

Como o projeto foi me foi apresentado

Tema *

Algoritmos de visualização de informação aplicados à dinâmica dos mercados financeiros

Objetivos *

O objetivo deste projeto será o de desenvolver ferramentas de visualização científica para visualizar a dinâmica dos mercados financeiros. O aluno deverá primeiro agregar a informação, usando ferramentas de webscraping. Depois aplicará métodos de IA (por exemplo clustering e técnicas de aprendizagem não supervisionada) e desenvolverá uma ferramenta que deverá funcionar como um GUI, uma App (desenvolvimento por exemplo em python + kivy ou tkinter) ou um website (html+css+d3js com django +python). O objetivo da ferramenta será o de emitir alertas para comportamentos anormais da evolução das cotações, o que terá de levar em conta com as comportamentos correlacionados.

Figura 1 Apresentação do projeto pelo orientador no ano anterior

Metodologia usada

A metodologia que utilizei para correlacionar as ações foi obter os preços de *open* e *close* para cada um dos dias, na altura utilizei 60 dias. Com estes dados calculei o *percentage change* da *stock* em cada dia.

$$percentage\ change_{i,j} = \frac{close_{i,j} - open_{i,j}}{open_{i,j}} * 100,$$

i representa o os dias (linha da matriz), *j* representa a stock (coluna da matriz)

Fiz isto para todos os dias, agrupava as *stocks* aleatoriamente criando um grupo, e calculava uma matriz que tinha os valores do *percentage change* de cada stock do grupo, ou seja, as colunas das matrizes representariam uma *stock*, e as linhas os dias.

Com esta matriz eu calculava o *percentage change* médio do grupo de cada dia.

$$percentage\ change\ médio\ (\mu_i) = \frac{1}{N-1} \sum_{j=1}^N percentage\ change_j,$$

j representa a stock (coluna da matriz), e *i* representa o dia (linha da matriz)

Depois utilizei o *percentage change* médio de cada dia previamente calculado para calcular os desvios de cada *stock* para cada dia.

$$\sigma_{i,j} = percentage\ change_{i,j} - \mu_i$$

Que por sua vez utilizei para calcular a variância.

$$Var_{i,j} = \sigma_{i,j}^2$$

E calculava a soma das variâncias.

$$Variância\ total = \sum_{i=1}^N \sum_{j=1}^M Var_{i,j}$$

Após isto o que eu fazia era escolher dois grupos X e Y , num dos grupos (X) escolhia a stock com os maiores desvios entre as stocks do grupo a que pertencia, e trocava com uma stock do grupo Y .

Depois tínhamos os seguintes casos que nos interessavam:

Variância Total (X) diminuía e Variância Total (Y) diminuía

Variância Total (X) diminuía e Variância Total (Y) aumentava, mas se:

$$Variância\ Total\ (X)_{atual} - Variância\ Total\ (X)_{antes} > \\ Variância\ Total\ (Y)_{antes} - Variância\ Total\ (Y)_{atual}$$

Isto significaria que a diferença da diminuição tem de ser maior que a diferença do aumento.

O inverso do último caso também interessa, Variância Total (Y) diminuía e Variância Total (X) aumentava, aplicando a mesma condição da diferença da diminuição ter de ser maior que a diferença do aumento.

Caso alguma destas condições se verificasse eu trocava as *stocks* de grupo, e passava para outro grupo.

Por fim era repetir isto até não haver trocas que diminuam a Variância total dos grupos.

