Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI:10.21474/IJAR01/2219
DOI URL: http://dx.doi.org/10.21474/IJAR01/2219
ISSN NO. 2320-5407

RESEARCH ARTICLE

A SURVEY ON FEATURE SELECTION FOR EFFICIENT ECONOMIC BIG DATA ANALYTICS.

Ms. DeepthiMogaparthi, Prof. Priyadarshani Kalokhe, Ms. Punam Patil, Ms. Pooja Shedutkar, Ms. Sharda Tenginkai.
Computer Department Alard College Of Engineering & Management, Pune.

Manuscript Info

Manuscript History

Received: 28 September 2016
Final Accepted: 30 October 2016
Published: November 2016

Key words:-
Feature selection, big data, clustering, economy, urbanization.

Abstract

Huge amount of data gets collected every day and there is also a need of technology to handle enormous amount of economic data. Hence there are various and huge number of opportunities for economic analysis.Lowquality, high-dimensionality and great volume pose great challenges on efficient analysis of economic big data. To overcome these challenges our paper presents a new structure for efficient analysis of high-dimensional economic big data based on innovative distributed feature selection. The presented framework combines the methods of economic feature selection and econometric model construction to discover the hidden patterns for economic development.

General Terms:-
Economic Big Data Analysis

Introduction:-

Big data is considered as very significant in solving social and e-commercial issues. [1]Every day 2.5 Quintillion bytes is being produced every day. Viewing all such issues big data offered a stupendous opportunity for the energy efficiency economy, and national security. Having just enormous amount of data is insufficient, because our interests are focused on the "prized" information.

When consumers purchase products through online, products information such as ratings, product reviews, product descriptions given by sellers are very useful for consumers to optimize their purchasing decisions. However, when a consumer purchases used products via online e-commerce sites, the consumer may consider much more attributes about the products than that for purchasing new products. This is due to the need for understanding instance-specific conditions before purchasing a used product and thus the available descriptions for a used product may differ in each other.

[20]Big Data analytics requires business processes to change and it must align with the organization's IT infrastructure to support the business initiatives. New ways of doing data analytics and business intelligence impact on technology infrastructure components.

Here we explore the hidden relations between economy and its response indicators from a new angle and extract the meaningful knowledge from economic big data in order to derive right insights and conclusions on an innovative distributed feature selection that integrates advanced feature selection techniques and econometric methods.

Corresponding Author:-Ms. Deepthi Mogaparthi.
Address:-Computer Department Alard College Of Engineering & Management, Pune.

Literature survey:-

2.12014-Transforming Big Data into Smart Data: Deriving Value via Harnessing Volume, Variety, and Velocity Using Semantic Techniques and Technologies.
Author :Amit P. Sheth:-
[1]Big Data has captured a much of interest in industry, with expectation of better decisions, efficient organizations, and many new jobs. Much of the attention is on the challenges of the four V's of Big Data: Variety, Veocity, Veracity and volume, and methods that control volume, containing storage and computational techniques to support analysis. The concept of Semantic Perception explains how to convert enormous amounts of data into information, meaning, and perception useful for human decision making.

According to study here referred idea to use the concept of Smart Data that is realized by extracting value from a heterogeneous data, and how Smart Data for expanding heterogeneous Big Data authorize a lot of larger class of applications that can profit not just big companies but each individual.

2.22014-A Group Incremental Approach to Feature Selection.
Applying Rough Set Technique
Authors:Jiye Liang, Feng Wang, Chuangyin Dang
Many real data increase dynamically in size. This case takes place in several fields including medical, population studies, and economic research. Since an effective mechanism to deal with such data, cumulative technique has been proposed in the literature.

According to study here referred idea is to use the concept of real data in databases are generated in groups, an effective and efficient group cumulative feature selection algorithm has been proposed which is an extremely important in research of data extraction and knowledge discovery.

2.3 2016-Combining Big Data Analytics with Business Process using Reengineering.
Authors: Meena Jha, Sanjay Jha, Liam O'Brien
Using data in a myriad new ways to drive business value Big Data can denote different things to different organizations, but one subject remains constant.Collaboration business process and Big Data analytics using reengineering can deliver the profit to companies and consumers. Big data analytics need to be integrate with business processes to upgrade operations and offer innovative services to customers.

According to study here referred idea is to use the concept of collaboration business process and Big Data analytics using reengineering can provide the profit to companies and consumer.

2.42005-Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance and Min-Redundancy.
Authors: HanchuanPeng, Fuhui Long, and Chris Ding
Feature selection is significant problem for pattern classification system. Making use of minimal-redundancy-maximal-relevance criterion (mRMR), for first-order incremental feature selection later a two-stage feature selection algorithm by collaborating minimal-redundancy-maximal-relevance and other more knowledgeable feature selectors.This permits us to choose a compact set of superior features at very low cost.

