Journal of Horticultural Science and Crop Research
Volume 1 | Issue 1

Review Article Open Access

Using Simulation and Data Mining for Engineering Knowledge Improvement

Ennouri K*

University of Sfax, Sfax, Tunisia

***Corresponding author:** Ennouri K, University of Sfax, Sfax, Tunisia, E-mail: 1karimennouri1@gmail.com

Citation: Ennouri K (2019) Using Simulation and Data Mining For Engineering Knowledge Improvement. J Horti Sci & Crop Res 1(1): 104

Received Date: August 22, 2018 **Accepted Date**: May 21, 2019 **Published Date:** May 24, 2019

Abstract

Significant work has been attempted to better comprehend the attributes and mechanisms suggested in simulation and data mining in precedent research. More prominent understanding causes not only give progresses in numerous fields but also enable experts to better consider the last target. The inquiries postured in this review is whether research is able for depicting and demonstrating data from a statistical point of view, along with investigating systems to enhance the precision of calculations that depend on this data. Simulation of data, statistical analysis and modelling of particular data contributes ads to more educated comprehension of the underlying characteristics of the desired output. Statistical analysis of data shows that data can be depicted utilizing particular design. Modelling method and is the perfect possibility for examiners and specialists.

Keywords: Data Mining; Simulation; Statistical Analysis; Modelling

Introduction

In recent years, data-mining (DM) has turned out to be one of the most significant instruments for extracting and controlling data [1] and for building up designs and patterns in order to generate helpful information and then a specific end goal for decision-making [2]. The disappointments of structures or materials in a domain are often either a consequence of obliviousness or the failure of individuals to take note of past problems or study the patterns of past incidents keeping in mind the end goal to settle on informed deductions and conclusions that can forestall future events [3]. Besides, almost all parts of life activities exhibit an analogous model. Whether the action is banking, selling, manufacture, population learning, employment, health area, observing of human or machines, art or teaching, all have approaches to record known data [4] but are incapacitated by not having the correct devices to utilize this known data to handle the vulnerabilities of the future [5]. In this paper, we begin by providing an exhaustive overview of simulation strategies. Subsequently, we review the data mining. We describe the used techniques of data mining. The goal of this paper is to resume the concepts enabling the researcher to identify which strategy or simulator to employ for a specified modelling problem.

Simulations and Modelling for Sciences and Engineering

Modelling is known as the process of building a model from an existing or intended system. It is surely the most significant characteristic of a simulation plan. Computer simulations are utilized broadly as models of realistic frameworks to assess output responses [6]. Uses of simulation are broadly found in various areas including manufacturing, and engineering design [7]. Besides, the determination of best simulation parameters can pilot to developed action, nevertheless arranging them well is a challenging complexity [8]. Customarily, the components are picked by choosing the best from an arrangement of applicant factor settings [9]. The principle of the simulation prototype is to provide insight into all performance metrics neighbouring the outpatient unit and how they are influenced by methodical modifications. Kuljis *et al.* [10] depict the six most important methods in simulation: 3-D and virtual reality simulation, discrete-event, agent-based simulation, continuous event, Monte Carlo and system dynamics, and how they have been employed in producing and how they could eventually be employed in healthcare. Moreover, employing 3D computer prototypes that permit the reconstitution of cardiac chambers offers the aptitude to replicate the organization of the ventricular anatomy in a virtual representation [11]. As illustration, incorporating computational prototypes of the heart with medical data establishes massive hope for increasing acts for cardiovascular illness [12].

It is noted that two principle sorts of dynamic modelling ways to deal a complex framework are simple analytical modelling [13,14], and numerical modelling [15]. In simple analytical modelling, the existent structure is extraordinarily simplified, so that explanation of the mathematical equations relating the field can be derived [16,17], and it frequently simulates one part of the response of a framework. On other side, the second approach is numerical modelling (Figure 1). Albeit comprehensive in its nature, the itemized numerical models frequently prompt computational high-priced cost, since the framework has to be divided into several repeated cells and equations are initiated to simulate the conduct of the framework [18].

