Data Mining with Big Data

Mr. K.Raju

Mr.KALSANISRAVANKUMAR

ABSTRACT:

Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

INTRODUCTION

DR. Yan Mo won the 2012 Nobel Prize in Literature. This is probably the most controversial Nobel prize of this category. Searching on Google with "Yan Mo Nobel Prize," resulted in 1,050,000 web pointers on the Internet (as of 3 January 2013). "For all praises as well as criticisms," said Mo

ecently, "I am grateful." What types of praises and criticisms has Mo actually received over his 31-year writing career? As comments keep coming on the Internet and in various news media, can we summarize all types of opinions in different media in a real-time fashion, including updated, cross-referenced discussions by critics? This type of summarization program is an excellent example for Big Data processing, as the information comes from multiple, heterogeneous, autonomous sources with complex and evolving relationships, and keeps growing.

PROPOSED SYSTEM:

➢ We propose a HACE theorem to model Big Data characteristics. The characteristics of HACH make it an extreme challenge for discovering useful knowledge from the Big Data.

➢ The HACE theorem suggests that the key characteristics of the Big Data are 1) huge with heterogeneous and diverse data sources, 2) autonomous with distributed and decentralized control, and 3) complex and evolving in data and knowledge associations.

➢ To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data.

ADVANTAGES OF PROPOSED SYSTEM:

☐ Provide most relevant and most accurate social sensing feedback to better understand our society at realtime.

SYSTEM ARCHITECTURE:
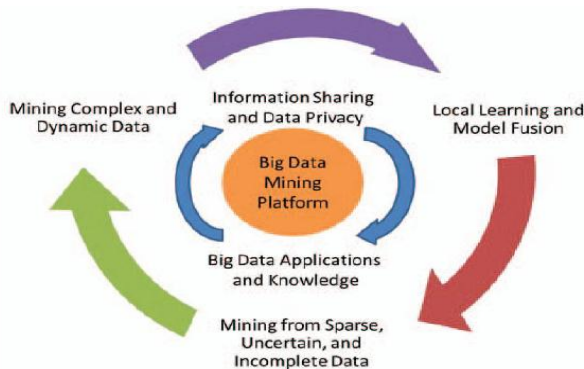
CONCLUSIONS

Driven by real-world applications and key industrial stakeholders and initialized by national funding agencies, managing and mining Big Data have shown to be a challenging yet very compelling task. While the term Big Data literally concerns about data volumes, our HACE theorem suggests that the key characteristics of the Big Data are 1) huge with heterogeneous and diverse data sources, 2) autonomous with distributed and decentralized control, and 3) complex and evolving in data and knowledge associations. Such combined characteristics suggest that Big Data require a "big mind" to consolidate data for maximum values [27].To explore Big Data, we have analyzed several challenges at the data, model, and system levels. To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data. At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values. In other situations, privacy concerns, noise, and errors can be introduced into the data, to produce altered data copies. Developing a safe and sound information sharing protocol is a

major challenge. At the model level, the key challenge is to generate global models by combining locally discovered patterns to form a unifying view. This requires carefully designed algorithms to analyze model correlations between distributed sites, and fuse decisions from multiple sources to gain a best model out of the Big Data. At the system level, the essential challenge is that a Big Data mining framework needs to consider complex relationships between samples, models, and data sources, along with their evolving changes with time and other possible factors. A system needs to be carefully designed so that unstructured data can be linked through their complex relationships to form useful patterns, and the growth of data volumes and item relationships should help form legitimate patterns to predict the trend and future. We regard Big Data as an emerging trend and the need for Big Data mining is arising in all science and engineering domains. With Big Data technologies, we will hopefully be able to provide most relevant and most accurate social sensing feedback to better understand our society at realtime. We can further stimulate the participation of the public audiences in the data production circle for

societal and economical events. The era of Big Data has arrived.

REFERENCES

[1] R. Ahmed and G. Karypis, "Algorithms for Mining the Evolution
of Conserved Relational States in Dynamic Networks," Knowledge
and Information Systems, vol. 33, no. 3, pp. 603-630, Dec. 2012.

[2] M.H. Alam, J.W. Ha, and S.K. Lee, "Novel Approaches to
Crawling Important Pages Early," Knowledge and Information
Systems, vol. 33, no. 3, pp 707-734, Dec. 2012.

[3] S. Aral and D. Walker, "Identifying Influential and Susceptible
Members of Social Networks," Science, vol. 337, pp. 337-341, 2012.

[4] A. Machanavajjhala and J.P. Reiter, "Big Privacy: Protecting
Confidentiality in Big Data," ACM Crossroads, vol. 19, no. 1,
pp. 20-23, 2012.

[5] S. Banerjee and N. Agarwal, "Analyzing Collective Behavior from
Blogs Using Swarm Intelligence," Knowledge and Information
Systems, vol. 33, no. 3, pp. 523-547, Dec. 2012.

[6] E. Birney, "The Making of ENCODE: Lessons for Big-Data

Projects," Nature, vol. 489, pp. 49-51, 2012.

[7] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock

Market," J. Computational Science, vol. 2, no. 1, pp. 1-8, 2011.

[8] S. Borgatti, A. Mehra, D. Brass, and G. Labianca, "Network

Analysis in the Social Sciences," Science, vol. 323, pp. 892-895, 2009.

[9] J. Bughin, M. Chui, and J. Manyika, Clouds, Big Data, and Smart

Assets: Ten Tech-Enabled Business Trends to Watch. McKinSey

Quarterly, 2010.

[10] D. Centola, "The Spread of Behavior in an Online Social Network

Experiment," Science, vol. 329, pp. 1194-1197, 2010.

Author's Profile

Mr. K.Raju received M.Tech degree from Jayamukhi Institute of Technological Science,Narsampet, Warangal affiliated to JNTUH, Hydearabad. He is currently working as Assistant professor, Department of CSE, in Vinuthna Institute of Technology & Science ,Hasanparthy, Warangal, Telangana, India. Hisinterest includes Data Base Management Systems.

Mr.KALSANI SRAVAN KUMARreceived M.Sc. Degree from Directorate of Distance Education,Sri Venkateswara University, Affiliated to Sri Venkateswara University (SVU), Tirupati.He is Currently pursuing M.Tech Degree in Software Engineering Specialization in Vinuthna Institute of Technology & Science Hasanparthy , Warangal, Telangana, India.