The International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2022)
DOI: 10.54254/2753-8818/2/20220175

Research on Intelligent Dialogue Systems

Siyi Hu

Arizona State University, Tempe, Arizona, 85281, United States

siyihu@asu.edu

Abstract. Intelligent dialogue systems, as a subfield of artificial intelligence, have very important research significance and application value. Today's AI dialogue systems are still in a relatively early stage, but they are developing very rapidly. In recent years, intelligent dialogue systems have been applied in many fields, such as intelligent customer service in online transactions, intelligent voice assistants in smartphones, and virtual chatbots. This paper introduces the background of intelligent dialogue systems and the current research status of key technologies and discusses some challenges in this field and some recent research to improve the system. Most of the current intelligent dialogue systems can perform effective human-computer interaction and respond accordingly. But for the next generation of intelligent dialogue system, more human characteristics are needed so that it can better understand and express human language, have its own personality, and maintain the consistency and logic of dialogue.

Keywords: Intelligent dialogue system, Artificial intelligence, Natural language processing, Chatbot.

1. Introduction

Artificial intelligence is a system or machine that mimics human intelligence to perform tasks and can iteratively improve itself based on the information it gathers. An intelligent dialogue system is a form of artificial intelligence. In recent years, with the continuous development of artificial intelligence technology, various forms of chatbots provide intelligent and efficient services to the public in a new and effective way of communication and have a wide range of applications in various scenarios, such as personal assistants and customer service. According to Nicola Bleu's article, approximately 23% of customer organizations are currently using intelligent dialogue systems, and an additional 31% of customer service organizations plan to begin using them in the next 18 months [1]. These data, on the one hand, illustrate the great business value of intelligent conversational systems in the future market, and on the other hand, show the demand of users for more convenient and faster services. Having an intelligent virtual assistant or chat partner is no longer a fantasy, but with current technology, understanding and expressing human language is a huge challenge for chatbots to achieve true anthropomorphic interaction. In simple terms, dialogue systems can be classified into two types, task-oriented and open-domain, depending on the goals and target users. A task-oriented dialogue system is usually confronted with specific goals and a limited scope of knowledge and only needs to focus on answering some specific questions. This also means that the inputs and outputs are limited and relatively simple to implement. For example, KLM was the first airline to use Messager chatbots, a service that

DOI: 10.54254/2753-8818/2/20220175

not only makes it easier for customers to get real-time responses to basic questions but also reduces significant labor costs to improve manual efficiency [2].

Open-domain dialogue systems are extremely free. Open and conversational systems usually do not have any limited topic or explicit goal, which means that the system needs to have a rich domain of knowledge and be able to understand human natural language and emotions accurately. In 2020, there are many mega pre-trained models, including Google's Meena, FAIR's Blender, and Baidu's PLATO. In these open-domain conversational systems, smooth and natural conversations can be generated. These pre-trained models also push the research on conversational systems to a new high point. Due to the huge demand and business value of intelligent dialogue systems, it is of great importance to improve the next generation of dialogue systems to give users a better experience. This paper contains research on intelligent dialogue systems, which will enable the public to have a clearer understanding of the principles and trends related to intelligent dialogue systems.

2. The development of intelligent dialogue systems

The story of the intelligent dialogue system can be viewed as starting way back in 1950 when Alan Turing asked the question, "Can machines think?" [3]. According to him, a machine is intelligent if it can imitate a person and behave in such a way that other people involved in a real-time conversation believe that they are interacting with a person but not a machine. In 1966, Joseph Weizenbaum developed a system called ELIZA, which is known as the first conversational system in history [4]. ELIZA transforms input to create responses by analyzing keywords in the user's input text. Obviously, this method could only throw the question back to the person asking it, but not solve it.

In 1995, Artificial Linguistic Internet Computer Entity(ALICE) came to life [5]. ALICE worked by applying heuristic patterns to match rules and an Artificial Intelligence Mark-up Language(AIML) knowledge base. ALICE has great language processing, and it is a three-time Loebner Prize winner (2000, 2001, 2004). However, such progress still suffers from huge linguistic deflection. Therefore, it is not uncommon to see inconsistencies in answers. ALICE can add personalization directly at the most basic level by manually creating <pattern> <template> pairs and integrating personality traits in <templates>[6]. However, this makes it difficult to change personalities unless a completely different script is used.

