The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Symbiotic Cognitive Systems: Technical Report WS-16-14

Interaction and Task Patterns in Symbiotic, Mixed-Initiative Human-Robot Interaction

Felip Martí Carrillo and **Elin Anna Topp**
Dept of Computer Science, Faculty of Engineering
Lund University, Sweden
e-mail: fmarti@swin.edu.au, elin_anna.topp@cs.lth.se

Abstract

In this paper we explain our concept of Interaction and Task Patterns, and discuss how such patterns can be applied to support mixed-initiative in symbiotic human-robot interaction both with service and industrial robotic systems.

1 Introduction

Recently, quite some research efforts have been made in the field of Human-Robot Interaction with industrial robotic systems, in many cases aiming at systems that would be easier to handle, to understand, and specifically to program by their users, reducing the need for robot experts on the shop floor and ultimately reducing the costs for deploying robotic systems in manufacturing for small and medium-sized enterprises. One example for such research efforts is *SMErobotics, the European Robotics Initiative for Strengthening the Competitiveness of SMEs in Manufacturing by integrating aspects of cognitive systems* (SMErobotics Project Consortium 2012). Within the SMErobotics consortium, one line of research is focused on the issue of symbiotic human-robot interaction, aiming to provide means for genuine mixed-initiative interaction. We assume this as crucial to support any type of human-machine symbiosis - both parts in the symbiotic relationship must be allowed and enabled to request and gather, but also to provide insight into their respective understanding of a situation so that they can benefit from these insights to solve their task in collaboration.

In this paper we explain our idea of Interaction and Task Patterns and their potential in HRI with both service and industrial robotic systems. Specifically for Interaction Patterns, we report results from a user study with 37 subjects, which allowed us to confirm the existence of such patterns in the interaction with a mobile service robot, as well as a prototypical implementation of an interaction monitoring system based on Bayesian Networks supporting identification, description and exploitation of the patterns. The idea of Task Patterns transfers the idea of identifying patterns in the interaction with a service robot into the perspective of industrial robotic systems - here it might not be the robot itself the human is interacting with directly, but there might be control systems and interfaces to be considered as interaction partners. The explanations of the concept and background ideas for Interaction Patterns are given in Section 2, Section 3 reports on the analysis of material from the user study conducted to confirm our hypotheses about Interaction Patterns. In Section 4 we discuss a prototypical implementation that was informed and evaluated with the data set gathered in the described study, and in Section 5 we elaborate on the transfer of the idea of Interaction Patterns to Task Patterns for HRI in industrial settings.

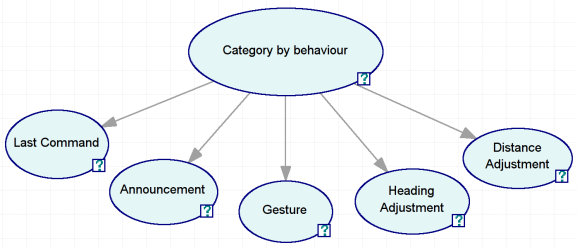

Figure 1: The core part of the Bayesian Network describing the relationship between observed behavioural features and the underlying category for the presented item of a SHOW-episode

2 Interaction Patterns

A significant part in mixed-initiative interaction is the understanding of the situation on the robotic system's side, including an understanding of the internal state and abilities of the system with respect to this particular situation, so that requests for additional information and clarification can be formulated appropriately, i.e., in a way that makes it easy for the human communication partner to provide the "right" (in terms of expected) response (Rosenthal, Veloso, and Dey 2012).

Now, looking at the human side of the communication, a rather natural assumption would be that the human who is interacting with the robotic system in a particular situation is perfectly capable of providing the necessary information to make the robotic system understand, e.g., in an instructing situation. However, when interacting with a robot in a situation in which the robot more or less replaces a human, people are certainly aware of the communication partner be-

