Semi-Supervised Training of Cell-Classifier Neural Networks

Gergely Pap, Tamás Grósz, László Tóth

Abstract: Nowadays, microscopes used in biological research produce a huge amount of image data. Manually processing the images is a very time-consuming and resource-heavy task, so the development and implementation of new automatic systems is required. Moreover, as we have access to a large amount of unlabeled data, while labels are only available for a small subset, these novel methods should be able to process large amounts of unlabeled data with minimal manual supervision. Here, we apply neural networks to classify cells present in biological images, and show that their accuracy can be improved by using semi-supervised training algorithms.

Keywords: DNN, RBM, semi-supervised learning

Introduction

The dataset used in this paper consists of features extracted from cells based on image processing considerations and biological requirements. The final goal is to automatically classify the cells by their properties regarding their gene-expression profile and other important cellular descriptors. As only a small subset of the database was annotated manually, here we seek training algorithms that allow us to exploit the unlabeled data to improve classification accuracy.

The processing of large datasets is a task which requires considerable manual labor and time, so usage of an automatic system is widely recommended. Artificial Neural Networks (ANNs) have made significant progress in the past few years within a wide area of scientific fields (for example: image processing, voice recognition, machine translation etc.). Neural networks generally require an adequately large annotated training database before they can be used to classify previously unseen or new data. However, creating an annotated database of considerable size is a labor-heavy and costly process that requires tedious manual work. Hence, it would be optimal if human supervision could be minimalised during the creation of the training data, moreover, if unlabeled data could be used to improve classification accuracy. The above-mentioned goals are achievable using Restricted Boltzmann Machines (RBMs) [2]. In the case of RBMs, an unsupervised learning method can be applied called Contrastive Divergence (CD), which pretrains the network using unlabeled data. Then, the subsequent regular supervised learning step using the labeled data usually attains better classification accuracies than the ones it achieved without pretraining.

During our research we used conventional "shallow" and deep artificial neural networks as a baseline trained via labeled data and then we compared their performance with two RBMs, which were configured to apply different learning algorithms (a generative and a discriminative one) using the unlabeled data set.

Data

The data used in our study comes from images of cells in different biological states. The cells were captured using fluorescent microscopy and 254 features based on biological properties (regarding gene-expression and other cellular descriptors) were then extracted using image processing. Biologists then manually labeled the cells in a small subset of the data, sorting them into 12 different classes. There are roughly 1,8 million entities in the dataset and only 2600 were labeled by hand. Classes included cells separated by their membrane features (normal, ruptured, vesical), by their place in the cells' life-cycle (normal, dying, apoptotic) and by their visual properties in connection to the microscopical and image processing traits (normal, segmentation error and well-defined contour). The features included the position of the cell, distance from other cells, properties of the different colour channels, biological attributes of cellular organs (e.g. membrane, nucleus, cytoplasm).

We split the labeled data into three subsets: 2100 were used for supervised training, 300 for validation (also known as development dataset) and 200 for testing.

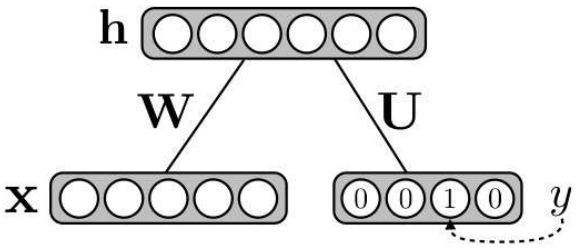

Figure 1: ClassRBM structure where x is the input vector, h denotes the units of the hidden layer and y represents the labels. Furthermore, W and U are the respective weight matrices.

Methods

Restricted Boltzmann Machines are often used in machine learning problems as feature extractors for another learning algorithm or as a good initialization strategy for deep neural networks. However, it has been shown that RBMs can also be utilized as a standalone discriminative classification method. In this work, Classification Restricted Boltzmann Machines (ClassRBMs) were pretrained using unlabeled data to improve classification accuracy.

RBMs have two layers of neurons (hidden and visible), where connections within the layers are forbidden. The information extracted by the hidden layer tries to model the joint distribution of the input and the hidden states, passing the data on to the next set of layers. Though standard RBMs are trained only to model the inputs of a task, the advantage of ClassRBMs is that they can also model the joint distribution of the inputs and the corresponding target classes, as Figure 1 illustrates. Traditionally, training is finished by converting the RBM into a regular feed-forward network which is trained on labeled data using backpropagation [3].