In the 21st century, the development of intelligent dialogue systems has come to a step closer, with several major technology companies in the system releasing voice-controlled types of intelligent dialogue systems, such as Apple's Siri, Amazon's Alexa, and Microsoft's Cortana. They have added new modules for converting speech into text to traditional text-based dialogue systems. Compared to ALICE and ELIZA, this type of system is in the learning-based open domain and has a richer set of features. They can not only provide more friendly dialogues but also anticipate the user's needs and provide relevant information. In recent years, Microsoft has also been constantly updating an artificial intelligence chatbot they released in 2014 called XiaoIce[7]. By 2018, Ice has been upgraded to its sixth generation, with a relatively high level of personalization and an ongoing commitment to improving the system's IQ and EQ. As a socially intelligent chat system, XiaoIce is able to establish a long-term emotional connection and interact effectively with users.

3. Foundational framework for intelligent dialogue systems

Several different models are also considered in the design of different classes of dialogue systems. Task-oriented dialogue systems generally have a pipeline model and an end-to-end approach. Retrieval-based models and generation-based models are typically used by open-domain dialogue systems.

3.1. Pipeline architecture

The pipeline architecture acts as an assembly line, defining each part of the basic process as a separate module. Each module is processed separately and then integrated together to accomplish a task. Figure 1 describes the flow of a traditional pipeline architecture task-oriented dialogue system, which mainly

DOI: 10.54254/2753-8818/2/20220175

includes automatic speech recognition, natural language understanding, state tracking in dialogue management, natural language generation, and speech output.

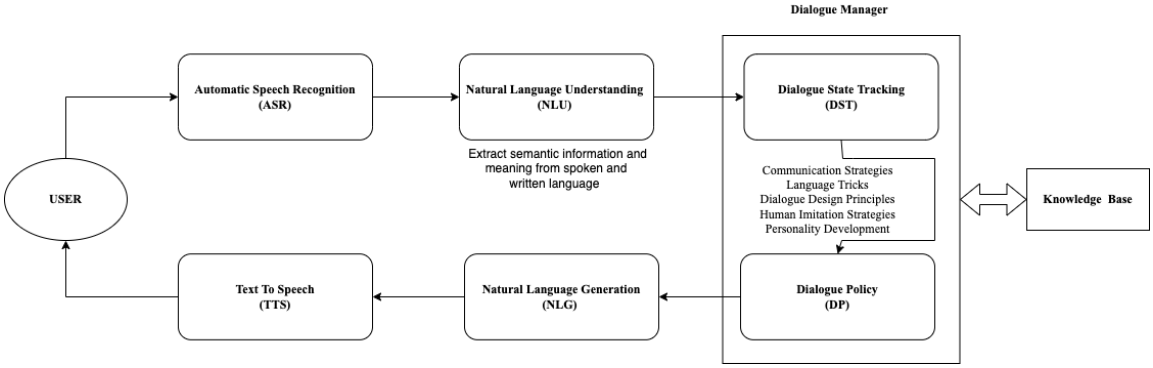

Figure 1. Intelligent dialogue systems framework.

3.1.1. Automatic Speech Recognition(ASR). The goal of automatic speech recognition technology is to convert the user's speech signal into textual information to solve a specific problem. There are many parameters that can affect speech recognition, such as pronunciation, accent, pitch, volume, or background noise. In the review by Karpagavalli and Chandra E, a traditional hybrid system for ASR is called "Generative Learning Approach-HMM-GMM", which combines the lexical model, acoustic model, and language model to perform transcription prediction [8]. Compared with the traditional ASR hybrid system consisting of HMM and GMM, the new generation of Deep Learning-HMM DNN models are easier to train, require less human effort, and are more accurate. Although the HMM DNN model has been very effective, it still cannot satisfy 100% accuracy. Improving the accuracy of speech recognition is one of the key challenges in developing intelligent dialogue systems.

