Weak Supervision with Incremental Source Accuracy Estimation

Richard Correro
rcorrero@stanford.edu

Weak Supervision Overview

Weak supervision approaches obtain labels for training data using noiser or higher-level sources than traditional supervision [1]. Recently proposed methods use generative models to combine labels from multiple noisy sources to generate probabilistic labels when true labels are unknown [2].

Varma et. al. [2] modeled the joint distribution of the weak supervision sources which produce noisy labels $\lambda_i$ for $1 \leq i \leq m$ and the latent true label $Y$ as a Markov random field associated with a graph $G = (V,E)$ where $V = \{\lambda_1, \ldots, \lambda_m\} \cup \{Y\}$. If $\lambda_i$ is not independent of $\lambda_j$ conditioned on $Y$ and the other sources then $(\lambda_i, \lambda_j) \in E$.

$$f_0(\lambda_1, \ldots, \lambda_m) = \frac{1}{Z(G)} \exp \left( \sum_{i \in V} \theta_i \phi_i + \sum_{(i,j) \in E} \theta_{ij} \phi_{ij} + \sum_{i \in V} \theta_i \phi_{i,Y} \right)$$

where $\phi_{ij} = 1$ if $i \leq j \leq m + 1$ define the canonical parameters associated with the supervision sources and the $Y$ and $X$ in a partition function. Using these parameters we may compute $f_0(Y|\lambda_1, \ldots, \lambda_m)$ for each of the training examples.

Problem

An incremental algorithm for learning source accuracies and graph structure would allow us to use weak-supervision labeling methods in out-of-core or online settings. For example, an incremental learning algorithm would allow for generating labels on streaming data in real-time. Although algorithm 1 estimates both the source accuracies $\mu$ and the dependency structure $\mu^*$, it requires the entire dataset and cannot be implemented iteratively. Algorithm 2 is much more efficient but sometimes not yield an estimate of the dependency structure of the labeling functions.

Solution

We develop a method which combines algorithms 1 and 2 to estimate both $\mu$ and $\mu^*$ incrementally.

Method

Given an initial batch of samples, we obtain $\lambda_0$ and $\mu_0$ using algorithm 1. From $\lambda_0$, we may determine the graph structure and dependency mask $\mu_0$. Let $\mu$ denote the new estimated accuracies of the supervision sources. Given $\mu$ we may estimate $\mu^*$ for a new batch $X'$ of examples using

$$\hat{\mu} = \arg\max_{\mu} \left( (xf)^T \mu^* + \xi(\mu) \right)$$

which may be solved by least-squares. From $\hat{\mu}$ we estimate the source accuracies on $X'$

$$\hat{x} = \sum_{k=1}^{m} (xf)^T \mu^*_k$$

$\hat{\sigma} = \sum_{k=1}^{m} \hat{x}^2 \mu^*_k

\hat{\mu}' = \hat{\mu} + \hat{\sigma}$$

Using $\hat{\mu}'$ we update $\mu$:

$$\mu' = (1 - \alpha) \mu + \alpha \hat{\mu}'$$

We thus estimate source accuracies using an exponentially-weighted moving average of the batch estimates $\mu'$.

Transfer Learning

Transfer learning deals with the same, but source and target domains are different [4].

Results

We test our model using tweets concerning a presidential debate labeled by sentiment ("positive", "neutral", and "negative"). We sort the tweets by timestamps, form a batch, and stream the model to the dataset to simulate the imagined use case in which the model is used to generate labels in real-time.

We compare the per-batch accuracy of the labels generated by our model with the accuracy of the non-incremental model. We test different values of $\alpha$ and compare accuracies.

Tests

We utilize our incremental algorithm in such a scenario. We attempt to label tweets by sentiment received in real-time, as they are broadcast, using the following weak supervision sources:

- Naive Bayes model trained on a set of movie reviews labeled by sentiment ("positive" or "negative")

- TextBlob Pattern Analyzer - A model which uses lookup tables to generate polarity and subjectivity estimates for a text.

- Naive Bayes model trained on a dataset consisting of tweets associated with US airlines and labeled by sentiment ("positive", "neutral", and "negative")

Data

Previous Work

(Algorithm 1 comes from [2] and algorithm 2 from [3]).

References


