Towards context-sensitive dialogue with robot companion

Jüri Vain
Department of Computer Science, TUT
Tallinn, Estonia
juri.vain@ttu.ee

Helena Sarapuu
ELIKO Tehnoloogia Arenduskeskus OÜ
Tallinn, Estonia

***Abstract*— Novel principles of implementing human-robot dialog using logic programming framework are outlined. Our inference mechanism makes use of semantic knowledge base and context sensitive clausal reasoning. Proposed normal forms and weak unification relation support efficient back- and forward chaining for dialog interpretation/control. Context driven dialog implementation is shown on autistic children training example.**

Keywords—human-robot interaction; emotional context; knowledge representation; clausal reasoning, weak unification

I. Introduction

The pilot project "Robot Companions for Citizens" in the European FET Flagship Program sets the future goals of cognitive robotics. In its *Manifesto* [1] it outlines "in order to assist humans, and to interact physically, socially and safely with them, new generation of sentient robots so called Robot Companions (RC) are needed". Sentience is the ability to integrate perception, cognition and action in one coherent scene and context in which action can be interpreted, planned, generated and communicated. In particular, sentience reveals in human-robot (HR) dialog that requires RCs to be cognizant of, reason about, and respond appropriately to the needs of people. *Sentient computing* is a form of ubiquitous computing that constructs a world model which allows location-aware and context-aware application of robot cognitive and functional capabilities. The need for RCs is capitalized for elderly and physically disabled people, autistic children, and for assisting as a guide at exhibitions and public buildings.

Contemporary dialog systems are mature enough to allow simple conversation in many applications: small talk with toy robots [2], voice command control of home appliances [3], robot guides [4], emergency reporting and control in power plants [5], etc. On the other hand, in most cases these dialog systems support conversation on very limited semantic domains almost without dynamic context switching and online learning capabilities. Specially, driving casual dialog through multiple contexts is still a long term research challenge to be studied by means of interdisciplinary approaches (AI, psychology of verbal communication, speech technology, etc).

The contribution of this paper is two-fold. At first, we provide a rule-based logic programming (LP) framework for building HR dialog core scenarios and for direct dialog control. Its strength lies in the context manipulation and knowledge representation rule structure that provides flexible dialog adaptation and extension mechanisms. Second contribution that is implemented by means of the dialog framework is integrating the dialog syntax and semantic rules with scenarios and dynamic contexts, including those that respond to estimates of human emotional states. Especially the latter provides opportunity for studying how RC can adapt to and influence the emotional state of humans in the course of dialog. As an application example we demonstrate how the rules and dialog schemes can be instantiated for an emotional context and intention when communicating with autistic kids.

The rest of the paper is structured as follows. Section 2 defines generic dialog structures and components needed for dialog control. Section 3 describes knowledge structures needed for scalable implementation of the dialog rule sets. Section 4 outlines clausal reasoning schemes for knowledge extraction/derivation from dialog and stored knowledge, and provides novel *weak unification* relation that improves the reasoning power on RC world model. In Section 5 the context manipulation operations are introduced and the dialog schemes driven by emotional contexts are demonstrated.

II. General structure and components of the human-robot dialog

Main component of the HR dialog systems (Fig. 1) is Dialog Manager that controls the state and flow of the dialog. Its activity cycle contains following phases [6]:

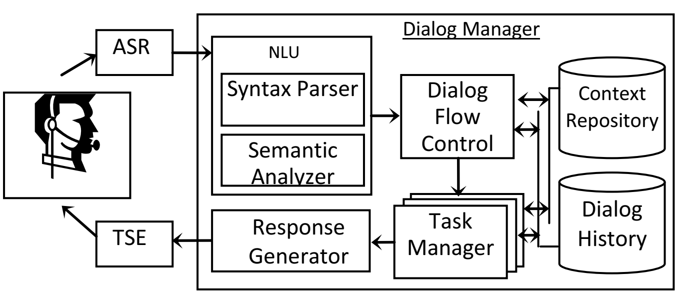

Fig. 1. *Human-robot dialog system architecture*

1. The user speech input is converted to text by the input recognizer that may include Automatic Speech Recognizer (ASR), gesture recognizer, and/or handwriting recognizer.
2. The text is analyzed by a Natural Language understanding Unit (NLU), which may include name identification, speech tagging, Syntactic Parser and Semantic Analyzer.
3. The semantic information is analyzed by Dialog Flow Control that keeps the Dialog History and state of the dialog and manages the general flow of the conversation.
4. The dialog manager activates Task Managers that operate using rule sets of specific contexts (Context Repository).
5. The dialog manager produces output using Response Generator, which may include natural language generator, gesture generator, and/or layout engine.
6. The output is rendered using an output renderer, which may include text-to-speech engine (TSE).

In the rest of paper we focus on steps 2 to 5 that contain components of the Dialog Manager, but at first, following general assumptions are made about the HR dialog.

Only one human and one robot (the dialog actors) are communicating at a time. They both are quiescent and non-interfering. It means that the dialog actors exchange phrases alternatively so that the number of phrases never exceeds some upper limit in each turn. We assume the human says only one phrase at a time and robot one or more phrases the total duration is not more than 10 seconds. Each phrase must be interpretable with limited number of contexts and related syntactic/semantic rule sets.

About the contexts needed for phrase interpretation and response generation we require the following:

- *Consistency* - no contradiction can be derived by means of rule sets of contexts that hold in the same time frame.
- *Minimality* - there must be at least one non-empty context active at a time to ensure responsiveness (no blocking/unbounded chaining while interpreting phrases and/or generating responses).
- *Completeness* and *connectedness of the universe of discourse* - active contexts include rule sets that cover necessary reasoning sequences. Incomplete contexts cause gaps in inference chains and the need for additional learning (robot has to question the human).

Common contexts are emotion-, space time, goal and situation related contexts that may combine in different types of dialog (emphatic, insisting, inquiring, and obeying dialog).

III. Knowledge base for Human-Robot dialog

Our HR dialog framework is built on the principles of logic programming (LP) and rule-based reasoning. The rules are defined as statements (Horn clauses [7]) of form $A \leftarrow B_1, B_2, \dots, B_n$, where $n \geq 0$, A denotes conclusion and $B_1, B_2, \dots, B_n$ conditions that in conjunction imply A. A logic program is a finite set of clauses. A clause is interpreted procedurally "To answer query A, answer the conjunctive query $B_1, B_2, \dots, B_n$". An existentially quantified goal G is a logical consequence of a program P if there is a clause in P with a ground instance $A \leftarrow B_1, B_2, \dots, B_n$, $n \geq 0$ such that $B_1, B_2, \dots, B_n$ are logical consequences of P, and A is an instance of G. A clause is called *fact* if $n = 0$, meaning that A is unconditionally true. A *query* is a conjunction of the form $A_1, A_2, \dots, A_n$? where $n \geq 0$ and A_i are goals. Variables in the query are understood to be existentially quantified. A computation of a logic program P finds an instance of a given query logically deducible from P. The set of general and application related rules, facts and queries constitutes the application knowledge base.

Our HR dialog knowledge base consists of rules of following categories: grammar rules for parsing syntactic structure of the input text (these rules are specified in Definite Clause Grammar [7]); lexical rules for grouping natural language words and expressions that have semantically similar meaning or pragmatically similar roles in the sentence; morphological rules to specify the relations between word categories; rules that encode semantic entities and relations between them, e.g. class-instance ("is_a") relation, causal/spatial/temporal relations, etc. that are determined by a conceptual meta-model of the universe of discourse; dialog scenario control rules; and dialog context manipulation rules.

Due to the large variety of rules needed in the dialog related reasoning we introduce *normal form* of clauses to minimize the knowledge base and to make the clausal reasoning more efficient. We implement the LP rules in Prolog syntax [8] of form "*Head* :- *Body*.", where *Head* denotes the conclusion predicate and Body condition predicates of the Horn clause. The format of normalized clauses knowledge/7 where 7 is the number of parameters (possibly empty lists - []) is following:

knowledge(A_n, CL_n, A_{inx}, CL_{inx}, A, CL, $CtxtL$). , where

A_n - conclusion of the Horn clause in normal form;
CL_n - list of conditions of the Horn clause in normal form;
A_{inx} - list of indexes of grammatical forms of conclusion;
CL_{inx} - indexes of grammatical forms of conditions;
A - the original form of the conclusion of the Horn clause;
CL - the original conditions of the Horn clause;
$CtxtL$ - indexes of contexts where the clause is applicable.

