Research on the Application and Optimization of Knowledge Graph in Automatic Machine Learning

Jiali He, Yong Liu

Henan University of Science and Technology, Luoyang 471000, China.

Abstract: Automatic Machine Learning (AutoML) refers to the use of machine learning techniques to automate the entire process of machine learning, including data preprocessing, feature selection, model selection, and hyperparameter optimization. As a structured method for representing and storing knowledge, knowledge graphs have broad application prospects in automatic machine learning. By fully utilizing the information in the knowledge graph, the intelligence and decision-making ability of automatic machine learning systems can be strengthened, promoting the development and application of machine learning technology in various fields.

Keywords: Knowledge Graph; Automatic Machine Learning; Application; Optimization

Introduction

With the rapid development of artificial intelligence technology, automatic machine learning has become an important means of achieving intelligence. As an efficient way of representing and storing knowledge, knowledge graphs have been increasingly applied and deepened in automatic machine learning. By graphing domain expert knowledge, data features, and model selection, it is possible to better support decision-making and optimization processes in automated machine learning. However, there are still challenges in the quality, data privacy, and computational efficiency of knowledge graphs, which require further research and optimization to promote the further application and optimization of knowledge graphs in automatic machine learning.

1. Definition of knowledge graph

Knowledge graph is a technique that represents and stores knowledge in the form of a graph structure. It is a semi-structured or structured data model used to organize and express entities, attributes, and relationships between entities, as well as their semantic meanings. Knowledge graphs can capture and present the essence of domain knowledge and provide machine understandable information. In a knowledge graph, knowledge is represented by nodes, and the relationships between nodes are represented by edges or links. Nodes can represent entities in the real world, such as people, places, events, etc. They can also represent concepts, attributes, etc. Relationships or links describe the relationships between these nodes, such as synonymous relationships, hierarchical relationships, and associative relationships. The process of establishing a knowledge graph involves knowledge acquisition and modeling by domain experts, and can also be supplemented and expanded through automated processes. An important feature of knowledge graphs is the use of semantic annotation, which utilizes semantic representation techniques to annotate nodes and relationships semantically. This enables machines to understand and infer the knowledge contained in the graph. Knowledge graphs are widely used in fields such as natural language processing, intelligent search, and recommendation systems, providing machines with more comprehensive, accurate, and semantically rich knowledge support.The construction process of knowledge graph is shown in Figure 1:

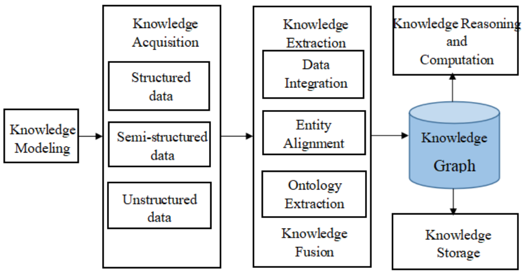

Figure 1.Knowledge Graph Construction Process

2. The application advantages of knowledge graph in automatic machine learning

2.1 Knowledge graph provides the integration and sharing of domain expert knowledge

A knowledge graph can serve as a unified knowledge base, integrating and sharing the knowledge of multiple domain experts. By transforming domain knowledge into a knowledge graph, knowledge from different domains can be structured and presented with semantic connections between entities, attributes, and relationships. This enables machine learning algorithms to better utilize both internal and external knowledge, thereby improving the training effectiveness and generalization ability of the model. Knowledge graphs can also help address the issues of missing and incomplete domain knowledge. By supplementing missing nodes and relationships, knowledge graphs can provide more comprehensive and accurate knowledge support, thereby improving the performance of automatic machine learning. At the same time, knowledge graphs can also store and update the knowledge of domain experts, achieving long-term accumulation and iterative improvement of knowledge, thereby enabling automatic machine learning systems to have the ability to continuously learn and adapt to changes.

2.2 Knowledge graph supports model selection and hyperparameter optimization

Automatic machine learning requires considering the selection of multiple models and hyperparameters to achieve optimal model performance. Knowledge graph can serve as a guide for selecting models and hyperparameters, providing effective knowledge reasoning and decision support. By organizing existing experimental data, domain knowledge, and model evaluation results into a knowledge graph, statistical relationships and inference rules in the graph can be utilized to assist in selecting the most suitable model and hyperparameters for a specific task. In addition, knowledge graphs can also support comparison and transfer learning between models. By modeling and abstracting the performance and features of different models, the similarity and correlation relationships between models can be defined in a knowledge graph. In this way, when encountering new tasks, suitable models can be quickly selected and adjusted based on task requirements and information in the knowledge graph, thereby accelerating the process of model development.

