Research on Intelligence Logging Interpretation Technology and System Based on Standard Big Data Platform

Ting-ting Li$^{(\boxtimes)}$, Hong-shu Zhang, Dao-jie Cheng, Ke Huang, Wen-mao Yu, and Hao Chen

China Petroleum Logging CO. LTD, Xi'an 710077, Shaanxi, China
`569710061@qq.com`, {`zhanghongshucq,chengdj_cpl,huangke_cpl,`
`zycjyuwm`}`@cnpc.com.cn`, `chenhao@cnlc.cn`

Abstract. "Massive" logging data assets, due to their insufficient storage methods and normalization, cannot be quickly and accurately called up, become a "data island", so that their value has not been fully explored. The current application scope of artificial intelligence is focused on single method research, with few system applications. However, intelligent interpretation requires the use of a large amount of logging data and related standard data. Based on a large number of documents related to large logging database and logging artificial intelligence, starting with supervised, unsupervised and semi-supervised intelligent algorithms, this paper expounds the application status quo and applicability of intelligent logging interpretation technology through machine learning for conventional logging lithology identification, automatic layering, sedimentary microfacies identification and reservoir identification. This paper briefly introduces the application status quo of logging data governance and mining technology. This paper summarizes the process of intelligent interpretation method, as well as the intelligent logging interpretation method and system based on a physical model under a standard big data platform. This paper discusses the existing problems in intelligent logging interpretation and evaluation and the feasible development direction of future research.

Keywords: Artificial intelligence · Algorithm · Lithology identification · Intelligent logging interpretation

1 Introduction

Geophysical logging is a method to measure geophysical parameters by using geophysical characteristics such as electrochemical characteristics, electrical conductivity, acoustic characteristics and radioactivity of rock strata. During oil drilling, logging must be carried out after drilling to the designed well depth, so as to obtain various petroleum geology data and engineering technical data as the original data for completion and development of oilfields, which is called completion logging or open hole logging. All logging after casing running in oil wells or during production are generally called production logging. The development of logging has generally experienced four stages such as analog logging, digital logging, numerical control logging and imaging logging.

Logging data processing and interpretation needs interpretation experts with rich regional geological experience, but even in the same area, the interpretation results of different experts are different. With the continuous development of interpretation and evaluation software, various logging interpretation software with complete functions has emerged at home and abroad, which has improved the accuracy and level of logging interpretation, promoted the progress of interpretation technology and solved some difficult problems of logging interpretation and evaluation. With the rise of big data and artificial intelligence, faced with more and more complex reservoir interpretation problems, industry experts have also realized the urgency of developing intelligent interpretation and evaluation systems and began to explore and study in this field [1].

Major international oil companies and service companies are also adjusting their development strategies, making continuous efforts in the fields of data science and artificial intelligence, and developing their own intelligent interpretation and evaluation systems. For example, Schlumberger has built its own intelligent logging processing and interpretation platform. Major oil companies and IT companies have cooperated in the field of intelligent application, and formed joint strategic research teams such as Shell + Microsoft and ExxonMobil + Microsoft, thus making many beneficial explorations in the construction of big data platform for oil and gas exploration and development, the creation of ecological environment for data sharing, and the improvement of data processing and interpretation quality, showing the great development potential of big data and artificial intelligence technology in the oil and gas industry [2].

The information abundance of data interpretation determines the application depth and breadth of logging technology, where the important link depends on the development of interpretation methods and software. As the largest producer and user of logging data in China, CNLC faces the problems such as diverse logging data types, strong professionalism and complex data operation [3].

Since 2010, China National Logging Corporation has been committed to the construction of a unified logging database with reference to the architecture of the CNPC's dream cloud platform for exploration and development. The construction of logging data resources has gone through three stages such as data management, data sharing and data application. CNLC has completed the centralized storage and management of scattered data, data sharing for individual applications and data mining applications. This has provided the required data for all kinds of professional software, helped the interconnection between systems and services, and supported the convenient sharing of data and results. On the basis of predecessors' research results, CNLC has established a data lake based on standard big data platform, followed the road of integration of logging interpretation software, realized the organic combination of three key elements: data, algorithm and scenario, improved the efficiency of logging analysis, and promoted the transformation of logging interpretation from single well interpretation to multi-well evaluation and reservoir analysis [4–8].

2   Data Governance Based on Logging Big Data Platform

Geophysical logging, as the "eye" of deep formations, has the advantages such as many methods, high resolution and large amount of information, and can provide continuous and accurate in-situ physical parameters such as electrical, acoustic and nuclear

parameters for reservoir evaluation. After the professional reorganization of CNLC, the logging data of 16 oil and gas fields have many types and different standards, leading to extremely complex logging data; in addition, these data cannot be directly applied, forming an "island" of data, which urgently needs data governance. Focusing on three major tasks: automatic logging data sorting technology, automatic data flow technology of logging data: sorting-data governance-data warehousing and logging big data analysis technology, CNLC has carried out logging data governance work to transform unstructured data into structured data.

