International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 18th Jul 2023. Link

:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 07

10.48047/IJIEMR/V12/ISSUE 07/05

Title **Multiple Similarity Assessment (MSA) Method for Data Optimization Based on Artificial Intelligence**

Volume 12, ISSUE 07, Pages: 47-53

Paper Authors **Hymavathi Sabbani, N Lavanya, Ramu V, S. Santhosh Kumar**

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Multiple Similarity Assessment (MSA) Method for Data Optimization Based on Artificial Intelligence

Hymavathi Sabbani[1], N Lavanya[2], Ramu V[3], S. Santhosh Kumar[4]
[1,2,3,4]Assistant Professor in CSE (AI&ML) Department, Balaji Institute of Technology and Science, Narsampet 506331.
hymavathisabbani@gmail.com[1], lavanyamatcha@gmial.com[2], ramubits2022@gmail.com[3]
santhu3513@gmail.com[4]

Abstract
Due to their capacity to enable Data users to browse the Internet in a personalized manner, similarity systems are widely used. For instance, the collaborative Similarity system is a potent data personalization tool that makes a variety of helpful recommendations to a particular user based on feedback gathered fromneighbors. One of the elements influencing the success of the collaborative Similarity system is the similarity measure. Additionally, computer science's branch of machine learning and artificial intelligence, both of which are designed to improve human intelligence, are related technologies. AI and ML can be utilized in e-healthcare to improve workflow, automatically handle volumes of medical data, and offer useful medical decision assistance. The authors of this work take several popular artificial intelligence models that are now available in the present research studies. This paper proposes a mechanism for allocating jobs to store enormous amounts of data load for cloud resources to balance the infrastructure platforms' demands for big data and artificial intelligence. Using Bayesian theory, the maximum posterior probability for each physical host is determined. The experimental results on the benchmark datasets show that the proposed data classifier is computationally affordable and comparable with cutting-edge methods. The significance of the experimental results has been compared using the conventional multi-task load balancing method.

Keywords—recommender system; artificial intelligence (AI); machine learning (ML); big data; data optimization

Introduction
Compare to recurrent neural, Boltzmann machine, etc., convolutional neural network (CNN) is a type of artificial neural network. Because the visual system is driven by neural mechanisms, the biological model of the convolutional neural network is capable of recognizing two-dimensional shapes. There are several options available to data consumers today when they browse the Web. To provide data users with personalizeddata they might use, similarity systems—as well as numerous other Web personalization tools—become important. Artificial intelligence models are used to make judgments based on huge data [1-8]. As artificial intelligence develops, both input and computation are required; hence, the underlying big data and real-time data necessity are known to boost the accuracy of the results obtained. The purpose of this study is to put into practice a heuristic technique called MSA uses large data processing and artificial intelligence. For each execution of the currently required task, the MSA approach, which is based on an artificial intelligence algorithm, will find the best possible selection of physical hosts. This encourages us to research the effectiveness of various similarity measures for collaborative recommender systems based on the available value cardinality. The comparison between the two consumers clearly shows that one of the 20 highly rated products is inferior. The second scenario is likewise more likely because it is certain that there would be small groups of neighbors. In this study, the performance of the

collaborative recommender system is assessed using three factors: the cardinality of the commonly rated data, the rating mean group, and the number of neighbors.

This work has three contributions:
(1) The idea of user data usage is presented as an alternative to user data aggregation.
(2) For the mean difference weights similarity measure, we provided a modified method, and
(3) Both artificial and actual data sets are used in our investigations.

The majority of available information on convolutional neural networks is unclear. The unique aspect of this study is how it conducts a brief test case for artificial intelligence model optimization for data classification and recognition using a convolutional neural network. It can give new-research to the industry advisors for how to begin working practically.

The remainder of this work is structured as follows: A review of the literature and an introduction to collaborative recommender systems are provided in Section 2, and Section 3 introduces the impact of the cardinality of the common set on the effectiveness of various similarity measures. Through discussions of the experiment's findings, Section 4 outlines the experimental approach utilized for evaluating various similarity metrics. In Section 5, we conclude our proposed work.

