

Prova setorial - Economia

Questão 1

Ewerton F M Araujo

1 Passo 1 - Indicador de Reputação da Empresa

Para o cálculo do indicador de reputação de cada empresa j , do setor k , **apenas no dia t** , denominado $\alpha_{j,k,t}$, será utilizada a média ponderada das notas de uma empresa dentro de cada dia. A fórmula matemática será:

$$\alpha_{t,j,k} = \frac{\sum_{i=1}^{n_{t,j,k}} S_{t,j,k,i}}{n_{t,j,k}}$$

onde:

- $\alpha_{t,j,k}$ é o indicador de reputação da empresa j , no setor k , no dia t .
- $S_{t,j,k,i}$ é o score atribuído pelo cliente **na venda i** para a empresa j , no setor k , no dia t .
- $n_{t,j,k}$ é o número de vendas da empresa j do setor k no dia t .

$\alpha_{t,j,k}$ nos dá o indicador de reputação da empresa considerando apenas o período de tempo t . Para obtermos um indicador dinâmico da empresa que, a cada dia, passe a considerar as avaliações mais recentes (dando mais peso a estas mas sem deixar de considerar as mais antigas) utilizaremos a fórmula abaixo:

$$A_{j,k} = \frac{\sum_{t=1}^T (\alpha_{t,j,k} \cdot Y_t)}{T}$$

sendo:

- T é o número de dias onde houve alguma avaliação.
- Y_t é o peso dado a cada dia avaliado. Seu cálculo é:

$$Y_t = \frac{y_t}{\sum_t y_t}$$

$$y_t = K \cdot e^{\frac{-td}{T}}$$

em que **td é o tempo decorrido entre a avaliação e a data de hoje** e K é uma constante (podendo inclusive ser unitária).

Dessa forma teremos tanto o indicador $\alpha_{t,j,k}$ que considera apenas as avaliações do dia quanto $A_{j,k}$, **que considera todos os dias anteriores onde houve avaliação da empresa e atribui maior relevância aos dias mais recentes.**

2 Passo 2 - Indicador Setorial

Para obtermos um indicador para cada setor por dia realizaremos a agregação dos $\alpha_{t,j,k}$ individuais de cada empresa através de uma média ponderada que terá como pesos a relevância da empresa dentro do setor, medida pela quantidade de avaliações por dia. Matematicamente:

$$\beta_{t,k} = \frac{\sum_{j=1}^{m_{t,k}} (\alpha_{t,j,k} \cdot W_{t,j,k})}{m_{t,k}}$$

onde:

- $\beta_{t,k}$ é o indicador de reputação do setor k , no dia t .
- $m_{t,k}$ é a quantidade de empresas do setor k avaliadas no dia t .
- $W_{t,j,k}$ é o peso relativo ao tamanho da empresa dentro do setor (total de vendas da empresa dividido pelo total de vendas do setor no dia t), calculado através da fórmula:

$$W_{t,j,k} = \frac{n_{t,j,k}}{\sum_j^{m_{t,k}} n_{t,j,k}}$$

Como já foi dito anteriormente: $n_{t,j,k}$ é o número de vendas da empresa j do setor k no dia t .

De forma semelhante ao caso do indicador da empresa, agregaremos temporalmente os indicadores setoriais $\beta_{t,k}$ utilizando os pesos diários Y_t .

$$B_k = \frac{\sum_{t=1}^T (\beta_{t,k} Y_t)}{T}$$

lembrando que T representa o total de dias onde houve avaliações.

3 Passo 3 - Indicador Agregado Geral

Para medir a reputação geral em um dia t agregaremos os indicadores setoriais $\beta_{t,k}$

$$\gamma_t = \frac{\sum_{k=1}^{q_t} \beta_{t,k} \cdot Z_k}{q_t}$$

onde:

- γt é o indicador geral de reputação no dia t .
- q_t é a quantidade total de setores avaliados no dia t .
- Z_k é o peso relativo ao tamanho do setor dentro das vendas totais (total de vendas do setor dividido pelo total geral de vendas no dia t), calculado por:

$$Z_k = \frac{n_{t,k}}{n_t}$$

Mais uma vez agregaremos os indicadores que consideram os dias individualmente em um **indicador geral que agrega todos os dias avaliados**:

$$\Gamma = \frac{\sum_{t=1}^T (\gamma_t Y_t)}{T}$$

4 Passo 4 - Incorporando Informações Qualitativas

Para incorporar os *reviews* e suas categorias na análise seria necessário utilizar técnicas relativas a Análise de Sentimento. trata-se de uma área que utiliza análise de texto e linguística computacional para transformar informações qualitativas em dados quantitativos.

O procedimento envolveria inicialmente a limpeza dos dados da coluna "review", removendo observações nulas, caracteres desconhecidos, espaços em branco, pontuações e palavras do tipo "e", "a", "o", entre outras.

Na sequencia, seria necessária a criação de um dicionário de palavras positivas (como: bom, excelente, ótimo, etc); negativas (péssimo, ruim, terrível, entre outras); e neutras. A lógica por trás do uso do dicionário é que clientes satisfeitos tendem a utilizar palavras positivas, enquanto que clientes insatisfeitos costumam criticar a empresa com o uso de palavras negativas.

Com uma grande quantidade de *reviews* à disposição, exceções a este comportamento se tornam pouco relevantes, de forma que uma empresa com muitas palavras positivas é, em geral, mais bem vista pelos clientes.

Por fim, a quantidade de palavras positivas e negativas associadas a uma empresa em determinado dia pode ser quantificada em um indicador de sentimento da empresa no formato:

$$S_t = \frac{\sum Pos_t - \sum Neg_t}{\sum Pos_t + \sum Neut_t + \sum Neg_t}$$

onde:

- $\sum Pos_t$ é a soma de todas as palavras positivas em avaliações da empresa
- $\sum Neg_t$ é a soma de todas as palavras negativas em avaliações da empresa
- $\sum Neut_t$ é a soma de todas as palavras neutras em avaliações da empresa

Caso este modelo de indicador seja seguido, teremos uma medida entre -1 e 1. $S_t > 0$ indica uma avaliação positiva, enquanto $S_t < 0$ seria uma avaliação negativa.

De posse de S_t , poderíamos realizar agregações semelhantes às discutidas nos passos 1 a 3 para obter indicadores setoriais e gerais.