Collaborative Problem Solving in Human-Robot Interaction

Chris Swetenham (s1149322)

First Draft - 1st Jan 2012

1 Introduction

1.1 Problem Statement

Collaborative Problem Solving 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. 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 (*collaborative problem solving* and *joint intention*). The robots must also be careful not to take actions which might endanger the human participants (*safety*). Speech output can be accompanied by facial expressions or gestures, such as looking at and/or pointing to the object being referenced. This an example of *multimodal interaction*.

1.2 Area

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.

1.3 Overview

In this report we will cover several projects in the area of collaborative problem-solving in human-robot interaction between a single robot and single human participant. We will in particular focus on approaches to natural language generation, intention recognition and the use natural conversation in shared planning and execution of a task.

In Section 2 we will look at Human-Robot communication in general. In Sections 3,4, 5 we will look at different types of communication. In Sections 6, 7,8 we will look at how the collaborative problem-solving aspect comes together. 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 [TODO] 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. Museum guide robots such as INDIGO typically feature speech but not gesture modes of communication [TODO] although the Robotinho project does use gestures [TODO].

3 Natural Language Processing and Generation

In general, Speech Recognition and Generation are treated as solved problems in human-robot interaction, with some leeway for the inevitable interpretation errors from either participant. Speech is then treated in terms of text. Natural Language Processing is a well established field dealing with processing the text into some more abstract model. 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.

4 Gesture Recognition and Generation

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.

In a robot, gesture recognition will be performed as a step after the visual interpretation of a scene from video data.

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 [TODO]. [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. [TODO] describes the MultiML language for representing multimodal actions in a dialogue.

6 Collaborative Problem-Solving

[LT00] describes the TRINDI dialog 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.

[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.

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. In the above system, joint intention is achieved by negotiating every change to a shared problem-solving state.

In the Leonardo project, the human participant sets the goal and teaches the robot the steps required [TODO]. 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 [TODO]. The human is taught sub-tasks which are then combined into a larger task.

[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.

7 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.

8 Mixed-Initiative Interaction

When both participants contribute towards the goal, 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.

9 Discussion

Collaborative Problem-Solving is a dynamic and promising area of Human-Robot Interaction research, with applications in space[TODO], medical and elderly care[TODO], 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[TODO] and urban search and rescue[TODO]. Collaborative problem-solving, mixed-intiative and multi-modal 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 is regarded less as a tool and more as a participant[TODO].

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.

[AF02] James Allen and George Ferguson, *Human-Machine Collaborative Planning*, International NASA Workshop on Planning (2002).

[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.

[BAD+] William Bluethmann, Robert Ambrose, Myron Diftler, Scott Askew, Eric Huber, Michael Goza, Fredrik Rehnmark, Chris Lovchik, and Darby Magruder, *Robonaut: a robot designed to work with humans in space*, Autonomous robots **14**, no. 2-3, 179–97.

[BBG+04] Cynthia Breazeal, Andrew Brooks, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, Andrea Lockerd, and David Chilongo, *Tutelage and Collaboration for Humanoid Robots*, International Journal of Humanoid Robotics **1** (2004), no. 2, 315–348.

[BBG+06] Tilman Becker, Nate Blaylock, Ciprian Gerstenberger, Ivana Kruijff-korbayov, Andreas Korthauer, Manfred Pinkal, Michael Pitz, Peter Poller, and Jan Schehl, *Natural and Intuitive Multimodal Dialogue for In-Car Applications : The SAMMIE System*, Language (2006).

[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.*

[CBC+] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, and Hao Yan, *Human Conversation as a System Framework: Designing Embodied Conversational Agents.*

[DA01] M A Diftler and R O Ambrose, *Robonaut: A Robotic Astronaut Assistant*, 1–8.

[DR94] Robert Dale and Ehud Reiter, *Computational interpretations of the Gricean maxims in the generation of referring expressions*, Cognitive Science **19** (1994), no. 2, 233–263.

