US-China Foreign Language, June 2025, Vol. 23, No. 6, 204-208
doi:10.17265/1539-8080/2025.06.002

The Innovative Application of Multimodal Fusion in Artificial Intelligence Decision-Making System

Haoxuan Lyu
The Chinese University of Hong Kong, Hong Kong, China

With the continuous development of artificial intelligence (AI) technology, multi-source fusion technology has gradually become an important research direction for system decisions. Multi-source fusion uses multiple data sources (such as images, sounds, texts, etc.) to improve the accuracy and flexibility of the decision system. This article first introduces the theoretical basis and technical environment of multi-source fusion and then explains the specific application of multi-source fusion in AI-determining systems in detail. Finally, it summarizes existing problems and looks forward to future development trends. Multi-source fusion not only improves the perception capabilities of AI systems, but also provides a large amount of information resources for determining the system, promoting the innovative application of AI technology in various fields.

Keywords: multimodal fusion, artificial intelligence, decision-making system, data fusion, deep learning, innovative applications, perception ability, cross-domain

Introduction

With the rapid development of artificial intelligence (AI), intelligent decisions are permeating all walks of life. However, traditional intelligent decisions are only made through one information channel and cannot completely extract multi-faceted information in the environment. In order to overcome this problem, multi-mode integration technology is introduced, which can unify multi-mode information, such as images, sound, and text to improve the ability and accuracy of the sensory system of the intelligent system. While summarizing the data, deep learning methods are used to analyze the deep laws between different data and further improve its intelligence. The article introduces the application, scientific foundation, and innovative research of multi-modal integration, and makes a prospect for future wide applications.

Current Situation and Challenges of Artificial Intelligence Decision-Making System

With the rapid maturity of artificial intelligence technology, intelligent decision-making systems have been widely used in all walks of life, especially in finance, medical care, manufacturing, and other industries, and have achieved obvious application results. Traditional decision-making mechanisms generally rely on a single data as an information source, such as mathematical data or literature data, and then use mathematical means to process the input data to obtain a definite result. However, practical problems in reality gradually become complicated; a single data source cannot cover all problems, and it is difficult to accurately express the levels in reality, which will limit the performance of decision-making accuracy and stability. When faced with complex data, chaotic

Haoxuan Lyu, Master's student, Faculty of Business Administration, The Chinese University of Hong Kong, Hong Kong, China.

data, and complex mixed heterogeneous data, traditional decision-making mechanisms will no longer be applicable. It lacks the flexibility to adapt to complex, dynamic, random, and other real applications.

Limitations of Traditional Decision-Making Systems

Traditional artificial intelligence decision-making systems often rely on a single source of information, namely data in fixed structures, simple rule engines, etc., but in real life, the sources of information are often diverse and in different forms, such as text information, image information, sound information, video information, etc. This information processing method is difficult to integrate multiple information sources at the same time, so the decision-making process of this method can only target one type of data; more importantly, there is a lot of noise in many real situations, such as news in social media, images in video surveillance, etc., and such machines cannot process this unstructured data, resulting in a reduction in the accuracy of its decision-making results. In addition, traditional decision-making mechanisms are often based on simple principles or linear models, and it is difficult to deal with complex and dynamic changes, resulting in misjudgment. These problems point out the shortcomings of traditional decision-making mechanisms in the current application of artificial intelligence, and it is urgent to introduce more advanced intelligent decision-making methods for data fusion.

How Multimodal Data Improves the Shortcomings of Decision-Making Systems

The multi-modal data integration method is proposed to solve the problems existing in the above decision-making system. It refers to the effective combination of data generated by different types (such as images, text, audio, etc.) to obtain more complete and accurate information. Diversified data provides a diverse perspective, and the decision-making system can interpret and analyze problems more comprehensively and fully. For example, in the smart medical system, doctors can analyze from multiple angles and integrate multi-modal data into different information sources, such as patient history archives, medical images (such as CT films), language expression (patient's description), etc., so as to achieve the diagnosis of the disease. For example, automobile autonomous driving systems need to integrate traditional radar information with picture information taken by cameras, as well as elevation information detected by internal sensors of the vehicle and other information on road signs, to make more appropriate decisions and deal with complex and changing road conditions. Through multi-modal data integration, decision-making systems can not only break through the information limitations of a single data source, but also integrate information from multiple angles to promote its intelligence and adaptability, so as to better cope with a complex and changing world.

