Smart E-commerce Recommendations with Semantic AI

Mohamed Badouch[1,*], Mehdi Boutaounte[2]

[1]Faculty of science. University of Ibn Zohr, Agadir, Morocco

[2]National School of Commerce and Management. University of Ibn Zohr, Dakhla, Morocco

Emails: mohamed.badouch@edu.uiz.ac.ma; m.boutaounte@uiz.ac.ma

Abstract

In e-commerce, web mining for page recommendations is widely used but often fails to meet user needs. To address this, we propose a novel solution combining semantic web mining with BP neural networks. We process user search logs to extract five key features: content priority, time spent, user feedback (both explicit and implicit), recommendation semantics, and input deviation. These features are then fed into a BP neural network to classify and prioritize web pages. The prioritized pages are recommended to users. Using book sales pages for testing, our results demonstrate that this solution can quickly and accurately identify the pages users need. Our approach ensures that recommendations are more relevant and tailored to individual preferences, enhancing the online shopping experience. By leveraging advanced semantic analysis and neural network techniques, we bridge the gap between user expectations and actual recommendations. This innovative method not only improves accuracy but also speeds up the recommendation process, making it a valuable tool for e-commerce platforms aiming to boost user satisfaction and engagement. Additionally, our system's ability to handle large datasets and provide real-time recommendations makes it a scalable and efficient solution for modern e-commerce challenges.

Keywords: Semantic analysis; Recommendation semantics; User feedback; BP neural networks; Web mining, E-commerce

1. Introduction

In today's digital age, e-commerce has fundamentally transformed the way consumers purchase goods, offering unparalleled convenience and a vast array of choices. As online shopping becomes increasingly prevalent, the need for effective web page recommendation systems has grown. These systems are crucial for suggesting products that align with customers' shopping habits, thereby enhancing shopping efficiency, user satisfaction, and overall experience. To meet this demand, numerous personalized recommendation systems based on web mining have been developed. Web mining techniques analyze vast amounts of data generated by user interactions on e-commerce platforms. For instance, previous research has introduced sophisticated algorithms that utilize semantic web data mining to delve into user-visited pages and log data. By examining the content and context of these pages, these algorithms can construct predictive models that accurately forecast product preferences.
Additionally, some studies have employed user feedback mechanisms to refine these recommendations further. By collecting both explicit feedback (such as ratings and reviews) and implicit feedback (such as browsing history and click patterns), researchers have been able to train BP (Backpropagation) neural networks [(3)]. These neural networks are adept at determining the relevance and importance of various e-commerce web pages, ensuring that the recommendations are both accurate and personalized.

For e-commerce platforms there is no organized web directory [(1)], users mostly rely on alerts from search engines to find suitable e-commerce pages. This can provide room for errors in sending relevance signals since most traditional search engines usually match queries based on statistical frequency and similarity in words which results in semantic mismatches and wrong retrieval results. It also means that those who advertise pay more money to be ranked high thus making it unworthy to trust such searches. A more trustworthy means lies within the do-main of semantic web technologies that provide a better classification of contents depending on core terms relating to each

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

other or any other logical relationships. To illustrate this point better let us examine how earlier studies have shown that semantic web reduces misdiagnosis rates and intensifies data extraction effectiveness. To address these challenges, this paper proposes an innovative web page recommendation algorithm that optimizes e-commerce page ranking. By integrating semantic web mining (2) with BP neural networks, our approach effectively handles the ranking of diverse web pages. We implement an intelligent meta-search engine designed to accurately deliver the required e-commerce pages to users, thereby improving the overall online shopping experience.

2. Methods

Backpropagation neural networks are a sophisticated class of artificial neural networks that utilize a specific learning process known as backpropagation. This method is crucial for training the network to minimize errors in its predictions, making it highly effective for tasks such as classification, regression, and recommendation systems (4).

Phases of BP neural networks

Backpropagation is a supervised learning algorithm used for training neural networks. It involves two main phases: the forward pass and the backward pass with error calculation (5).