In order to reduce manual participation as much as possible and improve the degree of automation in the process of data governance, a multi-source heterogeneous data governance architecture based on semi-supervised learning algorithm is very suitable for logging data governance (see Fig. 2). Its basic idea is to integrate heterogeneous data describing the same entity in the real world from different data sources into structured data. The specific process includes four parts such as information extraction, pattern matching, data matching and data fusion (see Fig. 1).The actual results show that the architecture can not only effectively solve the "data island" state, but also significantly improve the data quality with as little manual participation as possible. After governance, the data call is more convenient so as to meet the requirements of different logging interpretation and geological application [14, 40–43].

| Information extraction | → | Pattern matching | → | Data matching | → | Data fusion |

Fig.1. Multi source heterogeneous data governance scheme.

CNLC had formulated logging data warehousing specifications and logging data management specifications respectively for new wells and old wells. They standardize the storage file range, file format and file naming method of original logging data and result data and fully consider the current situation of the original logging database, Based on LEAD software, CNLC has developed a tool for automatically sorting, naming and exporting data. With the goal of building the brand of logging companies CNLC has unified the drawing style and formulated complete sets of drawing templates, headers and results table examples for conventional combination logging, imaging logging, production logging and engineering logging, unified curve line type, name and section filling; and standardized the symbols of logging interpretation conclusion and lithology. Moreover, CNLC has unified the mapping specifications and standardized the naming and dimensions of logging original logging curves and result curves. CNLC has established the standard name and used name, and directly managed the old well data in the original logging database to the big data platform. The new well data generated in real time is directly uploaded to the big data platform through the integrated application system.

The scope of logging data governance is defined in five logging types, which mainly include well basic information, logging curve data, map type data and table type data. It mainly evaluates the "six properties" of the data after governance, namely accuracy, completeness, standardization, uniqueness, consistency and timeliness. In view of the great differences in data storage formats of historical logging data, the non-uniform data
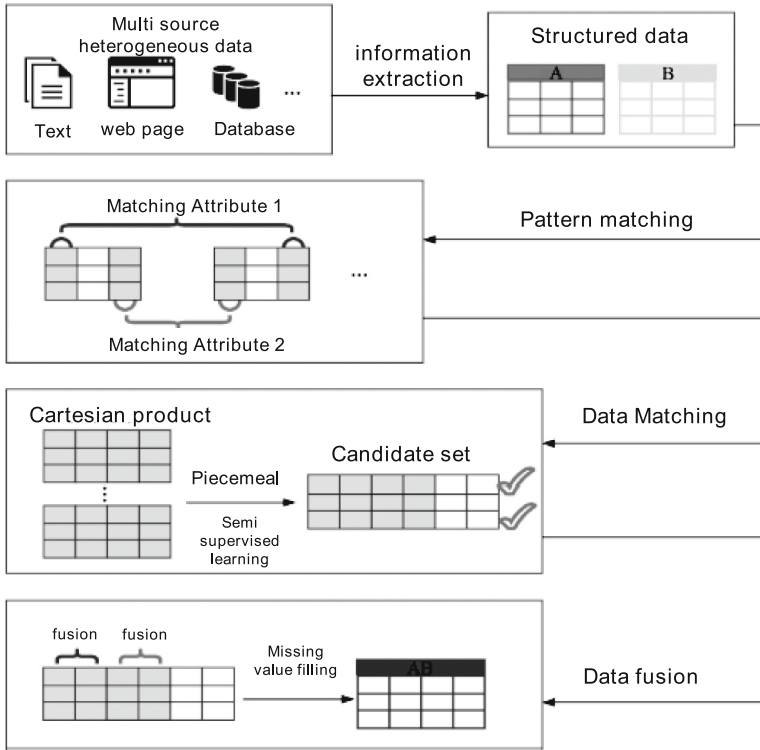
Fig.2. Multi-source heterogeneous data governance architecture based on semi-supervised learning (according to Wei-xiong Rao, et al.).
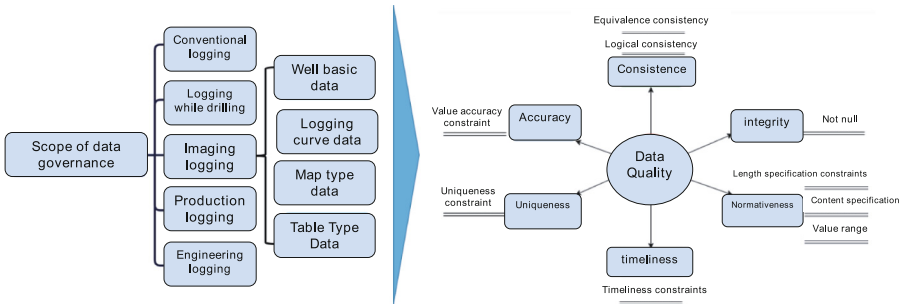

