ISSN: 2664-9640

Research on Knowledge-driven and Data-driven Fusion for University Scientific Research Management

Shuqin Li

North China University of Technology, Beijing, 100144, China

Abstract

Scientific research management is an important part of scientific research work in universities, which runs through the entire process of scientific research activities, and plays a role in integrating, guiding, encouraging, serving and supervising scientific research in universities. This paper analyzes the advantages and disadvantages of data-driven and knowledge-driven, combined with the advantages of knowledge-driven and data-driven methods, transforms the entities and relationships in the knowledge graph into low-dimensional continuous vectors, and uses the embedded results as the training data of the deep learning model to construct more powerful, interpretable and robust university scientific research management system. It promotes the innovative application of scientific research management in universities, and improves the efficiency and level of scientific research management.

Keywords

Scientific research management, knowledge-driven, data-driven.

1. Introduction

The high-quality, characteristic and connotative development of universities is inseparable from high-level scientific research support, and high-level scientific research is inseparable from scientific research management. Scientific and efficient scientific research management is an important guarantee for the development of scientific research in universities and the improvement of scientific research levels, and it is also an important point to improve the quality of scientific research in universities. So it is increasingly important to improve the efficiency of scientific research management and scientific decision-making in universities [1]. Information technologies such as deep learning and knowledge graphs provide new methods for the innovation of scientific research management in universities. Through the use of new technologies, the potential value behind data information is deeply explored, and information that is valuable for scientific research work and decision-making analysis is discovered, which is helpful to improve the level of scientific research management. It is of great significance to promote scientific research and innovation in universities.

Data-driven methods perform well in the case of large-scale data and complex tasks, and have been widely used in computer vision, speech recognition, natural language processing and other fields[2]. Data-driven methods such as machine learning and deep learning can approximate non-linear functions and have strong predictive capabilities. However, data-driven methods have higher requirements for data and are unexplainable (black box features), which can lead to reliability is low in some cases, and the judgment of expert experience is required.

Knowledge-driven methods use prior knowledge to make predictions or decisions, which are good at solving clearly defined logical problems. The process of knowledge reasoning is similar to the cognitive process of humans. The reasoning process is interpretable, but the number of features involved in knowledge-driven methods is usually relatively small. It often lacks

ISSN: 2664-9640

mathematical foundation, and unable to cope with complex reasoning tasks, and more difficult to use in high-level pattern recognition tasks (speech recognition or image classification).

Therefore, this article organically combines data-driven technology and knowledge-driven technology to build a university scientific research management system that integrates knowledge-driven and data-driven. Knowledge-driven methods provide the interpretability required for the reasoning process, and data-driven methods have better predictive capabilities. It can have both advantages at the same time, improve the efficiency and level of scientific research decision-making process.

2. Review of Related Research

In 2006, G. Hinton proposed the deep learning technology, which can effectively suppress the problem of gradient disappearance, support automatic extraction and representation of complex features, and make classification and prediction more accurate[3]. The neural network technology represented by deep learning can describe, identify, classify and explain things or phenomena by processing and analyzing various forms of information that characterize things or phenomena. Deep learning mainly includes networks such as CNN, ResNet, RNN and LSTM, and deep learning frameworks such as TensorFlow, Pytorch, and Paddle have emerged. On the basis of deep learning, enhanced learning, incremental learning, transfer learning are expanded. However, data-driven methods such as deep learning mainly input structured, semi-structured, and unstructured data, and the main calculation objects are data feature vectors rather than semantic vectors. It has certain limitations, mainly including: poor interpretability, poor robustness, poor generalizability and other shortcomings.

In 2012, Google officially released the Knowledge Graph, and built the next generation of intelligent search engine accordingly[4]. Knowledge Graph is a kind of knowledge base constructed by describing entity concepts and their relationships in the real world. With the support of Knowledge Graph, machines can realize human-like cognitive functions and discover new knowledge through conceptual reasoning and semantic computing. Different from data-driven methods such as deep learning, knowledge-driven knowledge graph technology can be promoted and applied in different tasks and different fields. Knowledge-driven adds semantic analysis, the reasoning results are interpretable, and the knowledge reasoning process is similar to human cognitive judgment, and the robustness is better.

In 2018, Zhang Bo proposed the theoretical framework system of the third generation of artificial intelligence[5]. In 2019 , Turing Award winner Yoshua Bengio pointed out that "deep learning should develop from perception-based to cognitive-based logical reasoning and knowledge expression" at the NeurIPS conference[6]. The third generation of artificial intelligence not only requires deep learning based on big data and corresponding perception recognition, but also requires machines to have cognition and reasoning capabilities. It also requires machines to have common sense and logic close to humans. This requires a two-wheel drive of data and knowledge fusion, using the four elements of knowledge, data, algorithms and computing power, to establish a general machine cognitive ability beyond the Turing test, so that the machine has the ability to reason, explain, and recognize, and let the machine "think" like a human. Cognitive intelligence not only needs to use data-driven methods to construct super-large pre-training models, but also needs to link user behavior, common sense knowledge, and cognition to actively "learn" and create.

