Knowledge-Driven Conversation for Social Robots: Exploring Crowdsourcing Mechanisms for Improving the System Capabilities

Lucrezia Grassi[a], Carmine T. Recchiuto[a] and Antonio Sgorbissa[a]

[a]*DIBRIS, Universitá di Genova, via all'Opera pia 13, 16145, Genova, Italy*

Abstract

Social robots and artificial agents should be able to interact with the user in the most natural way possible. This work describes the basic principles of a conversation system designed for social robots and artificial agents, which relies on knowledge encoded in the form of an Ontology. Given the knowledge-driven approach, the possibility of expanding the Ontology in run-time, during the verbal interaction with the users is of the utmost importance: this paper also deals with the implementation of a system for the run-time expansion of the knowledge base, thanks to a crowdsourcing approach.

Keywords

Social Robots, Autonomous Conversation, Ontology, NLP, Crowdsourcing

1. Introduction

Achieving a natural interaction between a human and a robot is a very complex task. A list of ten desired features that a conversational robot should have has been presented in [1]: among the most relevant aspects, the capability of *breaking the "simple commands only" barrier*, and having *multiple speech acts* should be pointed out.

The recent EU-Japan project CARESSES has dealt with some of these aspects [2]. In the context of the project, whose main focus was the development of *culturally-competent* robots, i.e., robots able to adapt verbal and non-verbal interaction to the user's cultural background, a framework for autonomous conversation has been developed [3], [4]. The framework was able to achieve mixed-initiative dialogues by exploiting the hierarchical structure of an ontology, thus implementing rich, knowledge-grounded conversations [5].

However, although significant progress has been achieved during the last years, assistive robots and chatbots still have many limitations. Some of the most common limitations are (i) failing to answer, (ii) not understanding the local language of the user, (iii) not giving the proper answer if there is some spelling mistake or some slang, (iv) having a limited knowledge, which can result in a repetitive conversation, and (v) not being coherent when answering to what the user says [6].

The work proposes an approach that mainly deals with (iv), relying on the framework for autonomous conversation developed within the CARESSES project but suggesting a *crowdsourcing approach* for adding knowledge as a consequence of the interaction with the users.

AIRO 2020

lucre.grassi@gmail.com (L. Grassi); carmine.recchiuto@dibris.unige.it (C.T. Recchiuto); antonio.sgorbissa@unige.it (A. Sgorbissa)

0000-0001-9550-3740 (C.T. Recchiuto); 0000-0001-7789-4311 (A. Sgorbissa)

2. Knowledge-Driven Conversation

Usually, knowledge-grounded conversation frameworks generate appropriate responses that reflect the acquired knowledge by relying on data-driven conversation models [7] or considering contextual information based on previous utterances [8]. In the proposed approach, the nucleus of the conversational framework is a Description Logic Ontology [9].

The Dialogue Tree is built starting from the Ontology structure, and the relation between topics is borrowed from the structure of the Ontology: specifically, Object Properties, Data Properties, and the hierarchical relationships among instances, are analyzed to define the branches of the Dialogue Tree.

Based on the Dialogue Tree, the policies for knowledge-driven conversation can be briefly summarized as follows [10]. Each time a user sentence is acquired:

1 A keyword-based Language Processing algorithm is applied to check if the sentence may trigger one of the topics in the tree;
2 If no topics are triggered, the conversation follows one of the branches of the tree, depending on the probabilities of each node (probabilities are encoded in the Ontology, and they depend on the user-specific preferences and on the user's cultural background);
3 Whatever node has been chosen, the system:
   (i) proposes some of the corresponding sentences (encoded in the Ontology as Class restrictions and Data Properties);
   (ii) acquires the user's feedback that can be used to update the Ontology and/or determine the next node to move to.

3. Crowdsourcing Mechanisms for the Run-time Expansion of the Knowledge Base

The possibility of acquiring knowledge in a systematic manner by relying on networked interactions with human users has been recently explored in different domains. For example, some museum collections have used a "social tagging" approach to enhance curatorial documentation [11], allowing users to assign tags to museum objects displayed on a website, while the New York Public Library has launched a web-based crowdsourcing project, asking people to transcribe menus from its historical menu collection [12]. Crowdsourcing mechanisms have also been investigated for collaborative Ontology construction projects: [13] proposed a two-phase methodology for allowing non-experts users to concurrently build an Ontology about dietary approaches, while a method to verify if automated techniques for building biomedical Ontology hierarchies are reliable, based on a Bayesian inference model, has been developed in [14].

