Towards Finding Appropriate Responses to Multi-Intents - SPM: Sequential Prioritisation Model

Jakob Landesberger
Speech Technology, Daimler AG
jakob.landesberger@daimler.com

Ute Ehrlich
Speech Technology, Daimler AG
ute.ehrlich@daimler.com

Abstract

Speech is an easily accessible and highly intuitive modality of communication for humans. Maybe that is the reason why people have wanted to talk to computers almost from the moment the first computer was invented. Today several consumer-level products developed in the last few years have brought inexpensive voice assistants into everyday use. The problem is that speech interfaces are mostly designed for certain simple commands. However, talking about several things in one utterance can make a dialogue more efficient. To find the appropriate reaction to such utterances, we propose prioritising one task according to certain criteria. Our sequential prioritisation model defines a six-step approach to address this problem.

1 Introduction

Utterances in dialogue serve often more than one communicative function. Like giving feedback about the understanding of a question and answering the question in a single utterance. The ability of humans to easily process such multiple communicative functions and to react accordingly, allows for a swift and effective communication (Lemon et al., 2002). This multifunctionality comes in a variety of forms. According to Allwood (1992), multifunctionality can be sequential or simultaneous. He gives an example where A's utterance contains the functions *feedback giving, request, request, request, statement,* and *response elicitation* in a sequential way.

> *A: Yes! Come tomorrow. Go to the church! Bill will be there, OK?* (Allwood 1992)

Bunt and Romary call these functional features such as request, statement, or promise dialogue acts and propose a formally definition:

> *A dialogue act is a unit in the semantic description of communicative behaviour produced by a sender and directed at an addressee, specifying how the behaviour is intended to influence the context through understanding of the behaviour.* (Bunt 2005)

Following the idea of multifunctionality, Bunt (1989, 2009) proposes the dynamic interpretation theory (DIT) which distinguishes dialogue acts in 10 dimensions where participation in a dialogue is viewed as performing several activities sequential and parallel. The First dimension is called Task/Activity. A dialogue act is labelled as Task/Activity if its performance contributes to performing the task or activity underlying the dialogue. Other dimensions cover dialogue acts like discourse structuring, turn management, or management of social obligations.
Utterances containing at least two sequential dialogue acts labelled as Task/Activity, which contributes to two different tasks or activities, are often called multi-intents (MI). Several Researchers used this expression in a human-machine interaction context. Kim et al. (2017) and Shet et al. (2019) propose algorithms to distinguish and segment such MIs like the utterances from speaker B and C.

> *B: Find the Big Bang Theory tv show and play it.*

> *C: What is the genre of big bang theory? Tell me the story about it.*

Such MIs are a useful mechanism to make a dialogue more efficient. Especially during demanding tasks like driving a car, it can be useful to talk about several things at once, to get back to the main task as fast as possible.

If both Tasks require further clarification, it can be difficult to define a proper reaction for a spoken dialogue system. Answering with a MI, too, can

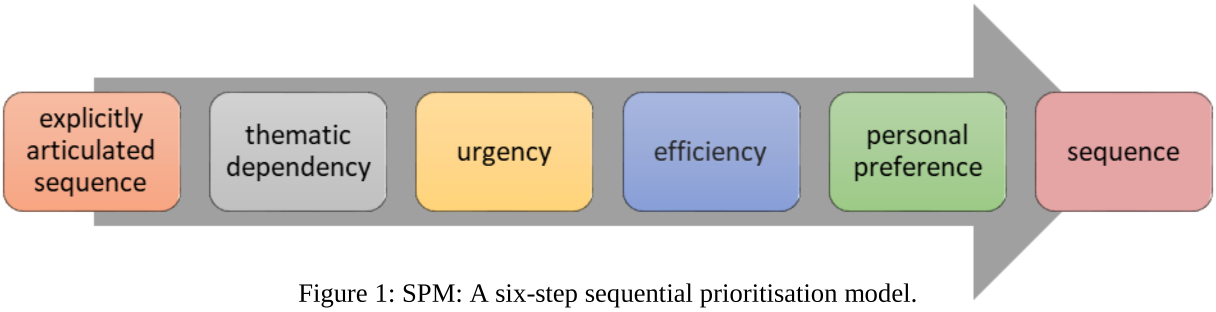

Figure 1: SPM: A six-step sequential prioritisation model.

produce long utterances, which can be cognitively very demanding. Prioritising a certain Task has to be logical and comprehensible. While a human conversation partner can easily decide if answering with a MI is appropriate and if not, identify the most important task, prioritise it and postpone the less important task, computers lack those skills. Therefore, we propose a six-step sequential prioritisation model (SPM) (see Figure 1).