Technical Principles of Multimodal Fusion

Multimodal data fusion is an important technical tool that people must apply in the process of building artificial intelligence decision-making systems. Its essence is to effectively and comprehensively organize several different types of data information in this process to provide guarantees for making more complete and accurate final decisions. Specifically, when integrating operations, we first need to master the characteristics of different types of integrated data themselves, carry out analysis work based on the specific characteristics of these data, and propose more reasonable data fusion measures on this basis, so that each different type of data characteristics can be organically combined, and further optimized and improved information value and decision-making efficiency under data integration. In addition, multimodal data fusion technology can not only smoothly process different types of data during actual operation, so that it can play its role as much as possible, and achieve

reasonable solutions to structured data, but also effectively deal with some unstructured data, such as text data, sound data, image data, etc., and have strong data application capabilities and expansion realization.

Characteristics and Classification of Multimodal Data

Multivariate information generally refers to a variety of sensory channels or information sources, and the information here may appear in the form of text, audio, images, video, etc. Information in different modes has its characteristics. For example, text provides a large amount of meaningful information, while images have spatial structure characteristics; audio information reflects emotions and tone, etc. Since the information of each pattern has its manifestation, some prior work is required when mixed together and extracting features. According to the type and source of data, we can divide multiple information into several major categories, such as image categories (such as pictures and videos), audio categories (such as recordings), written categories (such as text), sensor categories (such as data obtained by sensors), etc. Each category transmits different information and can guide the direction of solving problems. Images are suitable for image recognition and image analysis, audio is suitable for speech recognition and emotion analysis, and written ones are suitable for text analysis and emotion analysis.

Methods of Multimodal Data Fusion

The data fusion methods of each mode are divided into three types: pre-fusion, post-fusion, and intermediate process fusion. Pre-fusion refers to directly accumulating data in each mode and then fusion. The advantage of this data fusion method is that the information collection process can be fully developed at the beginning, but it may also cause problems in processing high-dimensional data. Post-fusion means that the data in each mode is processed separately first, and the results are combined again; its advantage is that the data processing can be performed using the optimal method in each mode, but the relationship information reflected between the modes may be lost. Intermediate process fusion is one of three fusion methods. In this mode, the data in each mode is initially fused, and features are extracted from it for fusion. This can ensure that the data information is complete without causing the data processing process to be too complicated. There are also some methods based on deep learning to perform multi-mode data fusion, such as multi-source neural networks. The network will automatically learn the relationship between each mode, so that the data of each mode can be effectively fusion; the deep learning process mode has the advantages of large amounts of data and many categories, and the characteristics of the multi-level mining mode are conducive to the results of fusion.

No matter what technical means are used, the fusion of multimodal data involves using a variety of information to improve the accuracy and stability of the decision-making system, thereby improving the overall performance of the artificial intelligence system.

Demonstration of a Multimodal Fusion Application of Intelligent Decision-Making AI

Currently, research focuses on in-depth exploration of single modes. The main characteristics are single modes, multi-information islandization, and a single goal. The elements that are realized are information in a single dimension in text, images, or speech.

Multimodal data fusion technology has wide application value in many industries, especially in industries, such as intelligent medical care and autonomous driving with high accuracy and safety requirements. Multimodal data fusion technology can integrate information from different types of modalities. In this way, the more information sources of the system and the deeper the understanding, the more correct the decision. The following

is a detailed analysis of the application of multimodal data fusion technology in intelligent medical care and autonomous driving.

