An Introduction to Positive and Unlabeled Learning

Edward Li

March 17, 2021

1 Introduction

Positive and Unlabeled Learning (PU Learning) is a semi-supervised learning technique designed to address scenarios where only positive examples are labeled, while the remaining data is unlabeled and may include both positive and negative instances. This paper introduces the fundamental concepts and methodologies of PU Learning, emphasizing key considerations for its practical implementation.

PU Learning is particularly valuable in situations where labeling negative examples is challenging, costly, or impractical, or where only positive labels currently exist due to privacy concerns or other constraints [Elkan and Noto, 2008]. The goal is to develop a classifier capable of effectively distinguishing between positive and negative instances by leveraging the labeled positive examples and the unlabeled data. Traditional supervised learning methods are unsuitable for such settings, as they require labeled examples from both classes.

2 Methodology

Let x denote a data point, and $y \in \{0, 1\}$ be its true label, where $y = 1$ indicates a positive instance and $y = 0$ indicates a negative instance. Let $s \in \{0, 1\}$ indicate whether x is labeled, where $s = 1$ if x is labeled (and therefore known to be positive), and $s = 0$ if x is unlabeled.

Since only positive examples are labeled, we have:

$$p(s = 1 \mid x, y = 0) = 0. \tag{1}$$

This means that negative instances are never labeled as positive.

PU Learning assumes that the labeled positive data are randomly selected from all positive examples. This assumption is known as the **Selected Completely At Random** (SCAR) assumption:

$$p(s = 1 \mid x, y = 1) = p(s = 1 \mid y = 1) = c, \tag{2}$$

where c is a constant representing the probability that a positive instance is labeled.

Under this assumption, s and x are conditionally independent given y. That is, the probability that a positive instance is labeled does not depend on its features x.

Our training set is a random sample from the distribution $p(x, y, s)$ that satisfies equations (1) and (2).

By training a classifier $g(x)$ to distinguish between the labeled positive instances ($s = 1$) and the unlabeled instances ($s = 0$), we obtain an estimate of $p(s = 1 \mid x)$. We refer to $g(x)$ as a **non-traditional classifier** because it is trained to predict the labeling probability rather than the class label.

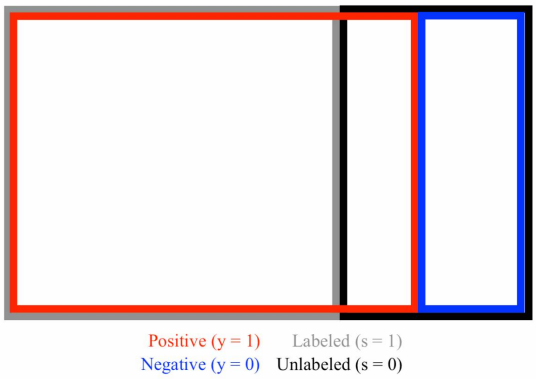

Figure 1: An illustration of the relationship between $p(y = 1 \mid x)$, $p(s = 1 \mid x)$, and c.

The key insight in PU Learning is that we can recover the probability that an instance is positive using:

$$p(y = 1 \mid x) = \frac{p(s = 1 \mid x)}{p(s = 1 \mid y = 1)} = \frac{g(x)}{c}. \tag{3}$$

Equation (3) allows us to compute $p(y = 1 \mid x)$ using the classifier $g(x)$ and the constant c.

In practice, $g(x)$ is an approximation of $p(s = 1 \mid x)$, learned from finite data, and may not be exact. Moreover, $c = p(s = 1 \mid y = 1)$ can be estimated from the data. One common estimator is:

$$\hat{c} = \frac{1}{n} \sum_{x \in P} g(x), \tag{4}$$

where P is the set of labeled positive instances, and n is the number of labeled positive instances.

The accuracy of PU Learning depends on accurately estimating c and on the performance of the classifier $g(x)$. The estimation of c improves with the number of labeled positive instances.

Selecting an appropriate model for $g(x)$ is crucial. The choice of classifier affects the performance of the PU Learning method. It is important to evaluate different classifiers to determine which one performs best in a given scenario.

3 Conclusion

PU Learning provides a framework for learning classifiers when only positive and unlabeled data are available. By leveraging the SCAR assumption and adjusting the output of a classifier trained on labeled and unlabeled data, we can estimate the probability that an instance is positive.

The general procedure for PU Learning is as follows:

1. Train a classifier $g(x)$ to approximate $p(s = 1 \mid x)$ by distinguishing labeled positive instances from unlabeled instances.

2. Estimate the constant $c = p(s = 1 \mid y = 1)$ using the labeled positive instances:

$$\hat{c} = \frac{1}{n} \sum_{x \in P} g(x).$$

3. For a new data point x_0, compute $p(y = 1 \mid x_0)$ using:

$$p(y = 1 \mid x_0) = \frac{g(x_0)}{\hat{c}}.$$

The performance of PU Learning depends on the accuracy of $g(x)$ and the estimation of c. Increasing the number of labeled positive instances improves the estimation of c. Additionally, choosing the appropriate classifier for $g(x)$ is essential for achieving good results.

References

Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, page 213–220, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605581934. doi: 10.1145/1401890.1401920. URL `https://doi.org/10.1145/1401890.1401920`.