LETTER

Benchmark Performance of MultiCASE Inc. Software in Ames Mutagenicity Set

Roustem D. Saiakhov* and Gilles Klopman

MultiCASE Incorporated, 23811 Chagrin Boulevard, Suite 305, Beachwood, Ohio 44122

Received March 4, 2010

Abstract: The predictive performances of MC4PC were evaluated using its learning machine functionality. Its superior characteristics are demonstrated in this following up study using the newly published Ames mutagenicity benchmark set.

To the editor: We are writing to you in relation to a paper published by Hansen, K. et al. *J. Chem. Inf. Model.* **2009**, *49* (9), pp 2077−2081,[1] which describes a comparison between a few commercial and noncommercial QSAR modeling tools. MultiCASE MC4PC was among them. Though the paper shed an interesting insight into the advantages and disadvantages of various methodologies, we found certain shortcomings in that study.

The authors used the same combination of training and testing sets to compare the predictive performance of various modules, retraining each model that they describe, with the exception of MC4PC and DEREK. However, MC4PC is a learning machine. Therefore, the logical procedure to follow for the evaluation of the performance of MC4PC is to use the same training sets as were used by the authors for the other methods to train the MC4PC system and to use the same external test sets to validate it. Unfortunately the authors chose to use one of the MultiCASE's pretrained expert systems based on a different learning set (AZ2) instead. To compare the predictive performance of the learning systems trained with the different learning sets is definitely not scientifically valid. We therefore decided to correct this shortcoming by using the technique described above.

The experiment was done in two stages. In the first stage, the entire learning set of 6512 structures, available from the American Chemical Society (ACS) Web site[2] was submitted to the MC4PC module builder. The module was compiled, producing 272 structural alerts for mutagenicity, and compared to 224 alerts derived using the Multicase AZ2 system, used by Halsen et al. We validated this expert system, using the leave-group-out approach with 10% of the learning set randomly selected and removed from the training, the remaining learning set retrained, and the removed molecules used as an external test set. The process has been repeated 10 times. Our results are presented in the Table 1 below.

It is obvious, that the learning set produced a very solid performing expert system.

In the second stage of our evaluation, we actually followed very closely the validation routine used by Hansen et al. We used the learning abilities of MC4PC to train the learning sets, as described in the paper, and used exactly the same test sets to validate them. The results are presented in the Table 2.

Table 1. Results of a Leave-Group-out Cross-Validation Experiment[a]

test *N*	concordance %	sensitivity %	specificity %	χ^2	*N* submit	not tested %
1	81.1	82.3	79.9	235	609	6.16
2	78.7	81.0	76.0	198	609	6.16
3	79.9	81.7	77.9	217	612	5.70
4	78.7	78.7	78.6	199	609	6.16
5	79.2	80.4	77.7	207	616	5.08
6	75.6	77.6	73.4	155	599	7.70
7	81.2	82.3	79.7	225	589	9.24
8	81.1	82.1	80.0	233	607	6.47
9	75.4	74.6	76.2	156	609	6.16
10	79.4	77.1	81.9	209	602	7.24
average	**79.0**	**79.8**	**78.1**	**203**	**606**	**6.61**

[a] Concordance, sensitivity, specificity are commonly used statistical parameters; χ^2 is χ squared; *n* submit is the number of chemicals randomly removed from the learning set and submitted later as a test set; and no test is the percentage of inconclusive test results.

Table 2. Results of the External Validation Test[a]

	concordance %	sensitivity %	specificity %	χ^2	coverage %
test 1	78.8	85.6	67.4	225	78.74
test 2	80.5	85.5	72.8	273	80.55
test 3	77.9	81.0	73.2	235	81.89
test 4	78.8	85.2	69.2	250	82.54
test 5	79.7	83.6	74.0	267	81.03
average	**79.1**	**84.2**	**73.2**	**250**	**80.95**

[a] Concordance, sensitivity, specificity are commonly used statistical parameters; χ^2 is χ squared; and coverage is the percentage of conclusive test results.

It is quite obvious, that in contrast to the conclusions drawn by Hansen et al., MC4PC performed very well in this experiment, and the reported predictivity statistical values are 6−15% better than those reported previously[1] (78% sensitivity and 57% specificity). We find that the Hansen et al.'s data is an excellent compilation of Ames assay data, and its availability helped to demonstrate the superior characteristics of MC4PC expert building system with this data, which we will gladly make available to the users of our methodology.

REFERENCES AND NOTES

(1) Hansen, K.; Mika, S.; Schroeter, T.; Sutter, A.; Laak, A.; Steger-Hartmann, T.; Heinrich, N.; Muller, K. Benchmark Data Set for In Silico Prediction of Ames Mutagenicity. *J. Chem. Inf. Model.* **2009**, *49*, 2077–2081.

(2) Hansen, K.; Mika, S.; Schroeter, T.; Sutter, A.; Laak, A.; Steger-Hartmann, T.; Heinrich, N.; Muller, K. Benchmark Data Set for In Silico Prediction of Ames Mutagenicity, Supporting Information. *J. Chem. Inf. Model.* **2009**, *49*, pp2077−2081; http://pubs.acs.org/doi/suppl/10.1021/ci900161g/suppl_file/ci900161g_si_001.zip. Accessed December 1, 2009.

* Corresponding author. E-mail: saiakhov@multicase.com.

CI1000899

10.1021/ci1000899
Published on Web 08/09/2010