

Data Mining Based Soft Computing Methods For Web Intelligence

Ankita Hatiskar

Student, Department of Information Technology

Sir Sitaram and Lady Shantabai Patkar College of Arts and Science, Mumbai, Maharashtra, India

Abstract: *Data mining is the procedure of extracting interesting knowledge from enormous amounts of data contained in databases, including such patterns, associations, changes, deviations, and prominent structures. Soft Computing Methods such as fuzzy logic, artificial neural network, etc. aims to uncover the potential for error and inaccuracy in order to accomplish scalability, durability, and reduced methods. In today's information age, the Web is the most common distribution medium. Due to its popularity on the Internet, it is widely used in commercial, entertainment, and educational purposes. Web Intelligence (WI) is engaged with the scientific study of the Web's new areas. It is a new area of computer science that integrates artificial intelligence with sophisticated information technology in the framework of the Web, expanding well outside each one of them. In online applications, data mining gives a plethora of possibilities. The biggest concern is figuring out how to identify relevant hidden patterns for improved application. Soft computing techniques such as neural networks, fuzzy logic, support vector machines, and genetic algorithms are used in evolutionary computation to solve this problem. We look at how soft computing approaches may be used to build web intelligences in this research.*

Keywords: Application, Artificial Neural Network, Data Mining, Fuzzy Logic, Genetic Algorithms, Soft Computing Methods, Web, Web Mining.

I. INTRODUCTION

Web mining involves applying data mining techniques to web data in order to retrieve information from databases, i.e., mining online content. Using this tool, you can fetch the useful information from web documents and web services. A web mining process is a method of extracting information from a web service or web data. A data mining project is launched off-line, while a web mining project is launched online. In data mining the data is being stored in the data ware houses and in web mining data is stored in server database and web logs [1]. Web intelligence, a term that was coined in the late 1999's, concerns about research and application of machine learning and information technology with a specific focus on the Web platforms. Online text classification is one of the most common Web Intelligence applications, Web recommender for e-commerce, Web document clustering, web usage profiling and similar knowledge discovery tasks are gaining the attention of multinational research firms. It is very challenging to make sense of the data originating from the Web, referred to as Web Intelligence data. WI (Web Intelligence) is studied carefully from different aspects [4]. WI exploits Artificial Intelligence (AI) and advanced Information Technology (IT) on the Web and Internet [4]. Computational Web Intelligence (CWI) is a hybrid technology of Computational Intelligence (CI) and Web Technology (WT) dedicated to increasing quality of intelligence of e-Business applications on the Internet and wireless networks [7]. WTI makes intelligent e-Business applications using Computational Intelligence (CI) and Web Technology (WT). Support Vector Machine (SVM) proposed by Vapnik is a newly developed technique which based on statistical learning theory [8,9], this method uses the Structure Risk Minimization approach, which prevents local minima and smoothly tackles the overlearning challenge while also maintaining adequate adaptability and clear that the results classification. It uses approximate logic rather than fixed reasoning. Fuzzy logic belongs to multivalued logic. A fuzzy logic variable will have a truth value that ranges from 0 to 1 rather than the binary values found in traditional sets (where variables are either true or false). Genetic algorithms imitate the process of natural selection through their search heuristics. Optimization and search problems are routinely solved by using this heuristic (also called a meta heuristic). Artificial neural systems, also known as neural networks, are physical cellular structures that also can learn, store, and utilize expertise information.

1. Web- The World Wide Web (abbreviated as WWW or W3, commonly known as the web) is a system of interlinked hypertext documents accessed via the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via hyperlinks.
2. Web Intelligence- Web Intelligence in the field of scientific analysis and innovation that analyses the capabilities of artificial intelligence and information technology in the development of new devices, applications, and platforms that are accessible by the Internet.
3. Methods in Data Mining
 - a. Artificial Neural Networks: Artificial Neural Networks often known as Neural Networks (NNs), are a set of algorithms that simulate the patterns of brain activity in order to identify patterns in data from multiple sources. Neural networks identify patterns in enormous amounts of data, helping the company to gain a better understanding of their consumers, allowing them to help regulate their marketing strategies, boost sales, and reduce spending.
 - b. Decision Tree: A decision tree is a tree-like model that illustrates decisions and their possible outcomes, impacts, and costs as a decision support technique.
A decision tree consists of 3 types of nodes.
 - Chance nodes- represented by circles.
 - Decision nodes- represented by squares,
 - End nodes- represented by triangles
 - c. Genetic Algorithm (GA): A genetic algorithm (GA) is a search heuristic that simulates the natural evolution process. This heuristic is widely designed to examine pragmatic approaches to optimization and search applications.

The format of this document is as follows: in Section II we present a Literature review covering Data Mining, Soft Computing, and Web Intelligence. Section III presents various applications of WI. Section IV explicates data mining-based soft computing approaches for WI. Finally, we conclude this paper in Section V.

II. LITERATURE REVIEW

2.1 Data Mining

Data mining is the procedure of extracting interesting knowledge from enormous amounts of data contained in databases, including such patterns, associations, changes, deviations, and prominent structures. In simple words, Data Mining is an undeniable phase in software engineering that analyses data in addition to identifying and portraying designs. Considering we live in a world where we are constantly bombarded with data, it is critical that we figure out how to organise it, find the data we need, illuminate structures, and reach judgments. And since the data around the important information is regarded useless, it is effectively referred to as noise, which is really wished to be absent. The source and volume of data are what defines data mining methods. The key characteristic of a data set attuned to mining is its complexities, and data mining methods isolate themselves from predefined analytical techniques by "learning."

2.2 Soft Computing

Soft computing is an emerging approach to computing which parallels remarkable ability of human mind to reason and learn in an environment of uncertainty and imprecision [3]. Soft computing approaches are divided into three categories, as follows. Artificial Neural Networks (ANN): for learning and adaptation; Fuzzy Logic (FL): for information representation via fuzzy If-Then rules; Genetic Algorithm (GA): for evolutionary computation. Hybridization of these concepts is used in Soft Computing. All of the benefits of constituent approaches would be acquired by a hybrid technique. Thus, the components of Soft Computing are complementary, not competitive, offering their own advantages and techniques to partnerships to allow solutions to otherwise unsolvable problems [5].

A. Distinction between Soft Computing and Hard Computing

| Soft Computing | Hard Computing |
|------------------------------|-----------------------------------|
| It uses fuzzy logic. | It uses two valued logics. |
| It can deal with noisy data. | It can only deal with exact data. |

| | |
|---------------------------|---------------------------|
| Computation time is less. | Computation time is more. |
| Gives approximate output. | Gives exact data. |
| Parallel Computation. | Sequential Computation |

Table 1: Distinction between Soft Computing and Hard Computing

2.3 Soft Computing Methods

A. Fuzzy Logic

As a core aspect of soft computing is fuzzy logic, which is an aspect of systems with high MIQs (machine intelligence quotients). In fuzzy logic's applications, two notions play a critical role. The first is a linguistic variable; that is, a variable whose values are words or sentences in a natural or synthetic language [5]. The other is a fuzzy if-then rule, in which the antecedent and consequents are propositions containing linguistic variables [5]. Fuzzy system proposes a mathematic calculus to translate subjective human knowledge of the real processes, which makes it possible to manipulate practical knowledge with some level of uncertainty. Soft computing has evolved as a result of the advent of fuzzy logic. In order to build relevant and usable correlations for both system variables, fuzzy models reflect a conservative and user-oriented filtering of data, qualitative insights, and calibration of common-sense theories. Due to the increasing flexibility of information discovery systems, any process of knowledge acquisition, modification, and processing requires a considerable amount of discussion and planning. Implicitly, fuzzy sets are oriented to dealing with linguistic domain knowledge and providing more interpretable results.

Benefits of Fuzzy Logic

- Unlike classical logic, fuzzy logic is mostly relevant to understanding real-world difficulties.
- Traditional Boolean logic algorithms claim additional hardware than fuzzy logic algorithms.
- With unclear or faulty input, fuzzy algorithms can perform adequately.

Drawbacks of Fuzzy Logic

- Fuzzy algorithms need to use a vast percentage of validation and verification.
- Human knowledge and abilities are essential for fuzzy control systems.

B. Artificial Neural Network

Artificial Neural Networks (ANNs) are used as a classifier due to their computational simplicity. The proposed method features a multilayer feed-forward network. Training is carried out using Back Propagation (BPN) algorithms. Input layers, hidden layers, and output layers must be present. Based on the error in the classification, the hidden and output layer nodes adjust their weight values. The signal flow in BPN is based on although the error propagates backward, the weights are updated to reduce errors. The weights are adjusted in accordance with the error curve's gradient, which leads in the vicinity of the local minimum. This makes the algorithm very reliable both for prediction and classification.

