A Cognitive Architecture for Understanding and Producing Natural Language in Support of Robotic Creativity

Arianna Pipitone, Vincenzo Cannella, Roberto Pirrone and Antonio Chella

Abstract—A novel cognitive architecture is presented, which is aimed at understanding and producing natural language utterances intended as single dialogue moves between a humanoid robot and the user. The presented system is part of a wider robotic architecture already proposed by some of the authors, and is devoted to the implementation of all the robot actions related to listening and speaking. The dialogue is managed according to the well known Speech Act Theory. The architecture relies on the Construction Grammar Theory for providing the robot with its linguistic competency. A suitable domain ontology coupled with a linguistic source are used to code the agent’s symbolic knowledge and lexicon for both understanding and production tasks. The whole architecture has been implemented for the Italian, and is presented as the support for creative tasks such as open-ended dialogues with humans, entertaining, and tutoring.

I. INTRODUCTION

A robot engaged in creative tasks is required to adopt very complex behaviors to gain success. In general, creativity can be connected with intuition, emotions, reflection on one’s mental state, and understanding the others’ feelings. All these abilities can be regarded as aspects of meta-cognition. In this paper we focus on the support in creative dialogic tasks such as entertaining people in a conversation, recitation, and tutoring. All these tasks can be regarded as the ability of interacting with humans managing open-ended dialogues related to different topics. In this scenario, the robot has to plan the dialogue moves for achieving its persuasive intent; it has to be able to switch the focus of the conversation between different topics, and it must understand the human’s intentions. The conversation environment is perceived mainly through the auditory, and visual channels. Visual perception are related to the face and gestures that can convey information about the emotional state of the people interacting with the robot. The main information is perceived through the auditory channel where the text of the conversation is contained along with pragmatic information. The environment outlined above is uncertain as regards both the perceptions and the information devised by the robot because it has no access to internal state of the human. In this paper we propose a novel architecture to support creative dialogic tasks. The proposed system relies on a meta-cognitive robot architecture already presented by some of the authors. Such an architecture is aimed at supporting the robot in tasks that require planning in uncertain environments [1]. The system implements all the actions and the internal state of the robot that are related to the tasks of understanding a statement, producing the appropriate response, searching for new information sources, and switching the context. The implementation relies on the Construction Grammar Theory (CxG) for providing the robot with its linguistic competency [2]. A suitable domain ontology coupled with a linguistic source are used to code the agent’s symbolic knowledge and lexicon both for understanding and production tasks. In this way a single dialogue move is implemented. The definition of the inner state of the robot, and the meta-cognitive actions needed to manage an entire conversation have been implemented through a proper “dialogic reward” function, and by selecting the proper speech act [3] to rise the human’s interest. The whole system is being currently implemented in Italian on an Aldebaran Robotics NAO platform. The rest of the paper is arranged as follows. Section II reports some remarks on Computational Creativity and NLP, and the general meta-cognitive architecture supporting the presented one. Section III details the presented architecture, while in section IV the discussion is reported about the advantages of the presented system in supporting creativity in dialogic tasks.

II. THEORETICAL REMARKS

A. Computational Creativity and Natural Language Processing

Computational Creativity (CC) embraces Natural Language Processing (NLP) techniques in many creativity applications domain: in particular, Computational Linguistic (CL) was the first NLP discipline with the aim to support creativity evaluations. Personality Recognition from Text (PRT) is the automatic classification of authors personality from text; in literature strategies based on CL are by Argamon [4] that classified authors characters by linguistic features of her written text; on same wake, Oberland [5] added other characters and used a Naïve-Bayes algorithm applied on text of blog. Other approaches derive from the community of social network analysis that falls in CL too, such as[6] and [7] that used facebook posts and tweets respectively; they differ to the purely linguistics approaches only for the usage of other social parameters (as friends and follower numbers). Other attempts were made in the Semantic Web field for weaving creativity; in [8] an ontology for making creativity machine-understandable is developed: a core lexicon consisting of words that are associated with discussions about creativity is identified, and through a similarity measures they extract the ontology from the the corpus (that are composed by
scientific papers). To make the agent creative is a focus in the Linguistic Creativity, where the verbal expression is a main artificial creativity and two basilar tasks are related on what and how the agent says something. An interesting point is expressed in [9] where an algorithmic perspective is presented as the underlying paradigm for linguistic creativity.

B. The Reference Meta-Cognitive Architecture

The meta-cognitive architecture in support of the presented system relies on a unified management of uncertainty in Markov Decision Processes (MDPs). We called “believability” the unified mathematical form expressing uncertainty expressed as probability, possibility, and fuzzy logics, while the model has been named “uncertainty based MDP” (u-MDP). The structure of the agent is made by two u-MDP layers: the “cognitive u-MDP”, and the “meta-cognitive u-MDP” that is laid upon the previous one. Figure 1 shows the conceptual design of this architecture.

III. THE PROPOSED ARCHITECTURE

Figure 2 shows the proposed architecture; continuous arrows show sensory I/O, and functional connection between the different parts of the system. Dashed arrows show the retrieval from both internal and external information sources. Finally, thick arrows show the actual perception-action cycles. White rectangles are the implemented software components, which hav to be regarded as actuators connected to perception/action nodes in the two MDPs.

A. Dialogic Reward

MDP plans are computed using finite-horizon backward induction algorithm. The key idea is that planning a creative dialogue relies on a proper “dialogic reward” function $DR$ related to the human’s interest degree perceived by the robot. As a consequence, the dialogue will be aimed at keeping the human’s interest high using a proper choice of the speech act to be used at each dialogue move. To this aim, the robot will be able to change the dialogue context, to search for detailed information on user’s request, and the dialogue length will be limited to some tenths of moves to prevent a loss of interest in the spokesperson. Particularly, speech act choices will be coded as suitable believability values in the MDP model matrices. $DR$ will be increased by user’s statements showing interest otherwise it decays slowly at each dialogue move. Explicit requests can trigger information retrieval and a quick decreasing in $DR$. Finally, context switching is triggered when $DR$ falls below a minimum threshold.

B. Perception-Action Cycles for Dialogic Tasks

Three main dialogic tasks have been implemented: understanding/production, search for new information, and context switching. All the implemented tasks make use of two knowledge bases. The Domain KB contains the internal representation of the dialogue domain owned by the robot in form of OWL ontology. MultiWordnet [10] and the Italian Verbs Source (IVS) form the Linguistic KB. They are used to expand the robot’s lexicon. We are currently expanding the IVS to cover the other parts of speech, and it will become the only Italian dictionary of the system. In this phase MWn is used for retrieving parts of speech other than verbs. All of the tasks are activated by cognitive uMDP perception nodes connected to visual and auditory channels. Perception triggers the Meaning Activator (MA) that implements a
graph similarity search between the query-graph and the conceptual-graph. The strategy adopted for implementing cognitive understanding in the MA relies on applying the Graph Edit Distance (GED) method [11] between the query-graph Q and the conceptual-graph C, so that their GED is no larger than a distance threshold \( \tau \). In particular, the query-graph is the triple \( Q = (N_q, E_q, L_q) \) where the nodes set \( N_q \) contains the macro-syntactic roles of the NL query, parsed by the Freeling parser [12]. These nodes are sequentially connected reflecting their position in the query. The labels set \( L_q \) are labels nodes, and correspond to the tokens of the query provided by the parser. The conceptual-graph is the 4-tuple \( C = (N_c, E_c, L_c, \sigma) \) where the nodes set \( N_c = C_n \cup R_n \) is the union set of the set \( C_n \) containing the concepts in the KB, and the set \( R_n \) that contains relations. An edge in \( E_c \) connects only a concept-node to a relation-node if the concept is either the domain or the range for the relation itself. The edge is labeled with a progressive number for tracing the entities involved in the relation. \( \sigma \) is a label function \( \sigma : N_c \rightarrow L_c \) that associates to each node \( n_c \in N_c \) a list of strings \( l_c \in L_c \) that are obtained by querying the linguistic sources on-the-fly to expand the base lexicon formed by the labels in the KB. The distance measure involves both a similarity distance between the labels in the graphs and a structural measure in terms of the number of edit operations needed to make the graphs isomorphic. The understanding/production task takes place at the cognitive level of the architecture. The sub-graphs detected by the MA, and enriched by the related expanded lexicon are fed into the Speech Act Execution module (SAE). Such a component annotates sub-graphs with a proper string expressing form that corresponds to the planned speech act for the response. Some examples are “negative” or “interrogative negative”. Annotated sub-graphs are passed to the Answer Composer module (AC) that in turn produces lexical and grammatical descriptions of the domain terms according to the CxG theory. Such descriptions are the input for the Fluid Construction Grammar engine (FCG) [13], and are represented as “constructions”, namely form-meaning couples, and are made by two “poles”. Form pole contains the syntactic features of terms, while the semantic one contains meaning. Lexical constructions are related to a single word, while conjunctions of lexical constructions generate the grammatical ones. The FCG engine contains lexical constructions for the Italian adverbs and articles, that were manually created. For adverbs, the features embed some commonsense knowledge about them to enable context disambiguation. The lexical construction for the adverb “dove” stores some features like “luogo”, “posto”, “destra”, “sinistra” and so on. Lexical and grammatical constructions about the domain form the linguistic base of the system and are generated by querying the KB and the linguistic sources. When annotated KB sub-graphs are fed to the AC, it builds the correspondent meaning poles, and the related constructions fire; constructions are filtered according to the annotated form of the response. Several candidates remain after filtering. Next, the adverb tokens in the query are parsed by the FCG engine, and their corresponding lexical constructions fire. AC makes final disambiguation by choosing the sub-graph that has a link to the adverbial features stored in the construction, and the corresponding statement is produced. Searching for new information and context switching are both meta-cognitive actions. They are triggered by perception nodes in the meta-cognitive uMDP when the spokesperson says explicitly that either he is not satisfied by the robot answer or he is bored. The MA performs the base understanding sub-task in both cases. The meta-cognitive actuator in the first case is the Semantic Annotator (SA) presented by some of the authors in [14] which retrieves external contents in support of disambiguation when the user expresses perplexity about the actual meaning of the robot’s sentence. The latter actuator is the Context Switching module (CS) which instantiates a new OWL ontology as the domain KB.

IV. CONCLUSIONS

A novel architecture in support of dialogic creative tasks for humanoid robots has been presented where the robot has to keep the human’s interest high. The perception-action cycles in support of the tasks for understanding/production, searching for new information, and context switching have been implemented as a set of actuators in an MDP based meta-cognitive reference architecture where dialogue planning relies on maximizing a proper dialogic reward function. Speech act theory is used for building the form of each dialogue move, while Construction Grammars are used to assemble sentences. The whole system is being implemented in Italian on a NAO platform, and has proved extremely versatile in support of tutoring and open-ended dialogue tasks.

REFERENCES


