Multi-View Active Learning Optimization Based on Genetic Algorithm and Gaussian Mixture Models for Hyperspectral Data

Nasehe Jamshidpour, *Student Member, IEEE,* Abdolreza Safari, Saeid Homayouni, *Senior Member, IEEE*

***Abstract*—In this paper, we proposed a novel optimal view generation framework based on the genetic algorithm (GA) and Gaussian mixture models (GMMs) to improve multi-view active learning (MV-AL). AL methods enlarge training data sets, by iteratively selecting the most informative samples, in order to improve classification performance. By using multiple views to build multiple classifiers, the information content of each unlabeled samples can be more accurately estimated. The MV-AL methods are more inherently suitable for high dimension data such as hyperspectral images. This hybrid framework simultaneously constructs the optimal number of diverse and sufficient views. The proposed algorithm has two main steps. In the first step, by applying a cluster distortion function-based GMMs, the actual number of available independent views is determined. In the next step, a hybrid GA approach selects the optimal combination of views using two different criteria. The experiments were conducted on two benchmark hyperspectral datasets, namely KSC and Indian Pines AVIRIS. The results demonstrated increasing in diversity, and sufficiency of the views compared to the traditional view generation methods. Furthermore, the performance of MV-AL has also been significantly improved.**

***Index Terms*—Active learning (AL), Gaussian Mixture Models (GMMs), Genetic algorithms (GA), Multi-view learning, View generation methods.**

(Corresponding author: Abdolreza Safari)

N. Jamshidpour and A. Safari are with the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. (E-mail: njamshidpour@ut.ac.ir, asafari@ut.ac.ir).

S. Homayouni is with the Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Quebec, Canada (E-mail: saeid.homayouni@ete.inrs.ca).

I. INTRODUCTION

The hyperspectral image is a high dimensional data cube consists of many narrow spectral bands. This high dimensionality imposes some difficulties for data processing such as overfitting, inaccurate model's parameters estimation, and the Hughes phenomenon. To overcome these issues highly sophisticated dimension reduction methods have been suggested so far [1, 2].

However, this high dimensionality makes them intrinsically suitable for multi-view (MV) [3] for supervised learning algorithms. However, there is also some successful unsupervised multi-view learning method such as [4]. Using multiple views, instead of a single view, for training a learning algorithm, can simultaneously reduce the generalization error rate and improve the learning performance. In addition to dividing the initial feature space into two or more subsets, the curse of dimensionality of the data can be significantly alleviated. However, MV learning methods, for being successful, need an efficient view generation method that produces sufficient and diverse views simultaneously [5]. In other words, each view must be sufficient to train a suitable and reliable classifier and also provide redundant information to solve the classification problem.

Active Learning (AL) has recently attracted much attention because of its outstanding performance to improve classification task with a limited number of labelled training data [6-10]. AL methods try to build a compact and well-chosen training data set by selecting the most informative samples for the current model at each iteration [11]. The selected instances are then labeled by an expert and added to the training set. The key difference between various AL methods is the type of query function, which estimates the amount of information content for all candidate samples.

The main concern of multi-view active learning (MV-AL) methods is how to choose the most valuable instances using multiple views of the original data. Although, many MV-AL methods have been proposed for hyperspectral image classification [12, 13]. They mostly used simple band partitioning methods based on the correlation between different, such as k-means, correlation, and uniform band slicing. In this way, the sufficiency of each view was not considered. Therefore in this paper, we suggest a hybrid view generation method for hyperspectral data based on the genetic algorithm (GA) that maximize diversity between views and sufficiency of each one.

GA is one of the evolutionary optimization algorithms that has been extensively employed to solve multi-criteria optimization problems [14]. The algorithm initializes with a random population which represents the possible solutions to the problem. Then, by the progress of the algorithm, the solution is to approach the optimal answer by considering the criteria.

In the current optimization problem, we have dual criteria, first maximizing the sufficiency of each view, and second, minimizing the dependency between views. Thus, by integrating these criteria, a hybrid optimization problem is formed, because the first criteria is a wrapper and the second one is a filter problem. The wrapper methods need to perform the classification is each time, whereas the filter method is independent of the classification task. Therefore, the filter methods are faster, but the wrapper methods produce a more model-based solution. The obtained experimental results demonstrated the efficiency of the hybrid solution in comparison of wrapper or filter methods.

The optimal number of views is another crucial issue in MV-AL which has been mostly neglected in the literature. Although in the earliest MV-AL methods, e.g., Co-EM [15], only two distinct views have been employed, the data structure might imply a different number of views. In this paper, we initially determined the optimal number of views using Gaussian mixture models (GMM) [16]. Secondly, the optimal views by considering the proposed new hybrid criterion were chosen.

II. METHODOLOGY

A. *Finding the Optimal Number of Views*

Gaussian mixture models (GMM) are employed to approximate non-gaussian and discrete variables by the weighted sum of a finite number of Gaussian mixture components [17].

$$P(x|\lambda) = \sum_{i=1}^{M} \omega_i g(x|\mu_i, \Sigma_i) \tag{1}$$

where x is the approximated input D-dimension data vector, ω_i and $g(x|\mu_i, \Sigma_i)$ are the mixture weight and the Gaussian density, respectively. To approximate the input feature vector, the GMM parameter set $\lambda = (\omega_i, \mu_i, \Sigma_i)$ must be estimated, with the constraint $\sum_{i=1}^{M} \omega_i = 1$. There are different popular methods such as maximum likelihood (ML) and expectation maximization (EM) [18]. In this paper, we used the ML method to solve the GMM approximation.

The number of these components must be determined as *a prior* which significantly affects the emerged components. Therefore, some studies are trying to find the optimal number of the components [19, 20] that also indicate the number of available clusters. Incremental *k*-means clustering is one of the mixtures' number optimization method [20]. In this method, *k* is gradually increased as long as all components are still independent. The mutual information is employed for determining the statistical independence. If by adding the k^{th} component, at least one component becomes dependent on the other components, the optimal number of components will be chosen as (*k*-1).

Another way to estimate the optimal number is the cluster distortion function-based approach [19]. In this paper, we used this algorithm which proposes the function $f(k)$ based on the sum of cluster distortion (I_j) for *k* number of clusters:

$$S_k = \sum_{j=1}^{k} I_j \tag{2}$$

The function $f(k)$ is defined as the ratio of S_k and S_{k-1} as follows:

$$f(k) = \begin{cases} 1 & k = 1 \\ \frac{S_k}{\alpha_k S_{k-1}} & k > 1 \end{cases} \tag{3}$$

The coefficient α_k is also defined by the coefficient value of the previous step α_{k-1}:

$$\alpha(k) = \begin{cases} 1 - \frac{3}{4D} & k = 2 \\ \alpha_{k-1} + \frac{1-\alpha_{k-1}}{6} & k > 2 \end{cases} \quad (4)$$

and D is the dimension of the data vector. The minimum value of $f(k)$ indicates the most concentrated clustering and also the optimal number of components. We used the incremental search of k to minimize the $f(k)$ function.

B. View Generation by Hybrid GA-FSS Method

The main characteristic of MV-AL algorithms is how to generate views. The constructed views should have some essential characteristics to ensure the superiority of the MV-AL over single view AL methods.

First, each view must impose redundant information compared to the other views, which is measured by the diversity of the views. Moreover, the views should be as independent as possible from each other. On the other hand, the sufficiency of each view to building a good classifier should be taken into account. Our proposed GA-based view generation method tries to find the best set of multiple-views by integrating all the above criteria. The general framework of the proposed GA-based method is given in Fig. 1.

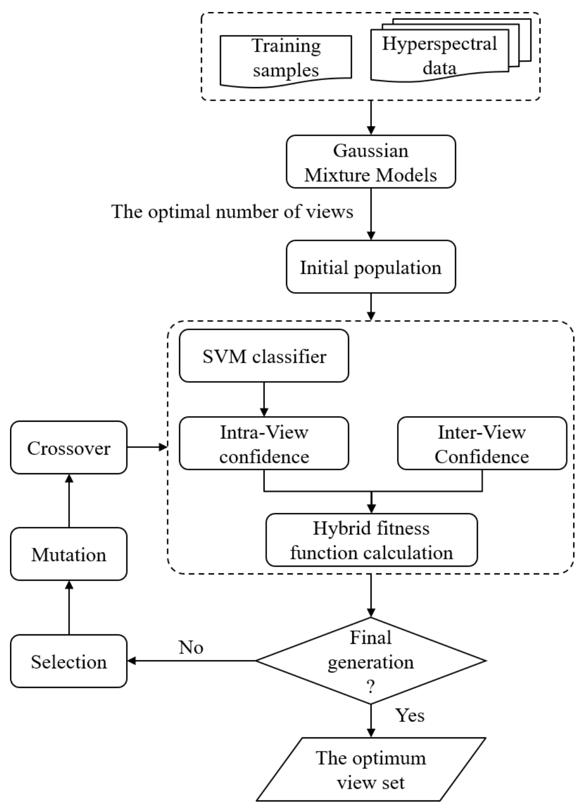

Fig. 1. Flowchart of the proposed hybrid GA-FSS method.

GA has been mostly used as a feature subset selection (FSS) method for dimension reduction of hyperspectral data [21]. The primary goal of FSS is to build a sufficiently unique and compact subset of features. We proposed a novel framework based on GA-FSS to construct multiple diverse, dependent and sufficient views. The FSS problem traditionally is solved in two ways using either wrapper or filter criteria. To find out a more appropriate solution, we employed a hybrid approach by integrating both filter and wrapper criteria. For this purpose, we introduce the *inter-confidence* criteria representing the diversity and independence between all the views, as the filter method:

$$C_{inter} = \sum_{i=1}^{N_v} \frac{1}{MI(X^i, X^j)} \quad (5)$$

$MI(X^i, X^j)$ is the mutual information between the i^{th} and j^{th} views. In the other hand, we used the *intra-confidence* as the wrapper criteria, which is computed by the mean of all views' sufficiency.

$$C_{intra} = \frac{1}{N_v} \sum_{i=1}^{N_v} P(X_L^i, X_U^i, Y_L) \quad (6)$$

where P represents the classification performance. Therefore, the hybrid fitness function is formed by the weighted summation of both filter and wrapper criteria using w_f and w_w weights respectively.

$$F = w_f * C_{inter} + w_w * C_{intra} \quad (7)$$

The weights are experimentally set as $w_f = w_w = 0.5$ using 2D grid search method with the search space {0.1, 0.2, 0.3, …, 0.9} for each parameter.

In the GA algorithm, each chromosome represents a solution, and at each generation, the population gets closer to the optimum solution by evolutionary operators such as selection, crossover and mutation and migration. Therefore, in our proposed GA scheme a binary chromosome with the size of $N_v \times D$ was used, where D and N_v are the initial dimension of hyperspectral data and number of views. Each D bits represent a single view, where the value of 0 or 1 for each bit shows the presence or absence of the corresponding band in the view.

III. Experimental Results

A. Hyperspectral Data

For the evaluation of the proposed method, two hyperspectral data set, provided over the Kennedy Space Center (KSC) and Indian Pines, were employed. These data sets, acquired by the AVIRIS sensor, but with different 18-m and 30-meter spatial resolutions due to their different altitude. After excluding noisy and water absorption bands, 176 and 185 bands remain for KSC and Indian Pines dataset, respectively. The KSC image contains 512×614 pixels which include 13 different land cover classes with overall

5,211 labelled samples. While the Indian Pines data set including 16 different land cover classes with 10,366 ground truth samples in a 145 by 145 scene.

B. Experimental Setup

The experiment was conducted using the training datasets that were produced by combining '5-fold' and 'hold-out' randomly selected cross-validation schemes to make experimental results robust to the training sets and prove the generalization ability of the proposed model. First, the initial labelled data set was divided into 5-fold complementary subsets. Then, the 'hold-out' cross-validation mechanism was employed. We have considered the worst scenario with the smallest size of the training set, i.e. only five labelled samples per class. The performance of the proposed and baseline methods was evaluated using the average overall accuracy ($\overline{OA}$) over five folds.

Support vector machine (SVM) was used as the classifier because of its excellent performance and it only two free parameters that need to be tuned. They were tuned using 2-D grid search with a search space of 0.05, 0.1, 0.15, …, 0.95.

At the first iteration of the proposed MV-AL framework, the SVM classifier was trained, then by using the AMD query function [12] the five most informative samples were added at each step. The AL algorithm stopped after reaching to 40 iterations and adding 200 samples to the training dataset.

There are also some parameters of GA which must be selected. The initial population was randomly generated with a size of 50 individuals. The crossover, migration and mutation rate were selected as 0.8, 0.2 and 0.05 respectively. The elitism selection strategy was implemented. The maximum number of generation was set to 100, and the number of consecutive unchanged iterations was 20.

C. Experimental Results

In this paper, two experiments were conducted to investigate the efficiency of our proposed method. First, we determined the maximum number of the available independent views in the datasets using GMMs. In the second experiment, the GA-FSS view generation was compared with six related methods.

The optimal number of independent views was estimated using the proposed cluster distortion function-based GMMs method. The experimental results for both datasets are reported in Table I.

