Direct comparison between support vector machine and multinomial naive Bayes algorithms for medical abstract classification

In 2011 Matwin *et al* published a letter to *JAMIA* entitled 'Performance of SVM and Bayesian classifiers on the systematic review classification task'.[1] This letter continued a discussion on the relative benefits of using support vector machine (SVM) and Bayesian techniques for performing systematic reviews.[2–4] In particular, it was suggested that the running time of algorithms must be taken into consideration when comparing their performances as it becomes very important for large datasets.

Following up on this idea, we attempted to directly compare the performance of a Bayesian method with the SVM algorithm used by Cohen in his original work.[4] The same SVM system (SVMlight[5]) with the same parameters as was described in the original paper[4] was tested versus the multinomial naive Bayes algorithm (MNB) implemented in WEKA.[6] The data were separated into features and processed following as closely as possible our understanding of the procedure used by Cohen except for the size of the feature vector dictionary which was set to 2500 words per features type. While this option did not affect the result of the SVM algorithm, it was very important to the performance of the MNB. We ran five trials of twofold cross-validation, and both algorithms were subjected to exactly the same sequence of splits.

This procedure allowed us to run statistical significant tests between the differences in the MNB and SVM performances. Columns three and four of table 1 show the area under the curve (AUC) values for the highest ranked feature system from Cohen[4] (unigram features based on abstract and title, MeSH-based features and 1,2-grams features based on abstract and title) and column one shows the percent difference in their averages. The parenthesized entries in the table indicate differences that are not statistically significant according to the Student t test at 95% confidence. Column two shows the ratios of the average cpu time of SVMlight[5] to the average cpu time of the MNB classifier. Finally, in the last column we reprint the results from Cohen.[4]

As we can see, not only does the MNB run two to six times faster than the SVM, but for most datasets (nine out of 13) it also performs equally well or better. Figure 1 shows the algorithm execution time versus the number of examples in the dataset for both MNB and SVM. We can clearly see that the MNB executes faster than SVM and, even more importantly, that this difference increases with the dataset size.

Other feature systems showed generally similar results; however, the magnitude and even the sign of the difference changed from one feature system to another for the same dataset.

We want to emphasize that we ran SVM experiments with all data and algorithm settings as close as possible to those of Cohen, but due to some nuances in the data pre-processing procedures we were unable to reproduce his exact results on some datasets.

Nonetheless, we would like to conclude by pointing out that for medical abstract classification using MNB as opposed to SVM appears to be appreciably faster without a significant loss in performance. This might prove essential when working with large datasets.[7]
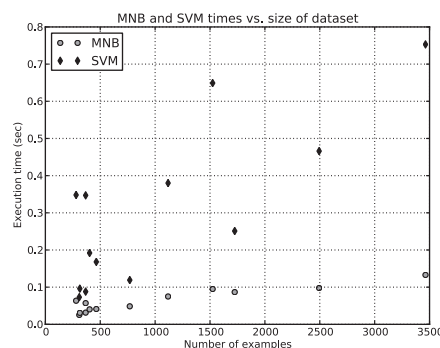
Table 1 Area under the curve comparison for ABTITLE+MESH+NGRAMS12 (abstract and title, MeSH terms and 1,2-ngrams built from abstract and title)

	% Difference (MNB−SVM)	t_{SVM}/t_{MNB}	MNB	SVM	Cohen
ACE inhibitors	4.67	4.8	0.8570	0.8169	0.9462
ADHD	2.10	2.5	0.9576	0.9375	0.9244
Antihistamines	−12.67	2.9	0.6776	0.7635	0.7228
Atypical antipsychotics	5.41	6.1	0.8286	0.7837	0.8180
β Blocker	7.19	6.8	0.8413	0.7808	0.8911
Calcium channel blockers	(1.41)	5.5	0.8408	0.8920	0.8738
Estrogens	−13.60	2.8	0.7636	0.8674	0.8876
Opioids	14.71	2.9	0.9349	0.7994	0.8972
Oral hypoglycemics	−11.66	4.1	0.6857	0.7657	0.7815
Proton pump inhibitors	(−0.83)	5.1	0.7798	0.7682	0.8600
Statins	14.24	5.7	0.9057	0.7767	0.9007
Triptan	(−0.44)	4.8	0.8602	0.8640	0.9097
Urinary incontinence	3.90	3.1	0.8353	0.8028	0.8915

ADHD, attention deficit hyperactivity disorder; MNB, multinomial naive Bayes algorithm; SVM, support vector machine algorithm.

Figure 1 Execution time versus dataset size for multinomial naive Bayes (MNB) and support vector machine (SVM) algorithms.

Stan Matwin, Vera Sazonova

School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada

Correspondence to Dr Stan Matwin, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada; stan@site.uottawa.ca

REFERENCES

1. **Matwin S,** Kouznetsov A, Inkpen D, et al. Performance of SVM and Bayesian classifiers on the systematic review classification task. *J Am Med Inform Assoc* 2011;**18**:104–5.
2. **Cohen A.** Performance of support-vector-machine-based classification on 15 systematic review topics evaluated with the WSS@95 measure. *J Am Med Inform Assoc* 2011;**18**:104–5.
3. **Matwin S,** Kouznetsov A, Inkpen D, et al. A new algorithm for reducing the workload of experts in performing systematic reviews. *J Am Med Inform Assoc* 2010;**17**:446–53.
4. **Cohen A.** Optimizing feature representation for automated systematic review work prioritization. *AMIA Annu Symp Proc* 2008:121–5.
5. **Joachims T.** Text categorization with support vector machines: learning with many relevant features. *Proceedings of the 10th European Conference on Machine Learning; 1998*. London: Springer-Verlag, 1998:137–42.
6. **Hall M,** Frank E, Holmes G, et al. The WEKA Data Mining Software: An Update. **Vol. 11**. SIGKDD Explorations, 2009, Issue 1.
7. **Frunza O,** Inkpen D, Matwin S, et al. Exploiting the systematic review protocol for classification of medical abstracts. *J Artifi Intell Med* 2011;**51**:17–25.