BUILDING EMERGING PATTERN (EP) RANDOM FOREST FOR RECOGNITION

Liang Wang[*†], *Yizhou Wang*[†‡], *Debin Zhao*[*]

[*]*School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China*
[†]*Nat'l Engineering Lab for Video Technology, Peking University, Beijing, China*
[‡]*Key Lab. of Machine Perception (MoE), School of EECS, Peking University, Beijing, China*
{*wangliang@jdl.ac.cn, Yizhou.Wang@pku.edu.cn, dbzhao@jdl.ac.cn*}

ABSTRACT

The Random forest classifier comes to be the working horse for visual recognition community. It predicts the class label of an input data by aggregating the votes of multiple tree classifiers. However, the classification performances of these tree classifiers are different. The random forest classifier ignores the difference by simply assigning them equal weights in voting for the final classification decision. Also, the random forest classifier only casts votes from individual tree classifiers without considering their compositions which would be more accurate. In this paper, we propose to tackle the two points by discovering weighted decision rules from the tree classifiers' output sets on training data. By treating the outputs of the tree classifiers on each data as a digital itemset, we want to find discriminative patterns (either containing the output of a single tree classifier or a set of tree classifiers) from the itemsets of training data. We employ an efficient data mining algorithm, the Emerging Pattern (EP) Mining, to search such discriminative patterns and weight them according to their discriminative powers. A set of decision rules are built from these discovered patterns and the final outputs of the Random Forest are made using these decision rules. We call the proposed classifier Emerging Pattern (EP) Random Forest. Experimental results on action categorization problems confirm that the proposed method really improve the performance of the traditional Random Forest classifier.

Index Terms— Random forest, Emerging pattern mining, Action recognition

1. INTRODUCTION

In recent years, random forest classifier is hot in computer vision community due to its excellent classification performance and the efficiency in training and testing. It has been widely used in object categorization [1], image labeling [2], etc. The Random Forest classifier predicts the class label of an input data by aggregating the predictions made by a set of tree classifiers. Each tree classifier divides the feature space into a number of partitions and its output for an input data is determined according to the partition the data lying in. As a result, the classification accuracy of a tree classifier strongly depends on the data class purity of its partitions (i.e. whether each partition contains purely one class of data or is filled with multiple classes of data). For example, for a three-class classification problem, the partition containing the portions of the three classes of training data as $class1$: 60%, $class2$: 20%, $class3$: 20% and the partition with the portions as $class1$: 40%, $class2$: 30%, $class3$: 30% should both assign the data lies in them with the class label 1. However, the classification accuracy of the two partitions are different. Random Forest ignores this difference and weights them equally to vote for the final classification decision. Moreover, it has been proved that the compositions of weak classifiers perform better than individual weak classifiers [3].

In this paper, we propose a new methodology to combine the outputs of the tree classifiers to produce the final decision of the Random Forest. Instead of producing the classification decision through voting the individual tree classifiers' outputs, we learn a set of decision rules from the tree classifiers' outputs on the training data set. A decision rule predicts the class label of an input data by checking the output(s) of a single or a composition of tree classifier(s). The decision rules are weighted according to their discriminative power and the final decision of the Random Forest classifier is made by aggregating the votes of these decision rules, and these votes are weighted using the corresponding classifiers' weight. To discover decision rules from the outputs of the tree classifiers on the training data, we employ a efficient contrast pattern mining method, the Emerging Pattern (EP) mining algorithm. The resulted classifier is called EP Random Forest since it combine the two approaches in a single framework.

The remaining of the paper is organized as follows. The algorithm to learn the contextual Random Forest classifier is presented in Section 2. Experimental results are shown in Section 3 and we conclude in Section 4.

This research was supported in part by NSFC-60872077, NSFC-60833013, the Scientific Research Foundation for Returned Scholars, and Major State Basic Research Development Program of China (973 Program 2009CB320904).

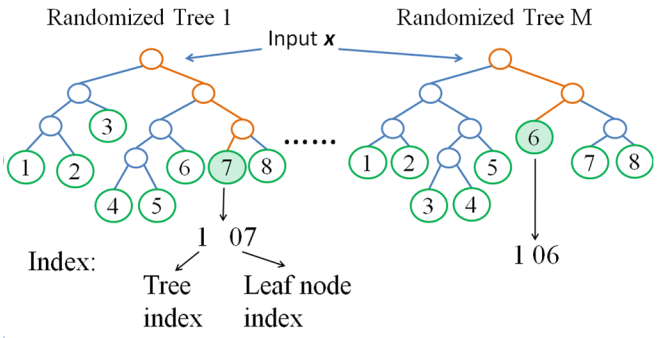

Fig. 1. Illustration of the method to assign index to each partition of a randomized tree. Each partition is associated with a leaf node of the randomized tree.

