EVALUATION OF MULTIPLE KERNEL LEARNING ALGORITHMS FOR CROP MAPPING USING SATELLITE IMAGE TIME-SERIES DATA

S. Niazmardi [a]*, A. Safari [a], S. Homayouni [b]

[a] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran - (s.niazmardi, asafari)@ut.ac.ir
[b] Dept. of Geography, Environment and Geomatics, U. of Ottawa - saeid.homayouni@uOttawa.ca

KEY WORDS: Satellite Image Time-series, classification, time-series data, Multiple Kernel Learning, Kernel learning, crop mapping.

ABSTRACT:

Crop mapping through classification of Satellite Image Time-Series (SITS) data can provide very valuable information for several agricultural applications, such as crop monitoring, yield estimation, and crop inventory. However, the SITS data classification is not straightforward. Because different images of a SITS data have different levels of information regarding the classification problems. Moreover, the SITS data is a four-dimensional data that cannot be classified using the conventional classification algorithms. To address these issues in this paper, we presented a classification strategy based on Multiple Kernel Learning (MKL) algorithms for SITS data classification. In this strategy, initially different kernels are constructed from different images of the SITS data and then they are combined into a composite kernel using the MKL algorithms. The composite kernel, once constructed, can be used for the classification of the data using the kernel-based classification algorithms. We compared the computational time and the classification performances of the proposed classification strategy using different MKL algorithms for the purpose of crop mapping. The considered MKL algorithms are: MKL-Sum, SimpleMKL, LPMKL and Group-Lasso MKL algorithms. The experimental tests of the proposed strategy on two SITS data sets, acquired by SPOT satellite sensors, showed that this strategy was able to provide better performances when compared to the standard classification algorithm. The results also showed that the optimization method of the used MKL algorithms affects both the computational time and classification accuracy of this strategy.

1. INTRODUCTION

Satellite image Time-Series (SITS) data are a collection of satellites images acquired from the same geographical area over a period of time (Jonsson and Eklundh, 2004). The SITS data, due to their ability to capture the dynamic spectral behaviour of plants and crops during their growing cycles, have been frequently used for different agricultural applications (Jamali et al., 2014). Among these applications, identification of crop types through classification is one of the most important ones (Verhegghen et al., 2014). This is because knowledge of crop types is required as the base information for several other agricultural studies, such as crop acreage estimation, yield forecasting, estimation of water requirements, and assessment of food security (Li et al., 2014; Löw et al., 2015; Simonneaux et al., 2008).

However, crop mapping through SITS data classification is a challenging task, due to particular characteristics of the data. The most discriminative characteristic of the SITS data is the dimensionality of its images' feature space. If the SITS images have a single feature, it is called as univariate SITS. The SITS is called as multivariate SITS if its' images have more than one feature (Adhikari and Agrawal, 2013). SITS data that consist of images acquired by multispectral or hyperspectral sensors are categorized as the Multivariate SITS (Adhikari and Agrawal, 2013). This type of SITS, in its original representation, is a four-dimensional data which cannot be

classified using the conventional classification algorithms (Baydogan and Runger, 2015).

The other SITS characteristic that affects its classification, is the fact that different images of SITS data have different statistical characteristics and contain different amount of information regarding the separability between different classes. These differences may happen as a result of changes in the sensor and atmospheric conditions between the acquisition times, as well as the changes of spectral characteristics of the crops (Li et al., 2014; Niazmardi et al., 2014).

To address issues caused by the four-dimensional representation of the data, in some studies the multivariate SITS is converted into a univariate SITS by extracting a Vegetation Index (VI) such as Normalized Difference Vegetation Index (NDVI) from the data. On the one hand, this strategy is appealing since the univariate SITS data can be classified using all the classification algorithms, but on the other hand, it has two main drawbacks. First, the VIs are calculated based on the information of few spectral bands of the data at each time, consequently, the univariate SITS generally contain less information content in comparison with multivariate SITS (Baydogan and Runger, 2015). Second, the user should select the best performing VI for the classification problem, which there is no method available to assist this choice (Gerstmann et al., 2016). Stacking the images of a multivariate SITS to create a single image is another common practice for multivariate SITS data classification. However, the stacked image can be very high dimensional (Keogh and

---

* Corresponding author

Pazzani, 1998). It should be noted that these two strategies cannot address the issues caused by the different statistical characteristics of the SITS images.

