Proceedings of the ICORR 2003(The Eighth International Conference on Rehabilitation Robotics), 22-25 April 2003

Use of Machine Learning based on Constructive Induction in Dialogs with Robotic Heads

Marek Perkowski [+*], Tsutomu Sasao[&], Atsumu Iseno[&], Uland Wong [$], Mikhail Pivtoraiko[*], Michele Folgheraiter [#], Martin Lukac[*], David Ng[^], Miranda Fix[*], Karl Kuchs[*]

+ Dept. of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu,Daejeon 305-701, Korea, & Department of Computer Science and Electronics, Kyushu Institute of Technology, Iizuka, Fukuoka 820, Japan, $ Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA, * Department of Electrical and Computer Engineering, Portland State University, Portland 97207-0751, Oregon, USA, #DEI Department, Politecnico di Milano, piazza Leonardo da Vinci 32, Italy, ^ Kit Model Design and Production Limited, Workshop No.5,3/F, Decca Industrial Centre, 12 Kut Shing Str. Chai Wan, Hong Kong

Abstract

Paper presents our research progress and experiences related to designing inexpensive natural-size humanoid caricature and realistic robot heads - actors for robot theatre, agents for advertisement, education and entertainment. We concentrate on Machine Learning techniques used to teach robots behaviors, natural language dialogs and facial gestures.

1. Introduction

Future robots that will work in human environments will be quite different from industrial robots. They will interact closely with non-sophisticated users, children and elderly, so the question arises, how they should look like? It is both a result of public opinion pools and our own observations that the robot should look more-or-less like a human, which property will enable users to understand its intentions and program its behaviors in natural and simple ways [8,9]. If human face for a robot, then what kind of a face? Handsome or average, realistic or simplified, normal size or enlarged [7,10]? The most famous example of a robot head is Kismet from MIT [9]. Why is Kismet so successful? We believe that a robot that will interact with humans should have some kind of "personality" and Kismet so far is the only robot with "personality". Everybody who has played and observed closely Kismet will testify to this. How to build robots that will be equipped with this kind of "personality"?

Moreover, we believe that a robot face should be not only friendly but also funny. The Muppets of Jim Henson [6] are hard to match examples of puppet artistry and animation perfection – we intend to build robots with this kind of looks and behaviors. Since 1999 we have been building animatronic humanoid robot heads for interactive robot puppet theatre [1], tourist guides and robotic foreign language teachers. While Kismet is a half-child-half-animal [11], we try to animate various kinds of humanoid heads with 4 to 12 DOF, expecting comical and entertaining values. Because it is difficult to build a humanoid face with more DOF, we are building several variants of head kinematics, and we are experimenting with animation of simpler heads on our path to a ultimate prototype with about 20 DOF. We are less interested in beauty and accuracy [8] and more in robot's personality as expressed by its behavior, facial gestures, emotions and learned speech patterns. Our robots have different capabilities: BUG (see below) is equipped with learning based on robot vision, and Professor Perky with automated speech recognition (ASR) and text-to-speech (TTS) capabilities. The newest robot, Maria (Fig. 4), will have both vision and speech faculties.

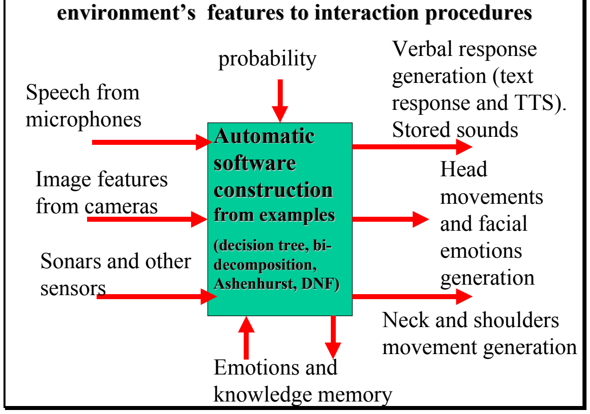
2. Hardware Design, Speech Recognition and Synthesis

We use inexpensive servos from Hitec and Futaba, plastic, playwood and aluminum for construction. The robots were either PC-interfaced, use simple micro-controllers such as Basic Stamp, or were radio controlled from a PC or by the user. Previous heads were equipped with microphones, USB cameras, sonars and CDS light sensors. Face Maria was designed by David Ng and is commercially available in two variants. Image processing and pattern recognition uses software developed at PSU, CMU and Intel (public domain software available on WWW). We compared several commercial speech systems from Microsoft, Sensory and Fonix. Based on experiences in highly noisy environments and with a variety of speakers, we selected Fonix for both ASR and TTS for Professor Perky and Maria robots, but in future we plan to use Vorero from Asahi Kasei because of its high quality operation in noise. We use microphone array from Andrea Electronics. Software is in Visual C++, Visual Basic, Lisp and Prolog [12].