Previous Works:
However, they chose the smallest number of shared features in advance and assessed their thoughts in light of that number [1–9]. DeepFixis one of them. Its methodology employs a hierarchical approach to automatically learn features and predict saliency maps from beginning to end [1]. Their model structure, however, is far too rigid. Numerous studies have discussed and proposed a wide range of similarity measurements. [3] researched established methods and recommended the power coefficient as a metric for similarity, even though considered the average set size to be more than or equal to five. A low size of the common set had an impact, as [9] found after doing an empirical investigation for numerous similarity metrics employed in collaborative filtering. To improve system performance, they advised default voting for unrated metrics. Numerous authors employ the same strategy to get around the lack of common objects [10]. This is consistent with the conclusions of [11], where they demonstrated that users of their system had higher levels of choice confidence when there was a substantial rating overlap between the recommender and the decision maker. To increase the recognition accuracy of noise recognition, [12] created a convolutional neural network-based architecture. They looked at the ideal configuration of filters, pooling, and input feature map size before proposing their architecture. The thorough examination of the architecture by [12], however, is still insufficient [2]. In[13], used a unique multi-dimensional convolutional neural network in the field of medicine. Use a network model to classify lung nodules for potential cancer-patients. Potential uses of the recommended strategy for nodes are provided by the way their approach captures multi-scale node salient qualities in a single network [3]. Using second-order statistics, [13] introduced a new class of convolutional neural networks. To accomplish this, they created severalnew layers that could be joined. Except for a minor accuracy issue, it replaces the fully linked layers of conventional convolutional neural networks [4]. Convolutional neural networks were used by [13] to learn distinguishing features for action recognition and suggested an efficient approach for encoding with sequential spatiotemporal information. However, its drawback is that the application's scope is too narrow, and the learning processes are very difficult [5]. The DLB is a purely combinatorial, classic NP-hard optimization problem [12]. Heuristic dynamic algorithms are frequently used in dynamic load balancing to provide effective load balancing in realtime by allocating resources and workloads. Light loads are gathered and scheduled first via the Load Receiver method (LRS) algorithm, which

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

employs a greedy method [13]. A fresh combinatorial model was created by [14] to dynamically optimize the objective. A heuristic load assignment approach based on the fusion of heavy and light loads was created by [14] to achieve dynamic load balancing with minimum communication overhead and use modest loads. The VMware virtualization product uses Distributed Resource Scheduling (DRS) as its load-balancing method [16]. By analyzing the load on each server, VMware chooses a placement strategy when using the DRS method to choose physical hosts that enhances overall load balancing [17]. To dynamically schedule virtual machines, DRS employs VMware VMotion technology to continuously analyze the load condition of all hosts. Through the dynamic movement of virtual machines, DRS achieves dynamic optimization of load balancing. Designed by [15] method for dynamically scheduling virtual machines that reduces communication costs and achieves load balancing. A strategy for deploying virtual machines that assumes significant application requirements is suggested by [17]. The SLB algorithm is used by the Eucalyptus platform to provide load balancing [10]. [11] used the SLB technique to determine host weights and deploy virtual machines on physical hosts with the lowest weight ratio. Based on the idea of speeding up data transfer, [17] created a network-aware virtual machine assignment method, which enhanced the performance of the cloud computing platform. To boost the processing capability of cloud computing systems, and created a hybrid technique combining dynamic and static resource pre-allocation scheduling.

Proposed Work:

Deep Learning (DL), a machine learning approach that relies on backpropagation and mathematical optimization models to deliver spectacular results, has significantly advanced and developed AI in the modern era. The aim of developing AI with human intelligence is still a long way off, and it is questionable whether the present AI paradigms can get us there. The SS (Similarity System) matches the active user to the training user database that is accessible during the similarity computation phase using the most appropriate similarity metric, as shown in Figure 1.