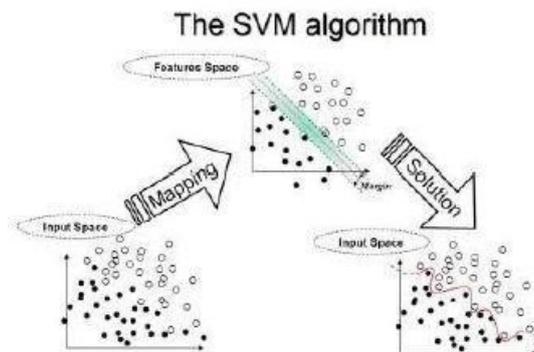


Figure 1 Support Vector Machine

C. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a statistical learning theory-based classification technique. It is focused on a hyper plane classifier approach. The objective of SVM is to generate a linear optimum hyper plane that maximises the distance between the two classes [2]. A diagram of the SVM procedure is shown in Figure 2.

D. Genetic Algorithm (GA)

This algorithm mimics natural selection, in which the fittest individuals are chosen for reproduction in order to generate the following generation's children.

1. Methodology proposed, selection as a parent The parent chromosome is the previous weight allocated to an error that occurred in the NN layer.
2. Dividing a parent into two parents.
3. Choose the parent's Crossover values, then swap the element to create new offspring.
4. Merge the two offspring and train NN with the newly produced child.
5. Set weights to the layer with the minimum amount of inaccuracy.
6. Selected values are then muted.
7. Repeat steps 1 through 6 for each NN Layers error that occurs.
8. Stop.

E. Web Intelligence

The implications of this definition are as follows: The basis of WI is AI and IT. The “I” happens to be shared by both “AI” and “IT”, although with different meanings in them, and “W” defines the platform on which WI research is carried out, [9]. WI's purpose is to deliver AI and IT combined on the Web's new platform. Related Topics in Web Intelligence-

1. Web Information Retrieval: Approximation query, conceptual web mining, picture retrieval, international information retrieval, multimedia collection, innovative retrieval methods, ontology-based knowledge discovery, and automatic Web content categorization and indexing are all examples of information retrieval techniques.
2. Web Mining and Farming: Web analysis and transformation, learning user profiles, multimedia data analysis, recognize patterns in Web surfing and Internet delays, information retrieval, Web-based ontology engineering, Web-based reverse engineering, Web farming, Web-log mining, and Web warehousing are all examples of web mining, personalised information management, semi-structured data management, use and management of metadata. Web knowledge management, Web page automatic generation and updating, and Web security are some of the topics covered.
3. Web Information Management: Information transformation, Web databases and OLAP (on-line analytical processing), Internet and Web-based data management, Data quality management, multimedia information management, new data models for the Web, object-oriented Web information management.
4. Web Information System Environment and Foundation: Web site competitive forces, evolving Web technologies, network community organization and maintenance, new Web information specification and query languages, smaller social Web theories, Web information system development mechanisms, and Web protocols are all topics covered.
5. Web Agents: Active information trends, e-mail sorting, email semi-automatic respond, global gathering information, information filtering, navigation advises, decision support systems, remembrance officials, reputation techniques, resource intermediary and coordination mechanisms, and Web-based cooperative problem - solving skills are just a few of the topics covered.
6. Web Human-Media Engineering: New website layout, multimedia content representation, multimedia data processing, Web data visualisation, and Web-based human computer interface are all examples of Web page design.

III. WEB INTELLIGENCE APPLICATION

3.1 Google Translate

- Google likes to make its language conversion tool more accurate and acceptable in its responses, therefore the company is turning to artificial intelligence (AI).
- In other words, it wants the program to do more than merely interpret phrases for you.
- The neural machine translation, according to Google, works by translating a whole piece at a go instead of just narrowing it down simultaneously.
- This enables it to better grasp the flow of a statement, delivering in more reliable and effortless translations.
- Google has also failed to convert interpret that is accessible to a larger user audience.

