International Journal of Advanced Research Methodology in Engineering & Technology, ISSN 2456 6446
Volume 1, Issue 3, May 2017

SOCIAL SET ANALYSIS FOR BIG DATA

S. Nandhini
UG Scholar, Department of CSE
Knowledge Institute of Technology
Salem, India

T. Karthikeyan
Assistant Professor, Department of CSE
Knowledge Institute of Technology
Salem, India
tkcse@kiot.ac.in

T. Rubika
UG Scholar, Department of CSE
Knowledge Institute of Technology
Salem, India

Abstract - In Computational Social Science, the analytical approaches can be defined by four paradigms namely social network analysis (theory of graph), social simulations (agent-based modeling and cellular automation), text analysis (extracting and classifying the information) and social complexity analysis (analysis complex systems). However for societal and organizational analysis, for analyzing big social data, there exists no approach to model, conceptualize, analyze, predict and explain social media interactions. To overcome this limitation, this paper undergoes another comprehensive approach to big social data analytics called **Social Set Analysis**. It comprises of theory of social data, philosophies of computational work, formal and conceptual models of social data and generative framework for combining the big social data with the societal and organizational data sets. It also undergoes four case studies covering the range of descriptive, predictive, prescriptive and visual analytical methods for analyzing the big social data sets. Current drawbacks of set-theoretical approach and the future implementations are discussed in this paper.

Index terms: Social Set Analysis, big social data, predictive, descriptive, prescriptive and visual analytical methods.

I. INTRODUCTION

Social media are fundamentally scalable communication technology that brings the internet-based to the interactive dialogue platform[1]. On the "demand-side", increasing number of customers and users are moving towards various types of social media for searching information and for making decisions related to products and public services [2]. On the "supply-side", social business and enterprises are used to describe birth of public institutions and private enterprises that uses the social media channels for improving their operation efficiencies, organizational effectiveness and empowering their employees. The usage of social media produces larger volumes of unstructured data called "Big Social Data". However, there are unsolved critical problems that are related to how the big data is integrated with the real datasets of organizations. To overcome this issue, this paper presents a research project to design, implement and examine the set-theoretical approach using the large amount of data from twitter, Facebook and many other social media channels. Specifically, this paper presents a research works in the domains of Computational social science and Data science with the practical applications.

II. CONCEPTUAL FRAMEWORK

A. Theory of Social Data

For the purposes of practically analyzing and examining big social data, we argue that any particular theory of social data must support mathematical and conceptual modeling of data at the software level. It is a fact that the outcomes from big social data collection from historic web crawling methods or modern web service calls are nothing more than software log entries and digital trace records. As the theories of social data are unfold in the actual-space and in the real-time, they would be operated at the micro-genetic level of social media interactions. For this Social Set Analysis, we have chosen the theory of Vatrapu as it undergoes perception and interacts in the actual-space and in the real-time.

The theory of social interactions is derived from the following resources:

1) the enactive approach to the philosophy of mind [4]
2) the ecological approach to perception and action [3]
3) the phenomenological approach to sociology [5], [6]

We also use the theory of social interactions [7]–[9] to describe how individual data items such as Facebook posts, comments, likes etc. come into existence.

B. Philosophy of Computational Social Science

The purpose of this philosophy is to present an argument that we require philosophies of Computational Social Science that discuss and explicitly outline their mathematical modeling, computational implementations, empirical analysis, and sociological assumptions. To the best of our knowledge, no other philosophy of Computational Social Science exists except Social Network Analysis [10] based on the sociology of relations [11] and mathematics of graph theory [12]. Dominant paradigms of computational social science such as Social Simulation [13] and Social Complexity have varying levels of modeling and philosophical unity and maturity. Computational frameworks for Big Data Analytics should inspire towards positive approach that go beyond the negative contributions of methods and assumptions that regularly feature in prominent recent criticisms (for instance, [14] and [15]). We argue that one class of positive contributions would be generative frameworks that provide external articulation of philosophical predictions underlying analytical approaches as well as a production system for creating evaluating new philosophies. The primary scientific objective of this paper is to mathematically model, empirically investigate and theoretically formulate an alternate holistic approach based on set theory, fuzzy set theory [16], social set analysis [17] and associational sociology. To achieve these objectives, the theory of social data is developed and discussed.

