Towards the Identification of Information Needs in Conversational Search Dialogues

Alexander Frummet

University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany
alexander.frummet@ur.de

Abstract

As conversational search becomes more pervasive, it becomes increasingly important to understand the user's underlying needs when they converse with such systems in diverse contexts. We report on an in-situ experiment to collect conversationally described information needs in a home cooking scenario. A human experimenter acted as the perfect conversational search system. Based on the transcription of the utterances, we present a coding scheme comprising 27 categories to annotate the information needs of users. Moreover, we use these annotations to perform prediction experiments based on random forest classification to establish the feasibility of predicting the information need from the raw utterances. We find that a reasonable accuracy in predicting information need categories is possible.

Keywords: conversational search; information needs; cooking

1 Introduction and motivation

Voice-based interaction systems are changing the way people seek information, making search more conversational. Spoken queries are very different to typed queries and by mining spoken interaction data, intelligent assistance can be provided. Information Science has a rich tradition of studying conversations to learn about user information needs (Taylor, 1962) and their behavior to address these (Belkin, 1987). Voice-based digital assistants such as Amazon Echo and Google Home show that information seeking conversations now take place in diverse situations embedded in users' everyday lives. One crucial feature for this kind of assistant is the ability to understand and

infer user needs. With conversational search tipped to dominate search in the future (Culpepper, 2018), it is crucial to understand how conversations vary in these diverse domains. If systems could move beyond simply providing results to offering intelligent assistance, complex tasks for which this mode of interaction is suited (Radlinski & Craswell, 2017) can be resolved better. According to Culpepper et al. (2018), accurately eliciting information needs is a key challenge for conversational search. The focus of this thesis is on need elicitation – specifically on understanding and predicting user information needs, which are important for systems to conversationally identify what a user requires, facilitate appropriate retrieval and attain relevance feedback (Trippas, 2018). We study information needs in the domain of home cooking, which, based on the literature, we believed would be a fertile context for the kinds of complex needs suited to conversational search (see Elsweiler, 2017; Cunningham & Bainbridge, 2013) and a situation where users simultaneously perform practical, sometimes cognitively challenging tasks that make searching in the traditional sense problematic.

We perform an in-situ study that facilitates a naturalistic cooking situation resulting in the organic development of information needs. We do this with respect to the following research questions:

RQ1: Which information needs occur in cooking situations?

RQ2: Can these needs be identified using conversational utterances from the user?

To answer these questions, we analyze the collected data qualitatively to learn about the diverse types of information needs which can occur in this context, and we utilize machine learning approaches to classify needs using the raw transcription of participant utterances.

2 Methods

2.1 Data collection

To establish a corpus of naturalistic conversational data large enough to perform machine learning prediction, we devised an in-situ user study. We simulated a natural cooking situation by gifting a box of ingredients to participants in their own kitchen. Participants were tasked with cooking a meal

which they had not cooked before based on as many of the contained ingredients as possible. To assist the process, they could converse with the experimenter who would answer questions and needs using any resource available to him via the Web. The experimenter communicated answers in a natural human fashion (arguably the optimal behavior for a conversational system). The advantage of this method was that it provided naturalistic speech utterances – we uncovered real information needs, described by participants in their own words while cooking in their own kitchen. Moreover, the method achieved a corpus large enough to run machine learning experiments.

2.2 Ingredients

To ensure divergent recipes and conversations the ingredient boxes varied across participants. The ingredients typically had a value of around €10 and were chosen based on guidelines by the German Nutritional Society (DGE). Typically, the box contained some kind of grain or starch (e.g., potatoes or rice), a selection of vegetables and a source of protein (e.g., eggs). Participants prepared diverse meals using the ingredients, a selection of which can be found in Figure 1.

Fig. 1 Some example meals cooked during the experiments

2.3 Participants

45 Participants (22 females, M_{age} = 24 years, *Range* = 19–71 years, 20% non-students) were recruited using a snowball sampling technique. This method offers two advantages. First, it generates a basis for trust among the participants and the experimenter which Castella et al. (2000) claim leads to more informal and open speech. Second, it allowed a relatively large sample to be achieved.

2.4 Transcription and identification of needs

In total, 38.75 hours of material were collected with the language spoken being German. The recorded conversations were transcribed and annotated by a trained linguist, who was also the experimenter, using the recommendations by Dresing and Pehl (2015). Thereafter, the utterances were split into queries whereby one query could represent one or more information needs. In total, trials yielded on average 36.93 queries (*Range* = 7–73, *Q1/Q3* = 22/50, *MED* = 36.0, *SD* = 17.48, *skewness* = 0.26, *curtosis* = 2.19). The overall number of queries extracted was N_q = 1662. Utterances could have the form of direct and indirect questions. They could also represent implicit and explicit actions.

3 Analyses

We analyzed the collected data both qualitatively and quantitatively. First, we examine the information needs identified to establish the variation of needs that occurred. The resulting set of information needs annotated with an appropriate category was used for quantitative experiments, to establish the feasibility of automatically categorizing the queries (information needs) using machine learning with the raw utterance text.

3.1 Coding scheme for information needs

The starting point for the coding scheme was the set of categories derived for cooking related questions posted on the Google Answers forum in Cunningham and Bainbridge (2013). Then, in an iterative process akin to the coding process by Strauss and Corbin (1996), each query was taken in turn, and a category from the existing scheme was attempted to be applied. When none of the existing categories were suitable, a new category was derived. The outcomes of this classification were 27 different information need categories. The distribution of these can be observed in Figure 2.

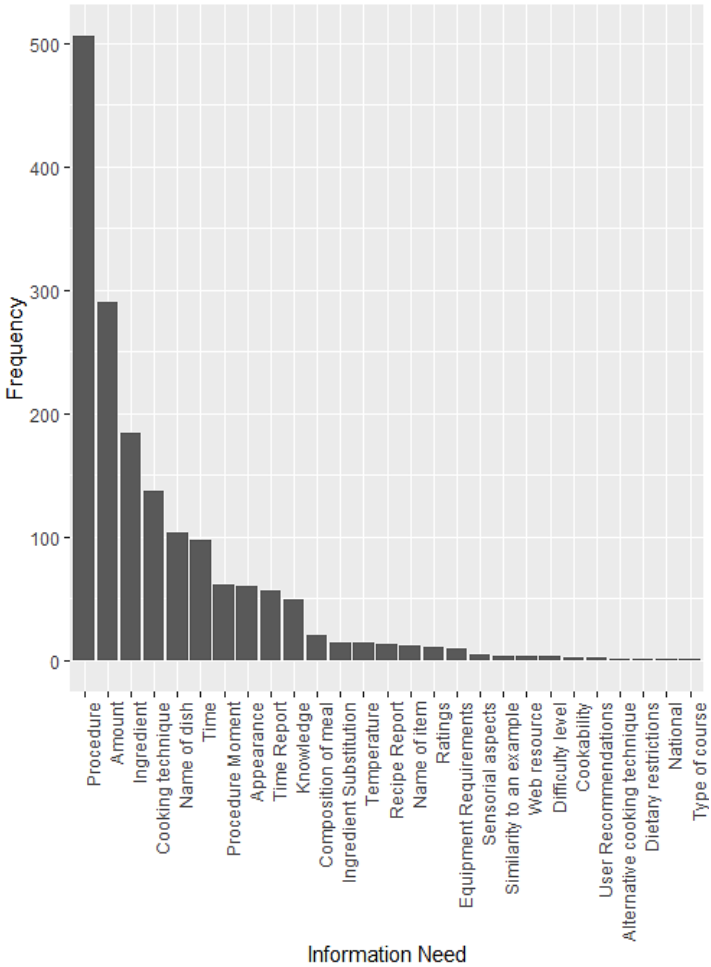

Fig. 2 Information need frequencies

3.2 Predicting information needs

The quantitative analysis was formulated as a prediction task, i.e., given a set of features derived from the raw conversational utterances and context information, is it possible to predict the category of information need. Since most information need categories contained only few turns, the prediction experiments are only concerned with the top 10 categories. We employed a random forest classifier for this purpose because it turned out to be an effec-

tive approach in Shiga et al.'s (2017) work. As the use of word embeddings was shown to be beneficial for predicting information need categories in Shiga et al., we used 200-dimensional word embeddings as a baseline feature for all classification experiments. By doing this, we achieved an average accuracy of 40%. Using resampling methods, we could significantly increase average accuracy up to 64%.

4 Conclusions

The results obtained by the prediction experiments show that the queries issued to the conversational search system are useful for distinguishing different information needs. Our results suggest that information need categories during conversation can be predicted with average accuracy values achieved of up to 64% when resampling is used. Even the non-resampled performance of approximately 40% is significantly larger than chance (which would be 10% with ten classes).

In the thesis, we discuss in detail the implications that our work has for the design of spoken conversational search systems in particular in specific domains such as the kitchen, thereby, making a contribution to the Information Science literature generally.

References

Belkin, N. J., Brooks, H. M, & Daniels, P. J. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies, 27*(2), 127–144.

Castellà, V. O., Zornoza Abad, A. M. Alonso, F. P., & Silla, J. P. (2000). The influence of familiarity among group members, group atmosphere and assertiveness on uninhibited behavior through three different communication media. *Computers in Human Behavior, 16*(2), 141–159.

Culpepper, J. S., Diaz, F., & Smucker, M. D. (2018). Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum, 52*(1), 34–90.

Cunningham, S. J., & Bainbridge, D. (2013). An Analysis of Cooking Queries: Implications for Supporting Leisure Cooking Ethnographic Studies of Cooks and Cooking. In Schamber, L. (Ed.), *iConference Proceedings* (pp. 112–123). iSchools.

DGE: Die deutsche gesellschaft für Ernährung e.V. (2018). https://www.dge.de/wir-ueber-uns/die-dge/

Dresing, T., & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende* (6th ed.). Marburg: Dr. Dresing und Pehl, Marburg.

Elsweiler, D., Trattner, C., & Harvey, M. (2017). Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 575–584). New York, NY: ACM Press.

Radlinski, F., & Craswell, N. (2017). A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway* (pp. 117–126). New York, NY: ACM Press.

Shiga, S., Joho, H., Blanco, R., Trippas, J., & Sanderson, M. (2017). Modelling information needs in collaborative search conversations. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 715–724). New York, NY: ACM Press.

Strauss, A., & Corbin, J. (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz.

Taylor, R. S. (1962). The process of asking questions. *American Documentation, 13*(4), 391–396.

Trippas, J., Spina D., Cavedon, L., Joho, H., & Sanderson, M. (2018). Informing the Design of Spoken Conversational Search: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA* (pp. 32–41). New York, NY: ACM Press.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 445–451. DOI: doi.org/10.5283/epub.44962.