

A Density-based Approach for Positive and Unlabeled Learning

Cholwich Nattee^{*1} Nirattaya Khamsemanan^{*1} Thanaruk Theeramunkong^{*1}
Masayuki Numao^{*2}

^{*1} Sirindhorn International Institute of Technology, Thammasat University, Thailand

^{*2} The Institute of Scientific and Industrial Research, Osaka University, Japan

Positive and Unlabeled learning (PU learning) is a group of supervised machine learning techniques that aims to construct a classification model from a set of positive and unlabeled examples without any negative example. PU learning can be used when it is very difficult or impossible to prepare a set of negative examples. We propose a technique to automatically extract a set of reliable negative examples from the collected unlabeled examples. The proposed technique is based on the idea that an unlabeled example has a high chance to be a negative example when it is located far from any positive examples and in an area with various unlabeled examples. From the extracted negative examples, we can then use an ordinary supervised machine learning technique to build a classification model. To evaluate the proposed technique, we conduct experiments based on sets of generated examples. The experimental results confirm the effectiveness of the technique.

1. Introduction

PU learning focuses on constructing classification models from only positive and unlabeled examples without negative examples. This learning technique plays an important role in the application domains where labeling negative examples requires a lot of effort or costs. It is very difficult or impossible to label negative examples in some applications, e.g. text categorization where one example may belong to multiple classes, and it is usually trivial for users to specify the classes that the example belongs to, but it is difficult to specify the class that the example does not belong to. A basic approach to handle this problem of labeling negative examples by using an ordinary supervised machine learning technique is to construct a classifier by using the examples in the target class as positive examples, and the examples in the other classes as negative examples. However, this approach may result in an inefficient classifier since the provided negative examples are not the actual negative examples. Ambiguities in the training examples then result in a classifier that is unable to efficiently discriminate the positive examples from the negative ones. A learning technique that efficiently builds a binary classifier from only positive and unlabeled examples would allow us to overcome this problem.

Up to now, various techniques for PU learning have been proposed [Liu 03, Li 10, Lee 03]. Most of the PU learning techniques are composed of two steps:

1. Identify a set of negative examples from the provided set of unlabeled examples.
2. Construct a classification model based on the extracted negative examples.

Some PU learning techniques perform both steps iteratively. They initially identify the most reliable negative examples and construct a classifier. The obtained classifier is then used to identify more negative examples before adjusting the classifier. Biased SVM [Liu 03] is a PU learning technique that combines both steps. It works similarly to the ordinary soft-margin support vector machines that accept two classes of examples. Here, the unlabeled examples are simply considered negative examples. However, the optimization constraint of the Biased SVM is set so that the penalty of misclassifying negative examples is very small, while that of misclassifying positive examples is high.

In this paper, we focus on the first step of the PU learning that identifies a set of negative examples. We propose a technique to extract a set of reliable negative examples from the unlabeled examples by using the concept of density in the space. The basic idea of this paper originates from the diverse density [Maron 98] proposed for multiple-instance learning.

Multiple-instance learning is a collection of machine learning techniques focusing on the classification problem that the positive examples cannot be exactly labeled, but a group of examples is labeled positive when the group is considered containing at least one positive example. In this setting, a group of examples is considered negative when all the examples in the group can be labeled negative. The multiple-instance learning is thus a supervised learning technique with ambiguities in positive examples. The diverse density is a technique proposed for solving multiple-instance learning. It measures the density of the search space based on the ambiguities in the provided positive examples. The area in the search has high diverse density when the examples from various positive groups located together, and it is far away from any negative examples.

We modify the diverse density so that it is appropriate for handling ambiguities in the PU learning. Here, an unlabeled example has a high chance to be negative when it is far from any positive examples, and several unlabeled ex-

amples are located nearby. By measuring the density of the unlabeled examples, we can extract a set of reliable negative examples. The extracted negative examples can later be used to build a classifier.

Next, we will explain the proposed technique including how to compute the proposed density for PU learning. After that, we conduct some experiments to confirm the effectiveness of the proposed technique.

