Verlag London Limited 20*05

Proceedings Papers:

[5] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure using Mobile Phone Data," *Proceedings of theNational Academy of Sciences (PNAS), vol. 106, no. 36, pp.15274–15278, September 2007*

[6] Huiqi Zhang; Dantu, R.; Cangussu, J.W. , " Socioscope: Human Relationship and Behavior Analysis in Social Networks", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Volume: 41, Issue: 6, Publication Year: 2011 , Page(s): 1122 – 1143*

[7] Zhang, Huiqi; Dantu, Ram, "Predicting social ties in mobile phone networks", *IEEEInternational Conference on Intelligence and Security Informatics(ISI)2010, pages: 25-30*

[8] Farrahi, K.; Gatica-Perez, D. , "Probabilistic Mining of Socio-Geographic Routines From Mobile Phone Data", *IEEE Journal of Selected Topics in Signal Processing, Volume: 4, Issue: 4 Publication Year: 2010 , Page(s): 746 – 755*

[9] Simoes, J. Magedanz, T., "Can you predict human behavior?", 14 th International Conference on Intelligence in Next Generation Networks (ICIN), 2010, *Page(s): 1 – 6*

[10] Xu Yang; Yapeng Wang; Dan Wu; Ma, A., "K-means based clustering on mobile usage for social network analysis purpose", *IEEE, 6 th International Conference on Advanced Information Management and Service(IMS), 2010, Page(s)-223-228*

[11] Huiqi Zhang; Dantu R., "Quantifying the presence of Phone Users", *5 th IEEE Conference on Consumer Communication and Networking Conference, 2008, Page(s): 883 - 887*

[12] Michal Ficek, Lukas Kencl, "Spatial Extension of the Reality Mining Dataset", *Proceedings of IEEE 7 th International Conference on Mobild Adhoc and Sensor Systems(MASS), 2010*

[13] Zhenhui Li, Cindy Xide Lin, Bolin Ding, Jiawei Han, "Mining Significant time intervals for relationship detection" *Proceedings of the 12 th international conference on Advances in spatial and temporal databases, 2011*

[14] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, Nikolai Nefedov, "A regularization framework for mobile social network analysis" *IEEE International Conference on Acoustics, Speech and Signal Processing, 2011*