Improving semi-supervised neural networks for scene understanding by learning the neighborhood graph

Jiwoong Im
University of Guelph
imj@uoguelph.ca

Graham W. Taylor
University of Guelph
gwtaylor@uoguelph.ca

1. Introduction

Scene understanding from remotely sensed images is a challenging task in computer vision. One key concern faced is the diversity in optical sensors such as multispectral and hyperspectral imagery (HSI). These open up new areas of application but pose new methodological challenges in data analysis [3]. Another key challenge is the relative scarcity of labeled training data for training and evaluation. To this end, semi-supervised learning (SSL) techniques are an attractive framework that can exploit both labeled and unlabeled information [8]. A number of SSL techniques have been applied to multispectral and hyperspectral scene understanding, including Transductive SVMs (TSVM) [1], Laplacian SVMs (LapSVM) [4], and Graph-based methods [2]. However, none of these methods are scalable to millions of unlabeled pixels and are therefore not suitable for large-scale remote sensing applications.

Ratle et al. proposed semi-supervised neural networks (SSNN) to classify ground cover classes from HSI [6]. Trained by stochastic gradient descent or other gradient-based methods, these networks are computationally efficient and were shown to outperform the aforementioned SSL methods on a number of benchmarks. These techniques require an explicitly constructed neighborhood graph based on Euclidean distance in input space. They are based on the smoothness assumption that examples that are similar in input space should also be similar in label space. A limitation with this approach is that Euclidean distance may not be an ideal measure of *semantic* similarity.

At first glance, it seems sensible to apply similarity learning to *learn* the neighborhood graph required by SSL approaches. However, these methods, which require labeled data, tend to overfit when labeled data is limited [7] as is often the case in remote sensing.

We describe a simple and novel alternative for learning the neighborhood graph for SSL based on binary classifiers which predict whether two inputs are from the same or different class. We experiment with two neural network-based architectures: a classical and multiplicative interaction network. We show that our approach markedly improves the performance of the SSNN when applied to scene understanding from HSI.

2. Background

The point of departure for our work is the SSNN which attempts to minimize the following cost

$$\mathcal{L} = \frac{1}{l}\sum_{i=1}^{l} V(\mathbf{x}_i, y_i, f) + \lambda_u \frac{1}{l+u}\sum_{i,j=1}^{l+u} L(f_i, f_j, w_{ij})$$

which is composed of two terms, the loss function of errors (labeled information) summed over l labeled data points $\{\mathbf{x}_i, y_i\}$, and the regularizer (unlabeled information) summed over the l labeled points and u unlabeled points. The coefficient λ_u is an empirically determined constant which controls the strength of the regularizer.

The function f is a neural network classifier which makes label predictions $\hat{y} = f(\mathbf{x})$ using multiple layers of nonlinear transformations of the data. We use the shorthand $f_i = f(\mathbf{x}_i)$. W_{ij} are edge weights which define pairwise similarity between unlabeled examples.

For V we use the hinge loss function, and for L we consider two options proposed by [6]. The first is the Laplacian SSNN (LapSSNN):

$$L(f_i, f_j, W_{ij}) = \begin{cases} ||f_i - f_j||^2, & \text{if } W_{ij} = 1 \\ \max(0, m - ||f_i - f_j||)^2, & \text{if } W_{ij} = 0 \end{cases}$$

where similar points in the input space ($W_{ij} = 1$) are forced to have similar outputs. Dissimilar points ($W_{ij} = 0$) that lie within a user-specified margin m are repelled. An alternative formulation of the unlabeled loss is to ensure that neighbours should have the same class assignment. For the case of binary classification, this can be achieved through the following loss

$$L(f_i, f_j, W_{ij}) = \begin{cases} \eta^{(+)} V(\mathbf{x}_i, f(\mathbf{x}_i), c), & \text{with } c = \text{sign}(f_i + f_j) \\ & \text{if } W_{ij} = 1 \\ -\eta^{(-)} V(\mathbf{x}_i, f(\mathbf{x}_i), c), & \text{with } c = \text{sign}(f_j) \\ & \text{if } W_{ij} = 0 \end{cases}$$

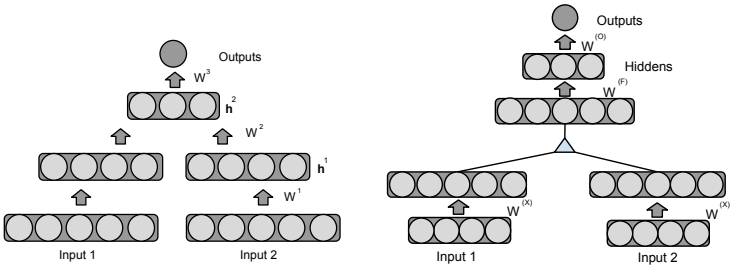
where $\eta^{(+)}$ and $\eta^{(-)}$ are the separate learning rates for the neighbouring pairs and non-neighbouring pairs. The extension to multi-class is straightforward [6].

A key consideration is that in this framework W is restricted to be a hard neighborhood graph defined according to a k nearest neighbor criterion in input space. In this work, we consider replacing the KNN criterion with a *learned* neighborhood graph.