ing "different", but they do not necessarily have an exact insight into the specific differences in understanding they might face. Hence, as we could observe in a number of studies, there are certainly adaptations made to compensate for assumed or perceived shortcomings on the robot's side, but at the same time, the human user might give implicit signals and information that she assumes to be understood by the robot just as if there was a human being adressed. We observed such situations in a "guided tour" scenario within several explorative user studies with a mobile service robot (Hüttenrauch, Topp, and Severinson Eklundh 2009, as an example) within the previously introduced framework for *Human Augmented Mapping* (Topp 2008). For example the user would point to the doorway leading into an office while presenting "the office" with the same type of utterance they used otherwise *inside* a room to introduce that, unaware that this information might lead to the robot storing its immediate surroundings (i.e, the corridor part close to the doorway) as "the office", instead of the area *beyond* the doorway. We consider such a situation as ambiguous and a definite source for misunderstandings and difficulties for the user to, firstly, understand what went wrong (why the robot stored, e.g., "the meeting-room" correctly in its map, but not "the office") and, secondly, find a way to correct the misunderstanding. To avoid such problems, we assume that the robotic system must be equipped with the means to detect ambiguities in a situation, i.e., any kind of deviation from expectations regarding user behaviour, surroundings, action or activity sequences should at least be evaluated and possibly lead to a request for disambiguation, before erroneous information is stored. Thus, to achieve genuine mixed-initiative interaction as a part of human-robot symbiosis, we need the means to describe both expectations and deviations from them to refine knowledge and contextual understanding on both the human and robotic system parts of this interaction.

Based on our experiences from the above mentioned user studies, where the investigations of potential patterns observable in the interaction were rather a sideline of research, we developed the concept of Interaction Patterns, i.e., reoccurring patterns in the observable interaction (including preparation / positioning of the robot and general movements around the robot), that might correspond to the underlying meaning, conceptual understanding, or even intention the user assumes for her utterances, which would give us at least some means to describe those expectations and deviations. Our investigations focused so far on patterns that would allow to hypothesise about the conceptual category (region, location, or object, see explanation in the following section) of an item presented to the robot beyond its label. E.g., while in the above example the utterance "this is the office" indicates that a room (or region) is presented, the user behaviour (pointing clearly) suggests that some specific location or large object (the door) is referred to. This should result in a mismatch of expected category and observations, which can further be used to trigger a request for clarification.

We confirmed the applicability of our idea in a more controlled study, that had explicitly the purpose to further investigate the concept of Interaction Patterns, of which we could so far report preliminary results from manual inspection of the material (Topp 2011). We can now report on a more thorough investigation of this particular study material in the following section.

3 The user study

Within the framework for Human Augmented Mapping a (partially) hierarchical generic model of space is assumed based on the three conceptual categories, *regions* (delimited areas, typically corresponding to rooms), *locations* (workspaces, defined by large, rather stationary objects, e.g., a fridge or a coffee-maker) and *objects* (small items that can be manipulated and are not stationary, e.g., cups or books). Further, the study was again conducted based on the guided-tour scenario. The coverage of all categories in each trial run of our study with 37 subjects was guaranteed through the instructions, that suggested items in different lists.

By inspecting the video footprint of the trials, the SHOW episode (Hüttenrauch et al. 2006) for each item presented during the trial runs was segmented into a *preparation phase* and a *show event* (corresponding to a gesture and / or an introducing utterance like "This is ..."). The observations were in an initial analysis categorised in a number of preparations and gestures according to our previously discussed hypotheses about Interaction Patterns (Topp 2011).

Specifically, we looked at the observable types of *preparations* and *gestures* applied before and during the SHOW-episode, such as "moving the robot to a specific position and/or pose", "fetching an item from a surface", "holding the item in front of the robot's camera", or "pointing with the whole hand".

We used these verbal and qualitative, human-comprehensible, descriptions of preparations and gestures to define a set of potentially machine-observable features, which were applied to annotate the video material. We also provided meta-annotations like the theoretically expected category for each item, information whether the subject had given explicit comments on their understanding and preferences to disambiguate the situation (e.g., "this is the printer-room, I am inside it, but you're outside") and an indication for the item being somewhat ambiguous in itself (coffee-makers can indeed be seen as both workspaces - a place to go to for picking up a cup of coffee - or objects, as at least the smaller, household-type can easily be lifted).

We consider the episodes independent from the actual timeline, i.e., we assume that order and timing of observations are irrelevant, only the presence or absence of a certain observation in connection with a SHOW-episode is important. We are aware that this might be a significant simplification in the general case, however, for the studied scenario our static approach seemed appropriate and allowed for much lower complexity in the analysis.

The annotated material was then quantitatively summarised, so that we could apply these statistics to train and evaluate a Bayesian Network (generated and trained with GeNie, https://dslpitt.org/genie as of 2015-10-30), where we assumed a causal relationship between the subject's "category assumption" and the observable behavioural features. We compared the (according to the network) most likely

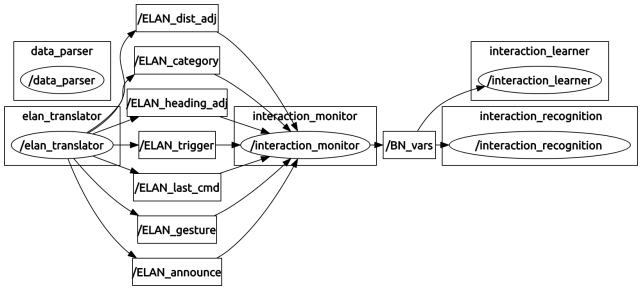

Figure 2: Our prototypical system for parsing and monitoring interaction events (features) for the interpretation of interaction patterns

