VINBRAIN

Exploring Diseases Dataset for Chatbot Dialogue Management

Tai Ng. (Applied Scientist)

April 10, 2021

VINBRAIN INTERNSHIP PROGRAM 2021

Characteristics

A task-oriented dialogue system is developed to perform a clearly defined task. Usually, the task involves finding information within a database and returning it to the user, performing an action, or retrieving information from its users.

Technologies

Usually, the user's input is processed by a natural language understanding (NLU) unit, which extracts the slots and their values from the utterance and identifies corresponding the dialogue act. This information is passed to the dialogue state tracker (DST), which infers the current state of the dialogue. Finally the output of the dialogue manager is passed to a natural language generation (NLG) component.

Dialogue Management (DM)

The DM could be connected to some external Knowledge Base (KB) or Data Base (DB), such that it can produce more meaningful answers.

The Dialogue Manager consists the following two components:

- The Dialogue State Tracker
- Policy Learning

Approach Dialogue Management

- Rule-based: Finite state machine
- Learning-based: Deep Q-Learning

Evaluation

Two main aspects are evaluated, which have been shown to define the quality of the dialogue:

- Task-success
- Dialogue efficiency

Diseases Dataset

A dataset to provide the students a source to create a healthcare related system.

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5
0	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic _patches	NaN
1	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches	NaN	NaN
2	Fungal infection	itching	nodal_skin_eruptions	dischromic _patches	NaN	NaN
3	Fungal infection	itching	skin_rash	dischromic _patches	NaN	NaN
4	Fungal infection	itching	skin_rash	nodal_skin_eruptions	NaN	NaN
...	...	...	...	...	...	...

Figure 1: Diseases Dataset

Insight

Investigate the correlation between each other diseases over the symptoms.

- Construct the chatbot script (usecase)
- Avoid waste of time to query KB/DB
- Define semantic frame for response task

Database

```
/* 1 */
{
    "_id" : ObjectId("606c0a179dadc241387f678c"),
    "Disease" : "(vertigo) Paroymsal  Positional Vertigo",
    "Symptom" : [
        "spinning_movements",
        "unsteadiness",
        "loss_of_balance",
        "vomiting",
        "headache",
        "nausea"
    ]
}

/* 2 */
{
    "_id" : ObjectId("606c0a179dadc241387f678d"),
    "Disease" : "Alcoholic hepatitis",
    "Symptom" : [
        "distention_of_abdomen",
        "history_of_alcohol_consumption",
```

Figure 2: Disease database

Dialogue tracking

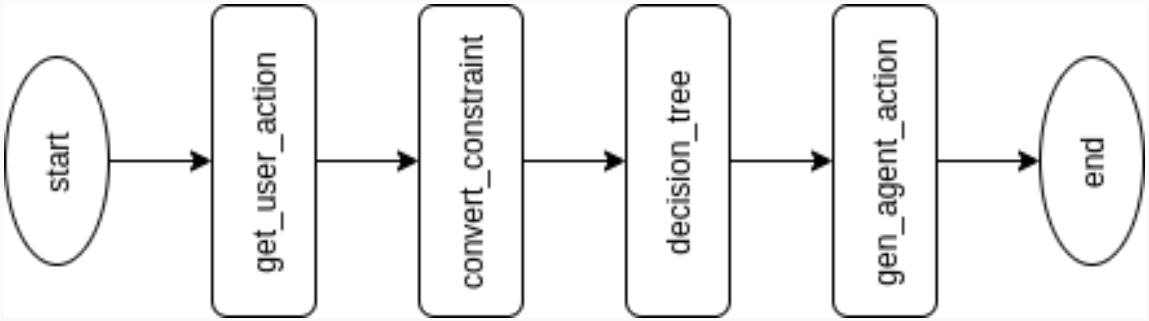

Figure 3: Flow dialogue management

Agent's action

1. Inform & Request: Randomly, amount of records greater than one.
2. Match found: Only one record satisfy user's constraints.
3. Done: End conversation if cannot found disease.

