Intelligent Question and Answer and Dialogue System

Xin Ma

Beijing Institute of Technology, Beijing 100081, China.

1120211358@bit.edu.cn

Abstract. This article discusses the development and application of question answering systems in the context of artificial intelligence and natural language processing. It highlights the importance of these systems in resolving user queries and providing accurate and relevant answers. This paper highlights the need to design question answering systems according to established rules and guidelines and use machine learning techniques to improve their performance. It also mentions the emergence of advanced question-answering systems such as IBM Watson, which leverage natural language processing and machine learning to provide more complex answers.

Keywords: Intelligent question and answer system; Machine learning; Knowledge graph; Deep learning; Sentiment value analysis.

1. Introduction

With the rapid development of modern society, the advancement of science and technology and the rapid circulation of information, people's lives are filled with various new high-tech products and services. In this process, the magnitude of data has also increased exponentially, so big data has become a topic of great concern. With such a huge amount of data, users usually have two main methods to obtain the required information: search technology and question and answer systems (Xiang Cheng et al., 2017).

Search technology is a technique used to find and retrieve relevant information in large-scale document collections or the Internet. It can find documents related to the query based on the query provided by the user and return them in order of relevance. This technology helps users quickly find the information they need through keyword matching, inverted indexing, relevance sorting and other methods. However, search technology relies more on keyword and document-level matching, which may lead to some irrelevant or inaccurate search results.

Different from search technology, the question-and-answer system can understand the semantics and intent of the question based on the specific questions raised by the user, extract relevant information from structured or unstructured knowledge sources, and finally give direct and accurate answers. Through technologies such as semantic understanding, knowledge extraction and reasoning, the question and answer system can more accurately understand the user's questions and provide more specific and in-depth answers.In addition, the question and answer system can also be customized and trained for specific fields, allowing it to provide more professional and field-specific answers. At the same time, with the continuous development of science and technology and the improvement of people's needs, the question-and-answer system will face more challenges and opportunities.

1.1. Q&A System Definition

The Q&A system is designed to understand the semantics and intent of user questions. Through natural language processing and semantic understanding technology, users can get the data and answers they want in a simple and easy-to-understand way without having to browse a large number of documents or perform complex search operations (Kushal Kafle, Christopher Kanan, 2017).

The question-and-answer system uses human-computer interaction to provide accurate and specific answers in natural language based on questions raised by users. The question-and-answer system can also extract information from structured or unstructured knowledge sources by combining natural language processing, information retrieval, knowledge representation and reasoning technologies, and transform it into data that users can read and understand.

1.2. Q&A system processing flow

The main process of the question-and-answer system is the system receives questions raised by users and performs preliminary analysis. In this step, the system will analyze the syntax and semantics of the question and identify the keywords, entities and relationships in the question to determine the type and intent of the question. Depending on the type and intent of the question, the Q&A system will retrieve information related to the question from relevant knowledge sources. The system can also use pre-built knowledge bases to obtain relevant knowledge. Based on the existing retrieved data, the system needs to extract specific answers from the retrieved data. This process will be applied to syntax analysis, named entity recognition and other technologies. Finally, the system will organize the extracted information into a complete and accurate answer. and respond in natural language.

2. The development of Q&A system

The intelligent question answering system was first born in 1950, when the British mathematician A.M. Turing (1950) proposed the famous Turing test in his paper "Computing Machinery and Intelligence" to test whether the machine has the ability to think.With the development of artificial intelligence, two more famous question and answer systems have emerged, namely BASEBALL born in 1961 and LUNAR born in 1973. The basic working principle of these systems is to save existing data in the database. When the user asks a question, the user's question is converted into an SQL statement, the relevant data is queried from the database, and the results are returned to the user.

In the 1970s, the question-and-answer system entered the period of computational linguistics. The representative system of this period was Unix Consultant. At this stage, the question-and-answer system mainly deals with and is limited to use in specific fields.

In the 1990s, with the rapid development of the Internet and the massive growth of document data, the question-and-answer system entered a new stage, that is, the open domain and text-based period. In particular, the establishment of the QA (Question Answering) competition has greatly promoted the development of question answering systems.