Forward Pass:
- During the forward pass, input data is fed into the network.
- The data passes through multiple layers of neurons, each performing computations and applying activation functions.
- The final output is produced at the output layer, representing the network's prediction.

Error Calculation:

- The network's output is compared to the actual target values (ground truth).
- An error is calculated based on the difference between the predicted output and the actual target. This error is often measured using a loss function, such as mean squared error or cross-entropy loss.

Backward Pass:

- The error is propagated back through the network from the output layer to the input layer.
- The goal is to adjust the weights and biases of the network to minimize the error.
- This is achieved using gradient descent, where the gradient of the loss function with respect to each weight and bias is computed.
- The weights and biases are updated in the opposite direction of the gradient, reducing the error in subsequent iterations.

Layers of BP neural networks

This strategy involves using a three-layered Back Propagation (BP) neural network as a classifier for training and recognition, the network is comprised of three parts, namely the input layer, the hidden layer, and the output layer (6):

- Input Layer: The first layer that receives the raw input data. Each neuron in this layer represents a feature of the input data.
- Hidden Layers: Intermediate layers where the actual computation and learning occur. These layers can have multiple neurons, each applying a weighted sum of inputs followed by an activation function.
- Output Layer: The final layer that produces the network's prediction. The number of neurons in this layer corresponds to the number of output classes or values.

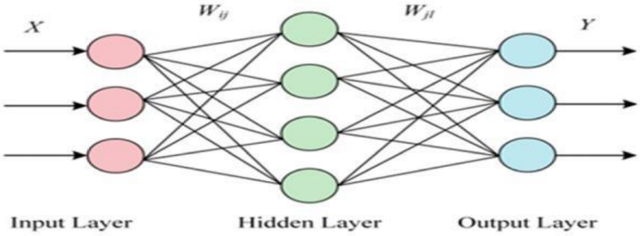

Figure 1. Simplified schema of BP neural networks.
Source: Own elaboration.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

Mathematic behind BP neural networks

We leverage the **traingd** function for training and the **tansig** function for inter-layer transfers [7]. The training process involves adjusting the network's parameters to minimize prediction errors. Specifically, we derive the weight matrix (W_{ji}) and sill matrix ($\emptyset_j$) from the input layer to the hidden layer. These matrices are crucial for capturing the relationships between input features and hidden neurons [18]. Additionally, we obtain the weight matrix (V_{ki}) and threshold matrix (φ_k) from the hidden layer to the output layer, which are essential for mapping the hidden layer's activations to the final output [8].

$$S_j = \sum_{i=0}^{I} W_{ji} X_j - \theta_j, j = 1,2,\dots,J$$

To begin with, compute the input (s,j) for the (j) neuron in the hidden layer. This computation uses the weight matrix (W_{ji}) and the bias term ($\emptyset_j$). The output (h_j) of the (j) neuron in the hidden layer is then derived by:

$$h_j = g(W_j) = \frac{2}{1+e^{-2sj}} - 1, j = 1,2,\dots,J$$

Next, determine the input (r_k) of the output layer using (V_{kj}) and (φ_k) by:

$$r_k = \sum_{j=1}^{J} V_{ji} h_j - \varphi_k, k = 1,2,\dots,K$$

Derive the output values (k) and (y) from the output layer by:

$$y_k = g(r_k) = \frac{2}{1+e^{-2rk}} - 1, k = 1,2,\dots,K$$

To wrap up the process, determine the recognition result by locating the maximum value among the recognition similarity (values (y_k)). This step ensures that the most similar match is identified accurately.

3. Development

3.1. Semantic web mining

Semantic web mining [9] in e-commerce represents a sophisticated intersection of semantic web technologies and data mining techniques, aimed at enhancing the functionality and user experience of online retail platforms. This approach enables the precise interpretation and organization of web data, facilitating more accurate search results and personalized product recommendations. By leveraging user behavior and preference data, semantic mining can significantly improve the intuitiveness and personalization of the shopping experience. Furthermore, it allows for the efficient extraction of valuable patterns from extensive datasets, thereby supporting data-driven decision-making processes within e-commerce businesses [10]. This methodology addresses the inherent limitations of traditional e-commerce systems, which often suffer from a lack of standardized data management practices. By providing a structured and standardized framework, semantic web mining ensures a more coherent and efficient handling of data, ultimately leading to enhanced operational efficiency and user satisfaction.

