Gathering Information and Engaging the User ComBot: A Task-Based, Serendipitous Dialog Model for Patient-Doctor Interactions

Anna Liednikova[&,#], Philippe Jolivet[&], Alexandre Durand-Salmon[&], Claire Gardent[†]

& ALIAE, # Université de Lorraine, † CNRS/LORIA

Introduction

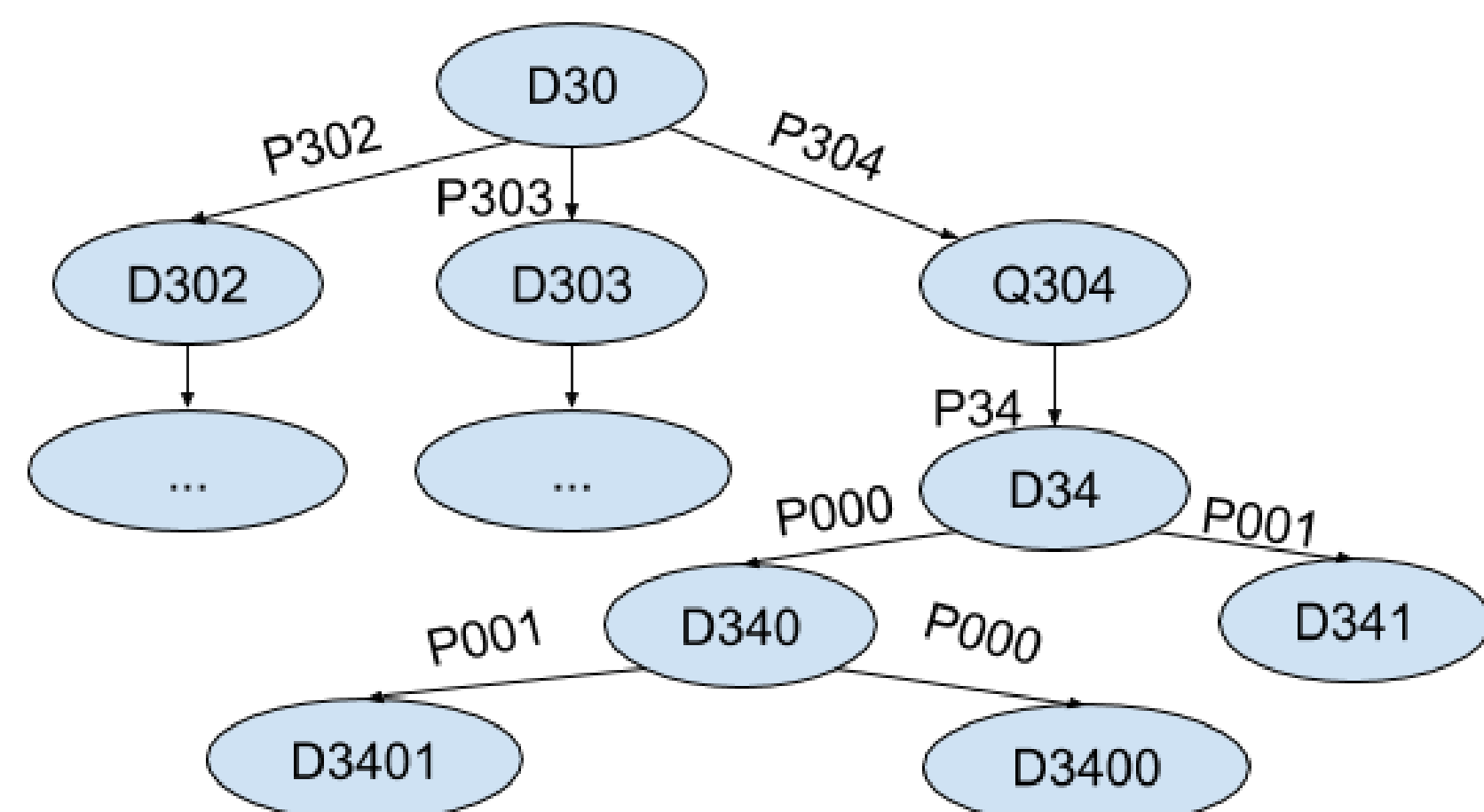**Context**: chronic pain patients follow-up

Goal: dialog system to collect information in more flexible and non-trivial way

Contribution: an ensemble model ComBot; a generic method to create a Follow-Up bot data; a detailed analysis of the three bots interaction

Medical Bot

Goal: starts the conversation and ask questions from medical tree as soon as the user mentions relevant problem

A zero-shot retrieval model based on ConveRT embeddings [Henderson et al.2019] and expert doctor-patient dialog tree. During the conversation we choose the most similar context from the dataset to our last three turns and output the corresponding question.

