

MAXIMUM MARGIN CLUSTERING OF HYPERSPECTRAL DATA

S.Niazmardi^a, A.Safari^a, S.Homayouni^{a,b}

^a Dept. of Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran

^b Dept. of Geography, University of Ottawa, Ottawa, Canada

KEY WORDS: Clustering, Maximum Margins, Hyperspectral Data, Support Vectors

ABSTRACT:

In recent decades, large margin methods such as Support Vector Machines (SVMs) are supposed to be the state-of-the-art of supervised learning methods for classification of hyperspectral data. However, the results of these algorithms mainly depend on the quality and quantity of available training data. To tackle down the problems associated with the training data, the researcher put effort into extending the capability of large margin algorithms for unsupervised learning. One of the recent proposed algorithms is Maximum Margin Clustering (MMC). The MMC is an unsupervised SVMs algorithm that simultaneously estimates both the labels and the hyperplane parameters. Nevertheless, the optimization of the MMC algorithm is a non-convex problem. Most of the existing MMC methods rely on the reformulating and the relaxing of the non-convex optimization problem as semi-definite programs (SDP), which are computationally very expensive and only can handle small data sets. Moreover, most of these algorithms are two-class classification, which cannot be used for classification of remotely sensed data. In this paper, a new MMC algorithm is used that solve the original non-convex problem using Alternative Optimization method. This algorithm is also extended for multi-class classification and its performance is evaluated. The results of the proposed algorithm show that the algorithm has acceptable results for hyperspectral data clustering.

1. INTRODUCTION

Hyperspectral imagery consists of the acquisition of the radiance of earth objects in the portion of the electromagnetic spectrum spanning from the visible to the long wavelength infra-red region in numerous narrow and contiguous spectral bands (Goetz et al., 1985). In comparison to multispectral data, hyperspectral data allow more accurate detection, classification, characterization and identification of land-cover classes (Camps-Valls et al., 2007). However, numerous spectral bands lead to new challenges for supervised and unsupervised classification algorithms (Jia and Richards, 2007).

One of the most successful supervised classification algorithms that show very good performance for hyperspectral data are large margin algorithms such as Support Vector Machines (SVMs) and Support Vector Domain Description (SVDD) (Khazai et al.; Melgani and Bruzzone, 2004). Nevertheless, the performances of these algorithms are highly affected by the quality and quantity of the training data. (Jia and Richards, 2007). In real-world applications, collecting of enough high quality training samples is an expensive and time-consuming task (Mingmin and Bruzzone, 2005).

On the other hand, the unsupervised classification algorithms can solve the problem of training data. However, there are many more issues, such as used similarity measure, heavy computational cost and unknown number of classes that limit their application for remotely sensed data (Niazmardi et al., 2012).

In recent years, there has been a growing interest in extending the large margin methods for unsupervised learning. One of the algorithms that has gained more attention is Maximum Margin Clustering (MMC), proposed by Xu *et al* (Xu et al., 2004). This algorithm performs clustering by simultaneously finding the separating hyperplane and labels. This algorithm and its variant have been very successful in many clustering problems (Zhang et al., 2009).

However, since the labels of pixels are unknown, the optimization is done over all possible class labels for each sample. This make the convex optimization problem of supervised large margin methods, a non-convex, hard, combinational problem and definitely more computationally expensive (Zhang et al., 2009). Moreover, since this algorithm optimizes the SVM's cost function, therefore it is a two-class algorithm. Regarding the optimization problem of this algorithm, Many researchers have tried to solve that by reformulating and relaxing it as a solvable convex problem (Valizadegan and Jin, 2007). Meanwhile, other researchers, have tried to extend the efficiency of this algorithm by proposing multi-class MMC algorithm (Zhao et al., 2008).

Most of proposed variants of MMC algorithms have very high computational complexity and are usually evaluated in small data set (less than 1000 samples). It is mainly due to use of numerical methods for relaxing and solving the obtained convex problem. This issue makes use of these methods impractical for remote sensing data. Zhang *et al.* tried to solve this issue by solving the original non-convex problem using the well-known alternative optimization method (Zhang et al., 2009). Although, their proposed method is a computationally affordable method, it is considered as a binary classification. The purpose of this paper is to extend the Zhang's algorithm to a multi-class clustering algorithm and evaluate its performance for clustering of hyperspectral data.