The normal form allows (*i*) unification of semantically close clauses that differ by their grammatical and word forms; (*ii*) computing their semantic distance function that is needed for weak unification process as will be described in Section IV.

Our RC target language is Estonian where normalization of phrases comprises two steps: (*i*) substituting the words with their base forms and indexes of their inflections; (*ii*) substituting base forms with their representatives in synonyms group. The clause is provided also with indexes of contexts where this clause is valid, e.g. the unconditional phrase A = "nemad töötasid seppadena" (in English "they worked as blacksmiths") in non-restrictive context 0 has normal form:

knowledge([nad, töötama, sepp], [], [1, 10, 26], [], A, [], [0]).

IV. Clausal reasoning with weak unification

Clausal reasoning with resolution [7] applies inference that leads to a refutation theorem-proving technique for sentences in propositional logic and first-order logic. In other words, iteratively applying the resolution rule in a suitable way allows for telling whether a propositional formula is satisfiable and for proving that a first-order formula is unsatisfiable.

There are two types of reasoning with rules - forward and backward chaining. Forward chaining matches the premises of each rule against the known facts to derive new facts. It is also known as “forward reasoning” or “data driven” reasoning. Backward chaining tests whether a given hypothesis is true by finding a rule which has the hypothesis as the conclusion, and testing its premises. It is also known as “backward reasoning”, or “goal-driven” reasoning. Both forward and backward chaining reasoning schemes have direct use in our dialog system. Forward reasoning deduces answers to questions of the form "What is the consequence of…" or "What if …?", where "…" denotes a conjunction of unconditional clauses. Backward chaining returns answers to questions of form "Why...?", "What is the cause of …?".

To apply the rules of LP reasoning we need to match (unify) terms of spoken phrases against the components of rules (premises, conclusions). Unfortunately, applying abovementioned reasoning schemes with standard unification rule (implemented by MGU algorithm [7]) is impractical due to the need for unreasonably many clauses in RC knowledge base. That is because of large variety of syntactical forms and their loose use in spoken language.

To match the spoken phrases with clauses of RC knowledge base we have extended the meaning of unification by adding semantic distance measure (similarity on scale 0 - 100) between them. That provides a possibility to define the parameter "similarity threshold" that is needed for deciding whether two phrases match semantically or not. We call it *weak unification* relation between clauses and use in both backward and forward clausal reasoning. The similarity function has two clauses *A* and *B* as arguments and it returns the maximum inclusion percent of semantic units of one clause in the other. Semantic inclusion is computed by finding the matches of semantic units of the first clause in the second clause. The maximum is taken over both orderings (*A*, *B*) and (*B*, *A*) of phrases. Two semantic units have similarity value 100 if they are synonyms or one is the instance (by "is_a" relation) of other. Otherwise, the semantic similarity between them is calculated as inverse of topological distance between the nodes of the Semantic Network the RC knowledge base constitutes.

By means of weak unification the clauses of RC knowledge base are searched for the responses to questions with inquisitive words "Where…?", "When…?", "How long…?", "What…?", etc. All these questions refer to some word or phrase in the knowledge base clauses that is semantic match to that inquisitive word. For instance, if the body of When…? question weakly unifies with some knowledge/7 fact the non-matching part of that fact is analyzed further for semantics matching with the inquisitive word and returned as an answer if it passes the semantic filter defined for given inquisitive word.

V. Clausal reasoning with Contexts

A. *Representation of dialog contexts*

The aim of HR dialog control is to make it close to natural human dialog. An important prerequisite for it is the adaptivity of the RC's reasoning mechanism to dialog contexts. In casual dialog the number of contexts may be large and they may change erratically in the course of conversation. To capture the current context and to define the context shifts adequately we propose the context representation to be knowledge structure that conforms to the Meta-Model (MM) depicted in Fig. 2.

The dialog manager keeps track of the contexts by matching the interpretations of phrases with the models of active contexts (instances of the meta-model). We say that the context changes in the course of dialog whenever the quantities observable by RC do not match with attributes of some current context. In fact, proper context sensitivity of RC needs wider environment monitoring and reacting also to sensor data and visual information. Due to the limited space of current paper we focus on the context changes observable in the dialog only.

B. *Context manipulation*

Contexts are either "hard coded" or learned by RC in the course of dialog. In both cases the knowledge is structured according to the context meta-model of Fig. 2. The contexts (instances of MM) are specified using `knowledge/7` clauses. Explicit inheritance relation between entities of MM and their instances allows applying same reasoning mechanisms over different levels of abstraction. Additionally, we have context signature clause `context/3 used` for context status monitoring and switching. These clauses have the role of defining the scope of applicability of knowledge and selecting the rules of inference most relevant in given context.

Dialog manger keeps the contexts up-to-date and chooses the right context(s) for current dialog by using the parameters of `context/3`: (*i*) context identifier/index, (*ii*) context (dynamic) priority, and (*iii*) the list of parent contexts (we allow polymorphic inheritance of contexts). The context priority parameter is refreshed when the context reuse is detected. The priority is increasing when the context is used, and decreasing otherwise. The dialog manager removes the inference rules of low priority contexts and loads the ones of active contexts in Prolog workspace.

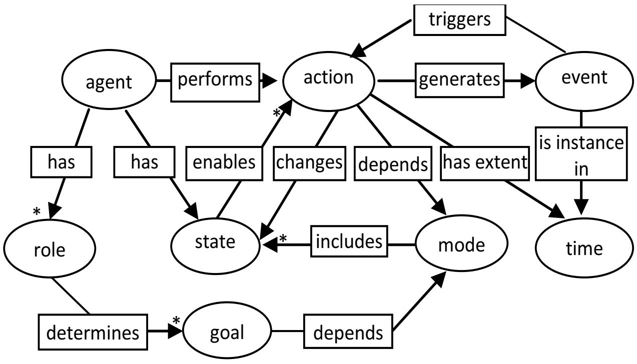

Fig. 2. *Agent's situation contexts meta-model*

This research has been supported by European Union through European Regional Development Fund

VI. Case-study: Emotional context Driven dialog with autistic children

The experiments with humanoid robot NAO have shown promising results in the use of autism therapy [9]. Autism spectrum disorder (ASD) denotes a group of complex disorders of brain development. ASD can be associated with intellectual disability, difficulties in motor coordination and attention and physical health issues. Treatment options for children include therapeutic activities for at least 25 hours per week. The therapy is guided by specific and well-defined learning objectives, and the child's progress in meeting these objectives need to be regularly evaluated and recorded.

As a case-study we develop the training aid for constructional abilities using a system [10] for constructing geometrical patterns from simpler shapes. NAO has a role of instructor who can observe the progress of task completion and guide it by means of dialog. In addition to speech input NAO gets data that support decision on emotional states from patient's affective physical condition monitoring system: skin conductance, heart rate and temperature. As demonstrated experimentally in [11] the variations of physiological characteristics correlate with human emotional states.

NAO's dialog manager is following two types of contexts: emotional state of the patient and the progress of completing the task. Combinations of these contexts dictate what rule sets of the dialog need to be activated. Current context is decided based on the monitored physiological characteristics and patient's oral responses. The progress of task completion is evaluated by the training system [10] as the ratio of correct and incorrect moves within a preset time intervals.

When inferring the speech command/response in the course of training session NAO uses Prolog clauses of form:

`knowledge(...,Phrase_i,...,Context_j):-`
`      Cond1, ..., Cond`$_n$`.`

where `Cond1,...,` `Cond`$_n$ denote conditions when the phrase `Phrase_i` can be used in the contexts specified in the parameter `Context_j`. When the conditions `Cond`$_k$ k=1,...,n need to refer to some earlier phrase in the dialog the history facts are referred in the terms of `Cond`$_k$. For instance the query "?- `history(Phrase)`." returns the latest phrase recorded, and "?- `nth_clause(history,N, Phrase_n)`." returns n-th latest phrase that unifies with `Phrase_n`.

