Intelligent (Task-Oriented) Conversation Assistant for Course Selection

Progress Report



Information Technology Capstone Project

COMP5703

Group Members

1. Shengyuan Sun (460257820)
2. Rui Chen (470111585)
3. Quan Chen (470199228)
4. Biying Wang (470067400)

Table of Contents

[Table of Contents i](#_Toc521413584)

[1. Progress Status 1](#_Toc521413585)

[2. Roles & Responsibilities 2](#_Toc521413586)

[3. Individual Achievements 2](#_Toc521413587)

[4. Group Collaboration 2](#_Toc521413588)

# Progress Status

|  |  |
| --- | --- |
| **Project Name** | CS17 Intelligence(Task-Oriented) Conversation assistant for course selection |
| **Project Start Date** | 6/ 3/ 2019 |
| **Project Manager** | Biying Wang |

|  |  |
| --- | --- |
| Project Description | The project focuses on using NLP and other technologies to build a dialog system to analyse and answer the questions that students ask in terms of course units on University of Sydney education system |

|  |  |  |
| --- | --- | --- |
| Project Status Report | # | Date: 22/ 3/ 2019 |

|  |  |  |
| --- | --- | --- |
| **Status Item** | **Status up to last week** | **Planned for this week** |
| **Major deliverables** | Initial project architecture  Initial Requirement list | Detailed project architecture  Added word emebedding into the intent classification mode |
| **Planned delivery date** | 13/ 3/ 2019 | 21/ 3/ 2019 |
| **Major issues** | Identify all interface and structure of the project  Understanding the special requirement of the project | Design a plan for whole part of work  Understand the input requerment and out requirement of intentclassficication |
| **Major risks** | Misunderstand the requirement  Designed an unachievable requirement | The accuracy and precision of the network mode.  Some part of the project should depend on the other part of work, the consisant should well defined. |
| **External dependencies** | Slack Github interview | Slack Github Zoom |
| **Estimated effort (hr)** | 20hr | 20hr |
| **Recorded effort (hr)** | 40hr | 20hr |
| **Status (R, Y, G)** | yellow | red |

# Roles & Responsibilities

Shengyuan Sun

Role: developer

Responsibility:

* system architecture design
* intent classification method implemention
* word emebedding implemention
* design the plan and work dependency
* analysis the intent and slot structure of ATIS dataset

Rui Chen

Roles: Analyser and Developer

Responsibility:

* Continously test online corpus (MIT-Moive corpus)
* Reading the rule-based chatbot paper and analyse intent-slot model
* Analysing and Design the top-level architecture of cassandra system, and discussing them with team members
* Design the working flow graph for further co-working

Biying Wang

Role: Project manager; Data collector

Responsibility:

* Check whether weekly deliverables is done by each member on time; make sure project progessing is successful
* Write meeting minutes as a recording
* Check meeting time, location and topic for group
* Collect information for dataset designing
* Design dataset structure
* Help and discuss with other team member for project work

Quan Chen

Role: Developer

Responsibility:

* Back-end database system building and data collecting
* Help to design the whole process of system (such as slot and intent)

# Individual Achievements

In this week, My job mainly force on the intent classification and word embedding, based on the last week work, we decide to use the Bi-Lstm and attention mode for natural language understanding, apart from the pre-processing part, we have test two dataset, ATIS dataset with result in 98% in test set, but when we use the model in another MIT-movie corpus dataset, the result only 85% after 50 episodes. The model has not the word embedding part, just give each word an index. Then I assume the word vector maybe can solve the problem.

I demo two word2vec method and one FastText method for word embedding all method provide by Gensim.

A screenshot of a social media post

Description automatically generatedA screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

The second work of this week is implemented the intent classification, based on the work in the last week the result of the MIT- movie corpus in the belowA screenshot of a social media post

Description automatically generated

The result of the MIT-movie corpus is much lower than the result in ATIS corpus. The reason could be the word embedding or the model itself, In the next week I will check it out.

One thing we are unclear is the dialog management part, because of there are some different idea in our group, so in the most simple way, we consider the initial version of our project should be like an information retrieval which only have one asking sentence in the last week. But in this week after the meeting we change our scope and requirement into dialog management, then we design another structure, the plan of our workflow also changed, I design a new plan of the implement.A close up of a piece of paper

Description automatically generated

# Group Collaboration

In this week, our group divided into two tasks, Ray and I work on the intent classification part, Biying and Quan work on the CUSP analysis and database design, so in this week, Ray drawing whole project structure, and I design the work plan, the main conversation between group is about the details of the structure, the rule-based system, which we also asking David have an extra meeting with us.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated