

# FAKE GAME

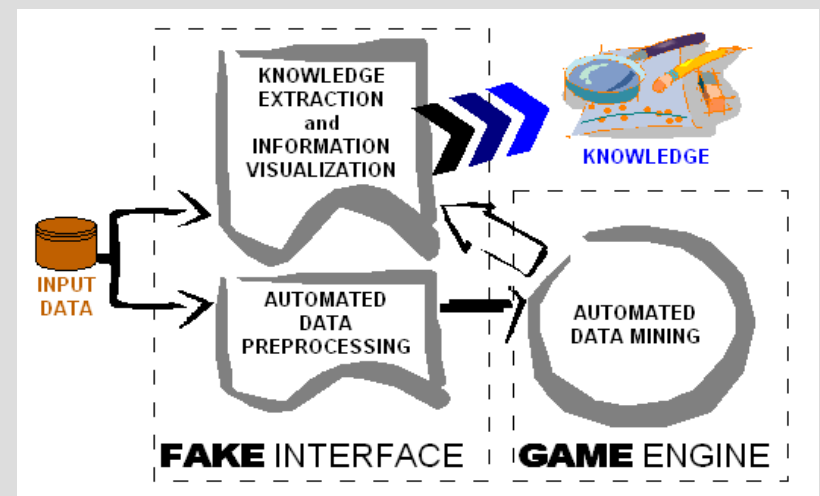
Vladimír Šťastný

a4m35osp (Open-source programování)