In this paper, a classification strategy based on Multiple Kernel Learning (MKL) framework is presented for proper classification of multivariate SITS data. This classification strategy obtains a kernel-bead representation of the data by combining the kernels constructed from each time of the SITS data. The optimal combination of the kernels is estimated using the MKL algorithms. Finally, the composite kernel, obtained from this combination, is used for SITS data classification. Using this classification strategy, the issues associated with different statistical characteristics of the SITS data can be addressed. Because different kernels are constructed from the images of each time of the SITS. In addition, using this strategy, the final representation of the multivariate SITS data is a kernel function which can be classified using all the kernel-based classification algorithms.

In this paper, we evaluated this classification strategy using several algorithms from different MKL categories. Although the MKL algorithms have been widely used for multi-modal and multi-feature classification of remote sensing images (Gomez-Chova et al., 2015; Niazmardi et al., 2017; Wang et al., 2016), they never have been used for multivariate SITS data classification.

2. METHODOLOGY

In this section, initially the MKL algorithms are briefly introduced and then the classification of SITS data using the MKL algorithms are discussed. Since most of the MKL algorithms are proposed based on the Support Vector Machines (SVM), the theory of the MKL algorithms are only proposed for binary classification problems.

2.1 Multiple Kernel Learning

MKL algorithms are a group of algorithms that aim to optimally combine a set of predefined kernels (known as basis kernel) into a composite kernel (Gönen and Alpaydın, 2011). The basis kernels can be constructed using different data modalities or from the same data by adopting different kernel functions.

Assume that n binary-labelled samples $\mathbf{x}_i \in \mathbb{R}^D, i = 1,...,n$ are available as the training data, from which M basis kernels $\{\mathbf{K}_1, \mathbf{K}_2,..., \mathbf{K}_M\}$ are constructed. The MKL algorithm combines these basis kernels with each other to create a composite kernel $\mathbf{K}_c$. The composite kernel provides a more flexible and more informative representation of the data as compared to those provided by each one of the basis kernels. The kernel combination is usually modelled as a linear weighted summation of the basis kernel as follows (Bucak et al., 2014):

$$\mathbf{K}_c = \sum_{m=1}^{M} d_m \mathbf{K}_m \qquad (1)$$

where $d_m \in \mathbf{d}$, $m = 1,...,M$ are the non-negative weights associated with the mth basis kernel, which should be estimated by the MKL algorithm.

Considering the method used by the MKL algorithms to estimate the kernel weights, they can be divided into two categories: i) fixed rule MKL algorithms and ii) optimization-based MKL algorithms. The detailed descriptions of these categories are given in the following sub-sections.

2.1.1 Fixed-Rule MKL algorithms
The algorithms of this category assign equally fixed values to the kernel weights, without any optimization. The MKL-sum algorithm, one of the most used algorithms of this category, sets the weights of each basis kernels equally to one (i.e., $d_m = 1$, $m = 1,...,M$). The fixed-rule MKL algorithms are very fast, but their performances are highly influenced by the presence of noisy and weak kernels among the basis kernels (Gönen and Alpaydın, 2011). The algorithms of this category are the most common algorithms in remote sensing literature (Camps-Valls et al., 2008; Camps-Valls et al., 2006; Tuia et al., 2010b; Zhou et al., 2015).

