

---

# Saving Evaluation Time for the Decision Function in Boosting: Representation and Reordering Base Learner (Supplement)

---

**Peng Sun**

SUNP08@MAILS.TSINGHUA.EDU.EDU

Tsinghua National Laboratory for Information Science and Technology(TNList), Department of Automation, Tsinghua University, Beijing 100084, China

**Jie Zhou**

JZHOU@TSINGHUA.EDU.CN

Tsinghua National Laboratory for Information Science and Technology(TNList), Department of Automation, Tsinghua University, Beijing 100084, China

## 1. Related Work

### 1.1. Imbalanced Class Prior

In the so-called detection problem, the number of positive examples is dominated by the number of negative examples in testing stage. However, positive and negative examples are made equal size in training stage to design a reasonable classifier. In this case, only negative training examples are used for the portion  $P_t$  in EPL at each level.

In the general case when class prior ratio in training stage differs from that in testing stage, one needs reweigh the positive and negative examples when calculating the portion  $P_t$ :

$$P_t = pr(P_t|y = +1) \times w_1 + pr(P_t|y = -1) \times w_2$$

instead of the original total probability theorem:

$$P_t = pr(P_t|y = +1)pr(y = +1) \\ + pr(P_t|y = -1)pr(y = -1)$$

### 1.2. BDD and Threshold Function

BDD is a data structure for general Boolean Expression (BE) (Bryant, 1986). The BE with special structure as explained in main paper was firstly studied in (Hosaka et al., 1994) under the name Threshold Function. BDD construction for Threshold Function were studied by researchers from VLSI CAD tool (Hosaka et al., 1994) and Integer Programming (Behle, 2010; Mayer-Eichberger, 2008). The interval division techniques for BDD construction, as explained in main paper, is credited to (Behle, 2010).

### 1.3. FastExit

FastExit (with weight sorting) is a by-product in (Kim et al., 2012), based on the idea of (Sochman & Matas,

2005). Interestingly, the same technique seemed to be found independently by researchers from Information Retrieval community (Broder et al., 2003).

## 2. Datasets in Experiments

We carry out experiments on several public datasets, as follows:

#1: optdigits_bin05	#2: pendigits80
#3: optdigits17	#4: pendigits49
#5: mnist10k06	#6: mnist10k37
#7: mnist10k49	#8: zipcode38
#9: covertype145k06	#10: poker25kT109

They are adopted from UCI and variations of MNIST, e.g., mnist10k means we exchange the training and testing data in original MNIST(60k training data, 10k testing data) such that the training data are now 10k. Since we are focusing on binary classification, two classes are selected from each of those multi-class datasets, e.g., mnist10k06 means that it is from class 0 and class 6 of mnist10k. We try to incorporate various datasets. The size in our adopted datasets ranges from several thousands to hundreds of thousands, while the number of features ranges from several tens to several hundreds. Some of the datasets are noisy (relatively high testing error), while the others are separable (0 testing error).

## 3. More Results for DBPSYN

In Figure 1 we give the Pareto fronts on all datasets for DBPSYN with three reordering methods: SORT, TD and DBPSYN.

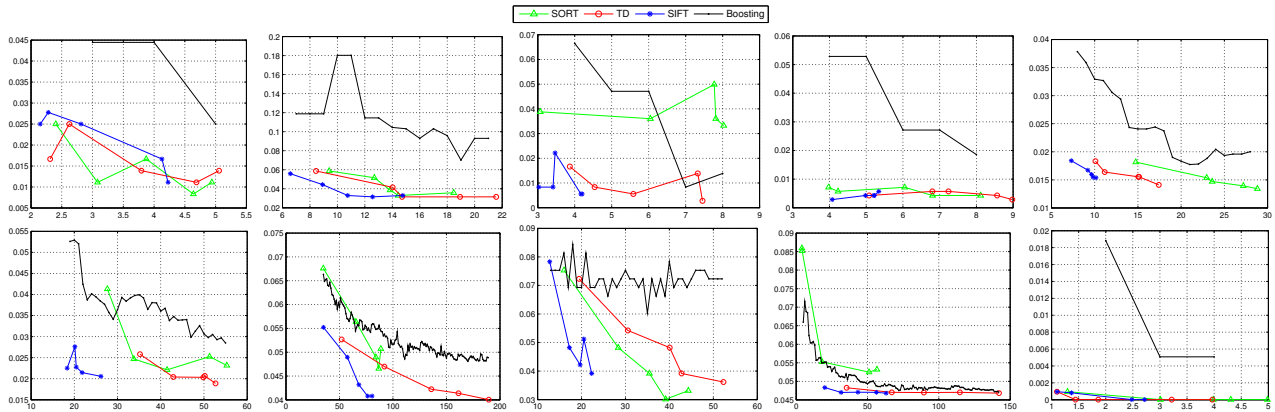


Figure 1. EPL (horizontal axis) v.s. test error (vertical axis) curve for all datasets. The first row: #1-#5. The second row: #6-#10.

## References

Behle, Markus. *Binary decision diagrams and integer programming*. PhD thesis, der Universität des Saarlandes, 2010.

Broder, Andrei Z, Carmel, David, Herscovici, Michael, Soffer, Aya, and Zien, Jason. Efficient query evaluation using a two-level retrieval process. In *Proceedings of the twelfth international conference on Information and knowledge management*, pp. 426–434. ACM, 2003.

Bryant, Randal E. Graph-based algorithms for boolean function manipulation. *Computers, IEEE Transactions on*, 100(8):677–691, 1986.

Hosaka, Kazuhisa, Takenaga, Yasuhiko, and Yajima, Shuzo. On the size of ordered binary decision diagrams representing threshold functions. *Algorithms and Computation*, pp. 584–592, 1994.

Kim, Tae-Kyun, Budvytis, Ignas, and Cipolla, Roberto. Making a shallow network deep: Conversion of a boosting classifier into a decision tree by boolean optimisation. *International Journal of Computer Vision*, pp. 1–13, 2012.

Mayer-Eichberger, Valentin Christian Johannes Kaspar. *Towards solving a system of pseudo boolean constraints with binary decision diagrams*. PhD thesis, Universidade Nova de Lisboa, 2008.

Sochman, Jan and Matas, Jiri. Waldboost-learning for time constrained sequential detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pp. 150–156. IEEE, 2005.