	Social Network Analysis	**Social Set Analysis**
Basic Theory	There exists a relation between social actor A and social actor B	There exists an association by actor A with some entity E which can be an actor or an artifact
Unit of Analysis	Dyadic	Monadic, Dyadic & Polyadic
Social Action	Interpersonal Relations	Individual Actions
Mathematics	Graph Theory	Set Theory
Social Structure	Networks	Associations

Fig 1. Two philosophies of computational science

C. Set-Theoretical Big Data Analytics

Based on Verkuilen [15] and Smithson there are five advantages for applying classical set theory [18] in fuzzy set theory [16] and in general particular to computational social sciences:

(1) Set-theoretical ontology is pertinent to conceptualize lack of preciseness which is a main aspect of social science construction. For instance, in the social science domain of marketing, concepts such as brand sentiment, brand loyalty and customer satisfaction are not precise.

(2) Set-theoretical epistemology is pertinent for the analysis of social science constructs that are both dimensional and categorical. That is, set-theoretical approach is well- suited for dealing with different types as well as degrees of that type. For example, social science constructs such as personality, emotion and culture are both dimensional and categorical and set theoretical approach can help for conceptualizing their inherent duality.

(3) Set-theoretical methodology can help analyze multiple associations beyond the general linear model and the conditional means. Besides, set theoretical approach analyzes human associations' antecedent to relations and it allows for both qualitative case study methods as well as quantitative variable centered analytical methods.

(4) Set-theoretical analysis has more theoretical fidelity with most of the social science theories that are expressed logically in set-terms. For example, theories on political preferences and market

segmentation are logically articulated as categorical exclusions and inclusions that lend themselves to set theoretical analytics and formalization.

(5) Set-Theoretical approach practically combines empirical analysis and set-wise logical formulation of social science theories using statistical models for continuous variables. For example, in the case of predictive analytics, it is possible to apply fuzzy set theory to dynamically construct data elements for independent variables such as brand sentiment (subjectivity, polarity, etc.).

D. Conceptual Model of Social Data

Social data consists of two types: Social Text and Social Graph. Social Text constitute the functions, structures and technological inter subjectivity (what the actors/users are trying to communicate with each other and how they are trying to involve each other through language). Social Graph constitutes the appropriation and perception of affordances (which actors/users act up on which technological methods to interact with other social actors in the systems). Social text consists of the communicative and linguistic aspects of the social media interaction such as the topics discussed, keywords mentioned, pronouns used and sentiments expressed. Social graph consists of the structure of the relationships emerging from the appropriation of social media actions such as posting, linking, tagging, sharing, liking etc. It targets on identifying the actors involved, the actions they take, the activities they undertake, and the artifacts they create and interact with.

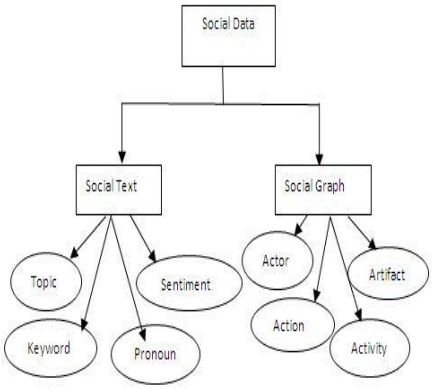

Fig 2. Social Data Model

The following is the research framework for the big social data analytics. It gives the visualization of how the data sets are arranged according to the social set theory.

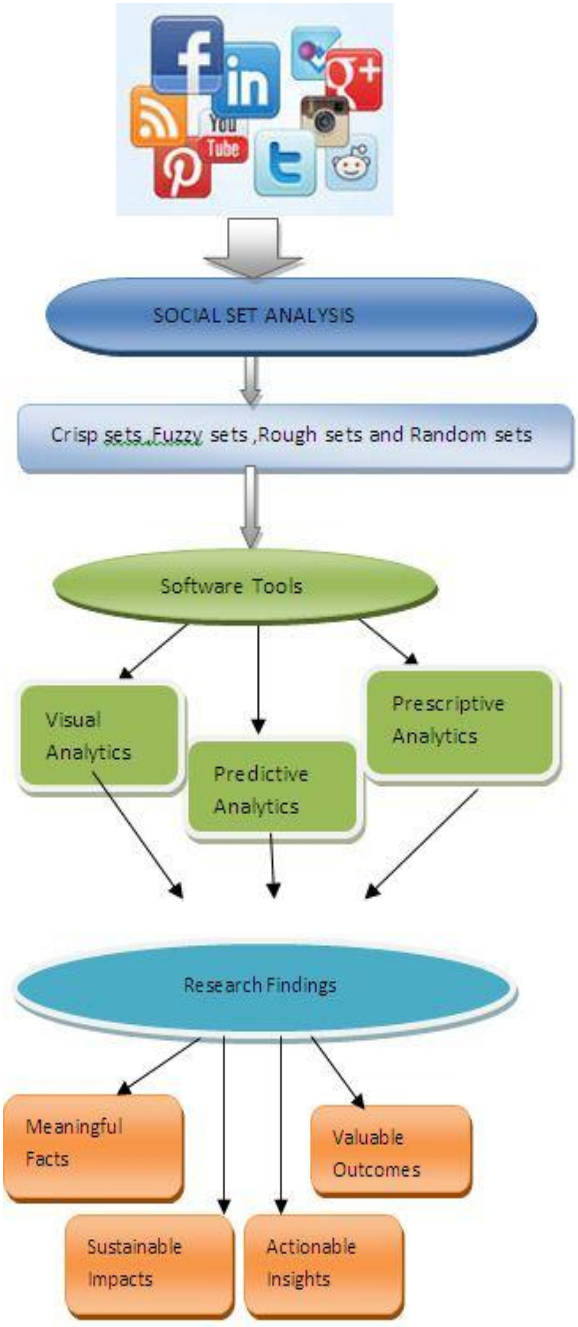

Fig 3. **Research Framework for Big Social Data**

III. CASE STUDIES

CASE STUDY: 1.PREDICTIVE ANAYTICS

Recent research in the computational social science field have shown that how the data resulting from the use of social media channels such as *twitter* can be used to predict the results such as localized moods, movie revenues and epidemic outbreaks. In this paper, we illustrate how social media data from twitter can be used to predict the quarterly sales of iPhones. Based on a conceptual model of social data consisting of and social text (topics, keywords, pronouns, and sentiments) and social graph (actors, actions, activities, and artifacts), we evaluate linear models that converts the iPhone tweets into a prediction of the quarterly iPhone sales.

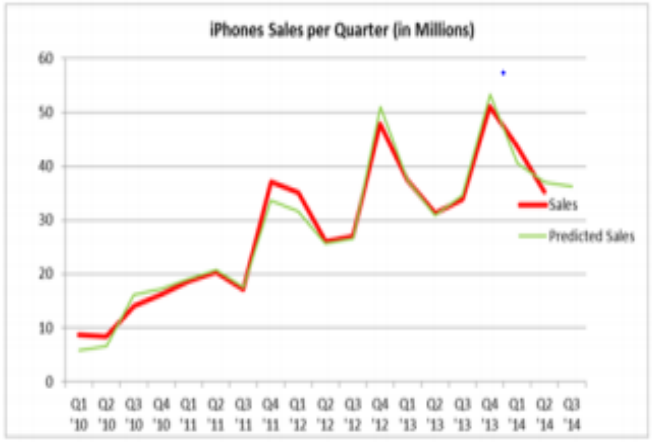

Fig 4. Predictive VS Actual analytics of iPhone sales

CASE STUDY: 2.VISUAL ANAYTICS

In 2005, Cook and Thomas from the National Visualization and Analytics Center coined the term *"Visual Analytics"*. This document states *"the need for analysis techniques for streaming data"*. Every very day, large volumes of data are producing in today's world either by the professionals like local and global news websites or by the public through social network, blogs and website commentary. In theory, all the data are accessible for real-time exploration, but text processing produces delays. Text streams contain both relevant and irrelevant information and it lacks semantic structure. To filter the higher-level semantic structure (stories, sentiments, events, topics and so on) and to extract the relevant information analysts need automatic text-processing method. Regardless of the structure, humans need to make sense of the outcomes and draw the conclusion for decision making. Interactive visualizations can fill the gap between human analysis requirements and computational methods - an idea that has made the evolution of *visual analytics.*

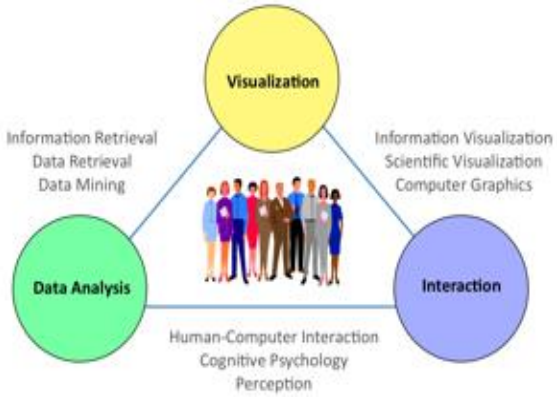

Fig 5. Visual-Analytics

CASE STUDY: 3. DESCRIPTIVE ANALYTICS

Descriptive Analytics is also known as *Statistics.* It tells exactly what the name implies i.e.) describe or summarize the raw data. Descriptive analytics are analytics that describe the past. The past in the sense the time at which the event has occurred that may be one minute ago or one year ago. This analytics are useful because they help us to learn from the past and make us to understand what might be going to happen in future. It has been found that more than 80% of the business analytics (say social analytics) are descriptive. Some social data contains fans, likes, page views, number of posts, pins, check-ins etc. General examples of descriptive analytics are reports that generate historical details regarding finance, customers, company's production, inventory, sales, operations etc.