API

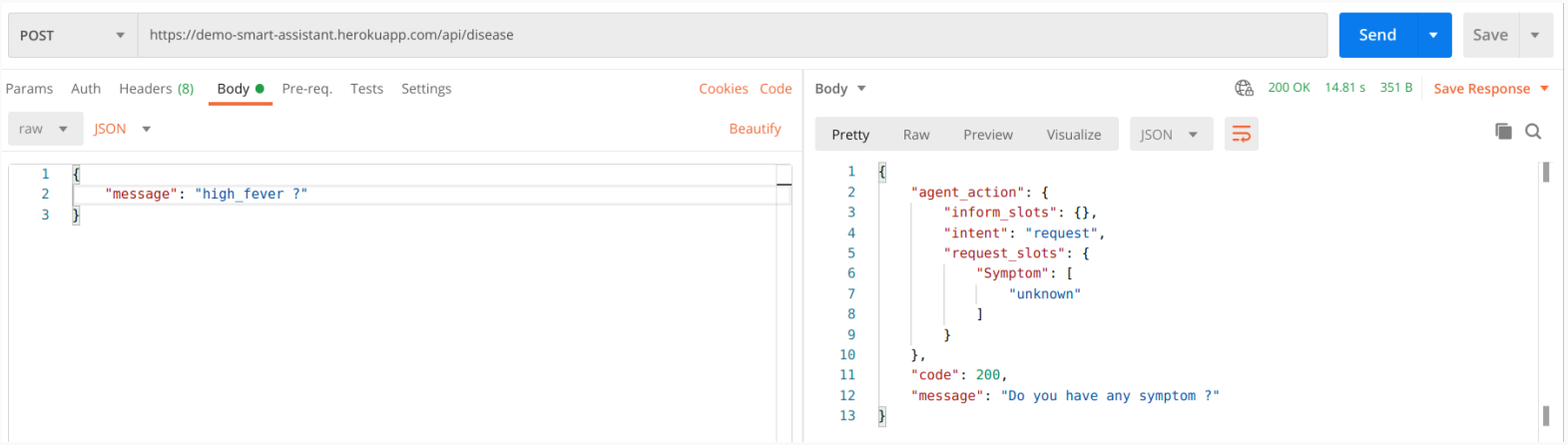

Figure 4: API

Demo

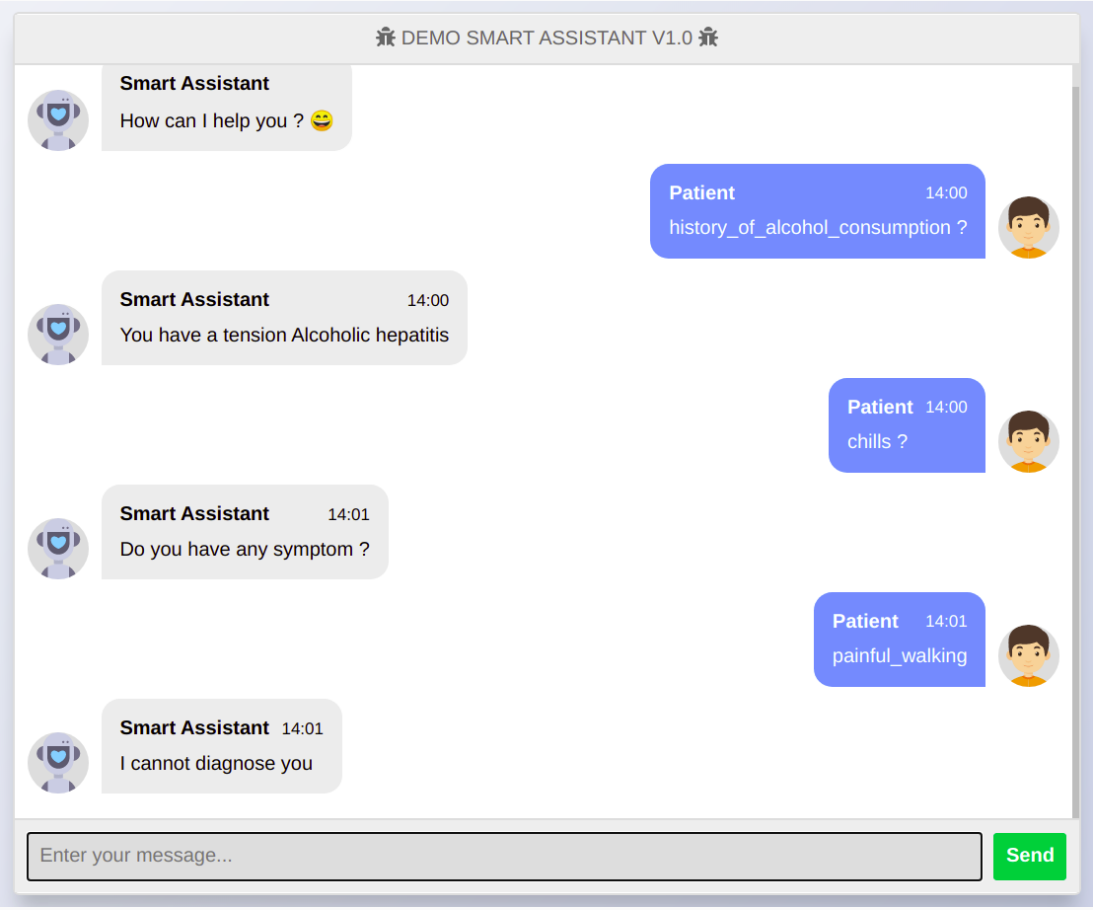

Figure 5: Demo conversation

Conclusion

Q&A