Application of self-organizing maps in astroinformatics

Lukaš Lopatovský[1], Petr Škoda[2]

[1] *Czech Technical University, Prague, Czech Republic*
[2] *Astronomical Institute of the Academy of Sciences, Czech Republic*

Abstract

The astronomy is source of big amount of data with possibility to be effectively examined by computer. For this purpose we focused on the machine learning algorithm called self-organizing maps and its use on big astronomical data. For scaling of algorithm we used well known data sets from the UCI repository and then concerned ourselves particularly with the classification of stars based on their spectral characteristics.

The self-organizing maps are excellent tool for clustering data in a new way. The unsupervised learning paradigm enabled us to find the groups of self-similar outliers and hidden patterns in data, hardly possible to find by eye. The algorithm shows very promising results.

Application of Self-Organizing Maps in Astroinformatics

Lukáš Lopatovský
lopatluk@fit.cvut.cz

Faculty of Information Technology
Czech Technical University in Prague

New challenges in astro- and environmental informatics in the Big Data era, 2014

Problem domain

- Clustering of stellar spectra
- Data mining
- Self-organizing maps.

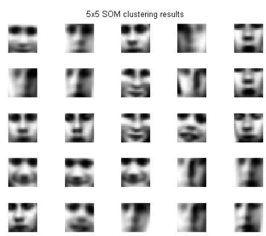
Self-organizing maps

- Similar to real brain
- Unsupervised learning paradigm
- Self-organization

Usage

- Pattern recognition (sounds, photos, countries, ...)
- Reduction of dimensionality
- Visualization for high-dimensional data
- Topological ordering of data space (clustering)

Usage



Topology of the map

- Input layer
- Grid of neurons
- Connection weights

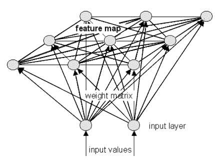


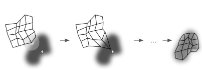
Figure: Topology of SOM

Algorithm

- Random initialization of weights
- Selection of random vector
- Finding of best matching unit
- Counting of neighborhood radius
- Actualization of weight to better fit to input vector

$$\mathcal{W}_{ij}(t) = \mathcal{W}_{ij}(t+1) + \Theta(t,i)\mathcal{L}(t)(\mathcal{V}_j(t) - \mathcal{W}_{ij}(t))$$

- Choose another random vector



Data

- Spectra of Be and normal stars from archive of the Astronomical Institute of the Academy of Sciences of the Czech Republic in Ondřejov
- Four shape depended classes
- Special class containing outliers

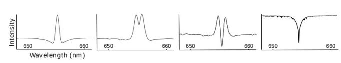


Figure: Characteristic shapes of spectra.

Application

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

- Wrapped more implemantations together
- Sequential reading of files
- Also special boosting methods (probing algorithm)
- Parallelization (10 threads, 8x) , MapReduce, GPU

	learning rate function	time [s]
Rapid Miner	linear	215
WEKA	linear	202
SOM_PAK	inverse time	116
SOM_PAK	linear	103
Our Implementation	exponential	68

Table: Time comparison

Application

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

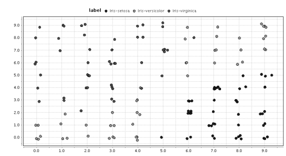


Figure: Association matrix of `Iris` data set after using exponential learning rate function.

Application

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

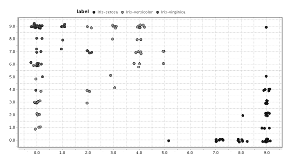


Figure: Association matrix of `Iris` data set after using linear learning rate function.

Application

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

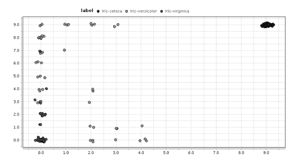


Figure: Association matrix of `Iris` data set after using inverse time learning rate function.

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

	learning rate function	q.err (iris)	q.err (stellar spectra)
SOM_PAK	inverse time	0.035	104.19
SOM_PAK	linear	0.018	86.92
Our Implem.	exponential	0.005	7.78

Table: Time comparison of implementations

Clustering of outliers

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

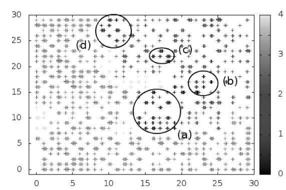
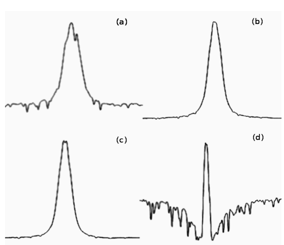


Figure: Four main clusters of outliers.

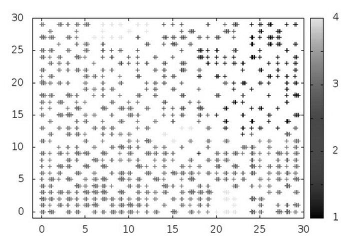
Clustering of outliers

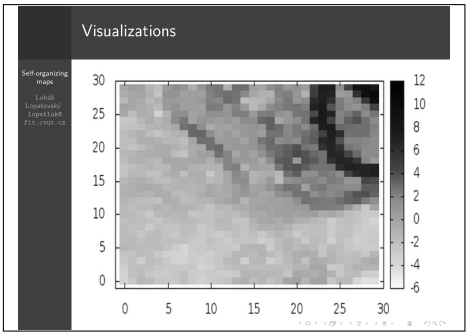
Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz

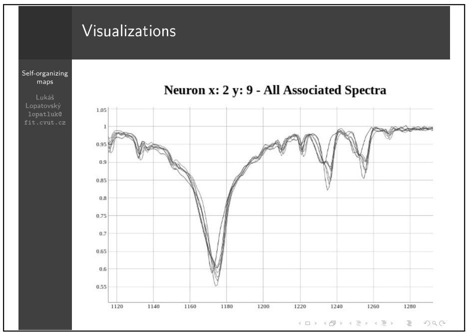


Visualizations

Self-organizing maps
Lukáš Lopatovský lopatluk@fit.cvut.cz




Visualizations
Self-organizing maps
Lukáš Lopatovský
lopatluk@ fit.cvut.cz
30
25
20
15
10
5
0
0 5 10 15 20 25 30
12
10
8
6
4
2
0
-2
-4
-6


Visualizations
Self-organizing maps
Lukáš Lopatovský
lopatluk@ fit.cvut.cz
Neuron x: 2 y: 9 - All Associated Spectra