

# Issues and Challenges in Complex Analysis of Variables

M.V.Mawale<sup>1</sup>, Dr.Vinay Chavan<sup>2</sup>

<sup>1</sup> *Researcher Department of Computer Science,  
Adarsha Science J.B.Arts and Birla Commerce, Mahavidyalaya  
Dhamangaon Rly-444709 (India) M.S*

<sup>2</sup> *Associate prof. Dept. Of Computer Science & IT,  
Seth Kesarimal Porwal College, Kamptee,  
Nagpur M.S.*

**Abstract**— The collection of data, its analysis, and presentation of complex data plays critical roles in research, business, and government activities. Complex analysis can be used to uncover the meaning behind complex data sets. The Data Mining and complex analysis useful for uncover the hidden information for approaching challenging data analysis problems. Data are prominent features of modern life. As the amount of data increases so do the challenges facing the statistician in trying to extract information from ever larger data sets which are also more complex and require flexible multiple analyses in order to reveal their structure and information efficiently extracted. The purpose of the data analysis and interpretation phase is to transform the data collected into meaningful information and performance can be analysed. The Paper analyse the issues and challenges in complex analysis of data variables. The paper also discusses importance of data analysis and relationship between variables.

**Keywords** : Complex analysis, Computational complexity , correlations and causalities, Data analysis , Data integrity, Precise prediction, Stability

## I INTRODUCTION

Complex systems are consisting of a large number of elements interacting each other in non-linear fashions which indicate a high importance of interactions for extracting meaningful information from a complex system. Extracting knowledge from data is very complex task as some data is semi-structured or unstructured. When data is in the form of image, such data are challenging to analyse. Some datasets are not in group having same properties and hence require to classify with certain techniques, some data contain some terms to make complex and other datasets must convert larger datasets into smaller ones to solve uncertainty problem[1]. Knowledge discovery process with different techniques helps to unveil some of the complex system characteristics Data analysis is consist of pre-processing of the data, visualization and interpretation of the results. Making relationships from raw data help to draw interpretation and judge their dependability also the scope of complex data analysis.

The Paper is focus on major issues and challenges in complex analysis of data and discuss its scope. This paper is structured as follows Second section discusses the major issue in complex data analysis, Third section discusses the

major challenges in in complex data analysis. Fourth section discusses Problem in complex data analysis fifth section discusses the Future scope of in complex data analysis.

## II ISSUES

When data analysis is carried out for qualitative research such data analysis must contain statistical procedures, when data collection is done in batch wise manner, data is continuously collected and analysed simultaneously. Researchers generally analyse for way or patterns in observations through the whole data collection phase [2].

As data integrity is important aspects in data analysis as it give the accurate and appropriate analysis of research findings. Statistical analysis always gives clear picture about cause of data analysis. Improper statistical analyses push far behind scientific findings, mislead casual readers [3].

- i. **Selecting proper data collection methods and appropriate analysis method:** There is various data collection method like survey method, questionnaires, group discussion, collecting data sample from particular organisation or agency. Data collection method is selected as per requirements and suitability of problem. Method of analysis is selected as per scientific discipline but analytic procedures must selected early in the research process The method of sampling and its design selection and its approximation must depend on Statistical advice at the stage of initial planning of an investigation Analysis is influenced by the method use in data recording [4].
- ii. **Drawing unbiased inference:** The main aim of analysis is to draw inference which make different between a situations reflecting a valid side versus invalid one. Appropriate method of collection of the data, or selection of method of analysis, will increase the chances of drawing a unbiased inference. Biased inference always shows default result [5]. More Emphasis on pattern of behaviour although knowledge of theories can define the processes used in the qualitative studies, many times way of of behaviours or

occurrences derived from analysed data can draw new theoretical standard[6] (Savenye,obinson, 2004). These patterns of behaviour are obtained from statistical analysis. Computer-based statistical packages can facilitate complex analytic procedures application . Inappropriate uses of statistical packages can draw wrong prediction result.

- iii. **Nature of the variables used:** Variables identity shows its nature either quantitative, Comparative, or qualitative. Quality variable is good sign for better accuracy in data analysis as each variable shows its significance individually. Quantity variables make relationship between each other having same properties.
- iv. **Determining significance :** Statistical significance makes standard of acceptability for user as statistical significance has a true practical meaning. Clinical significance is the potential for research findings to make a real and important difference to clients to any other problem identified [7].
- v. **Lack of clearly defined and objective outcome measurements :** Data analysis process must have clearly defined objectivity of problem as well as Clearly present outcome measurement. As correct goal is defined, other process to achieve that goal is done as per logic of process otherwise thus it potentially misleads the readers.
- vi. **Provide accurate analysis:** To decrease the statistical error is potential issue to get honest and accurate analysis result. Data must be related with analysis yet concise[8].
- vii. **Manner of presenting data:-** More psophiscated way of presentation of derived data makes increases a significant finding by determining how to present derived data rather than raw data. Graphical presentation of data makes more readable and how data was manipulated for future review [9].
- viii. **Data integrity:** Raw data is converted into derived data by partitioning the data to reduce or eliminate inconsistencies so that data integrity is not compromised.
- ix. **Objectivity& Subjectivity:** While applying various techniques, objectivity and subjectivity is raised when data is analysed to validate result and check its application ability.
- x. **Reliability and Validity:** While performing analysis on either quantitative or qualitative analysis there are big challenges of reliability and validity. Three factors that can affect the reliability of analysed data:
  - a) **Stability:** The tendency to follow same way of coding and re-code that same data by the same method over a period of time consistently.
  - b) **Reproducibility:** The tendency to classify categories membership without changing old way so that extension of process much more simplify.
  - c) **Accuracy:**The tendency of process to show result which similar corresponds to a standard or norm statistically [10] .