With the maturity and popularization of Internet technology, new data sources have emerged, such as frequently asked questions (FAQ) data. With the continuous emergence of online communities, frequently asked questions data has grown rapidly, providing a large amount of question-and-answer pair data for the question-and-answer system, thus enabling the question-and-answer system to further develop and enter the period of open fields and question-answer pairs.

2.1. Rule-based Q&A system

A rule-based question answering system is a system that answers questions through predetermined rules or logic. (Lehnert WG, 1977) The implementation steps of this question and answer system: first, use the question understanding component to convert the user's questions from natural language into a form that can be understood by the machine, and then use the rule matching mechanism to compare the converted questions with the data in the knowledge base After successful matching, the system generates answers based on existing data in the knowledge base or based on pre-established logic, and finally converts the answers into natural language and outputs them to the user.

Rule-based question answering systems can be traced back to the early stages of artificial intelligence. During this period, artificial intelligence researchers began to explore how to achieve dialogue with machines and complete the Turing test. One of the landmark events was the ELIZA program developed by American computer scientist Joseph Weizenbaum in 1963. ELIZA simulated the identity of a psychiatrist to complete a conversation with a patient. ELIZA parsed the user's input through a series of logical rules and pattern matching. , and generate corresponding answers (Heung-yeung Shum et al., 2018). Although the ELIZA program is not a rule-based question and answer system in the strict sense, the emergence of this program has greatly promoted the development of question-and-answer systems.

As time goes by, today's rule-based question and answer systems continue to improve and develop. Today's more famous rule-based question and answer system is IBM Watson: The working principle of IBM Watson is mainly divided into three parts, namely: natural language processing and machine learning, and advanced reasoning (Eliza Strickland, 2019; Ellen Riloff and Michael Thelen, 2000). The natural language processing part is mainly used to analyze unstructured data such as files, emails, conversations, etc.; machine learning allows the system to improve its analysis and understanding capabilities by analyzing data sets; advanced reasoning is a decision-making function based on available information. , greatly reducing the time to answer users (Sandeep A. Thorat, Vishakha Jadhav, 2020), Professor James Hendler also mentioned that IBM Watson is built based on "associated knowledge", and its implementation process is shown in Figure 1 below.

With the continuous development of artificial intelligence, the shortcomings of rule-based question and answer systems have gradually become apparent. The operation of this kind of question-and-answer system relies on pre-defined rules and templates, so a huge database is needed to store a large number of rules and templates to cover various questions. This storage requirement limits the scalability and flexibility of the system. In addition, the limitations of rules also lead to difficulties in improving the accuracy of question-and-answer systems. In order to ensure the accuracy of answers, the system needs to seek experts in relevant fields to write rules and templates. This limits the scope of application of the system, since rules can only be adequately covered in areas covered by experts. Due to data storage limitations, question answering systems can only answer questions in a small number of fields. Finally, due to its inherent design method, the rule-based question answering system also makes it difficult to develop into different fields. Due to the limitations of rules, when faced with new problems or areas, new rules need to be manually added to expand the coverage of the system. Such development and maintenance costs are high, and it cannot respond to knowledge updates and changes in a timely manner.

2.2. Q&A system based on statistical machines

The statistical machine-based question answering system is a type of question answering system built using statistical machine learning and natural language processing technology (Rohit Arora et al., 2021). Different from rule-based question answering systems, statistics-based question answering systems learn statistical patterns and correlations between questions and answers by analyzing large-scale corpus data (Paulina Adamska and Marta Ju´zwin, 2015). To build a statistics-based Q&A system, you first need to collect data from databases such as the Internet for training; during the training phase, the system will use feature value processing methods to convert the collected questions and answers into a form that can be understood by the computer; the system will use the processed The model is trained on questions and answers, and model parameters are learned to maximize the match between questions and answers. After completing the training of the model, it enters the testing phase. At this time, the user asks a question, and the system matches the similarity between the question and the model, then calculates the similarity between each question and the candidate answer, and then obtains the best answer based on the similarity ranking. Answer (Yihan Bian and Kaiwen Peng, 2021).

