A Survey of Collaborative Problem-Solving in Human-Robot Interaction

Chris Swetenham (s1149322)

19th Jan 2012

Abstract

Collaborative Problem-Solving brings together many different fields in computer science and artificial intelligence to enable humans and robots to collaborate on shared tasks. We investigate the challenges which need to be overcome, and detail several recent projects which start the work of combining all the required elements for true human-robot collaboration.

1 Introduction

1.1 Problem Statement

Collaborative Problem-Solving looks at a class of task in which human and robot participants must work together "shoulder to shoulder"[HB04] to cooperatively solve a problem, often using natural speech and gestures to communicate. This is a multidisciplinary problem in Human-Robot Interaction which brings together much of the current work in robotics, natural language, human-computer interaction, and artificial intelligence. While many parts of the required capabilities have been studied for some time, it is only relatively recently that these have been brought together into research platforms which attempt to investigate the entire end-to-end task of collaborative problem-solving.

This problem is important because it is how we interact with other humans and how we ultimately want to interact with intelligent, embodied agents. It encompasses tasks involving teams comprising members with different knowledge and skills, both human and robot.

It is a difficult problem, because it requires many capabilities which have no well-established solution, such as gesture recognition, or mature ones which are still error-prone, such as speech recognition.

1.2 Challenges

Human and robot participants must be able to understand each other without impediment to the task, must understand each other's roles and intentions in the task (*intention recognition*), and must agree on steps to be performed to complete the task (*joint intention*). The robots must also be careful not to take actions which might endanger the human participants (*safety*). The robot may need to understand human speech (*natural language processing*) and make itself understood in return (*natural language generation*). Speech output can

be accompanied by facial expressions or gestures, such as looking at and/or pointing to the object being referenced (*multimodal interaction*).

1.3 Related Work

Human-Agent Interaction is an extension of Human-Computer Interaction to interactions with software that exhibits some form of agency, potentially embodied and situated in a virtual world. Human-Robot Interaction is an extension of Human-Agent Interaction to interactions with embodied, situated agents in the physical world. [GS07] gives a survey of the field of Human-Robot Interaction.

Social Human-Robot Interaction studies robots which interact by social means with humans and each other, including gestures, facial expressions and speech. [Fon03] gives a survey of Social Human-Robot Interaction.

Human-Robot Collaboration studies social interaction between humans and robots in the context of a shared task to be performed. [BWB08] gives a survey of Human-Robot Collaboration.

1.4 Motivating Examples

We draw from the literature a number of motivating examples for cooperative problem-solving.

The first example is drawn from the JAST (Joint-Action Science and Technology) Project. In this project, a human and a robot collaborate to assemble structures from a toy construction kit. Both participants communicate using natural language and to a lesser extent using gestures. The robot can perform some steps of the assembly, instruct the human in the assembly steps required, and correct the human participant if they pick up the wrong piece[RKF+08].

The second example is drawn from the Leonardo Project. In this project, a small, highly-expressive robot and a human collaborate to activate a sequence of buttons. The human communicates with speech and gesture, and the robot entirely with gestures. The human teaches the robot to press a button, after which the human and the robot negotiate with each other the steps in pressing several buttons in turn[BHL04].

The third example is a simulated seam-welding exercise as part of the NASA Robonaut Project. This exercise simulates a team of astronauts and robots assembling a structure on a planetary surface. Rather than using real welding tools, the exercise uses spray paint for ease of experimentation. Two astronauts act as "master welders", placing panels and performing initial "tack welds". One additional astronaut acts as a remote supervisor. The highly articulated Robonaut handles a welding tool to finish the welds, and a mobile rover robot inspects the quality of the welds. When necessary, humans and robots can request additional assistance from each other with their individual tasks; for example, the astronauts can request the inspection rover to turn its spotlight towards the panel they are working on, or the inspector robot can request the advice of the supervisor if it cannot determine the quality of a weld[FKHB06].

