

LIBRARY meta-learning

LIBRARY meta-learning technique was originally suggested by Tetko (*Tetko IV et al. Large-scale evaluation of log P predictors: local corrections may compensate insufficient accuracy and need of experimentally testing every other compound. Chem. Biodivers. 2009 Nov;6(11):1837-1844.*). The following excerpt was taken from a doctoral work by Sushko (Sushko I., Applicability Domain of QSAR models. Doctoral work. 2011. <http://mediatum.ub.tum.de/node?id=1004002>)

Sushko I, Applicability Domain of QSAR models. Excerpt

Given a predictive model and an additional set of new experimental measurements (referred to as “library”), it is possible to correct the model by taking into account these measurements. The process is called LIBRARY model correction (because we complement a model with a “library” of experimental measurements) and is performed as follows.

To obtain a corrected prediction for a molecule J, we calculate the original (noncorrected) prediction $\tilde{y}(J)$ given by the original model and find K molecules from the “library” $\{J_i, i = 1..K\}$ nearest to the molecule being predicted. The nearest compounds are defined by the correlation coefficient in space of model predictions.

Then, we calculate the expected residual for the molecule J as the average residual for the K nearest compounds according to the following expression:

$$\Delta(J) = \frac{\sum_{i=1}^K (\tilde{y}(J_i) - y(J_i))}{K}$$

Finally, we correct the original prediction by subtracting the average residual:

$$\tilde{y}_{corr}(J) = \tilde{y}(J) - \Delta(J)$$

Thus, the LIBRARY correction assumes that for a new compound a model will behave similarly to compounds from the “library”. Such technique is especially useful in case if retraining of the original model is infeasible due to high computational complexity or unavailability of the original training set. The LIBRARY technique was introduced by Tetko and was shown to significantly increase the prediction accuracy for the lipophilicity and distribution coefficient models.

<http://mediatum.ub.tum.de/node?id=1004002>