<D30> <tree_pos> What is the most difficult for you about your sleep ?
<P304> <tree_pos> I wake up early in the morning
<D304> <tree_pos> What time did you wake up today ?
<P34> <tree_pos> I woke up at 5 am
<D34> <tree_pos> Do you wake up spontaneously but without enough sleep to feel rested all day long ?
<P000> <tree_pos> no
<D340> <tree_pos> Could you describe how do you wake up ?
<P001> <tree_pos> yes, the alarm clock wakes me up
<D3401> <tree_pos> Thanks to these informations, your physician will have more information at his disposal for your next visit

Figure 1: Fragment of dialog tree and a corresponding dialog

Future Research

- Dealing with negation
- Right time to stop
- Taking care about emotional balance
- Giving advice

Follow-up Bot

Goal: retrieve health-related questions which naturally keeps up with the user's topic in order to get more health information and come back to Medical bot

A zero-shot retrieval model based on ConveRT embeddings [Henderson et al.2019] and HealthBoard dataset, which consists of 3181 context-question pairs with length less than 100 characters and similarity between context and question higher than 0.6.

To select context-question pairs, we classify sentences in the thread with dialog act classifier and selected only assertions and questions.

During the conversation we select the candidate by comparing last three turns of the dialog with context in the dataset, we select top-20 contexts and corresponding question and with MMR and similarity score make sure that that are not repetitive to the questions already asked in the dialog.

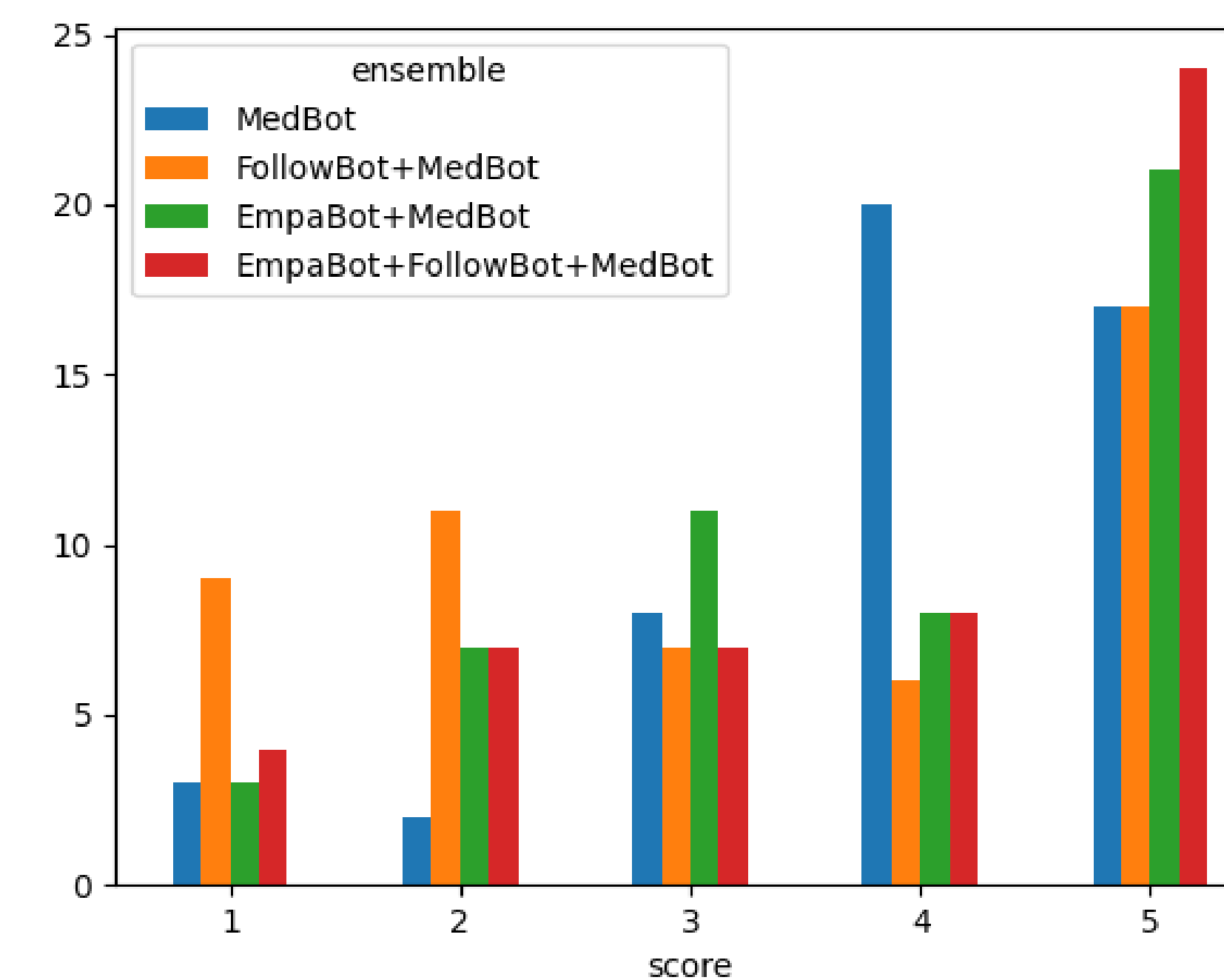| context	question
I can't fall asleep at all.	Are your sleep issues anxiety related?
I have nightmares every night too, and I don't know what to do.	If you have a reoccuring theme in your nightmares, it might help to come to terms with that phobia?
I don't think I slept at all last night.	Maybe you could take a nice nap?
I cannot sleep cause of pain 24/7.	Have you found any relief from any strong pain relievers the will let you sleep at night?

