Active Learning with Clustering and Unsupervised Feature Learning

Saul Berardo(✉), Eloi Favero, and Nelson Neto

Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, PA 66075-900, Brazil
{saulberardo,favero,nelsonneto}@ufpa.br

Abstract. Active learning is a type of semi-supervised learning in which the training algorithm is able to obtain the labels of a small portion of the unlabeled dataset by interacting with an external source (e.g. a human annotator). One strategy employed in active learning is based on the exploration of the cluster structure in the data, by using the labels of a few representative samples in the classification of the remaining points. In this paper we show that unsupervised feature learning can improve the "purity" of clusters found, and how this can be combined with a simple but effective active learning strategy. The proposed method shows state-of-the art performance in MNIST digit recognition in the semi-supervised setting.

Keywords: Active learning · Clustering · Unsupervised feature learning

1 Introduction

Active learning algorithms can exploit labeled and unlabeled data, as in the general semi-supervised setting, but instead of using a predefined set of labeled examples, the training algorithms is allowed to query an oracle during training (e.g. a human annotator) to obtain the labels of the training examples considered the most informative. This approach is particularly useful in problems in which unlabeled data is abundant, but obtaining labels is expensive, such as in protein classification [18].

There are two main strategies employed in active learning. The most commonly explored in previous works is based on the use of labeled examples near the decision boundary of a discriminative classifier to direct the search in the hypothesis space in a efficient way. This strategy assumes that points near the decision boundary are more informative and gives little importance to points farther away [5]. The second strategy is based on the manifold hypothesis, according to which points of real data concentrate in the neighborhood of a low-dimensional manifold embedded in a high-dimensional space [4,12]. In an ideal scenario, a clustering algorithm would isolate points of different classes in their own clusters and one labeled sample for each would be enough to correctly classify the remaining examples [6]. In practice finding good clusters in complex and high-dimensional data such as images is not easy.

D. Barbosa and E. Milios (Eds.): Canadian AI 2015, LNAI 9091, pp. 281–290, 2015.
DOI: 10.1007/978-3-319-18356-5_25

It has been long suggested that images could be recognized by multiple layers of feature detectors [15], but it has not been until recently that effective methods for training models with multiple layers have been devised [3,8]. The first successful techniques relied on a greedy layer-wise unsupervised procedure (pre-training) for initializing the network before applying backpropagation. One perspective of pre-training procedure is that each layer learns to extract good features from the layer below. In image recognition, for example, a linear classifier in the output layer can more easily separate different classes in terms of higher level features, such as object parts, than in terms of the original input, such as pixels.

In this work we explore the intuition that the same layer-wise unsupervised training procedure used in deep learning can also be used to transform the input from the original space to a feature space in which clusters are more easily identifiable. Our method is composed by the following steps:

1. We start by applying an unsupervised feature learning technique to extract features from unlabeled data.
2. We then convert the raw input into features and use a clustering algorithm to find clusters in the feature space.
3. Then we query the labels of cluster representatives and classify all points in the same clusters with the same labels.
4. We optionally use a probabilistic model to determine the quality of the classification and remove the "excess" of points classified with low probabilities.
5. Finally, we use the training examples with the labels found as the training set of a neural network.

Our experiments demonstrate that unsupervised feature learning can improve clustering purity and we show how the learned features can be combined in a simple but effective active learning strategy.

2 Unsupervised Feature Learning

Every clustering algorithm needs a measure of dissimilarity (or similarity), such as the euclidean distance used in k-means. In certain datasets it is easy to notice that the euclidean distance does not represent well our intuitive notion of what similar points are. In digit images for example a shifted version of a digit will stop sharing most pixels in common with the original version and thus their distance will be large, although they still belong to the same class. Unsupervised feature learning techniques can be used to convert the raw input to a feature space where the distance metric can more easily represent the actual notion of dissimilarity between data points. If we have features which represent objects parts, for example, it is easier to tell whether two images are similar or not by checking if they share the same object parts, rather than by comparing their raw pixels.