Nineteenth National Conference on Artificial Intelligence (AAAI-04), July 25-29, 2004, San Jose, California

Common Sense Data Acquisition for Indoor Mobile Robots

Rakesh Gupta and **Mykel J. Kochenderfer***
Honda Research Institute USA, Inc.
800 California Street, Suite 300
Mountain View, CA 94041
rgupta@hra.com, m.kochenderfer@ed.ac.uk

Abstract

Common sense knowledge can be efficiently collected from non-experts over the web in a similar fashion to the Open Mind family of distributed knowledge capture projects. We describe the collection of common sense data through the Open Mind Indoor Common Sense (OMICS) website. We restrict the domain to indoor home and office environments to obtain dense knowledge. The knowledge was collected through sentence templates that were generated dynamically based on previous user input. Entries were converted into relations and saved into a database. We discuss the results of this online collaborative effort and describe two applications of the collected data to indoor mobile robots. We discuss active desire selection based on current beliefs and commands and a room-labeling application based on probability estimates from the common sense knowledge base.

Introduction

The objective of this research is to enhance the intelligence of mobile robots so that they can autonomously accomplish tasks in a home or office environment. For these tasks, the robots must posses some common sense including knowledge about human desires, objects and their locations, and causality. Since common sense does not require expert knowledge, the data may be collected as part of a public online collaborative effort over the Internet.

Distributed online knowledge acquisition projects, such as those associated with the *Open Mind Initiative* (Stork 1999; 2000), have become quite popular. The Open Mind Common Sense project, led by Push Singh at the MIT Media Lab, has accumulated a corpus of 700,000 pieces of knowledge from 14,000 users (as of January 2004) over the past three years[1]. Other projects such as Open Mind Word Expert[2] and Open Mind 1001 Questions[3], have also been successful.

*Currently at Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom.

[1]http://commonsense.media.mit.edu

[2]http://www.teach-computers.org/word-expert.html

[3]http://www.teach-computers.org/learner.html

This paper describes the Open Mind Indoor Common Sense (OMICS) project that became publicly available at the beginning of August 2003[4]. In the next section we describe the knowledge base framework and templates used to capture data. We then report on our results and experiences with working with online contributors. The next section of the paper discusses two applications that use this data. We discuss inference based on anticipating the desires of the users. We also describe a room-labeling application based on simulated object recognition and simulated room and object labels given by a user. Finally, we discuss our conclusions and future work.

Data Collection Methods

In this section we first describe the framework of the knowledge base and the relations necessary to capture those types of common sense most useful to an indoor robot. We then describe how we built a website to capture this data in a user-friendly manner and how we converted the data into machine-understandable relations.

Knowledge Representation

The framework of this work is *object-centric*. Robot activities involve perceiving the environment and acting upon it. What the robot knows about the environment must include what objects are present and their state. The robot must manipulate these objects in such a way so as to put them in some desired state to accomplish its goals. Hence, everything that the robot knows about the world and can do in the world is grounded in objects and their properties (or state).

The robot can observe properties of objects in its vicinity and it can perform actions that change the properties of objects. In this system, a statement is a pair $\phi = (o, p)$ where o is some object and p is an adjective describing the property. Statements may be thought of as assertions about the property of an object in the world or actions to be taken on an object to achieve a particular effect (actions are referred to by the effect they achieve.) For example, the statement (*cup-of-coffee*, *hot*) can mean "a cup of coffee is hot" or represent the action "make a cup of coffee hot." Using the same notation allows us to make connections between beliefs, desires, and intentions.

[4]http://openmind.hri-us.com

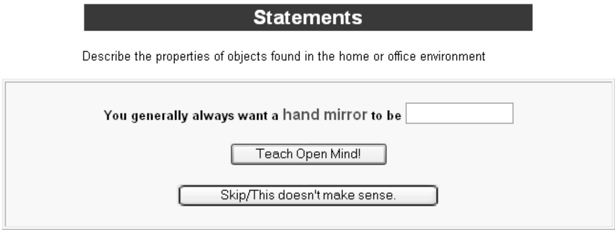

Figure 1: A template sentence.

Our representation allows us to capture such common sense knowledge as: (o_1, p_1) causes (o_2, p_2). For example, the statement (*fan*, *on*) causes (*room*, *cool*). We also wish to capture knowledge about human desires such as (o_1, p_1) indicating the human desire (o_2, p_2). For example, the perception (*cup-of-coffee*, *cold*) indicates that the desire (*cup-of-coffee*, *hot*) be fulfilled.

At any given point in time, the robot observes a set of statements that are true and can execute a set of statements. Using the common sense knowledge about causality and desires as previously described, the general problem is to decide which statements to execute in order to achieve perceived goals.

Indeed, this object-centric representation is limited. It cannot express everything that first order logic with a large array of predicates can represent. In addition to making inference simpler, this "object-centric" representation makes it much easier to collect data from sentence templates.

Activity Sentence Templates

We need some way to convert the common sense in the minds of non-expert users into relations in a knowledge base. Following the style of the Open Mind Common Sense website, we decided to use sentence templates. Users are prompted to fill in the blanks of sentences with words, as shown in figure 1.

Once a user logs on with their account, they are presented with a random activity. After a random number of entries for a particular activity, the system prompts them with a new activity. Users may also manually switch between the activities.

Different activities capture different types of knowledge. Below is a summary of some of the activities:

Objects In this activity, the user is asked to identify the types of objects commonly found in a home or office. The user may be prompted to type in an object name that comes to mind or simply label an image of an object. It is important to allow a user to simply type in any indoor object that comes to mind because we want to include all relevant objects in the database even if we do not have their picture.

Image labeling can link multiple labels to the same object (e.g. the labels "sofa" and "couch" might be both associated with "h0141.gif"). The images themselves can be used for training the object recognition system of the robot. When the website became public, the database initially contained a set of over 400 images of indoor objects selected by hand from a collection of stock photography.

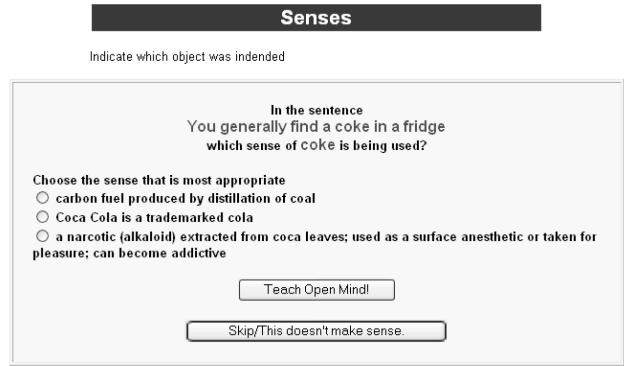

Figure 2: A word sense disambiguation form.

Statements In the 'statements' activity, the user is prompted with a question such as, "You often want a fan to be ______." This activity pairs objects with properties in the knowledge base. The objects that appear in these sentence templates come from the objects entered by users in the other activities such as the 'objects' activity.

Uses This activity associates objects with their uses. For example, the user might be prompted with the form, "A hanger is used to ______." Again the objects come from user input.

Causes This activity captures causality. For example, a form might ask, "A computer is off when a ______ is ______." If the user enters a new object or a new object-property pair, it will be entered into the object or statement table. The object and property that makes up the first part of the sentence is formed dynamically by selecting a random object from the knowledge base.

Desires This activity helps the robot determine what needs to be done in various situations. A template form might ask, "You might want a fan to be blowing if you notice that your ______ has become ______."

Locations This activity associates objects with the rooms where they are typically found. For example, the user might be prompted with, "A room where you generally find a dinner table is the ______."

Proximity This activity associates objects with each other based on proximity. A sample form would be, "You generally find a frying pan in the same room as a ______."

Senses This activity disambiguates the intended sense of various objects entered into the database by other users. Figure 2 shows a sample form. The objects to disambiguate are selected from previous user entries and the senses are from WordNet (Miller 1995).

People This one describes the activities of people in a home or office (e.g. People eat food when they are hungry). The template form is, "People ______ when they ______ ."

Paraphrase This activity tries to capture multiple ways of interacting with the robot in natural language to accomplish a task. A sample template is "Another way to say heat the food in the microwave is ______."

Tasks This activity tries to capture the steps required to accomplish a task like making coffee, answering the phone etc. We prompt the user with 7 short natural language steps to accomplish the task. A sample template is "The task water indoor plants involves the steps: ______."

Generalization This activity tries to generalize upon previous entries in the knowledge base. Instead of having to individually specify that bananas, oranges, apples, etc. are commonly found in the kitchen, it would be useful to know that all kinds of fruit are found in the kitchen. If we know that fruit is found in the kitchen, we can use the WordNet hierarchy to infer that all of the hyponyms of fruit are found in the kitchen. An example of a prompt for this activity is the following: "Are all types of writing implement (a generalization of 'marker') commonly found in the study?"

Freeform This activity allows users to type in any form of common sense knowledge that might not be captured by any of the other activities. Although it is quite difficult to convert freeform sentences into useful relations, it provides us with a sense of the types of knowledge the general public would like an indoor robot to understand. Analysis of freeform sentences will later lead to the creation of new activities.

Data Quality Review

It is important that there be some way of ensuring data quality since the data (such as names of objects and their properties) are used to generate new sentence templates. Sentence templates containing misspelled objects or objects that do not appear in a home or office environment would propagate errors in the knowledge base.

The completed sentence templates are stored in the database as raw sentences pending administrator review (see figure 3). It generally takes an administrator roughly half a minute to scan through a page with fifteen submissions. Once an administrator approves a set of entries, they are parsed into relations immediately. There is currently no need for part-of-speech tagging or lemmatization (as with the Open Mind Common Sense project) since the sentence templates are structured and designed in such a way that they implicitly cue the user as to what part-of-speech and tense they should use.

Data Collection Results

To advertise the Open Mind Indoor Common Sense website, a message was sent to the Open Mind Initiative mailing list on August 5th. With no other direct advertising, within three weeks we had 190 users and 18,000 submissions with about 17,000 of them accepted. As of March 2004 we have over 400 users with 29,000 submissions with over 26,000 of them accepted.

We have had two weekly contests (lasting four weeks each) in August 2003 and February 2004 where the top contributor was awarded an Open Mind t-shirt. Other Open

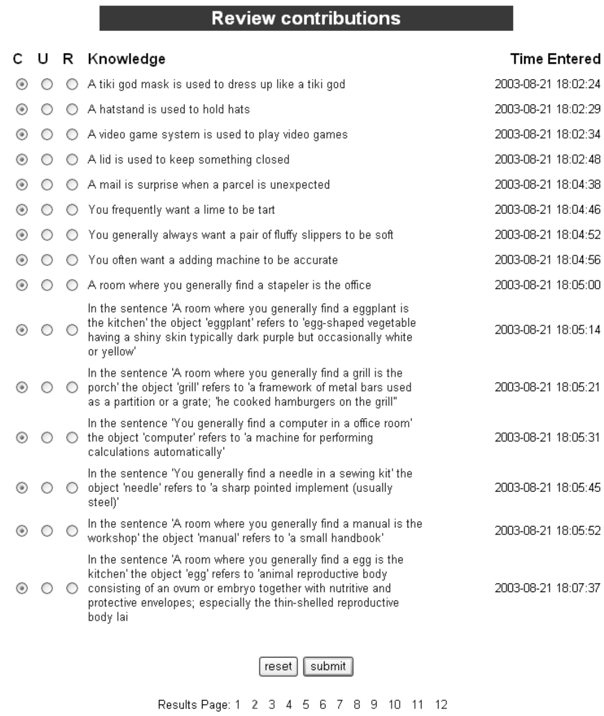

Figure 3: The review form used for administrators to commit, uncommit, or reject entries.

Mind projects have used similar contests to help motivate submissions. Winners were listed on the front page of the site.

Observations

The OMICS site was publicly announced on August 5th and the first t-shirt prize was awarded August 6th. A significant portion of the entries were submitted within two days of the announcement. Prizes were also awarded on August 12, August 19, and August 26. In general, submissions were greater close to contest deadlines.

The quality of the data was actually quite good. About 10% of the submissions were rejected. Entries that were rejected tended to fall within one of the following categories:

- Misspelling: e.g. "A room where you generally find a exercise bike is the bym."
- Unclear or loose statements: e.g. "People cry when they can't get it."
- Outside the scope of home and office environments: e.g. "A trap is set when a predator is hunting."
- Nonsense: e.g. "You generally want a green light to be addressed to you."
- Inappropriate: e.g. suggestive or obscene

The 'causes' activity had the highest rejection rate. Deciding how one object affects another object proved to be difficult for some users. Interestingly, almost all word sense activities were answered correctly. Even users that entered

Activity	Count
Objects	5804
Uses	3517
Locations	3400
Statements	3394
Proximity	2547
Freeform	1696
People	1667
Desires	1792
Causes	1558
Senses	2349
Generalization	718
Paraphrase	351
Tasks	120
Images	55

Figure 4: The number of submissions for each activity.

rogue data in other activities generally entered the correct sense of the words in the 'sense' activity.

Users that appeared to have hostile intentions, indicated by sentence completions that were of a crude or sexual nature, also submitted useful data. Surprisingly, a few users that might be classified as malicious were among the top contributors of good data.

Figure 4 shows the ranking of the various activities and the number of submissions. Not surprisingly, users spent significantly more time on the 'objects' activity. This is probably because the 'objects' activity requires the least amount of thought and because it was the only activity that involved images. Although users were allowed to submit their own images of indoor objects, very few users actually did. The second contest in February 2004, with more emphasis on Generalization, Paraphrase, and Tasks activities had much fewer submissions (about 2000). Fewer submissions may be due to higher degree of difficulty, thought, and typing required for submissions.

Feedback

Although we have received little feedback on the OMICS website, comments thus far have been largely positive. One of the weekly winners entered data with her seven-year-old son. She had the following to say about the site:

> As a teacher I think it is really great and put my son on it with me here at home—It was a great mom and kid project for several days. My little one who is 5 will be working on it too this week. It really forces us to do some critical thinking—the relationships and location section were great for him as were the free questions that he could come up with on his own.

Some users were concerned about their spelling errors and the spelling errors that were part of their sentence templates. Most grievous spelling errors were filtered out by the administrators, but some minor spelling errors were allowed into the database. Other users were concerned that the data they entered already existed in the database. One user commented:

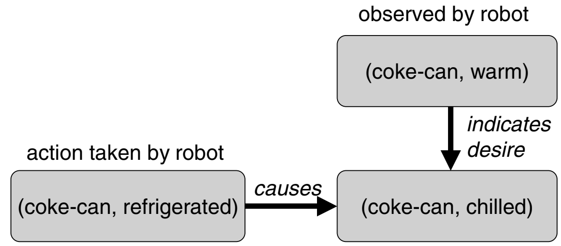

Figure 5: Example of using the OMICS knowledge base for active desire selection.

> I still don't really know what to put in the *freeform* category, or what needs to be worked through to be the most help in the project (I'm guessing something like *computer & mouse* gets overused, while other objects go ignored?).

The general tone of the e-mails we received were positive and indicated a desire by the users to see the project succeed in making robots more intelligent.

Data Applications

The data collected from the Open Mind family of projects have been applied to a wide variety of problems. Open Mind Word Expert has been applied to word sense disambiguation (Liu, Lieberman, & Selker 2003) and Open Mind Common Sense has been applied to textual affect sensing (Liu & Singh 2003). In this section, we describe two ways in which the data collected as part of OMICS is being used.