3.2 Recommended solution framework for e-commerce

This project aims to rank web pages based on user search information, semantic web mining, and a BP neural network, thereby recommending web pages that meet user needs and increasing user engagement time on the website. The proposed solution employs a modular design, comprising preprocessing and dictionary implementation modules, a time consumption priority module, a content priority module, a neural network priority module, and a semantic recommendation module [16].

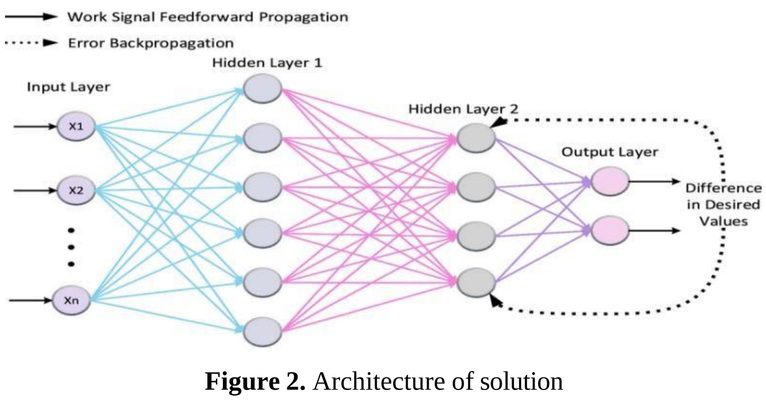

Figure 2. Architecture of solution
Source: Own elaboration.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

In the preprocessing and dictionary implementation module, candidate web pages retrieved by search engines undergo preprocessing to remove incomplete input entries and clean data and stems. The dictionary implementation module inputs the website URL to facilitate web log preprocessing, user navigation analysis, and the creation of a web page dictionary consisting solely of words from candidate web pages, which correlate with the length of words used by users to search for specific e-commerce products [17]. The content priority module utilizes web dictionaries and candidate web pages as input, employing web content mining technology to classify documents into an ordered structure. This process extracts knowledge from web document content or descriptions, thereby determining the relevance and priority of web pages and filtering out e-commerce website pages unrelated to service products.

In the time consumption priority module, web pages are input, and candidate web page timestamps are used to determine their priority. The timestamp reflects the statistical value of the time previous users spent on the same product web page, assigning higher priority to web pages with longer user engagement [22]. The semantic recommendation module continues to use web pages as input, applying the longest common subsequence algorithm to identify user session data from different semantic behavior files. This module determines the ontology category to avoid incorrect interpretations of user retrieval queries.

Finally, the overall priority of e-commerce web pages is determined through a BP neural network. The five input features of the BP neural network include content priority, time consumption priority, and explicit/implicit user feedback on candidate websites, recommendation semantics, and input bias. The neural network randomly assigns weights to all input nodes, compares the actual output with a manually set threshold, and adjusts the input node weights adaptively based on the error tolerance feedback until the correct output is achieved [19]. This fine-tuning is crucial for supervised learning and significantly improves the accuracy of prioritizing web pages relevant to user searches.

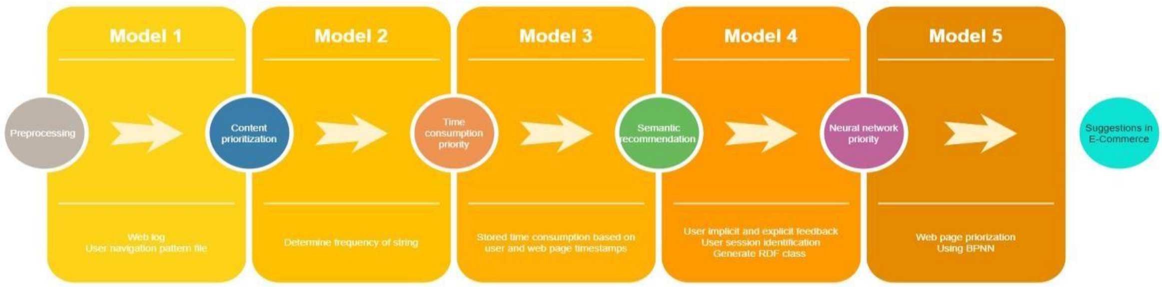

Figure 3. Models of the solution
Source: Own elaboration.

Given the five feature inputs, the neural network's input layer [13] consists of five neurons, while the output layer comprises a single neuron representing the correlation between the web page and the user's preferences. For the hidden layer, the number of neurons is optimized based on existing literature, balancing classification performance and speed. The optimal number of neurons in the hidden layer is determined by the formula: the number of neurons in the input layer multiplied by the square root of the number of neurons in the output layer plus an integer (L) (ranging from 0 to 5) [20]. This project selects the number of neurons in the hidden layer to be four, based on multiple experimental analyses, ensuring a balanced consideration of classification performance and speed.

3.3 Algorithm for recommendation (models' details)

The algorithm for the e-commerce recommendation scheme proposed in this article, which leverages semantic web mining and a BP neural network. We present the variables used in this algorithm along with their meanings.

- **DW**: Document word.
- **Wi**: search stage specific keywords.
- **Ts**: Web page timestamp.
- **Tan f**: Linear activation function for training in neural networks.
- **WDp**: Web dictionary corresponding top Web document.
- **Si**: E-commerce product string retrieved by the user.
- **Min**: minimum length of any keyword in Si.
- **Max**: maximum length of any keyword in Si.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

We will provide you all the stages in detail that we adopt in our paper:

1) Getting the retrieval string from the user.
2) Removing stems from the search string.
3) Recording the guided sequence pattern to the user file database.
4) Using a search engine to retrieve web documents.
5) Dividing the string into different distinct words $W1$, $W2$,..., Wn.
6) Determining the minimum and maximum length of different words within the search statement.

min: $=Strlen(W1)$, max: $=Strlen\ (W1)$
for i =2 to n do
if min > $Strlen\ (Wi)$ then
    min: = $Strlen\ (Wi)$
if max < $Strlen(Wi)$ then
    max: = $Strlen\ (Wi)$

7) Initializing Ti of each document to 0.
8) Retrieving the time database based on the keywords entered by the user and retrieve the same documents through the search engine when retrieving the Ti step.
9) Preprocessing each web document Dj in WDi. There are only DWkwords in Dj and satisfy the condition $Strlen\ (DWk)$ <= max.
10) Initialization:
    for p =1 to m do
    initialization $foundp$: = 0, $nfoundp$:=0
    if Wp is found in WDp then
    $foundp$: = $nfoundp$ + 1
    else
    $nfoundp$: = $nfoundp$ + 1
11) Clearing web page $nfoundp$> $foundp$.
12) Determining the timestamp of a web page Ts.
13) At the onset of the user session, determine the duration of the user session on the current web page tp, and determined according to the following equation Tp:
    if Tp =0 then
    $Tp = tp$
    else
    $Tp = (tp + Tp)/2$
14) If is lower and Tp is higher, a higher weight is assigned to the web page.
15) Updating time information with keywords, web addresses and Tptools.
16) User-directed sessions are identified through a comparative analysis between the user's search query and each entry within the user file database, conducted as follows:

LCS[i,j]=0, if i =0 or j =0 OR LCS[i,j] = LCS[$i-1$,$j-1$] +1, if i, j ≠0 and $S1i = S2j$ OR LCS[i,j] = max ($LCS[i-1,j]$, $LCS[i,j-1]$),
if i,j > 0 and $S1i \neq S2j$