2. EMERGING PATTERN RANDOM FOREST

In this section, we introduce the basic concept and implementation details of the proposed EP Random Forest classifier.

2.1. Mining decision rules from the outputs of the weak classifiers

A random forest classifier is a collection of a set of tree classifiers, denoted as $\mathcal{H} = \{h_j\}_{j=1}^{M}$, where M is the number of tree classifiers. For an input data, each tree classifier casts a vote for the final decision of the random forest and the class label of the input data is determined to be the one with the maximum number of votes.

Each tree classifier divides the feature space into a number of partitions. The output ρ_j of a tree classifier h_j for an input data $\mathbf{x}$ is assigned according to the partition $\mathbf{x}$ lies in. Thus, by passing $\mathbf{x}$ to the random forest classifier, the index set of the partitions it falling in is $\mathcal{T}(\mathbf{x}) = \{\rho_1, \rho_2, \ldots, \rho_M\}$. The method to index the partitions of tree classifiers are illustrated in Fig. 1. Instead of making classification decision for $\mathbf{x}$ directly from individual weak classifiers' outputs, we propose to discover discriminative patterns from the index sets of all the training data and make classification decision according to these discriminative patterns. Here, a pattern is in the form of a partition index set.

To mine discriminative patterns, we adopt the emerging pattern mining algorithm proposed in [4]. Intuitively, EP mining finds the patterns whose support ratios are significantly different from one dataset to the other. We use the notation in [4] to introduce the mathematical definition of EP mining. Let $I = \{i_1, i_2, ..., i_N\}$ be a set of N items. A *transaction* is a subset T of I. A *data set* D is a set of transactions. A subset S is called a *k-itemset* if $k = \|S\|$. If $S \subseteq T$, we say the transaction T contains the itemset S. The *support* of S in a data set D is defined as $\rho_S^D = \frac{\text{count}_D(S)}{\|D\|}$, where $\text{count}_D(S)$ is the number of transactions in D containing S. Given an itemset S and a pair of data sets D_1 and D_2, the *growth rate* of an itemset S from D_1 to D_2 is computed as

$$\tau_S^{D_1 \to D_2} = \begin{cases} 0, & \text{if } \rho_S^{D_1} = 0 \text{and } \rho_S^{D_2} = 0 \\ \infty, & \text{if } \rho_S^{D_1} = 0 \text{and } \rho_S^{D_2} \neq 0 \\ \rho_S^{D_2}/\rho_S^{D_1}, & \text{otherwise} \end{cases}$$

A pattern is said to be a *η-emerging pattern* from D_1 to D_2 if $\tau_S^{D_1 \to D_2} > \eta$.

According to the terminology of EP mining, we call the index set associated with each training data the *transaction* of the data. Then, a set of emerging patterns are mined for each data class. To mine the EPs for data class y, we use the transactions of class y's training data as the positive dataset, and the transactions of the rest training data are treated as the negative dataset. Then the EP set of data class y are obtained by perform EP mining from the negative dataset to the positive one, denoted as $\mathcal{P}_y = \{\mathbf{p}_{yj}\}_{j=1}^{n_y}$. We specify two threshold parameters for mining the EPs: the basic support ratio in the positive class and the growth rate. By defining the basic support ratio of the positive class, we maintain the descriptivity of the mined discriminative patterns for the positive data; and through the basic growth rate parameter, the discriminative power of the mined patterns is guaranteed.

Each EP $\mathbf{p}_{yj}$ can be used to form a decision rule γ_{yj} for data class y

$$\gamma_{yj} : \mathbf{p}_{yj} \to y \tag{1}$$

and according to [5], its confidence score $\text{S}(\gamma_{yj})$ for making the decision can be computed as

$$\text{S}(\gamma_{yj}) = \rho_{\mathbf{p}_{yj}} * \frac{\tau_{\mathbf{p}_{yj}}}{\tau_{\mathbf{p}_{yj}} + 1} \tag{2}$$

$\rho_{\mathbf{p}_{yj}}$ and $\tau_{\mathbf{p}_{yj}}$ refer to the support ratio and growth rate of emerging pattern $\mathbf{p}_{yj}$.

2.2. Classification by the mined decision rules

Based on the decision rule set of each data class, we determine the class label of an input data by aggregating the confidence score of the decision rules whose corresponding emerging pattern are contained in data's itemset.