2.3 Knowledge graph provides explanations and extensions of domain expert knowledge

In automatic machine learning, the interpretability and interpretability of models are important research directions. Knowledge graphs can provide explanatory support for model prediction and decision-making. By explaining the nodes and relationships in the knowledge graph, it is clear how the model utilizes domain expert knowledge for decision-making. This interpretability not only helps researchers understand the intrinsic mechanisms of the model, but also helps users evaluate and adjust their trust in the model results. In addition, knowledge graphs can provide more comprehensive and in-depth domain expert knowledge by expanding existing knowledge. When encountering new data or tasks, the transfer learning and reasoning abilities of knowledge graphs can be utilized to obtain relevant domain knowledge from the graph and apply it to model training and optimization. In this way, not only can the performance of automatic machine learning be improved, but it can also promote the extension and development of domain knowledge.

3. The application of knowledge graph in automatic machine learning

3.1 Application of knowledge graph in feature selection and data preprocessing

Feature selection and data preprocessing are important steps in automatic machine learning, and knowledge graphs can provide valuable support for these steps. The domain expert knowledge stored in the knowledge graph can be used to guide the process of feature selection. By matching features with entities and attributes in the graph, feature selection can be based on the relevance and importance of domain knowledge, thereby reducing data dimensions and improving the effectiveness and generalization ability of the model. Knowledge graphs can also be used for missing value filling and outlier handling in data preprocessing. By analyzing the correlation between missing and outlier values and the graph, similar instance attribute information can be used to fill or correct them, improving data quality and model reliability.

3.2 Application of knowledge graph in model selection and hyperparameter optimization

Model selection and hyperparameter optimization are two key tasks in automatic machine learning, and a knowledge graph can serve as a knowledge base that includes the performance and features of different models and hyperparameters. By storing and maintaining the evaluation results and experimental data of various models in the knowledge graph, users can choose the most suitable model according to task requirements. In addition, knowledge graphs can also be used for hyperparameter optimization. Hyperparameters are parameters that need to be manually adjusted in a model, and their selection often depends on domain knowledge and experience. By associating hyperparameters with model performance and attributes in the knowledge graph, statistical rules and inference methods in the graph can be utilized to assist in selecting and adjusting the optimal hyperparameters, thereby improving model performance and training efficiency. Figure 2 shows the AUC results for different data sets in different AutoML tools:

