Suppliment 1 to CLRC—TR—08—01
UKOPS: supplementary results 1

Dmitry Devetyarov, Ilia Nouretdinov, Brian Burford, Zhiyuan Luo,
Alexey Chervonenkis, Volodya Vovk, Alex Gammerman
Royal Holloway, University of London

November 05, 2008

Abstract

This report covers the results of analysis of UKOPS data (both Reading and UCL data sets) with Quality weighted towards specificity rather than sensitivity as it was done before. The experiments showed that sometimes we selected rules different from the ones selected when weighting towards sensitivity. But all of these new rules are inferior to the CA125 cut-off rule on the test set.

In addition, the analysis of Benign vs Malignant discrimination for samples with CA125 levels greater than 30 was carried out. For Reading data, we identified peak 52 (m/z = 3507.3 Da), that often appears in selected models for a single peak, 2 peaks, 2 peak + CA and provides stable performance on the test set. For UCL data, we could find peak 28 (m/z = 3450.3 Da), that often appears in selected models for a peak + CA, 2 peaks, 2 peak + CA. However, this peak does not provide stable results on the test set. Additional analysis of samples with CA between 30 and 600 Da proved to be similar to the results for CA greater than 30 Da.

1 Analysis with Quality weighted towards Specificity

We carried out the same analysis for both Reading and UCL data sets when considering

$$\text{Quality} = \frac{1}{3} \times \text{Sensitivity} + \frac{2}{3} \times \text{Specificity},$$

that is, weighting towards Specificity. The same types of discrimination were considered:

- Healthy vs Malignant and Borderlines;
- Healthy vs Malignant;
- Benign vs Malignant;
- Healthy + Benign + Borderlines vs Malignant.

The selected rules are represented in the attached file 'UKOPS weights 1 2.xls'. In summary, we sometimes selected rules different from the ones selected when weighting towards Sensitivity. But all of these 'new' rules are inferior to the CA125 cut-off rule on the test set.

For example, for Healthy vs Malignant (with or without Borderlines) on Reading data we select the same models 1 and 2, one of which matches the cut-off model, another - outperforms the cut-off model with respect to performance on the test set. In addition, one new model was selected, but the model is outperformed by the cut-off model as well as by models 1 and 2.

2 The analysis of Benign vs Malignant for samples with CA > 30

We considered the models consisting of 1 peak, 2 peaks, 1 peak and CA, 2 peaks and CA for Benign vs Malignant discrimination, for samples with CA > 30 and Quality = Accuracy. Detailed results are represented in the attachment 'UKOPS BvM Accuracy AC gr than 30.xls'.

In summary, for Reading data, the peak of high importance seems to be peak 52 (m/z = 3507.3 Da), that often appears in selected models for a single peak, 2 peaks, 2 peak + CA (but not as a single peak + CA) and provides stable performance on the test set (that is, performance close to the performance demonstrated on the training set). Table 1 shows performance of peak 52 as a single peak.

For UCL data, we could find peak 28 (m/z = 3450.3 Da), that often appears in selected models for a peak + CA, 2 peaks, 2 peak + CA (but not as a single peak). However, this peak does not provide stable results on the test set.

The distribution of Reading data samples on CA vs peak 52 plot can be seen in Figure 1. The plot shows that there are only two malignant samples with CA ≤ 30, and all the samples with CA ≥ 600 are malignant. For this reason, it seemed sensible to carry out Benign vs Malignant classification for this data only for samples with $30 <$ CA < 600.

The attachment 'UKOPS BvM Accuracy AC between 30 and 600.xls' with the results shows that when we apply weighted nearest neighbours, the results for CA between 30 and 600 Da are similar to the results for CA greater than 30 Da. We can see stable performance only when considering one peak without CA125, mostly with peak 52 selected. By stable performance we mean the performance on the test set which is almost as good as the performance on the training set. Stable selected rule are also demonstrated in Table 1.

The distribution plot also demonstrates that it may be reasonable to apply SVM with linear kernel (as Benign and Malignant samples seem to be more or less separable by the plane on CA vs peak 52 plot). The experiments carried out for 1 peak and 1 peak + CA125 showed that we always select the same peak 52. The performance on the training set is similar to performance for kNN, but it deteriorates on the test set. Thus, kNN seem to provide better results than simple separation by the plane.

Rule	Training set			Test set		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
CA > 30						
2 out of 5 NN	97.2%	50.0%	79.3%	86.2%	44.4%	76.3%
4 out of 8 NN	88.9%	59.1%	77.6%	82.8%	44.4%	73.7%
$30 <$ CA < 600						
1 out of 4 NN	100.0%	50.0%	75.0%	92.9%	44.4%	73.9%
2 out of 5 NN	100.0%	50.0%	75.0%	92.9%	44.4%	73.9%
SVM, linear kernel	77.3%	72.7%	75.0%	71.4%	44.4%	60.9%

Table 1: Stable selected rules for Reading data, Benign vs Malignant, peak 52 (m/z-value = 3507.3 Da) without CA-125

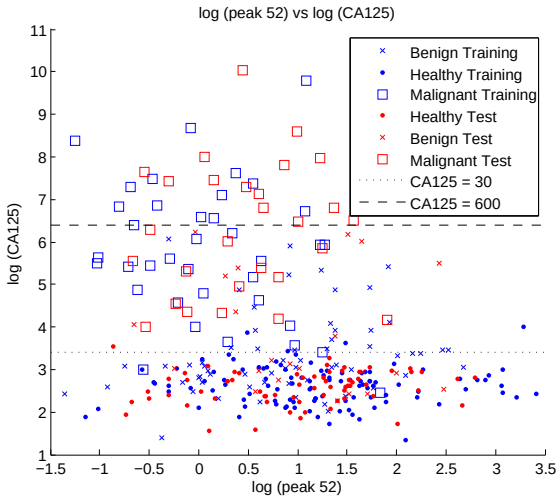

Figure 1: CA125 vs peak 52 plot for Reading data