Semi-supervised Learning for Natural Language Processing

John Blitzer
Natural Language Computing Group
Microsoft Research Asia
Beijing, China
blitzer@cis.upenn.edu

Xiaojin Jerry Zhu
Department of Computer Science
University of Wisconsin, Madison
Madison, WI, USA
jerryzhu@cs.wisc.edu

1 Introduction

The amount of unlabeled linguistic data available to us is much larger and growing much faster than the amount of labeled data. Semi-supervised learning algorithms combine unlabeled data with a small labeled training set to train better models. This tutorial emphasizes practical applications of semi-supervised learning; we treat semi-supervised learning methods as tools for building effective models from limited training data. An attendee will leave our tutorial with

1. A basic knowledge of the most common classes of semi-supervised learning algorithms and where they have been used in NLP before.

2. The ability to decide which class will be useful in her research.

3. Suggestions against potential pitfalls in semi-supervised learning.

2 Content Overview

Self-training methods Self-training methods use the labeled data to train an initial model and then use that model to label the unlabeled data and re-train a new model. We will examine in detail the co-training method of Blum and Mitchell [2], including the assumptions it makes, and two applications of co-training to NLP data. Another popular self-training method treats the labels of the unlabeled data as hidden and estimates a single model from labeled and unlabeled data. We explore new methods in this framework that make use of declarative linguistic side information to constrain the solutions found using unlabeled data [3].

Graph regularization methods Graph regularization methods build models based on a graph on instances, where edges in the graph indicate similarity. The regularization constraint is one of smoothness along this graph. We wish to find models that perform well on the training data, but we also regularize so that unlabeled nodes which are similar according to the graph have similar labels. For this section, we focus in detail on the Gaussian fields method of Zhu et al. [4].

Structural learning Structural learning [1] uses unlabeled data to find a new, reduced-complexity hypothesis space by exploiting regularities in feature space via unlabeled data. If this new hypothesis space still contains good hypotheses for our supervised learning problem, we may achieve high accuracy with much less training data. The regularities we use come in the form of lexical features that function similarly for prediction. This section will focus on the assumptions behind structural learning, as well as applications to tagging and sentiment analysis.

References

[1] Rie Ando and Tong Zhang. A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data. JMLR 2005.

[2] Avrim Blum and Tom Mitchell. Combining Labeled and Unlabeled Data with Co-training. COLT 1998.

[3] Aria Haghighi and Dan Klein. Prototype-driven Learning for Sequence Models. HLT/NAACL 2006.

[4] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-supervised Learning using Gaussian Fields and Harmonic Functions. ICML 2003.

Tutorial Abstracts of ACL-08: HLT, page 3,
Columbus, Ohio, USA, June 2008.