Data Management Method in the Petroleum Exploration Work Area

Ying Peng[1], Yicheng Gu[2], Pengtao Yang[1], Linfu Zhan[1]*

[1] Institute of Geophysical Prospecting, Sinopec Shengli Oilfield Company, China
[2] School of Software, Shanghai Jiao Tong University, China
pengy.slyt@sinopec.com, guyicheng98@sjtu.edu.cn, yangpengtao.slyt@sinopec.com, zhanlinfu.slyt@sinopec.com

Abstract

Petroleum exploration is an industry that generates a large amount of data, but the datasets used are highly correlated and complex to process. To achieve intelligent management of petroleum data, we propose a multi-model framework based on deep learning networks. This framework combines the advantages of Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) to identify hot data that are more likely to be accessed by voting. In addition, we compare the performance of three commonly used time series prediction models for spatial prediction of petroleum exploration work areas. Experiments show that the multi-model framework outperforms traditional solutions by 25.3% and exhibits a 7.0% performance improvement compared to the best-performing LSTM model in a single model. LSTM is more suitable than Least Squares Regression (LSR) and Support Vector Regression (SVR) for spatial prediction of petroleum data, and a simple offset processing of the prediction results can cover more than 90% of real scenarios.

Keywords: Petroleum exploration data, Intelligent data management, Machine learning

1 Introduction

Petroleum exploration is a data-intensive industry [1], accumulating massive oilfield data resources in various stages such as geology, logging, geophysical exploration, and development, including well data, logging data, seismic data, layer and fault data, etc. [2]. Through mainstream exploration analysis software such as GeoFrame, these data can form seismic profiles that provide data support for finding suitable petroleum exploration areas. Data collected in each petroleum exploration area, which is useful for locating oil-bearing areas, is stored in a petroleum exploration work area. The data are highly correlated and complex to process, all of which can have an impact on the final results during data analysis.

To maintain a balance between storage costs and application performance, it's crucial to implement intelligent data management. Employing a single storage medium for all exploration data leads to either substantial storage expenses or compromised application performance. Thus, predictive technologies like pre-fetching hot data and forecasting data storage space growth have become imperative. By prioritizing hot data and storing it on high-speed media, data utilization efficiency can be enhanced. Additionally, accurately predicting data volume growth enables proactive storage medium management, thereby optimizing storage space utilization.

To achieve the effects of these two technologies, this paper proposes a multi-model framework leveraging Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) to enhance pre-fetch selection analysis in petroleum exploration. By integrating the strengths of these deep learning models through a voting mechanism, the accuracy of prefetch results is improved. Moreover, this paper conducts a performance comparison among three commonly used time-series prediction models—Least Squares Regression (LSR), Support Vector Regression (SVR), and LSTM—for predicting storage space in petroleum exploration data. Through this comparison, a machine learning prediction method more suitable for managing petroleum exploration data is identified.

The rest of this paper is organized as follows. In Section 2, related work is discussed. Section 3 explains methodologies for prediction. In Section 4, we use a series of benchmarks to evaluate the performance of these methodologies. A summary is given in Section 5.

2 Related Works

2.1 Prefetching Data

Prefetching is a critical optimization technique in computer systems that aims to enhance system performance by accessing the required content quickly through prefetching-related content [3]. It plays a crucial role in improving the efficiency of computer systems. The accuracy of pre-fetching algorithms can be improved by using content and historical retrieval information.

Xu and his team adaptively prefetch a set of consecutive data stream fragments to the cache using a multi-armed bandit model, reducing memory overhead and achieving effective duplicate data deletion [4]. DeepUM combines new page prefetching strategies with relevant prefetching techniques to enable deep neural networks with super memory usage [5].

Deep learning can predict data access patterns and fea-

*Corresponding Author: Ying Peng; Email: pengy.slyt@sinopec.com
DOI: https://doi.org/10.70003/160792642026012701009

tures to improve pre-fetching effectiveness. Buyuktanir and Aktas applied LSTM and bidirectional LSTM to high-data-intensive network applications, modeling customer browsing data and predicting subsequent user behavior, minimizing data access latency and improving operational performance and data delivery speed [6]. Ganfure and his team proposed the DeepPrefetcher model, which uses distributed representation learning to learn block access pattern contexts and utilizes LSTM for context awareness to achieve data prefetching [7]. This method is also very effective in improving prefetching efficiency. Using multiple deep-learning models together for classification can also improve the prefetching accuracy of the database [8].

