Co-Training

Based on

"Combining Labeled and Unlabeled Data with Co-Training"
by A. Blum & T. Mitchell, 1998

Problem:

Learning to classify data (ex: web pages) when the description of each example can be partitioned in 2 distinct views.

Assumption: Either view of the example would be sufficient for learning if we had enough labeled data, but:

Goal: use both views to allow inexpensive unlabeled data to augment a much smaller set of labeled examples.

Idea: 2 learning algorithms are trained separately on each view. Then each algorithm's predictions on new unlabeled examples are used to enlarge the training set of the other.

Empirical result on real data: The use of unlabeled examples can lead to significant improvement of hypotheses in practice.

Not presented here:

(see the paper)

Theoretical goal: **Provide a PAC-style analysis for this setting.**

More general: **Provide a PAC-style framework for the general problem of learning from both labeled and unlabeled data.**

Example

Classify web pages at CS departments at some universities as belonging or not to faculty members.

Views:

1. **the text appearing on the document itself**
2. **the anchor text attached to hyperlinks pointing to this page from other pages on the web.**

Use weak predictors, like

1. **"research interests"**
2. **"my advisor"**

Pages pointed to by links having the phrase "my advisor" can be used as 'probably positive' examples to further train a learning algorithm based on the words on the text page, and vice-versa.

Co-training Algorithm

Input:

L, **a set of labeled training examples**

U, **a set of unlabeled examples**

Create a pool U' **of examples by choosing** u **examples at random from** U**.**

Loop for k **iterations:**

use L **to train a classifier** h_1 **that considers only the** x_1 **view of** x

use L **to train a classifier** h_2 **that considers only the** x_2 **view of** x

select from U' p **most confidently labeled by** h_1 **as positive examples**

select from U' n **most confidently labeled by** h_1 **as negative examples**

select from U' p **most confidently labeled by** h_2 **as positive examples**

select from U' n **most confidently labeled by** h_2 **as negative examples**

add these self-labeled examples to L

randomly choose 2p+2n examples from U **to replenish** U'

Working example

Classify course home pages

1051 web pages at CS departments at several universities: Cornell, Washington, Wisconsin, and Texas

22% course pages

263 (25%) were first selected as a test set;

from the remaining data it was generated L, the set of labeled examples, by selecting at random 9 negative examples and 3 positive examples;

the ramining examples form U, the set of unlabeled examples.

use a Naive Bayes classifier for each of the two views.

Results

	page-based classifier	*hyperlink-based classifier*	*combined classifier*
supervised training	12.9	12.4	11.1
co-training	6.2	11.6	5.0

Explanation: The *combined* classifier uses the naive independent assupmtion:

$$P(Y \mid h_1 \wedge h_2) = P(Y \mid h_1)P(Y \mid h_2)$$

Conclusion: The *co-trained* classifier outperforms the classifier formed by supervised training.

Onother suggested practical application

Classifying segments of TV broadcasts, for instance:
learning to identify televised segments containing the US president.

Views: X_1 **– video images,** X_2 **– audio signals.**

Weakly predictive recognizers:

1. **one that spots full frontal images of the president's face**
2. **one that spots his voice when no background is present.**

Use co-training to improve the accuracy of both calssifiers.

Onother suggested practical application

Robot training, recognizing an open doorway using a collection of vision (X_1), sonar (X_2) and laser range (X_3) sensors.