ČVUT, FEL 2010

# Popis aplikace

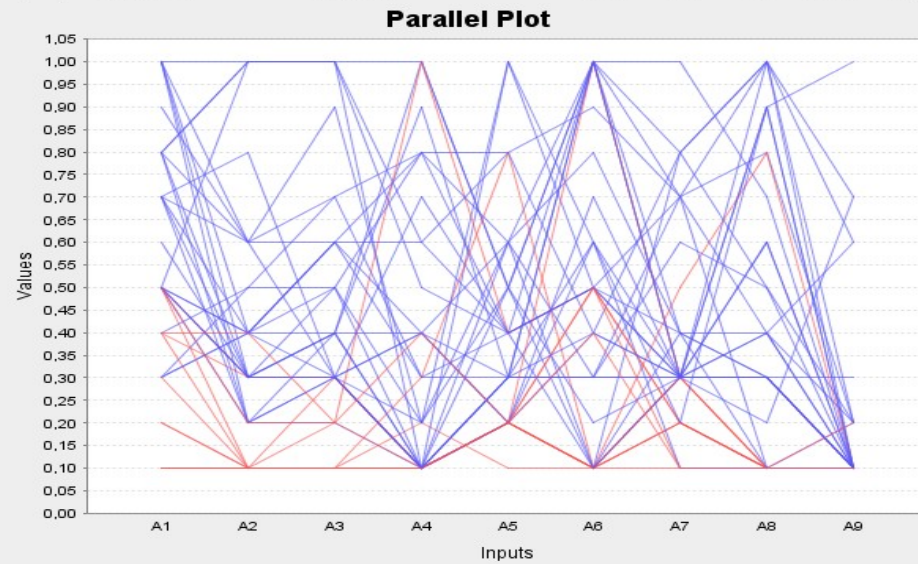
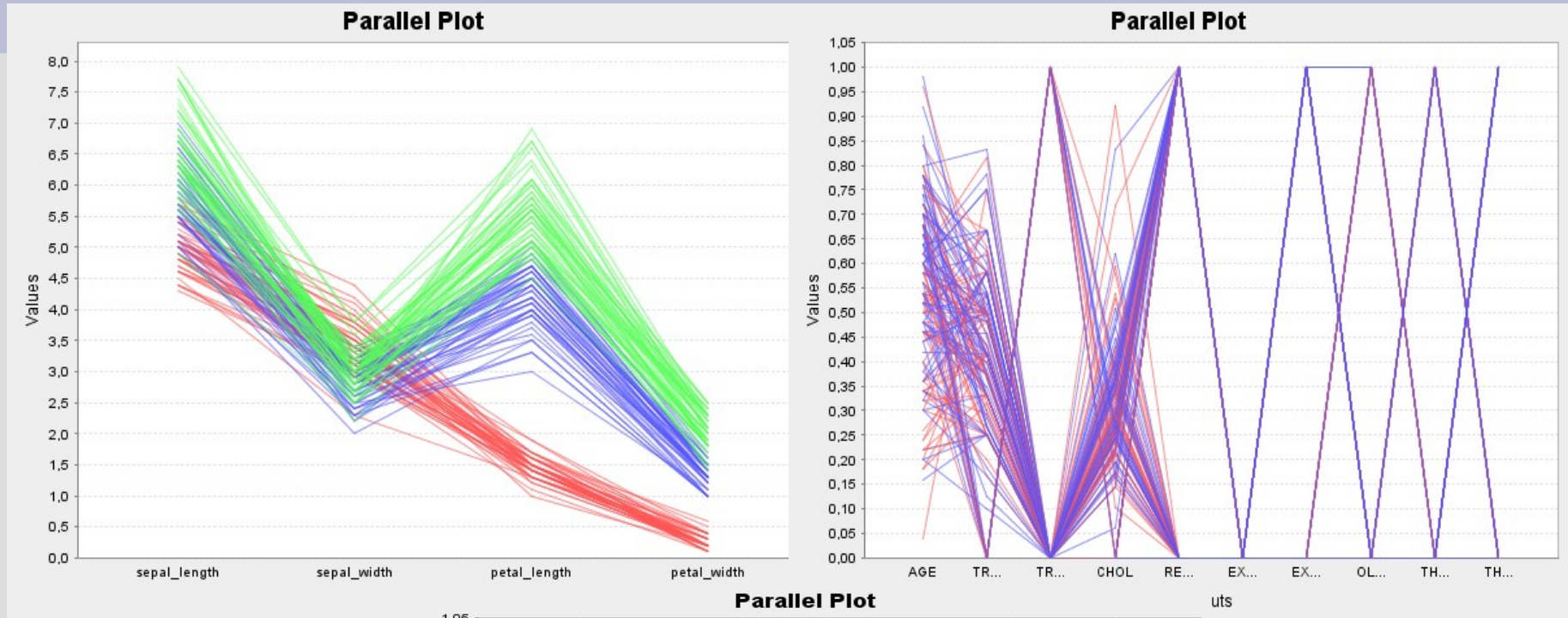
- data mining (DM) nástroj vyvíjený skupinou Výpočetní inteligence na Katedře počítačů ČVUT FEL (GPL licence)
- využívá k vytváření DM modelů algoritmy inspirované přírodou
- implementováno v Javě; cca 180 packages
- klasifikace, regrese, shluková analýza, predikce ad.
- režimy běhu aplikace:
  - GUI
  - „Fully Automated Knowledge Extraction“



# Moje činnost

- úprava pravidel pro tvorbu vizualizací modelů
  - operation flow: načtení datasetu, předzpracování, vytvoření a ohodnocení modelu, vizualizace
  - výsledkem je report výsledků
- test na vybraných datasetech
  - archiv datasetů Proben
  - různé vlastnosti dat
- zdrojové kódy dostupné na <http://fakegame.sourceforge.net/doku.php?id=research> (visualizations rules update)

# Generování spojnicového grafu



# Generování spojnicového grafu II.

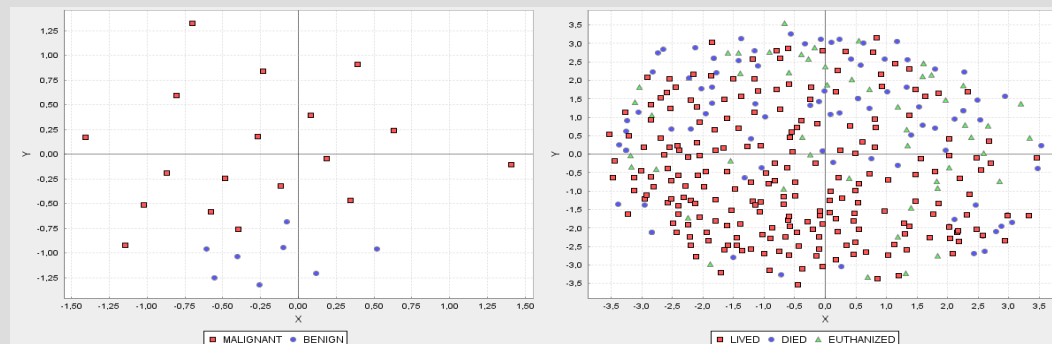
- křížení hodnot pro jednotlivé vstupní atributy
- hledání pořadí vstupních atributů tak, aby docházelo k co nejmenšímu křížení
- genetický algoritmus
  - řešení kódováno jako permutace vstupních atributů
  - fitness funkce: počet křížení