"intended category" with our annotations regarding our theoretically assumed category and found our network to be a satisfying description for the material.
However, all these analysis steps were performed manually, hence the results are quite error-prone due to subjective interpretation of timelines and observations. Approaches to automating also the identification of the features in the first place are subject to current investigations.To avoid at least the last layer of subjectivity we implemented a prototypical interaction monitor system, that was able to parse the annotations of the video material and feed specific annotations as "observed features" into the respective core part of the Bayesian Network, which is shown in Figure 1. We explain our prototypical implementation in the following section.

4 Automated identification of Interaction Patterns

Our prototypical implementation (Martí Carrillo 2015) was based on ROS (www.ros.org, as of 2015-10-30), both for compatibility reasons within the SMErobotics related research efforts in symbiotic HRI, but also to benefit from the improvements and further development of ROS in comparison to the tools and hardware abstraction previously used for implementing the Human-Augmented Mapping software. For the manual annotation of the video footprint from the user study we had used ELAN (https://tla.mpi.nl/tools/tla-tools/elan, as of 2015-10-30), a for research purposes freely available tool that produces XML-files from which it is possible to reconstruct the original timeline and organise the annotations accordingly. Hence, one part of the prototype was a parser for the ELAN-generated annotation files, that provides the core part of the system, the interaction monitor, with the stream of annotations. As indicated above, we assume here, that it would be possible to exchange the parser with online recognition tools for different types of perceptions, e.g., a tracking system (here, we already tested an approach that would not rely on the manually provided annotations but on actual tracker data (Martí Carrillo 2015)), or vision based gesture recognition. The interaction monitor provides in turn the Bayesian Network with observations, whenever a SHOW-episode can be assumed to be concluded. Figure 2 gives an overview of the ROS-nodes (data_parser, elan_translator, interaction_monitor, interaction_learner / interaction_recognition) and -topics of the prototype. In our previous analysis step we had identified the following features (variables of the Bayesian Network) as most relevant:

- *last_command* – the last explicit motion command the user had given to the robot before presenting an item. In cases where several SHOW-episodes occurred right after each other without the robot moving in between, this observation would be missing from the second episode.
- *gesture* – the type of gesture that was observed in connection with the SHOW-episode, if any gesture was observed at all.
- *announce* – an item was announced (or not) to be presented in the near future, before the actual SHOW-episode took place ("and now, we go to the meeting room" would be such an observation).
- *heading_adj* and *dist_adj* – the user adjusted her position relative to the robot regarding heading or distance respectively, shortly before or while presenting an item (each feature could be independently present or not).
- *category* – not used as an observation or feature for the Bayesian Network, but applied for evaluation. Contains the category ascribed the presented item by the experiment supervisor, i.e., *region*, *location/workspace*, *object*, or*unknown*, as in some cases the annotation files did not contain a clear categorisation matching the time period for the SHOW-episode.
- *trigger* – indicates when an episode can be considered terminated, i.e., the actual item presentation was identified in the stream of annotations of user utterances.

We ran several different tests, where we used different subsets of the data set (37 subjects, 548 SHOW-episodes overall), to train and evaluate the Bayesian Network. We evaluated the performance according to the number of matches between the inferred item category by behaviour and the a priori category assumed by the operator. We classified the results of the network into clear matches or mismatches, when the network inferred the same or a different category as the assumed one with a significantly higher probability than for other categories, a similarity between two categories (with one of them being the assumed one), a similarity of two, but none of them was the assumed one, or a similarity between three categories, i.e., the network result was very ambiguous, and finally, it could happen, that the network suggested a category for an item previously assumed *unknown*. For two test cases the results are shown in Figure 3. Although the number for clear matches are below 50% in both cases, we claim, that the network still produces applicable results. In general, this would mean, that *relying only* on the automated identification of interaction patterns to assign a category to an item would be risky, but to *detect ambiguities*, that would be found in mismatches and the other types of not quite clear matches, the results would be quite suitable.
A more specific look into the cases classified as *similar between two*, showed that this was very often the case for items with categories that in some sense are close to another; as stated before, workspaces and objects are often difficult to classify even for a human user, in particular chairs and

