Decoding in Unified Hypergraph Framework

Lane Schwartz

Department of Computer Science & Engineering
University of Minnesota
Minneapolis, MN 55455, USA
lane@cs.umn.edu

Common parsing techniques are often viewed as a process of chart or tree construction. Such systems can alternatively be modeled as deductive rule system. Klein & Manning (2001) show that parsing can also be viewed as a process of hypergraph construction. Their paper provides a unified algorithm which generalizes the top-down Early parsing algorithm with the bottom-up CKY+ parsing algorithm.

The use of hypergraphs opens the door to the use of interesting graph algorithms. Huang and Chiang (2005) present a series of increasingly efficient methods for extracting a k-best list of parse trees from a hypergraph. Chiang (2007) presents cube pruning as an efficient algorithm for integrating a language model in hierarchical phrase based decoding. Huang & Chiang (2007) show that this algorithm can be defined in terms of hypergraph construction.

At the Second Machine Translation Marathon, one track of the open source convention was "Foundations for Hierarchical Models and Modular Search." Researchers working on this track examined how to further generalize the hypergraph decoding algorithm presented in Klein & Manning (2001). We considered how the phrase based decoding algorithms used in the Moses decoder could be framed in terms of hypergraph construction. We considered how the general hypergraph decoding algorithm in Klein & Manning (2001) might need to be adapted to handle hypergraphs in phrase based systems. And we considered how other modules, such as cube pruning and the calculation of inside or forward probabilities, could be incorporated into a generic, re-usable hypergraph decoding framework.

The result of this research track is a prototype framework, written in Java, for general decoding using hypergraphs. The framework is designed so that modules such as cube pruning and inside probability calculation can be written once, and then shared. This framework provides a common code base for hypergraph decoding which can be reused for very different machine translation strategies. A module for this framework was written which implements a very basic phrase based decoder.

I propose additional extensions to this framework. In particular, I propose to assist in implementing the CKY+ search strategy module, as well as an additional module to calculate the inside probability by traversing the hypergraph. Over the past two years, I have developed an independent implementation of a hierarchical phrase based decoder, including cube pruning; this code implements CKY+ parsing and k-best extraction directly using hypergraphs. My experience with these concepts, as well as my existing code, will be extremely helpful in implementing the CKY+ algorithm within the Moses hypergraph prototype framework.

Once a working prototype of the Moses hypergraph framework has been completed, this code base can be ported to C++ and integrated into the main Moses branch. The result of these efforts will allow developers to incorporate new and diverse search strategies into a common open source code base with relatively little new effort.