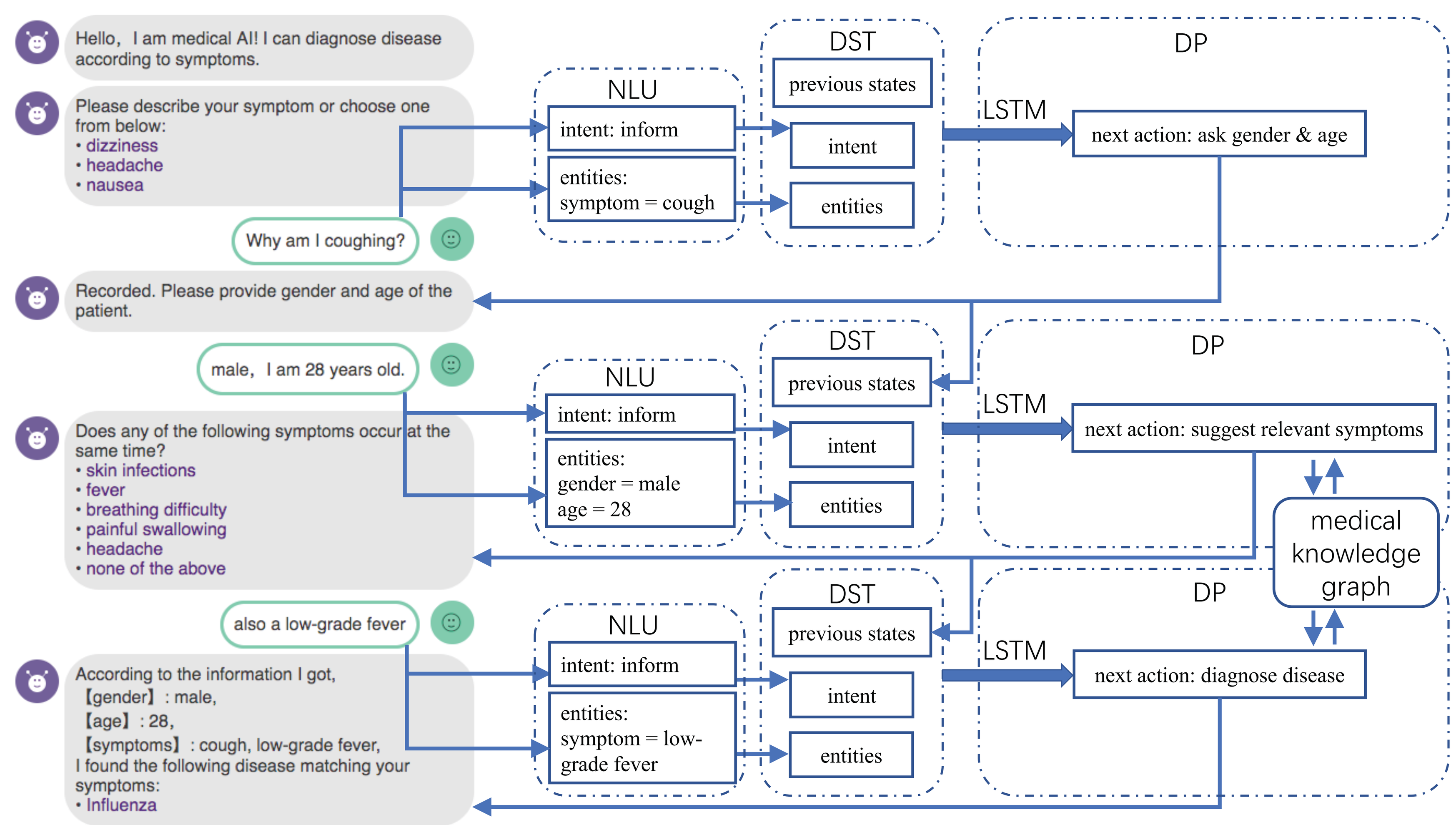
A Knowledge Graph Based Health Assistant

Lin Bo, Wenjuan Luo, Zang Li, Xiaoqing Yang, Han Zhang, Daxin Zheng
AI Labs, Didi Chuxing

KNOWLEDGE-BASED DIALOGUE SYSTEM

We introduce a medical knowledge-based dialogue system that acts as a health assistant that can help people analyze their medical symptoms and pay more attention to their health. With well-designed algorithms, it efficiently matches symptoms against its medical knowledge base and recommend a suitable diagnosis.