According to study mRMR can be effectively combined with other feature selectors such as wrappers to find a very subset from candidate features at lower. Experiments on both distinct and uninterrupted data sets and multiple types of classifiers reveal that the classification accuracy can be notably improved based on mRMR feature selection.

2.5 2012-Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach
Author: Bing Xue, Mengjie Zhang, Will N. Browne
Classification problems usually have many features in the data sets, but not all of them are useful. Feature selection intends to choose a small number of admissible features to achieve similar or even better classification performance than using all features. Feature selection algorithms serve the task as a single objective problem.

According to study feature selection intends to choose a small number of admissible features to achieve similar or even better classification performance than using all features.

2.62015-A Hybrid Method for Incomplete Data Imputation.
Author: Liang Zhao, Zhikui Chen, Zhennan Yang Yueming Hu
Detonative increase of data volume research of data quality and data usability had attracted large attention.Using these two stacked auto-encoder and incremental clustering (SAICI) following are the aspect of data usability which data can been used for increase of data research incompleteimpute data, and present a novel missing value imputation method has been implemented.

According to studyexperimental results exhibit that this scheme not only assigns the missing data value but also had better time performance scheme for data research. Moreover, the work is well suited for distributed data processing framework that can be applied to the imputation of incomplete big data**.**

2.72015-Embracing Information Explosion without Choking: Clustering and Labeling in Microblogging
Authors: Xia Hu, Lei Tang, Huan Liu.
Detonative popularity of microblogging services has produced a huge volume ofmessages related to microblogging .Helpful information is buried in scrambled, incomplete and unstructured text messages.Methods like labeling and clustering of microblogging messages are difficult because the of the messages length are much shorter compared to conventional text documents.

According to study here propose to form the hulk amount of messages regenerated into clusters that has meaningful cluster labels, that has provided an overview of the subject matter to accomplish users needs related to information and had become helpful in microblogging services that proved to be beneficial.

2.8 2015-Petuum: A New Platform for Distributed Machine Learning on Big Data
Authors: Eric P. Xing, Qirong Ho, Wei Dai, Jin Kyu Kim, Jinliang Wei, Seunghak Lee, XunZheng, PengtaoXie, Abhimanyu Kumar, and Yaoliang Yu.
Modern parallelization scheme use fine-grained operations and scheduling farther as the classic bulk-synchronous processing prototype generalized by MapReduce, and even the specialized graph-based execution that depend on graph representations of ML programs. Divert of approaches incline to pull systems and an algorithms design in different way, and that remains become hard to find out universal platform relevant to a wide range of ML programs.

According to study here MLprograms are basically optimization-centric and error-tolerant, iterative-convergent algorithmic solutions. All these presents unique opportunities for an collective system design, like bounded-error network synchronization and dynamic scheduling that is based on ML program structure.

2.9 2012-Big Data, Big Impact: New Possibilities for International Development.
Authors: Big Data Big Impact Briefing.
Deluge of data is created every day by the communication of billions of people familiar and using computers, GPS devices, mobile devices, and medical devices. These interactions take place through the use of mobile devices being used by people in the growing world. Researchers and policymakers are begin to clear the potential for channeling many torrents of data converted into actionable information that can be beneficial to identify needs cater services, and foretell and prevent crises for the welfare of low-income populations.

According to study it results into quicker Outbreak Tracking & Response and also has Improved the Understanding of Crisis Behavior Change and exact Mapping of Service that needs power to foretell Demand & cater Changes.

2.102014-Big-Data Applications in the Government Sector
Authors: Gang-Hoon Kim, SilvanaTrimi, and Ji-Hyong Chung.
Big data**,** for the large amount of digital data being collected from all sources is so large, raw or unstructured for analysis has been done through conventional relational database techniques.Big data meantnew challenges that have been involving complexity, security and risks to concealment for new technology.

[3]According to study concluded that Big data has given new definition to data management, from extract, transform and load processes to new engineering in big-data applications. Big-data applications affiliated with the

business sector and also had to compare these two sectors in forms of goals, missions, decision-making processes, decision actors and organizational structure.

Existing system:-

With promptly increasing popularity of economic activities, large number of records are involved in economic development.[20]Existing system involves limited indicator and requires prior knowledge of economist.[16]When compressing large varieties of economic data, existing methods gives unsatisfactory performance.There are some issues of economic data which pose great challenges:(1)The collected large amount of data contains incorrect,missing values and nonstandard items.(2)The huge dimensionality of economic indicators makes manual feature selection for economic model construction impossible.(3)[10]Statistical analysis software frequently generates runtime errors when dealing with the high-dimensionality and huge volume economic data.

Proposed system:-

We present a novel framework combining distributed feature selection methodsand econometric models for efficient economic analysis, which can reveal the valuable insights from the low-quality, high dimensionality, [1]and huge volume economic big data. We develop a subtractive clustering based feature selection algorithm and [6]an attribute coordination based clustering algorithm to select and identify the important features of data in horizontally and vertically. Also, we extend these two methods to distributed platform for economic big data analysis. We conduct correlative and collaborative analysis simultaneously to explore the direct and indirect relations between economy and its response indicators based on the identified economic features. We evaluate the proposed framework and algorithms on the economic development data. [16]Extensive experiments and analysis demonstrate that the designed framework and algorithms can distil the hidden patterns of economic development efficiently and the achieved results accord with the actual development situation.