- xii **User interaction issues:** The ability to mine knowledge by viewing multiple small amount of hidden information, the use of domain knowledge, ad hoc mining, and knowledge visualization is related with user interaction. Data analysis is one of the parts of data mining hence data characterization, discrimination, association, classification, clustering, tread and deviation analysis, and similarity analysis issue is important. The purpose of the data analysis is to transform the data collected into meaningful information and its use in development of the intervention and its performance.
- xiii **Efficient and scalable algorithms:**To effectively extract information from large data, data mining algorithms must be efficient and scalable. Computational complexity is to reduce size of data, divide data into partitions that can be processed in parallel using some data mining methods.
- xiiii **Handling of relational and complex types of data:** Complex data arise affect the reliability of analysed data. When dealing with complex systems, an important part of the information about them is combining within the relationships between the different elements of the system. Two levels of relationships can be defined
  - Correlations: Correlation is evolution of the dynamics of two elements share some common features.
  - Causality: Causality is the dynamics of one element is driving the evolution of a second element.

### III CHALLENGES

Data analysis is start with data Collection and ends with precise result. Hence challenges are start from data collection phase.

- i. **To Choose related variable for data collection**  
Data collection is done with quite accuracy by deciding which raw data is important or not. For soil fertility almost every parameter is included as it indirectly related with crop yield. Past study problem has concerned only one or two or three soil parameter which affect soil productivity.
- ii. **Precise prediction**  
Challenges of the higher reliability raised design of precision and prediction for control system faces nonlinearities in the system.
- iii. **Data mining challenges**  
As data analysis is part of data mining, data availability and complex representation models, are major challenges to reduce complexity.
- iv. **Classification of data**  
The datasets usually are very large, complex, heterogeneous, and hierarchical in nature and vary in quality concern. Data preprocessing and transformation are first step in data analysis. To convert the data into appropriate form is big challenges to get accurate result[11].
- v. **Size of Data**  
To effectively extract information from a huge amount of data in databases, data must be efficient and scalable[12]. Domain knowledge can be used to

eliminate repeated data or unnecessary data to reduce size of data[13].

**vi. To avoid Inconsistent data**

Inconsistencies due to data representation can exist if more than one model for expressing a specific meaning exists additionally, the data type does not always reflect the true data type. This plays an important role during statistical analysis (mean and variance). a substantial challenge for data.

**vii. Number of variables**

The computational complexity is not linear for certain data mining techniques. Large number of variable requires more maintenance. Techniques such as principle component analysis can be used to reduce number of variables without losing most of the original variability of data [14].

**viii. Missing/Incomplete data**

We think that Some data are not useful and Some data elements are not collected due to omission, irrelevance, excess risk or inapplicability in that particular application area.

**III PROBLEMS**

**Common problem:** Common challenges include the filling in missing data, altering or changing data, data mining, and developing graphical representations of the data[15]. Inconsistencies due to data entry errors are common problems.

**Assessment of correlations and causalities:** While the problem of detecting correlations and causalities has been extensively studied in recent research as both plays important role in determining inter dependability of variable.

**IV CONCLUSION**

After reviewing some techniques that used to mine and analyse the relationships existing between the constituting elements of a complex system, only structured alphanumeric digital data are taken into analysis. Other

types of data are not taken into consideration. special Techniques and specific software is needed to development platform for pre-processing data. Identification of the different data sources, finding of particular potential relationships among these sources and identification of potential realities that might be analysed through them are within scope and judging the correct data is one of the challenges. The paper discussed the major issues, challenges and scope in complex data analysis. Paper discussion will provide some way into the challenges that may encounter when using data mining to solve complex real-world problems.

**REFERENCE**

[1] D. Zighed et al. (Eds.), *Complex Data ,Studies in Computational Intelligence*, Vol. 165, Springer, 2009, 153-63.  
 [2] Savenye, W. and Robinson, R. (2003). *Qualitative research issues and methods*  
 [3] Shepard, R.J. (2002). *Ethics in exercise science research. Sports Med*, 32 (3): 169-183.  
 [4] Smeeton, N., Goda, D. (2003). *Conducting and presenting social work research: some basic statistical considerations. Br J Soc Work*, 33: 567-573.  
 [5] www.ori.hhs.gov/education/products/n\_illinois\_u/.../datopic.html[stst analysis]  
 [6] Savenye, W. and Robinson, R. (2003). *Qualitative research issues and methods*  
 [7] Jeans, M. E. (1992). *Clinical significance of research: A growing concern. Canadian Journal of Nursing Research*, 24, 1-4.  
 [8] Shamoo, A.E., Resnik, B.R. (2003). *Responsible Conduct of Research. Oxford University Press.*  
 [9] Shamoo, A.E. (1989). *Principles of Research Data Audit. Gordon and Breach, New York.*  
 [10] Gottschalk, L. A. (1995). *Content analysis of verbal behavior: New findings and clinical applications. Hillside, NJ : Lawrence Erlbaum Associates, Inc*  
 [11] www.cs.umb.edu/~dsim/papersps/dmmd.pdf  
 [12] *Challenges in Data Mining on Medical Databases* Fatemeh Hosseinkhah, Hassan Ashktorab, Ranjit Veen  
 [13] Owrang (2007) *Information Quality Management*  
 [14] <https://www.statsoft.com/Textbook/Data-Mining-Techniques>  
 [15] Shamoo, Resnik,( 2003) Shamoo, A.E., Resnik, B.R. (2003). *Responsible Conduct of Research. Oxford University Press.*