Finally, we demonstrate how a fragment of emotion driven dialog between NAO (*N*) and Child (*C*) is coded in Prolog (to keep the dialog simple only closed questions are used):

N: Are you feeling well? % PhraseN_1
C: No. % PhraseC_2
N: Are you bored? % PhraseN_3
C: Yes. % PhraseC_4
N: Please continue with the task. % PhraseN_5
C: I don't want to. *(confronting)* % PhraseC_6

Alternative emotional context dependent responses:

N: It's necessary to complete the task so that you could continue with playing. *(tutoring)* % PhraseN_71
N: You're doing great, go on. *(encouraging)* % PhraseN_72
N: We've done enough for today. Let's play now.% PhraseN_73

Detectable contexts are:

- Emotional state: cooperative| bored| confronting.
- Progress: satisfiable| slow| completed .

```
knowledge(…,PhraseN_1,…,[bored,slow]).
knowledge(…,PhraseN_3,…,[ bored,slow]):-
        nth_clause(history(PhraseN_1),2,_),
        nth_clause(history(PhraseC_2),1,_).
knowledge(…,PhraseN_5,…,[_,satisfiable]):-
        nth_clause(history(PhraseN_3),2,_),
        nth_clause(history(PhraseC_4),1,_).
knowledge(…,PhraseN_71,…,[confronting,slow]):-
        nth_clause(history(PhraseN_5),2,_),
        nth_clause(history(PhraseC_6),1,_).
knowledge(…,PhraseN_72,…,[cooperative,slow]):-
        nth_clause(history(PhraseN_5),2,_),
        nth_clause(history(PhraseC_6),1,_).
knowledge(…,PhraseN_73,…,[bored,completed]):-
        nth_clause(history(PhraseN_5),2,_),
        nth_clause(history(PhraseC_6),1,_).
```

VII. Conclusion and future work

We have shown that constructing and implementing a dialog system for robot companions presumes involvement of cognitive psychology aspects into the mechanisms of constraint programming. Proposed LP framework makes use of semantic knowledge base and context sensitive clausal reasoning with several novelties such as normal forms of clauses, weak unification relation that both support efficient back- and forward chaining for dialog interpretation/control. Context driven dialog implementation example is shown on ASD children training case study. Before proposing the solution to clinical practice an extensive set of tests needs to be conducted, e.g. to validate if the emotional state detection is adequate enough for children of different psychological types, if there are more comprehensive ways of detecting the emotion (by voice tone and intensity, prosody), etc.

References

[1] Accessed online at http://www.robotcompanions.eu/, 2014-05-04.

[2] T.Bickmore, "A computational model of small talk", 1999. Accessed online at http://media.mit.edu/, 2014-05-04.

[3] A.K. Gnanasekar, P. Jayavelu, V. Nagarajan, "Speech recognition based wireless automation of home loads with fault identification for physically challenged". Proc. of ICCSP, pp.128 - 132, 4-5 April 2012.

[4] Accessed online at http://japan.gadgetmeet.com/Meet-Tawabo-new-quadrilingual-guide-robot-at-the-Tokyo-Tower/#more-2990, 2014-05-04

[5] M. Waibel, M. Beetz, J. Civera, et al. "RoboEarth," in Robotics & Automation Magazine, IEEE , vol. 18, pp. 69 - 82, Issue: 2, June 2011.

[6] D. Jufarsky and J. H. Martin, Speech and language processing. Pearson International Edition, 2009, Chapter 24.

[7] L. Sterling, E. Shapiro, The art of prolog, Advanced Programming Techniques, 2nd ed, MIT Press, 1994.

[8] Swi-prolog homepage: http://www.swi-prolog.org/, 2014-05-04.

[9] S.Shamsuddin, et al. "Humanoid Robot NAO Interacting with Autistic Children of Moderately Impaired Intelligence to Augment Communication Skills," Procedia Engineering 41, 2012, pp. 1533-1538.

[10] O. Miyashita, S. Tsukamoto, A. Kuusik, D. Miyata, T. Yoshida, J. Vain, and S. Ishigami, "A Human-Adaptive-Mechatronics Assisted System for Training Constructional Ability," in 8th International Conference on Mechatronics Technology (ICMT2004), November 8 - 12, 2004, Hanoi.

[11] L. Nummenmaa, E. Glerean, R. Hari, and J. K. Hietanen, "Bodily maps of emotions," Proceedings of the National Academy of Sciences of the United State of the America, pp. 646–651, January 14, 2014.