In storage systems, the separation of frequently accessed data and infrequently accessed data into hot and cold data achieves a unified storage cost and system performance. The Least Recently Used (LRU) strategy is a classic method for identifying cold and hot data by replacing data with the least recently accessed data [9, 12]. To improve the accuracy of cold and hot data identification, various methods have been adopted domestically and abroad. At the data structure level, Bloom filters or Cuckoo filters have been proposed to enhance the accuracy of cold and hot identification [9]. The introduction of the Adaptive Robust Control (ARC) algorithm [10], Dynamic Data Clustering (DAC) algorithm [11], and temperature cooling model have continuously improved the accuracy of cold and hot identification. Additionally, machine learning methods such as those proposed in [10, 12], and [13] have proven to be effective in identifying cold and hot data.

2.2 Time-series Prediction

Time series prediction methods are widely used in various fields by forecasting future trends of the same event based on changes in it [14], such as in predicting COVID-19 cases [15-16], population forecasting [17], atmospheric environment prediction, etc. Traditional time series prediction models such as Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) [18] are effective for data problems with stationary time series characteristics. However, in reality, most events do not change at the same rate, and the prediction accuracy of such nonlinear time series is difficult to meet the needs of applications.

Therefore, nonlinear prediction methods have received more attention both at home and abroad. Pannakkong and Huynh combined ARIMA with discrete wavelet transform (DWT) and artificial neural network (ANN) to achieve better prediction results than a single model on three classical datasets [19]. Support Vector Machine (SVM) is a classical machine learning model [20], on which Luo, Jiang, and Zheng proposed a reconfigured training set SVM (RTS-SVM) to solve the classification problem in high noise scenarios [21]. Recurrent Neural Network (RNN) [22-23] is also a neural network model that uses historical data to efficiently predict sequential data. To address the problem of RNN gradient disappearance and explosion, LSTM was proposed to improve RNN [24-26]. Combining LSTM with CNN can effectively reduce the prediction error [27-28].

However, for scenarios in petroleum exploration, where the data has high correlation and complexity, there has not been sufficient research targeting this field. Prefetching and spatial prediction algorithms have not been tested for their effectiveness in petroleum exploration contexts.

2.3 Machine Learning in Petroleum Exploration

Machine learning is emerging as a highly effective tool in the realm of petroleum exploration, offering advancements across its upstream, midstream, and downstream sectors [29]. Numerous studies have highlighted machine learning's prowess in processing and analyzing vast amounts of data, tailoring its applications to the unique characteristics of petroleum exploration scenarios [30-31]. Al-Mudhafar et al. [32] conduct a comparative analysis of five machine learning algorithms aimed at classifying carbonate lithologies, with the goal of enhancing reservoir discrimination for improved fluid flow and storage capacity assessment. Sheykhinasab et al. [33] integrate the Least-squares support-vector machines (LSSVM) and multilayer extreme learning machine (MELM) with optimization algorithms such as cuckoo optimization algorithm (COA), particle swarm optimization (PSO), and genetic algorithm (GA) to predict permeability—a crucial factor in enhancing oilfield development efficiency. Despite these advancements, much of the current research primarily focuses on machine learning applications in oilfield exploration, with fewer studies addressing the storage of petroleum exploration data characterized by small datasets.

3 Methodology

In this section, the paper will introduce a traditional method suitable for petroleum exploration scenarios and a new machine-learning method for distinguishing between cold and hot data. Additionally, a brief introduction to spatial prediction models will also be presented.

3.1 Traditional Prediction Model

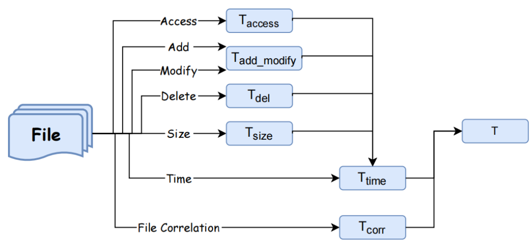

Figure 1. Procedure for calculating file temperature

The CT-LRU model is a traditional method for identifying hot data by predicting future file access based on information from various dimensions of historical files. The model calculates the dynamic hotness value of a file by weighting its information, such as access operations, modification operations, deletion operations, addition operations, and file size. The dynamic hotness value is then dynamically adjusted using the temperature cooling model, which takes the operation time as input. Finally, the static hotness value is obtained based on the file association. The total hotness value is the sum of the dynamic and static hotness values, and the higher the total hotness value, the higher the likelihood of accessing the file. The calculation process of file hotness is shown in Figure 1.