Table 1: Follow-Up Bot context-question pairs examples

References

[Henderson et al.2019] Matthew Henderson, Iñigo Casanueva, Nikola Mrkšić, Pei-Hao Su, Tsung-Hsien Wen, and Ivan Vulić. 2019. Convert: Efficient and accurate conversational representations from transformers.

[Li et al.2019] Margaret Li, Jason Weston, and Stephen Roller. 2019. Acute-eval: Improved dialogue evaluation with optimized questions and multi-turn comparisons. *arXiv preprint arXiv:1909.03087*.

[Roller et al.2020] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston. 2020. Recipes for building an open-domain chatbot.

Evaluation

For each system we collected 50 dialogs with AMT, asking workers to rate the system in the end with the satisfaction score. Later we compared MedBot and ComBot with Acute-Eval [Li et al.2019] setting.
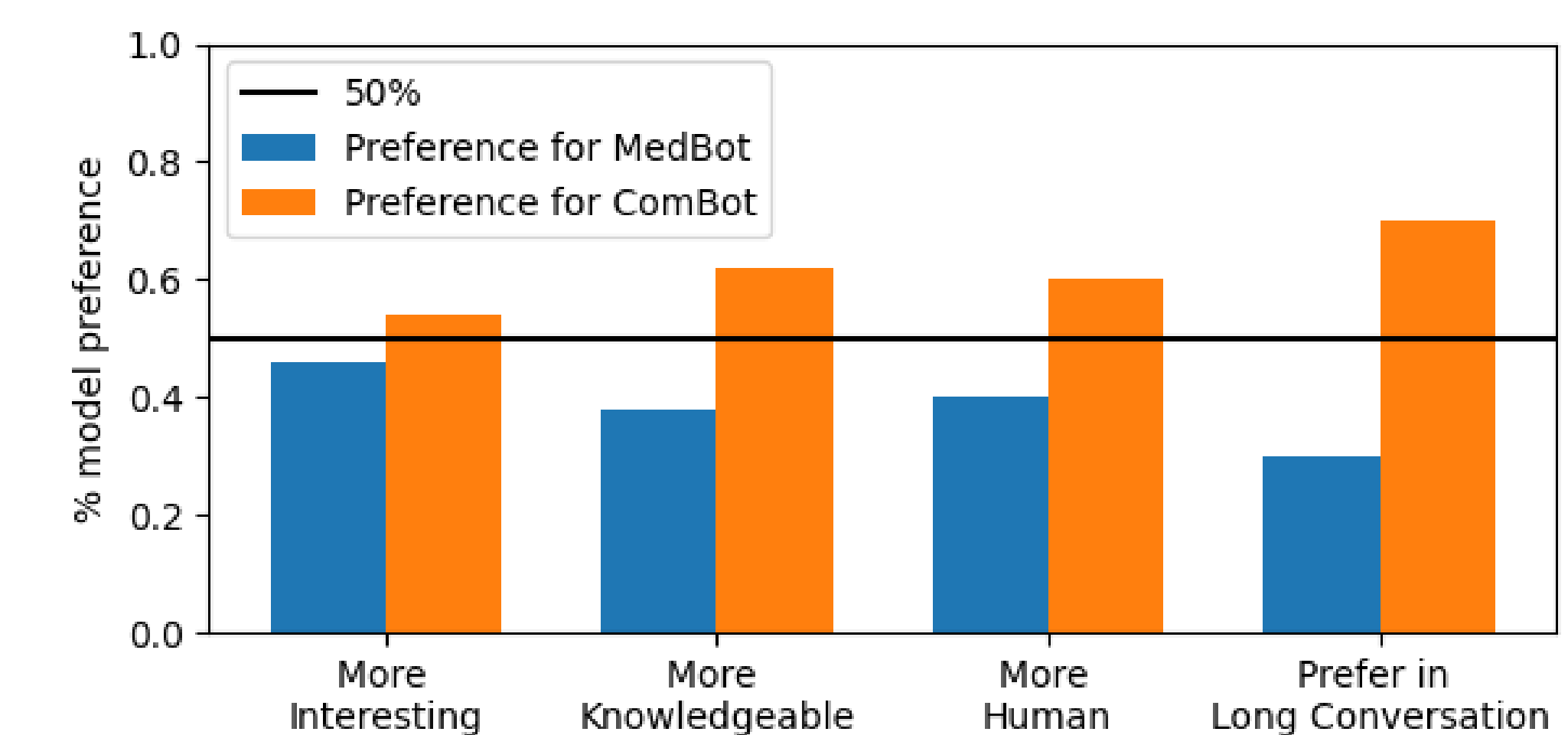
CoSim: Average Cosine Similarity between adjacent turns. Slots: Average Number of Medical Entities per dialog (in brackets: in the user turns). ConvLen: Average Number of turns per dialog. InfoGain: Average number of unique tokens per dialog (in brackets: normalised by dialog length). UserQ: number of questions asked by Human (in brackets: total number).

