Mach Learn
DOI 10.1007/s10994-013-5383-2

Learning perceptually grounded word meanings from unaligned parallel data

Stefanie Tellex · Pratiksha Thaker · Joshua Joseph · Nicholas Roy

Received: 16 May 2012 / Accepted: 4 May 2013

Abstract In order for robots to effectively understand natural language commands, they must be able to acquire meaning representations that can be mapped to perceptual features in the external world. Previous approaches to learning these *grounded* meaning representations require detailed annotations at training time. In this paper, we present an approach to grounded language acquisition which is capable of jointly learning a policy for following natural language commands such as "Pick up the tire pallet," as well as a mapping between specific phrases in the language and aspects of the external world; for example the mapping between the words "the tire pallet" and a specific object in the environment. Our approach assumes a parametric form for the policy that the robot uses to choose actions in response to a natural language command that factors based on the structure of the language. We use a gradient method to optimize model parameters. Our evaluation demonstrates the effectiveness of the model on a corpus of commands given to a robotic forklift by untrained users.

Keywords Robotics · Language · Machine learning · Probabilistic graphical models

1 Introduction

In order for robots to robustly understand human language, they must have access to representations capable of mapping between symbols in the language and aspects of the external world which are accessible via the robot's model of its environment. Previous symbolic approaches have represented word meanings as symbols in some specific symbolic language, either programmed by hand (Winograd 1971; MacMahon et al. 2006) or learned (Matuszek et al. 2010; Chen and Mooney 2011; Liang et al. 2011). Because word meanings are represented as symbols, rather than perceptually grounded features, the mapping between these symbols and the external world must still be defined *a priori*.

Editors: Antoine Bordes, Léon Bottou, Ronan Collobert, Dan Roth, Jason Weston, and Luke Zettlemoyer.

S. Tellex (✉) · P. Thaker · J. Joseph · N. Roy
32 Vassar St, Cambridge, MA 02139, USA
e-mail: stefie10@csail.mit.edu

Published online: 18 May 2013