

From Robust Spoken Language Understanding to Knowledge Acquisition and Management

Luís Seabra Lopes, António J. S. Teixeira, Marcelo Quinderé, Mário Rodrigues

IEETA, Departamento de Electrónica e Telecomunicações
Universidade de Aveiro, 3810-193 Aveiro, Portugal
{lsl, ajst, marcelo, mfr}@ieeta.pt

Abstract

The recent evolution of Carl, an intelligent mobile robot, is presented. The paper focuses on robust spoken language understanding (SLU) and on knowledge representation and reasoning (KRR). Robustness in SLU is achieved through the combination of deep and shallow parsing, tolerating non-grammatical utterances. The KRR module supports the integration of information coming from different interlocutors and is capable of handling contradictory facts. The knowledge representation language is based on semantic networks. Question answering is based on deductive as well as inductive inference. A preliminary evaluation of the efficiency of the SLU/KRR system, for the purpose of knowledge acquisition, is presented.

1. Introduction

Due to the successive advances in robotics-related technologies, robots are closer to humans. Intelligent service robots capable of performing useful work in close interaction with humans are the next generation. However, in order to reach that phase, it is necessary to include in their design such basic capabilities as linguistic communication, reasoning, reactivity and learning.

“Integrated Intelligence” identifies an approach to building intelligent artificial agents in which the integration of all those aspects is considered [10]. This is the scope of CARL (*Communication, Action, Reasoning and Learning in Robotics*), a research project started in our institute in 1999, in the framework of which a robot prototype was developed, Carl [12, 11, 7].

The software architecture of Carl is based on a set of Linux processes. One of them handles general perception and action, including navigation. Other processes are dedicated to speech processing and touch screen interaction. An animated face displays appropriate emotions. High-level reasoning and natural language generation are mostly based on the Prolog engine. The central manager is an event-driven system.

The human-robot communication process is modeled as the exchange of messages, much like is done in multi-agent systems. The currently supported set of performatives or message types includes *tell*, *ask*, *ask-if*, and *achieve*.

The latest developments of Carl concern the SLU and KRR modules. Designed to enable a future enhancement of the dialogue capabilities, the new KRR module adopts a semantic network representation style and is capable of integrating information obtained from different interlocutors, even if they are contradictory. The robust SLU module uses a new hybrid deep/shallow approach. This paper focuses on the application of the developed SLU/KRR integration for knowledge acquisition.

Some other projects have attempted to use spoken natural language for human-robot interaction, including Godot [2] and Jijo-2 [5]. In comparison with these works, we explore the knowledge acquisition and management aspects in greater depth, namely by using inductive, deductive and analogical inference. Although there are various works on KR for robotics, few of them are focused on supporting the dialogue system. Skubic et al developed a framework to use spatial language in human-robot dialogues [13]. However, they are not focused on acquiring information from different interlocutors and do not consider the acquisition of non-spatial knowledge.

The paper is structured as follows. Section 2 describes the KRR module. Section 3 presents the robust SLU module. Section 4 presents some preliminary experimental results. Section 5 concludes the paper with reference to future work.

2. Knowledge Representation and Reasoning Module

The scenario in which the Carl robot must act involves dialogue with various interlocutors and exchange of information with them. The module should provide functionalities for knowledge acquisition and question answering. It's important to note that the agent should be able to start with no prior knowledge.

Since the robot can acquire data from different sources, the KRR module must handle contradictory pieces of information. Moreover, since spoken language input will be used, the recognition confidence should also be taken into account.

These problems have, in part, been addressed in other contexts. Benferhat et al presented strategies for conflict resolution developed to deal with exception handling and iterated belief revision, but which could also be applied in merging information from different sources [1]. Brazdil and Torgo have developed a method to construct an integrated knowledge base from several separated ones [3].

2.1. Knowledge Representation Language

Complex domains, like the one of a dialogue system in an intelligent mobile robot, require a general and flexible knowledge representation (KR) [9]. The definition of our KR language is based on typical definitions of semantic networks and on class and object diagrams of UML (Unified Modeling Language).

Semantic networks [14] address the main representation requirements to support a high level dialogue. It is very easy to represent the entities with them. Inference is simple, all one has to do is to follow the relations between entities. Semantic networks, not only provide ways for efficient computing, they also provide a very intelligible layout. Using semantic networks, the

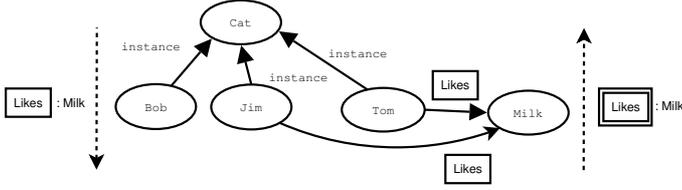


Figure 1: Analogy

rule of inheritance can be easily applied. In this rule, all properties of the supertype are copied to the subtypes, except if there is a redefinition of the property in the subtype.

This language assumes that objects are identified by system-generated identifiers. The predicates that provide information on specific objects include $instance(ObjID, Type)$, $name(ObjID, Name)$, $function(ObjID, Function)$ and $attribute(ObjID, Attribute, Value)$. Generalization is declared with $subtype/2$. Standard UML relations ($composition$, $association$) are also supported.

2.2. Basic Inference Mechanisms

There are three main types of inference: deduction, induction and analogy [6]. Take the logical entailment (1), in which P represents the premise, BK the background knowledge and C the consequence.

$$P \cup BK \models C. \quad (1)$$

Deduction is truth preserving. It derives consequence C from a given premise P and background knowledge BK . Here, this kind of inference is used when the type has an attribute with a default value and a question is made about the object's attribute.

If there is BK and C , induction can be used to hypothesize a premise P . In this work, inductive inference is used when the objects of a type have some attribute information that the type itself does not have.

Finally, analogy is a combination of deduction and induction. Take the knowledge given in Fig. 1, suppose someone asks "what does Bob like?". Since this information is given neither on the object *Bob* itself nor on the type *Cat*, the module has to: first use induction from the objects *Tom* and *Jim* to the type *Cat*; then use deduction from the type *Cat* to the object *Bob*.

2.3. The Reasoning System

The functionalities of the module are provided mainly by the following two procedures:

- $tell(Int, Fact, Conf)$ – the interlocutor *Int* tells *Fact* to the system; *Conf* is the interlocutor's confidence on *Fact*; in the integrated system, this is the Automatic Speech Recognizer (ASR) confidence;
- $ask(Fact, Answer, Conf)$ – the system is asked about *Fact*; the confidence in the *Answer*, *Conf*, is returned.

The procedure *tell* simply stores the information given by the interlocutors. Inference is used in *ask* in order to provide the most suitable answer.

Confidences are calculated as follows. Consider a property of an object or type for which a certain value is supported by N interlocutors and that a total of T interlocutors provided values

for this property. In this case, the confidence that the mentioned value is the correct one is given by:

$$Conf(N, T) = \frac{N}{T} \cdot \left(1 - \frac{1}{T+1}\right). \quad (2)$$

Note that the confidence of answers based on few statements is reduced.

If the question is about an attribute value in a type, Algorithm 1 is used to determine the value. The tree traversal step of the algorithm computes the frequency of occurrence of the possible values of the Attribute in the type and all its subtypes (and sub-subtypes, etc) and respective objects. If there is a supertype, the result of the tree traversal step is combined with a similar result inherited from the supertype.

Algorithm 1: GET_VALUE(Type, Attribute)

begin

$V_1, \dots, V_k \leftarrow$ possible values of Attribute

In a tree traversal, compute the frequencies of occurrence, N_1, \dots, N_k , of each possible value of Attribute in Type and all its subtypes (and sub-subtypes, etc) and respective objects

$T \leftarrow \sum_{i=1}^k N_i$ (total number of statements about values of Attribute)

for $i \leftarrow 1$ **to** k **do**

$C_i \leftarrow conf(N_i, T)$ (According to Equation 2)

if Type has no supertype **then**

$\text{return } ((V_1, C_1), \dots, (V_k, C_k))$

$ST \leftarrow$ supertype of Type

$((V_1, C'_1), \dots, (V_k, C'_k)) \leftarrow \text{GET_VALUE}(ST, \text{Attribute})$

for $i \leftarrow 1$ **to** k **do**

$C''_i \leftarrow (C_i + C'_i)/2$

return $((V_1, C''_1), \dots, (V_k, C''_k))$

end

In this algorithm, for simplicity, it is assumed that the input confidences are always 1.0. If that is not the case, each frequency of occurrence, N_i , should be replaced by the sum of the input confidences on value V_i .

If the question is about an attribute value in an object, confidences for all possible values are computed in the object (equation 2) and in the type (Algorithm 1). The value with the highest combined confidence is returned.

If the question is about a conjunction of facts, the global confidence is the multiplication of the confidences of the individual facts.

The implementation of this module was done in Prolog.

3. Robust Spoken Language Understanding

In order for a robot to acquire knowledge from human interlocutors through spoken language interaction, a robust spoken language understanding (SLU) capability is necessary.

SLU systems generally include a natural language understanding (NLU) module attached to an Automatic Speech Recognizer (ASR). The ASR decodes voice inputs to sequences of words. These sequences are then processed by the NLU module. NLU is usually performed in two stages: first a syntactic structure is built and then semantic information is extracted.

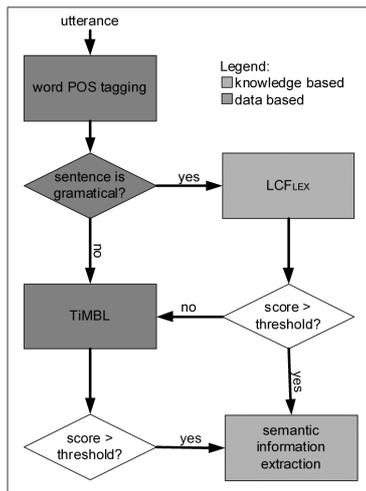


Figure 2: NLU system architecture

It is important to extract meaning from every ASR output instead of ignoring it, even if it is not grammatical. To be able to get the meaningful part of such utterances, the NLU module of the SLU system should be robust to several types of errors.

3.1. System Description

Carl uses LCFLEX [8], a flexible left-corner parser designed for efficient robust interpretation, based on the Lexical Functional Grammar (LFG) formalism for grammar definition. LCFLEX goal is to extract the most complete interpretation possible from the inputs. It supports word skipping, insertion of non-terminal grammar rules and selective flexible feature unification.

Although LCFLEX can interpret utterances not completely described by the system grammar, if the utterance has great deviation from it, results can miss the meaningful part of the utterance. For those utterances, a complementary memory based learning (MBL) approach is used.

MBL is based on the idea that intelligent behavior can be obtained by analogical reasoning. MBL algorithms take a set of examples as input, and produce a classifier capable of classifying new input patterns. If the system learns to detect important information from examples, it should be able to identify meaningful information in any utterance. The TiMBL [4] tool is used for this purpose.

The NLU system architecture is represented in Fig. 2. The priority of the system is to use LCFLEX for syntactic analysis. The parser can perform a deeper analysis than TiMBL, extracting more information from the utterances. If the parser fails or if the sentence is considered with great deviation to the grammar, TiMBL is used to extract meaningful information.

To build the syntactic structure of a sentence, the information used is the morfo-syntactic role of each word in the sentence instead of the word itself. So the lexical entries of the grammar used by LCFLEX are part-of-speech (POS) tags instead of words. The POS tags are assigned on-the-fly by two taggers.

The taggers use the Penn Treebank tagset that has 36 tags, so the lexicon of our grammar has just 36 entries which allows us to have a very small, easily maintained grammar.

If both taggers agree in more than 75% of the tags and the sentence is classified by TiMBL as grammatical, the sentence is

passed to LCFLEX for a deep syntactic analysis, otherwise it is passed to another instance of TiMBL for a shallow analysis.