3. Behavior, Dialog and Learning

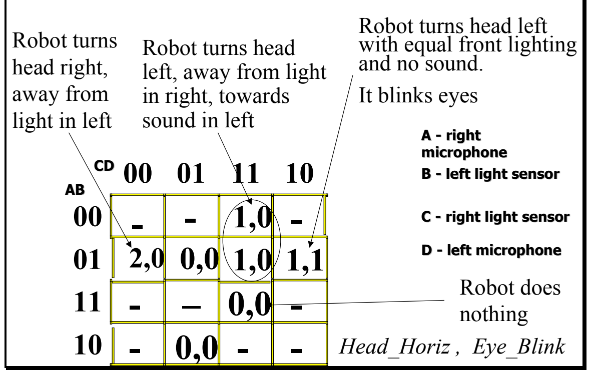
General diagram of our system is shown in Figure 1. We view robot activity as a mapping of the sensed environment and internal states to behaviors and new internal states (emotions, energy levels, etc). Our goal is to uniformly integrate verbal and non-verbal robot behaviors. Let us remember that words communicate only about 35 % of the information transmitted from a sender to a receiver in a human-to-human communication. The remaining information is included in: body movements, face mimics, gestures, posture, external view - so called para-language. Emotions, thoughts, decision and intentions of a speaker can be recognized earlier than they are verbalized. In brief, the dialog/behavior have the following components: (1) Eliza-like natural language dialogs. Commercial products like Memoni, Dog.Com and Heart, Alice and Doctor all use this technology very successfully – for instance Alice program won the 2001 Turing competition. This is a "conversational" part of the robot brain, based on pattern-matching, parsing and black-board principles. It is also a kind of "operating system" of the robot, which supervises other subroutines. (2) Subroutines with logical data base and natural language parsing (like CHAT). This is the logical part of the brain used to find connections between places, timings and all kind of logical and relational reasonings, such as answering questions about Japanese geography. (3) Use of generalization and analogy in dialog on many levels. Random and intentional linking of spoken language, sound effects and facial gestures. Use of Constructive Induction approach to help generalization, analogy reasoning and probabilistic generations in verbal and non-verbal dialog, like learning when to smile or turn the head away from the human. (4) Model of the robot, model of the user, scenario of the situation, history of the dialog, all used in the conversation. (5) Use of word spotting in speech recognition rather than single word or continuous speech recognition. (6) Avoidance of "I do not know", "I do not understand" answers from the robot. Our robot will have always something to say, in the worst case, over-generalized, with not valid analogies or even nonsensical and random.

Fig. 1. Learning Behaviors as Mappings from environment's features to interaction procedures



The feature values in Figure 1 are extracted from any kind of sensors such as inexpensive USB cameras, switches, microphones, sonars, etc. [2]. They are stored in a uniform notation of tables [5]; with rows corresponding to examples and columns to feature values of both input variables (sensors, text) and output variables (servos, actuators, TTS). This is a standard format in logic synthesis, Data Mining and Machine Learning. Such tables are created by encoding in the uniform way the data coming from the feature-extracting subroutines based on speech recognition, sensors and image processing. The tables have control parameters given to servos and TTS as their output data. Thus the tables store mapping information. If the input sample used in teaching is encountered again in table's evaluation, the same exactly output data from the table is produced as found in teaching (rote learning). But what if a new input data is given during evaluation, one that has never appeared before? Here the system makes use of analogy and generalization based on constructive induction.

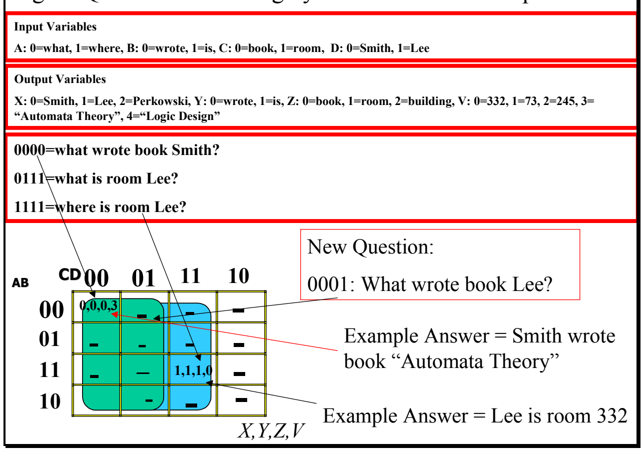
Fig. 2. Seven examples (4-input, 2 output minterms) are given by the teacher as correct robot behaviors



We will explain this concept on simple examples. Data in tables are stored as binary, and in general muti-valued, logic values (0, 1, 2, etc). The teaching examples that come from preprocessing are stored as multi-output (care) minterms (i.e. combinations of input/output variable values – Fig. 2). For illustration, we use here Karnaugh maps as data, but Binary Decision Diagrams and quite sophisticated logic synthesis algorithms such as Ashenhurst/Curtis hierarchical decomposition and Bi-Decomposition algorithms are used in our software [2,3,4,5]. These methods create a subset [2] of MVSIS system (developed under Prof. Robert Brayton, University of California at Berkeley). The entire MVSIS system can be also used. The system generates robot's behaviors (C program codes) from user-given examples. Berkeley's system is used for embedded system design, but we use it specifically for robot interaction. It uses a comprehensive Machine Learning/Data Mining methodology based on constructive induction and particularly on hierarchically decomposing decision tables of binary and multiple-valued functions and relations to simpler tables, until tables of trivial functions that have direct counterparts in behavior components are found. In contrast to Neural Nets or reinforcement learning, the constructive induction methods are based on logic, combinatorial optimization, and highly efficient data structures such as Binary Decision Diagrams [3,4,5]. Our approach can be thought of as a generalization of Decision Trees known from DM, ML and Robotics. It is similar to methods used in design automation of digital logic and it satisfies Occam Razor principle. In terms of quality (not speed) of learning it compares very well to NNs and Decision Trees in experiments [5].

A unified internal language is used to describe behaviors in which text generation and facial gestures are unified. Expressions (programs) in this language are either created by humans or induced automatically from user examples. Our approach includes deterministic, induced and probabilistic grammar based responses called from the language procedures. Figure 2 explains how controls for various behaviors can be automatically created from user-provided examples by using logic synthesis. In these examples we use simple DNF-like minimization of binary-input multiple-valued output functions only to illustrate the idea without going into technical methods that we use. The Kmap expresses controls of motors *Head_Horiz* for rotation of head and *Eye_Blink*. Value 2 is for right, 1 for left and 0 for no movement. Minimal mv expression found from the Kmap is converted to C code describing behavior of the head rotating and blinking eyes as a function of signals from left and right microphones and light sensors corresponding to binary inputs *A, B, C* and *D*.

Fig. 3. Question Answering by induction of answer parameters.



Question-answering example illustrates that exactly the same techniques and representations can be used to natural language processing and translation. The input sentences are encoded in multiple-valued logic and the output sentences are generated as a results of logic-synthesis-based generalization. There are several variants of this language learning technology, and here we use only the simplest possible example. Based on two examples the system infers incorrect but reasonable answer to the third question – *"What wrote book Lee?"* Observe that the answer will depend on the logic minimization procedure and MV prime implicants like *C'* or *D* found in this particular case. There are many possible answers to induce in this case, including 1000 = *"Lee wrote book 332"*. Example illustrates thus also certain randomness, that always exists in all generalization-based methods such as Constructive Induction. Various texts are generated as results of different runs and logic minimization approaches. This method automatically generalizes answers in case of insufficient information. This is only a simplified example that does not take into account parsing and syntax. Variable values are only parameters that are instantiated to correct syntax generators – so ungrammatical sentences as in Fig. 3 are avoided. If robot's response to question 0001 = *What wrote book Lee?* is wrong, it can be corrected by inserting to cell 0001 of Kmap a correct care minterm value, like 1004 = *Lee wrote book "Logic Design"*, which means, row 0001 1004 to the table.