1.5 Overview

In this report we will cover several projects in the area of collaborative problem-solving in human-robot interaction in teams with both human and robot partic-

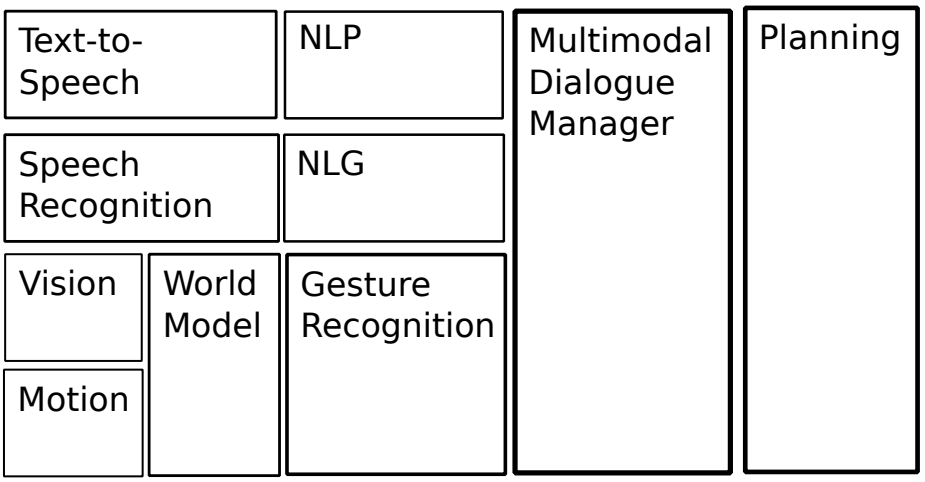

Figure 1: Example architecture of a robot's software for collaborative problem-solving

ipants. We will in examine each of the challenges listed in Section 1.2 in turn, and see how they are tackled by the projects listed above, as well as in related projects where applicable, including focus on approaches to natural language generation, intention recognition and the use natural conversation in shared planning and execution of a task.

Figure 1 shows the general software architecture for collaborative problem-solving robots.

In Section 2 we will look at forms of Human-Robot communication in general. In Sections 3,4, 5 we will look at different types and stages of communication. In Section 6, we will look at how the collaborative problem-solving aspect comes together in the dialogue between the participants. In Section 7 we will examine how performance in collaborative problem-solving tasks is evaluated. In Section 8 we will discuss recent and future directions in collaborative problem-solving. Finally in Section 9 we will discuss and summarise.

2 Human-Robot Communication

At the core of Human-Robot Communication, is the notion of having multiples *modes* of interaction or communication. These interactions can be generated from or interpreted to an internal grammar by the robot system. The 'sentences' in the grammar are operated on by a dialog engine, which interacts with the task planning system and the robot's internal model of the world.

In terms of communication from the robot to the human, there is some leeway in which input and output modalities are provided to the robot. The Leonardo project[?] focuses on an expressive robot capable of many facial and body gestures and without any natural language generation or speech synthesis capabilites. The JAST project [FBRK06] combines two industrial arms with the iCat expressive robot face, and features speech generation and recognition. [vB04] describes the Lino and iCat robots and the principles used to animate them.

3 Natural Language Processing and Generation

Speech is processed in several forms in a collaborate problem-solving system. The lowest level is raw sound samples captured from or output to the physical world. The next level is a textual representation, in sentences or sentence fragments. The next level after this is more abstract and consists of syntax trees representing the information communicated. At this level, the representation can be independent of the human language used by the system. The JAST project, for example, is able to converse in both English and German[FGI+09]. The final level of abstraction is whatever internal representation is used by the dialogue engine to represent the state of the dialogue.

Translation between these representations is handled by different components of the system. Translation between speech and text is done by Speech Recognition and Text-To-Speech. Translation between text and abstract sentences is done my Natural Language Processing and Natural Language Generation. Translation between abstract sentences and internal representations is done by the Dialogue Manager.

