ARTIFICIAL INTELLIGENCE IN CHEMISTRY & CHEMOMETRICS
CHEMOMETRICS

- Part of chemistry
- Extraction of useful information from multidimensional data
- MATHEMATICS + STATISTICS + ARTIFICIAL INTELLIGENCE

1971, Svante Vold
1974, „part of chemistry”
CHEMOMETRICS

- Design of experiments (DOE)
- Simplex optimisation
- Factor design
- Cluster analysis
- Relation modelling
- Classifiers
- PCA
- Pattern recognition
Artificial Intelligence (AI)

- technology and methods inspired by informatics and psychology
- Construction of machines, which way of acting can be considered as „human” (caused by „human” intelligence)

1955, John McCarthy
Artificial Intelligence

- Artificial Neural Networks (ANN)
- Fuzzy Logic
- Modeling of uncertainty, probabilistic estimation
- Expert Systems (ES)
- Machine learning
- Optimization methods
- Multidimensionalal Statistics
- Genetic Algorithms (GA)
- Pattern Recognition (PARC)
Artificial Intelligence – how far are we?

The progress is very hard and slow...

- Fuzzy logic – process control in industry
- Expert Systems – pharmacy and medicine
- Machine translation
- Neural Networks
- Optical recognition, speech recognition, handwriting recognition
- Deep Blue won with Gary Kasparov
- Economics – automatic systems able to estimate credit capability

...but:

- Human conversation
- Generating profits on stock exchange
- Proper translation of literature and common parlance
Artificial Intelligence in chemistry

The main goals of AI in chemistry:

Optimization and approximation

- Neural Networks
- Genetic Algorithms
- Expert Systems
Artificial Neural Networks (ANN)

- Adaptative structure
- Able to model complicated relations In-Out
- Generalization of obtained knowledge
- Proper processing of incomplete data
- Parallel computing

40’ - Warren McCulloch, Walter Pitts
ANN in chemistry

- Calibration of devices, sensors
- Development of new measurements methods
- Dynamic process on-line monitoring
- Signal processing
- “Shape classification”
- QSAR, QSPR
ANN in chemistry – QSAR, QSPR

QSPR = Quantitative Structure-Property Relationship

QSAR = Quantitative Structure-Activity Relationship
Genetic Algorithms (GA)

- Search the space of alternative solutions in order to find the best one

\[\rightarrow\text{OPTIMIZATION}\]

- The procedure emulates biological evolution

John Henry Holland
# Genetic Algorithms

<table>
<thead>
<tr>
<th>Natural</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>chromosome</td>
<td>string</td>
</tr>
<tr>
<td>gene</td>
<td>feature, character or detector</td>
</tr>
<tr>
<td>allele</td>
<td>feature value</td>
</tr>
<tr>
<td>locus</td>
<td>string position</td>
</tr>
<tr>
<td>genotype</td>
<td>structure, or population</td>
</tr>
<tr>
<td>phenotype</td>
<td>parameter set, alternative solution, a decoded structure</td>
</tr>
</tbody>
</table>
## The Coding

### Binary Representation
(the most frequently used)

<table>
<thead>
<tr>
<th>No.</th>
<th>STRING</th>
<th>PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100001</td>
<td>x=4; y=1</td>
</tr>
<tr>
<td>2</td>
<td>001100</td>
<td>X=1; y=4</td>
</tr>
<tr>
<td>3</td>
<td>110010</td>
<td>X=6; y=2</td>
</tr>
<tr>
<td>4</td>
<td>000100</td>
<td>X=0; y=4</td>
</tr>
</tbody>
</table>

**PARAMETER 1**

<table>
<thead>
<tr>
<th>z, b.</th>
<th>0,7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**PARAMETER 2**

<table>
<thead>
<tr>
<th>0,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

**PARAMETER 3**

<table>
<thead>
<tr>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

**PARAMETER 4**

<table>
<thead>
<tr>
<th>1,3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

**GENE**

<table>
<thead>
<tr>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
Selection

The main tool of evolution ;-) 

• Selection of individuals according to strict, determined criterion
• Criterion = evaluation function determined by the user
  • High value -> reproduction of the individual
  • Low value -> the individual dies
• Following generations are more and more adapted
Genetic Operators

Before recombination:

Before recombination:

After recombination:

After recombination:

Crossing-over Point

Crossing-over

1 0 0 0 1 0 1 1 0 1

1 1 0 1 1 1 0 1 1 1

1 1 0 1 1 1 1 1 1 0 1

1 0 0 0 1 0 0 1 1 1

1 1 0 1 1 1 1 1 1 0 1

1 1 0 1 1 1 1 1 1 0 1

1 1 0 1 1 1 1 1 1 0 1

1 1 0 1 1 1 1 1 1 0 1
Genetic Operators

000101101

Mutation

000111101
GA in Chemistry

- Curve fitting (IR spectra)
- Feature selection – multicomponent calibration
- Determination of the configuration of some systems (for example C60)
- The composition of complex materials (for example composites)
- Molecular structure optimization
- Protein folding (3D structure of proteins)
- Protein-ligand docking
Expert Systems (ES)

Program / set of programs

It aims to recall the use of knowledge and making decisions

Why?

• Costs
• No experts in many cases
• Work more fast
• Not get tired
• Consequent
• Objective
• Precise
• The analysis of huge amounts of data demands the use of a computer
Expert System - a scheme

The framework of ES:

- User interface
- Explanation mechanism
- Conclusion mechanism
- Knowledge base editor
- Variables’ database
- Knowledge base

User

Expert

Knowledge engineer
ES in Chemistry - DENDRAL

- Generation of chemical structures according to data obtained from MS, NMR, IR, UV
- From 1969 used in chemistry, many problems solved:
  - structure of organic esters,
  - hormones,
  - antibiotics,
  - impurities in chemical substances.

The results for mixtures are better than those obtained by experts!
ES in Chemistry - CRYSTALIS

- Program, which elaborates crystallographic data of proteins (costs of classical analysis – even 1,000,000 $).
- Interpretation of electron density maps and other data.
ES in Chemistry - SYNTHESIS

• **CASD** - Computer Aided Synthesis Design
• 20 atoms - $10^{18}$ various molecules!
• Number of basic substrates in synthesis ~500, in industry even more (~2000).
• Database – about 500 reactions, more than 100,000 concrete ones.
Summary

• Chemometrics & Chemistry <-> AI
• Analysis of multidimensional, complex data
• Optimization, approximation
• ANN, GA, ES

MORE

• Chemometrics and Intelligent Laboratory Systems
• Journal of Chemometrics
• Environmetrics
• Analytical Chemistry
• Analytical Letters
• Analytica Chimica Acta