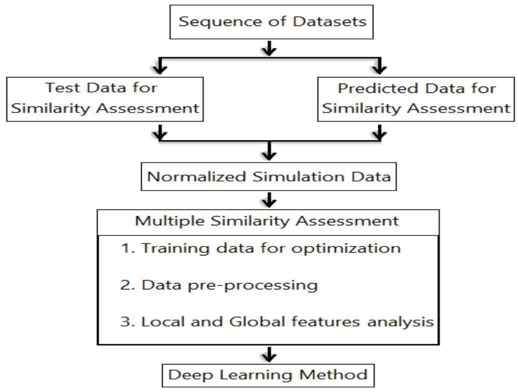

Figure 1: Similarity Assessment Model (SAM)

This number represents how similar two users are to one another as soon as similarity values are the algorithm computes a user's neighbors by ranking users based on how similar they are to the active user.From figure 1, we create an example dataset with 12 data for 19 users. 18 other users are training users while the first user is the active user. An item's zero rating means that the user has not yet given it a rating and can therefore suggest it to him. The sample data set must meet the following requirements: (i) It must include a large number of the common set cardinalities. So, we choose several values, including 2, 3, 6, 9, and 11. (ii) It is important to consider the small, medium, and high-ranked mean groups equally. (iii) The sample data is built up so that for each cardinality of the common set, one opposed user and three users with various rating means (low, medium, and high) are reachable. (iv) The final two users, who both have data 8 and 10, demonstrate individuals with opposing viewpoints to the one being used. (v) To study the effects of raising the cardinality of the common set without altering the previous collection of data, for each rating mean group, the user with the larger cardinality of the common set inherits the identical things from the user with the lower cardinality.The values of similarity between training and active

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

users comprise the example dataset for the four similarity indices discussed previously.

These are the specific SAM steps:

(1) Based on the local maximum and lowest values of the original signal, we represent the upper and lower envelopes, respectively.
(2) To get the mean envelope, we compute the average value of the upper and lower envelopes.
(3) To produce the intermediate signal, we permit the original signal to be subtracted from the mean envelope.
(4) Phases (1) to (3) will be repeated using the mean envelope as the original signal if the intermediate signal satisfies the two requirements ofi-date (intermediate data).

If condition is not satisfied, the intermediate signal will be taken into consideration as the original signal, and the subsequent processes (1) through (4) will be repeated. In Figure 1, the data signal is broken down and rebuilt, then normalized features are extracted from the reconstructed signal, and finally, the features are fed into the Deep Learning Method (DLM) network to train the deception detection classifier. Some of i-data's produced by DLM processing may be inefficient or even inhibiting in their ability to detect false data. Therefore, it is likely that some i-data's might be eliminated to enhance the detection of false data. This finding motivated as the input for our suggested scheme. After being chosen based on the correlation between the magnitude of the features and the threshold value, the components are then again merged to create the signal that was rebuilt. Figure 2 depicts the entire procedure.

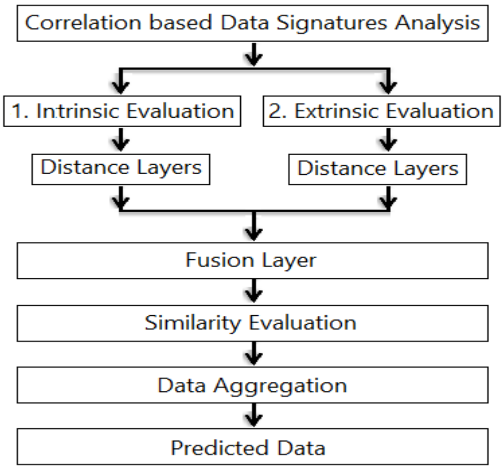

Figure 2: Proposed Model Analysis

As illustrated in Figure 2, after the signal processing previously discussed, the features are recovered and a classifier is trained.

Figure 2 depicts the entire procedure, and the list of stages below covers each stage in detail:

(1) Each date is multiplied by a Data Signature Analysis (DSA), following the division of the data signal into a layers-level representation.

(2) The data is transformed into anevaluation representation using the DLM.

(3) The evaluation output of each DLM is calculated after each data has been run through it.

