Towards Developing Probabilistic Generative Models for Reasoning with Natural Language Representations

Daniel Marcu[1] and Ana-Maria Popescu[2]

[1] Information Sciences Institute and Department of Computer Science,
4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292
marcu@isi.edu, http://www.isi.edu/~marcu

[2] Department of Computer Science, University of Washington,
Seattle, Washington 98105-2350, Box 352350
amp@cs.washington.edu, http://www.cs.washington.edu/homes/amp

Probabilistic generative models have been applied successfully in a wide range of applications that range from speech recognition and part of speech tagging, to machine translation and information retrieval, but, traditionally, applications such as reasoning have been thought to fall outside the scope of the generative framework for both theoretical and practical reasons. Theoretically, it is difficult to imagine, for example, what a reasonable generative story for first-order logic inference might look like. Practically, even if we can conceive of such a story, it is unclear how one can obtain sufficient amounts of training data. In this paper, we discuss how by embracing a less restrictive notion of inference, one can build generative models of inference that can be trained on massive amounts of naturally occurring texts; and text-based deduction and abduction decoding algorithms.

1 Introduction

Reasoning is one of the most studied areas of Artificial Intelligence (Russell and Norvig, 1995). In introductory courses to Artificial Intelligence, we teach students the elegant language of first-order logic (FOL) and spend significant time explaining how they can turn inference rules into logical formulas. At the end of an AI course, good students are capable to take naturally occurring conditionals, such as that shown in (1), and turn them into well-formed first-order formulas, such as the formula shown in (2).

"If a company cuts production, its stock price will fall." (**1**)

$$\forall c\ \forall t1\ \forall p1\ (company(c) \wedge cutProduction(c,t1) \wedge priceStock(c,t1,p1) \rightarrow \exists\ t2,p2\ (t2 > t1\ \wedge priceStock(c,t2,p2) \wedge p2 < p1)) \quad (\mathbf{2})$$

The power of FOL and other formal languages comes from their well-defined semantics. Given a formula such as (2) and a set of assertions about the world (3), we can use Modus Ponens and mechanically derive new statements that are in true in all interpretation in which formulas (2) and (3) are true. For example, we can