Active Desire Selection

Our knowledge base can be used for common sense and practical reasoning using Belief-Desires-Intentions (BDI) theory. BDI was originally developed by Bratman (1987) and is founded upon established theory of rational action in humans.

Given causality relations, observations, and human commands, the robot can use the desires relations from the OMICS knowledge base to deduce active desires (goals). These desires can then be used in action selection using the Belief-Desire-Intention (BDI) architecture (Rao & Georgeff 1995; Wooldridge 1999).

Figure 5 shows an example of active desire selection. Here the robot observes from its sensors that a coke can is warm. From the desires relations, the robot knows that (*coke-can*, *warm*) $\rightarrow_d$ (*coke-can*, *chilled*), and hence that it should pursue an action that causes (*coke-can*, *chilled*). The robot, however, does not know how to directly cause (*coke-can*, *chilled*), so it looks into its list of causality implications. The robot sees that (*coke-can*, *refrigerated*) $\rightarrow$ (*coke-can*, *chilled*). Thus it can infer that by refrigerating the warm coke-can, it can be chilled, and therefore the robot can add that desire to its list of active desires.

Topological Map with Room Labeling

Space can be labeled using terms that people typically use such as Large Conference Room, Small Conference Room,

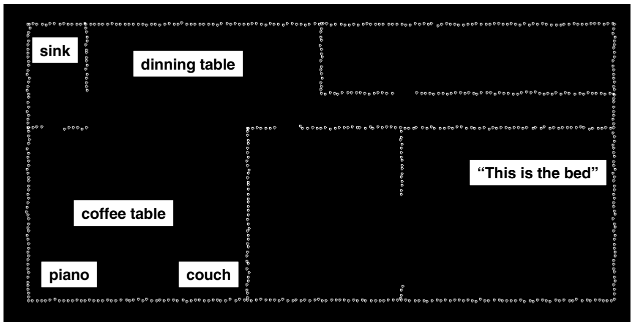

Figure 6: Room labeling showing input of simulated room and object labels at specific locations.

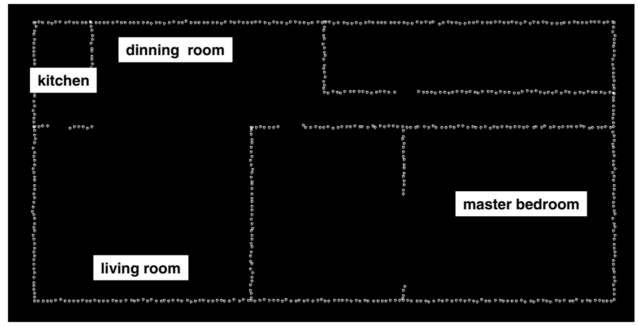

Figure 7: Room labeling showing output of our algorithm with most likely room labels and locations.

and Library. Information on the extent and connectivity of these regions can be used to generate a topological map with spatial labeling for providing natural interaction. This map can be used to plan a path to a specific room based on the current task.

We simulated data for a laser scanner to provide extent and connectivity 2D map information. Dots in figure 6 represent individual laser scan readings from a merged map generated during environment exploration. Humans provide a running description of the immediate surroundings as the robot explores a new home or office. These labels can be room types (e.g., *This is a kitchen* or *This is the living room*), or about objects (e.g., *This is a chair* or *This is a computer*). In addition, the robot might use an object recognition system to recognize objects with a confidence value given by the recognition system.

These simulated objects and room labels with associated probabilities (from a speech/object recognition system) and the 2D map are input to our system. We then use probability estimates given by statistical analysis of object location data from our knowledge base to label different rooms and open areas. We ran simulations to output topologically labeled maps of indoor home and office environments with different inputs.

In related work, Myers and Konolige (1992) generated and modified map layouts to reflect sentential information and common sense constraints for example:

- individuals own offices
- galleries or walkways are not owned by individuals
- reception is located at the entrance to the office

As shown in figure 6, we are given the walls and labels such as *This is the bed* and some objects recognized in the environment or pointed to by a user like sink, dining table, coffee table, piano and couch with an associated probability value. From this information, sensory data D, and the Open Mind Indoor Common Sense database we may compute the location probabilities.