2.1.2 Optimization-based MKL algorithms
Optimization-based MKL algorithms estimate the optimal weights of the basis kernels through optimizing a target function (Gönen and Alpaydın, 2011). The target function is a parametric function of the kernel weights that reaches its extremum on the best set of kernel weights. There have been various target functions proposed for different algorithms of this category (Gönen and Alpaydın, 2011). However, the SVM loss function is the most used target function (Niazmardi et al., 2016). In the MKL algorithms that adopts this target function, both the SVM parameters (i.e., support vector coefficient) and kernel weights are estimate by optimizing the following optimization problem (Bucak et al., 2014):

$$\min_{f \in \mathcal{H}_c} \quad \frac{1}{2}\|f\|_{\mathcal{H}_c}^2 + C\sum_{i=1}^{n} \ell\left(y_i f(\mathbf{x}_i)\right) \qquad (2)$$

Where f is the separating hyperplane in the Reproductive Kernel Hilbert Space (RKHS) of the composite kernel ($\mathcal{H}_c$). y_i and $\ell(.)$ are binary label of ith sample ($\mathbf{x}_i$) and loss function respectively. Trade-off parameter in this equation is shown by C. it can be proved that, this problem in the case of employing the Hing loss, can be cast into the following min-max optimization problem (Bucak et al., 2014):

$$\min_{\mathbf{d}} \max_{\mathbf{\alpha}} \quad \mathbf{1}^T\mathbf{\alpha} - \frac{1}{2}(\mathbf{\alpha} \circ \mathbf{y})^T K_c (\mathbf{\alpha} \circ \mathbf{y})$$

$$\text{subject to:} \quad \begin{aligned} &\sum_{i=1}^{n} \alpha_i y_i = 0 \\ &0 \leq \alpha_i \leq C \quad \forall i = 1,...,n \\ &\mathbf{d} \in \Delta \end{aligned} \qquad (3)$$

where $\mathbf{\alpha} = (\alpha_1,...,\alpha_n)^T$ is the vector of Lagrange multipliers, $\mathbf{y} = (y_1,...,y_n)^T$ denotes the set labels of land-cover classes (where $y_i \in \{+1,-1\}$ is the binary class label of the ith sample), while symbol $\circ$ and $\mathbf{1}$ shows the element-wise product between two vectors, and a vector whose all elements are one, respectively. Δ is convex set from which the kernel weights are selected. Different MKL algorithms use different strategies for solving the optimization problem of Eq.3 (Bucak et al., 2014; Niazmardi et al., 2016). One of the most successful strategies is using the alternative optimization. This two-step optimization strategy, after assigning initial values to kernel

weights, iterates a two-step procedure until a termination criterion is met. In the first step, it solves a classic SVM with the composite kernel and estimates the classifier parameters. Then in the second step, SVM parameters are fixed to the values obtained at the first step and the SVM objective function is optimized with respect to the kernel weights (Niazmardi et al., 2016).

Since all the SVM solvers can be used in the first step of this optimization, the differences between the MKL algorithms that use this strategy, (known as wrapper MKL algorithms), are the methods and assumptions used in the second step of their optimization (Gönen and Alpaydın, 2011).

Simple Multiple Kernel Learning (SimpleMKL), as one of the best-known wrapper MKL algorithms, uses a simple gradient descendant strategy to update the kernel weights on a simplex at the second step of this optimization (Rakotomamonjy et al., 2008). This algorithm, due to its simplicity and having acceptable performance, have been used in many studies in remote sensing literature (Gu et al., 2014; Gu et al., 2012; Tuia et al., 2010a). Generalized multiple kernel learning (GMKL) is another wrapper MKL algorithm which adds a regularization term to Eq.3 for estimating the kernel weights. GMKL can adopt any kind of regularization while it is a differentiable function of the kernel weights. This algorithm also uses the gradient descendent method for optimizing Eq.3 with respect to kernel weights (Varma and Babu, 2009). In remote sensing literature, the GMKL algorithm have been only used in (Gevaert et al., 2016) for the classification of unmanned aerial vehicle data.

In (Xu et al., 2010), using the group lasso regularization, an alternative formulation for the MKL optimization problem was proposed. This MKL algorithm, known as group lasso multiple kernel learning (GLMKL) uses a closed form solution to estimate the kernel weights at each iteration. This algorithm, have been recently used in remote sensing literature for combining kernels constructed form spatial and spectral features extracted from hyperspectral data (Liu et al., 2016).

Another alternative formularization of the MKL optimization problem was proposed by adding an L_p-norm (p>1) of the kernel weights vector as the regularization to the objective function of Eq.2. Using the block coordinate descent method for optimizing the obtained target function of this MKL algorithm, referred to as LPMKL, will lead to a closed-form solution for estimation of the kernel weights (Kloft et al., 2009; Kloft et al., 2011).

2.2 SITS classification using the MKL algorithms

SITS data classification using MKL algorithm, contains two steps. In the first step, basis kernels are constructed using the data acquired at each time of the SITS. Then in the second step, these kernels are combined into a composite kernel using an MKL algorithm, and the obtained composite kernel is used for classification. Figure 1 shows a flowchart for SITS data classification using the MKL algorithms.

The proposed strategy for SITS classification has several advantages. First, it is able to properly model the different statistical properties of the data acquired at different times. This is due to use of different kernels for each time of the data, which enable the method to separately model the information content of each data. In addition, using this method, the problem caused by the original four-dimensional representation of SITS data is addressed. Since, the composite kernel which is obtained as the final representation of the data, can be easily

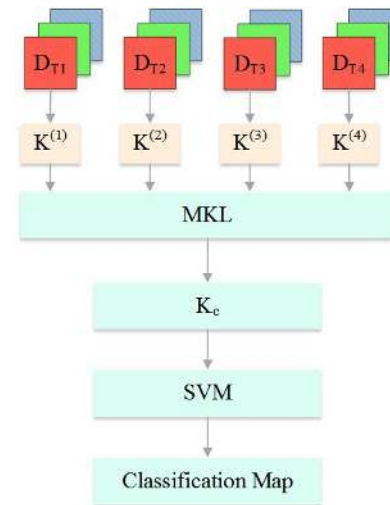used for its classification using any kernel-based classification algorithms.

Figure 1. Flowchart of the SITS data classification using MKL algorithms

3. DATA AND IMPLEMENTATION

3.1 SITS data set

The performance of different MKL algorithms in the framework of the proposed strategy, were compared according to their ability to classify different crops using two SITS data sets. The used data sets are two subsets of a large SITS data which were made up 4 multispectral images acquired by SPOT sensors during the 2012 growing season over southwest Winnipeg, Manitoba in Canada (see Table 1 for acquisition dates). All the images of both SITS data sets, which we referred to them as S1 and S2, were made up 1000×1000 samples with the spatial resolution of 20 m and were atmospherically corrected and orthorectified. Four different spectral bands of these images were used for the implementations, namely, green, red, near infrared and shortwave infrared. The false colour composite of both SITS data sets (using infrared, red and green bands) are presented in Figure 2 and Figure 3 for S1 and S2 data sets respectively.

From the available crop maps of each SITS data set, two different sets of samples were extracted as training and testing samples. The considered crops of each SITS, with the number of samples used for training and testing are presented in Table 2.

Number	Acquisition Date
1	23 June 2012
2	28 June 2012
3	5 July 2012
4	14 July 2012

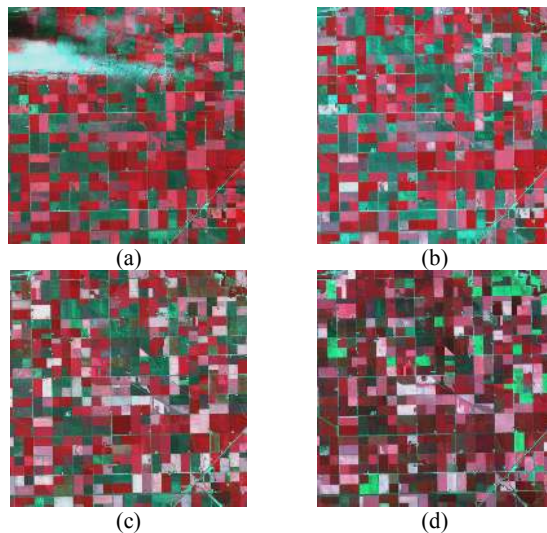Table 1. Acquisition dates of time-series data

Figure 2. False colour composite of the images of S1 time-series, a) 23 June; b) 28 June; c) 5 July; and d) 14 July.
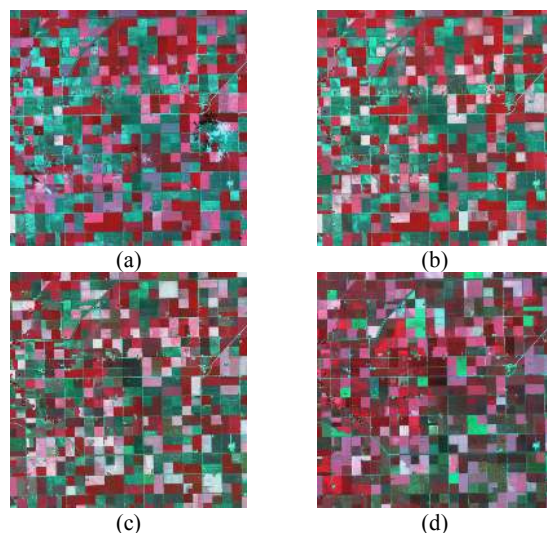

Figure 3. False colour composite of the images of S2 time-series, a) 23 June; b) 28 June; c) 5 July; and d) 14 July.

Crop name	S1		S2	
	Training	Testing	Training	Testing
Corn	108	1420	166	1132
Canola	180	2456	173	3607
Wheat	156	2054	153	3675
Soy	127	1987	162	4713
Oat	181	1978	148	1642
Sun flower	94	1453	-	-
Barley	115	1128	-	-
Forage	-	-	67	184
Total	961	12476	869	14953

Table 2. Number of samples per each crop used as training and testing samples for classification of each SITS data set.

3.2 Experimental Setup

In the experimental carried out in this paper, the basis kernels constructed from different images of the SITS data, have been combined into a composite kernel, using different MKL algorithms. The classification accuracy of an SVM algorithm trained with this composite kernel obtained from different MKL algorithms was used as the criterion for their comparison. MKL-sum (from the category of fixed-rule MKL algorithms) and SimpleMKL, GMKL (with L_1 norm), LPMKL (with L_2 norm), and GLMKL from optimization-based wrapper MKL algorithms, are considered for comparison in this paper.

In order to construct the basis kernels, the Radial Basis Function (RBF) was used as the kernel function, due to its excellent learning capabilities (Kim et al., 2005). The parameter of RBF kernel was selected from the values in the range of [0.01-10] with a step-size increment of 0.5 by using a 5-fold cross-validation technique.

The trade-off parameter of the SVM algorithm was tuned using a 5-fold cross-validation from the range of [0.01-2000] with a step-size of 50.

The trade-off parameters of the optimization-based MKL algorithms is another important parameter which should be tuned. However, there is no method available in the literature to estimate this parameter. Accordingly, in our experiments, all the MKL algorithms were run with different values in range of $\left\{10^{-4}, 10^{-3}, \ldots, 10^{3}\right\}$, and then the value associated to the highest classification accuracy was used.

The optimization-based wrapper MKL algorithms use an iterative optimization procedure. In the experiments carried out in this paper, the iteration of these algorithms terminated when the number of iterations reached to 50 or in the case that the difference between the kernel weight vectors between two successive iterations was less than 0.001.

For comparison, the performances of the MKL algorithms were compared with the results obtained from an SVM algorithm applied to the data cube obtained by stacking the images of different times of both SITS data sets. This method was called as standard SVM in this paper.

All the experiments were implemented in MATLAB on a standard Laptop PC with Intel Corei7 CPU 2.4 GHz, 12 GB RAM. The LibSVM library was used to implement the SVM algorithm (Chang and Lin, 2011).

4. RESULTS AND DISCUSSION

Obtained results from the classification of both SITS data sets, using composite kernels obtained from different MKL algorithms are presented in Table 3 and Table 4 for S1 and S2 data sets respectively. In these tables, the class accuracies and average class accuracies (ACA) are reported using the conditional kappa. In addition Overall Accuracy (OA) of the classification, kappa coefficient and the computational time of the MKL algorithms (in seconds) are also reported.

As it can be seen from these results, using the MKL algorithms provided much higher accuracies for SITS classification as compared to the standard method. As an example, for S1 SITS data set, the standard method yielded the accuracy of 68.59%, while the SimpleMKL provided the accuracy of 74.88%. The higher performances of the MKL algorithms for SITS classification is due to their ability to correctly model the different information level of the images acquired at different times.

Crop Name	Standard	MKL–sum	SimpleMKL	GMKL	GLMKL	LPMKL
Corn	0.86	0.87	0.86	0.86	0.86	0.88
Canola	0.68	0.78	0.85	0.85	0.79	0.79
Wheat	0.66	0.70	0.75	0.75	0.70	0.68
Soy	0.55	0.62	0.65	0.65	0.63	0.63
Oat	0.71	0.70	0.67	0.67	0.69	0.69
Sun flower	0.96	0.97	0.97	0.67	0.97	0.97
Barley	0.37	0.40	0.39	0.39	0.39	0.39
ACA	0.68	0.72	**0.74**	0.74	0.72	0.72
OA (%)	68.59	72.63	**74.88**	74.82	72.81	72.78
Kappa	0.63	0.68	**0.70**	0.70	0.68	0.68
Time (S)	-	0	26.0	13.9	3.1	2.5

Table 3. Obtained Results from classification of S1 using different algorithms.

Crop Names	Standard	MKL–sum	SimpleMKL	GMKL	GLMKL	LPMKL
Corn	0.76	0.77	0.78	0.78	0.77	0.77
Canola	0.97	0.94	0.94	0.94	0.94	0.94
Wheat	0.68	0.70	0.73	0.72	0.70	0.70
Soy	0.78	0.86	0.87	0.87	0.86	0.86
Oat	0.71	0.73	0.72	0.73	0.73	0.73
Forage	0.06	0.12	0.06	0.06	0.06	0.1
ACA	0.66	0.68	**0.68**	0.68	0.68	0.68
OA (%)	78.31	80.82	**82.16**	81.95	81.23	81.05
Kappa	0.72	0.75	**0.77**	0.76	0.76	0.75
Time (S)	-	0	25.5	10.3	4.0	2.3

Table 4. Obtained Results from classification of S2 using different algorithms.

Comparing the performance of the MKL algorithms showed that the optimization-based MKL algorithms (such as GMKL, LPMKL, …), were able to provide better performances than the fixed-rule MKL algorithm (such as the MKL-sum). This is because the optimization-based algorithms optimize the kernel weights, thus the influence of less informative kernels on the composite kernel will be decreased.

Among different optimization-based algorithms, the GMKL and the SimpleMKL provided better classification performances than the LPMKL and the GLMKL algorithms. For example for the S2 data set, these algorithms yielded the accuracies of 82.16%, 81.23%, 81.23%, and 81.05% respectively. However, since the objective functions of different optimization-based MKL algorithm are very similar to each other, these algorithms yielded similar performances.

The obtained results also showed that the class-specific accuracies obtained from most crops, particularly for soy and wheat, were dramatically increased in the case of using the MKL algorithms.

For a visual comparison the obtained classification maps of S1 data sets, using the standard method and MKL-sum,

SimpleMKL and GMKL are presented in Figure 4 and Figure 5.
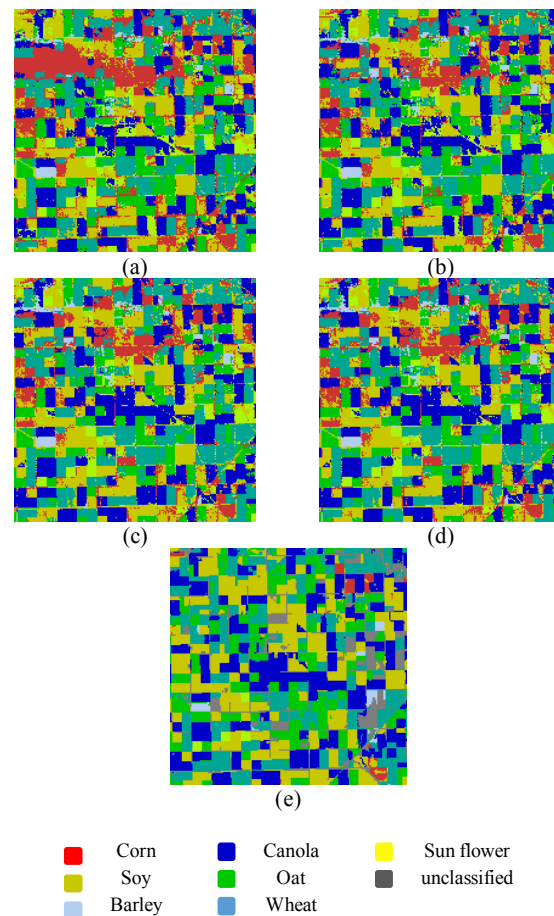

Figure 4. Classification maps of S1 data set obtained from a) Standard method; b) MKL-sum; c) SimpleMKL; d) GMKL; and e) ground truth.

Regarding the computational time, as mentioned before, the fixed-rule algorithms do not consider any optimization for estimating the kernel weights, thus the required computational time of the MKL-sum is zero in the tables. However, the computational times of different optimization-based algorithms vary as a result of using different optimization techniques. The SimpleMKL and the GMKL adopts a gradient descendent method for their optimization. This optimization method requires to estimate the gradient of MKL objective function at each iteration, which can be very time-consuming. The LPMKL and the GLMKL algorithms, due to using a closed-form solution for estimation of their kernel weights require less computational times than the other algorithms.

5. CONCLUSION

Crop mapping is among the most important application of SITS data. In this paper, the performance of different MKL algorithms for crop mapping using SITS data were evaluated. The theoretical and experimental comparison of the MKL algorithms led us to the following conclusions:

- MKL algorithms, due to their ability to model the different statistical characteristic of images of SITS data

set, can provide much better classification accuracy than other methods.

- The optimization-based wrapper MKL algorithms showed better classification performances than the fixed-rule MKL algorithms. The difference between the performances of these algorithms is higher if there are less informative kernels (e.g., kernel constructed from noisy or irrelevant to classification features) among the basis kernels. This is because in such cases, unlike the fixed-rule algorithms, the optimization-based algorithms can decrease the influence of these kernels on the composite kernels through optimization of the kernel weights.
- The SimpleMKL and the GMKL provided the best classification performances for both data sets. However, their optimization can be very time-consuming. The LPMKL and the GLMKL algorithms showed acceptable performances in term of both computational time and classification accuracy.

Although, the presented MKL algorithms in this paper provided acceptable performances, however, the other MKL algorithms for SITS data classification, should be studied. In addition, the effects of the parameters on the performances of the MKL algorithms needs to be studied further.

REFERENCES

Adhikari, R., Agrawal, R., 2013. An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.

Baydogan, M.G., Runger, G., 2015. Learning a symbolic representation for multivariate time series classification. Data Mining and Knowledge Discovery 29, 400-422.

Bucak, S.S., Jin, R., Jain, A.K., 2014. Multiple kernel learning for visual object recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 1354-1369.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Rojo-Álvarez, J.L., Martínez-Ramón, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. IEEE Transactions on Geoscience and Remote Sensing 46, 1822-1835.

Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., Calpe-Maravilla, J., 2006. Composite kernels for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters 3, 93-97.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2, 27.

Gerstmann, H., Möller, M., Gläßer, C., 2016. Optimization of spectral indices and long-term separability analysis for classification of cereal crops using multi-spectral RapidEye imagery. International Journal of Applied Earth Observation and Geoinformation 52, 115-125.

Gevaert, C., Persello, C., Vosselman, G., 2016. Optimizing Multiple Kernel Learning for the Classification of UAV Data. Remote Sensing 8, 1025.

Gomez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: a review and future directions. Proceedings of the IEEE 103, 1560-1584.

Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. Journal of Machine Learning Research 12, 2211-2268.

Gu, Y., Gao, G., Zuo, D., you, D., 2014. Model Selection and Classification With Multiple Kernel Learning for Hyperspectral Images via Sparsity. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 2119-2130.

Gu, Y., Wang, C., You, D., Zhang, Y., Wang, S., Zhang, Y., 2012. Representative multiple kernel learning for classification in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 50, 2852-2865.

Jamali, S., Seaquist, J., Eklundh, L., Ardö, J., 2014. Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. Remote Sensing of Environment 141, 79-89.

Jonsson, P., Eklundh, L., 2004. TIMESAT - a program for analyzing time-series of satellite sensor data. Computers & Geosciences 30, 833-845.

Keogh, E.J., Pazzani, M.J., 1998. An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback, KDD, pp. 239-243.

Kim, D.-W., Lee, K., Lee, D., Lee, K.H., 2005. Rapid and brief communication Evaluation of the performance of clustering algorithms in kernel-induced feature space. Pattern Recognition 38, 607 – 611.

Kloft, M., Brefeld, U., Laskov, P., Müller, K.-R., Zien, A., Sonnenburg, S., 2009. Efficient and accurate lp-norm multiple kernel learning, Advances in neural information processing systems, pp. 997-1005.

Kloft, M., Brefeld, U., Sonnenburg, S., Zien, A., 2011. Lp-norm multiple kernel learning. The Journal of Machine Learning Research 12, 953-997.

Li, Q., Cao, X., Jia, K., Zhang, M., Dong, Q., 2014. Crop type identification by integration of high-spatial resolution multispectral data with features extracted from coarse-resolution time-series vegetation index data. International Journal of Remote Sensing 35, 6076-6088.

Liu, T., Gu, Y., Jia, X., Benediktsson, J.A., Chanussot, J., 2016. Class-Specific Sparse Multiple Kernel Learning for Spectral–Spatial Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 54, 7351.

Löw, F., Knöfel, P., Conrad, C., 2015. Analysis of uncertainty in multi-temporal object-based classification. ISPRS Journal of Photogrammetry and Remote Sensing 105, 91-106.

Niazmardi, S., Demir, B., Bruzzone, L., Safari, A., Homayouni, S., 2016. A comparative study on Multiple Kernel Learning for remote sensing image classification, Geoscience

and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE, pp. 1512-1515.

Niazmardi, S., Homayouni, S., McNairn, H., Shang, J., Safari, A., 2014. Multiple kernels learning for classification of agricultural time series data, Agro-geoinformatics (Agro-geoinformatics 2014), Third International Conference on. IEEE, pp. 1-4.

Niazmardi, S., Safari, A., Homayouni, S., 2017. Similarity-Based Multiple Kernel Learning Algorithms for Classification of Remotely Sensed Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 2012-2021.

Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y., 2008. SimpleMKL. Journal of Machine Learning Research 9, 2491-2521.

Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A., Chehbouni, A.G., 2008. The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco. International Journal of Remote Sensing 29, 95-116.

Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M., 2010a. Learning relevant image features with multiple-kernel classification. IEEE Transactions on Geoscience and Remote Sensing 48, 3780-3791.

Tuia, D., Ratle, F., Pozdnoukhov, A., Camps-Valls, G., 2010b. Multisource composite kernels for urban-image classification. IEEE Geoscience and Remote Sensing Letters 7, 88-92.

Varma, M., Babu, B.R., 2009. More generality in efficient multiple kernel learning, Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 1065-1072.

Verhegghen, A., Bontemps, S., Defourny, P., 2014. A global NDVI and EVI reference data set for land-surface phenology using 13 years of daily SPOT-VEGETATION observations. International Journal of Remote Sensing 35, 2440-2471.

Wang, Q., Gu, Y., Tuia, D., 2016. Discriminative Multiple Kernel Learning for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing PP, 1-16.

Xu, Z., Jin, R., Yang, H., King, I., Lyu, M.R., 2010. Simple and efficient multiple kernel learning by group lasso, Proceedings of the 27th international conference on machine learning (ICML-10), pp. 1175-1182.

Zhou, Y., Peng, J., Chen, C.L.P., 2015. Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, 2351-2360.