ISSN: 2710-0588

BERT-based Outpatient Process Intelligent Consulting Model for Hospitals

Qi Fang[1, a]

[1]Chongqing Three Gorges University, Chongqing, 404100, China

[a]fangqi11203@163.com

Abstract

This study designs and implements an intelligent consultation model for hospital outpatient processes based on the BERT model. We collected and preprocessed data from various hospital outpatient processes and fine-tuned the pretrained BERT model. The model provides customized outpatient process information and recommendations based on user-input queries. We developed a user-friendly interface allowing patients to easily access guidance on appointments, registration, examinations, and medication collection. Experimental results demonstrate significant improvements in accuracy and practicality, offering a new intelligent solution for hospital management and patient services.

Keywords

BERT model; hospital outpatient processes; intelligent consultation; health information technology.

1. Introduction

In the context of rapid advancements in artificial intelligence technology, China, as the most populous country in the world, has enormous potential for the development of the healthcare industry. In recent years, China has vigorously advocated for the intelligent and modernized development of healthcare, promoting greater integration and practical application of medical services with artificial intelligence.

According to data from the National Bureau of Statistics, by the end of 2023, the population aged 60 and above in China had reached 297 million, accounting for 21.1% of the total population. Among them, the population aged 65 and above was 217 million, making up 15.4% of the total population. This data shows a year-on-year upward trend, indicating that the degree of aging in China is deepening. Aging not only exacerbates the issue of elderly care, but also generates tremendous demand for healthcare services. Due to factors such as physical degeneration, the high prevalence of chronic diseases, and limited understanding of medical knowledge, the elderly face numerous obstacles during the medical process. Therefore, finding ways to use technology to alleviate the difficulties elderly people face in accessing healthcare has become an urgent social issue that needs to be addressed.

Against this backdrop, many elderly people encounter difficulties when seeking medical consultations at hospitals. To address this issue, it is necessary to develop an intelligent medical knowledge question-answering model based on the BERT model to help solve the problem of elderly patients having difficulty with consultations in large hospitals.

Abroad, medical consultation systems based on natural language processing (NLP) and deep learning have been widely applied. Many hospitals and healthcare platforms have introduced intelligent customer service systems that use machine learning and NLP technologies to answer common patient questions in real-time. The application of these systems has not only improved the efficiency of medical services but also alleviated the problem of insufficient healthcare

ISSN: 2710-0588

resources to some extent. Especially for elderly patients, intelligent question-answering systems can provide real-time medical information support, reducing the burden on doctors.

Research in this field started relatively late in China. Domestic scholars have made some progress in the study of intelligent medical consultation systems, but most of the research has focused on online consultations related to patients' conditions. [1]Currently, domestic medical consultation systems still have relatively weak research and practical applications when it comes to the specific issues elderly patients face during in-person visits to hospitals. For example, elderly patients often ask seemingly trivial questions during in-person consultations, which can add pressure to busy large hospitals and contribute to increased consultation demands.

Therefore, this study aims to develop an intelligent medical consultation system based on the BERT model, focusing on addressing the challenges elderly patients face during hospital consultations. The system will provide personalized medical consultation services, helping elderly patients reduce the difficulty of seeking medical care and improving their healthcare experience.

This study not only has strong theoretical value, filling the gap in domestic research on medical consultation systems for the elderly, but also holds high practical value. Through the application of this intelligent consultation system, it can effectively alleviate issues such as information asymmetry and long waiting times that elderly patients face during medical visits, improve the efficiency of hospital operations, and promote the development of medical intelligence in a more widespread direction.

2. Data Source and Processing

The dataset is derived from the daily outpatient consultation records in hospitals. Specifically, it includes common question-and-answer dialogues between patients and doctors, nurses, guides, and other hospital staff during their interactions throughout 2016. The dataset was manually screened by hospital staff and mainly consists of the 14,000 most frequent outpatient consultation dialogues, covering various common topics such as registration, appointments, examinations, medication collection, and medical record inquiries.