Resultados obtidos usando este método

Para um determinado grupo

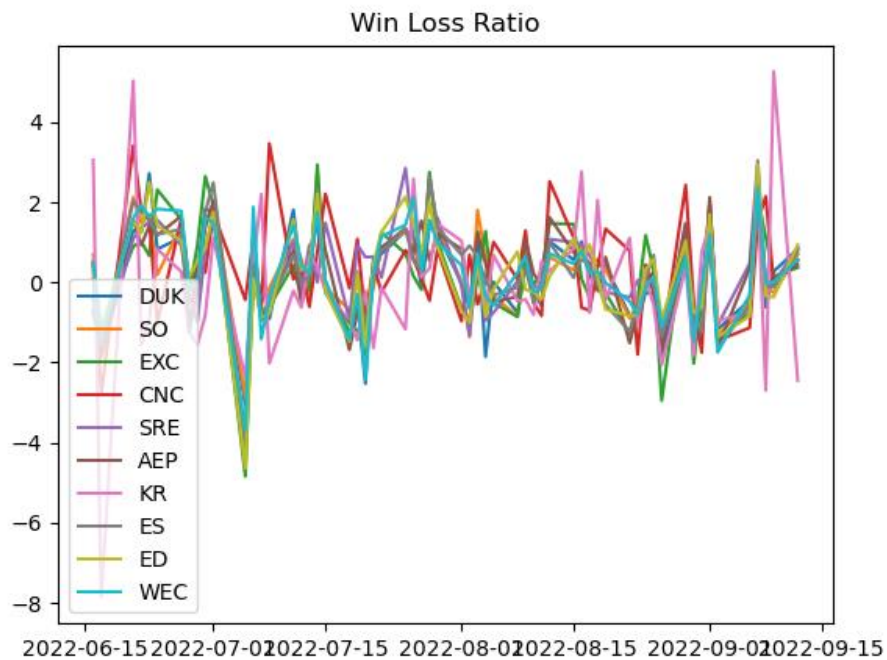


Figura 2 Variações do percentage change para um grupo

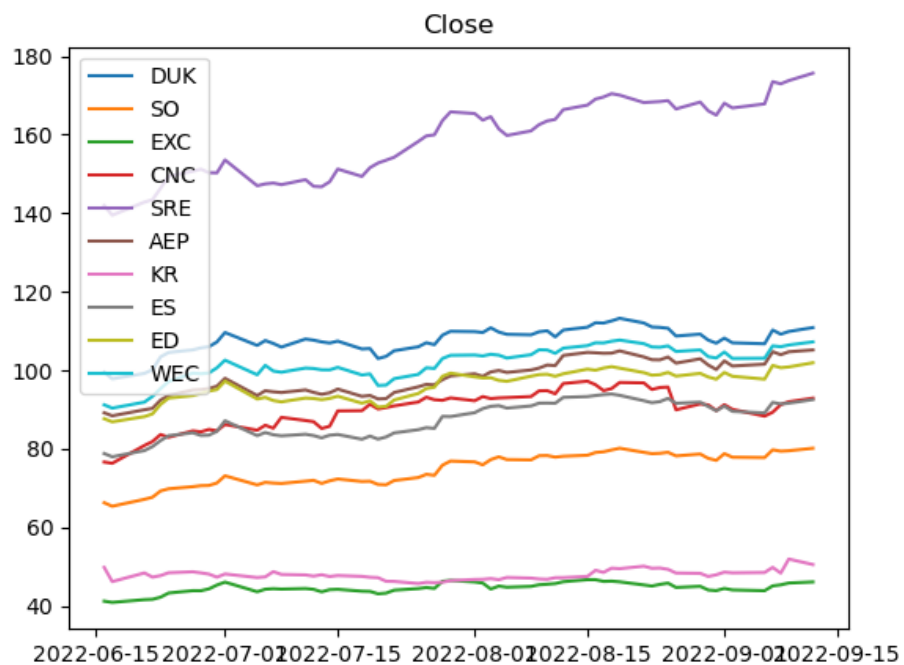


Figura 3 O preço das stocks do grupo no close

O método previamente explicado tem um problema, como podemos analisar na figura 3, existe *stocks* que se comportam da mesma maneira ao longo do tempo, mas existem outras que não. Isto acontece devido aos grupos terem um número específico de ações, e certas ações podem não estar correlacionadas, mas entre os grupos estabelecidos este é aquele que se correlaciona mais. Outro problema que pode surgir é que pode haver um número de stocks que não se correlacionam com nenhum dos grupos e então formam um grupo que tecnicamente em que as stocks não se correlacionam. Por fim existe um outro em que por exemplo posso estar a criar grupos de 10 stocks e haver 14 que se correlacionam, mas como um grupo só pode ter 10 as 4 restantes irão ser atribuídas a outros grupos.

Objetivos

Como está referido na figura 1, inicialmente teremos de fazer a obtenção dos dados, tanto como ponto de partida para correlacionar as stocks como para depois acompanhar a evolução diária das cotações ou outros parâmetros para detetar anomalias em tempo real.

Arranjar métodos para correlacionar as *stocks*, como foi mencionado previamente o método utilizado anteriormente tinha alguns defeitos.

Por fim teria que desenvolver uma aplicação que ajudasse na visualização dos dados, como possivelmente gráficos da evolução das cotações das stocks, visualização de stocks correlacionadas entre outros (dependerá das funcionalidades que poderei implementar e também informação útil para alguém que queira usar a aplicação).

Observações

Pelo que estive a ver na internet, a correlação de stocks é útil para pessoas que tenham o seu portefólio de investimentos. Uma vez que às vezes podem investir em várias ações a pensar que não estão correlacionadas quando na verdade estão. Tendo isto como ponto de partida eu poderia mudar a metodologia dos grupos (inicialmente era considerado um grupo de stocks elevado pelo que a nível computacional torna-se um bocado “pesado” analisar todos os detalhes). Mas se for para uma pessoa que tem um portefólio de cerca de 50 stocks deve ser possível correlacionar as stocks desse portefólio em grupos de tamanho dinâmicos (tamanho dos grupos aumenta se necessário) contrário ao método que foi previamente implementado em que o tamanho dos grupos era estático.

Envio em anexo um ficheiro com os resultados do método que utilizei anteriormente, eu utilizei 500 stocks e fiz grupos de 10. Os grupos estão numerados, dentro da pasta estão algumas informações relativas ao grupo e stocks guardadas em ficheiros de texto, nomeadamente país, setor e industria de cada stock (“stocks information.txt”), assim como percentagens destas informações do grupo (“group information.txt”).