The development of question answering systems based on statistical machines is inseparable from the advent of the data era. After 2000, with the explosive growth of data and the development of statistical machine learning methods, researchers began to use statistical models to establish the relationship between questions and answers. mapping relationship (Poonam Gupta and Vishal Gupta, 2012). As question answering systems are applied in different fields, statistics-based question answering systems gradually evolve into field-specific question answering systems. A typical representative is Microsoft MRC, which is built by Microsoft based on a statistics-based question and answer system that focuses on machine reading comprehension tasks (Arnab Sinha, 2015). Training is performed using large-scale reading materials and corresponding question-answer pairs to improve the ability to find answers in reading materials.

Statistical machine-based question answering systems have more advantages than rule-based question answering systems. First, the question answering system based on statistical machines can break through the limitations of rules and realize automatic learning of the semantic and contextual relationships between questions and answers (Wen Xu et al., 2006). This gives the system the ability to self-adjust and optimize, significantly improving system accuracy and performance. Secondly, the question answering system based on statistical machines is no longer constrained by rules, so it has better scalability and continuous improvement capabilities. To apply this question answering system to other professional fields, you only need to provide data in the corresponding field for training, and then combine the experience and knowledge of domain experts for model optimization. This flexibility enables the system to be expanded and applied in different fields and application scenarios. In addition, statistical machine-based question answering systems can achieve continuous improvement by increasing training data and optimizing models. Over time, the system can learn from user feedback and reviews, improving accuracy and performance.

2.3. Q&A system based on knowledge graph

The question-and-answer system based on knowledge graph is a question-and-answer system that utilizes knowledge graph technology. The difference between it and the question-and-answer system based on statistical machines is the way of data representation and the acquisition and expression of knowledge. In the question-and-answer system based on statistical machines, the relationship between questions and answers is mainly expressed through statistical models, while in the question-and-answer system based on knowledge graphs Data in the question-and-answer system is represented by a graph structure, where the nodes and edges in the graph represent entities, attributes and relationships respectively. Secondly, the question-and-answer system based on knowledge graphs and the question-and-answer system based on statistical machines are also different in the way of knowledge acquisition and expression. The question-and-answer system based on statistical machines obtains knowledge through statistics and natural language processing; matching is performed through statistical models. and predictions to express knowledge. The knowledge graph relies on building and maintaining the knowledge graph, reasoning and inference through entities, attributes and relationships in the graph, and using the knowledge in the graph for linking and reasoning to provide more accurate and comprehensive answers (Y.Wu et al. .,2020).

Question and answer systems based on knowledge graphs are becoming a promising research direction, and the more popular research now is to retrieve answers from existing large-scale knowledge graphs (S. Hu et al., 2018; R. Wang et al., 2019; W. Zheng et al., 2019). Previous research on knowledge graphs was mostly based on extracting intent from words and segmenting phrase semantics, which led to deviations in problem understanding. Therefore, in the article "Community answer generation based on knowledge graph", the way the question-and-answer system generates answers is improved (Yongliang Wu and Shuliang Zhao, 2021). The methods of generating answers in question-and-answer systems are divided into two categories: selecting existing answers and generating new answers. The main difficulty today is how to ensure the accuracy of generating new answers. The method CAGKG given in this paper improves the accuracy of questions by generating answers for users based on their background. CAGKG's approach is to understand post semantics and model user context through phrases, and then use the learned graph representation to generate answers. It can obtain answer entities related to the question by performing graph traversal on the graph or utilizing path information in the graph. Then, the final answer can be generated further based on the attribute information or association relationship of the entity. The paper also proposes an index used by the PASSE method to evaluate semantic similarity. Experiments on four real-world datasets show that CAGKG outperforms state-of-the-art question answering frameworks. Compared with other answer evaluation metrics, PASSE is a promising metric for evaluating semantic similarity.

2.4. Q&A system based on deep learning

Question answering systems have been a topic of widespread academic concern in the past few decades, among which machine learning and deep learning are popular methods in the field. Compared with machine learning, deep learning has some limitations, including large training sample size, high training model system configuration, and long model construction time. On the other hand, deep learning models usually require a large number of training samples to achieve good performance. Therefore, in order to train deep learning models, high configuration of computing resources and long model construction time are required. Traditional machine learning algorithms can complete model training using a small number of training samples at a faster training speed. When performing classification tasks, traditional machine learning methods usually require manually labeling the correct categories for data samples in advance. However, these labeled data may face unavailability situations, such as incorrect labeling, incomplete labeling, or lack of labeled data. This situation will have a negative impact on the accuracy of the machine learning model, resulting in relatively low classification accuracy (Lu Liu and Jing Luo, 2018; Yashvardhan Sharma and Sahil Gupta, 2018; Bo Song et al., 2108). Although deep learning and traditional machine learning each have their own advantages and disadvantages, in recent years deep learning has become more and more popular in research on question answering systems.