The formula for calculating the hotness value is shown below:

$$T_t = T_{t_{n-1}} e^{-k(t_n - t_{n-1})} + T_{acc} + T_{add\text{-}mod} + T_{del} + T_{size} \tag{1}$$

$$\begin{cases} T_{acc} = \alpha_{acc} f_{acc} \\ T_{add\text{-}mod} = a_{add\text{-}mod} f_{add\text{-}mod} \\ T_{del} = \alpha_{del} f_{del} \\ T_{size} = \alpha_{size} f_{size} \end{cases} \tag{2}$$

where t_n represents the most recent operation time, T_{t_n} is the temperature obtained from the temperature cooling model, $e^{-k(t_n - t_{n-1})}$ represents the heat decay caused by the time difference between the current operation t_n and the last operation time t_{n-1}. T_{acc}, $T_{add-mod}$, T_{del}, T_{size} respectively represent the heat values calculated from the access frequency, new creation/modification frequency, deletion frequency, and file size, which are obtained by optimizing the parameters f (the frequency or size of the actual application of dimension information) and α (the sensitivity of the corresponding dimension information).

$$T_{corr} = \alpha_{corr} C \tag{3}$$

where T_{corr} is the static hotness value generated based on the file association, C is the number of associated files, and α_{corr} is the sensitivity of file association.

$$T = T_{t_n} + T_{corr} \tag{4}$$

In the CT-LRU model, the final heat value T determines where the data will be placed according to its magnitude. Figure 2 illustrates the operational workflow of this strategy. Initially, when data is accessed, its current heat type is determined. New data receives a baseline heat value, whereas existing data undergoes an update in its heat value based on the temperature cooling model. After calculating the required heat value, if the data has been defined as hot or warm data, its heat value is updated. Conversely, If the data is not hot or warm data, it is assigned to the warm data category, with new heat and file relevance values computed. When storage for both warm and hot data reaches capacity, the model evaluates the oldest item in the hot data set (based on the least recent access) against the item with the highest heat value in the warm data. The data with a higher heat value is retained in the hot data, while the less critical data is relocated to the cold data. This method ensures that data with higher importance and utility are stored on media that allow faster access, optimizing retrieval times and storage efficiency.

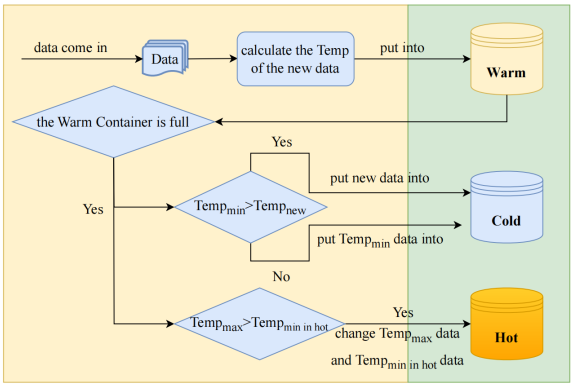

Figure 2. CT-LRU strategy

3.2 Multi-Model Prediction Framework

The model framework of this paper is shown in Figure 3. To forecast future data file accesses within the petroleum exploration work area, this framework integrates three well-established deep learning models: MLP, CNN, and LSTM. These models have been widely used in many fields and have been proven to perform well in handling different diverse data types.

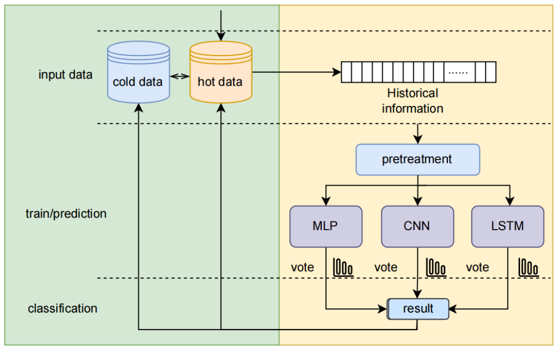

Figure 3. Multi model framework

MLP is a feedforward neural network model and one of the earliest widely used neural network models. It consists of multiple fully connected layers, each of which contains many neuron nodes. Each node processes the input of the previous layer through an activation function and passes it to the next layer. Due to the non-linear nature of MLP, it has powerful non-linear modeling capabilities and can model and classify highly non-linear data, and is suitable for many types of data. In this paper, a basic MLP model is implemented by establishing fully connected layers, and each model is trained using the Tanh activation function, mean squared error loss function, and Adam optimizer. The hyperparameters of the model are adjusted using the validation set. Compared with CNN and LSTM, MLP has a simpler structure, making it easier to implement and adjust.

CNN is a feedforward neural network model consisting of convolutional layers, pooling layers, and fully connected layers. The convolutional layer extracts the features of the input data through convolution and passes them to the next layer. The pooling layer reduces the size of the feature map, thus reducing the computation and parameter of the subsequent layer [34]. The dropout layer randomly drops the outputs of some neurons to reduce the overfitting of the model. The CNN model used in this paper is shown in Figure 4, and it extracts data features and reduces overfitting by using multiple convolutional layers, pooling layers, and dropout layers. Finally, the data is adjusted to the required output dimension through fully connected layers.