This work proposes the usage of crowdsourcing mechanisms for a run-time expansion of the knowledge base used for building the Dialogue Tree. A three-step approach has been implemented:

1. Recognition of relevant concepts in the user's sentence;
2. Insertion of the concepts in the Ontology;
3. Validation of the concepts.

Concerning 1, a Dialogflow Agent [15], i.e., a web service to manage functionalities for autonomous conversation, has been developed. Dialogflow Agents are characterized by *Intents*, which categorize end-user's intention for one conversation turn, and *Parameters*, which are values extracted from the sentence depending on the training of the Agent. In the proposed approach, four Intents have been defined: *memories-past*, *preferences*, *norms*, and *beliefs*, corresponding to conversation topics already present in the starting Ontology. All intents have been trained with example phrases for what end-users might say, which include some manually tagged parameters. As an example, the *preferences* Intent contains some training phrases of type "*I love <tagged parameter>*": whenever the user says "*I love <concept>*", this sentence is matched to the *preferences* Intent and whatever the value of *concept* is, it is returned as a response.

To collect appropriate training phrases for each Intent in such a way that Dialogflow can correctly perform the Intent matching, a vocal questionnaire has been created, where participants are required to answer, using their microphone, to 20 questions (5 for each Intent).

Regarding 2, different techniques for the insertion of new concepts (and related sentences) in the right place of the Ontology have been implemented. Such techniques involve again the usage of NLP tools for detecting the category of the user's sentence and the type of the recognized entity, and they are currently under evaluation. Finally, a procedure for validating the concepts inserted in run-time has also been developed (3). It is based on a revision process that indirectly asks users to independently revise others' information, to reach a consensual version. This *peer-review* approach, complemented by an external moderation, is also being evaluated.

4. Recognition of Relevant Concepts

As mentioned before, at the actual state, experimental tests with users have been focused on step 1. of the process, i.e., the recognition of relevant concepts in the user's sentence. A total of 30 participants have been recruited for the experiments: 20 questionnaires have been used to train the agent, while the answers of the remaining 10 questionnaires have been set aside to use them to test the Agent. To validate the approach, two independent *taggers* have tagged the 20 questionnaires of the training set, while a different person has tagged the evaluation set. It is worth saying that the Agent has been trained with pieces of the sentences (split by "and"). For this reason, to assess the performances of the Agent, the test answers have been split according to the same rules used for the training answers. The splitting also increases the probability of extracting at least one meaningful concept in the whole answer, which is what matters to expand the knowledge base.

Each piece of the test answers has been classified as:

- True Positive (TP): if Dialogflow correctly recognized the concept that was manually tagged;
- False Positive (FP): if Dialogflow recognized something that was not tagged;
- True Negative (TN): if Dialogflow did not recognize anything and nothing was tagged;
- False Negative (FN): if Dialogflow did not recognize anything but something was tagged.

The results of the classification of the pieces of each answer are summarized in the Confusion Matrix shown in Figure 1 (left). By looking at the tables, it is immediately clear that almost every time Dialogflow recognized something it was a tagged concept. However, in many cases, it did not recognize what had been tagged: as it will be clear in the following, this is not an issue if we reason in terms of answers and not in terms of pieces of answers. Starting from the Confusion Matrix, the most common parameters for a binary classifier have been computed and reported in Figure 1 (right). The same analysis has been carried out (Figure 2) considering the sentence as a whole, and by using this rationale, which allows to analyze if at least one concept has been correctly extracted from the whole sentence:

- If the answer contains at least one sentence classified as TP, the answer is considered as TP;

	PREDICTED YES	PREDICTED NO	Total
ACTUAL YES	**203**	119	322
ACTUAL NO	45	**91**	136
Total	248	210	**458**

(a) Confusion Matrix

Parameter	Formula	Value
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	0.64
Sensitivity (Recall)	$\frac{TP}{TP+FN}$	0.63
Specificity (SP)	$\frac{TN}{TN+FP}$	0.67
Precision	$\frac{TP}{TP+FP}$	0.82

(b) Parameters table

Figure 1: Confusion Matrix containing the classifications of the answers' pieces (a) and related parameters of the classifier (b).

	PREDICTED YES	PREDICTED NO	Total
ACTUAL YES	**151**	31	182
ACTUAL NO	17	**1**	18
Total	168	32	**200**

(a) Confusion Matrix

Parameter	Formula	Value
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	0.82
Sensitivity (Recall)	$\frac{TP}{TP+FN}$	0.83
Specificity (SP)	$\frac{TN}{TN+FP}$	0.06
Precision	$\frac{TP}{TP+FP}$	0.90

(b) Parameters table

Figure 2: Confusion Matrix containing the classifications of the answers (a) and related parameters of the classifier (b).

- If the answer does not contain any sentence classified as TP but it contains at least a sentence classified as FP, the whole answer is labelled as FP;
- If the answer does not contain any sentence classified as TP or FP, but it contains at least a sentence classified as FN, the whole answer is labelled as FN;
- If the answer does not contain any sentence classified as TP, FP or FN, but it contains at least a sentence classified as TN, the whole answer is labelled as TN.

5. Conclusions

From the analysis of the collected data, it may be observed that, working on the pieces of each answer, the system achieves a high Precision, which means that almost every recognized concept would have also been tagged manually. On the other side, the Accuracy (how often is the classifier correct?), the Sensitivity (when something is tagged, how often does it recognizes it?), and the Specificity (when something is not tagged, how often does it not recognize anything?) may be improved. However, when considering the whole answers to assess the capability of the system to extract at least one meaningful concept in each of them, all performance indicators are greater than 0.8, except the Specificity, due to a high number of FP. These results give positive insights of the reliability of the proposed approach aimed at recognizing relevant concepts in the user's sentence for a run-time expansion of the knowledge-grounded conversation system.

References

[1] N. Mavridis, A review of verbal and non-verbal human–robot interactive communication, Robotics and Autonomous Systems 63 (2015) 22–35.

[2] A. A. Khaliq, U. Köckemann, F. Pecora, A. Saffiotti, B. Bruno, C. T. Recchiuto, A. Sgorbissa, H.-D. Bui, N. Y. Chong, Culturally aware planning and execution of robot actions, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 326–332.

[3] A. Sgorbissa, I. Papadopoulos, B. Bruno, C. Koulouglioti, C. Recchiuto, Encoding guidelines for a culturally competent robot for elderly care, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018, pp. 1988–1995.

[4] B. Bruno, C. T. Recchiuto, I. Papadopoulos, A. Saffiotti, C. Koulouglioti, R. Menicatti, F. Mastrogiovanni, R. Zaccaria, A. Sgorbissa, Knowledge representation for culturally competent personal robots: requirements, design principles, implementation, and assessment, International Journal of Social Robotics 11 (2019) 515–538.

[5] C. Recchiuto, L. Gava, L. Grassi, A. Grillo, M. Lagomarsino, D. Lanza, Z. Liu, C. Papadopoulos, I. Papadopoulos, A. Scalmato, et al., Cloud services for culture aware conversation: Socially assistive robots and virtual assistants, in: 2020 17th International Conference on Ubiquitous Robots (UR), IEEE, 2020, pp. 270–277.

[6] H. Raval, Limitations of existing chatbot with analytical survey to enhance the functionality using emerging technology, International Journal of Research and Analytical Reviews (IJRAR) 7 (2020).

[7] M. Ghazvininejad, C. Brockett, M.-W. Chang, B. Dolan, J. Gao, W.-t. Yih, M. Galley, A knowledge-grounded neural conversation model, arXiv preprint arXiv:1702.01932 (2017).

[8] S. Kim, O.-W. Kwon, H. Kim, Knowledge-grounded chatbot based on dual wasserstein generative adversarial networks with effective attention mechanisms, Applied Sciences 10 (2020) 3335.

[9] N. Guarino, D. Oberle, S. Staab, What is an ontology?, in: Handbook on ontologies, Springer, 2009, pp. 1–17.

[10] C. T. Recchiuto, A. Sgorbissa, A feasibility study of culture-aware cloud services for conversational robots, IEEE Robotics and Automation Letters 5 (2020) 6559–6566.

[11] J. Trant, Tagging, folksonomy and art museums: Early experiments and ongoing research (2009).

[12] M. Lascarides, B. Vershbow, What's on the menu? crowdsourcing at the new york public library, Crowdsourcing Our Cultural Heritage (2014) 113–1137.

[13] M. Zhitomirsky-Geffet, E. S. Erez, B.-I. Judit, Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health, Journal of the Association for Information Science and Technology 68 (2017) 681–694.

[14] J. M. Mortensen, M. A. Musen, N. F. Noy, Crowdsourcing the verification of relationships in biomedical ontologies, in: AMIA Annual symposium proceedings, volume 2013, American Medical Informatics Association, 2013, p. 1020.

[15] DialogFlow Documentation, accessed October 7 2020S. URL: https://cloud.google.com/dialogflow/docs.