Conclusion:-

In our paper, we reviewed studies on the data analytics from the traditional data analytics to the recent big data analysis based on economic data. Herewe put forth a hardback feature selection based technology. [16]In order to reduce the noise and for the promotion of the data quality various techniques are approached for cleaning and transforming the collected economic big data.[1]Greatchallenges come in site for a user to quickly measurethe opinions when the UI is impressed by a huge amount of information. Our comprehensive experiments on both discrete and continuous data sets and multiple types of classifiers demonstrate the classification accuracy. [20]Organizing the large amount of messages and information into clusters with meaningful cluster tags or labels, provides an sketch of the content to fulfil users' information or data needs.

Combining the power of Big Data we extract the benefits of Big Data in the e-commercial sites. Various ways of doing economic or commercial data analytics and business intelligence provide an impact on technology framework components. Industries must focus on this now so that they can gain competitive advantage in the market place.

References:-

1. A. Sheth, Transforming Big Data into Smart Data: Deriving Value via HarnessingVolume, Variety, and Velocity Using Semantic Techniques and Technologies, in Proc. 30th IEEE Int. Conf. on Data Engineering, 2014, pp.2.
2. World Economic Forum, Big Data, Big ImpactNew Possibilities for International Development, http : ==www3:weforum:org=docs= WEF TC MFS Big- DataBigImpact Briefing 2012:pdf, 2012.
3. Big Data across the Federal Government, http : ==www:whitehouse:gov=sites=default=files=microsites=big data fact sheet final 1:pdf, 2014. [4] H. Giersch, Urban Agglomeration and Economic Growth, Springer Science & Business Media, 2012.
4. R. B. Ekelund Jr and R. F. Hbert, A History of Economic Theory and Method, Waveland Press, 2013.
5. B. Liddle, The Energy, Economic Growth, Urbanization Nexus across Development: Evidence from Heterogeneous Panel Estimates Robust to Cross sectional Dependence, The Energy Journal, vol.34, no.2, pp.223-244, 2013.
6. S. Ghosh and K. Kanjilal, Long-term Equilibrium Relationship between Urbanization, Energy Consumption and Economic Activity: Empirical Evidence from India, Energy, vol.66, no.3, pp.24-331, 2014.
7. S. H. Law and N. Singh, Does Too Much Finance Harm Economic Growth?, Journal of Banking & Finance, vol.41, no.4, pp.36-44, 2014.

8. D. Baglan and E. Yoldas, Non-linearity in the Inflation-growth Relationship in Developing Economies: Evidence from a Semiparametric Panel Model, Economics Letters, vol.125, no.1, pp.93-96, 2014.
9. Q. Ashraf and O. Galor, The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development, The American Economic Review, vol.103, no.1, pp.1-46, 2013.
10. V. Boln-Canedo, N. Snchez-Marono and A. Alonso-Betanzos, A Review of Feature Selection Methods on Synthetic Data, Knowledge and Information Systems, vol.34, no.3, pp.483-519, 2013.
11. S. Alelyani, J. Tang and H. Liu, Feature Selection for Clustering: A Review, Data Clustering: Algorithms and Applications, vol.29, 2013.
12. M. A. Hall, Correlation-based Feature Selection for Machine Learning, The University of Waikato, 1999.
13. M. Dash and H. Liu, Consistency-based Search in Feature Selection, Artificial Intelligence, vol.151, no.1, pp.155-176, 2003.
14. M. A. Hall and L. A. Smith, Practical Feature Subset Selection for Machine Learning, in Proc. 21st Australian Computer Science Conf., 1998, pp.181- 191.
15. L. Beretta and A. Santaniello, Implementing Relief Filters to Extract Meaningful Features from Genetic Lifetime Datasets, Journal of Biomedical Informatics, vol.44, no.2, pp.361-369, 2011.
16. Distributed Feature Selection for Efficient Economic Big Data Analysis Liang Zhao, Zhikui Chen, *Senior Member, IEEE,* Yueming Hu, Geyong Min, *Senior Member, IEEE,*and Zhaohua Jiang,2016
17. H. Peng, F. Long and C. Ding, Feature Selection based on Mutual Information Criteria of Max-dependency, Max-relevance, and Minredundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27, no.8, pp.1226- 1238, 2005.
18. I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2005.
19. J. G. Dy and C. E. Brodley, Feature Subset Selection and Order Identification for Unsupervised learning, in Proc. International Conference on Machine Learning, 2000, pp.247-254.
20. Combining Big Data Analytics with Business Process using Reengineering, Meena Jha,Sanjay Jha,Liam O'Brien,1-3 June 2016 IEEE Int. Conf on2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)