As it can be seen for each data set, there is an optimal number of views with the minimum value of the cluster distortion function *f(k)*. Also, to evaluate the obtained views, we computed the average mutual information ($\overline{MI}$) between each pair of views. As shown the optimal number of views provides the most diverse views with the minimum averaged mutual information value. On the other hand, the required CPU time by the proposed hybrid GA-FSS method are reported in the table. As expected, by increasing the number of views the computation cost also increases. Therefore, employing the proposed function-based GMMs method to find the optimal number of views is an essential and affordable preliminary step for MV-AL.

TABLE I
CLUSTER DISTORTION FUNCTION ($f(k)$), MUTUAL INFORMATION ($\overline{MI}$) AND REQUIRED CPU TIME (TIME) FOR THE DIFFERENT NUMBER OF GENERATED VIEWS BY THE PROPOSED HYBRID GA-FSS METHOD.

Number of views	KSC			Indian Pines		
	$f(k)$	$\overline{MI}$	Time (s)	$f(k)$	$\overline{MI}$	Time (s)
1	1.00	-----	-----	1.00	-----	-----
2	0.93	0.336	436.25	**0.19**	**0.179**	**498.53**
3	0.98	0.317	443.89	1.04	0.286	578.96
4	**0.32**	**0.109**	**523.86**	0.31	0.591	715.48
5	1.39	0.298	558.88	1.26	1.105	846.73
6	0.77	0.415	637.63	0.75	2.831	953.21
7	1.29	0.471	756.01	1.47	3.884	1132.8

We also investigated the effect of the number of views on MV-AL performance. The proposed GA-FSS method generated a different number of views. The learning curves of the performed MV-AL on these sets of views for KSC data set are given in Fig. 2. The best performance was achieved by the four views which have been already determined as the optimal number by our proposed method.

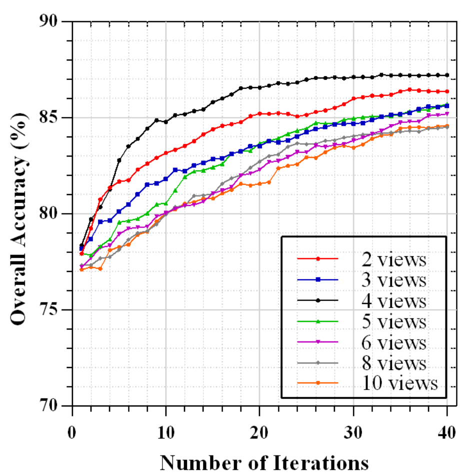

Fig. 2. Average learning curves of MV-AL for the different number of views using the proposed GA-FSS view generation method for KSC dataset.

In the second experiment, to demonstrate the effectiveness of the proposed GA-hybrid view generation method, the obtained results were compared to four traditional and also two other GA methods. The baseline methods were conventional band clustering algorithms including k-means, correlation, uniform, and random view generation methods which were proposed in [12]. Two single criterion GA-based methods were also implemented, a wrapper method as called GA-intra which only considers the intra-confidence. The other one was GA-inter method which is a filter method because of its independence of the classifier performance.

As can be seen in Fig. 3, the learning curve of the proposed method has the best performance in the KSC data set. As expected, random band clustering had the worst performance, while other band clustering methods produced similar results to some extent. Also, the numerical results of the proposed view generation methods and two state-of-the-art methods, view updating (VU) [12] and multiple morphological component analysis (MMCA) [13] for both KSC and Indian Pines data sets are presented in Table II.

TABLE II

ACHIEVED AVERAGED OVERALL ACCURACY ($\overline{\text{OA}}$) AFTER 40 ITERATIONS OF MV-AL METHOD AND REQUIRED CPU TIME BY DIFFERENT VIEW GENERATION METHODS FOR KSC AND INDIAN PINES DATASETS.

VG Method	KSC		Indian Pines	
	$\overline{\text{OA}}$ (%)	Time (s)	$\overline{\text{OA}}$ (%)	Time (s)
K-Means	85.50	11.31	81.76	15.64
GA-inter	86.14	420.34	82.28	312.77
GA-intra	86.74	445.71	82.60	365.03
VU [12]	85.63	356.12	82.01	298.47
MMCA [13]	85.47	207.98	82.63	143.28
GA-hybrid	**87.21**	**523.86**	**83.93**	**498.53**

Although the proposed hybrid view generation method implied a higher computation cost in comparison to the other methods, it was still feasible to implement. This higher required time was unavoidable due to the iterative nature of GA, which tries to find the optimal set of the views. However, the generated optimal views achieved the best performance of MV-AL for both data sets.

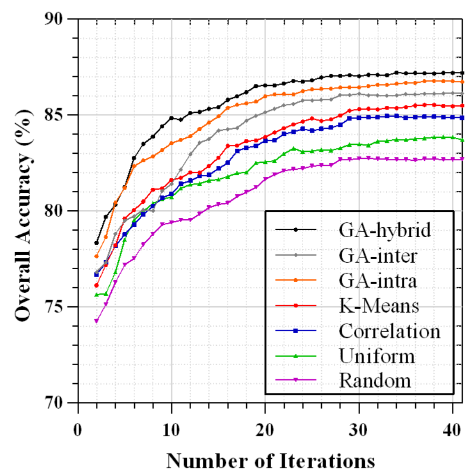

Fig. 3. Average learning curves of MV-AL for different view generation methods using four views and 5-fold cross-validation for KSC dataset.

IV. CONCLUSION

In this paper, we designed a novel framework to produce an optimal multi-view set to enhance hyperspectral multi-view active learning performance. The framework consisted of two steps. In the first step, the actual number of underlying independent views in the data was determined by GMMs. Then, a hybrid view generation method based on GA produced the optimal view set by considering two different filter and wrapper criteria. The filter criterion was defined to maximize the diversity between each pair of views, while the wrapper criterion tried to guarantee the sufficiency of the views for learning the classifier. The experiments have been conducted on KSC-AVIRIS and Indian Pines datasets. The experimental results confirmed that the proposed GMM function-based method efficiently selects the best number of views. By employing this method both the view intensity and the performance of the MV-AL improved. Therefore, the low computation cost and the observed efficiency of this method make it suitable as a basic step in all MV algorithms. On the other hand, a significant improvement achieved by the proposed GA-based view generation method compared to the conventional and also single-criterion GA-based methods. Despite the expected high computation cost of the proposed method, it was able to enhance the diversity and sufficiency of the views and also the final MV-AL method, consequently.

REFERENCES

[1] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proceedings of the IEEE*, vol. 101, pp. 676-697, 2013.

[2] Q. Wang, F. Zhang, and X. Li, "Optimal Clustering Framework for Hyperspectral Band Selection," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1-13, 2018.

[3] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[4] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[5] A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications*, vol. 49, pp. 31-47, 2016.

[6] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, "Random-Walker-Based Collaborative Learning for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 212-222, 2017.

[7] J. Guo, X. Zhou, J. Li, A. Plaza, and S. Prasad, "Superpixel-Based Active Learning and Online Feature Importance Learning for Hyperspectral Image Analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 347-359, 2017.

[8] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A Novel Semisupervised Active-Learning Algorithm for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[9] B. Du, Z. Wang, L. Zhang, L. Zhang, W. Liu, J. Shen, *et al.*, "Exploring representativeness and informativeness for active learning," *IEEE transactions on cybernetics*, vol. 47, pp. 14-26, 2017.

[10] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian Active Learning with Image Data," *arXiv preprint arXiv:1703.02910*, 2017.

[11] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *Selected Topics in Signal Processing, IEEE Journal of,* vol. 5, pp. 606-617, 2010.

[12] W. Di and M. M. Crawford, "View generation for multiview maximum disagreement based active learning for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 50, pp. 1942-1954, 2012.

[13] X. Xu, J. Li, and S. Li, "Multiview Intensity-Based Active Learning for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, pp. 669-680, 2018.

[14] A. P. Engelbrecht, "Introduction to Evolutionary Computation," in *Computational Intelligence: An Introduction*, 2nd ed Chichester: John Wiley & Sons, Ltd, 2007, pp. 127-142.

[15] W. Wang and Z.-H. Zhou, "Analyzing co-training style algorithms," in *European Conference on Machine Learning*, 2007, pp. 454-465.

[16] D. Reynolds, "Gaussian mixture models," *Encyclopedia of biometrics*, pp. 827-832, 2015.

[17] S. Tadjudin and D. Landgrebe, "Robust parameter estimation for mixture model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, pp. 439-445, 2000.

[18] S. Bonhomme, K. Jochmans, and J. M. Robin, "Non-parametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 78, pp. 211-229, 2016.

[19] B. R. Prusty and D. Jena, "Estimation of optimal number of components in Gaussian mixture model-based probabilistic load flow study," in *India Conference (INDICON), 2016 IEEE Annual*, 2016, pp. 1-6.

[20] Y. Lee, K. Y. Lee, and J. Lee, "The estimating optimal number of Gaussian mixtures based on incremental k-means for speaker identification," *International Journal of Information Technology,* vol. 12, pp. 13-21, 2006.

[21] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Systems,* vol. 24, pp. 40-48, 2011.