

US 18 - É possível verificar correlações entre notícias e solavancos?

Buscar relação entre o acontecido de notícias que aconteceram em um período imediatamente anterior ao início do solavanco e tentar verificar alguma correlação entre o início do solavanco e a notícia.

1) Coleta e Contagem de Notícias

Após coletarmos os títulos e links de notícias relacionados com a empresa Petrobrás para o jornais online [Estadão](#) e para a [Folha de S.Paulo](#), realizamos uma **contagem da quantidade de notícias** por **dia** e por **jornal**. Assim, foi possível montar uma série temporal de número de notícias para cada fonte de notícias¹.

2) Seleção da série de preços

Selecionamos a série temporal dos preços de fechamento (Preço_Último) das ações da Petrobrás (para o lote padrão). Assim, podemos comparar os valores do preço de fechamento com as notícias do mesmo dia.

3) Pré-processamento das séries

Para ser possível calcularmos a correlação entre as séries temporais realizamos a interpolação das séries. Nos preços de fechamento excluimos os gaps fazendo uma interpolação constante dos valores, ou seja, substituímos os gaps pelo último preço não nulo. Na interpolação das notícias substituímos os dias sem notícia por zero.

Além disso, removemos 8 notícias do Estadão.com.br que aconteceram em 2008 e 2009 e estavam dificultando a análise, pois quase todo os anos fora interpolado com zeros.

4) Detecção de Solavancos

Utilizamos o algoritmo de detecção de solavanco LocalBaseline, considerando uma janela de 15 dias para definir a existência de um solavanco caso a diferença entre os valores seja maior que 95-percentil dessa janela.

5) Cálculo da correlação

¹ A união não é indicada devido vários fatores subjetivos, como os diferentes interesses dos jornais, e objetivos como a diferença dos intervalos de notícias para a empresa.

Para cada ação da empresa (uma empresa tem vários ISINs, ou seja, várias ações) definimos o intervalo de interseção com a série de número de notícias e calculamos as correlações. Foram calculadas correlações entre 3 pares de séries:

- Número de Notícias (interpolada) vs. Preço de Fechamento (interpolado)
 - Chamada de: *plain*
- Número de Notícias (interpolada) vs. Preço de Fechamento (interpolado) apenas para os dias com solavanco
 - Chamada de: *solavanco only*
- Número de Notícias (interpolada) vs. Preço de Fechamento (interpolado) com o peso maior para os dias com solavanco (2x os dias sem solavanco)
 - Chamada de: *solavanco weighted*

6) Interpretação dos resultados

Aparentemente não existe correlação entre o número de notícias e preço de fechamento da cotação por empresa. Os gráficos das séries temporais com os solavancos marcados e os valores das correlações estão no Arquivo 1. No entanto, vemos que a correlação para os dias com solavanco apenas (*solavanco only*) é um pouco mais alta (em módulo) do que a série completa, indicando que deve existir um padrão para a empresa entre as notícias e os dias de solavanco e esse padrão é bem mais acentuado nas quedas.

Arquivo 1. Séries Temporais com Solavancos e Correlações.pdf

Possíveis causas da baixa correlação

Listamos abaixo, em ordem de prioridade, as possíveis causas dessa baixa correlação:

1. Contar o número de notícias não leva em consideração o peso da notícia (positivo ou negativo) para definir o crescimento/decrescimento do preço da ação
2. Falhas no método de busca:
 - a. As notícias podem não ter relação alguma com a variação da bolsa;
 - b. A seleção do caderno do jornal pode não ser a melhor.
3. O retorno da ação (similar a diferença diária de preços) pode ser mais correlacionado com o número de notícias do que o preço em si
 - a. Hipótese fraca: calculamos para a petrobrás e a correlação foi quase nula. No entanto, isso pode ter acontecido pelo problema citado acima, não existe um peso das notícias, por isso as correlações se anularam.
4. O tamanho da janela da detecção de solavanco pode ter negligenciado solavancos
 - a. Hipótese muito fraca: Quase toda a série foi marcada como solavanco, até as pequenas variações.
5. A função de correlação Pearson pode não ser a ideal.

Possível solução

Observar as notícias e o preço da ação **manualmente** em um intervalo de tempo considerável em busca da causa dessa baixa correlação. Idealmente daria-se pesos para as notícias e

reavaliaria a correlação com a nova série temporal das notícias, que poderia ser gerada com a média dos pesos por dia.