CONDUCTION OF CONFORMAL PREDICTION

Dinh Long Huynh – 000624-8397

GIVEN MODELS (see ML-REPORT)

rf_model = RandomForestRegressor(n_estimators= 184, bootstrap=False, max_depth=13, max_features='log2', min_samples_leaf=5, min_samples_split=4)	error_model = RandomForestRegressor(n_estimators= 184, bootstrap=False, max_depth=13, max_features='log2', min_samples_leaf=5, min_samples_split=4)

GENERAL PROCEDURE

The conformal prediction was developed based on the previous publication by Norinder et al. in 2014[1]. The standardized data was divided into three parts: the training set, the calibration set, and the testing set, in the ratio of 0.5, 0.2, and 0.3, respectively. Two models were built - the rf_model for predicting solubility and the error_model for predicting the absolute error. The training set was used for fitting both models, while the calibration set was used to calibrate the non-conformity scores and to choose the 0.9 quantile of these scores as the cut-off threshold for the interval. Finally, the testing set was used for generating the prediction and prediction interval based on the cut-off threshold and predicted error generated from the error_model.

DETAIL PROCEDURE

Step 1: Split the data into training set, calibrating set, and testing set (0.5: 0.2: 0.3, n/n/n).

Step 2: Conducted on training set
- Fit rf_model to training set.
- Generate predicted solubility for training set, used it for calculating the absolute error (AE):

$$AE = |\hat{y}_i - y_i| \text{ (equation 1)}$$

- Use the training AE as the target value for fitting the error_model.

Step 3: Conducted on calibrating set
- Generate predicted solubility for calibrating set using rf_model, calculate the AE for calibrating set.
- Generate predicted AE for calibrating set using error_model.
- Calculate non-conformity scores (alpha), which is equal to the AE for calibrating ser divided by predicted AE for calibrating set

$$\text{alpha} = \frac{|\hat{y}_i - y_i|}{\widehat{AE}_i} \text{ (equation 2)}$$

- Choose the cut-off threshold being equal to 0.9 quantile of non-conformity scores.

Step 4: Conducted on testing set
- Generate predicted solubility for testing set, using rf_model.
- Generate predicted AE for testing set, using error_model.
- Calculating prediction interval:

$$\text{Prediction interval} = \hat{y}_i \pm \text{cut-off threshold} * \widehat{AE}_i \text{ (equaion 3)}$$

Step 5: Evaluate model

$$\text{Prediction cappacity} = \#\text{successful predictions} / \text{total number of predictions (equation 4)}$$

RESULTS – DISCUSSION

Applying conformal analysis, we assumed that there is an existence of extreme AE, that represents outliers. Because of that, instead of considering the largest non-conformity score, we solely focus on the 0.9 quantile, which divided non-conformity scores into 90% counts on the left and 10% counts on the right.

In conformal prediction, the model predicts a range of values instead of a single point. Therefore, the usual evaluation metrics like RMSE, MAE, and R2 cannot be employed. Instead, the prediction capacity is used to measure the accuracy of

the model. This approach considers whether the actual solubility value falls within the predicted interval. If it does, the prediction is deemed successful, otherwise it is considered a failure. The prediction capacity is calculated by dividing the number of successful predictions by the total number of predictions.

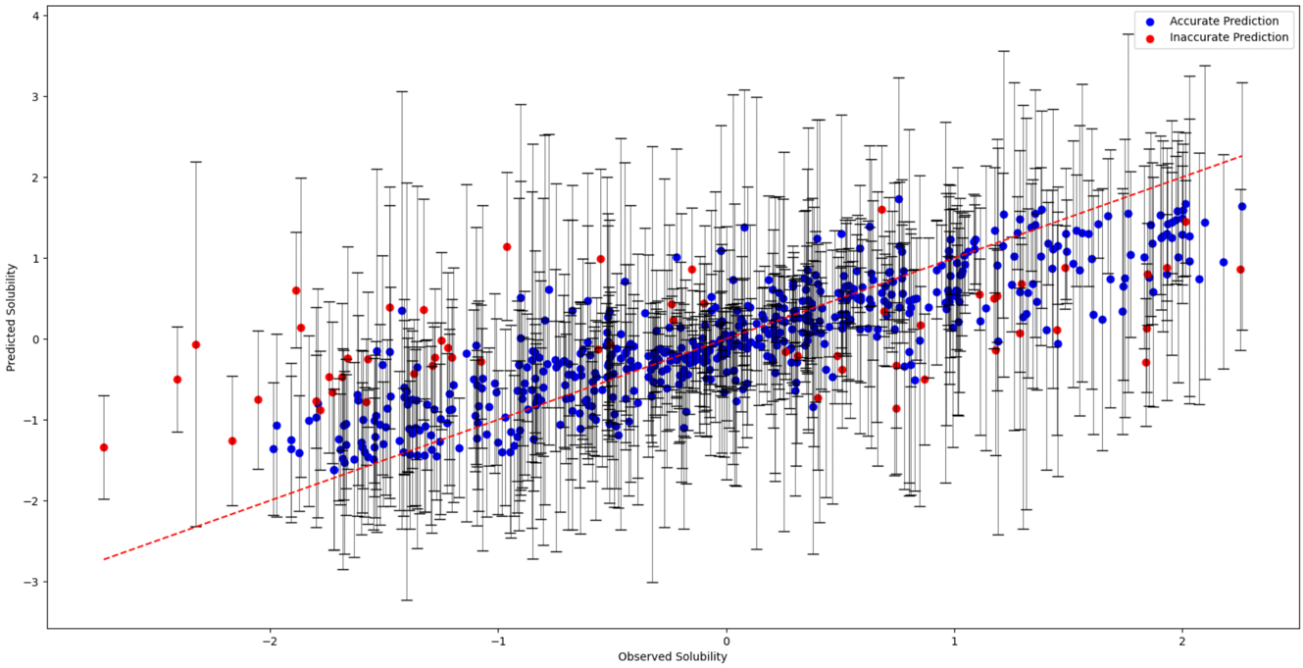

Figure 1. Model performance

After analyzing the non-conformity scores, it was found that the prediction capacity improved to 90.5% when the 0.9 quantile was used. It should be noted that choosing a larger quantile results in a wider prediction interval (equation 3) and a higher prediction capacity. However, this larger interval can be problematic as it does not provide a clear picture of the predicted solubility which can create issues when setting a solubility cut-off threshold for filtering out compounds during follow-up research.

CONCLUSION

With the conformal analysis, the prediction interval was built based on the non-conformity scores' 0.9 quantile and the predicted AE. The prediction capacity improved to 90.5%. The follow up step can be optimize the error_model, using hyperparameter tuning, optimize quantile choice and apply cross-validation to find the 95% CI of prediction capacity.

REFERENCE

1. Norinder, U., Carlsson, L., Boyer, S. & Eklund, M. Introducing Conformal Prediction in Predictive Modeling. A Transparent and Flexible Alternative to Applicability Domain Determination. *J. Chem. Inf. Model.* **54**, 1596–1603 (2014).