A Spoken Drug Prescription Dataset in French for Spoken Language Understanding

Ali Can Kocabiyikoglu1,2, François Portet1, Prudence Gibert2, Hervé Blanchon1, Jean-Marc Babouchkine3, Gaëtan Gavazzi4

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France, {ali-can.kocabiyikoglu, francois.portet, herve.blanchon}@univ-grenoble-alpes.fr
2. CHU Grenoble Alpes, Avenue Maujou-du-Gésièvre, 38700 La Tronche, France, pgibert@chu-grenoble.fr
3. Calystene SA, 16 Rue Irène Joliot Curie, 38320 Eybens, France, jm.babouchkine@calystene.com
4. Clinique de médecine gériatrique, CHU Grenoble Alpes, Équipe Grépi, EA 7408, CS 10217, 38700, La Tronche, France, ggavazzi@chu-grenoble.fr

Our dataset is available at: https://doi.org/10.5281/zenodo.6425586

1. Introduction

We present PxSLU corpus a drug prescription dataset comprising around 4h of transcribed and annotated dialogues of drug prescriptions in French acquired through an experiment with 55 participants experts and non-experts in prescriptions.

- Corpus acquired from human-machine interactions using a prototype of a goal-oriented dialogue system
- Publicly available through an Attribution 4.0 International (CC BY-4.0) license.

2. Data Collection Protocol

We have established a simple protocol, inviting the participants to follow the following steps:

1. Registration on the form for requesting to participate in the experiment
2. Receptions of the mobile application and follow the document explaining the installation and the course of the experiment
3. Receptions of prescription examples depending on the audience
4. Filling the metadata survey, agree to the terms of use, and complete the experiment

3. Corpus Overview

The experiment was performed between January 2021 and October 2021. The data represents 262 minutes of recordings when all the participants are included.

4. Data Preparation

We targeted two type of participants:

- Medical practitioners (doctors, pharmacists, biomedical engineers,etc.)
  - 10 pictographs
  - 10 reading exercises
- Non-experts
  - 20 reading exercises

5. Distribution of participants

6. NLU Model Evaluation

Our model evaluation builds on previous work (Kocabiyikoglu et al., 2019) where we have presented initial NLU systems trained on artificial and textbook data:

- Baseline CRF model
- Triangular CRF extension (Jeong et Lee, 2008)
- Bi-RNN with attention (Liu et Lane, 2016)
- Flaubert pre-trained transformer LM (Le et al., 2019)

<table>
<thead>
<tr>
<th>Model</th>
<th>Intent (acc)</th>
<th>Micro Avg</th>
<th>Macro Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF</td>
<td>92%</td>
<td>0.81</td>
<td>0.80</td>
</tr>
<tr>
<td>Tri-CRF</td>
<td>91%</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>Alt-CRF</td>
<td>93%</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td>Flaubert</td>
<td>94%</td>
<td>0.89</td>
<td>0.91</td>
</tr>
</tbody>
</table>

- Flaubert model gives the best results on both micro and macro averages
- Other models have comparable performance

7. K-Fold (K=5) Cross Validation on PxSLU

Fine-tuning Flaubert model using PxSLU (flaubert-base-cased)

- Similar performances with model trained with artificial+textbook data
- Data choice impacts the macro performance and the coverage of slots

8. Collected Dialogue Analysis

- Histogram of average time of session by number of participants
- Most of the participants completed the task in less than 1 minutes
- Medical practitioners on average less than 40 seconds
- Doctors have spent more time on dialogue sessions (avg 40-100 seconds)

References