

The self-training algorithm wraps around a base classifier and uses its own predictions through the training process. A base learner is first trained on a small number of labeled examples, the initial training set. The classifier is then used to predict labels for unlabeled examples (prediction step) based on the classification confidence. Next, a subset S of the unlabeled examples, together with their predicted labels, is selected to train a new classifier (selection step). Typically, S consists of a few unlabeled examples with high-confidence predictions. The classifier is then re-trained on the new set of labeled examples, and the procedure is repeated (re-training step) until it reaches a stopping condition. As a base learner, we employ the decision tree classifier in self-training. The most well-known algorithm for building decision trees is the C4.5 algorithm, an extension of Quinlan's earlier ID3 algorithm.

Algorithm 1 Outline of the Self-Training algorithm

```

Initialize:  $L, U, F, T$  ;  $L$ : Labeled data;  $U$ : Unlabeled data;
 $F$  : Underlying classifier;  $T$  : Threshold for selection;
 $Iter_{max}$  : Number of iterations;  $\{P_l\}_{l=1}^M$ : Prior probability;
 $t \leftarrow 1$ ;
while ( $U \neq \text{empty}$ ) and ( $t < Iter_{max}$ ) do
  -  $H^{t-1} \leftarrow BaseClassifier(L, F)$ ;
  for each  $x_i \in U$  do
    - Assign pseudo-label to  $x_i$  based on classification confidence
  - Sort Newly-Labeled examples based on the confidence
  - Select a set  $S$  of the high-confidence predictions according to  $n_l \propto P_l$ 
    and threshold  $T$  // Selection Step
  - Update  $U = U - S$ ;  $L = L \cup S$ ;
  -  $t \leftarrow t + 1$ 
  - Re-Train  $H^{t-1}$  by the new training set  $L$ 
end while
Output: Generate final hypothesis based on the new training set

```

The goal of the selection step in Algorithm 1 is to find a set unlabeled examples with high-confidence predictions, above a threshold T . This is important, because selection of incorrect predictions will propagate to produce further classification errors. At each iteration the newly-labeled instances are added to the original labeled data for constructing a new classification model. The number of iterations in Algorithm 1 depends on the threshold T and also on the pre-defined maximal number of iterations, $Iter_{max}$.