In general, Speech Recognition and Text-To-Speech are treated as solved problems in human-robot interaction, with some leeway for the inevitable interpretation errors from either participant. It will not be discussed further in this paper.

Natural Language Processing is a well established field dealing with processing the text into some more abstract model. It is a fairly well-solved problem in the context of collaborative problem-solving, since it is a restricted domain: the participants will be discussing only the shared world they are operating in, and their intentions with respect to the task.

Natural Language Generation is the counterpart of Natural Language Processing. It refers to the production of human text or speech from some internal data structure specific to the agent. In the context of this review the most important aspects are communicating the actions to be performed and the objects they should be performed on. In fact, generating descriptions of objects and their location is easy; the difficult part is deciding which information to include or not in the description. If there is not enough information, the description can be ambiguous or confusing for the human listener; if the information is too specific, it may seem to the listener as though the extra information was included for a specific reason even if it is irrelevant to the task. In [FGI+09] the JAST project has investigated different strategies for referring expressions and their impact on both task performance and user satisfaction. In [BKS+09] the GIVE challenge evaluates different strategies for guiding a user in an online game to move through an environment and perform actions. [BG08] find that task performance metrics do not correlate with metrics that measure how "human-like" the output of an NLG system is.

The Dialogue Manager is discussed in its own Section 6.

4 Gesture Recognition and Generation

An important part of collaborative-problem solving is the understanding and use of gestures. When humans communicate in a shared task they will often use gestures as well as speech. These may serve several purposes: to convey

intention or emotion, to direct the other participant's actions or attention, or to directly execute a step in the plan for the task. For example, a pointing gesture could direct the other participant's attention to a tool or area, or it could tell them to move themselves to the location pointed to.

Gesture recognition can be performed on an internal model of the world after computer vision has processed the raw video input. For example, in the software stacks available for the Microsoft Kinect sensor, the raw video and range data is segmented into regions, and the position of the joints and skeleton of any humans in the scene are inferred. From this joint data, gestures can be recognised by a variety of machine-learning methods.

4.1 Hand-over of objects

Extending an arm holding an object towards the other participants signals the intention to hand over the object while also being the first step of the execution of this intention. [KSS+06] and [HRK+08] explore handing-over gestures and approaches that are comfortable and recognisable to the human user.

4.2 Intent Recognition

[NDK+05] looks at classifying human gestures and recognising intent, and find that the context as well as the gesture itself need to be taken into account. The JAST robot can infer the intent of the human user when they pick up pieces and correct them when they pick up the wrong piece [RKF+08].

5 Multimodal Interaction

When performing a task, humans have many simultaneous modes of interaction: speech, facial expressions and stances, gaze direction, gestures, direct physical contact. It may be beneficial or even necessary to consider multiple modes at once to understand the overall meaning.

In the JAST project, the generation of referring expressions and the use of gestures were both studied and evaluated for their effectiveness in communicating with the human participant [FGI+09]. [Van05] gives an algorithm for the generation of multimodal referring expressions. In the Leonardo project, speech output was avoided entirely in favour of an expressive body and facial design. [GK08] describes the MultiML language for representing multimodal actions in a dialogue.

6 Dialogue

Collaborative problem-solving requires a dialogue engine which at minimum can navigate through pre-defined dialogues and keep track of the current state. More advanced dialogue engines include planning and plan or intention recognition components which allow the engine to reason about and react to the intentions of other participants.

Joint Intention refers to the state of affairs where several participants share a common goal and a common plan for achieving that goal. In order to reach

this situation, both participants must continually communicate their intentions as the execution of the task progresses.

[LT00] describes the TRINDI dialogue engine toolkit. It is based on the notion of a shared or individual *information state* which is updated by the *dialog moves* of the participants. The tookit defines the basic data structures and some dialog moves, but the precise information and the choice of dialog moves can be selected according to the task required. Simple examples of dialog moves are asking or answering a question. The TRINDI toolkit has been used in the JAST project.

A similar toolkit, Ariadne, is used by the NASA HRI/OS. [Den02] describes the Ariandne toolkit. [FKHB06] describes NASA's HRI/OS, developed as part of the Robonaut robotic astronaut project. HRI/OS allows for a wide range of interactions between humans and robots, from remote teleoperation to local collaboration. Robots using this system can request help from humans or other robots when they are unable to complete a task by themselves.

[RSL01] describes the highly influential COLLAGEN dialogue system for collaborative dialogue, developed in the context of collaborative problem-solving with a software agent rather than a robot.

[BA05] propose a different system which includes dialogue moves (which they call *interaction acts*) which are used to negotiate, accept, or reject changes to the shared problem-solving state, such as deciding to focus on a particular sub-problem or adopt a certain solution. In addition, there are wrapped in *grounding acts* which handle turn-taking in conversation, requests for acknowledgement, and requests for clarification. In this system, joint intention is achieved by negotiating every change to a shared problem-solving state.

The dialogue drives the shared execution of the task, and different ways of collaborating lead to different dialogue managers. For example, within the scope of collaborative problem solving, each member of the team, human or robot, can take on different roles. One example is the related roles of teacher and student. In the Leonardo project, the human participant sets the goal and teaches the robot the steps required[BHL04]. For example, the robot is taught to press a button, and then is given the task of pressing several buttons. In the reverse direction, in the JAST project, the robot participant sets the goal and teaches the human the steps required [FGI+09]. The human is taught sub-tasks which are then combined into a larger task.

7 Evaluation Methods

Evaluation of collaborative problem-solving tasks can be difficult since there are many possible metrics, and some of the interesting ones can be expensive to gather since they require human annotation of captured data. It is straightforward to capture video and audio data of the task performance, as well as tracking internal state of the robot and measuring the task success or failure and the time taken for the task. Others, such as the number of errors made or other properties of the dialog, may require human annotation.

The evaluation methods used in collaborative problem-solving can be traced back to ones used in Human-Computer Interaction and Natural Language Generation, for example the PARADISE evaluation framework[WLKA97].

In practice, evaluation is done in much the same way in collaborative problem-solving research. It is based on: task success or failure, time taken to complete the task, rate of errors based on human annotation of session recordings, and human participant satisfaction based on standard questionnaires[FNK+05, FGI+09, Sha11].

8 Future Directions

8.1 Teaching and Learning

In terms of working on a shared task, the Leonardo project focuses on the human teaching the robot how to participate in a task. The idea is that active tutelage can be much more effective than relying on blind experimentation by the robot or complex, brittle a priori knowledge in getting the robot to perform a new task[BHL04].

8.2 Mixed-Initiative Interaction

When both participants contribute towards the goal, rather than one participant driving the dialogue, this is termed a Mixed-Initiative system. [Gui96] describes an early model of mixed-initiative communication based on exchanging information and deductions between agents until a conclusion is reached. [BHL04] describe how Leonardo can suggest it takes the initiative or request help from its human partner. [BA05] describe a model for negotiating shared goals and plans between participants, and use it to analyse a planning discussion between two human participants. [FA07] describe an architecture for implementing a similar model in an agent.

8.3 Anticipation

Beyond the projects listed so far, some recent work has investigated the potential benefits of the robot participants anticipating the actions and needs of the human participants. [HB08] investigates a joint task between a human and a robotic lamp under time pressure. The robotic lamp could be commanded to move by pointing by the human, and could be told to change colour. The human and robot team had to complete a sequence of actions in the shortest time possible. They found that they could make the robot learn to anticipate the actions of their partner and start moving to help them rather than reacting after the fact. In some cases, the robot member became so effective that the human participant felt they were letting the team down with their own performance. This suggests that building anticipation into collaborative systems can greatly enhance their performance. [Sha11] applies these same ideas to an assembly task similar to that in the JAST project. By anticipating the required pieces it could bring them before the human participant requested them, significantly reducing the time the human participant spent idle waiting for the robot.