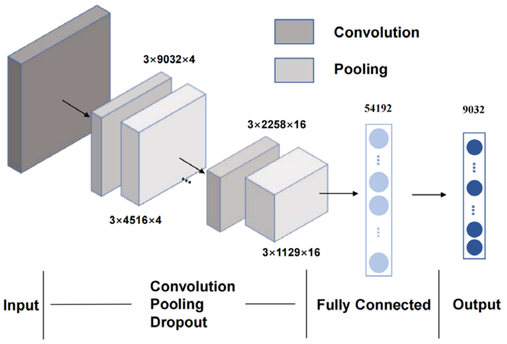

Figure 4. The CNN model

LSTM is a recurrent neural network model composed of many repeated LSTM modules. The LSTM module contains a forget gate, an input gate, and an output gate, which can control the flow of information in the LSTM module, allowing LSTM to remember and forget information in the input sequence. Unlike other recurrent neural networks, LSTM can avoid the problem of gradient vanishing or explosion in long sequence training, making it suitable for modeling sequence data. In this paper, the LSTM module is used for training, by setting the tanh activation function, mean squared error loss function, and Adam optimizer, adjusting the dropout rate, etc., to obtain more accurate predictions. LSTM is best at handling sequence data and has the greatest advantage in this scenario.

Each individual model is capable of extracting valuable information from the data to predict future access patterns of documents. To fully leverage the strengths of these three models, this paper employs a multi-model framework to integrate and weigh the results obtained from each model, ultimately deriving the final prediction outcomes. Different weights are assigned to documents predicted by each model based on their respective prediction results. The value assigned to each file is the aggregate of the weighted values from all three models. By comparing the values assigned to different files, those with higher values are deemed more likely to be accessed. This approach allows for a comprehensive and nuanced prediction of document access patterns by amalgamating insights from diverse models.

$$V_{file} = W_{mlp} * Pre_{mlp} + W_{cnn} * Pre_{cnn} + W_{lstm} * Pre_{lstm} \quad (5)$$

Where V is the value of files, W is the weight of each model, and Pre is the predicted result of each model for each file.

Specifically, this paper initially evaluates the accuracy of three models—MLP, CNN, and LSTM—when utilized individually for prediction. It was found that LSTM outperforms CNN, and CNN in turn outperforms MLP. However, using a single model in isolation does not yield optimal outcomes, as the potential of CNN and MLP is not completely harnessed. Therefore, the paper explores the accuracy improvement through model combination. When a set of data is predicted by two or more models simultaneously, this paper refers to it as being predicted by a combination of the models.

After integrating the models, there was a notable increase in the accuracy of the predictions, albeit with a corresponding reduction in the amount of data available. To compensate, the data were weighted and ranked. Each model is assigned a weight based on its accuracy; models with higher accuracy receive higher weights. The weight of a model is applied to its predictions by multiplying the model's weight by its confidence in each prediction. If data are predicted by multiple models, the total weight is the sum of the weights of these models. This approach, by utilizing the collective judgment of multiple models, reduces the risk of errors that might arise from relying on a single model and selects data based on the highest aggregate weights for predictions.

Based on the historical file access patterns, this paper establishes the relationship between the set of files in each cache and the set of files in the future cache, as shown in Figure 5. This approach predicts the next files to be accessed based on the order of past file operations, ensuring predictions are timely and relevant while minimizing the disruption caused by outdated data.

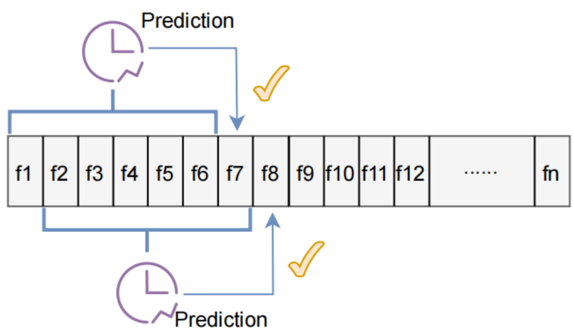

Figure 5. Relationship between data for prediction

3.3 Spatial Prediction Model

The spatial prediction model anticipates future space utilization based on past space usage at various points in time. In the scenario of petroleum exploration, the storage space in the work area remains constant during regular intervals and only expands when the area is actively utilized. However, the uncertainty of its variation makes predicting workspace hard. To predict the storage space occupied by each petroleum exploration workspace, this research focuses on implementing several commonly used time series prediction models, including LSR, SVR, and LSTM networks, for predicting petroleum exploration data storage space and performing comparative analysis to find the most suitable prediction method for petroleum exploration data management.