CASE STUDY: 4. PRESCRIPTIVE ANALYTICS

Prescriptive analytics is a new field that allows the users to *"prescribe"* different possible solutions for a particular problem. Prescriptive analytics surpass the descriptive and predictive analytics by providing more number of possible solutions. It usually predicts multiple features and allows the industry to get the possible results depending on the actions performed. Prescriptive analytics use the combination of tools and techniques such as algorithms, computational modeling procedures, business rules and machine learning.

These techniques are used across inputs from different data sets including transactional data, real-time data, historical data and big data. This prescriptive analytics are somewhat complex to administrator and so most of the companies are not using them yet in their business. If it is implemented in a right way and at right time, it makes larger impact on decision making. Some of the larger companies are using the prescriptive analytics in a successful manner in order to schedule, optimize production and inventory.

IV.CONCLUSION

Computational social science research has reached a level where social media data are ubiquitous though it is hard to gather and analyze data in domain-specific fields. In this context, as demonstrated by the four case studies above, social set analysis has covered the range of predictive, visual, descriptive and prescriptive analytics that are present to reach the required insights. The four illustrative case studies that offers main evidence that social set analysis as a new complete approach to social data analytics in particular and general in computational social science is feasible. Based on associational sociology and set theory, social set analysis can yield complementary insights to the graph theory, social network analysis and relational sociology.

References:

[1] R. E. Montalvo, ‘‘Social media management,’’ Int. J. Manage. Inf. Syst., vol. 15, no. 3, pp. 91–96, 2011.

[2] C. Vollmer and G. Precourt, Always On: Advertising, Marketing, and Media in an Era of Consumer Control (Strategy + Business). New York, NY, USA: McGraw-Hill, 2008.

[3] J. J. Gibson, the Ecological Approach to Visual Perception. Boston, MA, USA: Houghton Mifflin, 1979.

[4] A. Noe, Action in Perception. Cambridge, MA, USA: MIT Press, 2004

[5] A. Schutz, the Phenomenology of the Social World. Evanston, IL, USA: Northwestern Univ. Press, 1967.

[6] H. Garfinkel, Studies in Ethnomethodology. Englewood Cliffs, NJ, USA: Prentice-Hall, 1967.

[7] R. K. Vatrapu, ‘‘Technological intersubjectivity and appropriation of affordances in computer supported collaboration,’’ Ph.D. dissertation, Dept. Commun. Inf. Sci., Univ. Hawaii Manoa, Honolulu, HI, USA, 2007.

[8] R. K. Vatrapu, ‘‘Towards a theory of socio-technical interactions,’’ in Learning in the Synergy of Multiple Disciplines. Berlin, Germany: Spri
nger, 2009, pp. 694–699.

[9] R. K. Vatrapu, ‘‘Explaining culture: An outline of a theory of sociotechnical interactions,’’ in Proc. 3rd Int. Conf. Int. Collaboration (ICIC), 2010, pp. 111–120.

[10] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, ‘‘Network analysis in the social sciences,’’ Science, vol. 323, no. 5916, pp. 892–895, 2009.

[11] M. Emir Bayer, ‘‘Manifesto for a relational sociology,’’ Amer. J. Social., vol. 103, no. 2, pp. 281–317, 1997.

[12] J. L. Gross and J. Yellen, Graph Theory and Its Applications. Boca Raton, FL, USA: CRC Press, 2005.

[13] C. Cioffi-Revilla, Introduction to Computational Social Science: Principles and Applications. London, U.K.: Springer, 2013.

[14] Z. Tufekci. (2014). ‘‘Big questions for social media big data: Representativeness, validity and other methodological pitfalls.’’ [Online]. Available: http://arxiv.org/abs/1403.7400

[15] D. Boyd and K. Crawford, ‘‘Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon,’’ Inf., Commun. Soc., vol. 15, no. 5, pp. 662–679, 2012.

[16] F. Cusset, French Theory: How Foucault, Derrida, Deleuze, & Co. Transformed the Intellectual Life of the United States. Minneapolis, MN, USA: Univ. Minnesota Press, 2008.

[17] I. Hacking, The Social Construction of What? Cambridge, MA, USA: Harvard Univ. Press, 1999.

[18] J. Lovett, Social Media Metrics Secrets, vol. 159. New York, NY, USA: Wiley, 2011.