3.2 Google Scholar

- Find a list by googling for date, author, publication, or specific keywords and keywords- You might even narrow your relevant keywords and receive specific results by using the advanced search option. It includes a number of criteria to help you limit down your list. You can, for example, search for papers by author, journal, and/or publication date. You may also look for articles that include (or do not contain) certain keywords or phrases.
- Look for articles published in a certain language- You may also use Google Scholar to look for articles written in a given language. Sc holar offers results of search sites written in any language by default. This may be changed on the Settings tab on the left sidebar. Toggle the box next to 'Search exclusively for sites written in this language(s)' in Settings, then hit Save.
- Links to the library are also displayed- If you are a student or employee at an educational institution, you may link your institution's library to your Google Scholar account, and Scholar will display to you whether the retrieved results of the search contain library access links or not on each occasion you search for an article.
- Set up email notifications- Email notifications keep you informed about new writings on a specific topic. For example, you may set up an email alert for the topic 'digital literacies,' and each time an article on this topic is published, you will receive an email notification with a link to the new piece. You may also set up email notifications for your favourite writers and academics to receive emails with the newest news from them. Here's how to make a warning: Click 'Create alert' on the left-hand side of the Google Scholar search page, enter in your alert question and email address, and then click 'create alert'.
- Connect your bibliography manager- You might also connect Google Scholar to your preferred bibliography management and start receiving search results with links to significant citations in your bibliography manager. BibTex, EndNote, RefMan, and RefWorks are the ones that are presently supported.

IV. DATA MINING BASED SOFT COMPUTING APPROACHES FOR WEB INTELLIGENCE

Soft computing approaches, i.e., artificial neural networks are useful to handle the nonlinearities and unknown function approximation problems, based on fuzzy logic systems, expert's knowledge can be utilized to design intelligent systems, evolutionary algorithms are helpful to find global solutions in a complex search space [8]. Fuzzy systems is another alternative approaches for the identification and control of nonlinear systems due to the universal approximation capability of these systems [8]. The hybrid approach is proposed to discover hidden information and usage pattern trends which could aid Web managers to improve the management, performance and control of Web servers [6]. Consensus can be reached with the help of web intelligence and soft computing. Multiagent Web Services also take advantage of soft computing approaches for decision making. These are the possible implementations of data mining based soft computing approaches for web intelligence on e-commerce websites to track user behavior, identify users' needs, and deliver appropriate results.

V. CONCLUSION

In this paper, we specified the how Data Mining based Soft Computing has greatly impacted our today's world. Soft Computing methods, like Fuzzy Logic, Artificial Neural networks, Support Vector Machines and genetic algorithms recently, they've been utilized to tackle data mining challenges, and they're now being employed in web intelligence.

This research proposes that the soft computing approaches employed in data mining be applied to evaluating web forms based on the latest features, resulting in web intelligence.

ACKNOWLEDGMENT

I am grateful to 'Sitaram and Lady Shantabai Patkar College of Arts and Science, Mumbai India' for providing an opportunity to write a research paper in the form of a dissertation on the subject. "Soft Computing Methods For Web Intelligence Based On Data Mining."

I'd also want to thank my research professors for assisting me through the course of writing a research paper. Without their aid and continued role at every phase of the process, this task would not have been accomplished. I'd also like to thank my family and friends who helped me out with ideas and resources that supported me off tremendously.

REFERENCES

- [1]. A. Mathur and T. Agarawal "A Survey: Access Patterns Mining Techniques And ACO", International Journal of Engineering and Advanced Technology (IJEAT) ISSN:2249-8958, Vol. 2, Issue- 5, pp. 200-206, June-2013.
- [2]. C.J.C. Burges, " A tutorial on support vector machines for pattern recognition " , Data Mining and Knowledge Discovery , vol.2 , 1998, pp. 121-167.
- [3]. J. S. R. Jang, C. T. Sun and E. Mizutani, Neuro-Fuzzy and Soft Computing :A Computational Approach to Learning & Machine Intelligence, Prentice Hall, Saddle River, New Jersey, 1997.
- [4]. MAR. Deshmukh and PSR. Gupta, "Data Mining Based Soft Computing Methods For Web Intelligence", International Journal of Application or Innovation in Engineering and Management(IJAIEM), ISSN 2319-4847, Vol. 3, Issue- 3, pp. 376-382, March -2014.
- [5]. S.N. Sivanandam, S.N. Deepa, Principles of Soft Computing, Wiley India (P) Ltd, 2007.
- [6]. Xiaozhe Wang, Ajith Abraham and Kate A. Smith, "Soft Computing Paradigms for Web Access Pattern Analysis"
- [7]. YQ. Zhang and TY. Lin, "Computational Web Intelligence (CWI): Synergy of Computational Intelligence and Web Technology," Proceedings Of FUZZIEEE 2002 of World Congress on Computational Intelligence 2002: Special Session on Computational Web Intelligence, pp.1104-1107, May -2002.
- [8]. Yuehui Chen, "Hybrid Soft Computing Approach to Identification and Control Of Nonlinear Systems".
- [9]. YY. Yao, "Web Intelligence (WI): New Frontiers Exploration", University of Regina, Saskatchewan, Canada 2003