4 Performance and Analysis

The experimental platform used in this research is a PC equipped with an Intel(R) Core(TM) i7-10510U CPU processor with a clock speed of 2.30 GHz, 8 cores, and 16 GB of memory. The data used are records of operations, including accesses, additions, modifications, and deletions, for a specific work area in an actual petroleum exploration scenario in the Shengli Oilfield. There are 9032 files in this work area with a total size of 78.1GB, and 44016 operations were performed. After pre-processing the operation records, the files were divided into hot data and cold data, and a time series prediction of the workspace size was performed.

Figure 6 shows the distribution of the number of operations for each file. It can be observed that in the past six months, over 97% of the files in the area were operated on 11 times or less. Additionally, a small peak in the number of file accesses is observed at 9 and 10 times. From this, we can observe that most of the data files in this work area were accessed only a few times. The small peaks in access frequency are highly likely to indicate that these data were accessed together.

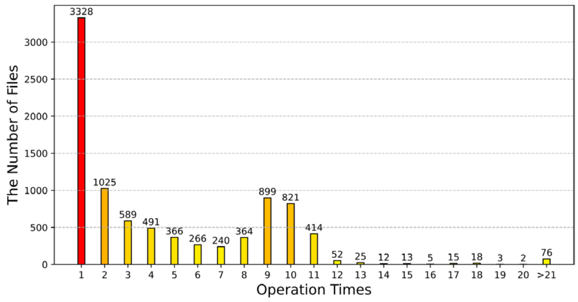

Figure 6. Distribution of the number of file operations

4.1 Hot Data Prefetching Strategy

Based on the application scenario of integrated exploration research, the experiment extracts the operation records of a petroleum exploration area within 6 months and records the various dimensional information of the files used in the area. Based on the file operations and various dimensional information, the performance of two traditional models (LRU, CT-LRU), MLP, CNN, LSTM, and multi-model framework (MMF) is compared.

4.1.1 Pre-processing

Based on the operation time of files in the workspace, a relationship is established between the history of the operated files and the next operated file. A new vector is formed by combining files with cache size quantities based on file names. By processing this vector through neural network layers, the final probability of accessing each file is obtained.

The traditional CT-LRU model optimizes its parameters through exhaustive grid search. As shown in Figure 7, in the case of petroleum exploration, when only one parameter is changed while the others remain fixed, each parameter change affects the final cache hit rate, with parameters α_{size} and α_{acc} having a larger effect and parameter α_{del} having a smaller effect. For the parameter α_{size} , the file size tends to zero for larger magnitudes. The optimal parameter configuration scheme can be found after training.

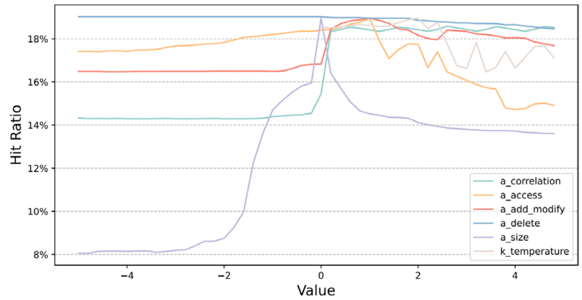

Figure 7. Relationship between coefficient and the hit ratio

4.1.2 Cache Hit Ratio

In this experiment, the hit rate of six strategies, LRU, CT-LRU, MLP, CNN, LSTM, and MMF, was compared under different cache size ratios. The variation of the hit rate is shown in Figure 8.

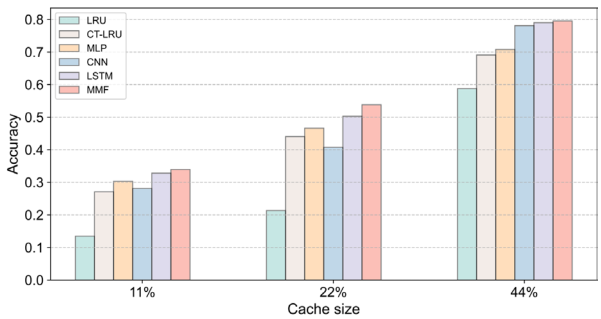

Figure 8. The Variation of hit ratio by the cache

According to Figure 8, it is evident that the cache hit rate remains low when the cache size is small. This phenomenon can be attributed to the low repetition rate and extended usage periods of the dataset within the petroleum exploration scenario. The traditional CT-LRU strategy performs well because it considers the information of various dimensions of each file. However, deep learning algorithms can mine more content and relationships of historical data, resulting in better performance in most cases. LSTM performs best in the single-model framework with up to 21.2% improvement compared to CT-LRU. The multi-model framework obtained by three-model voting performs the best, with up to 25.3% improvement compared to CT-LRU and up to 7.0% improvement compared to LSTM. This demonstrates the applicability of the multi-model framework in this case.