Applications in the Field of Intelligent Medical Care

Multimodal data fusion technology is also used in health management and disease diagnosis, and treatment. By analyzing multimodal data, such as images, text, voice, and physiological data, it can strengthen the scientific, efficient, targeted, and refined disease diagnosis and treatment, and achieve the improvement of medical service methods.

Disease diagnosis system based on image and text. Intelligent medicine mainly integrates pictures and literal information to help judge the condition and make an analysis. For example, medical information pictures, such as CT, MRI, and *X*-rays are used as one of the main references for medical doctors to diagnose patients' conditions. They can present a fine structure divided into a certain part of the human body, which can help doctors understand possible diseases in a timely manner. However, this method relies heavily on medical technicians, and the pictures will be interpreted differently due to the understanding of different personnel, resulting in certain shortcomings. Text information, such as medical record writing, doctors' opinions, and patients' complaints, is of great importance and can be used as a supplementary explanation of correct diagnostic data.

Health monitoring and decision-making of voice and physiological data. Combining voice information and physiological data can provide different health interventions based on the health status of each individual. Phonetic information reflects mental health issues, such as personal emotions, speech speed, and intonation. Physiological data presents us with the first-time physical health, such as heart rate, blood pressure, body temperature, etc. We can use comprehensive processing information to enable intelligent health management devices to better understand consumers' physical and mental states and emotions.

Applications in the Field of Autonomous Driving

Vehicle artificial intelligence is one of the important areas for the implementation of artificial intelligence technology, and it has high requirements for smart cars for data collection and decision-making of various real-time information. In order to ensure that correct and reliable decisions are made in complex road conditions, the car autonomous driving system needs to use various sensing devices (camera, radar, optical radar, various sensing devices, etc.) to cooperate with each other, and better perceive its own surrounding environment through the fusion of different types of data, improving driving safety and decision-making efficiency.

The integration decision of vision, radar, and sensor data. Camera, RADAR, LIDAR, etc. are commonly used sensors for autonomous driving. The camera mainly collects peripheral images and identifies road signs, obstacles, and pedestrians. RADAR and LIDAR mainly undertake the function of depth perception, and measure the distance, speed, and surrounding environmental characteristics of the target object through radio waves or lasers. The performance of a single sensor is often not excellent enough, such as the camera's dependence on illumination, the lack of image detail resolution capabilities, and the impact of rainy days on LIDAR performance.

Cross-modal perception and decision-making optimization in complex environments. How to make the best decisions in a complex, dynamic traffic environment is the core issue of the autonomous driving system. In complex traffic scenarios, different modal information often provides different types of information. How to fully integrate this information and make the best decisions is the key to autonomous driving technology.

Conclusion

After the artificial intelligence decision-making system is integrated into multi-mode fusion technology, it has greatly improved the development process of the decision-making system. Using multi-modal information, such as vision, hearing, and text makes the artificial intelligence decision-making system more comprehensive and accurate in analysis and judgment, especially in intelligent medical care, driverless cars, and intelligent customer service. However, since the current problems of data processing and model training still exist, it is hoped that with the continuous research of technologies, such as deep learning and edge computing, future multimodal fusion can play a role in more application scenarios. For the future, with the formation of various information processing capabilities of multi-mode integration, artificial intelligence will play a greater role in decision-making and promote the development of artificial intelligence technology.

References

Chen, P. F., Li, S. M., & Zhou, L. L. (2023). Innovative application of multimodal fusion in intelligent customer service systems. *Computer Science and Technology,* 4, 112-118.

Wang, X. D., Zhu, X. J., & Gao, Z. M. (2022). Research on the application of cross-modal information fusion in autonomous driving decision-making systems. *Intelligent Transportation System,* 8, 88-92.

Zhang, W. T., Wang, J. W., & Liu, X. F. (2023). Multimodal data fusion technology based on deep learning and its application in intelligent medical care. *Journal of Artificial Intelligence,* 12, 45-50.

Zhao, G. J., Sun, Z. Q., & Ma, L. N. (2022). Application and challenges of multimodal deep learning in video analysis. *Electronics and Information Technology,* 6, 67-71.