## **Multi-class Boosting for Early Classification of Sequences**

Katsuhiko Ishiguro ishiguro@cslab.kecl.ntt.co.jp Hiroshi Sawada sawada@cslab.kecl.ntt.co.jp Hitoshi Sakano keen@cslab.kecl.ntt.co.jp NTT Communication Science Laboratories NTT Corporation Kyoto, 610-0237, Japan

Consider the problem of driver behavior recognition from images captured by a camera installed in a vehicle [4]. Recognition of driver behavior is crucial for driver assistance systems that make driving comfortable and safe. One notable requirement for real applications is that we would like to predict and classify a behavior **as quickly as possible**: if we detect a sign of dangerous movements such as mobile phone use while driving, we would like to warn the driver quickly before the behavior causes any accidents. This kind of classification task is called **"early classification (recognition),"** and is important for many practical problems including on-line handwritten character recognition, and speech recognition systems.

In this paper, we focus one of the most famous discriminative models, i.e. Adaboost [1, 2], and extend it for early classification of sequences. While existing researches (e.g. [5, 6]) have studied only a binary classification problem, we present a multi-class extension of Adaboost for early classification, called Earlyboost.MH (Fig. 1). In this paper, we propose an efficient multi-class Adaboost for early classification by combining multi-class Adaboost.MH [3] and the early classification Boosting (Earlyboost [6]),

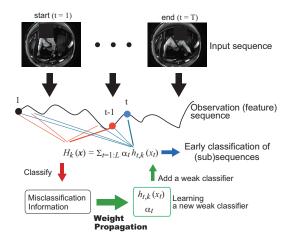


Figure 1: Overview of a concept of the multi-class early classification boosting. Final strong classifiers consists of time frame-wise weak classifiers. The weak classifiers are learnt through weight propagation technique to achieve early classification of (sub)sequences.

The training data consists of the ith sequence  $x_i = \{x_{i,t} \in \mathbb{R}^d\}$  and its class label  $y_i \in \{1,2,\ldots,K\}$ . The number of sequences is N: thus  $i \in \{1,2,\ldots,N\}$ . T is the length of time sequences, and  $t \in \{1,2,\ldots,T\}$  is the time index. A weak classifier  $h_{t,k}(x) : \mathbb{R}^d \to \{1,-1\}$  only accepts the samples on the t-th time frames, and returns 1 if x belongs to class k, and returns -1 otherwise. We also define  $g_k(y) : \{1,2,\ldots,K\} \to \{1,-1\}$  returns 1 if y = k, and returns -1 otherwise.  $H_k^t$  indicates the one vs. all type strong classifier which computes the likelihood of the sequence being the member of class k, and consists of t weak classifiers.

The loss to minimize in Earlyboost.MH is:

$$J(H^{t}) = \sum_{i=1}^{N} \sum_{k=1}^{K} \left( \exp\left(-g_{k}(y_{i})H_{k}^{t}(\boldsymbol{x}_{i})\right) \right) = \sum_{i=1}^{N} \sum_{k=1}^{K} \exp\left(-g_{k}(y_{i})\sum_{s=1}^{t} \alpha_{s}h_{s,k}(x_{i,s})\right).$$
(1)

And we seek for  $\{h_{t,k}, \alpha_t\}$  such that:

$$\alpha_{t}, h_{t,k} = \arg\min_{\alpha, h} \sum_{i=1}^{N} \sum_{k=1}^{K} \exp\left[-g_{k}(y_{i}) \left(H_{k}^{t-1}(\boldsymbol{x}_{i}) + \alpha_{t} h_{t,k}(x_{i,t})\right)\right].$$
 (2)

An optimal  $h_{t,k}$  and  $\alpha_t$  is computed as follows:

$$h_{t,k} = \arg\max_{h} r_{t,k}, \quad \alpha_t = \frac{1}{2} \log \left( \frac{1 + \sum_{k} r_{t,k}}{1 - \sum_{k} r_{t,k}} \right), \tag{3}$$

$$r_{t,k} = \sum_{i=1}^{N} g_k(y_i) h_{t,k}(x_{i,t}) D_t(i,k).$$
 (4)

 $r_{t,k}$  is a class- and frame-wise classification score which only depends on the observation at the t-th frame that will be large if the estimated label by a weak classifier and  $g_k$  match correctly.

These equations imply the benefits of Earlyboost.MH model. First, the optimal  $\alpha_t$  is computed via a sum of K class scores. This implies that the resultant weak classifiers are optimized for multi-class problem, not for K independent binary classifications. Second,  $D_t(i,k) \in \mathbb{R}$  is a weight of a sequence  $x_i$  for the k-th classifier at t-th frame, and computed as follows:

$$D_t(i,k) \propto D_{t-1}(i,k) \exp\left(-\alpha_{t-1}g_k(y_i)h_{t-1,k}(x_{i,t-1})\right).$$
 (5)

Eq. (5) implies that **each weak classifier**  $h_{t,k}$  **learns the classification boundary at time** t **to minimize the classification error induced by the information up to time** t-1. This interpretation of the weight  $D_t$  is first devised by Earlyboost [6] by using frame-wise weak classifiers h. Because of this "weight propagation" update rule, the Earlyboost.MH classifier will be good for early classification of sequences.

Our Earlyboost.MH is validated using two datasets: the online handwritten digits trajectory data, and the driver behavior data (Fig. 2). Experimental results showed the effectiveness of the proposed model in multiclass early classification of sequence data.

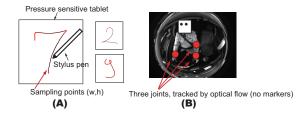


Figure 2: Dataset used in the experiments. (A) On-line handwritten digits data. Trajectories are collected through a pressure sensitive tablet and a wireless stylus pen. (B) Driver behavior data. Seven subjects are recorded their driving simulations by a consumer video camera. Three joints are tracked by optical flow, without any markers or special attachments.

- [1] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computing Systems and Science*, 55(1):119–139, 1997.
- [2] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting (with discussion). *Annals of Statistics*, 28(2):337–407, 2000.
- [3] R. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.
- [4] Y. A. Sheikh, A. Datta, and T. Kanade. On the sustained tracking of human motion. In *Proc. FG*, 2008.
- [5] J. Sochman and J. Matas. Waldboost learning for time constrained sequential detection. In *Proc. CVPR*, volume 2, pages 150–156, 2005.
- [6] S. Uchida and K. Amamoto. Early recognition of sequential patterns by classifier combination. In *Proc. ICPR*, 2008.