3.1.2. Natural Language Understanding(NLU). Natural language understanding is a subset of natural language processing that uses syntactic structure and intended meaning analysis of sentences in text and speech to determine the meaning of sentences. In other words, intention recognition and entity extraction are the key aspects of natural language understanding. The approaches to implementing natural language understanding have gone through three iterations: rule-based approaches, statistical-based approaches, and deep learning-based approaches. Early on, people determined the intent of natural language by summarizing the rules, and later, traditional machine learning natural language understanding models based on statistics emerged, such as support vector machines(SVM), Hidden Markov models(HMMs), Conditional random fields(CRF), N-grams and naive Bayesian, etc. [9][10]. After the rise of neural networks, some deep learning-based models have also gained explosive success. Convolutional neural networks(CNN), models combining CNN and Recurrent neural networks(RNN), and long short-term memory networks(LSTM) have become prime choices for natural language understanding. A new model called Transformer was proposed by Ashish Vaswani et al. in 2017. "This new simple network architecture, which is based solely on attention mechanisms, dispensing with recursion and convolution completely" [12]. Since the Transformer model processes all words in parallel, and each word can be linked to other words in multiple processing steps, it has higher computational performance and accuracy.

3.1.3. Dialog Manager. Dialog management captures the user's intention or goal by controlling and updating the content of the context. In this regard, dialog state tracking is used to continuously extract the required values from the user dialog to fill the pre-set states to achieve the user requirements. Dialog policy management is to determine which slots still need to be asked to generate the next system action based on the current system state output from DST.

DOI: 10.54254/2753-8818/2/20220175

3.1.4. Natural Language Generate(NLG). Natural language generation enables computing devices to generate text and language from data input. As with NLU, NLG applications need to consider language rules based on morphology, lexicons, syntax, and semantics to make choices on how to phrase responses appropriately. There are usually three models to organize appropriate response utterances: rule-based, retrieval-based, and generation-based models. The three natural language generation models are presented and compared in Figure 2. Since all three models have their own advantages and disadvantages, they are usually applied in intelligent dialogue systems for different purposes.

3.1.5. Text To Speech(TTS). The function of text-to-speech is to present system-generated text and synthesized human-like speech to customers.

Model	Description	Advantages	Disadvantages
Rule-based	By manually portable targeted templates for each scenario, so for the user the question is in selection rather than generation of new text.	Accurate responses in specific domains.	Poor portability and extensibility
Retrieval-based	Indexing of questions from an existing database containing rich conversational material. Keyword extraction by analyzing the questions asked by users, using API checks and thus fuzzy matching to find the appropriate response [13].	Flexible and easily expandable knowledge base.	Poor conversation continuity.
Generative	Generate new dialogs based on the large amount of dialog training data	Data-driven, incremental learning.	Requires a huge corpus for training.

Figure 2. Characteristics of three NLG models.

3.2. End-to-end architecture

End-to-End learning architectures are a hot topic in the field of deep learning. In particular, the increasing popularity of neural models in recent years has led to a growing interest in end-to-end learning models that jointly optimize multiple components [14]. It facilitates the deep neural network(DNN) architecture composed of multiple layers to solve complex problems. Unlike pipeline architectures, end-to-end architectures usually bypass all the intermediate steps and focus on the fact that they can handle the complete sequence of steps and tasks. In the training of an end-to-end deep learning model, a prediction is generated from the input to the output side of the input. This prediction result produces an error compared to the true result, and the error is then passed back through each layer in the model until the model reaches the desired result. All the operations in between are contained inside the neural network and are no longer processed in multiple modules. In simple terms, the end-to-end architecture can be seen as a black box test.

The end-to-end model circumvents the inherent drawback of multi-module models: error accumulation and reduces the complexity of engineering. Iulian, with his teammates, uses generative hierarchical neural network models to build an end-to-end dialog system [15]. The experimental results also show the excellent performance of the end-to-end architecture. Although the end-to-end architecture has many advantages, technical improvements are still needed to ensure logical and consistent dialogues.

