

Stock Market Prediction with Fuzzy Type-2 Clustering and Neural Networks

Marek Kukačka

MFF UK

1. prosince, 2011

Content

- 1 Úvod
- 2 Regrese
- 3 Clustering
- 4 Neuronová síť

Co a jak

Cíl:

- předpovídat budoucí hodnotu cen na trhu s cennými papíry (stock market)

Prostředky:

- trénovací data
- násobná regresní analýza - pro výběr vhodných proměnných
- fuzzy clustering - pro nalezení shluků v trénovacích datech, vhodných jako vstupy pro neuronovou síť
- neuronová síť - pro aproximaci funkční závislosti mezi hodnotami proměnných a budoucími hodnotami trhu

.. až na to, že všechno, kromě regresní analýzy, je Type-2 fuzzy ..

Data

- použít S&P 500 Index
- příklady sledovaných proměnných:
 - past prices, volume, technical indicators, T-bill rates, certificate of deposit rates, credit ratings, produce/consumer price indexes, industrial productino levels, money supply levels, ..
- hodnoty zaznamenávány měsíčně, použito 361 měsíců
- vyřazeny proměnné s nevhodnými t-statistikami a p-hodnotami

Vícenásobná regresní analýza

”Multiple regression analysis”

- podobná lineární regresi, ale pracuje s více proměnnými
- výsledkem je lineární rovnice více proměnných

Výsledky regresní analýzy

Identifikovány proměnné s největší korelací s budoucí hodnotou S&P 500 indexu:

- 3-month T-bill rate (T-bill3), 3-month Certificate of Deposit rate (CDR3), pst S&P500 Index (SP500) price level, past Money Supply (M1) level, Industrial Production (IP) reading, recent Producer Price Index (PPI)

$$SP500_{t+1} = -33.690 + (11.999 * T - Bill3) + (1.365 * IP) + (0.059 * M1) - (0.781 * PPI) - (9.124 * CDR3) + (0.955 * SP500)$$

R^2 hodnota výsledného modelu je 0.994, což znamená, že model vysvětluje 99.4% budoucích hodnot indexu

Clustering

Použitá metoda:

- Differential Evolution Optimization-based Fuzzy type-2 Clustering
- varianta FCM, kde fuzifikační parametr m je fuzzy interval $[m_1, m_2]$
- díky tomu jsou prováděny dvě klastrování, každá s jednou mezní hodnotou pro m
- klastrování je zlepšováno evolučně - jedincem v evoluci je kompletní vektor středů klastrů
- výsledkem jsou fuzzy clustery - kombinace středů klastrů z obou FCM verzí

Clustering (2)

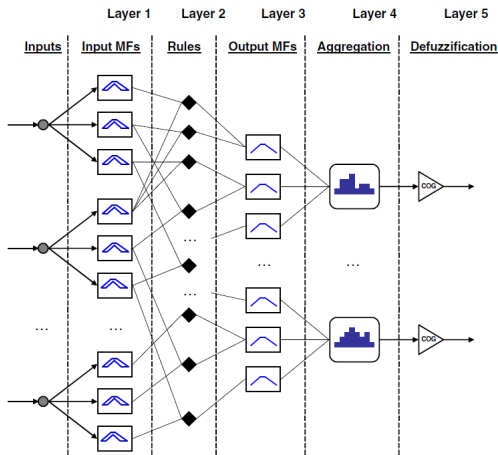
- praktický interval pro hodnotu m je $[1.4, 2.6]$
- výsledné fuzzy klastry jsou použity pro mapování vstupních hodnot neuronové sítě do fuzzy množin

Type-2 Fuzzy Inference Neural Network

Síť založená na Type-2 Fuzzy Logic systému, emuluje jeho komponenty:

- fuzzifier (enkodér) - mapuje vstupní hodnoty do type-2 množin
- báze type-2 fuzzy pravidel
- inferenční engine
- zpracování výstupu (defuzzifier, dekodér)

Type-2 Fuzzy Inference Neural Network (2)



Type-2 Fuzzy Inference Neural Network (3)

- parametry sítě: parametry vstupních a výstupních fuzzy membership funkcí
- doladěny diferenční evolucí
- optimalizovaná funkce: RMSE (Root Mean Square Error)
- dosaženo lepších výsledků než pomocí přístupu založeného na type-1 fuzzy logice
 - Type-1 approach RMSE: 0.948
 - Type-2 approach RMSE: 0.909