17) Generating categories using a web ontology language.
18) Normalizing all priority inputs in modules 2, 3, and 4.
19) Using the input and output data sets for training a BP neural network with a linear activation function, as described below:
    $\{O\} = tan\emptyset\{I\}$
20) In the hidden layer and output layer, use the sigmoid function to predict the output. The S-shaped function is
    $\{O\} = [1/(1+ e^{-I})]$, and the summation function is $\sum(I1WT1 + I2WT2 + I3WT3 + I4WT4 + I5WT5 + B)$
21) Calculating the error rate for weight adjustments in network neurons using the supervised BP neural network algorithm.
22) Suggestions in E-commerce by priority

4. Results and Discussion

4.1 Configuration of Experiments

In this study, the JavaEE framework is utilized to implement the proposed recommendation algorithm, with its performance compared to existing literature. An experiment was conducted on a website selling books, with a database containing 6,200 books categorized into five main groups: education, humanities and social sciences,

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

technology, literature, and life. Each main category is further divided into six subcategories. For example, the technology category includes subcategories such as popular science, computers, architecture, medicine, agriculture and forestry, and general science. Over time, more than 10,000 transaction events were recorded. The dataset was divided into five parts: one part served as the test set, while the remaining four parts were used as the training set. To evaluate the effectiveness of the web page recommendation system, two commonly used performance metrics—precision and recall—were employed. Precision measures the proportion of relevant documents in the recommendation results, i.e., the ratio of recommended products identified by the recommendation model to all products related to the recommended topic. The formula for precision is as follow:

$$PRE = \frac{\sum_{u \in U} R(u) \cap T(u)}{\sum_{u \in U} R(u)}$$

R(u) represents the list of recommendations generated by the algorithm for the test set. T(u) is the actual list of recommendations on the test set. U denotes the set of users, and I stand for the set of products. The recall rate quantifies the percentage of relevant documents included in the recommended results out of the total relevant documents in the entire collection. It evaluates whether the recommendation system has successfully identified all pertinent pages. The formula for recall is as follows:

$$PRC = \frac{\sum_{u \in U} R(u) \cap T(u)}{\sum_{u \in U} T(u)}$$

4.2 Comparison of Precision

R(u) The accuracy and recall rates of the proposed scheme and the scheme from literature on the test set are illustrated in Figures 4 and 5. In this context, the length of the user's access sequence is taken as the x-axis. The access sequence represents the process of a user's continuous navigation through the network. For example, if a user's access sequence is <A, B, C, D>, it means the user first visits URLA, URLB, then URLC and finally URLD [11]. The longer the access sequence, the more information it contains about user preferences.

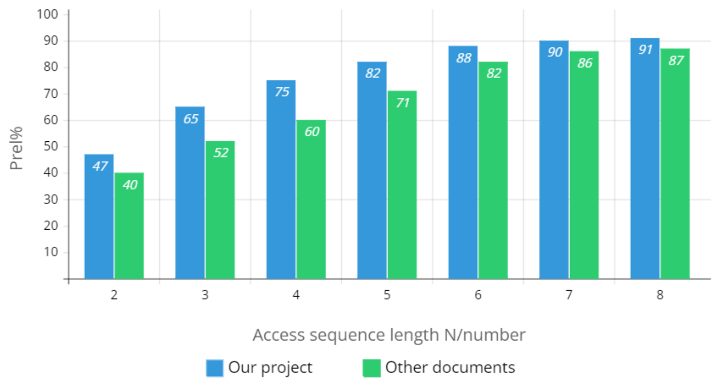

Figure 4. The rate of precision for the two schemes
Source: Own elaboration.