DOI: 10.54254/2753-8818/2/20220175

3.3. Retrieval-based architecture

Retrieval-based intelligent dialogue systems are typically used in closed domain scenarios and rely on a predefined set of responses to user messages. Retrieval-based intelligent dialogue systems perform three main tasks: intent classification, entity recognition, and response selection. Figure 3 shows the flowchart of a simple retrieval-based architecture for a question-and-answer system. The retrieval-based approach selects responses from candidate responses. The key to retrieval is message-response matching, and the matching algorithm must overcome the semantic gap between messages and responses. The basic idea of retrieval-based dialogue is to project the input and the candidate output into the same semantic space and determine whether they are similar. For the calculation of whether they are similar, the classical way is based on the unique hot encoding or bag-of-words model, which is the more traditional way of expression. When deep learning was developed, it started to use representation learning, that is, learning out vector embedding representations and finally matching similarity computation based on abstract representations.

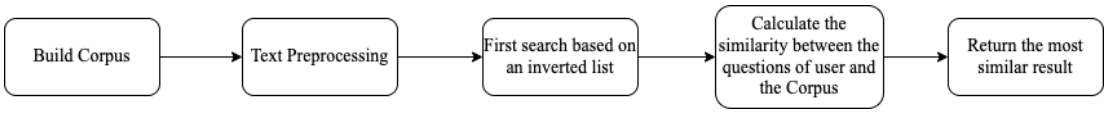

Figure 3. Flow of retrieval-based intelligent dialogue system.

3.4. Generative architecture

Generative model-based architectures, as the name implies, are usually able to generate responses based on linguistic competencies learned from a large corpus rather than relying solely on specific templates or answers. For now, the sequence-to-sequence model proposed by Ilya et al. in 2014 is the dominant model for generative dialogue systems[16]. The application of sequence to sequence in dialogue systems can be seen as a conversion from Question to Answer. This process in the dialogue system is to first encode the question X and then use the resulting encoded expression of X to predict the response Y.

4. Future and trends

Personalized conversation systems should be both highly capable of task completion and social connectivity. There is a market demand for everything that drives development. Whether it is Google's LaMDA, Microsoft's XiaoIce, or Apple's Siri, they all provide emotional value to people's lives and bring convenient services [17][18]. As a result, people have higher expectations for intelligent dialogue systems, and researchers continue to invest a lot of time in research to improve and upgrade them, thus forming a virtuous circle that promotes production development. Although intelligent conversational systems, supported by current research, have met the needs of many parties, today's conversational systems still suffer from problems such as inaccurate semantic understanding resulting in inaccurate answers, difficulties in gaining user trust due to inconsistencies between the identity and personality displayed in the conversation, and possible moral and ethical risks in conversational interactions. Therefore, next-generation dialogue systems should be knowledgeable, personal, and emotional [19].

4.1. Knowledgeable (IQ)

The IQ of a dialog system is primarily about the ability to get more information out of the user's response than just the output of a generic sentence. For example, in Figure 4, if a user says, "I have asthma since three years old," "the response: It's good for you to avoid triggers" is more attractive than simply "It's good for you to avoid triggers" would be more appealing than simply answering "Asthma is a disease"[19]. In Jaehun Jung et al.'s paper, they propose the concept of AttnIO(Attention Inflow and Out-flow), a new conversational conditional path traversal model that leverages the rich structural information in knowledge grounding based on two attention flow directions[20]. It simply means that the system is able to associate the extracted keywords with the semantic network associated with them(i.e., knowledge graph) for knowledge-aware encoding, and the attention flow for decoding[19]. This model can not only use knowledge graphs for inference but can also be used as a reference in

DOI: 10.54254/2753-8818/2/20220175

relation to the content of historical conversations. Therefore, after the AttnIO model is processed, it can understand the user's speech more and make it more reasonable and finer response.