2. The Proposed Density Function

We first formulate the PU learning problem as follows: Given a set of N examples composing of m positive examples, $P = \{(x_1, +1), \dots, (x_m, +1)\}$, and $N - m$ unlabeled examples, $U = \{x_{m+1}, \dots, x_N\}$, we want to find a hypothesis that classifies an unknown examples into either positive (+1) or negative (-1). Here, the number of positive examples is usually much smaller than that of unlabeled examples due to the cost of data preparation.

We propose a PU-based density d_q that weights an unlabeled example q . This density composes the densities from both positive and unlabeled examples as

$$d_q = (d_{q+})^\alpha (d_{qu})^\beta \quad (1)$$

where d_{q+} is the density from the positive examples, d_{qu} is the density from the unlabeled examples, and α and β are weighting parameters for combining two densities. Here, we set a criterion that $\alpha + \beta = 1.0$.

We apply the noisy-or model to compute the density from the positive examples. The density becomes higher when the considered example is located closely to the positive examples in the space. Therefore, we have

$$d_{q+} = 1 - \prod_{i=1}^m (1 - \text{distance}(x_q, x_i, s)) \quad (2)$$

where $\text{distance}(x, y, s)$ measures the distance between two examples x and y with a scaling factor s . In this paper, the distance is computed from

$$\text{distance}(x, y, s) = e^{-\|x-y\|_s^2} \quad (3)$$

We compute the density from the unlabeled examples from the idea that the density is high when the considered example is among several unlabeled examples. The density from the unlabeled examples is then computed from the average distance between the considered example, and the other unlabeled examples. We then have

$$d_{qu} = \frac{\sum_{i=m+1}^N \text{distance}(x_q, x_i, s)}{N - m} \quad (4)$$

Based on our preliminary experiment, we found that the density from the positive examples plays a more important role than that from the unlabeled examples. This is intuitive since the positive examples possess more information than the unlabeled examples. The density from the unlabeled examples is added to Equation 1 to show that the considered example is located in a collection of examples.

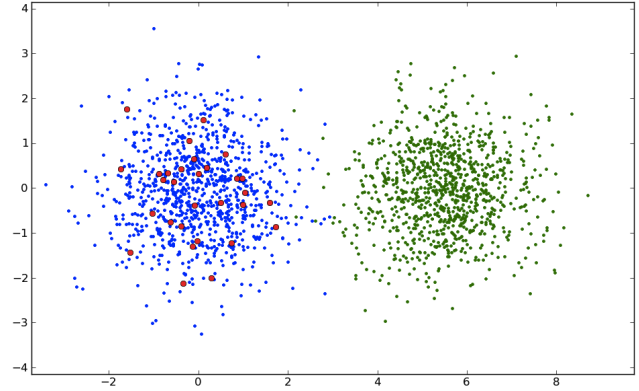


Figure 1: A simple example of a PU learning dataset.

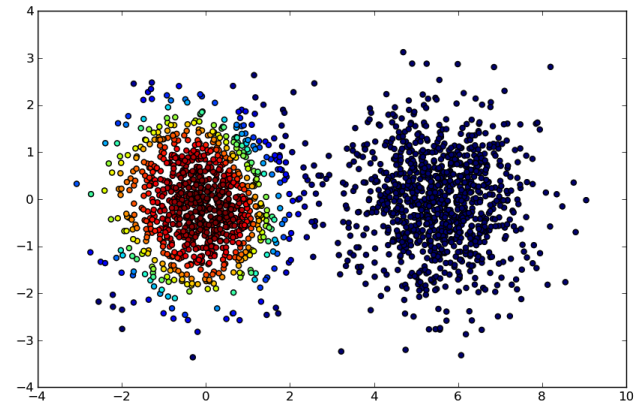


Figure 2: The simple dataset with density values.

Based on the observation, we therefore set the value of α to be higher than the value of β .

Figure 1 shows a dataset of PU learning. The bigger dots in the plot show the labeled positive examples sampled from the original distribution. After applying the proposed technique, we can plot the same dataset using different colors to show the density values in Figure 2. It can be seen that the proposed technique can discriminate between positive and negative examples.

3. Experiments

To evaluate the proposed technique, we conduct some experiments based on generated datasets.