Figure 2: Satisfaction scores distribution

Figure 3: Acute-Eval results for MedBot and ComBot

Model	Satisf.	CoSim	Slots	ConvLen	InfoGain	UserQ
MedBot	3.94	0.26	6.24 (1.68)	28.46	108.82 (3.82)	0.08 (4)
MedBot+ FollowUpBot	3.18	0.34	11.65 (3.22)	36.06	153.23 (4.25)	0.47 (23)
MedBot+ EmpathyBot	3.77	0.34	3.87 (1.46)	30.29	140.19 (4.63)	0.68 (33)
ComBot	3.72	0.36	7.12 (2.82)	21.96	124.82 (5.68)	0.48 (24)

Table 2: Satisfaction Scores (Satisf.) and Results of the Automatic Evaluation.

Empathy Bot

Goal: provide necessary empathy and understanding to make the user comfortable and eager to continue interaction

We used a generative model Blender [Roller et al.2020] which was trained to engage use of personality, knowledge and empathy.

Conclusion

- ComBot provides a better basis for collecting information than MedBot
- ComBot collects information in a more user-friendly way and more efficient manner
- ComBot allows for more coherent dialogs
- MedBot is triggered more often after Follow-up Bot (30 cases) than after Empathy Bot (12 cases)
- ComBot is adapting to the user intents of sharing or not medical information

Contact Information

Web http://aliae.io/
Email contact@aliae.io