Given an input data $\mathbf{x}$, for each class y, we compute the score of $\mathbf{x}$ belonging to y by aggregating the confidence score of the decision rules whose corresponding emerging pattern are contained in $\mathbf{x}$'s itemset $\mathcal{T}(\mathbf{x})$. The aggregation is performed in the same way as the emerging pattern based classification method *Classification by Aggregating Emerging Patterns (CAEP)* proposed by Dong et al. [5]. It classifies $\mathbf{x}$ by

$$y^* = \arg\max_y \mathcal{S}(\mathbf{x}, \Gamma_y) \tag{3}$$

where Γ_y is the decision rule set of y and $\mathcal{S}(\mathbf{x}, \Gamma_y)$ is the confidence score of $\mathbf{x}$'s itemset $\mathcal{T}(\mathbf{x})$ for class y

$$\mathcal{S}(x, \Gamma_y) = \frac{1}{Z_y} \sum_{\substack{\mathbf{p}_{yj} \subseteq \mathcal{T}(x) \\ \gamma_{yj} \in \Gamma_y}} \text{S}(\gamma_{yj}) \tag{4}$$

where $\mathbf{p}_{yj}$ is the emerging pattern corresponding to decision rule γ_{yj} (refer to Eq .1). Because different class has different number of emerging patterns, the score of an input data for each class is normalized by a normalization term to account for this difference. The normalization term Z_y of class y is computed as the median value of the scores of all the training data belongs to class y computed as in Eq. 3.

3. EXPERIMENTS

We conduct experiments on video action categorization to evaluate the performance of the proposed algorithm. The task of action categorization is to classify video sequences based on the action types they contain. The KTH human motion dataset [6] are employed in the experiments.

In the experiments, an action video is represented by the set of spatiotemporal interest points (STIPs) it contains. STIP is a type of sparse local feature which characterizes the local changes of appearance and motion in videos. For each video, we detect STIPs in eight different scales using the interest point detector proposed in [7], and extract a 3D spatial-temporal (ST) cuboid centered at each interest point. A cuboid's size is three times the scale of its corresponding STIP. Then we normalize all the cuboids into the same size, compute the gradients in x, y, t dimensions at each pixel, and build a gradient feature vector for each cuboid. The principle component analysis (PCA) is used to reduce the dimension of each feature vector to 100. Using these feature vectors, the detected STIPs are clustered into a predefined number (e.g. 120) of clusters (visual words) by the K-means algorithm.

Using the STIP based action video representation, the feature vector of an action video is the histogram counting the visual word occurrence frequency in the video. We utilize a cross-validation scheme to test the performance of the proposed approach. Specifically, each time, we use videos of randomly selected 20 actors as training data and the rest as testing videos. The basic support ratio and growth rate for mining the emerging patterns is 0.6 and 5 respectively. The Random Forest classifier with the same set of weak classifiers as our method is adapted as the baseline method to compare with the proposed approach.

3.1. Experiments on KTH Dataset

The KTH human motion dataset [6] contains twenty-five people performing six types of actions, namely, "boxing", "waving", "handclapping", "jogging", "running" and "walking". For each people, the actions were captured under four different environments with variations in scales, illuminations and camera motions, but all the videos were shot with simple backgrounds. Each type of actions contains about $96 \sim 99$ available sequences.

Twenty rounds of cross-validation are performed on this dataset. The average recognition accuracy of the proposed

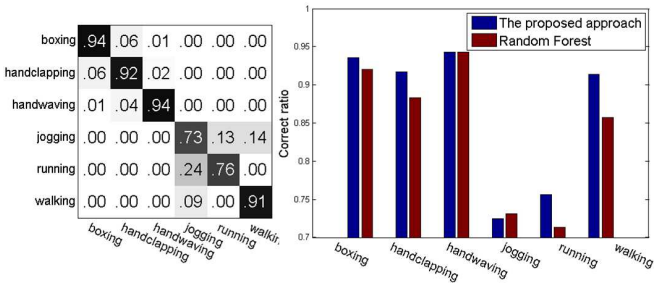

Fig. 2. Left: the confusion matrix illustrating the classification accuracy of the proposed method on KTH human action dataset. Right: comparison of the results of the proposed approach and the baseline method.

Method	Classifier	Accuracy (%)
Dollár, et al.[7]	SVM	81.2
Wong, et al.[10]	NNC	80.1
Niebles et al.[11]	pLSA	81.5
Nowozin et al.[8]	lpBoost	84.7
Schuldt et al.[6]	SVM	71.7
Liu et al.[9]	SVM	91.3
Ours	EP Random Forest	86.5

Table 1. The performance of the state-of-the-art approaches categorizing actions using STIP features.

method and the comparison with the baseline classification method are shown in Fig. 2. It can be observed that the proposed method outperforms the baseline method in most of the action classes. The results are obtained by using 120 clusters of visual words and 60 weak classifiers.

We also list other state-of-the-art approaches which also categorize actions based on STIP features in Tab. 1. Some of the results are not directly comparable with the proposed method due to the differences in experiment settings and the used features. For example, the method in [8] use the sequential distribution of the STIPs in a video as the basic features for the classifier, and Liu's method [9] uses an improved STIP clustering method and incorporates relative spatial-temporal distribution information of the interest points in forming the video features. These additional information (modifications) are not used in our method. Nevertheless, we still achieve a comparable result.