8.4 Safety

If humans and robots are to collaborate in close proximity, safety is an important issue[ABB+06]. In traditional industrial robots, humans and robots are simply

kept separated by distance or physical barriers and have a kill switch to stop the robot entirely in an emergency. This is the approach taken in the JAST project, where the robot only interacts with its own side of the table, except to hand an object to the human participant.

The second version of the Robonaut project tackled the issue of safe interaction directly, with a triply-redundant system using feedback from motors to monitor and limit the forces applied by the robot to its environment. As a result, collisions with a human or with another unexpected obstacle will simply stop or slow the motion of the robot to safe levels[DMA+11].

9 Discussion

Collaborative Problem-Solving is a dynamic and promising area of Human-Robot Interaction research, with applications in space[FNK+05], medical and elderly care[Fon03], and any context in which it is desirable to have autonomous robot participants perform tasks in a dangerous environment in collaboration with human participants, such as military applications and urban search and rescue[GS07].

In many cases the scope of the work on collaborative problem-solving is limited by the manipulation and visual understanding capabilities of the robot, but the nature of the tasks undertaken remains realistic and relavant to applications in the real world. Collaborative problem-solving, mixed-intiative and multimodal interactions allow for natural integration of robot members in a team without requiring special training of the human participants, and as the robot becomes more aware of intent and social cues, it becomes an efficient collaborator and is regarded less as a tool and more as a participant[HB07].

References

[ABB+06] R Alami, Antonio Bicchi, R Bischoff, R Chatila, A De Luca, and A De Santis, *Safe and Dependable Physical Human-Robot Interaction in Anthropic Domains: State of the Art and Challenges*, Society (2006), no. 1.

[BA05] Nate Blaylock and James Allen, *A Collaborative Problem-Solving Model of Dialogue*, Proceedings of the SIGdial Workshop on Discourse and Dialog, 2005, pp. 200–211.

[BG08] Anja Belz and Albert Gatt, *Intrinsic vs. Extrinsic Evaluation Measures for Referring Expression Generation*, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies Short Papers - HLT '08 (2008), no. June, 197.

[BHL04] Cynthia Breazeal, Guy Hoffman, and Andrea Lockerd, *Teaching and Working with Robots as a Collaboration*, Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3 (Washington, DC, USA), AAMAS '04, IEEE Computer Society, 2004, pp. 1030–1037.

[BKS+09] Donna Byron, Alexander Koller, Kristina Striegnitz, Justine Cassell, Robert Dale, Johanna D Moore, and Jon Oberlander, *Report on the First NLG Challenge on Generating Instructions in Virtual Environments (GIVE)*, Proceedings of the 12th European Workshop on Natural Language Generation - ENLG '09 (2009), 165–173.

[BWB08] Andrea Bauer, Dirk Wollherr, and Martin Buss, *Human-Robot Collaboration: A Survey.*

[Den02] Matthias Denecke, *Rapid Prototyping for Spoken Dialogue Systems.*

[DMA+11] M A Diftler, J S Mehling, M E Abdallah, N A Radford, L B Bridgwater, A M Sanders, R S Askew, D M Linn, J D Yamokoski, F A Permenter, B K Hargrave, R Platt, R T Savely, R O Ambrose, General Motors, and Oceaneering Space Systems, *Robonaut 2 - The First Humanoid Robot in Space*, Robotics **1** (2011).

[FA07] George Ferguson and James Allen, *Mixed-Initiative Systems for Collaborative Problem Solving*, AI Magazine (2007), 23–32.

[FBRK06] Mary Ellen Foster, Tomas By, Markus Rickert, and Alois Knoll, *Human-Robot Dialogue for Joint Construction Tasks*, ICMI (2006).

