A New Model for Question-Answer based Dialogue System for Indian Railways in Hindi Language

Lovely Sharma¹*, Vijay Dhir² and Kamaljeet Kaur³

¹Punjab Technical University, Kapurthala Road, Jalandhar – 144601, Punjab, India; lvlysharma.85@gmail.com
²St. Soldier Institute of Engineering and Technology, Jalandhar – 144622, Punjab, India
³Department of Computer Science, D. A. V. College, Amritsar – 143001, Punjab, India

Abstract

Objective: Currently, the user is rely on the results of the Google search i.e. lengthy documents, web links etc. So, Question Answering (QA) dialogue system becomes more vital to access the succinct answers in a jiffy. Methods: The main focus is on Query Processing based on partial semantic and syntactic analysis and keyword based approach. The major module is Dialogue Manager which is responsible to communicate between the User and the System to resolve anaphora and Co-reference problems. Findings: The existing Indian Railway system is based on clicking system which is more time consuming. Thus, proposed QA dialogue system provides specific and concise answers to users. The results are encouraging and a motivation for further research in this field. The proposed system restricted to provide answers of related queries but further extension of the system can also include replies related arrival time, departure time over the phone etc.

Keywords: Dialogue based Question-Answering System, Dialogue Manager, Semantic Analysis

1. Introduction

Question Answering (QA) Dialogue system¹ is a technique of Natural Language Processing (NLP)². Users often have specific queries in their mind, for which they expect concise, comprehensible and accurate answers from the web in their native language without being restricted to a specified query language, query formulation rules, linguistic knowledge etc. while search engine provides ranked list of web links, documents etc. So, the need of QA system becomes more vital.

The existing website of Indian Railway i.e. www.indianrail.gov.in does not support Question-Answer dialogue system and query system. So, there is a need to develop a Question-Answer based dialogue system in Hindi which will help the users to search about Indian railways, while sitting at home. Its database is available online and is also one of the India’s largest Online Database.

Dialogue System (DS) in any computer system is that which interacts with a human using natural language like English, Hindi, and Punjabi etc. The QA dialogue system is proposed in Hindi language for the domain of Indian Railway. This system is implemented on a relatively restricted domain i.e. Jalandhar district is selected according to convenience as well as major junctions of North zone of Punjab State. The list of all the trains which passes through and having stoppage at Jalandhar city or Cantt will be generated and it will be tried. Thus, the domain is:

- List of trains between two stations, but constraint is to input only those trains which have stoppage at Jalandhar.
- User can ask about fare of that particular train as mentioned in step (1).

There are approximately 250 trains having stoppage at Jalandhar. System is considering only Mail/Express/Superfast, DMUs, Passengers trains, Shatabdi, Garibrath and Jan-Shatabdi. User can ask following types of questions such as:

a) स्वर्ण शताब्दी ट्रेन का क्या कोड है?
Swaran shatabadi train ka kya code hai?

*Author for correspondence
2. Review of Literature

The early systems are very restricted to their nature such as Green et al. conducted a study on BASEBALL and found that it provides answers about US baseball league of one year. The classical work has been done by Eliza. It works as chatter bots and it naturally communicates between human and computer. Similarly, Stephen Wolfram has studied Wolfram Alpha based on the computational platform Mathematica. Additionally, Quarc is a rule based system that reads the comprehensions and provides the answers from the story, yet it attains 40% accuracy to their experiments.

In addition to this, Denecke et al. introduces instance based approach based on memory based learning approach which again following the case-based reasoning. Recently, Belinkov et al. proposed a model using linear support vector machine that attains better results as compared to competition winner.

3. System Architecture

There are three models of DS such as Frame based model, Finite model and Agent based model. The requirement is to use both the models as Frame based and Finite based model i.e. Hybrid Model. In Finite Model, the User follows the predetermined steps/stages while in Frame based model, the System posed the queries to the User in accordance to fill slots in a frame to accomplish the task. If requirement is not fulfilled to fetch the accurate answer from database then system will enter into dialogue with the User.

Moreover, the foremost concern of Railway QA architecture is to design the Railway database and the Knowledge base. The other various modules of QA architecture are Input decoder (Section 1), Query Frame Analyzer (Section 2), Dialogue Manager (DM) (Section 3), Domain Specific Component (SQL Generation) (Section 4) and Response Generation (Section 5). These systems are described in subsequent sections. The proposed three major types of databases are as follows:

1. Railway Database: It provides the information about trains between two stations and fare of related trains.

2. Knowledge base: It stores the knowledge related to language understanding.

In addition to this, dialogue system will accept the query string posed by user in Hindi language by two methods. One of them is using Hindi keyboard and other is Hindi Transliteration. The next language understanding module analyse the real text with the help of Translation database and search for the keyword in the query. It converts a sequence of words into a semantic representation using Knowledge base. e.g.