3.1.1. Weak classifiers' importance variations for different data classes

We check the variations of the importance of the tree classifiers in performing prediction for different action classes using our method. The importance $W(h_i)_y$ of a weak classifier h_i for an action class y is computed as the frequency of h_i's

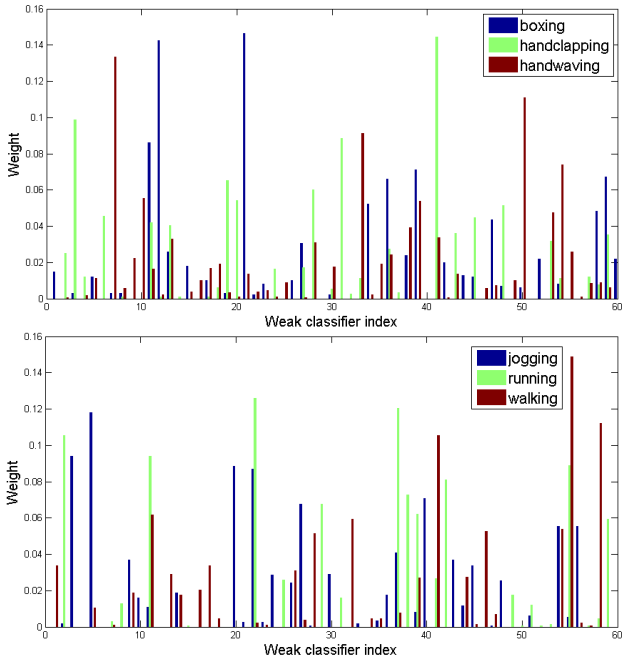

Fig. 3. Illustration of the weak classifiers' importance variations when making prediction for different data classes. The importance of a weak classifier for a data class is computed according to the frequency of its partitions participating in forming the decision rules for this class.

outputs participating in forming the decision rules of action class y

$$W(h_i)_y = \frac{\|\{\gamma_{yj}|h_i \in \gamma_{yj},\ j = 1, ..., n_y\}\|}{n_y} \quad (5)$$

where n_y is the decision rule number of action class y, and $\{\gamma_{yj}\}_{j=1}^{n_y}$ is class y's decision rule set. The importance distributions of the weak classifiers for different action classes are plotted in Fig. 3. As can be seen, the weak classifiers' importance distributions vary drastically between different action classes, which suggests that the discriminative power of a weak classifier for different data classes are really different. The proposed EP Random Forest classifier accounts for this discriminative power variations by mining decision rules for each data class individually.

4. CONCLUSION

In this paper, we proposed a new methodology to integrate the tree classifiers' classification votes to make the final decision for the random forest classifier. The method discovers decision rules by checking the statistical properties exhibited in the outputs of the tree classifiers on training data. The result classifier is named EP Random Forest since the decision rules are discovered using the emerging pattern (EP) mining algorithm. Comparing with the traditional Random Forest classifier, two improvements are made: first, instead of make classification prediction by each tree classifier independently, we discover the tree classifiers composition to form decision rules; second, we account for the weak classifiers' discriminative power variations for different data classes by mining decision rules for each data class individually.

5. REFERENCES

[1] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1632–1646, 2008.

[2] Jamie Shotton, Matthew Johnson, and Roberto Cipolla, "Semantic texton forests for image categorization and segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[3] J. Yuan, J. Luo, and Y. Wu, "Mining compositional features for boosting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[4] G. Dong and J. Li, "Efficient mining of emerging patterns: discovering trends and differences," in *Proc. ACM Int'l Conf. Knowledge discovery and data mining*, 1999, pp. 43–52.

[5] G. Dong, X. Zhang, L. Wong, and J. Li, "Caep: Classification by aggregating emerging patterns," *Discovery Science*, vol. 1721, pp. 737–747, 1999.

[6] T. Lindeberg, A. Akbarzadeh, and I. Laptev, "Recognizing human actions: A local svm approach," in *Proc. Int'l Conf. Pattern Recognition*, 2004, pp. 32–36.

[7] P. Dóllar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. IEEE Int'l Workshop on PETS*, 2005, pp. 65–72.

[8] S. Nowozin, G. Bakir, and K. Tsuda, "Discriminative subsequence mining for action recognition," in *Proc. Int'l Conf. Computer Vision*, 2007, pp. 1–8.

[9] J. Liu and M. Shah, "Learning human actions via information maximization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[10] S. F. Wong and R. Cipolla, "Extracting spatiotemporal interest points using global information," in *Proc. IEEE Int'l Conf. Computer Vision*, 2007, pp. 1–8.

[11] J. C. Niebles, H. Wang, and L. Fei-fei, "Unsupervised learning of human action categories using spatial-temporal words," *Int'l J. Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.