2 Sequential Prioritisation Model:

Explicitly articulated sequence:
The first step is checking if the user mentions an explicit order of tasks. Speaker D explicitly structures the conversation by saying which part of the utterance he wants to take on first.

D: Call my mom and first tell me what the weather's gonna be like.

Thematic dependency:
If no discourse structuring hints are given in the speaker's utterance, there can be dependencies, which predefine the order of sequence.

E: I want to take a break. I am hungry. Isn't Berlin coming soon?

Speaker E wants to take a break and if Berlin is near, it seems like a good opportunity to stop there. Nevertheless, if Berlin is too far away E may need something to eat earlier. Therefore, before looking for restaurants the first task to approach is the last part of the utterance.

Urgency:
If none of the above mentioned criteria is met, there is a chance that one task is more urgent than the other is.

F: What do I do with the zucchini? Oh, the pan is hot. What is coming in now?

A task is urgent if the task has to be completed in a short amount of time, because if not, it loses relevance or other negative consequences occur. Certainly, speaker F would be frustrated if the first mentioned task is considered before the second one. Urgency seems to play an especially important role in an environment with rapidly changing situations.

Efficiency:
If the tasks are both not urgent and equally important, maybe one of the tasks can be done faster e.g. because it needs less turns to complete. If the first part of G's utterance is done first, the speech channel is blocked by the call. Therefore, the second task cannot be completed until the call has ended.

G: Call my mom and tell me what the weather's gonna be like.

Personal Preference:
If a user-model is present which represents the likings and preferences of the user, the task that is preferred by the user can be prioritised.

Sequence:
The last strategy is the fall back solution, where the system talks about the tasks in the sequence they were mentioned.

3 Conclusion

To overcome the Problem of creating cognitively too demanding dialogues, while reacting to MIs, we present a six-step sequential prioritisation model. Each step defines criteria for the prioritisation of one task and has to be considered before going on to the next one.
Our future research will deal with testing and evaluating the model in real world scenarios with a special focus on the role of urgency. Additionally, we will research the role of explicit discourse structuring in the system's response to clarify the decision in a logical and comprehensible way.

References

Allwood, J. (1992): On dialogue cohesion.Gothenburg Papers in Theoretical Lin-guistics 65. Gothenburg Univrsity, Department of Linguistics.

Bunt, H. (1989): Towards a dynamic interpretation theory of utterances in dialogue. In H. Bouma and B. Elsendoorn (eds) Working models of human perception.New York: Academic Press, pp. 419 – 456.

Bunt, H. (1994): Context and Dialogue Control.Think Quarterly3 (1), 19–31.

Bunt, H. (2005): A framework for dialogue act specification. Proceedings of SIGSEM WG on Representation of Multimodal Semantic Information.

Bunt, H. (2009): The DIT++taxonomy for functional dialogue markup. D.Heylen, C. Pelachaud, R. Catizone and D. Traum (eds.)Proc. AMAAS 2009 Work-shop "Towards a Standard Markup Language for Embodied Dialogue Acts', Bu-dapest, May 2009.

Bunt, H. (2011). Multifunctionality in dialogue. Computer Speech & Language, 25(2), pp. 222-245.

Kim, B., Ryu, S., & Lee, G. G. (2017): Two-stage multi-intent detection for spoken language understanding. Multimedia Tools and Applications, 76(9), pp. 11377-11390.

Lemon, O. et al. (2002): Multi-tasking and collaborative activities in dialogue systems. In Proceedings of the Third SIGdial Workshop on Discourse and Dialogue.

Shet, R., Davcheva, E., & Uhle, C. (2019): Segmenting multi-intent queries for spoken language understanding. Elektronische Sprachsignalverarbeitung 2019, 141-147.