Efficient Implementation of SVM for Large Class Problems

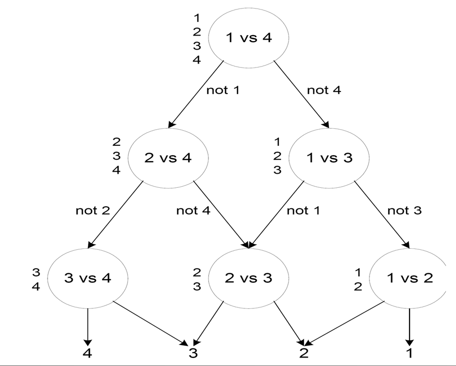
P. Ilayaraja, Neeba N.V. and C.V. Jawahar

How to scale SVM solutions for large class classification, in terms of their space and computational requirements?

• SVMs are popular and accurate binary classifiers with high generalization capability

• Direct multiclass extension of SVM is not attractive

• Binary pair-wise classifiers combined using DAG or BHC architectures are generally used



• As the number of pair-wise classifiers increases, the space and computational needs becomes an overhead

• For large classes, the solutions are not scalable.

Complexities

• Complexity of SVM is proportional to number support vectors in the solution

$$f(x) = \sum_{i=1}^{r} \alpha_i y_i K(x, s_i) + b.$$

How to reduce the complexities without compromising on the accuracy?

• An effective and easy to implement data structure for efficiently storing SVs

• An algebraic method for simplifying hierarchical SVM solutions *exactly.*

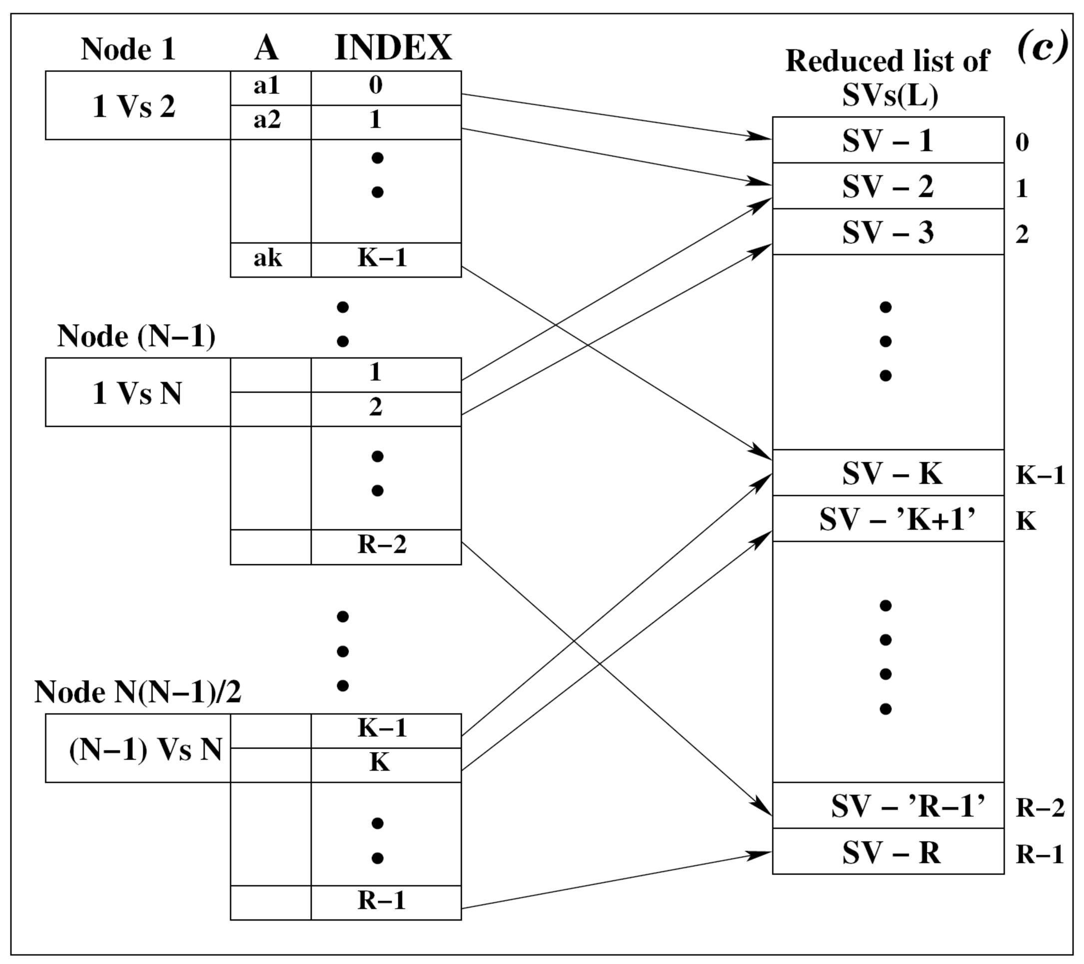
Properties of proposed MDS

• Breaks the independence assumption – SVs are samples on class boundaries

• Exploits the redundancies in SVs across the pair-wise classifiers.

• As number of classes increases, the redundancy also increase.

MDS: Multiclass Data Structure



The kernel computation for a SV once computed is *reused* in computing at other nodes.

Algebraic Exact Simplification

• *Step 1: Multi-class extension*

- Apply *exact simplification* proposed by *T. Downs et al. in JMLR, 2001* to each node independently
- Each node is reduced to a set of linearly independent support vectors
- Can we reduce the number of support vectors in each set further?

• *Step 2 : Hierarchical Exact Simplification (HES)*

- Union of two linearly independent sets need not be independent
- Add SVs from nodes that are above in a decision path to the one below
- Reduce the obtained extended set by exact simplification method
- Apply *HES* along each decision path independently

Experiments and Results

With the use of MDS, we achieved 98.5% of reduction in SVs and 60% reduction in classification time using linear and polynomial kernels on a 300-class data set in comparison to a naïve implementation.

Data set Name	Kernel Type	No. of SVs	
		IPI(S)	MDS(R)
PenDigits (10-class)	Linear	5788	2771
	Poly.	3528	1777
	RBF	67450	7494
Letters (26-class)	Linear	113249	15198
	Poly.	80553	12961
	RBF	482975	18666

MDS Vs IPI on UCI datasets

Dataset (# Class)	#Dim.	Reduction(%)		
		Step 1	*Step* 2	Overall
PenDigits (10)	16	85.42	71.49	95.84
Letters (26)	16	94.87	17.78	95.60
OptDigits(10)	64	59.25	54.92	81.63
Vowel(11)	10	76.89	68.90	92.81

HES Results

The utility of MDS, increases with the size of the problem.

With the use of HES the time complexity of multiclass problems can be reduced considerably.

Future Direction

• *Exploring a method to simplify the set of unique support vectors in the master list and compare that against HES.*

Center for Visual Information Technology
International Institute of Information Technology Hyderabad
http://cvit.iiit.ac.in