2.1. Diversity and Representativeness of the Data

The collected data covers a wide range of scenarios that patients may encounter during their outpatient visits. Based on the hospital's outpatient volume and patient distribution, the data is highly representative. The questions in the dataset are diverse, including common inquiries about registration information, departmental guidance, appointment information, medication collection instructions, and treatment processes for common diseases. This variety ensures that all stages of the outpatient process are covered. Furthermore, since the data volume is large and collected from different departments' actual outpatient processes, it accurately reflects the real needs and question-answering patterns of patients.

2.2. Data Quality and Reliability

The quality of the dataset is high for the following reasons:

Reliable Source: The data comes from the hospital's actual outpatient processes and has been manually selected and verified by hospital staff, ensuring its authenticity and accuracy.

Data Cleaning and Preprocessing: The data has been cleaned and standardized during the collection process, removing irrelevant noise such as incorrect inputs or invalid dialogues.

High-Frequency Consultation Phrases: The dataset focuses on the most common consultation phrases from 2016, making it highly representative and capable of reflecting the core needs of outpatient consultations (as shown in Figure 1).

ISSN: 2710-0588

Figure 1. Database Screenshot from a Certain Hospital

2.3. Data Preprocessing and Cleaning

In order to ensure the quality of the data and make it suitable for fine-tuning the BERT-based model, multiple preprocessing and cleaning steps were applied to the collected outpatient consultation data. The goal of data preprocessing is to improve the data's consistency and effectiveness, and ensure its applicability in model training. The specific preprocessing steps are as follows:

2.3.1. Removal of Irrelevant Information and Noisy Data

Some irrelevant information in the dataset, such as duplicate sentences, meaningless dialogues, spelling errors, or grammatical mistakes, were removed. In real hospital consultation records, patient inquiries may sometimes contain spelling errors or unclear expressions due to accent or speech issues. Such information needs to be cleaned and corrected.

2.3.2. Text Normalization

To ensure consistency and standardization in the sentences, the text was normalized.

2.3.3. Data Labeling and Alignment

Question and Answer Alignment: In each data entry, the patient’s question and the hospital's standard answer were accurately aligned. Each question and its corresponding answer were divided into pairs of inputs and outputs. During model training, the input is the patient's question, and the output is the system's response.

Labeling Intent and Entities: For some complex queries, intent labels and entities were annotated. For example, when a patient asks, “Can I schedule a chest X-ray today?”, the labels might include the intent “Schedule Examination” and the entity “Chest X-ray”.

2.3.4. Handling Data Imbalance

In real hospital outpatient data, certain categories of questions may be imbalanced, such as more queries from certain departments compared to others. To avoid bias during model training, the following methods were applied:

Oversampling/Undersampling: For certain rare types of inquiries, oversampling was applied to increase their representation. For categories with an excessive number of queries, undersampling was used to ensure balance across different question categories.

Data Augmentation: To improve the model's generalization ability, data augmentation techniques were used, such as synonym replacement and word substitution, to increase the diversity of the training data.

ISSN: 2710-0588

2.3.5. De-identification Processing

Since the data used comes from real hospital outpatient records and contains sensitive patient information (such as names and contact details), de-identification was performed on all data before use. This ensures that the data does not contain any information that could directly identify the patients.

3. BERT Model Establishment and Adjustment

3.1. BERT Model

The BERT (Bidirectional Encoder Representations from Transformers) model is a significant innovation in natural language processing (NLP) technology, released by Google AI in 2018. Its core breakthrough lies in the introduction of a bidirectional encoding mechanism, which allows the model to understand contextual information from both the left and right sides of a sentence simultaneously. [2] This bidirectional feature overcomes the limitations of traditional unidirectional language models (such as left-to-right or right-to-left models) in understanding context, significantly improving the model's ability to comprehend language in a more holistic manner.

The training process of BERT is divided into two main stages: pre-training and fine-tuning. In the pre-training stage, BERT uses two primary tasks to learn language representations: the Masked Language Model (MLM) and Next Sentence Prediction (NSP). The Masked Language Model randomly masks some words in a sentence and asks the model to predict these masked words. This method helps the model understand relationships between words and context. The Next Sentence Prediction task requires the model to determine whether the second sentence is a natural continuation of the first sentence, enhancing the model’s understanding of relationships between sentences.

3.2. BERT Model Training

3.2.1. BERT Model Establishment

Model Initialization: The BERT model is initialized by loading a pre-trained BERT model from the Hugging Face Transformers library (such as bert-base-uncased or bert-base-chinese).