The rest of this paper is outlined as follows. In the second section, the proposed algorithm is presented. In the third section, the implementation of them is presented. The results and the discussion are the subject of section four. Lastly, in the last section, we draw the conclusion and the subjects that should be studied further.

2. THE PROPOSED METHOD

2.1 Binary Maximum Margin Clustering

The binary MMC algorithm can be considered as unsupervised SVMs, which the class labels are missing for all samples and should be estimated. Thus, the purpose of MMC algorithm is to find the optimum large margin hyperplane from the set of all possible hyperplanes, separating the samples in two classes with every possible way. Since the class labels are unknown in here, the trivial solution can be obtained by assigning all samples to a single cluster. To prevent this trivial solution, another constraint should be added to constraint of SVMs algorithm, which is the class balance constraint. With this constraint, we put constraint on the minimum number of samples for each class.

Assume that $x_i \in \mathcal{X}$ are input samples and $y_i \in \{+1, -1\}$ are output cluster labels, the cost function of MMC can be written as follows:

$$\min_y \min_{w,b,\xi} \|w\|^2 + 2C\xi^T e \quad (1)$$

Which should be optimized under following constraints:

$$\begin{aligned} y_i(W^T \varphi(x_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0, y_i \in \{+1, -1\} \\ -l &< e^T y < l \end{aligned} \quad (2)$$

In these equation W and b are the normal and bias term of the hyperplane, ξ_i are slack variables, e is vector of all ones. The last constrain is the equation 2 is the class balance constraint. In this constraint $l \geq 0$ is a user defined constant controlling class balance.

It can be seen that the constraint $y_i \in \{+1, -1\}$ make this optimization problem a nonconvex one. As mentioned, different methods are proposed to solve this nonconvex problem, which most of them tries to convert the original nonconvex problem to a solvable convex problem. This procedure is usually done by some assumption and simplifications.

2.2 Alternative Optimization for MMC

Alternative Optimization (AO) is a well-known optimization algorithm that being used in many clustering algorithms, e.g. K-Means. In this optimization, one group of parameters is optimized by holding the other group(s) fixed, and vice versa. This iterative updating scheme is then repeated (De Oliveria and Pedeycz, 2007).

The original non-convex MMC cost function has two sets of parameters, the first set contains the hyperplane parameters and the second set contains the labels of samples. Therefore, by starting from an initial set of labels for each sample, one can estimate the hyperplane parameters (i.e. W , b) and estimate labels of samples using them. This procedure then iterates for a predefined number of iteration or until a termination criterion is met. As it is obvious, in this method the parameters of hyperplane can be estimated with a regular optimization method of SVM. Thus, the problem of non-convexity of the algorithm is solved. To estimate label of samples the following optimization should be solved.

$$\min_{y,\xi} \xi^T e \quad (3)$$

Under the following constraints:

$$\begin{aligned} y_i(W^T \varphi(x_i) + b) &\geq 1 - \xi_i \\ \forall i, \xi_i &\geq 0 \\ \forall i, y_i &\in \{+1, -1\} \\ -l &\leq e^T y \leq l \end{aligned} \quad (4)$$

It can be proved that for a fixed b the optimal strategy to determine labels (i.e. solve the optimization problem of equation 3) is to assign all y_i as -1 for those samples that

$W^T \varphi(x_i) + b \leq 0$, and assign all y_i as +1 for those with $W^T \varphi(x_i) + b \geq 0$ (Zhang et al., 2009).

One of the biggest problems of using AO algorithm is its premature convergence and its high probability of getting stuck in poor local optimum (Zhang et al., 2009). To reduce this probability and applying the class balance constraints another strategy should be used to estimate the hyperplane bias term.