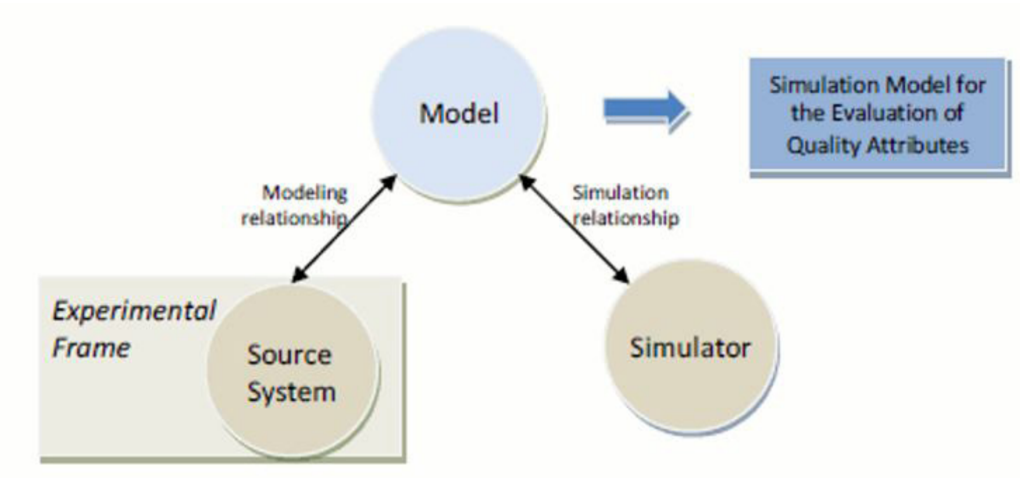

Figure 1: Simulation model for the analysis of quality attributes [19,20]

At this phase, it is beneficial to express that extracting the correct information from an arrangement of data using data-mining methods is reliant on the procedures themselves as well as on the creativity of the expert [21]. The examiner characterizes the issues and objectives, has the correct understanding and knowledge of the tools accessible and makes comparative, intuitive trial of which device to use to accomplish the best outcome that will meet the objectives [22]. There are additionally constraints for diverse users of data-mining for the reason that the software packages employed by analysts are habitually custom-designed to meet particular applications and may have partial and restricted ease of use outside those specific situations [23].

Data Mining: Techniques and Concepts

Data mining can be defined as a method to find out strategic information hidden in huge databases. The procedure of data mining comprises of three successive phases: the initial investigation, model construction or prototype recognition with approval/check, and finally arrangement [24]. The definitive purpose of data mining is expectation and also prediction and prescient data mining is the most widely recognized sort of data mining and one that has the most efficient applications [25]. Basically, the endeavour is to utilize all the data a user directly or indirectly has, as a major aspect of a scientific way to accepting setting and conditions, as well as to make an opinion loop in which the expert can make incremental modifications based not only on involvement and, not merely from the data reports, but based on much larger but more unclear information footprint which is rather catching the reality of concealed associations with all inputs and output(s) [26] (Figure 2). In a biotechnological manufacture or manufacturing situation, we fight with the issues of process optimization, cellular organization, sequencing, human factors, and quality control [27,28].

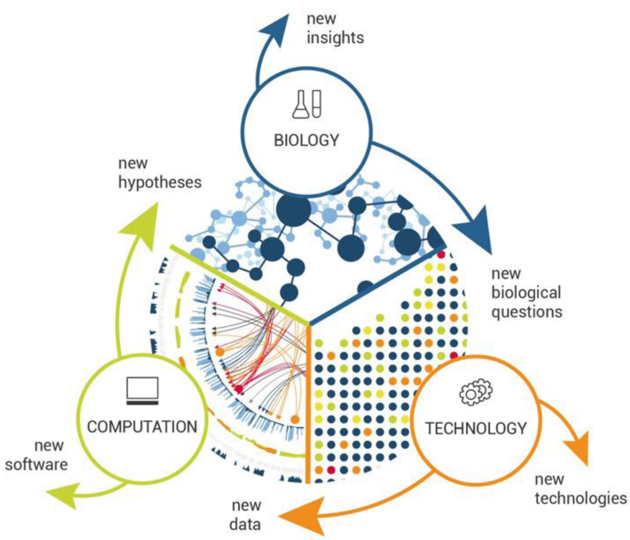

Figure 2: Steps for understanding biology at the system level [29]

The natures of these data sets are investigated and their single characteristics defined. It is named data pre-processing and arrangement [30]. To a substantial degree, this data preparing helps the analyst/expert to settle on a decision of the prescient method to apply [31]. The huge problem is how to diminish the variables to a smallest number that can totally anticipate the response variable [32].