Figures 4 and 5 illustrate that both the accuracy and recall of website recommendations improve as the access sequence length increases. This is because longer access sequences provide more representative features of user preferences through web data mining. However, when the access sequence length reaches 6, various performance metrics tend to stabilize. Notably, the scheme proposed in this paper outperforms the one in literature. When the sequence length is 8, the accuracy and recall rates of this scheme are 91.7% and 90.3%, respectively, which are 6.2% and 4.9% higher than those in literature. Experimental results demonstrate that this solution can better identify the web pages needed by users and is well-suited for ranking and recommending e-commerce websites.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

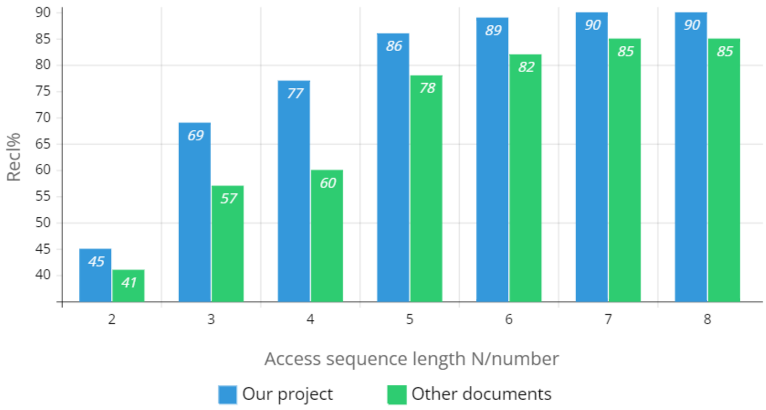

Figure 5. The rate of recall for the two schemes
Source: Own elaboration.

4.3 Comparison of usage time

The computational efficiency of this method is compared with that of the method described in literature. From the collected web log files of over 6,200 book transaction events, subsets of 10K, 50K, 100K, 500K log files were extracted. The time taken by the two algorithms to process these files is shown in Table below.

Table 1: Response time result for both schemes.

File size	10,000	50,000	100,000	500,000
Existing	1.63 min	23.22 min	1.55 h	18.61 h
Our schema	58 s	14.47 min	59.32 min	12.38 h

Source: Own elaboration.

Table 1 shows that, for the same file size, the time taken by the model in existing project is significantly higher than that of the solution proposed in this paper. This trend becomes more pronounced as the file size increases. Existing project attempts to predict current user access behavior by analyzing the access records of many web users. However, since web access is influenced by various factors such as cultural background, interests, hobbies, and browsing purposes [15], and is not limited by space and time, each user's web browsing behavior varies greatly. Using a semantic model to describe the browsing behavior characteristics of all users and make predictions inevitably results in low accuracy and high time and space complexity.

5. Conclusion

In this research work we describe a new system for recommending products from e-commerce portals using semantic web mining and BP neural networks. The importance of user preferences may help in designing better website structures for webmasters, who seek to create visually dynamic sites that incorporate design patterns which optimize user experience. To compare its results with the existing methods stated in other works, an experimental database containing book sales websites was built; hence, performance measures include accuracy of page recognition together with recall rates. Results show that our proposed technique is noble; it can learn from customer input while keeping low power consumption levels. The rise of deep learning has seen its application areas diversify into such fields as image processing, medical image analysis and image classification among others [21]. There is no doubt that the future of e-commerce will be defined by deep learning technologies. Face recognition and future trend pre-diction based on past financial data can also be accomplished using neural networks and deep learning technologies [12]. When merged with machine learning, these technologies are also helpful in climate prediction. Additionally, algorithmic techniques are employed for classification and autonomous learning [14]. Integrating these algorithms with user purchasing and browsing data to predict e-commerce sales trends represents a promising research direction.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024

References

[1] Gonçalves, R.; Rocha, T.; Martins, J.; Branco, F.; Au-Yong-Oliveira, M. Evaluation of e-commerce websites accessibility and usability: an e-commerce platform analysis with the inclusion of blind users. Universal Access in the Information Society 2018, 17, pp. 567-583.

[2] Bashir, F.; Warraich, N. F. Systematic literature review of Semantic Web for distance learning. Interactive Learning Environments 2023, 31(1), pp. 527-543.

[3] Liu, Y.; Jiang, C.; Lu, C.; Wang, Z; Che, W. Increasing the accuracy of soil nutrient prediction by improving genetic algorithm backpropagation neural networks. Symmetry 2023, 15(1), pp. 151.