In this strategy first, all $W^T \varphi(x_i)$ are sorted and then the midpoints between any two consecutive values are assumed as candidate values for b . From this set, the first $(n-l)/2$ and the last $(n-l)/2$ can be discarded since they violate the class balance constraints. Among the remaining values, the value that the labels are estimated as $\text{sign}(W^T \varphi(x_i) + b)$ and then the following objective function is computed.

$$f = \sum_{i=1}^n |W^T \varphi(x_i) + b - y_i| \quad (5)$$

The bias value for which this objective function is minimized is the optimum bias value. The whole algorithm can be summarized as follows.

- Step 1:** initial labels by another clustering algorithm such as K-Means or fuzzy C-means.
- Step 2:** perform standard SVM training and compute W from Karush-Kuhn-Tucker (KKT).
- Step 3:** compute bias term as described.
- Step 4:** Assign class labels as $y_i = \text{sign}(W^T \varphi(x_i) + b)$
- Step 5:** Repeat 2-5 until convergence.

2.3 Multi-class MMC

Different methods are proposed to extend the binary MMC to multi-class MMC. In here, we used a one against one strategy. In this strategy, the clustering algorithm that is used for estimating the initial labels is set to more than two clusters. After this, the binary MMC is used to classify each possible pair of clusters and the final labels are obtained by majority voting.

3. IMPLEMENTATION

3.1 Dataset

Indian Pine data set is used to evaluate the performance of the proposed algorithm. In order to obtain a better accuracy assessment and less computational cost, in this study only the labeled samples are used. To reduce the effects of spectral bands with higher values on those having lower values (due to the spectral/radiometric variations), the data are linearly normalized and mapped into the [0,1] interval.

Indian Pine Dataset was taken over northwest Indiana's Indian Pine test site in June 1992 (1992) by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS). The image data consists of 145×145 pixels at 20m spatial resolution with 220 bands. 20 water absorption bands and 15 noisy bands were removed, resulting in 185 spectral bands (Mojaradi et al., 2008). Five classes from this dataset were selected for the evaluation of the performance of the proposed algorithm (Jia and Richards, 2007; Samadzadegan and Naeini, 2011). The selected classes are listed in Table 1.

Class		No. of samples
No	Name	
1	Corn no till	1434
2	Hay Windrowed	489
3	Grass-tree	747
4	Wood	1294
5	Soybean no till	968

Table 1. Names and the number of pixels in each class

3.2 Parameter Setting

The MMC algorithm is a parametric algorithm; the open parameters of this algorithm include parameter of used kernel function, the trade off parameter for the SVM algorithm and the value of l for class balance constraint. In this study the Gaussian kernel is used as kernel function, which has one open parameters. The value of 1.5, 0.5, 300 are used for parameter of kernel, trade off parameter and the value of l respectively.

3.3 Evaluation criteria

In this study, the kappa coefficient of agreement is used as the validity index. Kappa is a supervised validity index that compares the results of classification algorithm to some known labelled samples or test data. However, since the clustering algorithms assign random numbers as clusters' labels, we need to assign each label to its corresponding class within the ground truth data. This can be done either by user or by using some mapping function. In this study, we use a mapping function between class numbers in ground truth and the classified map of the clustering algorithm. After this step, the kappa coefficient can be calculated.

4. RESULTS AND DISCUSSION

The proposed algorithm is performed with the mentioned parameters and its results are compared with FCM and K-Means clustering algorithms. The kappa coefficient of each algorithm is presented in table 2.

Algorithm	Kappa coefficient
MMC	0.72
FCM	0.64
K-Means	0.65

Table 2. Kappa coefficients of various algorithm

Based on these results, it is obvious that the proposed MMC algorithm have shown marginally better performance in comparison to the classic clustering approaches. There are different factors affecting the performance of the MMC algorithms. Two of the most important ones are the parameters and initial labels of the algorithm. Parameter selection for unsupervised algorithms has always been a challenging problem

due to lack of labelled samples. In this paper, a try and error method was used for parameter selection. However, the grid search method, which is usually used for parameter selection of the SVM algorithm, can be used for parameter selection of the MMC algorithm as well. Nevertheless, this will cause to high computational cost of the algorithm. Moreover, using the initial labels for parameter selection does not make sense because of not correct labelling.