Moreover, diverse data-mining strategies, including multiple linear regression (MLR) [33], inspired from the common least-square approach; principal component regression (PCR), an unsupervised method inspired from the principal component analysis [34]; the Partial Least Squares (PLS), defined as a supervised method [35], and the Nonlinear Partial Least Squares (NLPLS), which utilizes various neural network functions [36] to plot nonlinearity into models, were related to each of the data sets. Each procedure has diverse strategies for use; these distinctive techniques were utilized on each data set first and the best strategy in every system was noted and utilized for global examination with other methods for similar data set [37].

Such data-analysis strategies include data extraction, change, association, gathering, and investigation to see designs in order to make forecasts [38]. Miller et al. [39] indicate that joining more complication to a dataset does not automatically attach value to the ultimate examination. They go on to declare that too much complication is in reality counterproductive since it necessitates much more moment and endeavour to guarantee the model is executing like the authentic organism.

To industrial engineers, whose work it is to devise the best methods of optimizing procedures in order to create more value from the framework, data-mining turns into an intense apparatus for assessing and settling on the best choices based on records so as to generate supplementary value and to avoid failure [40]. The capability of data-mining for engineering supervisors has yet to be completely exploited [41]. Halamek et al. [42] suggested a revival training concept that employed high-accuracy coaching as a foremost constituent to improve team cooperation and technical qualifications.

In addition, the prescient part of data mining is likely its most expanded component; it has the best potential result and the most exact description [43]. In data mining, the decision of procedure to employ in analyzing a data set relies upon the comprehension of the examiner. Generally, many times is squandered in attempting each and every forecast strategy (bagging, boosting, and meta-learning) in a bid to discover the greatest explanation that fits the expert's requirements [44]. Hence, with the approach of enhanced and adjusted forecast systems, there is a requirement for an expert to know which device performs best for a specific category of data set [45].

Conclusion

Planning for the future is very essential in engineering sciences. Approximations of potential values of experimental variables are required. The manufacturing requires forecast or predicting of furnish and claim for production planning, and engineering resolutions. The understanding of data-mining devices that could diminish the frequent a *priori* ideas in these domains must be generally accessible in order to improve effectiveness of obtained results.

References

1. Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier (3rd Edn), USA.

2. Stéphane T (2012) Data mining et statistique décisionnelle: l'intelligence des données, France.

3. Zyt J, Klosgen W, Zytkow J (2002) Handbook of data mining and knowledge discovery, Germany

4. Wu X, Zhu X, Wu GQ, Ding W (2014) Data mining with big data. Trans Knowl Data Eng 26: 97-107.

5. Yang Q, Wu X (2006) 10 challenging problems in data mining research. Int J Inf Tech Decis 5: 597-604.

6. De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9: 67-103.

7. Jensen KF (2001) Microreaction engineering-is small better? Chem Eng Sci 56: 293-303.

8. Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: Review and development recommendations. Simulation 82: 609-23.

9. Simpson TW, Poplinski JD, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17: 129-50.

10. Kuljis J, Paul RJ, Stergioulas LK (2007) Can health care benefit from modeling and simulation methods in the same way as business and manufacturing has? Winter Simulation Conference.

11. Andreu D, Berruezo A, Ortiz-Pérez JT, Silva E, Mont L, et al. (2011) Integration of 3D electroanatomic maps and magnetic resonance scar characterization into the navigation system to guide ventricular tachycardia ablation. Circ- Arrhythmia Elec 4: 674-83.

12. Sermesant M, Peyrat JM, Chinchapatnam P, Billet F, Mansi T, et al. (2008) Toward patient-specific myocardial models of the heart. Heart Fail Clin 4: 289-301.

13. Islam R, Husain I (2009) Analytical model for predicting noise and vibration in permanent magnet synchronous motors. Energy Conversion Congress and Exposition (ECCE).

14. Cheng PJ, Lin SC (2001) An analytical model to estimate angle formed by laser. J Mater Process Technol 108: 314-9.

15. Rabczuk T, Akkermann J, Eibl J (2005) A numerical model for reinforced concrete structures. Int J Solids Struct 42: 1327-54.

16. Veeraragavan A, Lenert A, Yilbas B, Al-Dini S, Wang EN (2012) Analytical model for the design of volumetric solar flow receivers. Int J Heat Mass Transf 55: 556-64.

17. McCarthy CT, Gray PJ (2011) An analytical model for the prediction of load distribution in highly torqued multi-bolt composite joints. Compos Struct 93: 287-98.

