

# 集合知プログラミング

## 第10章 特徴を発見する(3)

05T4007T 江口晃

# 株式市場のデータを使用する

- NMF(非負値行列因子分解)で名義データを扱う

例: 単語のカウント

- NMFで数値データを扱う

例: 株式市場の取引量のデータに対しての適用

-重要な取引日のパターンが見える

-複数の株が取引される要因が見える

# 取引量とは何か？

- 株の取引量

- 与えられた期間内(通常は1日)に売られたり、買われた株券の数のこと。

- 取引量は株価に大きな変化があった場合に大きくなる傾向にある

- 例:会社の重大な発表、会社や業界へのニュース

- 外的なイベントがなければ、通常は一定の値を保つ。

- 株の取引量を扱う

- 複数の株に影響を与える取引量の変化パターンを調べる

- 他への影響が強いイベントを調べる

# Yahoo!Financeからデータをダウンロードする(1)

- Yahoo!Finance

- すべての種類の財務データをダウンロードできる

- ダウンロードできるデータには、株価、オプション、為替レート、社債利息のレートなどがある。

- CSV形式のダウンロード

Date,Open,High,Low,Close,Volume,Adj. Close\*

22-Dec-06,25.67,25.88,25.45,25.55,14666100,25.55

21-Dec-06,25.71,25.75,25.13,25.48,27050600,25.48

20-Dec-06,26.24,26.31,25.54,25.59,24905600,25.59

- それぞれの行は日付、初めと終わりの値段、取引量、調整後終値を含んでいる。

# Yahoo!Financeからデータをダウンロードする(2)

- CSV形式の利用
  - それぞれのチェッカーシンボルのURLを開きデータをダウンロードする
  - それぞれの行をコンマで分割し、株式の取引量のリストを作る。
- 対象としたティッカーシンボル
  - YHOO、AVP、BIIB、BP、CL、CVX
  - DNA、EXPE、GOOG、PG、XOM、AMGN

# 行列の準備

- 行列への変形

- ティッカーシンボルの形式

```
tickers=['YHOO','AVP','BIIB','BP','CL','CVX',  
         'DNA','EXPE','GOOG','PG','XOM','AMGN']
```

- リストの作成

```
[[4453500.0, 842400.0, 1396100.0, 1883100.0, 1281900.0, ...]  
 [5000100.0, 1486900.0, 3317500.0, 2966700.0, 1941200.0, ...]  
 [5222500.0, 1383100.0, 3421500.0, 2885300.0, 1994900.0, ...]  
 ...]
```

# NMFを走らせる

- NMFを走らせる

- 探し出す特徴の数を決める(本実験では5個)

- factorize関数を呼び出す

- 例: 写真の集合中から顔の特徴を自動的に決定する

- 重みと特徴の行列

- 特徴の行列中のそれぞれの行は特徴(株の取引量)を表す

- 重みの行列中のそれぞれの行は特定の日を表す

# 結果を表示する

- 結果の表示

- 行列を直接解釈するのは難しいため、もっと良いやり方で表示する必要がある。

- 今回の実験で見たい情報

- それぞれの特徴のそれぞれの株に対する取引量の分布

- それぞれの特徴にもっとも強く関連のある日付たち

# 結果の表示例1

- 結果の表示例1

Feature 4

(74524113.213559602, 'YHOO')

(6165711.6749675209, 'GOOG')

(5539688.0538382991, 'XOM')

...

[(7.950090052903934, '19-Jul-06'),

(4.7278341805021329, '19-Sep-06'),

(4.6049947721971245, '18-Jan-06')]

- 特徴から読み取る

-この特徴はYHOOのみに当てはまる

-2006年7月19日に強く当てはまる

-この日はYahoo!が収益予想を発表した日だった

# 結果の表示例2(1)

- 結果の表示例2

Feature 2

(46151801.813632453, 'GOOG')

(24298994.720555616, 'YHOO')

(10606419.91092159, 'PG')

...

[(2.9757765047938824, '20-Jan-06'),

(2.8627791325829448, '28-Feb-06'),

(2.356157903021133, '31-Mar-06')]

- 特徴から読み取る

-この特徴はいくつかの会社に均等に当てはまる

-この特徴はGoogleの取引量が非常に大きかった日を表している

# 結果の表示例2(2)

- 特徴から読み取る(続き)
  - 1月20日はGoogleが検索エンジンの使用についての情報を政府には渡さないことをアナウンスした日だった
  - このイベントはYahoo!にはほとんど関係ないことであるにも関わらず、Yahoo!の取引量にも大きな影響が出ている
  - 2月28日はGoogleの最高責任者が成長が鈍くなっていると発表した時にあった取引量の急激な山である
  - チャートによるとこの日Yahoo!の取引量も増加している

# まとめ

-二つの特徴の例は、GoogleとYahoo!は時には同じように、時には別々に動くこともあることを示している。相関だけを見ていては、これらの関係はすべて平均化されてしまう。

-もっと多くの株のデータを利用して、探すパターンを増やすことで、さらに複雑な相互関係が明らかになるだろう。