3. Method

We propose to construct the neighborhood graph by using a binary classifier. The idea is simple: using only the labeled examples, we first train a classifier on pairs of pixels and attempt to predict whether they are from the same class or not. We then use the confidence output by the classifier to define the weight W_{ij} between points i and j.

We introduce two neural network architectures for constructing the dual-input classifier (Figure 1). As the networks nonlinearly transform the inputs through multiple layers of learned representation, inputs do not need to be Euclidean-similar to yield a high affinity.



(a) Classical net for neighborhood graph learning (NGL)

(b) Multiplicative net for neighborhood graph learning (MNGL)

Figure 1: Two dual-input neural net architectures for predicting whether two pixels have the same class label.

The first architecture consists of two representational pathways joined by a common hidden layer. The common layer is connected to a single sigmoid output. Note that the first two sets of weights are tied between pathways as shown in Figure 1a. The second architecture is based on a third-order model which permits multiplicative interactions between the two representational pathways. The third-order model learns relations between the transformed pixel intensities [5] whereas in the classical model the two pathways contribute additively to the penultimate layer.

For each pair of input pixels i and j, our classifier outputs a real-valued probability of same class, which we adopt as W_{ij}. The LapSSNN and TSSNN regularizers are updated accordingly.

# Labeled Examples	**5**	**10**	**15**	**30**
TSSNN	29.9 ± 16.13	25.9 ± 7.21	23.8 ± 3.14	21.2 ± 1.92
LapSSNN	**27.3 ± 6.46**	24.0 ± 1.08	23.4 ± 1.58	21.6 ± 1.79
TSSNN+NGL	30.7 ± 11.33	26.7 ± 2.73	25.0 ± 2.38	21.5 ± 2.86
LapSSNN+NGL	29.4 ± 12.62	26.8 ± 3.65	25.9 ± 7.41	20.1 ± 4.52
TSSNN+MNGL	31.0 ± 10.11	27.4 ± 2.59	25.2 ± 2.45	21.0 ± 2.06
LapSSNN+MNGL	30.1 ± 18.26	**23.7 ± 1.78**	**22.6 ± 4.49**	**18.9 ±1.62**

Table 1: IP6 test set - classification error rates (%).

# Labeled Examples	**5**	**10**	**15**	**30**
TSSNN	26.8 ± 2.79	24.7 ± 2.09	23.7 ± 5.24	21.6 ± 1.46
LapSSNN	27.3 ± 2.82	25.9 ± 1.35	25.1 ± 4.07	24.2 ± 1.28
TSSNN+NGL	27.5 ± 3.84	24.7 ± 1.23	23.0 ± 3.08	20.8 ± 2.38
LapSSNN+NGL	**23.2 ± 1.45**	**22.2 ± 1.84**	**20.5 ± 2.48**	**19.7 ± 0.87**
TSSNN+MNGL	23.6 ± 4.65	22.2 ± 3.15	21.3 ± 6.62	21.3 ± 1.02
LapSSNN+MNGL	26.4 ± 5.01	25.1 ± 1.38	23.7 ± 3.33	21.0 ± 4.15

Table 2: PaviaU test set - classification error rates (%).

4. Experiments and Discussion

Experiments conducted on two hyperspectral datasets: the Indian Pines 6-class dataset (IP6 – 220 bands) and Pavia University dataset (PaviaU – 115 bands) reveal that our proposed methods (NGL, MNGL) can improve the state-of-the-art SSNN (see Tables 1 and 2).

Future work intends to look at learning a single rather than two-stage architecture for jointly learning the neighborhood graph and HSI classification. We also intend to investigate unsupervised representation learning for building the neighborhood graph.

References

[1] Lorenzo Bruzzone, Mingmin Chi, and Mattia Marconcini. A novel transductive svm for semisupervised classification of remote-sensing images. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(11):3363–3373, 2006. 1

[2] Gustavo Camps-Valls, Tatyana V. Bandos, and Dengyong Zhou. Semisupervised graph-based hyperspectral image classification. *Geoscience and Remote Sensing Letters, IEEE*, 45(10):3044–3054, 2007. 1

[3] Gustavo Camps-Valls, Devis Tuia, Lorenzo Bruzzone, and Jon Atli Benediktsson. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(10):45–54, 2014. 1

[4] Luis Gómez-Chova, Gustavo Camps-Valls, Jordi Munoz-Mari, and Javier Calpe. Semisupervised image classification with laplacian support vector machines. *Geoscience and Remote Sensing Letters, IEEE*, 5(3):336–340, 2008. 1

[5] Roland Memisevic. Gradient-based learning of higher-order image features. In *International Conference on Computer Vision*, 2011. 2

[6] Frédéric Ratle, Gustavo Camps-Valls, and Jason Weston. Semisupervised neural networks for efficient hyperspectral image classification. *Geoscience and Remote Sensing Letters, IEEE*, 48(5), 2010. 1, 2

[7] Liu Yang, Rong Jin, Rahul Sukthankar, and Yi Liu. An efficient algorithm for local distance metric learning. In *Proceedings of AAAI*, 2006. 1

[8] Xiaojin Zhu. Semi-supervised learning literature survey. *Computer Science, University of Wisconsin-Madison*, 2:3, 2006. 1