Based on the above analysis, this research adopts "two-wheel drive", namely data-driven + knowledge-driven. According to the characteristics of scientific research data resources in universities and the different requirements for decision-making results, the large amount of data and high prediction accuracy requirements are based on big data resources and deep learning methods are used to solve them, and the requirements with high interpretability

ISSN: 2664-9640

requirements are based on knowledge graphs and knowledge reasoning methods. To solve more needs, the data-driven and knowledge-driven methods are used to solve the problem.

3. Key Technologies of Knowledge-driven and Data-driven Integration

3.1. Construct Scientific Research Management Knowledge Graph

The knowledge graph is logically divided into data layer and conceptual layer. The data layer refers to a collection of entities and relationships in the form of triples, represented by <entity, relationship, entity> and <entity, attribute, attribute value>. The concept layer is built on the data layer and is a collection of accumulated knowledge[7]. The construction of knowledge graph is an iterative process of continuous updating, including processes such as knowledge extraction, knowledge fusion, knowledge processing and knowledge reasoning. The source data is converted into a triple form through knowledge extraction, then through the entity and the ontology, the data model is added to form a standard knowledge representation, and then a new relationship combination is generated through knowledge inference. All knowledge is evaluated by quality to form a complete knowledge graph.

3.1.1. Knowledge Extraction

Knowledge extraction is oriented to semi-structured data of tables and lists, and unstructured data of texts. The available knowledge is extracted through automatic or semi-automatic technology, including entity extraction, relationship extraction, attribute extraction and time extraction. Entity extraction automatically recognizes named entities from unstructured text data to form "nodes" in the knowledge graph. After the named entities are extracted from the unstructured text data, the association relationship between the entities is obtained through relation extraction, forming the "edges" in the knowledge graph, thus forming a networked knowledge structure. Attribute extraction is to extract the features and properties of entities in the information source. For example, for scientific researchers, information such as their name, age, title, research direction, and educational background can be obtained. Event extraction is to extract event information from information sources, including time, location, personnel, and related actions. Through data integration and knowledge extraction, multi-source heterogeneous source data is unified into standard structured data to facilitate the use of knowledge graphs.

3.1.2. Knowledge Integration

Knowledge fusion is a high-level knowledge organization. Through the process of heterogeneous data integration, disambiguation, processing, reasoning verification, and updating of knowledge from different data sources, it achieves the fusion of information, data, experience, methods, and human wisdom. Form a high-quality knowledge base. Knowledge fusion includes ontology alignment and entity alignment. Ontology alignment is knowledge fusion at the conceptual level. It is the process of determining the mapping relationship between concepts, relationships, attributes, and so on. It is usually achieved by calculating the similarity between ontologies through machine learning algorithms. Entity alignment is the knowledge fusion of the data layer. Entity alignment unifies and connects the same entities in different source data. Through knowledge fusion, the integration of knowledge bases is realized, forming a more dense and unified new knowledge graph.

3.1.3. Knowledge Processing

Knowledge processing includes processes such as ontology construction, knowledge reasoning, and quality evaluation. Ontology construction is the semantic basis for the connection of entities in the knowledge graph. It is mainly presented in a network structure composed of "points, lines and planes". Points represent different entities and lines represent the relationship between entities, and planes represent the knowledge network. The ontology can

ISSN: 2664-9640

be constructed manually by manual editing, or it can be constructed automatically driven by machine learning, and then revised and confirmed by a combination of quality assessment methods and manual review. Quality assessment is to evaluate the knowledge data that has been generated and import the data that meets the standards into the knowledge graph. Quality assessment is a key step to ensure that the content of the knowledge graph is correct and usable. It is the final "quality inspection" link of knowledge processing to ensure that the ontology is constructed knowledge obtained by reasoning with knowledge is reasonable.

3.1.4. Knowledge Reasoning

Knowledge reasoning finds new associations and acquires new knowledge or conclusions by calculating the relationships between existing entities and semantic analysis of triples, thereby expanding and enriching the knowledge graph network. The objects of knowledge inference can be entities, attribute values of entities, relationships between entities, hierarchical structure of concepts in ontology library, etc. Knowledge reasoning includes entity classification, relationship recognition, graph-based reasoning and logic-based reasoning. For example, given <author A, published, paper A> and <author B, published, paper A>, <author A, co-author, author B> can be inferred.