[DSDB07] Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi, *An atlas of physical human-robot interaction*, Mechanism and Machine Theory **43** (2007), no. 3, 1–18.

[EMCB07] Wolfram Erlhagen, Albert Mukovskiy, Fabian Chersi, and Estela Bicho, *On the Development of Intention Understanding for Joint Action Tasks*, 2007 IEEE 6th International Conference on Development and Learning **2007** (2007), no. July, 140–145.

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

[FBG+08] Mary Ellen Foster, Gurman Bard, Markus Guhe, Robin L Hill, Jon Oberlander, and Alois Knoll, *The Roles of Haptic-Ostensive Referring Expressions in Cooperative, Task-based Human-Robot Dialogue*, Human Communication Research (2008).

[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.

[FGK09] Mary Ellen Foster, Manuel Giuliani, and Alois Knoll, *Comparing objective and subjective measures of usability in a human-robot dialogue system*, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - ACL-IJCNLP '09 **2** (2009), no. August, 879.

[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.

[FM06] Mary Ellen Foster and Colin Matheson, *Following Assembly Plans in Cooperative , Task-Based Human-Robot Dialogue*, Assembly (2006).

[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.

[Fos08] Mary Ellen Foster, *Automated Metrics That Agree With Human Judgements On Generated Output for an Embodied Conversational Agent*, Proceedings of the Fifth International Natural Language Generation Conference on - INLG '08 (2008), no. 2006, 95.

[Fos09] ______, *Human-Robot Joint Action in the JAST project*, 2009.

[Fos11] ______, *Roles of a Talking Head in a Cooperative Human-Robot Dialogue System.*

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

[GKP03] Jennifer Goetz, Sara Kiesler, and Aaron Powers, *Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation*, Psychology (2003).

[GM06] Leonardo Giusti and Patrizia Marti, *Interpretative Dynamics in Human Robot Interaction*, ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication (2006), 111–116.

[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.

[GSL08] Kristof Goris, Jelle Saldien, and Dirk Lefeber, *The Huggable Robot Probo, a Multi-disciplinary Research Platform.*

[Gui95] Curry I Guinn, *Meta-Dialogue Behaviours: Improving the Efficiency of Human-Machine Dialogue: A Computational Model of Variable Initiative and Negotiation in Collaborative Problem-Solving*, Ph.D. thesis, 1995.

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

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

[HM06] Melita Hajdinjak and France Mihelič, *The PARADISE Evaluation Framework: Issues and Findings*, Computational Linguistics **32** (2006), no. 2, 263–272.

[HRJ04] Pamela J Hinds, Teresa L Roberts, and Hank Jones, *Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task*, 151–181.

[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.

[KBC+09] Alexander Koller, Donna Byron, Justine Cassell, Robert Dale, Johanna D Moore, Jon Oberlander, and Kristina Striegnitz, *The software architecture for the first challenge on generating instructions in virtual environments*, Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session on - EACL '09 (2009), 33–36.

[Kes07] S Keshav, *How to read a paper*, SIGCOMM Comput. Commun. Rev. **37** (2007), no. 3, 83–84.

[KHZ97] Alois Knoll, B Hildenbrandt, and J Zhang, *Instructing cooperating assembly robots through situated dialogues in natural language*, Proceedings of International Conference on Robotics and Automation **1** (1997), 888–894.

[KK06] John D Kelleher and Geert-Jan M Kruijff, *Incremental generation of spatial referring expressions in situated dialog*, Computational Linguistics (2006), no. July, 1041–1048.

[KPHL04] Jinsul Kim, Jihwan Park, Yong K Hwang, and Manjai Lee, *Advanced Grasp Planning for Handover Operation Between Human and Robot : Three Handover Methods in Esteem Etiquettes Using Dual Arms and Hands of Home-Service Robot Grasp Planners for Handover Operations*, Autonomous Robots (2004), no. c, 34–39.