[FGI+09] Mary Ellen Foster, Manuel Giuliani, Amy Isard, Colin Matheson, Jon Oberlander, and Alois Knoll, *Evaluating description and reference strategies in a cooperative human-robot dialogue system*, Proceedings of IJCAI, 2009, pp. 1818–1823.

[FKHB06] Terrence Fong, Clayton Kunz, Laura M Hiatt, and Magda Bugajska, *The human-robot interaction operating system*, Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06 (2006), 41.

[FNK+05] Terrence Fong, Illah Nourbakhsh, Clayton Kunz, Lorenzo Fl, John Schreiner, Reid Simmons, Laura M Hiatt, Alan J Schultz, J Gregory Trafton, Magda Bugajska, and Jean Scholtz, *The Peer-to-Peer Human-Robot Interaction Project*, System (2005), 1–11.

[Fon03] Terrence Fong, *A survey of socially interactive robots*, Robotics and Autonomous Systems **42** (2003), no. 3-4, 143–166.

[GK08] Manuel Giuliani and Alois Knoll, *MultiML - A General Purpose Representation Language for Multimodal Human Utterances*, Interfaces (2008), 165–172.

[GS07] Michael A Goodrich and Alan C Schultz, *Human-Robot Interaction: A Survey*, Foundations and TrendsÂ® in Human-Computer Interaction **1** (2007), no. 3, 203–275.

[Gui96] Curry I Guinn, *Mechanisms for Mixed-Initiative Human-Computer Collaborative Discourse*, Computer (1996), 278–285.

[HB04] Guy Hoffman and Cynthia Breazeal, *Collaboration in Human-Robot Teams.*

[HB07] ———, *Effects of Anticipatory Action on Human-Robot Teamwork Efficiency , Fluency , and Perception of Team*, Artificial Intelligence (2007), 1–8.

[HB08] ———, *Anticipatory Perceptual Simulation for Human-Robot Joint Practice: Theory and Application Study*, Artificial Intelligence (2008), 1357–1362.

[HRK+08] Markus Huber, Markus Rickert, Alois Knoll, Thomas Brandt, and Stefan Glasauer, *Human-robot interaction in handing-over tasks*, RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication (2008), 107–112.

[KSS+06] K L Koay, E A Sisbot, D S Syrdal, M L Walters, Kerstin Dautenhahn, and R Alami, *Exploratory Study of a Robot Approaching a Person in the Context of Handing Over an Object*, Proceedings of the IEEE (2006).

[LT00] Staffan Larsson and David R Traum, *Information state and dialogue management in the TRINDI dialogue move engine toolkit*, Natural Language Engineering **6** (2000), no. 3&4, 323–340.

[NDK+05] C L Nehaniv, Kerstin Dautenhahn, J Kubacki, M Haegele, C Parlitz, and R Alami, *A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction*, ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. (2005), 371–377.

[RKF+08] Markus Rickert, Alois Knoll, Mary Ellen Foster, Manuel Giuliani, and M Thomas, *Combining Goal Inference and Natural-Language Dialogue for Human-Robot Joint Action*, 25–30.

[RSL01] Charles Rich, Candace L Sidner, and Neal Lesh, *COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction*, AI Magazine (2001), 15–26.

[Sha11] Julie A Shah, *Fluid Coordination of Human-Robot Teams*, Ph.D. thesis, 2011.

[Van05] Ielka Francisca Van Der Sluis, *Multimodal Reference: Studies in Automatic Generation of Multimodal Referring Expressions*, Ph.D. thesis, 2005.

[vB04] A J N van Breemen, *Bringing Robots To Life: Applying Principles Of Animation To Robots*, Proceedings of Shapping HumanRobot Interaction workshop held at CHI **4** (2004), 1–5.

[WLKA97] Marilyn A Walker, Diane J Litman, Candace A Kamm, and Alicia Abella, *PARADISE: A Framework for Evaluating Spoken Dialogue Agents*, 271–280.