18. Huang MJ, Eames PC, Hewitt NJ (2006) The application of a validated numerical model to predict the energy conservation potential of using phase change materials in the fabric of a building. Sol Energy Mater Sol Cells 90: 1951-60.

19. Bogado V, Gonnet S, Leone H (2011) A Discrete Event Simulation Model for the Analysis of Software Quality Attributes. CLEI Electr J 14: 1-18.

20. Zeigler B, Praehofer H, Kim T (2000) Theory of Modeling and Simulation. In: Integrating Discrete Event and Continuous Complex Dynamic Systems, Academic Press, USA.

21. Kantardzic M (2011) Data mining: concepts, models, methods, and algorithms, Wiley-IEEE Press, USA.

22. Geng L, Hamilton HJ (2006) Interestingness measures for data mining: A survey. CSUR 38: 9.

23. Torra V (2003) Trends in information fusion in data mining. In: Information fusion in data mining, Springer, Berlin, Heidelberg, Germany.

24. Olson DL, Delen D (2008) Advanced data mining techniques. Science & Business Media, Springer, USA.

25. Seni G, Elder JF (2010) Ensemble methods in data mining: improving accuracy through combining predictions. In: Synth Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers, USA.

26. Mirkin B (2005) Clustering for data mining: a data recovery approach, Chapman and Hall/CRC, USA.

27. Kusiak A (2006) Data mining: manufacturing and service applications. Int J Prod Res 44: 4175-91.

28. Maulik U, Bandyopadhyay S, Mukhopadhyay A (2011) Multiobjective Genetic Algorithms for Clustering: Applications in Data Mining and Bioinformatics, Springer, USA.

29. Omicscouts (2012) Disease and systems biology. Proteomics: a tool for understanding biology at the system level, Germany.

30. Wang L, Fu X (2006) Data mining with computational intelligence, Science & Business Media, Springer, USA.

31. Roiger RJ (2017) Data mining: a tutorial-based primer, Chapman and Hall/CRC, USA.

32. Hill T, Lewicki P, (2006) Statistics: methods and applications: a comprehensive reference for science, industry, and data mining, StatSoft, Inc, USA.

33. Ennouri K, Ben Hassen H, Zouari N (2013) Optimization of bioinsecticides overproduction by Bacillus thuringiensis subsp. kurstaki using linear regression. Pol J Microbiol 62: 287-93.

34. Ennouri K, Ben Ayed R, Ercisli S, Ben Amar F, Triki MA (2017) Evaluation of Variability in Tunisian Olea europaea L. Accessions using Morphological Characters and Computational Approaches. Not Bot Horti Agrobot Cluj-Napoca 45: 262-9.

35. Ennouri K, Ben Ayed R, Mazzarello M, Ottaviani E, Hertelli F, et al. (2016) Classical and Bayesian predictions applied to Bacillus toxin production. 3 Biotech 6: 206.

36. Ennouri K, Ben Ayed R, Triki MA, Ottaviani E, Mazzarello M, et al. (2017) Multiple linear regression and artificial neural networks for delta-endotoxin and protease yields modelling of Bacillus thuringiensis. 3 Biotech 7: 187.

37. Ennouri K (2017) Procédés du Génie Technologique. Éditions universitaires européennes, Germany.

38. Liu H, Motoda H (2012) Feature selection for knowledge discovery and data mining, Springer Science & Business Media, USA.

39. Miller M, LaVecchia R, Shahi N (2008) Allocating outpatient clinic services using simulation and linear programming. Winter Simulation Conference.

40. Harding JA, Shahbaz M, Kusiak A (2006) Data mining in manufacturing: a review. J Manuf Sci Eng 128: 969-76.

41. Marbán O, Segovia J, Menasalvas E, Fernández-Baizán C (2009) Toward data mining engineering: A software engineering approach. Inform Syst 34: 87-107.

42. Halamek LP, Kaegi DM, Gaba DM, Sowb YA, Smith BC, et al. (2000) Time for a new paradigm in pediatric medical education: teaching neonatal resuscitation in a simulated delivery room environment. Pediatrics 106: E45.

43. Witten IH, Frank E, Hall MA, Pal CJ (2016) Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

44. Chawla NV (2009) Data mining for imbalanced datasets: An overview. In: Data mining and knowledge discovery handbook, Springer, Boston, MA, USA.

45. Tufféry S (2011) Data mining and statistics for decision making. Chichester, Wiley, USA.