Data-driven Human-Robot Interaction: A Frontier Exploration in Robotics

Luo Yujie

School of Pharmaceutical Sciences, Main Campus, Pulau Pinang, 11700 Penang, Malaysia

luoyujie556@gmail.com

Keywords: Data-driven; Human-computer Interaction; Technology Development

Abstract: This thesis synthesizes cutting-edge explorations of data-driven human-robot interaction technologies in different fields. Through technological tools such as deep learning, applications in the fields of intelligent assistants, home robots, medical care and rehabilitation robots, as well as education and training are having a profound impact on people's lives, work and learning. However, with the development of technology, we are facing challenges such as data privacy and security, algorithmic transparency and fairness. Therefore, we need to emphasize the ethical, legal and social policy building of technology to ensure the sustainable development of HCI technology. This paper calls for focusing on user experience and social impacts while technological innovation, so as to jointly promote the healthy development of human-computer interaction technology and make positive contributions to the construction of an intelligent and humanized society.

1. Introduction

In today's society, with the rapid development of artificial intelligence and robotics, human-robot interaction has become one of the important research directions in the field of robotics. Traditional human-robot interaction is often limited by predefined rules and instructions, and it is difficult to adapt to complex and changing real-world scenarios, so there is an urgent need for more intelligent, flexible and adaptive interaction methods to enhance user experience and robot performance. Data-driven human-robot interaction technology has emerged to enable robots to better understand user intentions, learn autonomously and optimize interaction strategies through the use of large-scale data and machine learning methods, so as to achieve a more natural, efficient and intelligent human-robot interaction experience.

The purpose of this research is to deeply explore the cutting-edge technologies, methods and applications of data-driven human-robot interaction in the field of robotics, aiming at solving the limitations of the traditional human-robot interaction methods and improving the robot's level of intelligence and the quality of user experience. Specific objectives include: analyzing the fundamentals and key technologies of data-driven techniques in human-robot interaction, and exploring their application scenarios and effects in robotic systems. Explore cross-domain data fusion and learning methods to realize the effective use of cross-modal data, and improve the robot's ability to understand and respond to user intentions.

2. Fundamentals of Human-Robot Interaction

2.1 Overview of Human-Computer Interaction

Human-computer interaction (HCI) is the process of information exchange, instruction transfer and task execution between human users and intelligent systems such as computers or robots. Traditional HCI mainly relies on input devices such as keyboards and mice and output devices such as monitors, and users interact with the system through graphical interfaces or command lines. However, with the rapid development of artificial intelligence and machine learning technologies, HCI is gradually evolving in the direction of being more intelligent, natural and flexible. Intelligent robots, as one of the important carriers of HCI, have the ability of perception, reasoning, decision-making and execution, and are able to interact with human users in a more natural and intelligent

  DOI: 10.25236/semihs.2024.012

way[1]. In the process of HCI, robots need to be able to understand multimodal inputs such as user's language, gesture, and emotion, and be able to accurately recognize the user's intentions and needs, and communicate and cooperate effectively with the user through voice, vision, touch, and other modalities.

2.2 The role of robots in human-robot interaction

Robots play an important role in human-computer interaction, and their role is mainly reflected in the following aspects:

Robots can communicate with users through voice recognition, natural language processing and other technologies, and provide personalized services, such as intelligent assistants and entertainment companions, according to users' needs and preferences. Robots are equipped with perception and execution capabilities, and can interact with users through a variety of sensory modalities such as vision, hearing, and touch to achieve a richer and more natural communication experience. Robots can provide auxiliary services for humans in industrial production, medical care, education and training, etc., performing repetitive, dangerous or tedious tasks to improve work efficiency and safety[2]. Through emotion recognition and emotion modeling technologies, robots can recognize the user's mood and emotional state and adjust interaction strategies accordingly to establish emotional connections and interactions with the user.