3.2. Knowledge-driven and Data-driven Fusion Realization

On the basis of constructing the knowledge graph of scientific research management, there are two ways to integrate knowledge-driven and data-driven[8]. One is the integration at the application level, such as: (1)Analyze and calculate the same problem using data-driven and knowledge-driven methods, which can achieve mutual cross-validation of the two analysis results; (2)Use knowledge graph technology to semantically explain the data-driven analysis results. Improve the interpretability of forecast results. The second is the fusion of the technical level. The entities and relationships in the knowledge graph are extracted and transformed into low-dimensional vector embedding vector space, and deep learning methods are used to train them. Using knowledge graph embedding, the entities and relationships in the triples are projected into a continuous low-dimensional vector space. Typical models include the Trans series. Through the above embedding, not only the original structure of the knowledge graph is retained, but also the problems of sparse data and low computational efficiency in the knowledge graph are solved.

This article is based on the structured data of scientific researchers, projects, academic achievements, standards, patents, awards, etc. in the scientific research management system of universities, and forms related entities, attributes and relationships through knowledge extraction, and then constructs three related entities through knowledge fusion. Tuples, after knowledge processing and quality evaluation, construct a knowledge graph of scientific research management. The entities and relationships in the knowledge graph are extracted and converted into low-dimensional continuous vector embedding vector space, the feature word vector and knowledge entity vector are input into the deep learning model, and the deep learning method is used to train them to obtain the prediction results (Figure 1). On this basis, the entity search and relational reasoning functions are realized.

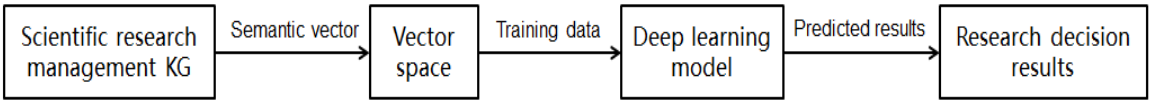

Figure 1: Framework structure of scientific research management system based on knowledge-driven and data-driven integration

ISSN: 2664-9640

4. Conclusion

Under the new situation, the important task of scientific research in universities is to carry out forward-looking research, strengthen high-precision technology research, produce more innovative results, and realize the effective combination of production, teaching and research. The development of scientific research in universities is inseparable from scientific and efficient scientific research. Management work, scientific research management must use new technologies to innovate management methods, improve scientific research management capabilities, and promote the sustained, healthy and rapid development of efficient scientific research.

At present, perception is done with data-driven methods, and cognition is mainly done with knowledge-driven methods. It is necessary to combine knowledge-driven and data-driven to give full play to the four elements of knowledge, data, algorithms and computing power to establish an interpretable robust artificial intelligence. This paper combines the advantages of knowledge-driven and data-driven methods, deeply integrates the knowledge graph and deep learning algorithms, converts the entities and relationships in the knowledge graph into low-dimensional continuous vectors, uses the embedding result as the training data of the deep learning model, and obtains the prediction result. To realize the semantic interpretability of the research and judgment results. The integration of knowledge-driven and data-driven innovation provides new opportunities for the innovation of scientific research management in universities. By digging into the potentially valuable information behind the data, it is of great significance to improve the level of scientific research management and promote scientific research innovation in universities.

Acknowledgements

The research results of this paper were funded by “North China University of Technology Yuyou Talent Support Program Project (20XN213/011)”.

References

[1] B. Zhang, J. Zhu, H, Su. Toward the Third Generation of Articial Intelligence,SCIENTIA SINICA Informationis,2020,50(09),1281-302.

[2] A. Gorka. Extending Knowledge-driven Activity Models Through Data-driven Learning Techniques, Expert Systems With Applications, 2015. 42(6).

[3] S.A. Abdul. A Hybrid Approach of Knowledge Driven and Data-driven Reasoning for Activity Recognition in Smart homes,Journal of Intelligent & Fuzzy Systems, 2019.36(5).

[4] V. Vapnik, R. Izmailov. Knowledge Transfer in SVM and Neural Networks, Annals of Mathematics & Artificial Intelligence,2017, 81(1-2),3-19.

[5] X. Zhang,. Know Risk: An Interpretable Knowledge-Guided Model for Disease Risk Prediction, In: 2019 IEEE International Conference on Data Mining (ICDM). 2019.

[6] X. Chai.Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning,IEEE Access, 2020. PP(99), p. 1-1.

[7] D.L. Jian, Y.W. Michael. An Ontology-based Hybrid Methodology for Image Synthesis and Identification with Convex Objects, The Imaging Science Journal, 2018. 66(8).

[8] K. Xu, C. Li, J. Zhu, et al. Understanding and Stabilizing GANs’ Training Dynamics Using Control Theory, In: Proceedings of the International Conference on Machine Learning (ICML), Vienna, 2020.G. Bräuninger: *Proc. International Workshop on Diamond Tool Production* (Turin, Italy, November 8-10, 1999). Vol. 1, p.154.