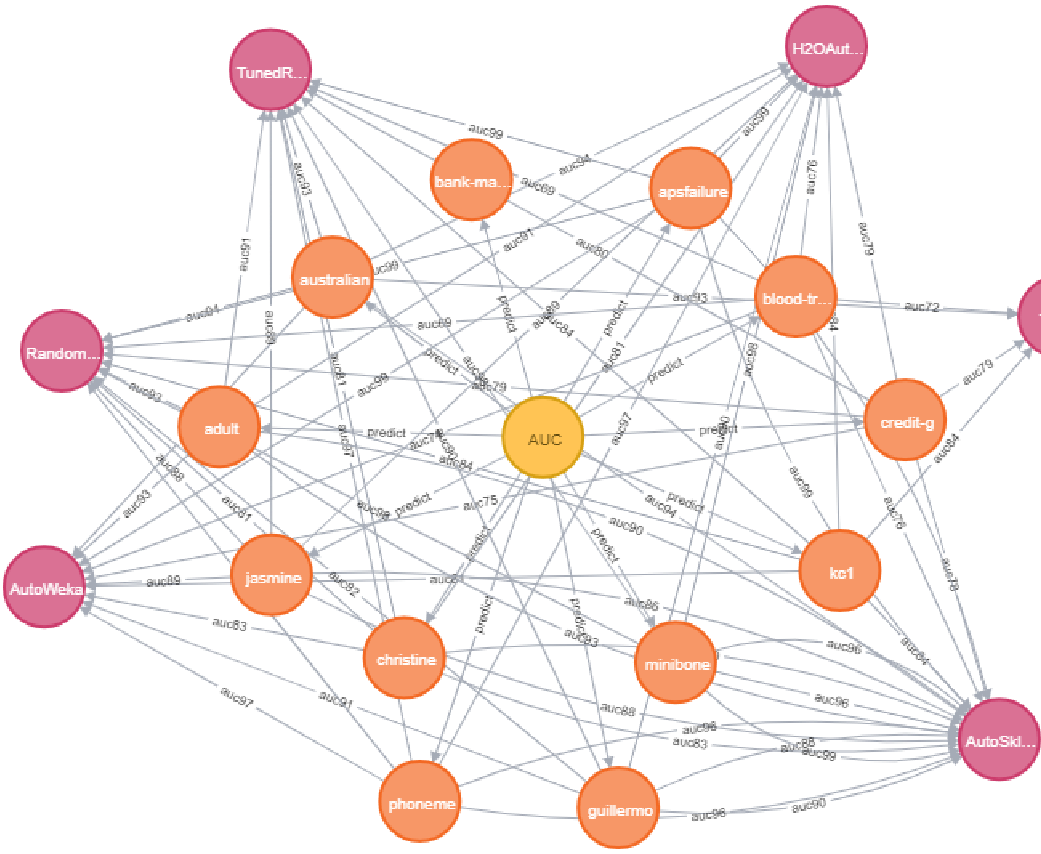

Figure 2.the AUC Results for Different Data Sets in Different AutoML Tools

3.3 Application of knowledge graph in model interpretation and interpretability

Knowledge graphs can help users understand the reasons and basis for model decisions by providing explanations and supplements of knowledge. By explaining the nodes and relationships in the knowledge graph, interpretable support for model decision-making can be formed, enabling users to clearly understand how the model makes decisions based on domain knowledge. In addition, knowledge graphs can further enhance the interpretability of models. By comparing the output and prediction results of the model with relevant domain knowledge in the graph, the consistency of the model's decisions with human experts can be evaluated. This enhanced interpretability helps to increase user trust and recognition of the model, thereby promoting its promotion and application in practical applications.

3.4 Application of knowledge graph in domain knowledge expansion and transfer learning

With the continuous changes in tasks and data, machine learning systems need to have the ability to learn new knowledge and transfer previous knowledge. The knowledge graph plays an important role in this regard. By storing and indexing knowledge graphs, it is easy to obtain new knowledge and experience from domain experts, and quickly adapt to new tasks and data. In a knowledge graph, not only is the relevant information of the model and data stored, but also the relationships and hierarchical structure between them are preserved. Through transfer learning algorithms and knowledge transfer techniques in graphs, existing knowledge and experience in graphs can be utilized to quickly transfer learning to new tasks, reducing data requirements and training time, and improving learning effectiveness.

3.5 Application of knowledge graph in open domain knowledge integration and distribution

A knowledge graph can serve as an open platform to integrate and distribute the knowledge of domain experts. By collaborating and sharing resources with experts from different fields, various knowledge sources can be integrated into a unified knowledge graph, building a comprehensive and multi domain knowledge base. In addition, knowledge graphs can promote the sharing and exchange of domain knowledge. By publishing knowledge graphs in an open manner and allowing users to participate in knowledge supplementation and revision, diverse perspectives can be integrated and collaborated. This open knowledge sharing model helps to establish a global knowledge community, promoting the rapid development and innovation of automated machine learning.

3.6 Application of knowledge graph in feature engineering

In automatic machine learning, feature engineering is a crucial step, and knowledge graphs can provide valuable domain knowledge and semantic relationships. In the future, with the continuous development of knowledge graphs, their application in feature engineering will become more intelligent and automated. Through the development of technologies such as deep learning and graph neural networks, intelligent feature selection can be achieved by utilizing the correlation between entities and attributes in knowledge graphs. For example, graph neural networks can dynamically determine the importance of each feature based on the rich information in the graph, thereby automatically selecting the most representative feature. Knowledge graphs can also be combined with natural language processing techniques to construct text features. By matching text data with entities and relationships in the graph, more semantically related text features can be generated. In the future, with the development of natural language processing technology, we can more accurately extract and construct features with domain expertise from text data.

4. Optimization strategies for the application of knowledge graph in automatic machine learning

4.1 Data preprocessing

In automatic machine learning, data preprocessing is one of the key steps to ensure model accuracy. And knowledge graphs can provide rich information sources for data preprocessing, thereby improving the quality and availability of data. Specifically, entities, relationships, and attributes in a knowledge graph can be used to discover and process noise and missing values in data. By classifying and clustering entities and relationships, inconsistencies and duplications in the data can be identified, and then data cleaning and organization can be carried out. Meanwhile, the attribute information in the knowledge graph can also help fill in missing values in the data. By calculating the similarity of attributes and reasoning relationships, existing data can be expanded to improve the diversity of training data and enhance the model's generalization ability.