These numbers can be used to compute the probability of the most likely room using Bayes' formula. Figure 7 shows the output of our algorithm. For example we have determined that based on available information, the most likely room label for the top-left room is the kitchen.

$P(\mathit{sink}\vert D) = 0.70$	(from object recognition)
$P(\mathit{microwave}\vert D) = 0.90$	(from speech recognition)
$P(\mathit{sink}\vert \mathit{kitchen}) = 0.13$	(estimated from OMICS)
$P(\mathit{microwave}\vert \mathit{kitchen}) = 0.11$	(estimated from OMICS)
$P(\mathit{sink}\vert \mathit{bedroom}) = 0.003$	(estimated from OMICS)
$P(\mathit{microwave}\vert \mathit{bedroom}) = 0.002$	(estimated from OMICS)

Further work will involve using laser data instead of simulated 2D maps. This data can also be extracted from cameras by building a sparse map of the unknown environment. Simulated place and object labels supplied by the user will be replaced by real data from a speech recognition system. Simulated object labels will be replaced by an object recognition system with associated confidence values.

Probability Estimation from common sense knowledge base The robot's sensory perceptions are combined with the priors and conditional probabilities for object location estimated from the common sense knowledge base to determine the most likely room label.

Given a set $\omega \in \Omega$ of rooms, objects x_i, and sensory data D, the robot sensory perception provides $P(x_i|D)$, and the location data in our knowledge base is used to estimate $P(x_i)$ and $P(x_i|\omega)$. The robot collects information about the objects in the room, perhaps through speech recognition (e.g. a human says "this is a chair") or through an object recognition system. These observations D induce a conditional probability distribution $P(x_i|D)$ over the objects $x_i \in X$. We wish to combine these probability distributions with $P(x_i|\omega)$ and $P(x_i)$ probability estimations from our knowledge base to calculate the room that is most likely, namely:

$$\omega' = \arg\max_{\omega\in\Omega} P(\omega|D)$$

Let $\mathbf{x}$ denote a vector indicating the presence or absence of the objects x_i. Assuming a generative Bayesian model where ω influences $\mathbf{x}$ and $\mathbf{x}$ influences D, we calculate the joint distribution:

$$\begin{aligned} P(\omega, \mathbf{x}, D) &= P(\omega)P(x|\omega)P(D|\mathbf{x}) \\ &= \frac{P(\omega)P(\mathbf{x}|\omega)P(\mathbf{x}|D)P(D)}{P(\mathbf{x})} \end{aligned}$$

We use this to calculate the most likely room, ω':

$$\arg\max_{\omega\in\Omega} P(\omega|D)$$

$$= \arg\max_{\omega\in\Omega} \sum_{\mathbf{x}} P(\omega, \mathbf{x}|D)$$

$$= \arg\max_{\omega\in\Omega} \sum_{\mathbf{x}} \frac{P(\omega)P(\mathbf{x}|\omega)P(\mathbf{x}|D)}{P(\mathbf{x})}$$

$$= \arg\max_{\omega\in\Omega} P(\omega) \sum_{x_1} \frac{P(x_1|\omega)P(x_1|D)}{P(x_1)} \cdots \sum_{x_n} \frac{P(x_n|\omega)P(x_n|D)}{P(x_n)}$$

$$= \arg\max_{\omega\in\Omega} \ln P(\omega) + \sum_{i=1}^{n} \ln \sum_{x_i} \frac{P(x_i|\omega)P(x_i|D)}{P(x_i)}$$

Since the x_i's are binary valued, the computational complexity is linear in the number of possible objects.