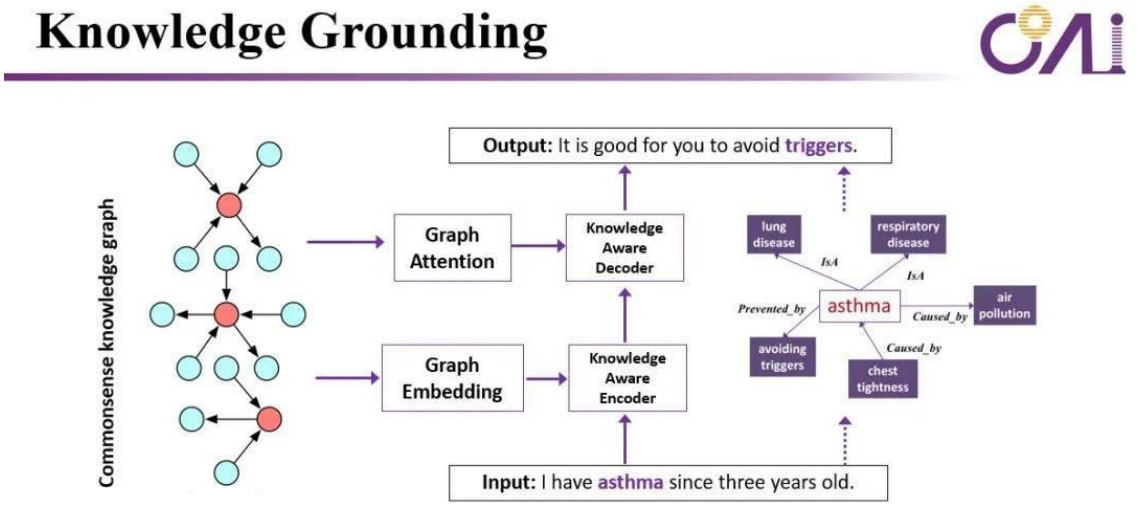

Figure 4. Knowledge Grounding [19].

4.2. Personality

High-quality dialogue activities need to win the trust of the other party, and having a fixed, consistent personality and identity is one of the key factors. A lack of consistent identity and personality during the conversation can make it difficult for the system to gain the trust of the user during the conversation and therefore make effective social interaction difficult. Nowadays, intelligent dialogue systems are usually pre-trained with dialogues of various personalities in the corpus, so there will be a lack of personality consistency. Personalizing an intelligent dialogue system can usually set a fixed personality for the dialogue system or match the user's personality by dialogue prediction. Zheng et al. used character sparse dialogues for pre-training the dialogue generation model and controlled the number of character-related features during the inference process, which would be good to produce coherent persona-consistent responses conditioned on explicitly represented personas [21].

4.3. Emotional (EQ)

Emotion is a state that integrates human feelings, thoughts, and behaviors and plays an important role in dialogue. Therefore, how to achieve emotion recognition as well as emotion expression in intelligent dialogue systems is a very important research topic nowadays. An effective emotion recognition algorithm is necessary to make intelligent dialogue systems more emotional and to obtain more fluent conversations. Microsoft's XiaoIce then uses an empathy computation module that can encode the user's sentiment and state by querying the empathy vector and specifying the response to the sentiment vector[18]. This empathy module includes features such as contextual query understanding, user emotion understanding, and personality response of XiaoIce characters. In the present day, it seems that XiaoIce's empathy framework is already very effective for responsive dialogue systems, and future intelligent dialogue systems may break through in emotional spontaneity.

5. Conclusion

This paper focuses on the history as well as the development of intelligent dialogue systems and highlights several frameworks for designing intelligent dialogue systems. Finally, it discusses how to improve and enhance intelligent dialogue systems. The next generation of intelligent dialogue systems should be rich in knowledge and human emotions to meet human social, informational, and emotional needs. Theories and technologies of dialogue systems are becoming more mature, and research on neural networks, machine learning, affective computing frameworks, empathy models, etc. is the way to

DOI: 10.54254/2753-8818/2/20220175

achieve these. This paper only introduces and discusses intelligent dialogue systems based on responding to text, but the future of intelligent dialogue systems is complex and multi-disciplinary. With the recent rise of the concept of the metaverse, the intelligent dialogue system of the distant future should be multisensory. Humans communicate with the world through multiple senses such as touch, smell, sight, hearing, and taste. In normal human conversations, other senses are involved in addition to text comprehension and response. Therefore, after improving the knowledge and emotion of intelligent dialogue systems, future research on intelligent dialogue systems will not be limited to text response.