2.3 Introduction to data-driven human-robot interaction

Data-driven human-robot interaction is an interaction mode based on large-scale data and machine learning technology, aiming to enable robots to interact with users in a more intelligent, adaptive and personalized way by analyzing and utilizing multimodal data such as user behavioral data, linguistic data and visual data. In data-driven human-robot interaction, the robot collects, processes, and analyzes user interaction data, from which it learns user preferences, behavioral patterns, and semantic understanding, and on the basis of which it can optimize its interaction strategy, improve interaction efficiency, and the quality of user experience. Commonly used data-driven technologies include: speech recognition and natural language processing: converting the user's verbal commands into text through speech recognition technology, and then utilizing natural language processing technology to understand the user's intent and semantics, realizing voice interaction and intelligent Q&A.

3. Application of data-driven technology in human-computer interaction

3.1 Speech recognition and natural language processing

Speech recognition and natural language processing (NLP) are crucial components in data-driven human-computer interaction. Speech recognition technology converts the user's verbal commands into textual form by analyzing the speech signal, providing a basis for machines to understand the user's intent. With the development of deep learning methods, speech recognition technology has made significant progress, not only in terms of accuracy, but also in terms of robustness in noisy environments. Natural language processing, on the other hand, is responsible for understanding and processing textual information input by users. Its core tasks include named entity recognition, lexical annotation, syntactic analysis, and semantic understanding. Through natural language processing technology, the machine can understand the user's instructions, answer the user's questions, and communicate with the user in a natural and smooth dialog[3]. In recent years, natural language processing models based on deep learning, such as BERT and GPT, have achieved great success, providing more powerful tools for semantic understanding and generation.

3.2 Motion planning and behavior generation

In data-driven human-robot interaction, motion planning and behavior generation are one of the key aspects for robots to achieve natural interaction with users. Motion planning is responsible for generating the robot's trajectory in physical space to realize the execution of user commands or interaction with the user. Through a data-driven approach, a large amount of motion data and machine learning algorithms can be utilized to enable the robot to learn and optimize motion planning strategies to adapt to different environments and task requirements, and to improve the efficiency and accuracy of motion execution. Behavior generation, on the other hand, generates a sequence of robot behaviors based on motion planning, according to user commands and changes in the environment. Through the data-driven approach, historical interaction data and machine learning models can be used to predict the user's behavioral intentions and adjust the robot's behavioral strategies accordingly, enabling the robot to interact and cooperate with the user more intelligently.

3.3 Emotion recognition and emotion modeling

Emotion recognition and emotion modeling are important components in data-driven human-robot interaction, aiming to enable the robot to recognize and understand the user's emotional state, and adjust the interaction strategy accordingly to improve the emotional connection and experience quality of the interaction. Emotion recognition involves recognizing and inferring the user's emotional state, including joy, anger, sadness, surprise, etc., from a variety of input data such as speech, facial expression, and verbal text. Emotion modeling, on the other hand, involves building models to describe the relationship between user emotions and behaviors based on the recognized emotion data so that the robot can make emotionally intelligent feedback and responses. Data-driven emotion recognition and emotion modeling usually rely on large-scale emotion labeling data and machine learning algorithms. Through methods such as deep learning, the correlation laws between emotions and speech features, facial expressions, linguistic texts, etc. can be learned from the data to achieve automatic recognition and modeling of emotions. Physiological signals such as heart rate, skin electrical response, etc[4].We can also be combined to assist emotion recognition and improve accuracy and robustness.

4. Data-driven human-computer interaction system design

4.1 System architecture and components

The design of a data-driven HCI system usually includes the following key components: sensors, which are used to collect user input information and feedback data from the environment, including cameras, microphones, touch sensors, and so on. These sensors are capable of capturing multimodal data such as user's voice, gesture, expression, etc., providing rich input signals for the system. The data processing module is responsible for processing and analyzing the raw data collected by the sensors, extracting the feature information therein and pre-processing it. This feature information can be used for subsequent data-driven model training and inference processes. Machine learning models, machine learning models are the core of data-driven human-computer interaction systems, including speech recognition models, natural language processing models, emotion recognition models and so on. By learning from a large amount of labeled data, these models are able to mine the user's behavioral patterns, semantic information, etc. from the data and make intelligent decisions and responses accordingly. System architecture and components are illustrated in Figure 1.