Model Fine-Tuning: Fine-tuning is performed using the preprocessed dataset. By training on the question-and-answer data, the BERT model's parameters are adjusted so that it can provide accurate answers to common questions in the hospital outpatient process.

3.2.2. Training Strategy

Learning Rate Adjustment: A smaller learning rate is used to ensure smooth fine-tuning and avoid over-adjusting the pre-trained weights.

Batch Size: The batch size of 16 was selected as optimal for the experiment.

Epochs: The number of epochs is set to 5. This is adjusted based on the model's performance and overfitting conditions.

Optimization Method: The Adam optimizer is used to update the model's parameters. Combined with learning rate decay and early stopping strategies, this approach helps prevent overfitting and ensures the model's generalization ability during the training process.

ISSN: 2710-0588

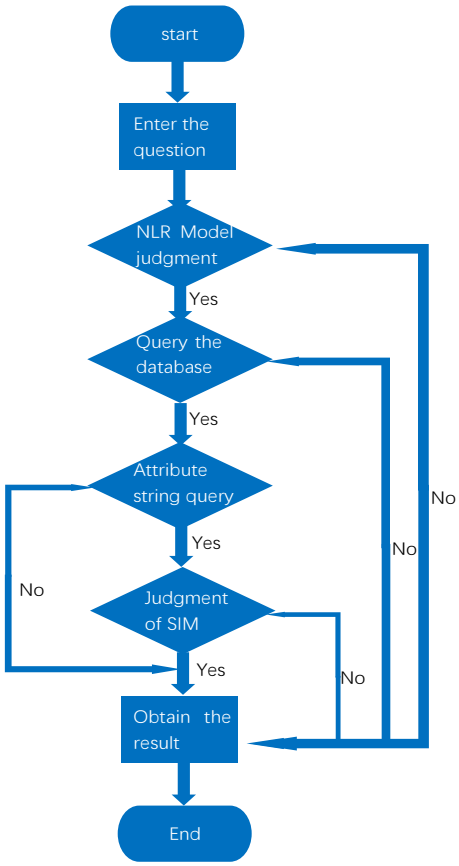

Figure 2. Training Principles of the BERT Model

After fine-tuning, the model's performance needs to be validated through experimental evaluation. Standard question-answering evaluation metrics, such as accuracy, were used, and the model performed well (as shown in Figure 3).

```
# test
# loss 0.0333103257204748
# question_acc 0.9319302392344928
#label_acc  0.9884675308641976

# dev
# loss 0.0234971376811594
# question_acc 0.9558620689655173
#label_acc  0.9932816475936721

# train
# loss 0.0168365984823491
# question_acc 0.9748273051047039
#label_acc  0.9923957448974632
```

Figure 3. Test Results of the Dataset and Test Set

4. Technical Challenges and Solutions

In applying the BERT-based intelligent consultation model to the hospital outpatient process, several technical challenges were encountered. These challenges spanned data, model, and system implementation aspects. [3] To ensure the system runs efficiently and provides accurate

ISSN: 2710-0588

consultation services, targeted solutions were implemented. Below are the main technical challenges and their corresponding solutions.

4.1. Handling Medical Terminology

Challenge: The medical field is rich in specialized terms, abbreviations, and complex language expressions, many of which are not adequately addressed in the general-purpose pre-trained BERT model. For example, patients may mention complex medical terms, disease names, or examination items, and the BERT model might struggle to accurately understand these terms. [4]

Solution:Domain-Specific Fine-Tuning: During the fine-tuning process, the BERT model was trained on real medical question-and-answer data provided by the hospital, especially data containing a large number of specialized terms. This domain-specific training allowed the model to better understand common expressions in the healthcare industry. [5]

4.2. Handling the Diversity and Ambiguity of User Queries

Challenge: User queries can be highly diverse, especially in outpatient consultations, where patients' questions are often informal and contain a lot of ambiguous or incomplete sentences. This presents a challenge for natural language understanding models, which must comprehend the meanings behind various ways of expressing the same query.

Solution:Data Augmentation: To improve the model's generalization ability, data augmentation techniques were used, such as synonym replacement and sentence rephrasing. This expanded the training dataset, enabling the model to recognize a wider variety of query formulations.