(4) The standard DSA values are then created using the Fusion Layers (FL's).

(5) The first and second-order DSA values are calculated.

To obtain the necessary DSA values, difference features are computed and merged with standard features. The process of determining the priority probable value is required with that used by DLM method since it offers an effective way of modifying the prime estimate using the information received. Subjects assess each hypothesis before choosing a host, known as the a priori probability value. The chosen host may be unable to complete the desired task. The MSA approach determines the ideal host set for each work assignment by analyzing the artificial intelligence algorithm. Deliver effective long-term load-balancing services in response to customer requests for big data processing.

Comparative Results and Discussions:
By comparing the recently proposed method with the dictionary-based, longest matching, and new methods, this study evaluates its effectiveness. The length of the immediate context, which was set to 4, 8, and 16, was the first parameter to be assessed. The size of the DSA, which was set to 512, 1024, or 2048, was the second parameter. The greatest results came from running the experiment with anDSA size of 2048 and a surrounding context length of 8. The term TP (True Positive) in the equation above stands for the quantity of correctly detected word segments, whereas FP (False Positive) stands for the number of word segments that were incorrectly identified. The number of unrecognized cases is represented by FN (False Negative) word groups.

The evaluation parameters used are:

$$\text{Precision} = \frac{\text{the number of correct words}}{\text{the number of word predictions}} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Recall} = \frac{\text{the number of correct words}}{\text{the number of words in the ground truth}} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$F1 - \text{Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Precision} + \text{Recall}},$$

According to the first assessment, which was done on the dataset Best2010, comparisons are made. Proposed work performs slightly better than Deepcut in terms of precision and F1-score even if its recall performance is greater at 96%. Proposed work's processing time is also longer than Deepcut's since it has more rules to verify. These findings, however, demonstrate that proposed work offers a different methodology and is on par with cutting-edge techniques. It should be mentioned that proposed work have as moderate precision because of the problem of unknown words, which even the suggested solution cannot fully solve, leading to some inaccurate segmentation. A different dataset, LST20, was used for the second assessment in Table 1.

The Proposed Work results (have strong recall but lower precision than those in the first evaluation. Although Proposed Work performed somewhat better than Deepcut, it still outperformed the longest and new techniques. This series of experiments was examined by comparing the incremental percentages of DSA load balancing standard deviation values as the number of task requests increased. The load balancing effect of DSA's cloud computing platform is more pronounced for large-scale data processing computing operations, as shown in Table 1. This collection of tests does not use the FL's assignment approach since it is good in heuristic data and adaptable capabilities. It is important to emphasize this as well. Therefore, the significance of compression participation in validation for this simulation test is negligible.

Table 1: Data Level Comparison of Data Sets with Different Models

Method/DataSet	Best2010			LST20		
/Parameter	Precision(%)	Recall(%)	F1-Score(%)	Precision(%)	Recall(%)	F1-Score(%)
Longest	77.01	67.05	71.68	80.60	72.06	76.09
DeepCut	96.11	96.57	96.34	96.67	97.13	96.90
Proposed Work	96.25	96.72	96.42	96.87	97.34	97.12

Conclusions and Future Scope:
In this research, a load-balancing technique for processing tasks on cloud platforms is created using big data analytics and artificial intelligence. Two novel modeling approaches for DLM-based and DSA-based learning are presented in this research. Compared to methods that rely on literature survey, both strategies are more accurate. In certain instances, the first strategy even produces results that are superior to the most cutting-edge Deepcut. Even so, it still takes longer than other methods to

process the information. With an average improvement of 0.15% accuracy and 0.1% of F1 score, the application of DSA analysis to rebuild the data signal enhances the original data's quality and raises the recognition rate for several conventional classification methods. Additionally, this paper's methodology provides many benefits over related research. Many techniques have been developed to reduce the impact of small collections of comparable data; some of them try to predict missing data while others try to reduce the contribution of the relevant users. Even though the article only makes use of DSAs from DLM, it would be useful to look at how well other DLM learning and segmentation prediction techniques, like the spatial and temporal memory, work. We think that new approaches should be presented in the future that exclusively employ information stored by the cloud-user to provide security in the missing values of data.

