Integer Linear Programming Inference for Conditional Random Fields

Dan Roth, Wen-tau Yih

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

Review by: Jake Pitkin

INTRODUCTION

In this paper the authors introduced the novel idea of using integer linear programming (ILP) in place of the Viterbi algorithm for inference when working with Conditional Random Field (CRF) models. Traditionally inference is performed by the Viterbi algorithm and the CRF operates under the first-order Markov assumption. This assumption coupled with the dynamic programming approach of the Viterbi algorithm provides an efficient inference step. The caveat of this approach is it prevents explicit modeling of general constraints such as long distance dependencies. Using ILP for inference enables us to introduce general constraints in a natural way. The authors empirically showed that this method provides a performance gain on the common task of semantic role labeling while maintaining the efficiency of traditional approaches.

STRENGTHS OF THIS APPROACH

This work presented an important formulation of using ILPs for inference in sequential labeling tasks. The authors give a nod to previous work (Punyakanok et al. 2004) which demonstrated that ILPs can be used for inference in the context of large scale NLP problems. They make improvements to this work by providing an elegant way to augment CRFs with general constraints while preserving the local and sequential constraints.

The authors do a strong job of setting of the stage for future work in this area. They demonstrate the solution that the Viterbi algorithm outputs is identical to the shortest path graph problem. They prove that the shortest path problem can be solved in polynomial time using linear programming. I found this to be vital to this paper's argument as it shows this approach is a sound replacement for the Viterbi algorithm.

A useful primer on how to convert general constraints into linear inequalities is presented in this work (which can be generalized to other tasks). There is a nice balance of introducing a new idea that can be generalized to any sequence labeling task and experimentally providing improvements over the state-of-the-art for semantic role labeling. It's clear the authors put in effort to make every section of their work easy to follow while remaining technical in their approach.

EXPERIMENTS

The experimental results are introduced by an explanation of the evaluation task of semantic role labeling (SRL). The explanation is helpful as this paper is applicable outside the realm of NLP and the authors give a concise but sufficient explanation. The explanation of syntactic features used in their experimental models are presented as a paragraph, where I think it would been more clear to present them in a table.

The goal of the experimental results is to show how general constraints improve the results of the common but hard task of SRL. The authors demonstrate this in what I think is the best way possible: by adding general constraints (five in total) one at a time during training. The standard precision, recall, and F1 score is provided at each step for two different CRF models and a performance gain is demonstrated over traditional Viterbi approaches.

Performing global inference during training is known to be not computationally convenient. To bypass inference during training, the authors explore a local training step followed by a post-training inference step to apply general constraints. They demonstrate that this approach (using all five general constraints) with a voted perceptron system provides better results than the CRF models that used inference during training.

FURTHER DEVELOPMENT

The authors demonstrate in their experimental results that local training can be performed and an inference step can be applied post-training with good results. I think this is an important direction to take further research as we don't induce the cost of inference during training while still benefiting from the performance gains of general constraints. It was demonstrated that local training with a final inference step provides a performance improvement on the task of semantic role labeling. Follow-up work that empirically explores other common NLP tasks with this approach could provide improvements over the current state-of-the-art. Many problems in NLP include a sequential labeling step somewhere in the problem's pipeline. Language has many constraints that we could apply with a post-training ILP inference step to prohibit non-sensical labelings and improve performance.

REFERENCES

Punyakanok, V., Roth, D., Yih, W., & Zimak, D. (2004). Semantic role labeling via integer linear programming inference. *Proc. of COLING-2004.*