Towards Expectation-based Failure Recognition for Human Robot Interaction

Michael Karg [1], **Martin Sachenbacher** [1], **Alexandra Kirsch** [1]

[1] *Institut für Informatik, Technische Universität München, Garching bei München, Bayern, 85748, Germany*
{*kargm, sachenba, kirsch*}*@in.tum.de*

ABSTRACT

With robots becoming increasingly autonomous helpers in human environments, recent development goes from semi-autonomous robotic systems, that need to be directed, towards autonomous partners that can cooperate with humans in joint activities. When operating autonomously in the real world, the probability of unexpected events dramatically increases and failures might make it necessary for the robot to adapt its behavior to be able to fulfill its goals. In our complex and constantly changing real world, it is not possible for the programmer to anticipate all situations a robot might encounter. So the diagnosis of cooperative plans for human robot interaction is a particular challenge since for a robot it is often unclear what is to be considered as an error. A general understanding of normality, based on the validity expectations would enable a robot to detect unexpected events and failures that have not been foreseen by the programmer, thus leading to a more robust and flexible behavior of the robot. We propose the combination of different learned models and common-sense knowledge to generate expectations, that could improve failure detection and enable us to detect and react upon unexpected events. In this paper, we formulate the key challenges for failure detection in human robot interaction, which we see in the representation of expectations, the modeling efficiency and the execution efficiency. We also provide a stack of possible knowledge-based solutions.

1 INTRODUCTION

Robots generally were designed to perform monotonous and exhausting tasks to make life easier for humans. While they have been mostly used in factories so far, the achievements in the area of robotics in the last decades have made it possible to think about robots helping humans also in domestic environments. There are already research projects with robots assisting humans in medical care, elder care and even in cooperating with humans in everyday household tasks.

When performing complex tasks in continually changing environments like a human household, a robot will at some time be confronted with unexpected situations, which can lead to errors that might prevent the robot to fulfill its goals. To be able to deal with an error in an appropriate way, we have to detect and isolate it to initiate an adequate response.

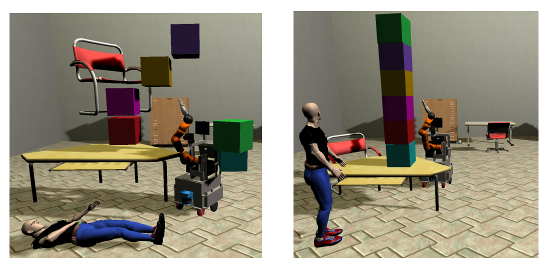

Figure 1: Left: An "abnormal" simulated scene in human robot interaction. Right: A "normal" simulated scene in human robot interaction.

In most model based diagnosis approaches, explicit models of the robot behavior are used (Carlson and Murphy, 2003), so every possible failure state has to be modeled. These approaches mostly focus on the internal faults of the robots that are caused by errors in the components of the robot itself. In our high-level planning domain, we are more interested in plan failures, that are caused despite perfectly functioning components of the robot, also referred to as external faults (Akhtar, 2011). For fault detection, we need knowledge about what a fault is. In complex robotic systems with human agents this is not necessarily always clear. Consider for example a robot and a human cooking a meal together, when suddenly the human partner leaves the room. Should this already be treated as an error and the robot stop its actions because the human might have abandoned the plan or did the human just get a telephone call and will return soon, expecting the robot to continue the common cooking plan? Or has the person forgotten that he/she was in the middle of a cooking task (which typically happens to people suffering from dementia)?

To diagnose unexpected events like this, the robot could use an abstract understanding of what is considered as "normal" in its environment at the given time. This general understanding of normality is a key aspect of human cognition and failure diagnosis, and has to our knowledge not sufficiently been addressed in the field of robotics and artificial intelligence. For example, the left picture in figure 1 shows a scene in a simulated environment that a human will quickly identify as "not normal" in contrast to the right scene. This understanding of normality of humans is generated by validating a combination of expectations based on different learned models and common sense knowledge. In the example of figure 1, flying objects will be classified abnormal because of the violation of physical models, but other observations, like the human lying on the ground, can not be explained this way. A human lying on the ground is physically possible, but might indicate an abnormal behavior of a human and could require the robot to react. When it comes to cooperative tasks of the human and the robot, there are even more events that can happen unexpectedly for the robot and might make reactions necessary. A human that goes to the couch and lies down might be perfectly normal, but not if the human and the robot are in the middle of cooking a meal together. Here an appropriate action for the robot would be to ask after the humans well-being, especially when dealing with elderly people. Also the fact that humans consider a situation as "not normal" influences their actions since in a situation that is "not normal" one might act more carefully or react differently than in a "normal" situation.