References:

1. Zhu, Haizhen, et al. "A structure optimization method for an extended belief-rule-based classification system." Knowledge-Based Systems 203 (2020): 106096.
2. D. Davis and C. -C. Chiang, "Natural Language Processing for Detecting Undefined Values in Specifications," 2022 17th Annual System of Systems Engineering Conference (SOSE), Rochester, NY, USA, 2022, pp. 191-196, doi: 10.1109/SOSE55472.2022.9812647.
3. B. Wen, M. Xiao and X. Chen, "Extended Belief Rule Base Parameter and Structure Optimization Based on Similarity Measure," 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2021, pp. 1009-1012, doi: 10.1109/ICAICA52286.2021.9498019.
4. P. Danenas and T. Skersys, "Exploring Natural Language Processing in Model-To-Model Transformations," in IEEE Access, vol. 10, pp. 116942-116958, 2022, doi: 10.1109/ACCESS.2022.3219455.
5. R. Chandra, A. Shukla, S. Tiwari, S. Agarwal, M. Svafrullah and K. Adiyarta, "Natural language Processing and Ontology based Decision Support System for Diabetic Patients," 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Jakarta, Indonesia, 2022, pp. 13-18, doi: 10.23919/EECSI56542.2022.9946601.
6. J. Zhou, X. Xing, D. Lin and X. Zhang, "Semantic Similarity Evaluation Method Based on Text Generation Data Augmentation," 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT), Changzhou, China, 2022, pp. 1-7, doi: 10.1109/ACAIT56212.2022.10137987.
7. Zhu, Haizhen, et al. "A structure optimization method for extended belief-rule-based classification system." Knowledge-Based Systems 203 (2020): 106096.
8. Debauche, O.; Mahmoudi, S.; Mahmoudi, S.A.; Manneback, P.; Lebeau, F. A new edge architecture for ai-iot services deployment. Procedia Comput. Sci. 2020, 175, 10–19.
9. Chang, Z.; Liu, S.; Xiong, X.; Cai, Z.; Tu, G. A Survey of Recent Advances in Edge-Computing-Powered Artificial Intelligence of Things. IEEE Internet Things J. 2021, 8, 13849–13875.
10. Rajavel, R.; Ravichandran, S.K.; Harimoorthy, K.; Nagappan, P.; Gobichettipalayam, K.R. IoT-based smart healthcare video surveillance system using edge computing. J. Ambient. Intell. Humaniz. Comput. 2021, 1–3.
11. Sha, K.; Yang, T.A.; Wei, W.; Davari, S. A survey of edge computing-based designs for iot security. Digit. Commun. Netw. 2020, 6, 195–202.
12. Amutha, J.; Sharma, S.; Sharma, S.K. Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review, taxonomy, research findings, challenges and future directions. Comput. Sci. Rev. 2021, 40, 100376.
13. Hu, K.; Li, Y.; Xia, M.; Wu, J.; Lu, M.; Zhang, S.; Weng, L. Federated Learning: A Distributed Shared

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Machine Learning Method. Complexity 2021, 8261663.
14. Zhu, Haizhen, et al. "A minimum centre distance rule activation method for extended belief rule-based classification systems." Applied Soft Computing 91 (2020): 106214.
15. Lin, F.P.C.; Brinton, C.G.; Michelusi, N. Federated Learning with Communication Delay in Edge Networks. arXiv 2020, arXiv:2008.09323.
16. Wang, X.; Wang, C.; Li, X.; Leung, V.C.M.; Taleb, T. Federated deep reinforcement learning for internet of things with decentralized cooperative edge caching. IEEE Internet Things J. 2020, 7, 9441–9455.
17. Zhou, Z.; Yang, S.; Pu, L.; Yu, S. CEFL: Online admission control, data scheduling, and accuracy tuning for cost-efficient federated learning across edge nodes. IEEE Internet Things J. 2020, 7, 9341–9356.