In this study, the FCM algorithm is used to assign the initial labels. However, any other clustering algorithm can be used instead of the FCM algorithm. It should be noted that, the random labels could not be used for initialization of this algorithm. Because, the proposed MMC algorithm actually correct the initial labels of samples and cannot discriminate between different classes in case of using random labels.

The proposed MMC algorithm, despite its marginal improvement of the accuracy has many potential and its accuracy will be increased in case of addressing the mentioned issues. Moreover, this algorithm can be reformulated as a semi-supervised algorithm with only small changes in its training phase. In this case, the results are less depending to the initial labelling, and consequently can lead to the higher accuracy.

For a better comparison the ground truth of the dataset and the classed map of the algorithms are shown in fig.1.

5. CONCLUSION

In this paper, a new multi-class maximum margin clustering algorithm has been developed and its performance has been evaluated for hyperspectral data clustering. The proposed MMC algorithm is an unsupervised SVM algorithm that simultaneously estimates the hyperplane parameters and the label of samples. The alternative optimization method is used to solve this non-convex optimization of the MMC algorithms, as well as, a one-against-one strategy for multi-class clustering. The obtained results of this study lead us to the following conclusions:

- The proposed algorithm shows better accuracy in comparison to other clustering algorithms. However, using better initial labels and adding a parameter selection step, will certainly leads to higher accuracies.
- As mentioned, the performance of the MMC algorithm is affected by the used parameters. On the other hand, the usual parameter selection methods, such as grid search method, cause dramatic increase of computational cost of the algorithm. Therefore, a suitable unsupervised parameter selection method should be proposed and used for the MMC algorithm.
- The alternative optimization method used in proposed algorithm is highly sensitive to the local optimums. The meta-heuristic algorithms can address this problem more properly. Nevertheless, the high computational cost of these algorithms makes their application for real cases difficult.

Based on these results and the subsequent discussion thereof, the proposed approach seems a promising clustering method that can outperform other algorithms. However, some issues such as parameter selection of this algorithm need more investigation.