"सरयु यमुना एक्सप्रेस में मेरा कितना किराया लगेगा?"

In the above example, the keyword is किराया. On the basis of keywords and tokens the appropriate Query Frame is chosen and each Query Frame has slot filling format. According to the foregoing example:

"सरयु_यमुना_एक्सप्रेस_किराया (ट्रेन_नाम/कोड, चलना, पहुंचना, आयु_श्रेणी)"

At this stage, two cases arise. A). Slot is complete. B). Slot is not complete.

Case A:

According to first case, if slot is complete as shown in following example:

**English Version**

Saryu Yamuna expresses main Amritsar se ambala tak 30 saal ke vayakti ka sleeper shreni main kitna kiraya lagega?

Saryu_yamuna_Express.kiraya(Train_Name/Code,source,destination,age,class)

**Hindi Version**

सरयु यमुना एक्सप्रेस में अमृतसर से अमला तक 30 साल के व्यक्ति का स्लीपर श्रेणी में कितना किराया लगेगा?

सरयु_यमुना_एक्सप्रेस.किराया (ट्रेन_नाम/कोड, चलना, पहुंचना, आयु_श्रेणी)

As specified above, if the slot is complete then corresponding SQL query will be triggered on the Railway database and concise and accurate answer will provide to the User in a simplest manner.

2 Conversion of a text from one script to another. In our proposed system conversion is from English to Hindi.
Case B:

According to second case, if slot-filling is incomplete, then system will enter into dialogue with the User with the help of Dialogue Manager and system is supposed to ask the queries from the user in accordance to fill the slots. Dialogue manager will manages all the aspects of dialogue and figures out how the text will fit into the overall context. After the completion of Query Frame SQL query will be triggered on the Railway database and succinct answer will present to the User in an interactive manner.

Demo

The step by step procedure processed by Dialogue Manager is as follows:

1. रेलवे.सरयु_यमुना_एक्स. (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)
   मैं: 14650

2. रेलवे.सरयु_यमुना_एक्स. (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)
   मैं: अमृतसर

3. रेलवे.सरयु_यमुना_एक्स. (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)
   मैं: अम्बाला

4. रेलवे.सरयु_यमुना_एक्स. (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)
   मैं: 30

5. रेलवे.सरयु_यमुना_एक्स. (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)
   मैं: स्लीपर

The diagram as shown in Figure 1 and example (Section 3.1-3.5) will help us to understand above methodology:

3.2 Query Frame Analyzer

Converts a sequence of words into a Semantic Representation will be used by the DM. It involves the use of Railway Database and Morphology, Syntax and Semantics. e.g.

सरयु_यमुना_एक्स, किराया
रेलवे. सरयु_यमुना_एक्स (ट्रेन नाम/ कोड, चलना, पहुंचना, आयु, श्रेणी)

3.3 Dialogue Manager

It manages all the aspects of the dialogue between the User and the System and takes a semantic representation of the user's text and figures out how the text fits in the overall context and creates a semantic representation of the System response.

3.4 Domain Specific Components (SQL)

The dialogue manager usually needs to interface with some external software such as a database or an expert system. Thus, the query or plans have to be converted from the internal representation used by the dialogue manager to the format used by the external domain specific system (e.g. SQL). This interfacing is handled by the domain specific components using Railway Database. e.g. select (columns name, column name.) from table_name where (Condition). Select fare from train where (14650, Amritsar, Amritsar, 30, sleeper).

The performance of the application mainly depends on the live practical systems and the different approaches are available to recognize performance.

3.5 Response Generation

It involves constructing the message that is to be shown to the user and requires the syntactic structure and
relatively simple way to represent the answer. According to foregoing example:

\[ \text{e.g. सिस्टम: 230 रुपए} \]

4. Conclusion and Future work

On the basis of above facts, it is said that the proposed Railway QA architecture is practical in use by the End user. The pipeline of various modules is interacting with each other in such a way that the accurate and brief answers are provided to the User. In future, the real text stated in natural language can be converted into speech input and system can provide the information over the phone using speech recognition system.

5. Acknowledgement

One of our authors, Lovely Sharma wants to thanks to Punjab Technical University (PTU), Kapurthala, India for their help and support.

6. References