ISSN: 2959-6386 (Online), Vol. 1, Issue 1

Journal of Knowledge Learning and Science Technology
journal homepage: https://jklst.org/index.php/home

Data Engineering Evolution: Embracing Cloud Computing, Machine Learning, and AI Technologies

Jawaharbabu Jeyaraman[1] Muthukrishnan Muthusubramanian[2]
[1]TransUnion – USA
[2]Discover Financial Services-USA

Abstract

The evolution of data engineering has been greatly influenced by advancements in cloud computing, machine learning (ML), and artificial intelligence (AI) technologies. This paper explores the intersection of these domains and discusses how data engineering practices have adapted to leverage the capabilities offered by cloud platforms and intelligent systems. It provides insights into the integration of ML and AI techniques into traditional data management processes, highlighting the challenges and opportunities associated with this evolution. Additionally, the paper discusses the impact of these technologies on data processing, storage, analysis, and decision-making, emphasizing the need for organizations to embrace these innovations to stay competitive in today's data-driven landscape.

Keywords:
Data engineering, Cloud computing, Machine learning, Artificial intelligence, Data management, Data processing, Data analysis, Decision-making, Technology integration, Innovation.

ArticleInformation:
Article history: *Received:*05/06/2023 *Accepted:*10/06/2023 *Online:*25/06/2023 *Published*: 25/06/2023
Doi: https://doi.org/10.60087/jklst.vol1.n.p89
Corresponding author: Jawaharbabu Jeyaraman

Introduction

Introduction:

The rapid advancement of technology has brought about significant changes in the field of data engineering. With the emergence of cloud computing, machine learning (ML), and artificial intelligence (AI) technologies, data engineering practices have undergone a profound evolution. This evolution has enabled organizations to harness the power of large-scale data processing, storage, and analysis in ways that were previously unimaginable.

In this paper, we delve into the evolving landscape of data engineering, focusing on the intersection of cloud computing, ML, and AI technologies. We explore how these technologies have revolutionized traditional data

management processes, providing organizations with new tools and capabilities to extract insights from vast amounts of data. Moreover, we examine the challenges and opportunities associated with integrating ML and AI techniques into data engineering workflows.

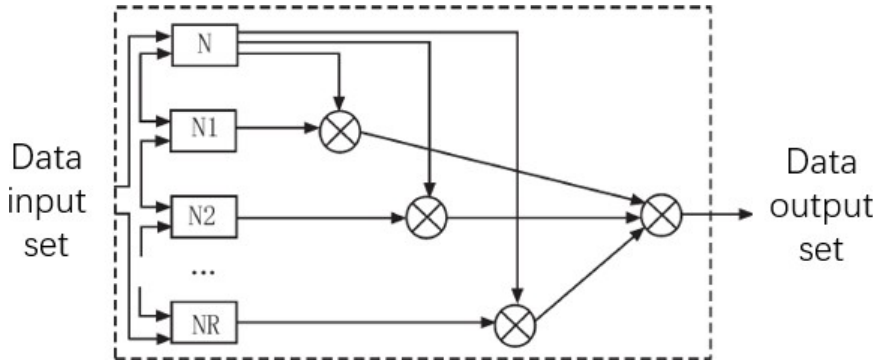
By embracing these innovations, organizations can gain a competitive edge in today's data-driven world. This paper aims to provide insights into the evolving role of data engineering and the transformative impact of cloud computing, ML, and AI technologies on data management practices. Through a comprehensive exploration of these topics, we aim to equip readers with a deeper understanding of the ongoing evolution of data engineering and the opportunities it presents for organizations across various industries.

Artificial Intelligence-based Data Mining Algorithms

The utilization of artificial intelligence (AI) in data mining involves the extraction of valuable insights from vast engineering datasets. However, this process encounters challenges such as data volume, noise, and rule extraction complexity due to the abundance of recorded data. A comprehensive data mining approach typically encompasses various stages including sampling selection, data processing, conversion, modeling, and evaluation.

In the realm of power engineering, several data mining techniques are commonly employed to tackle data processing and information extraction tasks. These include statistical analysis, decision trees, neural networks, fuzzy logic, and genetic algorithms. Particularly, the fuzzy neural network algorithm, which amalgamates the learning capabilities of neural networks with the fault tolerance of fuzzy systems, finds significant applications in power engineering for data processing and cost prediction.

In the application of neural networks for processing and predicting power engineering data, key challenges revolve around acquiring fuzzy system parameters and identifying fuzzy rule parameters. This entails determining the number of fuzzy rules and calculating membership, which are critical for effective model development. Given the complexity and variability of real-world scenarios, the fuzzy neural network algorithm necessitates the division of data space into fuzzy datasets through clustering methods. Subsequently, membership functions are derived through training to achieve the desired output. Figure 1 illustrates a simplified schematic diagram of a fuzzy neural network model



Data Preprocessing

In the analysis of power engineering data, a meticulous screening process is necessary to handle the large volume of

data involved in the project. The primary objective is to identify a purposefully selected dataset that accurately represents the attributes of the engineering data while retaining the integrity of the original information.

