SEMI-SUPERVISED LEARNING FOR CLASSIFICATION OF POLARIMETRIC SAR-DATA

R.Hänsch, O.Hellwich

Berlin University of Technology (TUB), Computer Vision and Remote Sensing Group
Franklinstrasse 28/29, FR3-1, D-10587 Berlin, Germany.
Tel.: +49-30314-73107, Fax: +49-30314-21104, E-Mail: rhaensch@fpk.tu-berlin.de

1. INTRODUCTION

In the last decades Synthetic Aperture Radar (SAR) technology gained more and more importance in remote sensing. Although often more difficult to interpret than optical data, SAR data has certain advantages, like independence from daylight or less influence of weather conditions. Furthermore the data contain information, which cannot be provided by other remote sensing technologies. Today, many modern SAR sensors like TerraSAR-X are available and provide a huge amount of data to the scientific and commercial communities. This data need to be analysed and interpreted. However, manual inspection is impracticable due to the detail, size and number of contemporary SAR images. That is why there is a great need of automatic algorithms, which are able to interpret SAR images with high accuracy. These methods should be robust regarding slight changes of acquisition circumstances, flexible and adaptable to problem specific tasks. Modern algorithms of machine learning have shown these properties in many areas and are therefore very promising.

2. SEMI-SUPERVISED LEARNING

The most algorithms of machine learning can be devided in two groups: supervised and unsupervised methods.
In the case of supervised learning exists a specific problem definition provided by a training set $\mathbf{X} = \{(\mathbf{x}, \mathbf{y})^\alpha\}_{\alpha=\mathbf{1},\ldots,\mathbf{p}}$ of p samples x^α and associated class labels y^α. The goal is to derive system parameters which produce y^α as system output given x^α as system input. Each data point x has to be assigned with the label y by a human expert, which is very expensive. Therefore, in general one cannot assume, that there is an infinite or even large amount of labeled data. But, a sufficiant amount of data is very important in supervised learning and the number of data points needed to find a robust solution is proportional to the number of dimensions of the data space.
Unsupervised learning does not need any kind of labeling, but uses only the given samples or extracted features to discover structures within the data. It is possible to influence how these structures are learned by changing the parameters of the method, but in contrast to supervised learning, it is not possible to define exactly, what should be learned. The advantage of unsupervised learning is, that all data available are used and structures within the data are revealed, which may help to solve the given problem. However, there is no guarantee, that the discovered structures solve the problem, the user is really interested in.
Especially in interpretation of SAR images there is often a specific problem defintion like classification of different landuses. Unsupervised learning cannot be used to solve this problem directly, because there are maybe more dominant structures as those caused by different landuses or the discovered classes does not fit to the classes the user is interested in. Furthermore there is no use of given labels and class definitions, which are important a priori knowledge. On the other hand, there are often only a small number of labeled data, but a huge amount of unlabeled data, which are ignored by supervised learning methods. Semi-Supervised Learning makes use of both types of data. One possibility is to combine supervised and unsupervised methods in a way, that both benefit from the advantages of the other.
In this paper the data is first clustered by an unsupervised technique named Deterministic Annealing (DA), which was proposed in [1] and whose ability to successfully analyse SAR data was shown in [2]. This method was slightly modified to make use of the available information provided by the given class labels. The output of DA is an association probability for each cluster and data point. This probability distribution is used afterwards as input of a Multi-Layer Perceptron (MLP).

3. DETERMINISTIC ANNEALING

Deterministic Annealing is an unsupervised clustering technique. It assumes that the data lie in a vector space with a certain structure, which is representable through a specific number of representatives or more precisely cluster centers. The membership of a data point to a cluster is represented by an association probability, which is governed by the similarity of the data point and cluster center. DA tries to partition the vector space into a number of clusters by minimizing an energy function E, which not only considers the spectral distance between a data point and a cluster center, but also the entropy of the association probability according to Jaynes's Maximum Entropy Principle (see [3] for more details). This ensures, that only distributions, which represent as much certainty as provided by the data, are used and have an influence on the clustering. The energy function E was extended to include the a priori knowledge contained in the class labels assigned to some part of the data. Furthermore, DA starts with only one single cluster to which all the data points belong at the beginning. During the annealing process more and more clusters are generated if and only if needed to better represent the discovered data structure.

4. MULTI-LAYER PERCEPTRON

Multi-Layer Perceptrons are very powerful learning frameworks. Their properties have been studied very well and their ability to solve difficult learning tasks has been proven in the past in a lot of different fields of machine learning. A MLP consists of a number of simple units called neurons which are organized in a number of layers. The first layer is called input layer, the last one output layer. All layers between these two are named hidden layers.
During the DA part of the proposed learning scheme a probability distribution over all clusters is learned for each data point. This distribution is now used as input of the MLP.

5. COMBINATION

As mentioned above there is only one cluster at the beginning of the clustering process. That is why the input layer consists only of one single neuron. As soon as a new cluster is created, a new neuron is generated, too, and the weights are adopted to the new structure of the input. As long as different user-defined classes are represented by similar association probabilities it is impossible for the MLP to seperate them. However, as soon as there are differences in the association probabilities the MLP will adapt the weights in the right way and learn the correct mapping from the cluster association to the user defined classes. The learning process will stop when the estimated generalisation error is smaller than an user-defined threshold or begins to increase due to overfitting.
The proposed learning scheme provides not only the classification of the data points, but also the optimal number of clusters during the unsupervised learning part. The obtained clustering is optimal in the sense of best classification result and gives therefore a very robust representation of the data regarding the defined classes.

6. EVALUATION

Although the proposed method can be applied to any kind of data, provided that a suitable distance measure exists, it will be evaluated on polarimetric SAR data and its performance will be compared with other classification schemes.

7. REFERENCES

[1] K. Rose: *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems* Proc. IEEE, vol. 86, pp. 2210-2239, 1998.

[2] R. Hänsch, M. Jäger, O. Hellwich: *Clustering by deterministic annealing and Wishart based distance measures for fully-polarimetric SAR-data* Proceedings of EUSAR 2008, Volume 3, pp. 419-422, 2008.

[3] E.T. Jaynes: *Information theory and statistical mechanics* The Physical Review. 106, Nr. 4, pp. 620-630, 1957.