4.2 Workspace Prediction

In this paper, we predict the storage space required for a given petroleum exploration work area based on its storage space variation. Due to the uncertainty of work area usage and the lack of a clear model, three methods, namely LSR, SVR, and LSTM, are chosen in this paper to analyze and predict the storage space of a work area in the context of oil exploration.

The storage space in the work area is at a scale of 10^{10}, which may lead to significant prediction errors. Therefore, the Max-Min normalization method is used to normalize the size of the storage space before further prediction.

LSR, SVR, and LSTM all predict the next storage space size based on the last three storage space sizes. The prediction results are shown in Figure 9, and the predicted R^2 values are shown in Table 1.

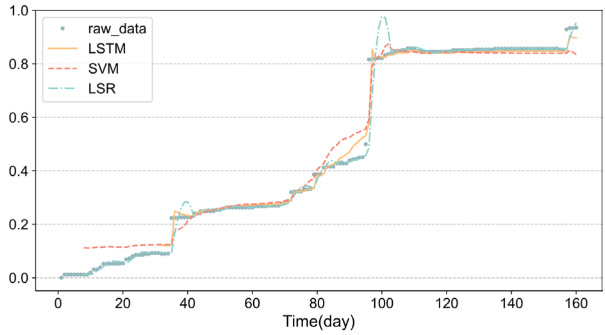

Figure 9. Results of spatial prediction of work area

Table 1. Results of space size prediction

Model	R^2
LSR	0.988
SVR	0.965
LSTM	0.992

As shown in Figure 9 and Figure 10, LSTM achieves better predictive performance than LSR and SVR. Without any processing, using LSTM to allocate storage space in advance can cover 78.3% of the cases. In this paper, by performing a simple offset processing on the LSTM prediction results and allocating an additional 3MB of space, the correct prediction space coverage rate can be increased by 17.8% to over 96% in O(1) time.

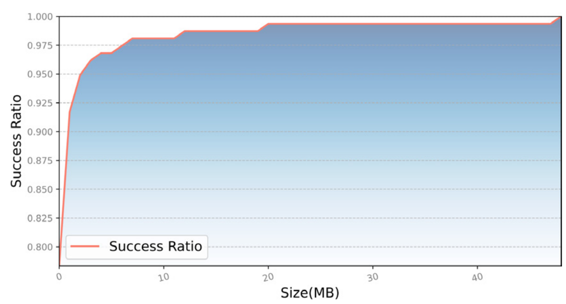

Figure 10. Variation of prediction success rate with offset space size

As shown in Figure 11, we compare the 281 work areas where the size of the work area will change, and we can intuitively see from the depth of color comparison that the LSTM algorithm generally outperforms the LSR and SVR algorithms on each work area.

Therefore, the LSTM model can be better applied to the spatial prediction of intelligent data management in the petroleum exploration scenario with simple subsequent processing.

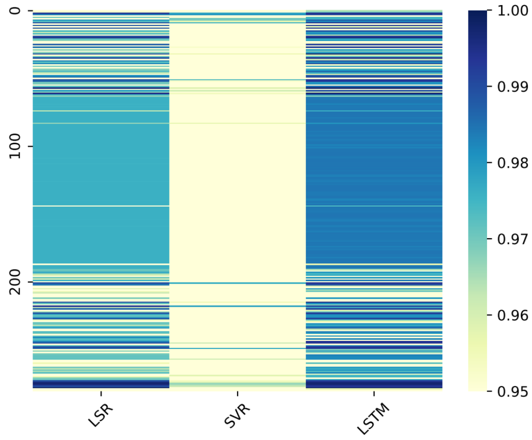

Figure 11. Heatmap of models for all work areas

5 Conclusion

To achieve intelligent management of oil exploration data, this study proposes a multi-model framework based on MLP, CNN, and LSTM for analyzing prefetch selection in petroleum exploration scenarios. The multi-model framework can take advantage of various single models to improve the success rate of prediction. The experimental results indicate that the multi-model framework obtained by voting on the three models performs the best, with a maximum improvement of 25.3% compared to the traditional CT-LRU model and up to 7.0% compared to the LSTM. These results demonstrate the applicability of the multi-model framework in the current scenario. A comparison of LSR, SVR, and LSTM spatial prediction models in the scenario of petroleum exploration, where LSTM performs the best. With a simple constant time modification, it can predict 90% of the results. Future work will explore improvements to the machine learning model in the petroleum exploration scenario to improve the accuracy and robustness of spatial prediction.