As illustrated above, the architecture of the medical knowledge-based dialogue system consists of three components:

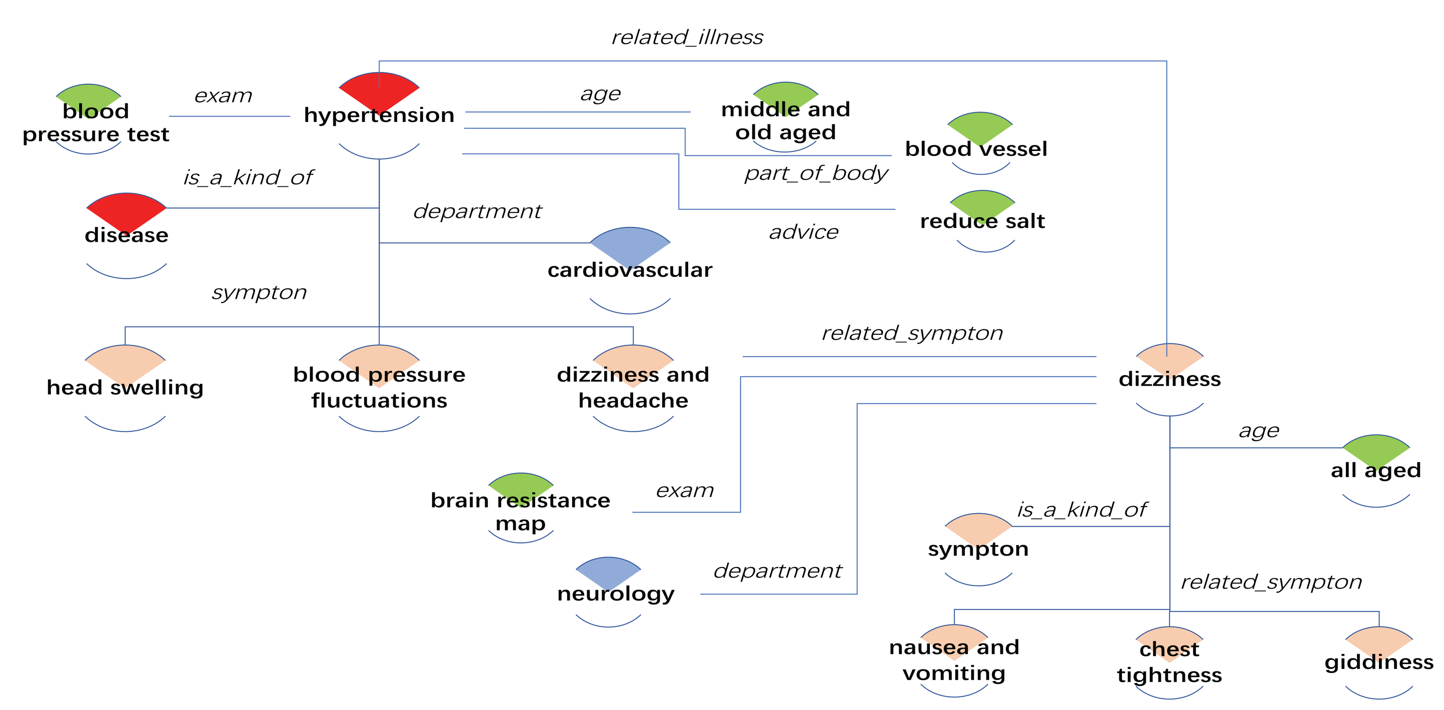
Natural Language Understanding (NLU):
This module analyzes the user's text-input and extracts useful information. Specifically it can be divided into two parts: an intent classifier and an entity extractor.

Dialogue state tracking (DST):
This module memorizes the intent and entities extracted at each dialogue turn during current conversation.

Dialogue policy (DP):
This module determines the next action of the chatbot and sends a proper response to the user. The next chatbot action is determined by long short-term memory (LSTM) network with sequence of historical dialogue states and actions as the model input. Moreover, equipped with medical knowledge graph, the chatbot system is able to identify symptoms that frequently co-occur with the ones reported by the user. In the situation where there is not sufficient symptoms information to determine a reliable diagnosis, the chatbot will ask about relevant symptoms in the next dialogue turn to clarify user's health condition.

MEDICAL KNOWLEDGE GRAPH CONSTRUCTION

The construction of a knowledge graph requires multiple triplets in the form of (S, P, O), where S stands for Subject, P stands for Predicate, and O stands for Object. The technologies involved include web crawlers, data cleansing, structured information extraction, entity identification and relation classification. Specifically, we selectively construct a medical knowledge base that includes common diseases and symptoms of drivers and corresponding medical examinations and diet recommendations.



Left figure shows a subgraph of our knowledge base. Red nodes indicate diseases and pink nodes stand for symptoms. Blue nodes denote the diseases/symptoms' medical specialties, and green nodes represent other attributes of the entities.