4. Conclusion. What did we learn

We designed and evaluated several early prototypes of robot heads. The main principles that we learned in this process are: (1) the more degrees of freedom the better the animation realism, (2) synchronization of spoken text and head (especially jaw) movements are very important, lessons from puppet theatre should be learned, (3) the eyes, jaws, and other movable head components should be of an exaggerated size rather than of natural size; similarly to achieve better interaction, gestures and speech intonation of the head should be slightly exaggerated, (4) the sound should be laud to cover noises coming from motors and gears and for a better theatrical effect, (5) noise of servos can be also reduced by appropriate animation and synchronization, (6) best available ATR and TTS packages should be applied, especially those that use word spotting, (7) the designer should look to new materials and in general learn from puppet theatre experiences in many areas, such as puppet design, stage design and audio, movement animation, speech, facial gestures, (8) because of a too slow learning, improved parameterized learning methods will be developed, but also based on constructive induction, (9) although robot face should be funny, it should not be ugly – based on these experiences with Professor Perky, we build now a new robot Maria.The bi-decomposer of relations and other useful software used in this project can be downloaded from http://www-cad.eecs.berkeley.edu/mvsis/. One of our future goals is the "Turing Test for Robotic Puppets": *"can the audience distinguish if the robot head is animated remotely by a human or is its language/gesture behavior totally autonomous?"*

5. Literature

1. M. Perkowski, Oregon Cyber Theatre, *Proc. 3rd Oregon Symposium on Logic, Design and Learning*, May 2000.
2. U. Wong and M. Perkowski, A New Approach to Robot's Imitation of Behaviors by Decomposition of Multiple-Valued Relations, *Proc. 5th Intern. Workshop on Boolean Problems*, Freiberg, Germany, Sept. 19-20, 2002, pp. 265-270.
3. A. Mishchenko, B. Steinbach and M. Perkowski, An Algorithm for Bi-Decomposition of Logic Functions, Proc. DAC 2001, June 18-22, Las Vegas, pp. 103-108.
4. A. Mishchenko, B. Steinbach and M. Perkowski, Bi-Decomposition of Multi-Valued Relations, Proc. 10th IWLS, pp. 35-40, Granlibakken, CA, June 12-15, 2001. IEEE Computer Society and ACM SIGDA. M.
5. Perkowski, and S. Grygiel, Decomposition of Relations: A New Approach to Constructive Induction

in Machine Learning and Data Mining – An Overview, Proc. Workshop of National Inst. Of Telecommunications and Polish Association for Logic and Philosophy of Science, May 25, 2001, pp. 91-111, Warsaw, Poland.
6. Ch. Finch, Ch. Finch and J. Henson, Jim Henson: The Works: The Art, The Magic, the Imagination, see also other books by and about Jim Henson on Amazon.com.
7. A. Edsinger, U-M O'Reilly, and C. Breazeal, A Face for a Humanoid Robot, MIT Memo, 2002
8. F.Hara, H.Kobayashi, F.Iida, and M. Tabata, Personality characterization of animate face robot through interactive communication with Human, Proc. 1st Int'l Workshop on Humanoid and Human Friendly Robots, pp. 1-10, 1998.
9. C. Breazeal, Sociable Machines: Expressive Social Exchange Between Humans and Robots*, Ph.D. Thesis,* MIT, 2000.
10. Edsinger and U-M. O'Reilly, Designing a Humanoid Robot Face to Fulfill a Social Contract*, MIT Memo,* 2002
11. C. Breazeal and B.Scassellati, Infant-like social interactions between a robot and a human caretakes, In *Adaptive Behavior,* August 2000.
12. M. Perkowski et al, "Listening and Talking Robot Heads," KAIST Technical Report, 2003.

Our robot heads chronologically, from left to right, top-down: Furby head with new control, Alien, Skeleton, Mister Butcher, Jonas, Adam, Marvin the Crazy Robot, Max of Mark Medonis, Virginia Woolf, BUG (Big Ugly Robot), Professor Perky and Maria. Arrows in last photo of Maria (no hair) show locations of servos and control rods.