Figure 1 System architecture and components

4.2 Data Collection and Preprocessing

Data collection and preprocessing is a crucial link in the data-driven human-robot interaction system, and its quality and efficiency directly affect the subsequent model training and interaction effects. The data collection phase involves collecting raw data of user-robot interaction, including voice, image, text and other forms of data. These data usually need to be cleaned, labeled, and other processing to ensure the accuracy of the data and the consistency of the labeling. In the data preprocessing stage, the raw data first need to be formatted and normalized to ensure the consistency and comparability of the data. Second, feature extraction and feature engineering are needed to convert the raw data into feature representations that can be understood and processed by machine learning models. This includes extracting speech features, image features, text features, etc., as well as performing operations such as dimensionality reduction and normalization.

4.3 Model Selection and Training

Model selection and training are crucial steps in a data-driven HCI system, determining the system's ability to understand and respond to user input. In the model selection phase, it is necessary to choose appropriate machine learning models, including deep neural networks, support vector machines, decision trees, etc., according to the task requirements and data characteristics. For different tasks, such as speech recognition, emotion recognition, behavior generation, etc., different types of models may need to be selected.

5. Frontier Technologies and Challenges in Robotics

5.1 Application of deep reinforcement learning in robot control

In robot control, deep reinforcement learning is mainly applied to the following aspects: path planning and motion control, where robots need to plan and execute paths in complex environments to realize autonomous navigation and obstacle avoidance. Deep reinforcement learning is able to learn optimal action strategies through interaction with the environment, as well as adaptive control strategies for different scenarios. Grasping and manipulation, in fields such as industrial automation and service robotics, robots need to learn how to grasp and manipulate various objects[5]. Deep reinforcement learning can help robots learn effective grasping strategies from trial and error, and optimize and generalize them for different scenarios. Task execution and collaboration, in collaborative robots and team robots, deep reinforcement learning can help robots learn how to perform complex tasks and collaborative strategies for teamwork and task division.

5.2 Cross-modal data fusion and learning

Cross-modal data fusion and learning refers to the effective integration and learning of multiple modal data from different sensors or data sources to improve the robot's understanding of user intent and environment perception. This approach enables robots to sense and understand the environment more comprehensively, leading to more intelligent and flexible human-robot interactions. In cross-modal data fusion and learning, multimodal data from different sensors first need to be fused and integrated to form a comprehensive data representation. For example, speech data, image data and text data are jointly processed to obtain richer and more comprehensive user input information[6]. Then, methods such as deep learning are used to learn and model the fused data, mining the intrinsic correlation and semantic information between the data, so as to achieve a deeper understanding of the user's intention and the state of the environment. Figure 2 shows the cross-modal data fusion and learning.

Figure 2 Cross-modal data fusion and learning

5.3 Human-computer collaboration and co-learning

Human-robot collaboration and co-learning refers to the process of co-operation and co-learning between human users and robots. In this model, the robot is not only a passive tool to perform tasks, but also actively cooperates and interacts with the user to complete complex tasks and learn together[7]. This mode of collaboration and co-learning helps to improve the robot's intelligence level and interaction effect, and lays the foundation for the further development of human-robot relationship. In human-robot collaboration and co-learning, robots need to have the following capabilities: context understanding and communication, robots need to be able to understand the user's intentions and context, adjust their behavioural strategies according to contextual changes, and communicate effectively with the user through natural interaction. Task division and collaborative execution, the robot needs to plan and assign tasks with the user, and collaborate with the user according to the task requirements to achieve efficient completion of the task. Joint learning and knowledge sharing, the robot needs to be able to continuously learn and accumulate knowledge from interaction with the user, and share the learned knowledge with the user to achieve common progress and improvement.