Fig.3. Scope of logging data governance and evaluation of "six properties" after logging data governance.

standards, and large data governance workload, there are 16 data formats and various variants of logging data achievement files, and so on, relevant decompilation tools and warehousing tools are correspondingly matched. At present, the governance of logging

data of hundreds of thousands of wells has been completed, which has laid a data foundation for artificial intelligence based logging interpretation (see Fig. 3).

3 Present Situation and Applicability of Intelligent Logging Interpretation

The development of artificial intelligence has a long history, and it is a science based on computer technology. In logging, most machine learning models are shallow learning, such as linear classifier, BP neural network, logistic regression, K-Means clustering, support vector machine, principal component analysis, Gaussian mixture model, gradient thruster and so on. There is usually only one hidden layer in these shallow learning structures, that is, only one nonlinear feature extraction layer. Shallow learning can only be effective for some simple or limited problems in general, but it is obviously at a disadvantage in the face of complex and huge data [9].

The intelligent logging interpretation method is mainly a deep learning algorithm. Deep learning network is an extension of traditional artificial neural network. Because of its multiple hidden layers, the deep learning network can realize the mapping transformation from low-dimensional space to high-dimensional space through multi-layer nonlinear transformation, thus distinguishing complex input data features in high-dimensional space and realizing the identification and classification of complex input information. According to the characteristics of an algorithm learning task, it can be divided into supervised learning, unsupervised learning and semi-supervised learning. According to its function, supervised learning can be divided into regression and classification. Regression is to predict the occurrence probability of an object, and classification is to classify the pattern class attribution of an object. According to whether there are labels in the input data, it can be divided into supervised learning and unsupervised learning. Unsupervised algorithms, such as clustering algorithm and dimension reduction algorithm, such as fisser discriminant method, are effective in identifying complex oil-water layers. Logical regression, support vector machine, proximity regression and decision tree algorithms in supervised algorithms have high accuracy in identifying complex lithology. On the basis of computer algorithm, semi-supervised algorithm is added with human experience, so that the clustering effect is significantly improved by using a small amount of labeled data and a large amount of unlabeled data. The semi-supervised classification task combined with unsupervised learning dimension reduction method can improve the classification effect of supervised learning under the condition of insufficient labels. Swarm intelligent optimization algorithm is an algorithm that combines the behaviors of animals such as foraging and avoiding obstacles, including bat algorithm, ant colony algorithm and firefly algorithm, and is used to study the wave impedance inversion method [46–53].

Industry scholars have used vector machine, neural network, fuzzy recognition and traditional decision tree methods to identify lithology, and achieved good application results [10–12].Wang Hua et al. deeply analyzed the applicability of applying artificial intelligence in logging data processing and interpretation from the traditional data modeling method and machine learning algorithm in geophysical logging field. Chen Xi et al. expounded that artificial intelligence based logging interpretation is feasible from three

cores such as data model, physical simulation algorithm and artificial intelligence based logging ecology, which can help logging analysts solve deeper geological problems [1, 13].

4   Artificial Intelligence Based Interpretation of Logging Data

4.1   Lithology Identification

The existing lithologic identification method is mainly to calibrate the logging curve through a small number of logging cores, and use the obtained logging curve data to identify the lithology of the whole interval. Intelligent interpretation method is combined with logging processing and interpretation to identify lithology. Its general idea is to select logging curves sensitive to lithology identification as input curves based on core data and lithology sensitivity analysis of logging curves, so as to realize lithology identification based on intelligent algorithm.

Decision Tree Algorithm is a kind of supervised learning. According to the weight of logging parameters in clastic rock research area in lithology identification model, the sensitivity of each parameter to lithology change is determined, so as to identify lithology. The decision tree method of C5.0 has effectively improved the accuracy of lithology identification. Decision tree algorithm also has high accuracy in identifying complex carbonate rocks. For the model with huge data, XGBoost algorithm can be applied, which adopts multithreading and distributed computing methods, greatly shortening the training time. It has a good recognition effect on limestones and dolomites, followed by argillaceous limestones, dolomites and argillaceous dolomites, and the recognition rate of calcareous dolomites is low. Boosting Tree algorithm can also effectively determine the lithology of complex glutenites [15–18].