Our knowledge base contains a collection of tuples of objects and rooms. We estimate $P(x_i)$ by counting the number of times x_i is mentioned in the database and dividing by the number of entries in the database. We estimate $P(x_i|\omega)$ by counting the number of times the tuple (x_i, ω) appears and dividing by the number of tuples that mention ω, i.e. $P(x_i|\omega) = C(x_i, \omega)/C(\omega)$. However, this assigns zero probability to $P(x_i|\omega)$ in cases where the database never mentions the tuple (x_i, ω). This is rectified by using Lidstone's law to redistribute some of the probability mass assigned to the observed tuples to the unobserved tuples. Lidstone's law uses a parameter $\lambda < 1$ to control how much probability is distributed to unseen tuples. We then have

$$P(x_i|\omega) = \frac{C(x_i, \omega) + \lambda}{C(\omega) + \lambda n}$$

Unseen instances are assigned probability λ/n, where n is the number of objects.

Conclusions and Future Work

The Open Mind Indoor Common Sense project has successfully captured thousands of pieces of common sense knowledge about home and office environments. Our contributions to common sense data collection include the restriction of the domain to enhance the density of the knowledge, dynamic prompting of data based on prior data in the knowledge base, and object-centric data collection focusing on objects and their properties. We use comprehensive manual data review to ensure the quality of the collected knowledge.

The indoor home and office focus of our data collection and the structured activities have lead to a dense knowledge base. This knowledge base was useful in determining active desires from which actions could be selected. We computed probability estimates from our common sense knowledge base and used them in a Bayes formulation to compute the most likely room label given room and object labels with confidence levels. These labels were combined with a 2D map to build topologically labeled maps.

Distributed knowledge capture results in messy knowledge and one of the challenges is to convert this into a usable form. Further research will be required to clean up the data using statistical probability estimation techniques. Research is also required to represent, maintain and update this knowledge on a real robot.

Although the robot can use the common sense knowledge at a very high level to determine which desires to pursue, it is not yet intelligent enough to actually select and execute the mid-level actions that accomplish these desires. It remains to be seen how common sense can be used in the actual execution of various tasks, such as cleaning a bathtub. One might use the teleo-reactive program framework, as proposed by Nils Nilsson (1992; 1994), to accomplish such basic tasks as *make a cup of coffee hot* or *pop some popcorn*.

Acknowledgments

This work was done while Mykel Kochenderfer was a summer intern at Honda Research Institute USA, Inc. The authors would like to thank David Stork and Push Singh for their helpful comments and suggestions. Nils Nilsson reviewed an earlier version of this paper and provided invaluable feedback. Thanks are also due to anonymous reviewers and the users of the Open Mind Indoor Common Sense website for their data and feedback.

References

Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Liu, H., and Singh, P. 2003. OMCSNet: A commonsense inference toolkit. Technical Report SOM02-01, MIT Media Lab Society Of Mind Group.

Liu, H.; Lieberman, H.; and Selker, T. 2003. Exploiting agreement and disagreement of human annotators for word sense disambiguation. In *Proceedings of the Seventh International Conference on Intelligent User Interfaces (IUI 2003)*, 125–132.

Miller, G. A. 1995. WordNet: A lexical database for english. *Communications of the ACM* 38(11):39–41.

Myers, K., and Konolige, K. 1992. Reasoning with analogical representations. In Nebel, B.; Rich, C.; and Swartout, W., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR92)*. San Mateo, CA: Morgan Kaufmann Publishers Inc.

Nilsson, N. J. 1992. Towards agent programs with circuit semantics. Technical Report STAN–CS–92–1412, Department of Computer Science, Stanford University, Stanford, CA 94305.

Nilsson, N. J. 1994. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research* 1:139–158.

Rao, A. S., and Georgeff, M. P. 1995. BDI-agents: from theory to practice. In *Proceedings of the First Intl. Conference on Multiagent Systems*.

Stork, D. G. 1999. The Open Mind Initiative. *IEEE Expert Systems and Their Applications* 14(3):19–20.

Stork, D. G. 2000. Open data collection for training intelligent software in the open mind initiative. In *Proceedings of the Engineering Intelligent Systems (EIS2000)*.

Wooldridge, M. 1999. Intelligent agents. In Weiss, G., ed., *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA, USA: The MIT Press. 27–78.