

# Hurtownie Danych i Data Mining

Prezentacja prac nad projektem

Wpływ wydarzeń politycznych, ekonomicznych i naturalnych na kształtowanie się cen akcji na rynku polskim

Tomasz Kupczyk, Patryk Obara, Karol Stosiek

Instytut Informatyki Uniwersytetu Wrocławskiego

11 maja 2010

# Plan prezentacji

- 1 Definicja problemu
- 2 Plan badań
- 3 Stan prac
- 4 Plany na przyszłość
- 5 Używane narzędzia

## Definicja

Obserwacje wahań kursów akcji pojedynczych spółek na giełdach papierów wartościowych wykazują pewną korelację z wydarzeniami politycznymi, ekonomicznymi i naturalnymi. Celem projektu jest próba zgrupowania tych spółek giełdowych w grupy wykazujące podobne zachowanie i zbadanie, które z tych grup są podatne na wydarzenia historyczne z uwzględnieniem kategorii tych wydarzeń.

- 1 Zebranie danych badawczych: notowań giełdowych wybranych spółek oraz stworzenie bazy wydarzeń historycznych.
- 2 Zbudowanie wiarygodnego narzędzia do grupowania notowań spółek giełdowych oraz analizy wydarzeń historycznych
  - w oparciu o algorytm k-means;
  - z możliwością tworzenia wykresów analizowanych danych;
  - z naniesionymi wydarzeniami historycznymi.
- 3 Dobranie odpowiednich parametrów do wiarygodnej analizy otrzymanych danych;
- 4 Zanotowanie obserwacji i podjęcie próby wyciągnięcia pierwszych wniosków;
- 5 Rozbudowa narzędzia o analizę opartą na sieciach Kohonena;
- 6 Ponowne dobranie parametrów do analizy otrzymanych danych;
- 7 Zanotowanie obserwacji i wyciągnięcie wniosków;
- 8 Porównanie z wynikami osiągniętymi podczas analiz opartych o wyniki algorytmu k-means.

- 1 **Zebranie danych badawczych: notowań giełdowych wybranych spółek oraz stworzenie bazy wydarzeń historycznych.**
- 2 **Zbudowanie wiarygodnego narzędzia do grupowania notowań spółek giełdowych oraz analizy wydarzeń historycznych**
  - w oparciu o algorytm k-means;
  - z możliwością tworzenia wykresów analizowanych danych;
  - z naniesionymi wydarzeniami historycznymi.
- 3 **Dobranie odpowiednich parametrów do wiarygodnej analizy otrzymanych danych;**
- 4 **Zanotowanie obserwacji i podjęcie próby wyciągnięcia pierwszych wniosków;**
- 5 Rozbudowa narzędzia o analizę opartą na sieciach Kohonena;
- 6 Ponowne dobranie parametrów do analizy otrzymanych danych;
- 7 Zanotowanie obserwacji i wyciągnięcie wniosków;
- 8 Porównanie z wynikami osiągniętymi podczas analiz opartych o wyniki algorytmu k-means.

# Aktualne osiągnięcia

Ilość iteracji: 100; Ilość grup: 5, 10, 20; Miara odległości: euklidesowa.

- Uruchomienie k-means dla parametru  $k=20$  daje coraz rozsądniejszy podział na grupy spółek zachowujących się podobnie (ca. 10 grup spółek ma sens)
- Uruchomienie k-means dla wektorów cen daje bardzo liczne grupy, ale spółki zachowują się podobnie w obrębie grupy;
- Uruchomienie k-means dla wektorów różnic cen daje w zasadzie jedną bardzo liczną grupę spółek, co uniemożliwia sensowną analizę;
- W obydwu przypadkach przynajmniej połowa otrzymanych grup to grupy zawierające niewielką ilość spółek (1,2,3).

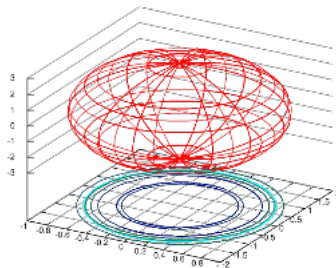
- 1 Zebranie danych badawczych: notowań giełdowych wybranych spółek oraz stworzenie bazy wydarzeń historycznych.
- 2 Zbudowanie wiarygodnego narzędzia do grupowania notowań spółek giełdowych oraz analizy wydarzeń historycznych
  - w oparciu o algorytm k-means;
  - z możliwością tworzenia wykresów analizowanych danych;
  - **z naniesionymi wydarzeniami historycznymi.**
- 3 Dobranie odpowiednich parametrów do wiarygodnej analizy otrzymanych danych;
- 4 Zanotowanie obserwacji i podjęcie próby wyciągnięcia pierwszych wniosków;
- 5 **Rozbudowa narzędzia o analizę opartą na sieciach Kohonena;**
- 6 **Ponowne dobranie parametrów do analizy otrzymanych danych;**
- 7 **Zanotowanie obserwacji i wyciągnięcie wniosków;**
- 8 **Porównanie z wynikami osiągniętymi podczas analiz opartych o wyniki algorytmu k-means.**

## Pycluster





## gnuplot



Dziękujemy za uwagę.