[KPvB+03] B J A Krose, J M Porta, Albert J N van Breemen, K Crucq, M Nuttin, and E Demeester, *Lino, the user-interface robot.*

[KSG+08] Alexander Koller, Kristina Striegnitz, Andrew Gargett, Donna Byron, and Justine Cassell, *Report on the Second NLG Challenge on Generating Instructions in Virtual Environments (GIVE-2)*, Challenge (2008), 243–250.

[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).

[KW05] Alfred Kranstedt and Ipke Wachsmuth, *Incremental Generation of Multimodal Deixis Referring to Objects*, Artificial Intelligence (2005).

[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.

[Mat] Colin Matheson, *Human-Robot Dialogue: JAST and INDIGO.*

[MGPR05] Patrizia Marti, Leonardo Giusti, Alessandro Pollini, and Alessia Rullo, *Experiencing the flow: design issues in human-robot interaction*, Science (2005), no. october.

[Moo95] Johanna D Moore, *The Role of Plans in Discourse Generation*, Discourse (1995).

[MSII04] Takashi Minato, Michihiro Shimada, Hiroshi Ishiguro, and Shoji Itakura, *Development of an Android Robot for Studying Human-Robot Interaction.*

[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.

[OASY09] Tetsushi Oka, Toyokazu Abe, Kaoru Sugita, and Masao Yokota, *RUNA: a multimodal command language for home robot users*, Artificial Life and Robotics **13** (2009), no. 2, 455–459.

[OD00] Bernard Ogden and Kerstin Dautenhahn, *Robotic Etiquette : Structured Interaction in Humans and Robots*, no. 1998.

[OR07] Jeff Orkin and Deb Roy, *The Restaurant Game: Learning Social Behavior and Language from Thousands of Players Online*, Network (2007), 1–39.

[RFG+07] Markus Rickert, Mary Ellen Foster, Manuel Giuliani, Tomas By, and Giorgio Panin, *Integrating Language, Vision and Action for Human Robot Dialog Systems*, Access (2007), no. ii, 987–995.

[RJ00] Jeff Rickel and W Lewis Johnson, *Task-Oriented Collaboration with Embodied Agents in Virtual Worlds*, Communication (2000), 1–29.

[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.

[SDG+11] Kristina Striegnitz, Alexandre Denis, Andrew Gargett, Konstantina Garoufi, Alexander Koller, and Mariet Theune, *Report on the Second Second Challenge on Generating Instructions in Virtual Environments (GIVE-2.5)*, World (2011).

[SG09] Glenda Shaw-Garlock, *Looking Forward to Sociable Robots*, International Journal of Social Robotics **1** (2009), no. 3, 249–260.

[SHSW05] Oliver C Schrempf, Uwe D Hanebeck, Andreas J Schmid, and Heinz Worn, *A Novel Approach To Proactive Human-Robot Cooperation*, Robot and Human (2005).

[SMS05] Amanda Stent, Matthew Marge, and Mohit Singhai, *Evaluating Evaluation Methods for Generation in the Presence of Variation.*

[SSBFL06] Laura Stoia, Darla Magdalene Shockley, Donna K. Byron, and Eric Fosler-Lussier, *Noun phrase generation for situated dialogs*, Proceedings of the Fourth International Natural Language Generation Conference on - INLG '06 (2006), no. July, 81.

[SSK06] Matthias Scheutz, Paul Schermerhorn, and James Kramer, *The Utility of Affect Expression in Natural Language Interactions in Joint Human-Robot Tasks*, Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06 (2006), 226.

[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.

[vB05] Albert J N van Breemen, *iCat: Experimenting with Animabotics*, 27–32.

[vBYM05] Albert J N van Breemen, Xue Yan, and Bernt Meerbeek, *iCat: an animated user-interface robot with personality*, 143–144.

[WLKA93] Marilyn A Walker, Diane J Litman, Candace A Kamm, and Alicia Abella, *PARADISE: A Framework for Evaluating Spoken Dialogue Agents*, 271–280.