Learning to Interact with Humans by Lifelong Interaction with Humans

LIHLITH introduces a new **lifelong learning** framework for the interaction of humans and machines on specific domains with the aim of improving the quality of existing **dialogue systems** and lowering the cost of

deployment in new domains

State of the art

- > Producing training data for each dialogue domain is very time consuming
- > Industrial dialogue systems are based on rules

Key ideas

- > LIHLITH will focus on goal-driven question-answering dialogue
- > LIHLITH will be able to learn continuously from the dialogues
- > Dialogues will be designed to get feedback from the user: "OK, thanks!" "I meant X"
- "No, that's not right"
- > All modules will learn from the feedback down in the pipeline

Technologies

- > Lifelong learning, deep learning
- > Natural language processing, knowledge induction
- > QA system, semantic textual similarity

Lifelong Machine Learning. Z. Chen and B. Liu. Morgan Clayton. 2016



Research areas **RA1: Lifelong learning for dialogue**

- » User feedback
- > **Proactive capabilities** to the system to ask for » New knowledge
- » Performance feedback to the user when
 - its past reactions have been rejected, what the user has said is too ambiguous, the possible answers are too numerous or have too
 - similar confidence scores

RA2: LLL for knowledge induction and question answering (QA)

- » Based on word embeddings (learned from domain corpora) and concept and relation embeddings (learned from existing ontologies previously adapted to the domain)
- » Feedback will be used to provide supervised signal in these learning systems (i.e. tune the parameters of the deep learning systems)
- » Powered by a semantic textual similarity system (STS)
- generate low-dimensional representations of terms, concepts and relations
- » Capturing propositional knowledge from background texts and ontologies to » Focused on community question answering sites (Yahoo Answers,
 - StackExchange, Answers.com...)

> Release of all **experimental conditions to the public**

- > Use **crowdsourcing platforms** in order to test the systems
- » Crowdsourcing templates will include online access to the dialogue systems, allowing to measure the performance and the subjective opinion of live users
- > Use semi-automatic techniques to produce **artificial large amounts of training** data in order to train deep learning systems
- » Manually converting QA datasets in interactive question-answering dialogues based on crowdsourcing
- » Using a novel technique where semi-automatic expansion is used to produce variations in the dialogues (e.g. hiding parts of the dialogues or using lexical variability)

> Autonomous reconfiguration of dialogue strategies based on

» Online lifelong learning algorithms

> Use user feedback to improve knowledge induction and QA

> Knowledge induction

> Question answering

RA3: Evaluation of dialogue

LIHLITH is a CHIST-ERA project selected for the call 2016 "Lifelong Learning for Intelligent Systems (LLIS)"



