



USEDAT: USA-Europe Data Analysis Training School

MOL2NET, International Conference Series on Multidisciplinary Sciences

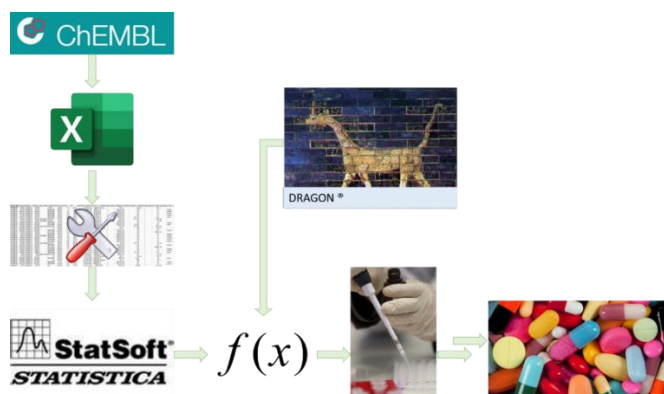


PTML-LDA One-Condition model for the design of new anti-cancer compounds

Harbil Bediaga Bañeres

¹Department of Organic Chemistry II, Faculty of Science and Technology, University of Basque Country (UPV/EHU), 48940, Leioa, Biscay, Spain.

Graphical Abstract



Abstract.

Data from preclinical assays of ChEMBL are obtained. This data is treated mathematically and a PTML-LDA model is obtained in which the conditions are analyzed individually. With this equation and the values of the descriptors of new molecules, the activity of the compound to be tested could be predicted.

Introduction

The models that analyze the Quantitative Structure-Activity Relationship (QSAR) are very useful when designing new compounds. Using the ALogP and PSA descriptors to define the structure of the compounds, we obtained a model based on the perturbation theory (PT) and machine learning (ML). The model is based on the method of classification by linear discriminant analysis (LDA).

Materials and Methods

ChEMBL data preprocessing

We obtained the results of many assay types expressed by the value of the measured activity from ChEMBL. These values were analysed by groups of conditions to get de moving average o every assay. The conditions analyzed are c_0 = activity, c_1 = protein, c_2 = cell line, c_3 = target type, c_4 = assay organism, c_5 = target organism and c_6 = assay type.

PTML-LDA model

The PTML-LDA technique is used to obtain the equation with better results in predicting the activity of new compounds.

Results and Discussion

Using the PTML-LDA One-Condition model, the values of the training set data are calculated in order to obtain the values of $S_n = 72.8$, $S_p = 88.4$ and $A_c = 86.4$. The values obtained with the model that analyzes the conditions individually have worse results when predicting the compounds compared with the Multi-Condition model (Bediaga, 2018).

Conclusions

Although the results of the PTML-LDA One-Condition model has worse results compared to previous multi-condition models, when using the model it may be interesting to be able to select the variables individually without having the assay condition imposed as in the Multi-Condition model.

References

- Bediaga, H., Arrasate, S., & Gonzalez-Diaz, H. (2018). PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Combinatorial Science*, 20(11), 621-632. doi:10.1021/acscmbosci.8b00090
- Gonzalez-Diaz, H., Arrasate, S., Gomez-San Juan, A., Sotomayor, N., Lete, E., Besada-Porto, L., & Ruso, J. M. (2013). General Theory for Multiple Input-Output Perturbations in Complex Molecular Systems. 1. Linear QSPR Electronegativity Models in Physical, Organic and Medicinal Chemistry. *Current Topics in Medicinal Chemistry*, 13, 1713 - 1741. doi:10.2174/1568026611313140011
- Gonzalez-Díaz, H., Arrasate, S., Gomez-San Juan, A., Sotomayor, N., Lete, E., Besada-Porto, L., & Ruso, J. M. (2013). MIANN models in medicinal, physical and organic chemistry. *Current Topics in Medicinal Chemistry*, 13(5), 619-623. doi:10.2174/1568026611313050006
- Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3), 1065-1076. doi:10.1214/aoms/1177704472