6. Application Case Analysis and Evaluation

6.1 Intelligent assistants and home robots

Intelligent assistants and home robots, as typical application scenarios of data-driven human-computer interaction, are increasingly becoming an indispensable part of people's lives. Intelligent assistants such as Siri, Alexa and Google Assistant provide users with voice interaction, information query, schedule management, smart home control and other functions through speech recognition, natural language processing and other technologies, realising the vision of natural human-machine interaction. Home robots, on the other hand, are intelligent robots designed for the home environment, aiming to provide services that are more personalised and close to users' needs. Home robots can help users with daily household tasks, such as household cleaning, entertainment companionship, and smart home control. Through interaction with family members, home robots can gradually learn the user's preferences and habits and provide a more personalised service experience.

6.2 Industrial Automation and Intelligent Manufacturing

Industrial automation and intelligent manufacturing is an important application of data-driven human-robot interaction technology in the industrial field. Through data-driven intelligent control systems and robotics, factories and production lines are able to realise more efficient, flexible and intelligent production processes. In industrial automation, data-driven control systems are able to monitor and adjust parameters in the production process in real time, thus enabling automated control and optimisation of production lines. The application of robotics enables factories to automate a wide range of complex tasks, improving production efficiency and product quality.

6.3 Medical care and rehabilitation robots

Medical care and rehabilitation robots are one of the important applications of data-driven human-robot interaction technology in the medical field. These robots can provide diverse medical services, including remote diagnosis, rehabilitation training, nursing services, etc., providing patients with a more personalised and convenient medical care experience. In the field of telemedicine, medical care robots can make use of sensors and cameras and other equipment to achieve remote monitoring and diagnosis of patients, and healthcare personnel can communicate with and inspect patients through remote control robots, obtaining the health status of patients in a timely manner and providing appropriate medical advice and guidance.

7. Conclusion

The application of data-driven human-computer interaction technology in various fields has shown a booming momentum. Through technical means such as deep learning, natural language processing, and machine vision, robots are able to understand and respond to user needs more intelligently, achieving a more natural and efficient way of interacting with humans. From intelligent assistants and home robots to medical care and rehabilitation robots, to applications in education and training, data-driven human-robot interaction technologies are bringing great change and convenience to people's life, work and learning. However, with the continuous development of the technology, we are also facing some challenges and problems, such as data privacy and security, transparency and fairness of algorithms, and humanisation and popularisation of HCI. Therefore, we need to pay attention to user experience and social impacts along with technological innovation, and strengthen the construction of ethical, legal and social policies for technology, so as to ensure that the development of human-computer interaction technology is in line with the long-term interests of human beings and social values.

References

[1] Admoni H. Toward a data-driven generative behavior model for human-robot interaction[J]. ACM, 2014.

[2] Perez-D'Arpino C, Shah J A. Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification[C]//2015 IEEE International Conference on Robotics and Automation (ICRA).IEEE, 2015.

[3] Kanda T, Ishiguro H. Human-robot interaction in social robotics[M]. CRC Press, 2017.

[4] Rooker M N. Communicative exploration: a multi-robot exploration approach[D]. Jacobs University Bremen, 2005.

[5] Senarathne, P. G. C. N. Wang, Danwei. Incremental algorithms for Safe and Reachable Frontier Detection for robot exploration[J].Robotics and Autonomous Systems, 2015, 72(Null).

[6] Keidar M, Kaminka G A. Efficient frontier detection for robot exploration.[J].International Journal of Robotics Research, 2011, 33(2):215-236.

[7] Yu C, Smith T G, Hidaka S, et al. A Data-Driven Paradigm to Understand Multimodal Communication in Human-Human and Human-Robot Interaction[J].Springer-Verlag, 2010.