3.1 Experiment 1:

Portions of Positive Examples

This experiment aims to show the effectiveness of the proposed technique with different portions of positive examples. We generate a multivariate normal distribution dataset composing of 1,000 positive, and 1,000 negative examples. Each example has 10 features. We randomly split the set of positive examples into two subsets with different portions i.e. 5% : 95%, 10% : 90%, 30% : 70%, 50% : 50%, and 80% : 20%. Then, the first subset is used as the positive examples, while the second subset is merged with the negative examples to produce the set of unlabeled exam-

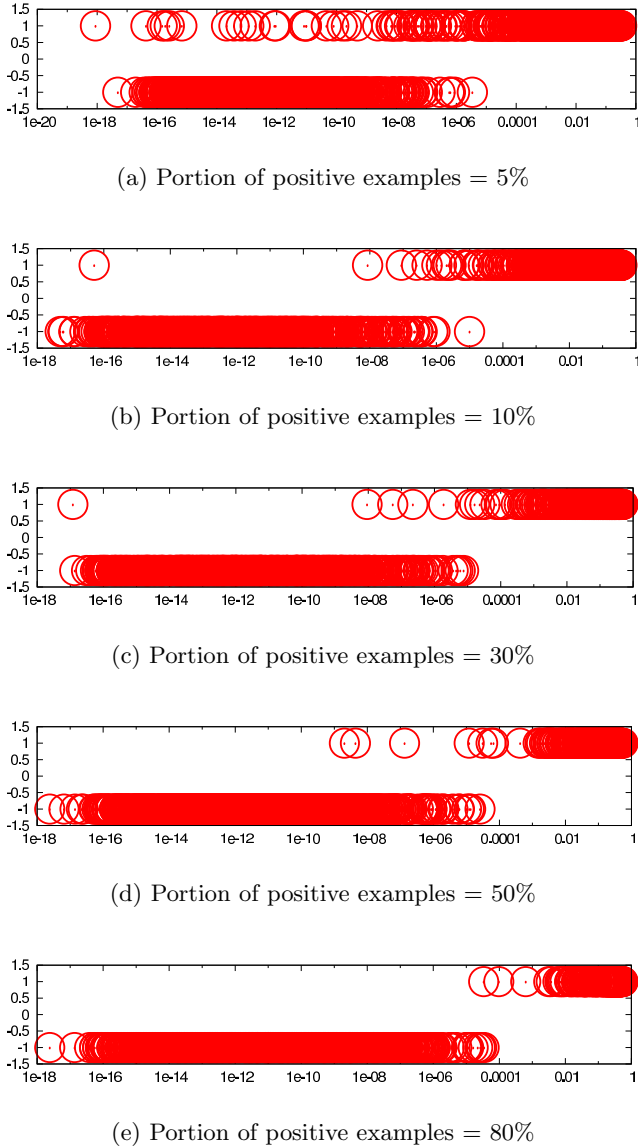


Figure 3: Density values on different portions of positive examples

ples. We then apply the proposed technique to compute the density of each of the unlabeled examples.

Figure 4 shows the plots of the obtained densities of the positive and negative examples in the set of unlabeled examples. All the densities are plotted in a log-scale. It is obvious that when the portion of the positive examples increases, the difference between the densities also increases. The experimental results also show that the proposed density function is able to identify the positive examples, even if the portion of positive examples is small (e.g. 5%). We can set a threshold value based on the group of examples with the highest density values for extracting the reliable positive examples, as well as the reliable negative examples. This technique can also be considered a transductive learning technique working on the set of unlabeled examples. Moreover, the technique can be used with any supervised learning technique by extracting the reliable positive

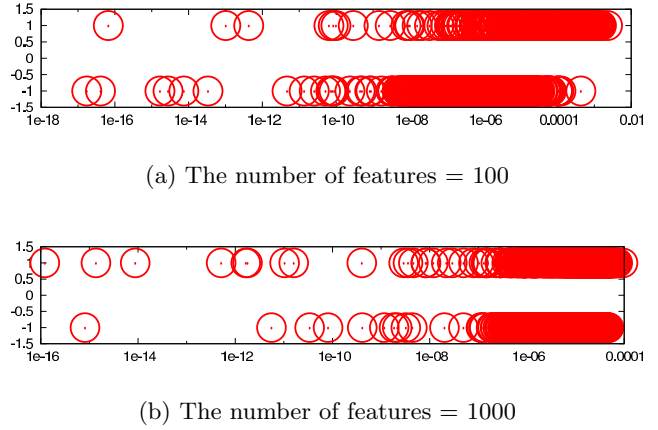


Figure 4: Density values on different numbers of features

and negative examples before constructing a classification model.

