

Concept Learning from Natural Language Interactions

Abstract:

Humans can efficiently learn about new concepts and phenomena using natural language communications. For example, a human can learn the concept of a phishing email from natural language explanations such as 'phishing emails often request your bank account number'. On the other hand, purely inductive learning systems typically requires a large collection of labeled data for learning such a concept. If we wish to make computer learning as efficient as human learning, we need to develop methods that can learn from natural language interactions.

Learning from language presents two key challenges. The first is that of learning from interpretations, which refers to the mechanisms through which interpretations of language statements can be used to solve learning tasks in the environment. The second is the more basic problem of learning to interpret language, which refers to an agent's ability to map natural language explanations in pedagogical contexts to formal semantic representations that computers can process and reason over. We address aspects of both these problems, and provide an interface for guiding concept learning methods using language.

For learning from interpretation, we focus on concept learning (binary classification) tasks. We demonstrate that language can formulate learning tasks by defining rich and expressive features (e.g., 'Does the email ask me to click a hyper-link?'), and show that methods for concept learning can benefit substantially from such explanations. We propose to address assimilation of declarative knowledge expressed in language explanations that implicitly specifies model constraints for learning tasks (e.g., 'Most emails are not phishing emails'). In particular, we focus on quantifier expressions (such as usually, never, etc.), which denote generality of specific observations, and can be incorporated into training of classification models. We also propose to explore the use of language for mixed-initiative interactions with a teacher to reduce the sample complexity of learning.

Apart from developing computational machinery that uses language explanations to guide machine learning methods, we also develop complementary algorithms for learning to interpret language by incorporating different types of environmental context, including conversational history and sensory observations. We show that such context can enrich models of semantic interpretation by not only providing discriminative features, but also reducing the need for expensive labeled data used for training them.



Speaker:

Shashank Srivastava

Thesis Committee:

Tom Mitchell (Chair)
Taylor Berg-Kirkpatrick
William Cohen
Dan Roth (University of Pennsylvania)

Mon., Oct. 16, 2017

9:00am

3305 Newell Simon Hall

Link to draft document:

<http://www.cs.cmu.edu/~shashans/thesis-proposal.pdf>