4.2 Feature selection

Feature selection is another important step in automatic machine learning, aimed at reducing dimensions and eliminating redundant information, in order to improve the training efficiency and prediction accuracy of the model. The feature selection method based on knowledge graph can utilize information such as relationship paths and attribute similarity in the knowledge graph to evaluate and rank features. By analyzing the interactions between entities and relationships in the knowledge graph, it is possible to identify attributes that have decision-making ability and discrimination in feature representation. In addition, by analyzing the importance and credibility of relationship paths, feature combinations with high predictive ability can be found to improve the model's generalization ability to new data. Therefore, feature selection methods based on knowledge graphs can help automatic machine learning systems select and utilize key features more accurately, improving model performance.

4.3 Model training

During the model training process, knowledge graphs can provide rich semantic information for automatic machine learning systems to better understand and represent data. On the one hand, by linking entities and relationships in the knowledge graph to data samples, a more diverse and abstract representation of the samples can be constructed. This can better capture the semantic similarity and correlation between samples, and improve the model's ability to understand the data. On the other hand, inference algorithms in knowledge graphs can also be applied to model training. By inferring relationships and attributes, more background knowledge and prior information can be provided to the model, thereby enhancing its generalization ability and adaptability. In summary, the application of knowledge graphs in model training can improve the model's ability to express complex data and prediction accuracy, promoting the development and application of automatic machine learning systems.

4.4 Model evaluation

In the process of model evaluation, knowledge graphs can be used to evaluate the interpretability and stability of the model. The accuracy of the model can be evaluated by comparing the predicted results of the model with the true labels in the knowledge graph. The entity and relationship information in the knowledge graph can serve as reference standards to verify and analyze the prediction results of the model. If the predicted results of the model are consistent with the labels in the knowledge graph, it can be considered that the model has high accuracy. The attribute information in the knowledge graph can also be used to interpret and interpret the prediction results of the model. For example, for a text classification model, if the model predicts that an article belongs to a certain category, the relevant attributes in the knowledge graph can explain that the classification result is caused by specific keywords or other features of the article. This explanation can enhance the credibility of the model and enable users to better understand and accept the results of the model.

4.5 Model optimization

In knowledge graph based automatic machine learning, model optimization mainly includes two aspects: model parameter tuning and model structure optimization. Based on the relationship and attribute information in the knowledge graph, the parameters of the model can be adjusted and optimized. For example, in graph neural networks, by utilizing the node attributes and edge relationships in the knowledge graph, the weights and bias parameters of the network can be adjusted to improve the predictive accuracy of the model. The graph structure information based on knowledge graphs can also optimize the structure of the model. For example, by utilizing neighboring node information in a knowledge graph, more complex feature transfer methods can be designed to further improve the expressive and generalization abilities of graph neural networks. Through these model optimization measures, the performance and effectiveness of automatic machine learning models can be effectively improved.

5. Conclusion

In summary, the application of knowledge graphs in automatic machine learning plays an important optimization role. By utilizing the relationships and attribute information of knowledge graphs, improvements can be made in data preprocessing, feature selection, model training, model evaluation, and model optimization. These application measures can improve the accuracy, efficiency, and interpretability of the model, promoting the development and application of automatic machine learning technology.

References

[1] Cao Shujin, Li Ruijing. Construction and Application of Innovation Knowledge Graph Based on Patent Literature Abstracts [J]. Intelligence Theory and Practice. 2022; 45 (11): 21-28.

[2] Yang Guang. (2022) Research on Entity Recognition and Relationship Extraction Based on Biomedical Knowledge Enhancement

[D]. Shanghai: East China Normal University.

[3] Tang Wangjing. A Science Popularization Text Classification Algorithm Based on Attention Mechanism and Knowledge Graph Enhancement [D]. Beijing Jiaotong University, 2022

[4] Du Junlei. (2022) Design and Implementation of a Learning Resource Recommendation System Based on Knowledge Graph [D]. Beijing: Beijing Jiaotong University.

[5] Jin Zhe, Zhang Yin, Wu Fei, Zhu Wenwu, Pan Yunhe. Artificial Intelligence Algorithm Model Combining Data Driven and Knowledge Guided [J]. Journal of Electronics and Information Science. 2023; 45 (07): 2580-2594.

[6] Yi Xiaoyu, Yi Mianzhu. Research on Knowledge Graph Fusion Technology [J]. Computer Programming Skills and Maintenance. 2022; (08): 144-146+164.

[7] Deng Liang, Cao Cungen. A method for constructing a patent knowledge graph [J]. Computer Science. 2022; 49 (11): 185-196.

Author Introduction: He Jiali (1998.11-), female, Han nationality, Anyang city, Henan province, majors in computer technology at Henan University of Science and Technology, with a research focus on knowledge graphs and automatic machine learning.