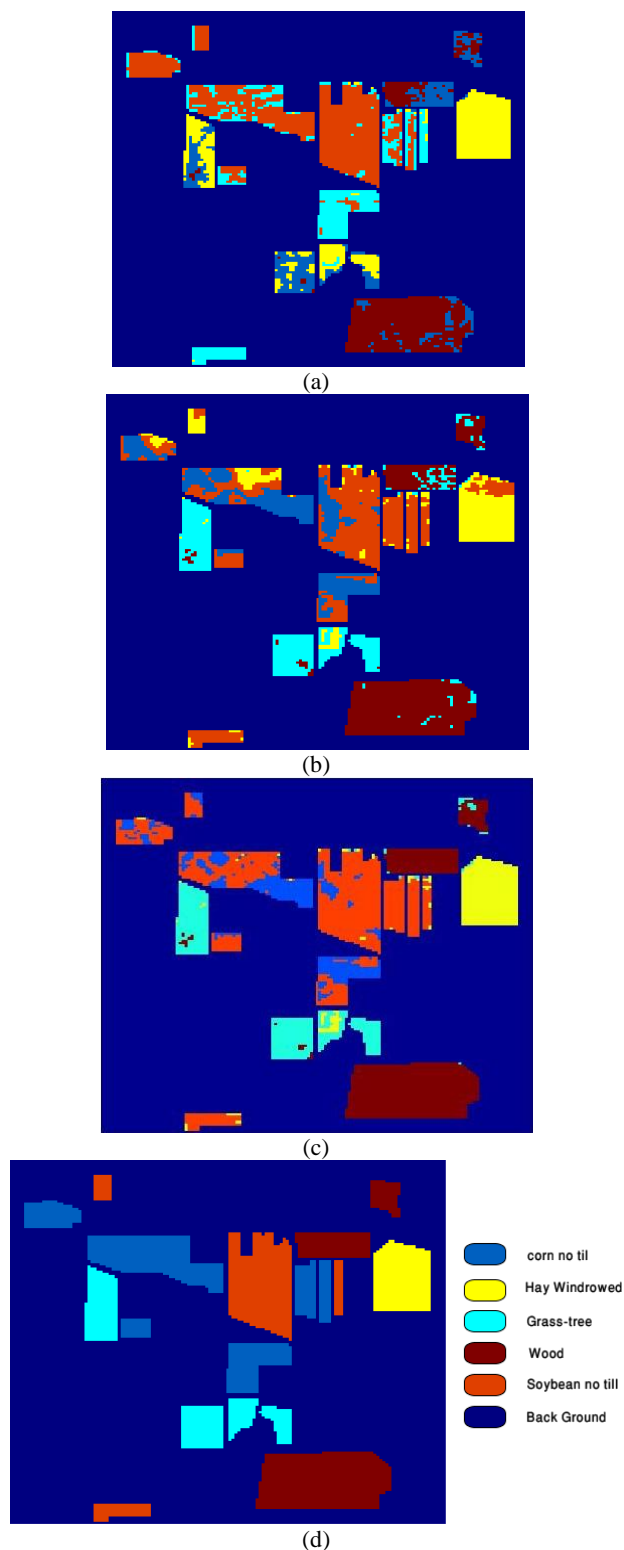


Fig 1. Ground truth and the classified maps obtained from different algorithms. (a) FCM, (b) K-Means, (c) MMC, (d) Ground truth

Reference

1992. AVIRIS image Indian Pine Test Site. <http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/hyperspectral.html>.
- Camps-Valls, G., Bandos Marsheva, T., Zhou, D., 2007. Semi-Supervised Graph-Based Hyperspectral Image Classification.

Geoscience and Remote Sensing, IEEE Transactions on 45, 3044-3054.

De Oliveria, j.V., Pedeycz, w., 2007. Advances in Fuzzy Clustering and its Applications. John Wiley & Sons.

Goetz, A.F.H., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228, 1147-1153.

Jia, X., Richards, J., A., 2007. Hyperspectral data representation, in: Chang, C., I (Ed.), Hyperspectral Data Exploitation: Theory and Applications. John Wiley and sons, inc, Hoboken, New Jersey, pp. 205-225.

Khazai, S., Safari, A., Mojaradi, B., Homayouni, S., Improving the SVDD Approach to Hyperspectral Image Classification. Geoscience and Remote Sensing Letters, IEEE, 1-5.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. Geoscience and Remote Sensing, IEEE Transactions on 42, 1778-1790.

Mingmin, C., Bruzzone, L., 2005. A semilabeled-sample-driven bagging technique for ill-posed classification problems. Geoscience and Remote Sensing Letters, IEEE 2, 69-73.

Mojaradi, B., Emami, H., Varshosaz, M., Jamali, S., 2008. A novel band selection method for hyperspectral data analysis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37.

Niazmardi, S., Alizadeh Naeini, A., Homayouni, S., Safari, A., Samadzadegan, F., 2012. Particle Swarm Optimization of Kernel based Fuzzy C-Means for Hyperspectral Data Clustering. Journal of applied remote sensong 6.

Samadzadegan, F., Naeini, A.A., 2011. Fuzzy clustering of hyperspectral data based on particle swarm optimization, Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on, pp. 1-4.

Valizadegan, H., Jin, R., 2007. Generalized maximum margin clustering and unsupervised kernel learning. Advances in Neural Information Processing Systems 19, 1417.

Xu, L., Neufeld, J., Larson, B., Schuurmans, D., 2004. Maximum margin clustering. Advances in Neural Information Processing Systems 17, 1537-1544.

Zhang, K., Tsang, I.W., Kwok, J.T., 2009. Maximum margin clustering made practical. Neural Networks, IEEE Transactions on 20, 583-596.

Zhao, B., Wang, F., Zhang, C., 2008. Efficient multiclass maximum margin clustering, Proceedings of the 25th international conference on Machine learning. ACM, pp. 1248-1255.