Through research papers in statistics-related fields, it is found that the following methods are commonly used in question answering systems based on deep learning.

2.4.1 Convolutional Neural Network (CNN)

CNN is a deep learning model that is particularly suitable for processing data with a grid structure, such as images, videos, and audios (Kan Chen et al., 2015). How it works in the question-and-answer system: First, the input layer of CNN accepts text data as input and creates a matrix for each question and answer with the help of embedding technology. Then the convolutional layer of CNN uses N-GRAM filter to convolve the sentence matrix and extract relevant text word features from it. Then, the corresponding feature values are pooled in the pooling layer to reduce the spatial size of the feature map. Finally, feature mapping is performed in the fully connected layer, so that the neurons in the fully connected layer are connected to all neurons in the previous layer, and their weights will be learned through training (Dan Tian et al., 2023).

2.4.2 Long Short-Term Memory (LSTM)

LSTM is a special type of recurrent neural network. The LSTM model is able to remember long sequences, so it is particularly suitable for solving time series problems, such as stock market price prediction, weather forecast, etc. In recent years, it has also been used to solve question and answer questions. LSTM used in question-and-answer systems mainly has four components, namely forgetting gate, input gate, cell state update, and output gate (Yashvardhan Sharma and Sahil Gupta, 2019). First, the forgetting gate controls whether the state information in the cell state at the previous moment is retained or forgotten. It uses the input data and the hidden state at the previous moment as input, and outputs the degree of retention of the information in each cell state. The input gate calculates the importance of each input based on the input data and the hidden state of the previous moment. Then the cell state is updated based on the input of the input gate and the input data to complete the training of the question-and-answer system model. Finally, the output gate evaluates and outputs the updated candidate answers.

3. Q&A system face problems and challenges

An important difference between artificial intelligence and humans is the presence and ability to understand emotions. Emotions play an important role in human communication and understanding, so how to accurately understand and respond to users' emotions becomes an important challenge in the development of question-and-answer systems. An emerging sentiment analysis task now is ASTE (Aspect-based Sentiment Analysis). ASTE extracts three groups of emotions in an end-to-end manner

to identify and analyze topics in text and the emotional polarity associated with them. However, this method ignores the relationship between the three groups of elements and has the problem of error propagation. The latest ASTE research paper focuses on the relationship between three groups of elements and improves the extraction model of emotional triples.

3.1. Aspect-based Sentiment Analysis(ASTE)

Sentiment triples are a structure used to represent text emotional information, usually composed of three elements: theme (Aspect), sentiment polarity (Sentiment Polarity) and context (Context).

The goal of the ASTE task is to accurately identify and classify topics and emotional polarities in text in combination with the context of the topic, so as to deeply understand the user's emotional attitude towards a specific topic. The analysis of user emotions is mainly completed through the following steps: first, aspect extraction, which is a process of identifying aspects or entities related to emotions from the text; and then the emotion classification task: judging the aspect through classification methods such as Naive Bayes Whether the evaluation is positive, negative, or neutral; the third step is to use relationship extraction technology to extract triplet information from the text; finally, the ASTE method usually uses labeled data for model training.

3.2. Algorithm improvements

Traditional ASTE has many problems in the process of sentiment analysis and triple extraction, so improvements will be made to address related issues.

Aiming at the problem that the ASTE method ignores the mutual information between triple elements in the relationship extraction process, a semantic and syntactically enhanced emotional triple extraction model (S3E2) is proposed (Lu Xu et al., 2021). This model fully exploits the syntactic and semantic relationships between triplet elements and jointly extracts them. It leverages the graph sequence dual representation and modeling paradigm, combining graph neural networks (GNNs) for encoding semantic and syntactic relationships between word pairs in sentences and LSTMs for preserving order information. The feature information of triples cannot be fully utilized during the model training process, so this paper proposes an enhanced multi-channel graph convolution network model (EMC-GCN) to improve the sentiment analysis model ((Hao Chen et al., 2022), EMC - The GCN model utilizes the dual-quasi-attention module to embed the relationships between words as neighbor tensors. It treats the word and relationship neighbor tensors as nodes and edges respectively, converting sentences into multi-channel graphs. language features, including part-of-speech combinations, syntactic dependency types, tree-based distance and relative position distance, to enhance the EMC-GCN model.