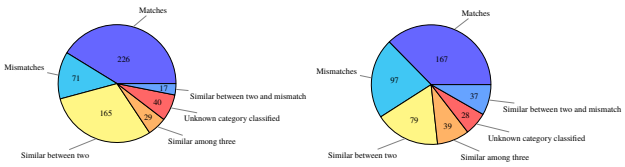

Figure 3: (left) Results for 100% of the data for training and evaluation and (right) Results for training with 18.5% and evaluation with 81.5% of the data

coffee-makers would end up in this unclear area quite often. Hence, it seems safe to say, that in simple cases, the network produced quite often a clear match, in ambiguous cases, it produced rather ambiguous results, which would help us to identify situations in which the robotic system should request a clarification from its user. Here, we plan to apply a more thorough analysis of specific cases to the results we could achieve so far to give a clear statement on the performance of the approach. However, we consider our results as a further confirmation for the possibility of identifying and applying Interaction Patterns (or Task Patterns as discussed in the following section) to enhance the understanding of an interaction situation.

5 Task Patterns

We assume the previously discussed Interaction Patterns to be mainly applicable in the realm of service or personal robots, where the interaction as such is somewhat more central than for industrial systems. With the idea of Task Patterns we target rather the area of industrial systems. We assume here that reoccurring patterns in task descriptions can be applied in several parts of the system and also in different contexts of interaction. For example, the sequence "washer - screw" is presumably part of many descriptions for the assembly of two parts by screwing them together (rather than snap-fitting or glueing). Given sufficiently many examples of different assembly tasks it would be possible to identify this sequence as a reoccurring Task Pattern. This and other previously learned patterns can then be applied to match a suggested sequence for an assembly with the available collection of expected patterns. If the sequence does for example contain "place a washer", but this is never followed by "place the screw", the system should request a clarification. Another application would be in actually supervising the user. If it was possible to identify the actual actions or activities connected with the features of the Task Patterns (picking and placing a washer, picking and placing a screw) from the observation of a human worker performing an assembly, potential erroneous assemblies (missing a washer) could be identified and corrected, before a failure-prone work piece leaves the production system. We are currently starting to investigate this idea by analysing annotated data from user studies performed in industrial (however laboratory) settings within the efforts of the SMErobotics research (Roitberg et al. 2015).

6 Conclusion

With this paper we explained our idea of Interaction Patterns, how we assume they can be exploited to support mixed-initiative human-robot interaction as a part of symbiotic interaction in terms of the identification of ambiguous situations, and discussed our prototypical implementation of an interaction monitoring approach that would allow us to automatically identify the patterns, potentially also in an online system. Our results allowed us to confirm the existence of Interaction Patterns in a specific scenario, and also their applicability in presumably unclear situations, so that confusion and errors in interaction can be reduced, if not avoided. We additionally explained our thoughts on how the concept of Interaction Patterns can be extended to Task Patterns, which we consider more relevant to industrial settings than Interaction Patterns. We consider these investigations as future work.

7 Acknowledgements

The user study was conducted at Lund University in collaboration with CITEC, Bielefeld University, Germany and partially funded through ENGROSS (Swedish SSF contract RIT08-0075). The further research leading to the reported results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n 287787. All support and funding efforts are gratefully acknowledged.

References

Hüttenrauch, H.; Severinson Eklundh, K.; Green, A.; Topp, E.; and Christensen, H. 2006. What's in the Gap? Interaction Transitions that make HRI work. In *Proceedings of IEEE RoMan, Hatfield, UK*.

Hüttenrauch, H.; Topp, E.; and Severinson Eklundh, K. 2009. The Art of Gate-Crashing: Bringing HRI into Users' Homes. *Interaction Studies* 10(3).

Martí Carrillo, F. 2015. Monitoring and Managing Interaction Patterns in HRI. Master's thesis, Lund University, Dept of Computer Science.

Roitberg, A.; Somani, N.; Perzylo, A.; Rickert, M.; and Knoll, A. 2015. Multimodal human activity recognition for industrial manufacturing processes in robotic workcells. In *Proceedings of ACM International Conference on Multimodal Interaction (ICMI), 2015*.

Rosenthal, S.; Veloso, M.; and Dey, A. K. 2012. Acquiring accurate human responses to robots' questions. *International Journal of Social Robotics* 4(2):117–129.

SMErobotics Project Consortium. 2012. SMErobotics. EU FP7, FP7-287787, http://www.smerobotics.org.

Topp, E. 2008. *Human-Robot Interaction and Mapping with a Service Robot: Human Augmented Mapping*. Doctoral Dissertation, KTH School of Computer Science and Communication (CSC), Stockholm, Sweden.

Topp, E. A. 2011. Understanding Spatial Concepts from User Actions. In *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM.