Yanchao Yu

Heriot-Watt University
School of Mathematics and Computer Science
Edinburgh Campus
Edinburgh EH14 4AS
United Kingdom

y.yu@hw.ac.uk
sites.google.com/site/yanchaoyu147/

1 Research Interests

My primary area of research interest lies in **Perceptual Semantic Grounding** using **Natural Language Interaction**. More specifically, I am working on building a multimodal teachable system that can interactively learn to ground, identify, and talk about objects and low-level features (visual attributes) in the real environment through dialogue from the human tutors. The system allows learning novel knowledge abut objects from scratch based on semantic and perceptual processing in the dialogue.

I am interested in exploring what linguistic approaches and dialogue strategies may help to improve the system/robot's learning performance, and also encourage humans to engage in a life-long learning game/task. It may ranges over different fields of **Incremental Dialogue Management**, **Semantic Analysis of Perceptual Scenes** as well as **Continuous Perception Learning**. On the current research stage, I am working on a fully adaptive learner dialogue strategy using Reinforcement Learning (RL) to achieve a best trade-off between learning performance and the effort by human tutors.

1.1 Semantic Language Grounding

There are two levels of language grounding that a multimodal system should consider when learning from human tutors through dialogue:

1) *Dialogue Level*: language grounding at the dialogue level is related to dialogue context from all participants and corresponding contextual analysis. It reflects whether the previous context has been agreed by participants or not and who has agreed with it.

2) *Perceptual Level*: language grounding at the perceptual level takes information from different modalities into account. It focuses on aligning words/phrases/sentences in NL to a perceptual scene (e.g. images and events) in the physical world.

In terms of Human-Robot Interaction and collaboration, an intelligent system/robot needs to consider both levels as it expects to interact with humans in a more natural human-like way within a real environment.

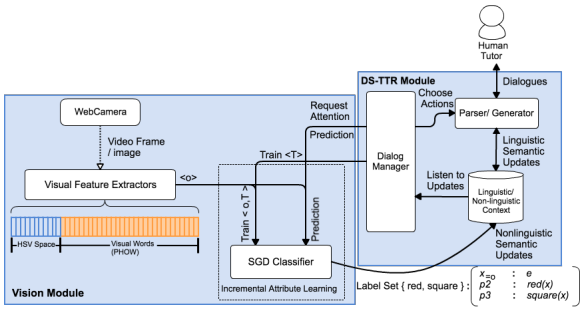

Figure 1: Architecture of the Teachable system

1.2 Previous Work

My current research mainly focuses on language grounding at the perceptual level. I will briefly describe my work in this section.

1.2.1 Integrating Semantic Processing with Vision

Yu et al. (2015) developed a multimodal teachable system for interactively learning to identify novel visual attributes through dialogue from human tutors. This system consists of two key components – a *vision system* and the *DS-TTR parser/generator* (see Fig. 1). The vision module classifies a (visual) scene, i.e. deems it to be of a particular type, expressed as a TTR Record Type (RT). This is done by deploying a set of binary and incremental attribute classifiers (Logistic Regression SVMs with Stochastic Gradient Descent (see Yu et al. (2016a)), which ground the simple types (atoms) in the system (e.g. 'red', 'square'), and composing their output to construct the total type of the visual scene. This representation then acts as a) the non-linguistic context of the dialogue for DS-TTR and b) the logical database from which answers to questions about object attributes are generated. The system can also generate questions to the tutor (Yu et al., 2016b) about the attributes of objects based on the entropy of the classifiers that ground the semantic concepts, e.g. those for colour and shape. The tutor's answer then acts as a training instance for the classifiers (basic, atomic types) involved.

1.2.2 Investigating Dialogue Strategies for Interactive Learning

Yu et al. (2016a, 2016b) have investigated the effectiveness of different tutor/learner-driven dialogue policies and capabilities (e.g. initiative, uncertainty, context-dependency as well as knowledge-demanding) in an interactive learning process with humans. We intend to explore an appropriate policy with best trade-offs between learning performance (i.e. recognition accuracy) and effort required by the tutor. Through comparisons, a dialogue strategy that the system takes initiative and consider uncertainty from the visual classifiers has demonstrated a better performance than the others.

1.3 Current & Future Works

Currently, we are focusing on optimising the Learner dialogue strategy with an adaptive uncertainty threshold using RL (Yu et al., 2016c). The partially adaptive dialogue strategy may determine when and how to express novel visual objects or ask for help from human tutors based on its previous knowledge and experience by adapting the uncertainty threshold.

In the future work, we intend to 1) work on a data-driven, incremental dialogue management at the lexical level; 2) use the similar setup to collect multimodal data (language and vision examples) from a set of human-human game interactions. The goal of these works is to explore and learn a human-like dialogue policy that may handle the interactive learning problem with humans in a more natural way.

2 Future of Spoken Dialog Research

In my opinion, there are many possibilities for spoken dialogue systems to keep growing in the near future, e.g. open-domain, data-driven dialogue management and incremental processing in dialogue system. I expect a continued growth in this field related to grounding problems at both dialogue and perceptual levels. An effective spoken dialog system need to be able to not only understand what users are talking about, but also ground situated context into other modalities from the real environment. To achieve this goal requires more close collaborations between researchers from different fields, such as physics, computer vision, multimodal interaction, and robotics.

In the more distant future, dialogue systems will be deeply involved in Human-Robot Collaboration, instead of the standard interaction. Apart from grounding, the important aspects of these collaborations might also include identifying and distinguishing user preferences and also detecting the final purpose through dialogue. A unified dialogue framework needs to be proposed to be deployed for different collaboration tasks and then expanding and sharing experience from one task to another.

3 Suggestions for Discussion

These are the four topics I propose for discussion:

- The new position of spoken dialogue systems in the future technology with new mediums (e.g. the 3D virtual reality);
- The importance of learning perceptual information online through Dialogue with humans.
- The possibility of training a dialogue manager with data from multiple modalities, rather than only speech examples.

References

Yanchao Yu, Arash Eshghi and Oliver Lemon 2016. *Comparing attribute classifiers for interactive language grounding*. The workshop on vision and language (VL15) associated with EMNLP. Lisbon, Portugal.

Yanchao Yu, Arash Eshghi and Oliver Lemon 2016. *Training an adaptive dialogue policy for interactive learning of visually grounded word meanings*. The Seventeenth Annual SIGDial Meeting on Discourse and Dialogue (SIGDial). Los Angeles, USA.

Yanchao Yu, Arash Eshghi and Oliver Lemon 2016. *Interactive Learning of Visually Grounded Word Meanings from a Human Tutor*. Fifth ACL workshop on Vision and Language (VL'16). Berlin, Germany.

Yanchao Yu, Arash Eshghi and Oliver Lemon 2016. *Comparing dialogue strategies for learning grounded language from human tutors*. Twentieth Workshop on Semantics and Pragmatics of Dialogue (SEMDIAL). New Jersey, USA.

Biographical Sketch

Yanchao Yu is a Research Assistant and part-time PhD student under the supervision of Professor Oliver Lemon in the Interaction Lab at Heriot-Watt University in Edinburgh, Scotland. He received his Master's degree in Software Engineering from Heriot-Watt University. During his Masters studies, his first encounter with spoken Dialog system was his Master project that built a mobile intelligent assistant for emergency aid. After that, he worked on several mobile spoken dialogue systems (e.g. PARLANCE and SpeechCity) in the Interaction Lab and started his part-time PhD in the lab in October 2014. Other than research, he is also interested in entrepreneurship as a career plan in the future.