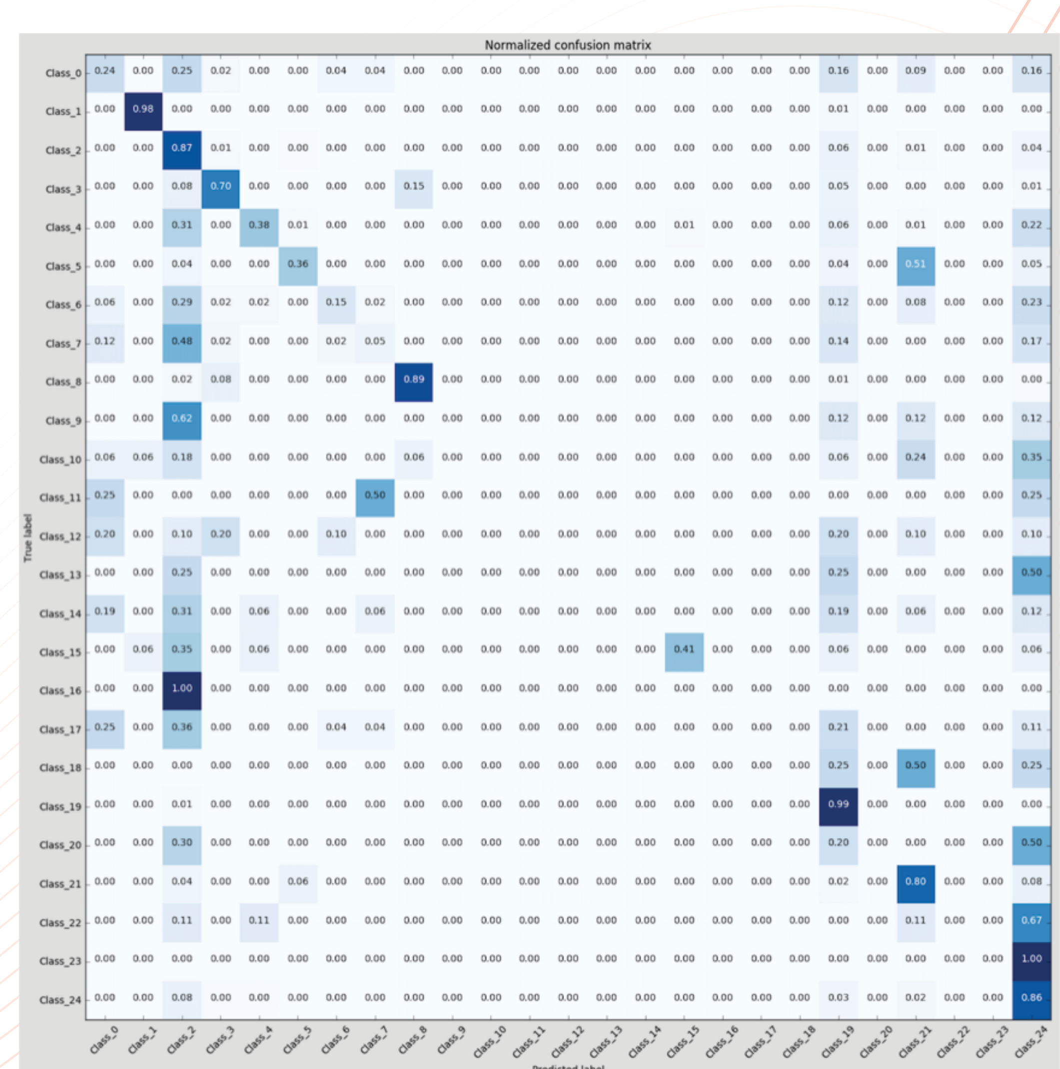
As some attributes of entities are missing, especially the disease prevention and treatment advice, we use search engine to get information in order to fill these attributes. In detail, we specify the entity as the keyword, and obtain related corpora. After data cleansing, we perform entity identification and relation classification to generate SPO candidates.

EXPERIMENTS

For relation classification, we design our model based on a bidirectional GRU with attention (BIGRU+ATT) and XGBOOST, where BIGRU with attention is used to train sentence embedding with respect to relation. XGBOOST utilizes the aforementioned sentence embedding and pre-trained entity embedding for final relation classification. On the right we show results for relation classification and the confusion matrix.

Table 1: Results for Relation Classification

Method	avergePrecision	averageRecall	averageF1
BIGRU+ATT	0.6478	0.7790	0.7074
Ours	**0.8820**	**0.8806**	**0.8813**



QUESTION ANSWERING OVER KNOWLEDGEG BASE

In our system, question answering mainly includes three steps: problem understanding, query construction and answer calibration. Firstly, problem understanding consists of entity identification and predicate recognition. Secondly, the query construction is the process of translating user question into query clause in the knowledge system. For example, suppose the back-end knowledge base is built on Elastic Search (ES), the system needs to rewrite the user question into a ES query statement. Finally, answer calibration refers to re-ranking the relevant results and generating final answers.

CONCLUSION

The health assistant chatbot has been deployed on DiDi's mobile application and WeChat official accounts platform since 2018 and has served hundreds of thousands of users who are ride-share drivers. This healthcare chatbot, under the premise of enhancing health awareness of the drivers, also demonstrates the potential of artificial intelligence technologies in making ride-share driver a safe and respected profession.