[4] Abdolrasol, M. G.; Hussain, S. S.; Ustun, T. S.; Sarker, M. R.; Hannan, M. A.; Mohamed, R.; Milad, A. Artificial neural networks based optimization techniques: A review. Electronics 2021, 10(21), pp. 2689.

[5] Wang, X.; Lin, X.; Dang, X. Supervised learning in spiking neural networks: A review of algorithms and evaluations. Neural Networks 2020, 125, pp. 258-280.

[6] Kumaraswamy, B. Neural networks for data classification. Artificial intelligence in data mining 2021, pp. 109-131.

[7] Alali, A. M.; Abughaban, M. F.; Aman, B. M.; Ravela, S. Hybrid data driven drilling and rate of penetration optimization. Journal of Petroleum Science and Engineering 2021, 200, pp. 108075.

[8] Dawani, J. Hands-On Mathematics for Deep Learning: Build a solid mathematical foundation for training efficient deep neural networks. Packt Publishing Ltd 2020.

[9] Kumar, B.; Roy, S.; Sinha, A.; Iwendi, C.; Strážovská, Ľ. E-commerce website usability analysis using the association rule mining and machine learning algorithm. Mathematics 2022, 11(1), pp. 25.

[10] Necula, S. C.; Păvăloaia, V. D.; Strîmbei, C.; Dospinescu, O. Enhancement of e-commerce websites with semantic web technologies. Sustainability 2018, 10(6), pp. 1955.

[11] Abad, S.; Gholamy, H.; Aslani, M. Classification of Malicious URLs Using Machine Learning. Sensors 2023, 23(18), pp. 7760.

[12] Zhou, L. Product advertising recommendation in e-commerce based on deep learning and distributed expression. Electronic Commerce Research 2020, 20(2), pp. 321-342.

[13] Yamashita, R.; Nishio, M.; Do, R. K. G.; Togashi, K. Convolutional neural networks: an overview and application in radiology. Insights into imaging 2018, 9, pp. 611-629.

[14] Shaheen, K.; Hanif, M. A.; Hasan, O.; Shafique, M. Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. Journal of Intelligent & Robotic Systems 2022, 105(1), pp. 9.

[15] Yap, A. Y. Enabling E-Commerce Growth through the Social Construction of a Virtual Community's Culture. Journal Electron Commer Res. 2002, 3(4), pp. 279-294.

[16] Maher, M.; Ngoy, P. M.; Rebriks, A.; Ozcinar, C.; Cuevas, J.; Sanagavarapu, R.; Anbarjafari, G. Comprehensive empirical evaluation of deep learning approaches for session-based recommendation in e-commerce. Entropy 2022, 24(11), pp. 1575.

[17] Opara, C.; Chen, Y.; Wei, B. Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics. Expert Systems with Applications 2024, 236, pp. 121183.

[18] Liu, T.; Yang, L. Financial Risk Early Warning Model for Listed Companies Using BP Neural Network and Rough Set Theory. IEEE Access 2024, 12, pp. 27456-27464.

[19] Ramanujan, V.; Wortsman, M.; Kembhavi, A.; Farhadi, A.; Rastegari, M. What's hidden in a randomly weighted neural network?. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020, pp. 11893-11902.

[20] Shen, Z.; Yang, H.; Zhang, S. Neural network approximation: Three hidden layers are enough. Neural Networks 2021, 141, pp. 160-173.

[21] Badouch, M.; Boutaounte, M. Design and Implementation of a Hotel Recommendation System Using Deep Learning. AI and Data Engineering Solutions for Effective Marketing, edited by Lhoussaine Alla, et al., IGI Global, 2024, pp. 388-408.

[22] Zhang, X.; Sen, S.; Kurniawan, D.; Gunawi, H.; Jiang, J. E2E: embracing user heterogeneity to improve quality of experience on the web. In Proceedings of the ACM Special Interest Group on Data Communication 2019, pp. 289-302.

DOI: https://doi.org/10.54216/FPA.170120
Received: November 29, 2023 Revised: March 25, 2024 Accepted: August 04, 2024