The random Forest Algorithm, which shows great advantages in thin layer identification, has strong generalization ability, insensitivity to feature loss, fast training speed and simple implementation. Based on the lithology sensitivity analysis of logging curves, a suitable logging curve is selected as the input curve, and the lithology identification model of complex carbonate rocks is established by using random forest algorithm, which is accurate for lithology identification [10].

Another advantage of random forest algorithm is lithology identification of volcanic rocks. The lithology of volcanic oil and gas reservoirs is changeable, so it is difficult to identify them accurately by conventional methods. Different types of volcanic rocks include volcanic breccia and fused breccia. Lava mainly includes basalt, andesite, dacite and rhyolite. Due to their differences in chemical composition, mineral composition and physical properties, there are some changes in their corresponding logging response characteristics, thus identifying lithology [21].

Principal Component analysis (PCA) is one of unsupervised learning, and the key to identification is to convert the comprehensive response characteristics of various logging curves to the principal component of prominent lithology, so the identification accuracy of alternate thin layers of volcanic rocks or shale with complex lithology is high [20, 22]. By combining BP neural network lithologic prediction model and Dropout mechanism, Dropout-BP neural network combines conventional logging parameters, upgrading the

conventional two-parameter crossplot to a multi-parameter neural network, and integrating the composition, structure and electrical properties of volcanic rocks to carry out lithologic prediction, which is more effective [23].

Data mining method of logging lithology identification based on emergent self-organizing mapping. Large-scale neurons and borderless torus mapping are used, visualized by U matrix, and finally clustered and classified by manual interaction. This method can effectively find hidden patterns in high-dimensional data, and is especially suitable for logging identification of complex lithology [24].

The application of multivariate statistical algorithm needs to preprocess logging data, including logging parameter selection, logging data normalization and dimensionality reduction. Its application effect is good [25], but it can only solve the simple linear relationship problem.

By constructing the technology of recovering the missing core picture information and combining with migration learning, mine lithology is identified. The corresponding core sample information is automatically synthesized from the logging curve data of non-coring wells, and the existing data is automatically learned and analyzed by using migration learning technology, so that the core sample information of coring wells is migrated to non-coring wells, and a logging lithology identification model aided by core samples is established. The establishment of intelligent identification model of cores based on migration learning is helpful to improve the accuracy of lithology identification of oil and gas reservoirs with complex cores, and logging curves are used to predict lithology quickly and accurately [26].

In order to integrate the algorithms and modules into the unified software, the data resource Lead software has been developed by CNLC, which includes reservoir parameter calculation module based on conventional logging data, single porosity calculation module based on clastic rock, CRA module based on carbonate rock calculation and CLASS module based on lithology classification. Its advantage is that it can choose the appropriate calculation model according to the background of different regions, which is convenient for rapid processing and interpretation of logging data. But the model coverage is not comprehensive enough.

4.2 Automatic Layering and Identification of Reservoirs

There are three main guiding ideas for automatic layering: (1) variance analysis of logging values and finding inflection points and half amplitudes on curves. The guiding ideology of variance analysis is that the intra-layer difference is small and the inter-layer difference is large. At the same time, the inflection point and half amplitude point are found on the logging curve by differential and slope extreme point. (2) According to the logging data, judge the rock attributes or calculate the membership degree of rocks, and merge the same lithology, so as to realize layering (see Fig. 4). (3) Divide strata by flow unit method based on fluid properties. In actual interpretation work, automatic layering is carried out according to the priority order of fluid > lithology > curve [29, 30].

Three kinds of methods: mathematical statistical methods include intra-layer difference method, ordered cluster analysis, extreme variance clustering method and change point analysis method (least square method and maximum likelihood estimation method); Non-mathematical statistical methods include activity function method

and wavelet transform method; Artificial intelligence methods include cluster analysis, fuzzy mathematics and neural network methods. These methods have their respective advantages and disadvantages.

Mathematical statistical method is strict in mathematics, which can keep the uniformity inside the rock strata, the difference between the rock strata is great, and the calculation amount is large. In addition, It has a very high requirement for the one-to-one correspondence between logging information and geological information. If it can't be achieved (in fact, it can't be completely achieved), the layering result is a perfect mathematical result, which is not easy to meet the requirements of geological application. Among the non-mathematical statistical methods, the activity function method has good application effect and can quickly identify various types of curves; Wavelet transform can simulate the artificial interpretation process of "from coarse to fine, layering step by step" through multi-scale analysis, so as to avoid layering on a visual level and being unable to distinguish between the local and overall information of strata [27].
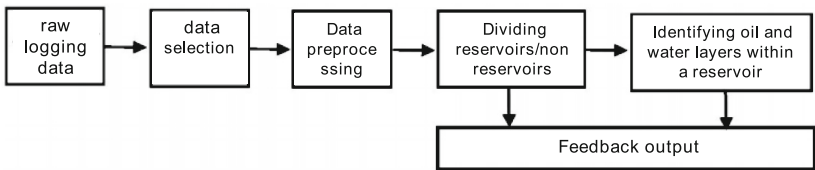