Typically, power engineering data is stored in numerical formats, prompting the use of Bayesian classifiers as evaluation functions in this paper. The Bayesian method is a traditional approach for classifying extensive datasets based on mathematical statistical principles. Assuming the dataset is denoted as S, with sample attributes represented by X1, X2, ..., Xn, and corresponding data types represented by C1, C2, ..., Ck, the probability that a new sample data belongs to a specific category Cj can be determined.

$$P(X|C_j) = (n_{C_j} h)^{-1} \sum_{X:C(X)=C_j} K(\frac{X-\mu_j}{h})$$

In the above formula, h = σ is the bandwidth, ncj is the number of samples, and K = g (X, 0, 1) is theGaussian distribution function. The posterior calculation probability calculation formula of the improved Bayesian classification algorithm is:

$$P(X|C_j) = \frac{1}{n_{C_j}} \sum_i g(X,0,1)$$

$$g(X,0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}$$

G is a Gaussian density function utilized to depict the probability distribution of data. With a mean of 0 and a variance of 1, it serves as a standard reference in statistical analysis. To illustrate, let's consider a specific power engineering dataset consisting of 100 data nodes. In this dataset, the attributes corresponding to each node undergo regularization. Consequently, the input and output data attributes for the power engineering dataset are structured as follows:

Input Set:
1. Voltage level
2. Line number
3. Transport distance
4. Terrain coefficient
5. Line length

Output Set:
1. Transportation engineering
2. Basic engineering
3. Wiring engineering
4. Attachment engineering

Data Model Establishment:

Based on the analysis findings of power engineering data discussed earlier, the multifaceted nature of engineering data necessitates a multi-dimensional approach to data analysis. Consequently, the challenge lies in addressing the nonlinear mapping problem inherent in such diverse datasets. To tackle this, the fuzzy neural network algorithm emerges as a viable choice as it serves as the cornerstone algorithm for engineering data processing, facilitating the development of a predictive model for power engineering costs.

The fusion of the fuzzy neural network algorithm with conventional neural network algorithms offers a unique advantage. It not only enables the acquisition of data patterns but also demonstrates robust network fault tolerance capabilities, particularly adept at handling complex nonlinear datasets.

Model Simulation:

The simulation process involves the manipulation of power engineering data. Beginning with 200 pieces of historical data post preprocessing, the dataset comprises 5 input attributes and 4 output attributes. Utilizing the K-means method, the power engineering data is classified. Following classification, adjustments to the membership function are made using the fuzzy neural network.

During the computation, the dataset is divided into 3 parts, consisting of training, testing, and verification sets with sample counts of 120, 40, and 4, respectively. The iterative process halts after 500 iterations. Additionally, to validate the results, a linear relationship between the output of the fuzzy neural network algorithm and the actual network output is established.

By employing the fuzzy neural network algorithm, power engineering data can be effectively analyzed and processed. Moreover, the data rules derived from the fuzzy neural network can aid in predicting the cost level of power engineering projects. Leveraging the fuzzy neural network, data rules are extracted from 200 sets of historical data, enabling accurate project cost predictions for specific projects.

Conclusion:

In this study, we introduced a fuzzy neural network algorithm leveraging artificial intelligence technology for the analysis and processing of power engineering data. Initially, the K-means algorithm was employed to conduct cluster analysis on the sample input space, generating the corresponding membership matrix. Subsequently, the neural network algorithm was utilized to train the data and conduct regression analysis on the sample data.

The proposed artificial neural network-based fuzzy neural network algorithm holds significant application potential and guiding significance for the comprehensive analysis and processing of power engineering data.

References:

1. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system; Manubot: 2019.

2. Mantelero, A. J. C. L.; Review, S., AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. 2018, 34 (4), 754 - 772.

3. Singh, M. P.; Huhns, M. N. J. I. E., Automating workflows for service order processing: Integrating AI and database technologies. 1994, 9 (5), 19 - 23.

4. Kowalski, R., AI and software engineering. In Artificial Intelligence and Software Engineering, Ablex Publishing: 1991; pp 339 - 352.

5. Artikis, A.; Bamidis, P. D.; Billis, A.; Bratsas, C.; Frantzidis, C.; Karkaletsis, V.; Klados, M.; Konstantinidis, E.; Konstantopoulos, S.; Kosmopoulos, D. In Supporting tele-health and AI-based clinical decision making with sensor data fusion and semantic interpretation: The USEFIL case study, International workshop on artificial intelligence and NetMedicine, 2012; p 21.

6. Yao, X. J. P. o. t. I., Evolving artificial neural networks. 1999, 87 (9), 1423 - 1447.

7. Basheer, I. A.; Hajmeer, M. J. J. o. m. m., Artificial neural networks: fundamentals, computing, design, and application. 2000, 43 (1), 3 - 31.

8. Zhang, G.; Patuwo, B. E.; Hu, M. Y. J. I. j. o. f., Forecasting with artificial neural networks:: The state of the art. 1998, 14 (1), 35 - 62.

9. Bratko, I., Prolog programming for artificial intelligence. Pearson education: 2001.

10. Luger, G. F., Artificial intelligence: structures and strategies for complex problem solving. Pearson education: 2005.

11. Bond, A. H.; Gasser, L., Readings in distributed artificial intelligence. Morgan Kaufmann: 2014.