

オンライン SVM に関する調査報告
(2010/6/17 株式会社数理システム 知識工学部 阿部裕介)

1. 概要と目的

オンライン学習が可能な SVM の可能性について、現在もっとも標準的な SVM 実装である LIBSVM、学習時間の大幅な削減に成功した BVM、NEC の研究者の手によるオンライン学習可能な LASVM の三者を比較することで探ることを目的とする。

2. SVM アルゴリズム

- ・ LIBSVM...台湾の研究者の手による、SMO アルゴリズムを用いたスタンダードな SVM ライブラリ(<http://www.csie.ntu.edu.tw/~cjlin/libsvm>)

- ・ BVM...Core Vector Machine の発展形で、minimum enclosing ball (MEB)に基づいて、巨大なデータセットでの学習が高速に行えるように工夫されている(<http://www.cse.ust.hk/~ivor/cvm.html>)

- ・ LASVM...オンライン学習と能動学習を用いた SVM(<http://leon.bottou.org/projects/lasvm>)

3. データセットを用いた比較結果

data	LIBSVM			BVM			LASVM		
	#SV	Acc.	Time	#SV	Acc.	Time	#SV	Acc.	Time
Web	3515	99.32	119	1084	97.93	7	7187	91.44	308
IJCNN	3181	98.99	47	1095	97.63	10	3014	98.89	124
Zero-One	1471	99.53	1198	648	99.45	30	1371	99.55	741
Forest	91868	98.24	127992	-	-	-	-	-	-
KDD-CUP	1665	94.2	9277	125	91.99	2	-	-	-

(DucDung Nguyen et al. Condensed Vector Machines: Learning Fast Machine For Large Data より)

4. 考察

オンライン学習可能な LASVM では Forest および KDD-CUP の各データセットで学習が終了していない。また、Web データセットでは LIBSVM の約 2 倍のサポートベクターを要している。これは判別タスク時により時間がかかることを意味している。判別精度も Zero-One データセットを除いて、LIBSVM より低い。

5. 結論

日本におけるカーネル法の第一人者である赤穂昭太郎氏の 2006 年発表の資料（カーネル法の最前線）によれば、「SVM でもオンライン学習の方法はいくつか提案されているが、スパースネスを保ちつつ軽い計算量で行うのが困難」とある。上の結果はそれを裏づけており、2010 年のいまでも状況はそれほど変化していない。オンライン化の代償として、スパースネス・精度の低下が見合うかどうか、微妙なところである。上の結果を見るかぎりでは、若干の精度低下を受け入れる代わりに、学習の高速化とスパースネスを実現している BVM が健闘している様子が伺える。結論としては、オンライン SVM の実務使用はまだ時期尚早であり、もっと安定して枯れた理論と実装の出現を待つべきと思われる。

最後に、オンライン学習ではない SVM 学習高速化の方向としては、Keerthi らのグループが Parallel SMO を提案しており、32 コアマシン使用時に約 21 倍の学習高速化が達成されたという報告がなされている。

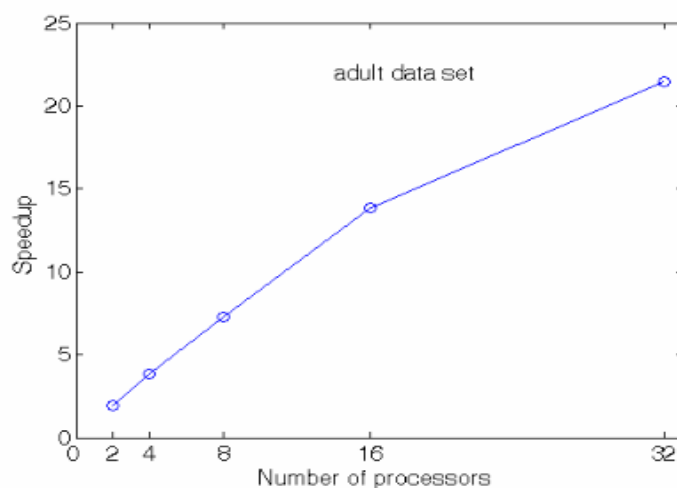


Fig. 1. The speedup of the parallel SMO on the adult data set.

(Cao and Keerthi et al. Parallel Sequential Minimal Optimization for the Training of Support Vector Machines より)