

# Training and Inference Methods over Translation Forests

Zhifei Li

## Abstract:

A hypergraph (or "packed forest") is a compact data structure that represents exponentially many trees. It can be used to represent the hypothesis space considered (for a given input) by a monolingual parser or a tree-based machine translation system (e.g., tree-to-string, string-to-tree, tree-to-tree, or string-to-string with latent tree structures). While a forest provides a much bigger (and hopefully better) hypothesis space than an n-best list, it also raises significant challenges. In this talk, I will give a systematic introduction to training and inference methods that work with hypergraphs. (These methods were developed by various people including myself.)

As background, I will first review different training procedures on the hypergraph -- perceptron, conditional random field (CRF), minimum error rate training (MERT), minimum-risk annealing, and max-margin training (e.g., MIRA). I will also highlight the pros and cons of these training methods and argue that minimum-risk annealing should be preferred for machine translation.

The key challenge in performing training and decoding on a hypergraph is to do efficient inference (e.g. computing the feature expectations during training time or finding the minimum risk translation during decoding time). Inference can be exact or approximate, depending on the problems (for example, Viterbi decoding can be done exactly, but MAP decoding with spurious ambiguity is intractable and thus approximations are required.) For exact inference, I will present a general semiring framework and discuss many useful semirings (e.g., Viterbi, K-best, and expectation semirings). For approximate inference, I will describe principled techniques (e.g., variational inference, belief propagation, and Gibbs sampling) as well as heuristic ones (e.g. beam- and cube-pruning).

In the end, I will give a brief introduction to our open-source toolkit Joshua, which implements most of the above training and inference methods. I will also give a live demo using Joshua.

## Bio:

Zhifei Li is currently a Ph.D. student in Johns Hopkins University's Computer Science Department and Center for Language and Speech Processing (CLSP). His interests are in machine translation, machine learning, applied algorithms and wireless networks.