# Zkrácení doby výpočtu

- ukázka výsledků pro vybraný dataset

Operation	Time	Description
Register subreports	00:00:00:006	Registers and defines order of subreports.
Import file	00:00:00:082	Imports dataset file.
Prepare reports	00:00:00:002	Prepares a report directory.
Basic Preprocessing	00:00:00:006	Performs missing values imputations, ...
Data reduction	00:00:00:001	Reduces the number of instances, if the dataset is too large.
Feature ranking	00:00:01:642	Feature ranking of dataset attributes
Visualizes dataset	00:00:02:670	Visualizes dataset using histograms, parallel plots, Andrew's curves, etc.
Project dataset to 2D	00:13:50:911	Projects dataset to 2D using Sammon's.
Multiple model iterator	00:00:17:374	If more than one classifier/model config is given, this iterates over all of them.
Operation Flow Info	00:00:00:000	Constructs this subreport.
Total	00:14:12:694	Total time spent.

- nejvíce času zabírá Sammonova projekce
- pro větší počet vstupních atributů se ztrácí vypovídací hodnota
- pokud počet vstupních atributů přesahuje určitou hodnotu, Sammonova projekce se neprovede



# Zpřehlednění reportu

- každý model popsán rovnicí - výstupní proměnná funkcí vstupních atributů

HYPER=46.170\*exp(-(power(A1-0.587;2)+power(A2-0.442;2)+power(A3-0.708;2)+power(A4-0.396;2)+power(A5-0.261;2)+power(A6-0.413;2)+power(A7-0.817;2)+power(A8-0.664;2)+power(A9-0.448;2)+power(A10-0.597;2)+power(A11-0.632;2)+power(A12-0.384;2)+power(A13-0.339;2)+power(A14-0.536;2)+power(A15+0.431;2)+power(A16-0.458;2)+power(A17-3.259;2)+power(A18+0.136;2)+power(A19+3.397;2)+power(A20-0.496;2)+power(A21+4.230;2))/power(6.625;2))-16.128;  
NORM=3.050\*exp(-(power(A1-1.019;2)+power(A2-0.153;2)+power(A3+1.212;2)+power(A4+0.492;2)+power(A5-0.040;2)+power(A6-0.777;2)+power(A7+0.469;2)+power(A8+0.974;2)+power(A9-1.247;2)+power(A10-0.617;2)+power(A11-0.219;2)+power(A12-1.237;2)+power(A13+0.448;2)+power(A14-0.558;2)+power(A15+0.194;2)+power(A16-0.496;2)+power(A17-0.400;2)+power(A18+0.214;2)+power(A19-0.551;2)+power(A20-0.522;2)+power(A21-0.499;2))/power(2.754;2))-0.605;  
SUB=40.469\*exp(-(power(A1+0.101;2)+power(A2-0.963;2)+power(A3-1.109;2)+power(A4-0.854;2)+power(A5-1.339;2)+power(A6-0.479;2)+power(A7-0.298;2)+power(A8-1.551;2)+power(A9-0.402;2)+power(A10-0.186;2)+power(A11-0.612;2)+power(A12-0.741;2)+power(A13-1.322;2)+power(A14-0.860;2)+power(A15+0.491;2)+power(A16-0.315;2)+power(A17+4.578;2)+power(A18-0.400;2)+power(A19-4.665;2)+power(A20+0.974;2)+power(A21-6.369;2))/power(8.989;2))-12.069;

- potlačení výpisu, pokud rozsah delší než určitý počet řádků

Děkuji za pozornost

a4m35osp (Open-source programování)