References

[1] J. N. Desai, S. Pandian, R. K. Vij, Big data analytics in upstream oil and gas industries for sustainable exploration and development: A review, *Environmental Technology & Innovation*, Vol. 21, Article No. 101186, February, 2021.

[2] Z. Tariq, M. S. Aljawad, A. Hasan, M. Murtaza, E. Mohammed, A. El-Husseiny, S. A. Alarifi, M. Mahmoud, A. Abdulraheem, A systematic review of data science and machine learning applications to the oil and gas industry, *Journal of Petroleum Exploration and Production Technology*, Vol. 11, No. 12, pp. 4339–4374, December, 2021.

[3] T. J. Chen, J. P. Sheu, Y. C. Kuo, Prefetching and caching schemes for IoT data in hierarchical edge computing architecture, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 33, No. 2, pp. 109–121, February, 2020.

[4] G. Xu, B. Tang, H. Lu, Q. Yu, C. W. Sung, Lipa: A learning-based indexing and prefetching approach for data deduplication, *2019 35th Symposium on mass storage systems and technologies (MSST)*, Santa Clara, CA, USA,

2019, pp. 299–310.

[5] J. Jung, J. Kim, J. Lee, Deepum: Tensor migration and prefetching in unified memory, *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, Vancouver, BC, Canada, 2023, pp. 207–221.

[6] T. Buyuktanir, M. S. Aktas, A deep learning-based prefetching approach to enable scalability for data-intensive applications, *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 2716–2721.

[7] G. O. Ganfure, C. F. Wu, Y. H. Chang, W. K. Shih, Deepprefetcher: A deep learning framework for data prefetching in flash storage devices, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 39, No. 11, pp. 3311–3322, November, 2020.

[8] Y. Chen, Y. Zhang, J. Wu, J. Wang, C. Xing, Revisiting data prefetching for database systems with machine learning techniques, *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, Chania, Greece, 2021, pp. 2165–2170.

[9] Y. Wang, Y. Yang, X. Qiu, Y. Ke, Q. Wang, CCF-LRU: hybrid storage cache replacement strategy based on counting cuckoo filter hot-probe method, *Applied Intelligence*, Vol. 52, No. 5, pp. 5144–5158, March, 2022.

[10] G. Vietri, L. V. Rodriguez, W. A. Martinez, S. Lyons, J. Liu, R. Rangaswami, M. Zhao, G. Narasimhan, Driving cache replacement with ML-based LeCaR, *in HotStorage'18: Proceedings of the 10th USENIX Conference on Hot Topics in Storage and File Systems*, Boston, MA, USA, 2018, pp. 928–936.

[11] N. V. Hiep, J. W. Hsieh, Timestamp-based hot/cold data identification scheme for solid state drives, *Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems*, Honolulu, Hawaiipp, 2018, pp. 255–259.

[12] Y. Wang, Y. Yang, C. Han, L. Ye, Y. Ke, Q. Wang, LR-LRU: A PACS-oriented intelligent cache replacement policy, *IEEE Access*, Vol. 7, pp. 58073–58084, April, 2019.

[13] T. Miura, K. Kourai, S. Yamaguchi, Cache replacement based on lstm in the second cache in virtualized environment, *2020 Eighth International Symposium on Computing and Networking Workshops (CANDARW)*, Naha, Japan, 2020, pp. 421–424.

[14] Z. Han, J. Zhao, H. Leung, K. F. Ma, W. Wang, A review of deep learning models for time series prediction, *IEEE Sensors Journal*, Vol. 21, No. 6, pp. 7833–7848, March, 2021.

[15] V. K. Kaliappan, R. Thangaraj, P. Pandiyan, K. Mohanasundaram, S. Anandamurugan, D. Min, Real-time face mask position recognition system using YOLO models for preventing COVID-19 disease spread in public places, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 42, No. 2, pp. 73–82, January, 2023.

[16] M. C. Younis, Evaluation of deep learning approaches for identification of different corona-virus species and time series prediction, *Computerized Medical Imaging and Graphics*, Vol. 90, Article No. 101921, June, 2021.

[17] K. H. Poon, P. K.-Y. Wong, J. C. Cheng, Long-time gap crowd prediction using time series deep learning models with two-dimensional single attribute inputs, *Advanced Engineering Informatics*, Vol. 51, Article No. 101482, January, 2022.

[18] I. Paliari, A. Karanikola, S. Kotsiantis, A comparison of the optimized lstm, xgboost and arima in time series forecasting, *2021 12th International Conference on Information, Intelligence, Systems &Applications (IISA)*, Chania Crete, Greece, 2021, pp. 1–7.

