

Tutorial for heat map coloring of kazius compounds

The tutorial assumes, that you have downloaded the following files from the site:

FingerPrinter: FingerPrinter.jar

LearningLibrary: LearningLibrary.jar

HeatmapViewer: HeatmapViewer.jar

Kazius data set: kazius.standardized.sdf

List with compounds: compoundsOfInterest.txt

The tutorial contains of three steps:

1. Calculation of Fingerprint encodings
2. Calculation of virtual screening model with LIBLINEAR
3. Visualization of compounds of interest with heat map bond coloring

Calculation of fingerprint encodings

The library has several options, which are accessible by typing:

```
java -jar FingerPrinter.jar
```

To gain a ECFPVariant encoding of depth 4 with hash space size of 2^{22} , the label of the AMESTEST, and Daylight invariant atom typing:

```
java -jar FingerPrinter.jar -p 3 -d 4 -a 6 -x 22 -l AMESTEST -f  
kazius.standardized.sdf
```

The result is a file, which contains the fingerprints of all compounds.

Calculation of virtual screening model

For calculation of the virtual screening model we employ a modified version of the large-scale linear support vector machine LIBLINEAR. It is possible to optimize for different virtual screening metrics in the model selection step. Additionally, different types of cross-validation are possible.

To get a help on all options, type:

```
java -jar LearningLibrary.jar
```

We train a model with a 2-fold cross-validation (-vG CV 2) optimizing the accuracy (-m ACC). The grid is chosen as $\log_2(C) = \{-5, -4, \dots, 7, 8\}$ and $\log_2(W_{-1}) = \{-4, -2, 0\}$. Features have to occur at least 3 times to be included in the training (-f UCF 3). The model is saved in the same directory (-sm). The command is:

```
java -jar LearningLibrary.jar -vG CV 2 -G C=-5,8,1 W-1=-4,0,2 -f UCF  
3 -m ACC -sm -F  
kazius.standardized.sdf.22bits.ECFPVariant_4.dayR.bit.fp
```

Visualization with heat map bond coloring

The command line help is accessible by

```
java -jar HeatmapViewer.jar
```

The depth and hashspace size should be chosen as in the fingerprint calculation step (depth 4 and hashspace size= 2^{22}). The hash space size is entered as the exponent x of 2^x and, thus, $x=22$. The sdf file and the model file are the data set file and the virtual screening model. The compounds of interest are the two compounds visualized in the publication. Finally, a boolean has to be set to enable the full set normalization. If you want to save the image as png, do not forget to add additional heap space by `-Xmx` java option.

The complete command is:

```
java -jar HeatmapViewer.jar kazius.standardized.sdf  
kazius.standardized.sdf.22bits.ECFPVariant_4.dayR.bit.fp.ACC.LL.mode  
l 4 22 compoundsOfInterest.txt True
```

The result should look like this:

