

Análise de sentimentos

Universidade Estadual de Maringá - Ciência da Computação
(Departamento de Informática)

Professor: Igarashi

Alunos

Marcos Henrique Souza Bosco - RA117873

Maria Fernanda Almeida Oliveira - RA118597

Maringá

2023

Introdução

Este relatório detalha um projeto de análise de sentimentos aplicado a notícias relacionadas a empresas, com ênfase na investigação da influência de técnicas de pré-processamento no modelo de análise de sentimentos. Além disso, a análise de correlação é realizada não apenas no mesmo momento da notícia, mas também em defasagens temporais, permitindo uma visão mais completa da relação entre os sentimentos das notícias e os valores das ações.

Etapas de implementação

O projeto inicia carregando informações de notícias de um arquivo YAML denominado 'noticias.yaml', no qual possuem 10 notícias de cada empresa extraídas do site yahoo.finance. Essas notícias estão organizadas por empresa e incluem dados de título, conteúdo e data. Há também uma cópia desse arquivo, denominada noticias_by_time.yaml, que possui as mesmas notícias porém as datas separadas por hora também, para poder fazer a correlação por período.

Os textos das notícias passam por uma etapa de pré-processamento para remover caracteres não alfabéticos e palavras irrelevantes. Após o pré-processamento, a análise de sentimentos é realizada usando a biblioteca `TextBlob`. Ela atribui uma polaridade aos sentimentos das notícias, classificando-os como 'Positivo', 'Negativo' ou 'Neutro'.

Os valores das ações das empresas são coletados por meio da biblioteca `yfinance`. Os dados de ações correspondem ao período de 1º de janeiro de 2023 a 10 de outubro de 2023. O período de análise por período é do dia 05 de outubro de 2023 das 9:30h às 15:30h.

Para cada empresa, a análise calcula a correlação entre os valores de polaridade dos sentimentos das notícias e os valores das ações correspondentes, essa correlação leva em conta a data da notícia e o valor das ações nessa data. Esse cálculo permite determinar se há alguma relação entre os sentimentos expressos nas notícias e o desempenho das ações da empresa.

Abaixo serão apresentadas essas funções e sua respectiva descrição:

```
def process_company_news(company_name, news_list):
    news_data[company_name] = []
    for news_item in news_list:
        news_title = news_item['Título']
        news_content = news_item['Conteúdo']
        news_date = news_item['Data']
        news_data[company_name].append({
            'Título': news_title,
            'Conteúdo': news_content,
            'Data': news_date
        })
```

Imagem 1 - Função que capta as notícias

É projetada para processar notícias de uma empresa específica e armazenar essas notícias em um dicionário chamado `news_data`, associando-as à empresa pelo nome da empresa (`company_name`)

```
def preprocess_text(text):
    text = re.sub('[^a-zA-Z]', ' ', text)
    words = text.lower().split()
    stop_words = set(stopwords.words('english'))
    words = [word for word in words if word not in stop_words]
    return ' '.join(words)
```

Imagem 2 - Função que formata as notícias

Realiza a limpeza e formatação de um texto, removendo caracteres especiais, convertendo-o para letras minúsculas, removendo stop words e retornando o texto processado como uma única string, que pode ser usada para análise de sentimentos ou outras tarefas de processamento de texto.

```
for company, news_list in news_data.items():
    sentiments[company] = []
    stock_data[company] = []
    for news in news_list:
        preprocessed_news = preprocess_text(news['Conteúdo'])
        sentiment = TextBlob(preprocessed_news)
        sentiment_polarity = sentiment.sentiment.polarity
        if sentiment_polarity > 0:
            sentiment = 'Positivo'
        elif sentiment_polarity < 0:
            sentiment = 'Negativo'
        else:
            sentiment = 'Neutro'
        sentiments[company].append({'Título': news['Título'], 'Data': news['Data'], 'Sentimento': sentiment})
```

Imagem 3 - Função que analisa o sentimento das notícias

Esse código processa notícias para várias empresas, realiza uma análise de sentimentos das notícias, classifica os sentimentos como positivos, negativos ou neutros, e coleta dados de ações para essas empresas. Esses dados podem ser usados posteriormente para avaliar a correlação entre os sentimentos das notícias e o desempenho das ações das empresas.

```
for empresa, sentiment_values in sentiment_polarity_by_company.items():
    print(sentiment_values, stock_value_by_company)
    correlation[empresa] = np.corrcoef(sentiment_values, stock_value_by_company[empresa])[0,1]

    print(f"Empresa: {empresa}")
    print("Sentiment Polarity Values:", sentiment_values)
    print("Stock Value Values:", stock_value_by_company[empresa])
```

Imagem 3 - Função que faz a correlação por empresa

Esse código calcula a correlação entre os sentimentos das notícias e os valores das ações para várias empresas, imprime os valores de polaridade dos sentimentos e os valores das ações para cada empresa e exibe a correlação resultante. Essa correlação pode ser útil para avaliar se existe alguma relação entre os sentimentos das notícias e o desempenho das ações das empresas.

Para a correlação por período, as funções são basicamente as mesmas, mas pegando os valores das ações no decorrer do dia 05 de outubro. Verifique melhor no arquivo **t2-correlation2.py**.

Resultados obtidos

Os resultados dos testes consistem em duas partes principais: os valores de polaridade de sentimentos atribuídos a cada notícia e os valores das ações correspondentes. Além disso, é calculada a correlação entre esses dois conjuntos de dados. A correlação é um número que varia de -1 a 1, onde -1 indica uma correlação negativa forte, 1 indica uma correlação positiva forte e 0 indica ausência de correlação.

Aqui estão os resultados da correlação por data para cada empresa:

- Empresa: AAPL
 - Valores de Polaridade de Sentimento: [0.1429805871212121, -0.013777777777777785, 0.09431818181818181, 0.09330909090909088, 0.12327297116029512, 0.08192995529061105, 0.08593358395989972, 0.25766195792511587, 0.008604917095483135, 0.08224448315911731]
 - Valores das Ações: [1.25, 1.25, 1.25, 2.5800018310546875, 2.5800018310546875, 1.25, 1.25, 1.25, 1.25, 1.260009765625]
 - Correlação: 0.08978303077195819

- Empresa: META
 - Valores de Polaridade de Sentimento: [0.15484848484848482, 0.07759398496240602, 0.07083333333333333, 0.5, 0.04545454545454545, 0.06818181818181818]
 - Valores das Ações: [4.639984130859375, -0.78997802734375, -0.78997802734375, -0.78997802734375, -0.78997802734375]
 - Correlação: 0.005710553194267927
- Empresa: MSFT
 - Valores de Polaridade de Sentimento: [0.03827838827838828, 0.14988372093023256, 0.21727272727272728, 0.05526973026973027, 0.09537337662337662, -0.0106060606060606, 0.09204545454545456, -0.0529658284977434]
 - Valores das Ações: [7.9000244140625, 7.9000244140625, 0.399993896484375, 0.399993896484375, 0.399993896484375, 5.569976806640625, -8.40997314453125, 6.04998779296875]
 - Correlação: -0.29986052203792907

Aqui estão os resultados da correlação no período de 05 de outubro para cada empresa:

- Empresa: AAPL
 - Valores de Polaridade de Sentimento: [0.12584595959595957, 0.025151515151515147, 0.10395256916996046, 0.10424242424242423, 0.12360544217687076, 0.07495941558441557, 0.09276061776061773, 0.2546903661310441, 0.005940012368583801, 0.08935249709639953]
 - Valores das Ações: [1.25, 1.25, 1.25, 2.5800018310546875, 2.5800018310546875, 1.25, 1.25, 1.25, 1.25, 1.260009765625]
 - Correlação: 0.10888135185394972
- Empresa: META
 - Valores de Polaridade de Sentimento: [0.15484848484848482, 0.060680272108843546, 0.07083333333333333, 0.5, 0.04545454545454545, 0.06818181818181818]
 - Valores das Ações: [4.639984130859375, -0.78997802734375, -0.78997802734375, -0.78997802734375, -0.78997802734375]
 - Correlação: 0.013517804299261605
- Empresa: MSFT
 - Valores de Polaridade de Sentimento: [0.027777777777777778, 0.15115384615384617, 0.18106060606060606, 0.09301484230055658, 0.08823051948051948, -0.0106060606060606, 0.09204545454545456, -0.05583887657058389]

- Valores das Ações [7.9000244140625, 7.9000244140625, 0.399993896484375, 0.399993896484375, 0.399993896484375, 5.569976806640625, -8.40997314453125, 6.04998779296875]
- Correlação: -0.3400473628326918

Análise dos resultados obtidos

Os resultados indicam que, em muitos casos, a aplicação de técnicas de pré-processamento melhora a qualidade do modelo de análise de sentimentos. Notavelmente, o pré-processamento tende a tornar as polaridades de sentimentos mais distintas, facilitando a identificação de correlações significativas.

Baseado na correlação por data:

- A empresa AAPL apresenta uma correlação positiva moderada (0.0898), indicando que notícias com sentimentos positivos estão associadas a aumentos nos valores das ações.
- A empresa META possui uma correlação próxima de zero (0.0057), sugerindo que não há uma relação clara entre os sentimentos das notícias e os valores das ações.
- A empresa MSFT mostra uma correlação negativa considerável (-0.2999), indicando que notícias com sentimentos positivos podem estar associadas a quedas nos valores das ações dessa empresa.

Baseado na correlação por período:

Empresa: AAPL (Apple Inc.)

- Valores de Polaridade de Sentimento: A média de polaridade de sentimentos das notícias é positiva, com uma média de aproximadamente 0.103.
- Valores das Ações: O valor das ações da AAPL teve uma variação de 1.25 para 2.58 durante o período analisado.
- Correlação: A correlação entre os valores de polaridade de sentimentos das notícias e os valores das ações é positiva, indicando que notícias com sentimentos mais positivos estão associadas a aumentos nos valores das ações. A correlação é de aproximadamente 0.109, o que sugere uma relação moderada.

Empresa: META (Meta Platforms, anteriormente conhecida como Facebook)

- Valores de Polaridade de Sentimento: As notícias relacionadas à META também têm uma média de polaridade positiva, com uma média de aproximadamente

0.09. No entanto, a variação nas polaridades é menor em comparação com a AAPL.

- Valores das Ações: Os valores das ações da META tiveram uma variação considerável, começando em 4.64 e caindo para -0.79 durante o período analisado.
- Correlação: A correlação entre os valores de polaridade de sentimentos das notícias e os valores das ações é próxima de zero (0.014), indicando que não há uma relação clara entre os sentimentos das notícias e o desempenho das ações da META. Isso pode sugerir que outros fatores estão mais influentes.

Empresa: MSFT (Microsoft Corporation)

- Valores de Polaridade de Sentimento: A média de polaridade de sentimentos das notícias relacionadas à MSFT é positiva, com uma média de aproximadamente 0.084.
- Valores das Ações: Os valores das ações da MSFT tiveram uma variação de -8.41 para 7.90, mostrando uma flutuação significativa.
- Correlação: A correlação entre os valores de polaridade de sentimentos das notícias e os valores das ações da MSFT é negativa, indicando que notícias com sentimentos mais positivos podem estar associadas a quedas nos valores das ações. A correlação é de aproximadamente -0.340, sugerindo uma relação negativa considerável.

Conclusão

Este projeto fornece insights valiosos sobre a análise de sentimentos aplicada a notícias de empresas e a correlação com os valores das ações. A utilização de técnicas de pré-processamento pode melhorar a qualidade do modelo de análise de sentimentos, tornando as polaridades de sentimentos mais distintas. Além disso, a análise de correlação em diferentes momentos, incluindo defasagens temporais, fornece uma visão mais completa da relação entre notícias e ações.

É importante ressaltar que a correlação não implica causalidade, e outros fatores podem influenciar os valores das ações. Portanto, ao tomar decisões de investimento com base em notícias, é fundamental considerar uma variedade de informações relevantes. Este projeto oferece uma base sólida para análises mais aprofundadas e pode ser aprimorado com a inclusão de mais dados e o uso de modelos de análise de sentimentos mais avançados.