[19] W. Pannakkong, V.-N. Huynh, A hybrid model of arima and ann with discrete wavelet transform for time series forecasting, *Modeling Decisions for Artificial Intelligence: 14th International Conference, MDAI 2017*, Kitakyushu, Japan, 2017, pp. 159–169.

[20] F. Liu, Z. Yu, Y. Wang, H. Feng, Z. Zha, R. Liao, Y. Zhang, Falsified data filtering method for smart grid wireless communication based on SVM, *International Journal of Internet Protocol Technology*, Vol. 13, No. 4, pp. 177–183, October, 2020.

[21] C. Luo, Z. P. Jiang, Y. J. Zheng, A novel reconstructed training-set svm with roulette cooperative coevolution for financial time series classification, *Expert Systems with Applications*, Vol. 123, pp. 283–298, June, 2019.

[22] J. Wang, Y. Zou, P. Lei, R. S. Sherratt, L. Wang, Research on Recurrent Neural Network Based Crack Opening Prediction of Concrete Dam, *Journal of Internet Technology*, Vol. 21, No. 4 , pp. 1161–1169, July, 2020.

[23] D. Biswas, J.-M. Gil, Design and Implementation for Research Paper Classification Based on CNN and RNN Models, *Journal of Internet Technology*, Vol. 25, No. 4, pp. 637–645, July, 2024.

[24] S. Kumar, A. Damaraju, A. Kumar, S. Kumari, C.-M. Chen, LSTM Network for Transportation Mode Detection, *Journal of Internet Technology*, Vol. 22, No. 4 , pp. 891–902, July, 2021.

[25] S. Lei, L. Wang, W. Cui, M. Woźniak, D. Polap, Research on a LSTM Based Method of Forecasting Primary Frequency Modulation of Grid, *Journal of Internet Technology*, Vol. 21, No. 3 , pp. 791–798, May, 2020.

[26] C.-Y. Tsai, M.-Y. Su, C. Chuang, C.-Y. Chang, D.-S. Roy, TEPM: traveller enrolment prediction mechanism using BERT-based feature clustering and LSTM models, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 46, No. 1, pp. 14–26, May, 2024.

[27] Y. Xiao, H. Yin, K. Xia, Y. Zhang, H. Qi, Utilization of CNN-LSTM model in prediction of multivariate time series for UCG, *Machine Learning for Cyber Security: Third International Conference, ML4CS 2020*, Guangzhou, China, 2020, pp. 429–44.

[28] S. N. Hussain, A. A. Aziz, M. J. Hossen, N. A. A. Aziz, G. R. Murthy, F. B. Mustakim, A novel framework based on cnn-lstm neural network for prediction of missing values in electricity consumption time-series datasets, *Journal of Information Pr*ocessing Systems, Vol. 18, No. 1, pp. 115–129, February, 2022.

[29] A. Lawal, Y. Yang, H. He, N. L. Baisa, Machine Learning in Oil and Gas Exploration-A Review, *IEEE Access*, Vol. 12, pp. 19035–19058, February, 2024.

[30] R. K. Pandey, A. K. Dahiya, A. Mandal, Identifying applications of machine learning and data analytics based approaches for optimization of upstream petroleum operations, *Energy Technology*, Vol. 9, No. 1, Article No. 2000749, January, 2021.

[31] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, H. Oza, Application of machine learning and artificial intelligence in oil and gas industry, *Petroleum Research*, Vol. 6, No. 4, pp. 379–391, December, 2021.

[32] W. J. Al-Mudhafa, M. A. Abbas, D. A. Wood, Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate

reservoirs, *Marine and Petroleum Geology*, Vol. 145, Article No. 105886, November, 2022.

[33] A. Sheykhinasab, A. A. Mohseni, A. B. Bahari, E. Naruei, S. Davoodi, A. Aghaz, M. Mehrad, Prediction of permeability of highly heterogeneous hydrocarbon reservoir from conventional petrophysical logs using optimized data-driven algorithms, *Journal of Petroleum Exploration and Production Technology*, Vol. 13, No. 2, pp. 661–689, February, 2023.

[34] T. Neela, S. Namburu, ECG signal classification using capsule neural networks, *IET Networks*, Vol. 10, No. 3, pp. 103–109, May, 2021.

Biographies

Ying Peng (1970-), male. PhD, Senior Engineer, his research interests include information security, information system development.

Yicheng Gu (1998-), male. He is currently a master candidate in the Shanghai Jiao Tong University. His research interests include virtualization technology.

Pengtao Yang (1976-), male, bachelor, Senior Engineer, his research interests include information system development.

Linfu Zhan (1973-), male, master, Senior Engineer, his research interests include information system development.