3.2 Experiment 2: The Curse of Dimensionality

Since the proposed technique is based on the distance between two examples, the curse of dimensionality may cause the difference between the densities to become insignificant. We then conduct an experiment to show the effect of the curse of dimensionality to the proposed technique. We generate a dataset using the same parameters as the previous experiment but each example is composed of 100 and 1,000 features. Then, 10% of the positive examples are randomly selected to be used as the set of positive examples. All the rest are merged to produce the set of unlabeled examples. We use the proposed technique to calculate the densities of the unlabeled examples.

Figure ?? shows the plots of the densities. It can be seen that the difference between densities cannot be recognized when the number of features becomes very high. Thus, the curse of dimensionality affects the proposed technique. We need to apply some dimensionality reduction to the dataset with a high number of features before applying the proposed technique.

3.3 Experiment 3: Multiclass Problem

As stated in the Introduction, PU learning can be applied to multiclass problems. Naturally, we would like to determine whether or not our proposed technique is able to handle a multiclass problem as well. Toward this end, we have conducted the following experiment.

We generate a dataset composing of seven classes of examples. Each class is a multivariate normal distribution composing of 100 examples. Each example has 40 attributes. We label the first class as positive, and the other classes as negative. Then, we evaluate the proposed technique by randomly selecting 10 examples from the first class to be the positive set. Table 1 shows the experimental results which are the average density value obtained for each class of examples.

The results show that the average density value of the first

Table 1: Average density values for the seven-class problem

Class Number	Average Density Value
1	2.41×10^{-4}
2	1.41×10^{-6}
3	4.89×10^{-7}
4	2.04×10^{-7}
5	8.78×10^{-7}
6	3.65×10^{-6}
7	6.38×10^{-6}

class of examples is significantly higher than the averages from the other classes since only 10 examples are used as the positive examples. This shows the effectiveness of the proposed density.

4. Conclusion and Future Work

In this paper, we proposed a technique to calculate density values for the unlabeled examples in PU learning. The proposed technique is based on the distance between two examples. The density of an unlabeled example is high when the example is located near some positive examples, and it is located in an area with several unlabeled examples. We conducted experiments to evaluate the performance of the proposed techniques. The experimental results showed that the proposed technique is able to discriminate between positive and negative examples mixed in the set of unlabeled examples. However, the proposed technique suffers from the curse of dimensionality since it is based on distance between two examples, and the distance becomes insignificant when the number of features becomes very high. We need to apply a dimensionality reduction technique to preprocess the dataset before applying the proposed technique.

For future work, we plan to combine the proposed technique with some supervised machine learning techniques in order to construct a classifier from the positive and unlabeled examples.

Acknowledgment

This work was performed under the Cooperative Research Program of “Network Joint Research Center for Materials and Devices” at the Institute of Scientific and Industrial Research, Osaka University, Japan.

References

- [Lee 03] Lee, W. S. and Liu, B.: Learning with Positive and Unlabeled Examples using Weighted Logistic Regression, in *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, DC, USA (2003)
- [Li 10] Li, X.-L., Zhang, L., Liu, B., and Ng, S.-K.: Distributional similarity vs. PU learning for entity set expansion, in *Proceedings of the ACL 2010 Conference Short*

Papers, pp. 359–364, Stroudsburg, PA, USA (2010), Association for Computational Linguistics

- [Liu 03] Liu, B., Dai, Y., Li, X., Lee, W. S., and Yu, P.: Building Text Classifiers Using Positive and Unlabeled Examples, in *Proceedings of the Third IEEE International Conference on Data Mining (ICDM-03)*, Melbourne, FL, USA (2003)

- [Maron 98] Maron, O. and Lozano-Prez, T.: A Framework for Multiple-Instance Learning, in *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pp. 570–576, MIT Press (1998)