4. Summary and Outlook

Question and answer systems are of great practical significance in various fields such as information retrieval, customer support, and knowledge management. These systems can help users quickly find relevant information and provide accurate answers, thereby improving the efficiency and accuracy of information search. Especially in large-scale document analysis and network search, question and answer systems can be applied to process and analyze massive data, helping users quickly locate specific content in massive information.

In the future, the question-and-answer system will further improve its accuracy and efficiency. Among them, multi-modal answering capabilities are crucial, making the question-and-answer system more practical and applicable. By combining non-text data such as images and videos with natural language questions, multi-modal question answering systems can achieve cross-modal intelligent question answering. This capability enables the system to more fully understand the user's questions and provide more accurate and detailed answers. For example, when users search for questions related to buildings, they can supplement the description of the question by uploading

photos or videos of buildings, and the Q&A system can leverage these multi-modal data to provide more accurate and specific answers.

Through multi-modal answering, the question-and-answer system can better meet the needs of users and provide more comprehensive and accurate information. Additionally, this feature helps address the limitations of traditional text-based Q&A systems in understanding and answering questions involving visual elements. With the continuous increase of image and video data, multi-modal question answering systems will play an important role in various fields.

References

[1] Arnab Sinha,Zhihong Shen,Yang Song, et al. An Overview of Microsoft Academic Service (MAS) and Applications. Proceedings of the 24th International Conference on World Wide Web. Florence Italy, 2015, p.243-261

[2] Bo Song,Yue Zhuo,Xiaomei Li. Research on Question-Answering System Based on Deep Learning. International Conference on Swarm Intelligence. Shanghai,2018, p.522-529

[3] Dan Tian,Mingchao Li,Qiubing Ren,et al. Intelligent question answering method for construction safety hazard knowledge based on deep semantic mining. Automation in Construction,2023, Volume 145

[4] E.Strickland. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE,2019. p. 24-31

[5] Kushal Kafle, Christopher Kanan.Visual question answering: Datasets, algorithms, and future challenges.Computer Vision and Image Understanding,2017,p.3-20

[6] Lu Liu,Jing Luo .A Question Answering System Based on Deep Learning. 14th International Conference. Wuhan, China,2018, p.173-181

[7] Hao Chen, Zepeng Zhai, Fangxiang Feng, et al.Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction.Association for Computational Linguistics，2022,p. 2974-2985

[8] R. Arora, P. Singh, H. Goyal, et al. Comparative Question Answering System based on Natural Language Processing and Machine Learning.2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). Coimbatore, India,2021

[9] Ruijie Wang,Meng Wang,Jun Liu.et al. Leveraging Knowledge Graph Embeddings for Natural Language Question Answering.24th International Conference.Chiang Mai, Thailand,2019,p. 659-675

[10] S. Hu, L. Zou, J. X. Yu.et al. Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs. IEEE,2018, p. 824-837

[11] Weiguo Zheng,Hong Cheng,Jeffrey Xu Yu.et al.Interactive natural language question answering over knowledge graphs.Information Sciences,2019,p. 141-159

[12] X. Cheng, S. Zhu, S. Su and G. Chen.A Multi-Objective Optimization Approach for Question Routing in Community Question Answering Services.IEEE,2017, p. 1779-1792.

[13] Y. Bian and K. Peng. Question Answering System Analysis Based on Machine Learning. IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Fuzhou, China, 2021

[14] Yashvardhan Sharma,Sahil Gupta .Deep Learning Approaches for Question Answering System.Procedia Computer Science Volume 132,2018,p. 785-794

[15] Yongliang Wu,Shuliang Zhao,Wenbin Li.Phrase2Vec: Phrase embedding based on parsing. Information Sciences Volume 517,2019, p. 100-127

[16] Yongliang Wu,Shuliang Zhao. Community answer generation based on knowledge graph. Information Sciences Volume 545,2021, p. 132-152