Acknowledgment

I am very grateful to my teachers for their help and guidance in writing my thesis. Thank you to all friends and family for your encouragement and support. Without all their inspiring guidance and impressive kindness, I would not have been able to complete my dissertation.

References

[1] Nicola Bleu. (2021) 29 Top Chatbot Statistics For 2022: Usage, Demographics, Trends. Blogging Wizard. Retrieved June 30, 2022 from https://bloggingwizard.com/chatbot-statistics/

[2] Ciarán Daly. (2018) KLM: Chatbots Are The Future Of Customer Support. AI Business. Retrieved June 30, 2022 from https://aibusiness.com/document.asp?doc_id=760517

[3] Turing, A. M. (2012) Computing machinery and intelligence (1950). The Essential Turing: the Ideas That Gave Birth to the Computer Age, 433-464.

[4] WEIZENBAUM J. (1983) ELIZA-a computer program for the study of natural language communication between man and machine[J]. Communications of the ACM，26(1): 23-28.

[5] Bayan AbuShawar and Eric Atwell. (2015) ALICE Chatbot: Trials and Outputs. Computación y Sistemas 19, 4. DOI:https://doi.org/10.13053/cys-19-4-2326

[6] Cahn, J. (2017) CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.

[7] Shum, H. Y., He, X. D., & Li, D. (2018) From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, 19(1), 10-26.

[8] Karpagavalli, S., & Chandra, E. (2016) A review on automatic speech recognition architecture and approaches. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(4), 393-404.

[9] Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman. (2011) Natural language processing: an introduction, Journal of the American Medical Informatics Association, Volume 18, Issue 5, Pages 544–551

[10] Cambria, E., & White, B. (2014) Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine, 9(2), 48-57.

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017) Attention is all you need. Advances in neural information processing systems, 30.

[12] Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019) Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 7, 100943-100953.

[13] Jia Xibin, Li Rang, Hu Changjian, Chen Juncheng. (2017) A Review of Research on Intelligent Dialogue Systems.

[14] Gao, J., Galley, M., & Li, L. (2019) Neural approaches to conversational AI: Question answering, task-oriented dialogues and social chatbots. Now Foundations and Trends.

[15] Serban, I., Sordoni, A., Bengio, Y., Courville, A., & Pineau, J. (2016) Building end-to-end dialogue systems using generative hierarchical neural network models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).

[16] Sutskever, I., Vinyals, O., & Le, Q. V. (2014) Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.

[17] Eli Collins. (2021) LaMDA: our breakthrough conversation technology. Google. Retrieved July 3, 2022 from https://blog.google/technology/ai/lamda/

DOI: 10.54254/2753-8818/2/20220175

[18] Li Zhou, Jianfeng Gao, Di Li, Heung-Yeung Shum. (2020) The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. Computational Linguistics 46 (1): 53–93.
[19] Jiqizhixin (2021) How long is there to go for the next generation of intelligent dialogue systems that speak as naturally and fluently as people?. Retrieved July 3, 2022 from https://www.jiqizhixin.com/articles/2021-04-27-6
[20] Jaehun Jung, Bokyung Son, and Sungwon Lyu. (2020) AttnIO: Knowledge Graph Exploration with In-and-Out Attention Flow for Knowledge-Grounded Dialogue. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP), pp. 3484–3497, Online. Association for Computational Linguistics.
[21] Zheng, Y., Zhang, R., Huang, M., & Mao, X. (2020, April) A pre-training based personalized dialogue generation model with persona-sparse data. In Proceedings of the AAAI Conference on Artificial Intelligence.Vol. 34, No. 05, pp. 9693-9700.