Contextual Understanding: The BERT model’s bidirectional encoding feature ensures that the model can understand patient queries based on contextual information, especially in multi-turn conversations, allowing it to recognize shifts in user intent.

4.3. System Response under High Concurrency

Challenge: The number of patients in outpatient departments can be extremely large, particularly during peak hours. If the system cannot effectively handle a high volume of concurrent requests, it may result in response delays, which would affect user experience and system stability.

Solution:Caching Mechanism: A caching mechanism was introduced for frequently asked questions and their answers. By caching high-frequency queries and answers in memory, the system can quickly respond to user requests without needing to perform model inference each time, improving response speed.

5. Data Privacy and Compliance

Challenge: Since the medical data used involves sensitive patient information, ensuring the privacy and compliance of the data is a critical issue. Without proper privacy protection measures, the system could violate relevant laws and regulations, damaging patient trust.

Solution:De-identification and Encryption: All medical data was subject to strict de-identification processing before being collected and used. Patients’ personal information (such as names, ID numbers, etc.) was removed or encrypted to ensure that the data could not be traced back to any specific individual. [6]

6. Model Inference Efficiency

6.1. Challenge

Although BERT performs exceptionally well in natural language processing tasks, its inference speed is relatively slow. When processing large-scale data and handling high-concurrency

ISSN: 2710-0588

requests, long response times can occur. To meet the hospital outpatient consultation system's real-time and efficiency requirements, the model's inference efficiency needed to be improved.

Solution:Quantization and Pruning: To accelerate inference speed, the model underwent quantization (e.g., converting floating-point weights into low-precision integers) and pruning (removing unnecessary neural network connections). These methods reduced the computational load and improved processing efficiency, making the model more suitable for real-time queries and high-concurrency environments.

6.2. Results and Validation

After fine-tuning, the model's performance was evaluated through multiple rounds of testing and assessments to ensure it met the expected objectives. Techniques such as cross-validation were used to assess the model's adaptability across various scenarios, ensuring that it could provide accurate and real-time intelligent consultation services in actual hospital outpatient settings.

6.3. Conclusion

In this study, we designed and implemented an intelligent consultation system based on the BERT model for hospital outpatient processes, aimed at improving hospital management efficiency and patient service quality. By combining the pre-trained BERT model with domain-specific outpatient data, we successfully built an intelligent system capable of accurately understanding and responding to patient inquiries. This system covers various outpatient service aspects, such as appointments, registration, examinations, and medication collection, providing personalized consultation services to patients and effectively supporting hospital administrators in optimizing and improving outpatient workflows.

Although certain achievements were made, there is still room for improvement in this study. Future research could focus on further optimizing model performance, especially in handling long-tail problems, complex queries, and multi-turn conversations. Moreover, with the advancement of technology, both model inference efficiency and data privacy protection measures can be further enhanced.

Overall, the BERT-based intelligent consultation model proposed in this study offers a new intelligent solution for hospital outpatient processes. As artificial intelligence and natural language processing technologies continue to develop, intelligent question-answering systems will become an essential component of hospital services and management, driving the healthcare industry towards a more intelligent, efficient, and personalized future.

References

[1] Chen, Y., & Chen, S. (2019). Smart healthcare services for elderly patients: A review. International Journal of Medical Informatics, 128, 38-48.

[2] Liu, Y., Wang, J., & Chen, K. (2020). A Review of the BERT Model and Its Applications in Natural Language Processing. Computer Engineering and Applications, 56(22), 1-10.

[3] Zhou, C., Wang, X., & Tang, T. (2021). Design and Challenges of Medical Intelligent Question Answering Systems.

[4] Chen, X., Li, L., & Zhang, Q. (2019). Application of Natural Language Processing in Medical Intelligent Question Answering Systems. Modern Computer, 29(11), 48-55.

[5] Zhang, X., Gao, M., & Xing, Y. (2021). Research on BERT-Based Medical Text Processing Methods. Computer Science and Exploration, 15(6), 1025-1036.

Information and Control, 50(5), 629-635.

[6] Wang, M., Xu, C., & Chen, L. (2021). Research on Medical Data Privacy Protection Technologies. Computer Engineering and Applications, 57(7), 75-82.