Classification Restricted Boltzmann Machines

In a generative setting, the RBM uses an energy function that models the joint distribution of every possible visible and hidden vector pair

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)}, \tag{17}$$

where the denominator Z converts the result of equation (1) into a probability, it is calculated as the summation of all possible hidden and visible vector pairs as follows

$$Z = \sum_{v,h} e^{-E(v,h)}. \tag{18}$$

The RBMs energy function is given by

$$E(v,h) = -b^T v - c^T h - v^T W h \tag{19}$$

where b is the bias of the visible layer and c is the bias of the hidden layer [1].
We then can calculate an estimate of the gradient regarding the model parameters for any $\theta \in \Theta$

$$\frac{\delta logp(y_i, x_i)}{\delta \theta} = -E_{h|y_i,x_i}[\frac{\delta}{\delta\theta} E_{y_i,x_i,h}] + E_{y,x,h}[\frac{\delta}{\delta\theta} E_{y,x,h}]. \tag{20}$$

As we were interested in the prediction of the correct labels, we used Discriminative Restricted Boltzmann Machines to model $p(y|x)$ instead of $p(y,x)$

$$L_{(disc)}(D_{train}) = -\sum_{i=1}^{|D_{train}|} logp(y_i|x_i). \tag{21}$$

The RBMs minimizing the function $L_{(disc)}$ are called Discriminative Restricted Boltzmann Machines (DRBMs). An RBM containing enough layers can be considered as an universal approximator for binary

inputs, which means that DRBMs are also universal approximators for the joint distribution of binary inputs and class labels y.

DRBMs use an algorithm called Contrastive Divergence (CD) for training, however in the model $p(y|x)$ can be calculated directly, we can then compute the exact gradient using Algorithm 1:

Algorithm 1 Contrastive Divergence for discriminative RBM training

1: INPUT: training pair (y_i, x_i) and learning rate λ
2: a $\leftarrow$ b means a is set to the value of b
3: a $\sim$ b means a is sampled from p
4: Positive Phase
5: $y^0 \leftarrow y_i, x^0 \leftarrow x_i, \hat{h}^0 \leftarrow sigm(c + Wx^0 + U\overrightarrow{y}^0)$
6: Negative Phase
7: $h^0 \sim p(h|y^0, x^0), y^1 \sim p(y|h^0), x^1 \sim p(x|h^0)$
8: $\hat{h}^1 \leftarrow sigm(c + Wx^1 + U\overrightarrow{y}^1)$
9: **Update**:
10: **for** $\theta \in \Theta$ **do**
11: $\quad \theta \leftarrow \theta - \lambda(\frac{\delta}{\delta\theta}E(y^0, x^0, \hat{h}^0 - (\frac{\delta}{\delta\theta}E(y^1, x^1, \hat{h}^1)$
12: **end for**

A semi-supervised approach introduces constraints to the model, in this case the decision surface of the model is modified by the requirements imposed upon the RBM, as it tries to approximate a good generative model of the unlabeled data.

For our experiments we used an RBM implemented in MATLAB. This toolbox also contained (an ANN) implementation and the code to execute both discriminative and generative training of RBMs in a semi-supervised setting [4].

Results

To have a baseline result first we used a DNN with 3 hidden layers each containing 500 nodes with sigmoid activation function. The network was trained using the standard backpropagation algorithm. The DNN was trained for 60 epochs with 0.1 learning rate. The baseline's classification accuracy was around 67% on average. One classRBM had 100 neurons in three layers with 0.01 learning rate, the other 500 neurons in one layer and a learning rate of 0.05. The classRBM trained only on the labeled data performed a little better, averaging around 74%. However when the discriminative classRBM was used in conjunction with the unlabeled data in a semi-supervised learning scenario, the numbers of the correctly classified cells increased to 85%. After experimenting with the parameters of the training, we found that the values presented here yielded the best results in each particular case.

Conclusions

Here, we studied machine learning methods that allow us to make use of unlabeled data, besides the conventional training step that requires labeled data. We achieved improvements in classification accuracy by using large amounts of unannotated data for pretraining the network with a discriminative approach. After pretraining, training was finalized using a smaller set of labeled data. After the comparison of the results produced by the generative and discriminative methods, we experienced better performance in case of the latter classRBM. Altogether we found that the accuracy of DNNs can be improved by using semi-supervised training algorithms.

Method	Hid. Structure	Epochs	Learning Rate	Dev	Test
Baseline ANN 1	3x100	45	0.5	68%	65%
Baseline ANN 2	3x500	60	0.1	70%	67%
Gen. classRBM 1	3x100	45	0.01	68%	71%
Gen. classRBM 2	500	60	0.05	73%	69%
Disc. classRBM 1	3x100	45	0.01	77%	73%
Disc. classRBM 2	500	60	0.05	79%	83%
Semi-sup. Gen. classRBM 1	3x100	45	0.05	67%	68%
Semi-sup. Gen. classRBM 2	500	60	0.05	69%	64%
Semi-sup. Disc. classRBM 1	3x100	45	0.05	79%	78%
Semi-sup. Disc. classRBM 2	500	60	0.01	88%	85%

Table 1: Results

References

[1] Srikanth Cherla, Son N. Tran, Tillman Weyde, and Artur S. d'Avila Garcez. Generalising the Discriminative Restricted Boltzmann Machine. *CoRR*, abs/1604.01806, 2016.

[2] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[3] Hugo Larochelle and Yoshua Bengio. Classification using Discriminative Restricted Boltzmann Machines. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 536–543, New York, NY, USA, 2008. ACM.

[4] Soren Kaae Sonderby. RBM toolbox is a MATLAB toolbox for online training of RBM and stacked RBM's. 2014.