Intention recognition for robotic applications

Michele Persiani
Umea University
Umeå, Sweden 901-87
michelep@cs.umu.se

Thomas Hellström
Umea University
Umeå, Sweden 901-87
thomash@cs.umu.se

Suna Bensch
Umea University
Umeå, Sweden 901-87
suna@cs.umu.se

ABSTRACT

Intention recognition is a pillar of human-human communication and has also been extensively studied for human-robot interaction. In this position paper we propose an approach to merge NLP intention recognition with task planning. Our goal is to obtain a robotic system that supports older adults in daily tasks by predictively inferring intentions from spoken utterances.

ACM Reference format:
Michele Persiani, Thomas Hellström, and Suna Bensch. 2017. Intention recognition for robotic applications. In *Proceedings of ECCE 2017 Workshop on Robots in Contexts, Umeå, Sweden, September 2017,* 2 pages.
DOI: 10.475/123_4

1 INTRODUCTION

One of the pillars of human-human communication is intention recognition. The ability to predict and understand what others will do gives us clear advantages compared to other species [12]. The same holds for a robot involved in human-robot-interaction - "reading the mind" of the human makes it possible for the robot to be pro-active and act on incomplete or incorrect information. It also makes it possible for a social robot to assist a human without explicit commands. Robots capable of this will give rise to "a fundamentally new kind of collaboration between humans and robots" [3]. Intention recognition for robots in general has been thoroughly addressed in our earlier research, both at the sensory motor level [1], and at higher levels [4] with several interaction modalities. Speech is often a preferred mode of interaction for applications with robots in health and eldercare [11], yet it often requires integration with other sensorial modalities in order to ground information to physical reality.

ECCE 2017 Workshop on Robots in Contexts, Umeå, Sweden

DOI: 10.475/123_4

Numerous methods to infer human intention from sentences of varying linguistics types have been suggested, and deep learning methods using word embeddings and Recurrent Neural Networks represent state of the art for this task.

Word embeddings techniques such as GloVe [9] and Word2Vec are recent approaches to create word representations as dense vectors that capture both semantic and syntactic information. Word2Vec uses local context windows methods to learn the representations, while GloVe combines such methods with global matrix factorizations, such as the words' co-occurrence matrix. Both algorithms are data-driven and rely on statistical information of words in the training corpus.

Supervised learning methods for intention recognition can be implemented on top of such word embeddings. Given an utterance, intention can be extracted as a desired user action and associated parameters or slots. Yoon Kim [7] shows how convolutional neural networks trained using pre-trained word vectors can be utilized for sentence classification, achieving results comparable to other state of the art methods. Joo-Kyung Kim at al[6] enrich word embeddings with information taken from different semantic lexicons, such as WordNet, PPDB and the Macmillan Dictionary. They modify initial off-the-shelf word embeddings by taking into account synonyms, antonyms and context related words. They then use a bidirectional Long Short Term Memory (LSTM) model for intention, slots and domain classification, leading to increased classification performance compared to methods based non-enriched word embeddings.

Intention recognition can be combined with task planning in the context of human-robot collaboration. In this case intention recognition is carried out in the task's execution context, and the robot has to keep an updated belief on how the collaboration with the user will proceed in order to decide how to act in the most useful way. The collaboration can then be modeled in a shared planning space in which both the user and robot plans are merged. Also for this task it is necessary to estimate the user's intent, represented by the user's goal and current plan. Work in this direction has been carried on by Saffar et al. [10]. They generate an Activation Spreading Network (ASN) starting from an Hierarchical Task Network (HTN). The activation that flows through the ASN represents the recognized user intention, and the path with the highest activation will be considered as the most probable user plan.

Planners have been previously used in robotics together with imperative sentences that define new planning goals[5, 8]. Imperative sentences can be used to direct human collaboration with a robot, yet a true adaptive and intent-driven collaboration still has to be achieved for complex and shared tasks.

2 OBJECTIVES

Our overall goal is to design an intention recognizing system that combines speech understanding with task planning.

In an assumed scenario, an older adult wants to eat, gets up from the chair and moves towards the kitchen. The adult may use an explicit imperative command such as "Make me a sandwich", or an implicit declarative sentence such as "I am hungry". Both sentences could represent the same underlying intention of asking for help with a sandwich, but require different inference mechanisms.

Further investigation will be carried on in the context of dialog management. The user can interact with the system through dialogs in order to clarify its intent or to change the current state of the system, for instance by changing parameters associated to the current behavior or by issuing new commands.

The system's goal is to keep the user satisfied. An initial set of user needs will be hardcoded, and the user will also be able to add new needs. In order to have successful dialogs with the user, the user intent will be inferred and a belief for the user's mental state will be continuously updated. The robot will pro-actively carry out dialogs to check if it should execute a certain task, when uncertainty is present. This could also be carried out in order to learn over time the desired behavior associated with the user. Grounding to physical objects will be done by extending earlier work[2] in which priming in semantic networks was shown to reduce perceptual ambiguity in intention recognition. Visual-sensor based intention recognition will later be integrated.

3 RESEARCH TASKS

We will develop algorithms to infer human intention from sentences of varying linguistics types, and deep learning methods for NLP focused on word embeddings and neural networks will be employed for this task. Further investigation will be carried out in the context of dialog management, having the goal of adjusting the system's behavior from dialog with the user.

Intention recognition and dialog management will be incorporated in the robot's task planner. In a possible approach, the planner will be equipped with static goals (such as keeping the user satiated), and uttered declarative sentences will add facts to the current state. The planner will schedule the robot's actions depending on context (e.g. hour of day) and current value of user's estimated state. New goals and recurrent goals and procedures are added by the means of dialogs with the user. The planner will then generate a plan that takes into consideration the user's intention. In this way, by taking into account the dialog management, the planner is used as a reasoning tool for implicit intention recognition, rather than a problem solving task planner.

REFERENCES

[1] Benjamin Fonooni and Thomas Hellström. 2015. Applying a priming mechanism for intention recognition in shared control. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2015 IEEE International Inter-Disciplinary Conference on*. IEEE, 35–41.

[2] Benjamin Fonooni, Thomas Hellström, and Lars-Erik Janlert. 2016. Priming as a Means to Reduce Ambiguity in Learning from Demonstration. *International Journal of Social Robotics* 8, 1 (01 Jan 2016), 5–19. DOI: http://dx.doi.org/10.1007/s12369-015-0292-0

[3] Andreas G Hofmann and Brian Charles Williams. 2007. Intent Recognition for Human-Robot Interaction.. In *Interaction Challenges for Intelligent Assistants*. 60–61.

[4] Aleksandar Jevtić, Guillaume Doisy, Yisrael Parmet, and Yael Edan. 2015. Comparison of interaction modalities for mobile indoor robot guidance: direct physical interaction, person following, and pointing control. *IEEE Transactions on Human-Machine Systems* 45, 6 (2015), 653–663.

[5] Thomas Keller, Patrick Eyerich, and Bernhard Nebel. 2010. *Task Planning for an Autonomous Service Robot.* Springer Berlin Heidelberg, Berlin, Heidelberg, 358–365. DOI: http://dx.doi.org/10.1007/978-3-642-16111-7_41

[6] Joo-Kyung Kim, Gokhan Tur, Asli fjelikyilmaz, Bin Cao, and Ye-Yi Wang. 2016. Intent detection using semantically enriched word embeddings. In *SLT*.

[7] Y. Kim. 2014. Convolutional Neural Networks for Sentence Classification. *ArXiv e-prints* (Aug. 2014).

[8] Thomas M. Howard, Stefanie Tellex, and Nicholas Roy. 2014. A natural language planner interface for mobile manipulators. (05 2014), 6652–6659.

[9] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *In EMNLP*.

[10] Mohammad Taghi Saffar, Mircea Nicolescu, Monica N. Nicolescu, and Banafsheh Rekabdar. 2015. Intent Understanding Using an Activation Spreading Architecture. *Robotics* 4 (2015), 284–315.

[11] António Teixeira. 2014. A critical analysis of speech-based interaction in healthcare robots: Making a case for the increased use of speech in medical and assistive robots. *Speech and automata in health care* (2014), 1–29.

[12] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28, 5 (2005), 675fi?!691. DOI: http://dx.doi.org/10.1017/S0140525X05000129