If the sentence syntactic structure returned by LCFLEX includes more than 75% of the total words of the sentence, the analysis is passed to the semantics extraction module. Otherwise the system considers that the sentence has a great deviation from the grammar and the final analysis is shallow and made by the second instance of TiMBL. This instance of TiMBL can also ignore the sentence if it cannot get a valid analysis.

With this hybrid architecture, it was developed a system capable of performing deep analysis if the sentence is mostly acceptable by the grammar, but also capable of performing a shallow analysis if the sentence has severe errors. A more detailed explanation of the SLU system is given on [7].

3.2. Semantics

In order to integrate the SLU module with the new KRR module into the existing software architecture, it is crucial that the two modules use the same language. Therefore, a new semantics analysis module was developed, which receives as input the syntactic structure given by the ASR and parsing modules.

As an example, if an interlocutor tells Carl that “Professor James is in France”. The system extracts the following semantics: $[name(X, james), function(X, professor), association(be_in, X, Y), name(Y, france)]$.

It is important to note that the semantics analysis module is robust with respect to some misrecognition problems. For instance, consider the relation $subtype(SubType, Type)$. Suppose a phrase is misrecognized and instead of “a car is a machine” the ASR gets “the car is a machine”. This would generate a semantics $[instance(X, car), subtype(X, machine)]$. However, since the relation $subtype$ is only supposed to link two types, the misrecognized instance X is ignored and the semantics becomes: $subtype(car, machine)$.

4. Experimental Results

In Table 1, the results of an experiment performed with the SLU system are presented. One speaker read aloud 41 sentences, repeating 3 times each one, in a total of 123 sentences with 312 relations. The sentences were selected in order to generate a minimum of 20 samples of each semantic relation. The average number of words per sentence is 5.0 and the average number of semantic relations correctly representing the semantics of the sentences is 2.54. This experiment was carried out in the Carl robot in an environment with some background noise (a common research laboratory). As ASR system, Nuance 8.0 with trigrams language model was used.

Two performance measures are used. Precision is the number of correct relations extracted by the system divided by the total number of relations extracted ($= Right/(Right+Wrong)$). Recall is the number of correct relations extracted divided by the total number of correct relations that could have been extracted ($= Right/Target$). The system had a Recall of 84.3% and a Precision of 80.9%.

Looking carefully to results not presented, sometimes the final result does not contain enough information for the robot to take a decision, but at least it gives an idea about the sentence content allowing the robot to ask for future clarification.

Another experiment was carried out to test the performance of the integrated SLU/KRR system. First, an “Utopian” state of the knowledge base was created, using the same 41 sentences of the previous test. For every phrase, the right semantics was gen-

Table 1: *Semantic Analysis Results.*

Relation	Target	Right	Wrong	Precision	Recall
Association	87	74	19	79.6%	85.1%
Composition	21	7	0	100.0%	33.3%
Name	93	82	1	98.8%	88.2%
Instance	39	36	39	48.0%	92.3%
Subtype	21	21	1	95.5%	100.0%
Function	51	43	2	95.6%	84.3%
Total	312	263	62	80.9%	84.3%

erated manually, and added to the base with a call to the predicate *tell*, using 100.0% as a fictitious confidence of the ASR. Then, 41 questions based on the original phrases were submitted using the predicate *ask*. Of course the system could respond to all of them, but since the confidence of the answer depends on the number of statements that supports it, the average confidence is considerably lower than 100%.

To finish the second experiment, we have made the same 41 questions to the base that actually was produced after the first recognizer test. Table 2 shows the results of this comparison between reality and utopia. If we take the confidence result of reality (29.4%) and divide by the respective result of utopia (43.7%), we have a good measure of the efficiency of the system (67.2%). The Performance numbers indicate the percentage of questions that could be answered.

Despite the small test size, the results show that the system is suitable to support knowledge acquisition for an intelligent mobile robot.

Table 2: *Results of the KRR on the second experiment - utopia vs reality.*

Data	Questions	Answers	Performance	Confidence
Utopia	41	41	100.0%	43.7%
Reality	41	37	90.2%	29.4%
Efficiency			90.2%	67.2%

5. Conclusion and Future Work

In this paper, the recent evolution of the intelligent mobile robot Carl was presented. The paper focuses on the robust SLU module and on the new KRR module. The KR language is based on semantic networks and incorporates some notions from UML. The reasoning system combines deductive and inductive inference.

With respect to SLU, a new hybrid approach is used, which combines a robust parser with case-based shallow parsing (MBL). With this hybrid architecture we have developed an interface capable of performing deep analysis if the utterance is completely or almost completely accepted by the grammar, but also a shallow analysis if the utterance has severe errors.

Experiment results presented show that the system had a recall of 84.3% of the total relations spoken by the user, during a test in laboratory. Another experiment compared the performance of KRR module after the first test to an utopian state. This showed that the KRR module could answer to 90.2% of the questions, with an average confidence of 29.4%, which is 67.2% of the utopian confidence.

With respect to the KRR system, some important aspects

have not yet been addressed. One of them is induction on attributes with continuous values. Another problem is that inheritance does not take into account the confidence on the generalization links (subtype). A third problem is that when evaluating the values of an attribute it is assumed that only one is correct, which is not always the case.

6. References

- [1] Benferhat, Salem, et al.: Weakening conflicting information for iterated revision and knowledge integration. *Artif. Intell.* 153(1-2), (2004) 339–371
- [2] Bos, J., E. Klein, T. Oka Meaningful Conversation with a Mobile Robot Proceedings of the EACL (2003).
- [3] Brazdil, P. and L. Torgo: Knowledge Acquisition via Knowledge Integration. *Current Trends in AI*, B. Wielenga et al.(eds.), (1990) IOS Press, Amsterdam
- [4] Daelemans, W., J. Zavrel, A. van den Bosch, and K. van der Sloot.: TiMBL: Tilburg Memory-Based Learner Reference Guide, version 4.2. Technical report, P.O. Box 90153, NL-5000 LE, Tilburg, The Netherlands, June 2002
- [5] Matsui, T., et al. Integrated Natural Spoken Dialogue System of Jijo-2 Mobile Robot for Office Services Proceedings of AAAI (1999), 621-627.
- [6] Michalski, R. S.: Inferential Theory of Learning: Developing Foundations for Multistrategy Learning. *Machine Learning: A Multistrategy Approach*, 1994
- [7] Rodrigues, M., A. Teixeira and L. Seabra Lopes. An Hybrid Approach for Spoken Natural Language Understanding Applied to a Mobile Intelligent Robot. *Proc. Natural Language Understanding and Cognitive Science (NLUCS)*, (2004) 145-150
- [8] Rosé, C. P., and A. Lavie.: LCFlex: An Efficient Robust Left-Corner Parser. Technical report, University of Pittsburgh, 1998
- [9] Russell, Stuart and Peter Norvig *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1st edition (1995); 2nd edition (2003).
- [10] Seabra Lopes, L. and Connell, J.H. (eds.): *Semisentient Robots*: special issue of *IEEE Intelligent Systems*. vol. 16, n. 5. Computer Society (2001) 10-14.
- [11] Seabra Lopes, L., A. Teixeira, M. Rodrigues, D. Gomes, C. Teixeira, L. Ferreira, P. Soares, J. Girão & N. Sénica: Towards a Personal Robot with Language Interface. *Proc. Eurospeech'2003*, p. 2205-2208 (2003).
- [12] Seabra Lopes, L., A. Teixeira, M. Quinderé, A Knowledge Representation and Reasoning Module for a Dialogue System in a Mobile Robot. *Proc. Natural Language Understanding and Cognitive Science (NLUCS)*, (2005), to appear
- [13] Skubic, M., D. Perzanowski, et al.: Spatial Language for Human-Robot Dialogs. *IEEE Transactions on Systems, Man, and Cybernetics*, C 34(2): (2004) 154-167
- [14] Sowa, John F.: *Semantic Networks*. *MIT Encyclopedia of Cognitive Science*, www.jfsowa.com/pubs/semnet.htm