



Step by Step from Machine Learning Algorithm to Making QSAR model:

Present a Combined Feature Selection with a Case Study

Mazaher Maghsoudloo¹, Masoud Arabfard*, Sajjad Gharaghani², Kaveh kavousi²

¹University of Tehran, Kish International Campus, Iran

² Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

* Corresponding E-mail Address: arabfard@ut.ac.ir

Abstract: Due to the large amount of data in various bioinformatics sciences, such as drug design, gene selection, QSAR, etc., the use of machine learning techniques such as feature selection has become an essential requirement for the construction of the model .

Feature selection is one of the most important steps in pattern recognition, machine learning and data mining. The purpose of the feature selection is to select the most optimal subset of the feature from the entire space of the main features problem, so that, while reducing the dimensions, the descriptive accuracy of the machine's learning techniques can be achieved. Optimization algorithms, random search, evolution, etc. In selecting features, new and effective methods are used to find optimal solutions for problems. The randomness of these algorithms prevents them from falling into local optimal points. Therefore, the use of feature selection methods in various bioinformatics sciences is inevitable.

In this paper, we have used a combination of classic and random methods to select important descriptors in a set of QSAR data in drug design. In this paper, the combination of Grey number theory methods and firefly algorithm has been used to select the feature, so that we first use the Grey number theory algorithm to rank the features belonging to our database. Then, using some of these features that have the highest priority and rank, as well as the random firefly algorithm, we choose the best feature among these priority characteristics. The combination of these methods suggests acceptable results, which can be used for feature selection algorithms in bioinformatics.

Keywords: QSAR, Feature Selection; Grey Number algorithm; Drug Design; Machine Learning.

References

1. Binitha S, Sathya SS. A survey of bio inspired optimization algorithms. International Journal of Soft Computing and Engineering. 2012;2(2):137-51.
2. Zang H, Zhang S, Hapeshi K. A review of nature-inspired algorithms. Journal of Bionic Engineering. 2010;7:S232-S7.
3. Yang X-S. Nature-inspired metaheuristic algorithms: Luniver press; 2010.