CONSENSUS MODELING OF ANXIOLYTIC ACTIVITY OF CHEMICAL COMPOUNDS BY CONVOLUTIONAL NEURAL NETWORKS

Vassiliev Pavel Mikhailovich
Depression and anxiety disorders

635 million people
Goal

Build classification consensus ensemble model of dependence of anxiolytic activity of chemical compounds from their structural QL descriptors using method of convolution artificial neural networks
Tasks

- Formation of verified training set for known anxiolytic substances
- Calculation of QL descriptor representation for training set compounds
- Line convolution of find structural parameters
- Training of neural networks and formation of consensus ensemble model
- Prediction of anxiolytic activity for new compounds
Finding information about known compounds

Anxiolytic activity

1,939 entries
### Unification and aggregation of techniques

**Anxiolytic** — 707

#### Clustering by k-means method

4 activity classes:
- High
- Moderate
- Low
- Inactive

#### Unification of techniques and activity clustering

- Clustering by k-means method
- 4 activity classes
- High, Moderate, Low, Inactive
537 verified structures of anxiolytic substances: 273 pronounced activity 264 low activity
**Structural descriptors**

- $\text{-N<}$ heteroatomic
- $\text{-CH3}$ carbonic
- $\text{Cyc05}$ cyclic

**Length descriptors**

- $03$ path length
- $-1$ inclusion index
- $p..1$ Bond descriptors
Composite descriptors of QL language

4 rank
Basic

3 rank

2 rank
8686 species of QL descriptors of 11 different types
Correlation of one sequence with the reversed second sequence.

\[ y(n) = \sum_{m=0}^{n} h(n - m) x(m) \]

8686 QL descriptors → 66 convolution variables
Example of QL descriptor convolution

<table>
<thead>
<tr>
<th>SD</th>
<th>-NH2</th>
<th>&gt;NH</th>
<th>-N&lt;</th>
<th>-N=</th>
<th>#N</th>
<th>&gt;N+=</th>
<th>-OH</th>
<th>&gt;O</th>
<th>=O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BD</td>
<td>...0</td>
<td>...1</td>
<td>..n1</td>
<td>.a.0</td>
<td>.a.1</td>
<td>.A.0</td>
<td>.A.1</td>
<td>p..0</td>
<td>p..1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>30</td>
<td>0</td>
<td>10</td>
<td>23</td>
<td>15</td>
<td>30</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Initial data

<table>
<thead>
<tr>
<th>SD</th>
<th>-NH2</th>
<th>&gt;NH</th>
<th>-N&lt;</th>
<th>-N=</th>
<th>#N</th>
<th>&gt;N+=</th>
<th>-OH</th>
<th>&gt;O</th>
<th>=O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BD</td>
<td>p..1</td>
<td>p..0</td>
<td>.A.1</td>
<td>.A.0</td>
<td>.a.1</td>
<td>.a.0</td>
<td>..n1</td>
<td>...1</td>
<td>...0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>15</td>
<td>23</td>
<td>10</td>
<td>0</td>
<td>30</td>
<td>12</td>
</tr>
</tbody>
</table>

Data for convolution

\[
SD\cdot BD = (0\times0) + (0\times0) + (2\times30) + (3\times15) + (0\times23) + (0\times10) +
(1\times0) + (0\times30) + (0\times12) = 0 + 0 + 60 + 45 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 105
\]
Formation of training set

Anxiolytics with pronounced activity

Anxiolytics with low activity

66 neurons

7 samplings
Neural network with bottle-neck

Input neurons, I

Hidden layer, H

Output neurons, O

\( f_1(x) \)

\( f_2(x) \)
7 sampling options

4000 trained neural networks for each sampling option

50 automatically selected neural networks

1 best neural network

Ensemble of 7 neural networks

~30 000 networks were trained
Accuracy of neural network ensemble

Accuracy \( F_0 = 78\% \)

Sensitivity \( F_a = 85\% \)

Specificity \( F_n = 70\% \)
Prediction results

Diazepinobenzimidazole

Benzoimidazopyrazine

Quinoxaline

Triazolobenzodiazepine

Mercaptobenzimidazole

97 new compounds – 53 active
Laboratory for Information Technology in Pharmacology and Computer Modeling of Drugs

~40 Tflops
Thank You for Your attention!