We want to equip robots with this general understanding of *normality* of a situation thus improving failure diagnosis in cooperative high level plans for human robot interaction and allowing the detection of unexpected events and failures that could not be detected with other approaches. It could also be used to regulate the level of error detection thus improving the computational efficiency of the diagnosis system, which is an important factor because an autonomous system in human robot interaction has to run in real-time.

Humans consider a situation as normal when it fulfills their *expectations* about the world, themselves and the situation. To classify a situation as "normal", we have to combine different expectations that have different properties and might depend on each other. We see expectations as a combination of common sense knowledge and models that implicitly and explicitly include predictions of plans, the robots environment, the human partner(s) and the robot itself. The combination of such expectations would allow failure diagnosis of external faults in robotic systems and give the robot a general understanding of what is considered as normal or not. This would enable us to diagnose unexpected events and errors without explicitly modeling every possible failure state.

There are already approaches in model-based diagnosis that model nominal behavior of technical systems, but for human robot interaction we face additional challenges when generating, combining and validating different expectation models that are applicable in this domain. The nominal behavior of a human for example should only be modeled on a certain level and always depends on factors like his plans, his actions, the environment and attributes of the individual human being. Thus for notions of normality, not just the events themselves but often also their durations matter. When for example the human in the aforementioned cooperative cooking task lies down on the sofa, the robot would classify this as an unexpected behavior when validating the expectations about the plan and knowing that lying down while cooking is not normal. It could trigger a reaction like asking the human why he lies down and how to proceed. Also the robot could be aware that a human sitting down for a few minutes, then continuing cooking could still be treated as normal but lying down for a longer time in the middle of cooking could be an evidence for the human to be sick or not feeling well.

In the next chapter, we will specify the challenges we see in designing such a flexible system that is able to use general expectations for failure diagnosis and recovery.

2 RELATED WORK

Prior work in the field of diagnosis for robotic systems mostly addresses the problem of internal faults of single components of a system. For example Gerald Steinbauer (Steinbauer and Wotawa, 2005) proposes observers that perform model based diagnosis on single components of a robot without affecting the behavior of the control system.

Kuhn et al. (Kuhn *et al.*, 2008) introduce the novel paradigm of pervasive diagnosis that enables active diagnosis and model based control simultaneously. Using predictions of uncertainties of given plans and a heuristic to calculate potential information gain of plans, they select the plan that maximizes the information gain while still achieving the goal of the plan. This provides them with the ability to identify the actions that caused the plan to fail.

Wiliams et al. (Williams *et al.*, 2003) suggest the use of model based autonomy to allow for state and fault aware autonomous systems in uncertain environments. They use an explicit model that encodes the nominal behavior of a system and common failure modes to perform extensive reasoning for failure recognition and recovery. However they see the future in "(...) cooperative robotic networks, in which robotic systems act together to achieve elaborate missions within uncertain environments.". In (Shah *et al.*, 2009) his group addresses planning with agents that can interact and adapt to humans by supporting just-in-time task assignment and scheduling. They rely on logic reasoning to predict the outcome of possible task allocations and scheduling decisions thus finding an optimal plan.

For the prediction of external faults of autonomous systems, Naveed Akhtar (Akhtar, 2011) uses qualitative reasoning on naive physics concepts for diagnosis. Here only external faults, that occur in the absence of any external agent, are considered and the focus of the work lies in the reasoning part of the system.

An example for the use of expectations in autonomous robots can be found in (Maier and Steinbach, 2010). Here Maier et al. work with expectations for diagnosis in the robotics domain.They use surprise detection based on a dense map of images of the robot's environment and comparisons of luminance and chrominance values of the images at different times. This enables them to detect changes in the robot's environment and make assumptions about the expectations and uncertainties in the environment.

These approaches work well for diagnosing faults in specific domains. But to our knowledge there is no approach that incorporates a combination of different models, including a human agent, to enable robots performing diagnosis on cooperative everyday tasks in household environments.

3 CHALLENGES

Most fault diagnosis approaches in robotics address internal faults of the robot, meaning failures in the robot's components, like sensors or actuators, and model based diagnosis has been successfully been applied in this domain. In robotics, complex nonlinear systems with noisy sensors and high uncertainties are used, while the aforementioned approaches are restricted to devices with well behaved processes (Akhtar, 2011). Naveed Akthar already addresses diagnosis of external faults of robotics systems in the absence of humans, but diagnosis of high level plans of a robot in the presence of humans and also in cooperation with them is still an open problem. For example, our group focuses on robots performing everyday household tasks together with a human, mainly in kitchen-environments. Therefore the robot uses reactive plans that can be adapted to changes and allow for reasoning about themselves. The human agent brings even more uncertainties in the continually changing and uncertain world and thus the use of only one model of plans, the environment or the system is not suitable for the detection of some errors that arise in cooperative plans in human robot interaction. The idea of working with expectations in this domain by combining different abstract models raises certain challenges that we want to split into three categories: the representation of expectations, the modeling efficiency and the execution efficiency.

3.1 Representation of expectations

To provide a robot with a measure of normality, we have to combine different models of expectations. These expectations should allow the robot to evaluate robot plans, recognize unexpected events and predict failures to avoid them or trigger appropriate reactions. We claim that for human robot interaction it makes sense to consider four categories: the environment, the human agent, the plans of the robot and the robot itself. If we think about the cooperative cooking task, a rule-based model of the environment could for example tell the robot that an object moves that is supposed to be static, which could be due to the robot or human accidentally pushing it or a leg of the table breaking off and objects slipping. On the other hand, a human suddenly leaving the room could be labeled as unexpected by having a set of areas where the human is expected to be. But this event should not yet be treated as an error, so we also need knowledge that sometimes human beings can do non task-related actions like going to the restroom in the middle of another task. This could for example be addressed by ontological knowledge about human beings that could also tell us that a human lying on the ground should be considered "not normal" (in most situations).

Expectations about all of the categories should be probabilistic because the real world is full of uncertainties and they should be adaptable since expectations can change over time. If we think about the human changing something in our world, we can see that our categories have dependencies among each other. Consider the human picking up an object and bringing it to another place. This action can involve changes in the environment, the human and even can affect the robot's plans if the object was part of the plan. So we need to find a representation that allows us to model different uncertain expectations that depend on each other and offers a way to combine them.

3.2 Modeling efficiency

Considering we have found a way to represent expectations, we still need to generate them. In our complex world, we do not want to model every possible state our world could be in, neither do we want to model every possible failure since this would result in an infinite amount of data very fast. Think about a robot trying to put a cup onto a table when suddenly the table collapses because one leg broke off. In a case like that we would like to be able to recognize that something went wrong, but we do not want to explicitly have to model that a leg of a table (or every other thing in the world that is not indestructible) can break, since this would result in an infinite number of models. Our expectations need to be general enough to give our robot an understanding of what is considered "normal" in our world and what should be treated as an "abnormal" event. So we have to find a way of combining different kinds of information to generate a measure of normality. This measure of normality could for example be generated by using abstract ontological and geometrical knowledge from available data bases like OpenCyc or the TUM kitchen data set (Tenorth *et al.*, 2009). But also learned models of the robot's experiences could be useful to address this challenge. Therefore we propose to use the Robot Learning Language (RoLL) (Kirsch, 2009), which provides us with learned data about the robots plans and environment as well as logged experiences of the robot.

3.3 Execution efficiency

To be able to react to unexpected events, we need to validate our expectations. In our domain of human robot interaction, it is obvious that the execution of our plans runs in real-time and the validation of the expectations should on the one hand also run in real time and on the other hand not prevent the actual planning from running in real time. One could imagine a diagnosis component that validates the current expectations during plan execution like the approach of Gerald Steinbauer (Steinbauer and Wotawa, 2005). The performance in our robotics application is affected by the huge state space we have when modeling the real world. To achieve a performance suitable for real-time applications, a solution could be not to consider the whole state space, but only parts of it. However this raises the question, which parts are relevant for the expectation that should be validated right now. It might also be useful for the robot not to validate expectations for its whole plan, but only for a short time in the future. So the challenge here is to find a set of expectations and their adequate level of detail they should have to be useful for diagnosis in human robot interaction.

4 CONCLUSION

An autonomous system that incorporates expectations into its high level planning to diagnose and prevent errors would be an important contribution. The enhancement of robustness and flexibility in combination with real time capabilities could improve the productivity and usefulness of future robotic systems. We see the main challenges for such a system to become reality in the representation of expectations, the modeling efficiency and the execution efficiency.

To generate high level plans for our autonomous systems, we use the CRAM plan language proposed by Beetz et al. (Beetz *et al.*, 2010). This allows autonomous robots to reason about their plans and infer control decisions thus making them more flexible, reliable and general than robots without that ability. We see our future contribution in extending the CRAM plan language with a general representation of expectations to allow for a more adaptive, anticipatory robot that adapts to the human and foresees possible problems to avoid them. Currently we have already set up a system and an simulation environment, that will enable us to test different expectation models for plan based robots and evaluate them in cooperative human robot interaction tasks in simulated household environments. We use the MORSE simulator (Echeverria *et al.*, 2011) and a simulated human that can be controlled as in modern 3D computer games. Figure 2 shows a simulated environment for human robot interaction using the MORSE simulator.

In our future work, we want to provide robots with a general understanding of normality and enable them to "expect the unexpected".

Figure 2: A simulated environment for human robot interaction using the MORSE simulator

ACKNOWLEDGMENTS

With the support of the Technische Universität München - Institute for Advanced Study and Cluster for Excellence Cognition for Technical Systems, funded by the German Excellence Initiative.

REFERENCES

(Akhtar, 2011) Naveed Akhtar. *Fault reasoning based on naive physics*. Hochschule Bonn-Rhein-Sieg, 2011.

(Beetz *et al.*, 2010) Michael Beetz, Lorenz Mösenlechner, and Moritz Tenorth. CRAM – A Cognitive Robot Abstract Machine for Everyday Manipulation in Human Environments. In *IEEE/RSJ International Conference on Intelligent RObots and Systems.*, 2010.

(Carlson and Murphy, 2003) J. Carlson and R.R. Murphy. Reliability analysis of mobile robots. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 1, pages 274–281. IEEE, 2003.

(Echeverria *et al.*, 2011) G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan. Modular openrobots simulation engine: Morse. In *Proceedings of the IEEE ICRA*, 2011.

(Kirsch, 2009) Alexandra Kirsch. Robot learning language — integrating programming and learning for cognitive systems. *Robotics and Autonomous Systems Journal*, 57(9):943–954, 2009.

(Kuhn *et al.*, 2008) L. Kuhn, B. Price, J. de Kleer, M. Do, and R. Zhou. Pervasive diagnosis: Integration of active diagnosis into production plans. In *proceedings of AAAI*, 2008.

(Maier and Steinbach, 2010) Werner Maier and Eckehard Steinbach. A probabilistic appearance representation and its application to surprise detection in cognitive robots. *IEEE Transactions on Autonomous Mental Development*, Vol. 2, No. 4, pp. 267 - 281, 2010.

(Shah *et al.*, 2009) J.A. Shah, P.R. Conrad, and B.C. Williams. Fast distributed multi-agent plan execution with dynamic task assignment and scheduling. *Proceedings of ICAPS-09*, 2009.

(Steinbauer and Wotawa, 2005) G. Steinbauer and F. Wotawa. Detecting and locating faults in the control software of autonomous mobile robots. In *19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 1742–1743. Citeseer, 2005.

(Tenorth *et al.*, 2009) Moritz Tenorth, Jan Bandouch, and Michael Beetz. The TUM Kitchen Data Set of Everyday Manipulation Activities for Motion Tracking and Action Recognition. In *IEEE Int. Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS). In conjunction with ICCV2009*, 2009.

(Williams *et al.*, 2003) B.C. Williams, M.D. Ingham, S. Chung, P. Elliott, M. Hofbaur, and G.T. Sullivan. Model-based programming of fault-aware systems. *AI Magazine*, 24(4):61, 2003.