Fig.4. Technical route of automatic layering and identification with logging curves.
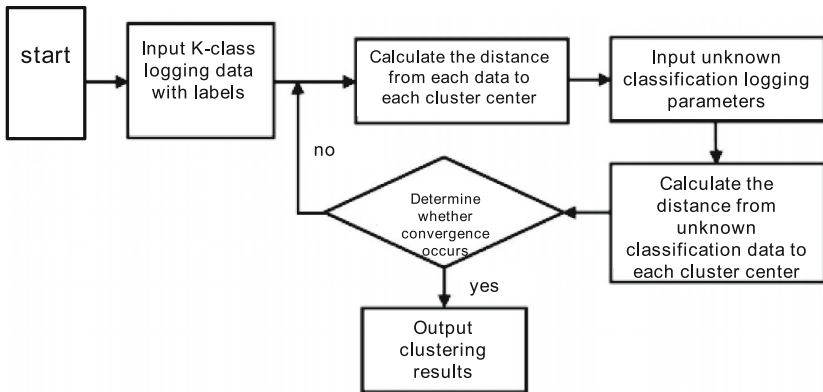

Fig.5. Flow chart of multi-granularity clustering method (according to Ji Qingqing).

In the application of artificial intelligence methods, the multi-granularity clustering algorithm with good effect belongs to supervised learning (see Fig. 5). This intelligent algorithm can quickly and accurately solve various classification problems, extract the characteristics of different layered logging curves by learning standard logging curves and layering results, and then identify oil-water layers on the basis of dividing reservoirs. In the case of standard big data, firstly, the original logging curves are analyzed by

principal component analysis, and then the relationship between each original logging curve and principal component is analyzed by principal component load matrix, and then the logging curves used for automatic layering with logging curves are selected [28, 44, 45].

The other is a knowledge-driven neural network reservoir evaluation model (KPNFE) based on the knowledge map of reservoir logging. Its functions mainly include: (1) multi-dimensional and multi-scale extraction of characteristic parameters that describe oil and gas reservoirs in detail; (2) The entities, relationships and attributes associated with these characteristic parameters are represented as vector characteristic graphs by graph embedding technology; (3) Realizing intelligent identification of oil and gas reservoirs; (4) Organically integrate expert knowledge into intelligent computing, and establish an evaluation system and optimization algorithm for potential layer recommendation [32]. The KPNFE model inherits and promotes expert knowledge and experience, effectively solves the problem of robustness in oil and gas reservoir identification, and its calculation results are highly interpretable and accurate, and it is an effective method for re-logging evaluation of old wells in old areas with high efficiency and high quality.

4.3 Sedimentary Microfacies Identification

The traditional method of identifying sedimentary microfacies is achieved manually by geologists according to their own knowledge and experience. This manual interpretation is subjective and time-consuming, and may introduce human bias. The method of identifying sedimentary microfacies based on logging curves usually includes three steps: logging curve layering, feature extraction and classification [31].Typical classification algorithms include Bayesian criterion, linear discriminant analysis, fuzzy logic, convolutional neural network method, K nearest neighbor algorithm, SVM, ANN and so on. The process of depth learning method based on logging curve is as follows: (1) data preprocessing: (2) data marking and division: (3) model training: (4) model verification. However, due to its own limitations, a single intelligent method is difficult to complete the task of sedimentary microfacies identification alone.

Convolutional neural network method takes into account the morphological change characteristics of logging curves in depth direction and the need to integrate the three steps of curve layering, feature extraction and classification. Aiming at the multi-scale and time series of logging curves, a logging sedimentary microfacies identification model, Improved U-net, with multi-scale characteristics constraints has been established, which can well identify distributary channel, channel side margin and distributary bay with different scales. KD-SegCaps, a logging sedimentary microfacies identification model with time series constraints, can well identify sedimentary microfacies such as sand flat, sand mud flat and mud flat [31]. Using DMC-BiLSTM, an intelligent identification method of sedimentary microfacies based on feature construction (DMC) and bidirectional long-term and short-term memory network (Bilstm), the geological trend characteristics, median filtering characteristics and clustering characteristics have been constructed. Compared with the convolutional neural network method, this method is helpful to extract the hidden features of logging curve sequence, and has better recognition performance for sedimentary microfacies such as distributary bay, front sheet sand, distributary channel, estuary bar and channel side maigin [39].

5   Intelligent Logging Interpretation Method Process and Data Architecture

Intelligent logging interpretation integrates all kinds of deep learning algorithms combining the characteristics of logging interpretation business, so that intelligent algorithms are integrated with traditional logging interpretation concepts. Its steps are intelligent model training, model combination and automatic recommendation (see Fig. 6).

The Digital Reservoir Research System (RDMS) pioneered by Changqing Oilfield is divided into four layers such as data layer, data link, support layer and application layer. Functionally, it includes five platforms: basic management, data service, collaborative research, decision support and cloud software [1, 13, 34, 35]. Inspired by this model, CNLC has established a big data ecology based on logging data lake (see Fig. 7). Taking logging data as the main body, CNLC has built up a regional lightweight lake by gathering logging data at home and abroad. Data are transmitted to various professional libraries by means of automatic collection of the Internet of Things and manual standardized collection, and then merged into the data lake after cleaning and processing. Real-time data and video data of industrial control are stored nearby [4]. CNLC has studied key technologies such as data integration and professional software interface, and developed and integrated exploration and development business model, multi-source data of oil and gas reservoirs, multidisciplinary professional software and online analysis tools. Realizing the coupling and integration of professional software, intelligent application and data lake. In the application scenario, data loading can be completed. On the basis of core analysis, the interpretation conclusion has been re-recognized, the logging characteristic values of the target horizon of each well have been marked, and the sample data have been submitted in batches by layers. In addition, these data are stored in the local computer work area, and can be adjusted and updated to the sample library at any time. The mode from big data platform to data lake plus interface can meet the requirements of different logging geological structure analysis scenarios.
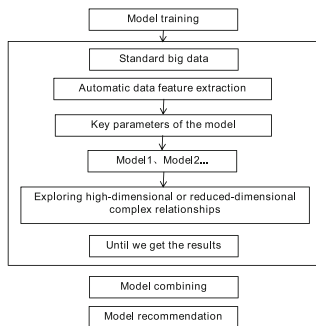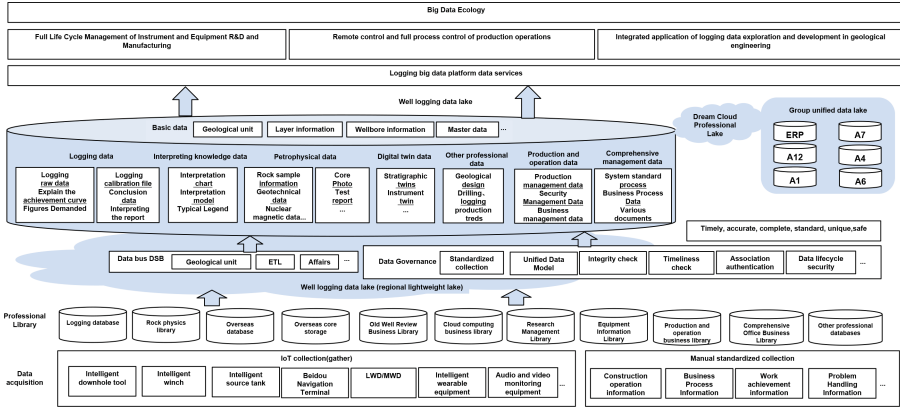

Fig.6. Intelligent Interpretation Method Process.

Fig.7. Intelligent Logging Interpretation Data architecture diagram Based on Standard Big data Platform(CNLC).

6 Summary and Prospect

There are many kinds of intelligent interpretation methods, and different methods have their own advantages and disadvantages in lithology identification, automatic stratification, hydrocarbon reservoir identification and sedimentary microfacies identification of clastic rocks, complex carbonate rocks, shale and volcanic rocks. Through model training, the optimal method can be obtained so as to improve the interpretation accuracy and efficiency of complex reservoirs.

Big data is the foundation of intelligent interpretation. In practice, a high variable dimension may not have high analytical accuracy, and sometimes it may even have the opposite effect. Through the logging data management and data quality evaluation of the big data platform, the logging data of hundreds of thousands of wells have been managed to ensure the accuracy, completeness and standardization of the data and facilitate data call. In the application scenario, the interpretation conclusion has been re-recognized on the basis of core analysis. The mode from big data platform to data lake plus interface can meet the requirements of different logging geological structure analysis scenarios.

Generally speaking, the infrastructure layer realizes IOT perception and resource support, the data sharing layer realizes data entering the lake and comprehensive management, the middle platform layer builds shared and reused data and business service capabilities, and the application layer builds lightweight and agile intelligent application scenarios. It has realized the transformation of production and operation from man-machine combination to intelligent cooperation, business management from process-driven to data-driven, and business decision-making from experience management to intelligent analysis, thus building a digital logging ecology and building a digital enterprise.

References

1. Hua, W., Yushun, Z.: Research status and prospect of artificial intelligence in logging data processing and interpretation [J]. Well Logg. Technol. **45**(4), 345–356 (2021)

2. Yong, Y.: Application progress of big data & AI technologies in exploration and development of Shengli Oilfield [J]. Petrol. Geol. Recov. Effi. **29**(1), 1–10 (2022)
3. Ailin, J., Jianlin, G.: Key technologies and understandings on the construction of smart fields. Pet. Explor. Dev. **39**(1), 118–122 (2012)
4. Yujiang, S., et al.: Development and application of intelligent logging interpretation system based on big data [J]. China Petrol. Explor. **26**(2), 113–126 (2021)
5. Yong, Y., et al.: Research on construction of data interlinked lakes of E & P dream cloud [J]. China Petrol. Explor. **25**(5), 82–88 (2020)
6. Zhiyong, L., et al.: Technology research and construction scheme of unified data lake [J]. Telecommun. Sci. **37**(1), 121–128 (2021)
7. Zhou, J., et al.: Research and application of well logging database system [C]. In: China Petroleum and Petrochemical Enterprise Information Technology Exchange Conference, pp. 393–398 (2016)
8. Zhou, J., et al.: Construction and application of unified logging database [J]. Well Logg. Technol. **46**(6), 757–761 (2022)
9. Fan, H.: Research on intelligent recognition of indicator diagrams based on artificial intelligence algorithms [D]. In: Henan University of Science and Technology, Luoyang (2019)
10. Wang, Q., et al.: Identification of complex carbonate lithology based on random forest algorithm [J]. Chin. J. Eng. Geophys. **17**(5), 550–558 (2020)
11. Teng, M.: Research on Wave Impedance Inversion Method Based on Improved Firefly Algorithm [D]. Jilin University, Jilin (2021)
12. Sun, Y., et al.: Logging identification of complex carbonate rock lithology based on XGBoost algorithm [J]. Lithological Reservoirs, **32**(4), 98–106 (2020)
13. Cheng, X., Song, X., Li, G.: Design and implementation of ecological clouds for artificial intelligence logging driven by big data and algorithms [J]. Well Logg. Technol. **45**(3), 233–239 (2021)
14. Rao, W., et al.: Multi-source heterogeneous data governance based on semi-supervised learning [J]. J. Tongji Univ. (Natural Science), **50**(10), 1393–1404 (2021)
15. Miao, T., et al.: Lithology identification method for clastic rock based on data mining technology and its application [J]. Complex Reservoirs, **14**(1), 39–44 (2021)
16. Rui, W., Xiaomin, Z., Lichang, W.: Identifying carbonate rock lithology by data mining [J]. Well Logg. Technol. **36**(2), 197–201 (2012)
17. Kai, J., et al.: A logging lithology identification model based on boosting tree algorithm [J]. Well Logg. Technol. **42**(4), 395–400 (2018)
18. Wang, Z., Liu, J., Ren, L.: The method for lithology classification in geophysical well logging based on the k-means dynamic clustering analysis [J]. J. East China Inst. Technol. (Natural Science Edition), **32**(2), 152–156 (2009)
19. Mao, R., et al.: Identification of mixed rock lithology based on lithology scanning logging: taking the Fengcheng formation of Mahu sag as an example [J]. Xinjiang Petrol. Geol., **43**(6), 743–749 (2022)
20. Liu, Y., et al.: Application of principal component analysis method in lithology identification for shale formation [J]. Fault-Block Oil Gas Field, **24**(3), 233–239 (2017)
21. Kai, L., et al.: Intelligent identification and prediction of volcanic rock lithology based on machine learning [J]. Spec. Res. **29**(1), 38–45 (2022)
22. Li, S., et al.: Lithology identification of carboniferous volcanic rocks in Xiquan area, eastern Junggar Basin [J]. Lithological Reservoirs, **33**(1), 258–266 (2021)
23. Yiming, H., et al.: Identification of volcanic rock lithology using neural networks based on conventional logging data: a case study of the changling fault depression in the southern songliao basin [J]. World Geology **40**(2), 408–418 (2021)

24. Haifeng, G., et al.: Data mining based on emergent Self-organizing mapping: a new method for logging lithology identification [J]. J. Petrol. Nat. Gas (J. Jianghan Petrol. Inst.) **31**(2), 67–70 (2009)
25. Li, Z., et al.: Application of data mining methods in logging lithology identification [J]. Fault Block Oil Gas Field, **26**(6), 713–718 (2019)
26. Ren, Y., et al.: Intelligent identification technology for lithology based on comprehensive consideration of core images and logging curves [J]. Tech. Appl. 78–80 (2021)
27. Bo, X., et al.: Review and prospect of automatic layering methods with logging curves [J]. Prog. Geophys. **25**(5), 1802–1810 (2010)
28. Ji, Q., et al.: Automatic layering recognition method for logging curves based on multi-granularity clustering [J]. Chin. High Technol. Lett. **30**(12), 1215–1224 (2020)
29. Xianmin, Z., Jianhua, W.: Artificial intelligence of the lithologic auto-recognition from digital well-logging data in engineering [J]. J. Tianjin Univ. **34**(5), 633–635 (2001)
30. Hui, Y., et al.: Application of wavelet transform characteristics of logging curves in automatic stratification. J. Geophys. **43**(4), 568–573 (2000)
31. Yongxiang, H.: Intelligent Identification Method of Sedimentary Microfacies from Logging Curves [D]. University of Electronic Science and Technology of China, Chengdu (2021)
32. Guoqiang, L., et al.: Construction of well logging knowledge graph and intelligent identification method of hydrocarbon-bearing formation [J]. Pet. Explor. Dev. **49**(3), 502–512 (2022)
33. Li, H., et al.: Research on data mining methods for logging evaluation of complex reservoirs [J]. J Acta Petrolei Sinica, **30**(4), 542–548 (2009)
34. Yang, H., et al.: Construction and application of an integrated information platform for reservoir research and decision-making [J]. China Petrol. Explor. **20**(5), 2–8 (2015)
35. Tan, M., et al.: Progress in research on committee machine logging interpretation driven by multiple source data by the commission [J]. Geophys. Prospect. Petrol. **61**(2), 233–239 (2022)
36. Di, Q., et al.: Intelligent steering drilling technology based on cloud big data [J]. J. Eng. Geol. **29**(1), 162–170 (2021)
37. Changchun, Z., et al.: Research on ROP prediction model based on fusion of Feature selection algorithm [J]. Drill. Eng. **49**(4), 233–239 (2022)
38. Jianguo, C., Wei, M., Ye, L., et al.: Research of sandstone reservoir physical properties estimation based on Elman neural networks with hybrid dimensionality reduction [J]. Sci. Technol. Eng. **14**(3), 24–28 (2014)
39. Renze, L., et al.: Intelligent identification of sedimentary microfacies based on DMC-BiLSTM. Geophys. Prospect. Petrol. **61**(2), 253–261 (2022)
40. Wang, X., et al.: Koko: a system for scalable semantic querying of text[C]. In: Proceedings of the VLDB Endowment, pp. 2018–2021. VLDB Endowment, Rio de Janeiro (2018)
41. Mudgal, S., et al.: Deep learning for entity matching: a design space exploration [C]. In: Proceedings of the 2018 International Conference on Management of Data, pp. 19–34. ACM, New York (2018)
42. Trivedi, R., et al.: LinkNBed: multigraph representation learning with entity linkage [C]. In: Proceedings of the 56th Annual Meeting of the Association for Computational, pp. 252–262. ACL, Melbourne (2018)
43. Konda, P., et al.: Technical perspective: toward building entity matching management systems [J]. ACM SIGMOD Rec. **47**(1), 33 (2018)
44. Amanipour, V., Ghaemmaghami, S.: Median filtering forensics in compressed video [J]. IEEE Signal Process. Lett. **26**(2), 287–291 (2019)
45. Green, O.: Efficient scalable median filtering using histogram-based operations [J]. IEEE Trans. Image Process. **27**(5), 2217–2228 (2018)
46. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning [M]. MIT Press, Cambridge (2016)

47. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neuralnetworks [C]. In: The 25th International Conference on Neural Information Processing Systems, New York, USA (2012)
48. Vincent, P., et al.: Stacked denoising autoencoders:learning useful representations in a deep network with a local denoising criterion [J]. J. Mach. Learn. Res. **11**(12), 3371–3408 (2010)
49. Hinton, G.E., Osindero, S., The, Y.W.: A fastlearning algorithm for deep belief nets [J]. Neural Comput. **18**(7), 1527–1554 (2006)
50. Wagstaff, K., et al.: Constrained K-means clustering with background knowledge[C]. In: Eighteenth International Conference on Machine Learning, San Francisco, USA (2001)
51. Basu, S., Banerjee, A., Mooney, R.J.: Active semi-supervision for pairwise constrained clustering [C]. In: Proceedings of the SIAM International Conference on Data Mining, Lake Buena Vista, USA (2004)
52. Ren, Y., et al.: Semi-supervised deepembedded clustering [J]. Neurocomputing **325**(1), 121–130 (2019)
53. Bergen, K.J., et al.: Machine learning for data-driven discovery in solidearth geoscience [J]. Science **363**(6433), eaau0323 (2019)