

Atividade Final - Processamento de Linguagem Natural com Deep Learning | Espec. DL 2023/1:

Análise de Sentimentos e Classificação de Recomendações para o Jogo Dead By Daylight disponíveis na plataforma de jogos Steam.

Julierme Silva e Willian Oliveira

Roteiro

- Motivação e Objetivos
- Coleta dos dados
- Pré-processamento e limpeza dos dados
- Aplicação das técnicas de NLP
- Resultados e Conclusões

Motivação

Este projeto faz parte da avaliação do curso de Processamento de Linguagem Natural com Aprendizado Profundo da especialização em Aprendizado Profundo do CIn-UFPE. O objetivo é utilizar técnicas de **Processamento de Linguagem Natural** (PLN) na análise de avaliações do jogo Dead by Daylight na plataforma Steam. O foco principal é na análise de sentimentos, categorizando as avaliações **Recomendadas** como positivas e as **Não Recomendadas** como negativas. Foi realizado neste trabalho:

- Coleta de todas as avaliações em inglês do jogo Dead by Daylight na plataforma Steam utilizando a API de avaliações da Steam.
- Pré-processamento e limpeza de dados.
- Aplicação de modelos tradicionais e avançados de PLN:
 - SVM com Bag of Words
 - SVM com Word2Vec
 - BERT
 - GPT-2.
 - Aprendizado em Contexto (GPT-4)

Coleta dos dados

- Scraping realizado diretamente pelo json gerado no 'appreviews' da Steam.
- 173236 comentários avaliativos coletados.
- As avaliações já vieram binarizadas para a variável Recommended.

Exemplos extraídos da base:

review_id, review_text, recommended

136009306,this game is going downhill.,False

136009209,Nice game!,True

136009145,anxiety simulator. very fun. scary. amazing dlc >:3,True

Pré-processamento e limpeza dos dados

Para os modelos baseados em SVM, foi realizado um pré-processamento tradicional construído em 5 etapas:

- Remoção de caracteres especiais e transformação do texto em minúsculo.
- Tokenização do texto.
- Remoção de palavras comuns (stopwords).
- Lematização das palavras.
- Junção das palavras novamente em uma string.

Os modelos BERT e GPT-2 possuem tokenizadores e funções de pré-processamentos próprias que são disponibilizadas por meio do HuggingFace e foram utilizadas tendo como input os dados originais.

Pré-processamento e limpeza dos dados

Como nosso conjunto de dados está desequilibrado ($\approx 81\%$ Positivo), usaremos o **RandomUnderSampler** da biblioteca **Imbalanced-Learn** para reduzir aleatoriamente a quantidade de avaliações positivas em nossos conjuntos de treinamento, obtendo avaliações perfeitamente balanceadas (50% por classe). Os conjuntos de Validação e Teste, quando usados para avaliação do modelo, não foram balanceados.

Todos os modelos e splits do dataset também foram realizados com seed global definido em 0, para garantir reprodutibilidade do código apresentado no github.

Aplicação das técnicas de NLP

Primeira abordagem: SVM com Bag of Words

Nós juntamos os conjuntos de treinamento e validação para realizar uma validação cruzada usando o SVC e pipelines do Scikit-Learn. Nesse caso, obtivemos **59248** amostras de treinamento e **17324** amostras de teste. Em nossos testes de ajuste, descobrimos que o kernel Sigmoid demonstrou convergir mais rapidamente e alcançar melhor desempenho do que o linear e o rbf para este problema de classificação.

Hiperparâmetros:

```
kernel="sigmoid"  
tol=1e-3  
max_iter=-1
```

Métricas em teste:

```
Accuracy: 0.66  
Precision: 0.88  
Recall: 0.68  
F1: 0.77  
ROC AUC: 0.65
```

Aplicação das técnicas de NLP

Segunda abordagem: SVM com Embeddings (Word2Vec)

Novamente, juntamos os conjuntos de treinamento e validação para realizar uma validação cruzada usando o SVC e pipelines do Scikit-Learn. Nesse caso, obtivemos **59248** amostras de treinamento e **17324** amostras de teste. Em nossos testes de ajuste, também descobrimos que o kernel rbf se mostrou melhor do que outras opções para este problema de classificação.

Hiperparâmetros:

```
kernel="rbf"  
tol=1e-3  
max_iter=-1
```

Métricas em teste:

```
Accuracy: 0.58  
Precision: 0.93  
Recall: 0.52  
F1: 0.67  
ROC AUC: 0.68
```


Aplicação das técnicas de NLP

Terceira abordagem: Bidirectional Encoder Representations from Transformers (BERT)

Para esta abordagem, não juntamos os conjuntos de treinamento e validação, pois não estávamos usando métodos de validação cruzada devido à complexidade do modelo e aos altos custos computacionais. O conjunto de treinamento balanceado tinha **46178** avaliações, e os conjuntos de validação e teste tinham **25985** avaliações. Nesse caso, utilizamos o tokenizador pré-treinado e as configurações do BERT da versão bert-base-cased disponível na HuggingFace para performar um pequeno tuning no modelo.

Hiperparâmetros:

```
optimizer="Adam"  
learning_rate=2e-5  
loss="sparse_categorical_crossentropy"  
epochs=3,  
batch_size=16
```

Métricas em teste:

```
Accuracy: 0.86  
Precision: 0.96  
Recall: 0.86  
F1: 0.91  
ROC AUC: 0.93
```

Aplicação das técnicas de NLP

Quarta abordagem: Generative Pre-Trained Transformers (GPT-2)

Nesta abordagem, como se trata de um modelo generativo, ele não é adequado diretamente para classificações, mas concatenamos o texto da avaliação com as etiquetas para que o modelo pudesse aprender os contextos em que a avaliação recomendava ou não o jogo. Seguimos os mesmos passos que fizemos nas configurações do BERT, mas neste caso usamos o GPT2Tokenizer e a versão pré-treinada do gpt-2, também disponível na biblioteca HuggingFace transformers.

Como este é um modelo grande, levou quase **3 horas** para realizar uma **única época** de treinamento, usando o **otimizador AdamW** com uma **taxa de aprendizado muito pequena ($5e-4$)** com um agendamento linear que usa warmup.

Com apenas esta época, o modelo foi capaz de **gerar avaliações que não estavam presentes** em nenhum conjunto de dados e classificá-las de maneira razoável:

- Review: Great game, especially the killer. But they can make it super difficult at times > Recommended
- Review: Pretty good job so far > Recommended
- Review: This game is so bad for survivors that I can't join game. There's so many bugs and glitches. The game doesn't work at all. If you're looking for an even remotely enjoyable experience, I really think the game is for you > Not recommended

Aplicação das técnicas de NLP

Quarta abordagem: Generative Pre-Trained Transformers (GPT-4)

Também realizamos o aprendizado em contexto com métodos de prompt usando o GPT-4 por meio do [ChatGPT](#).

Nessa abordagem demos dois exemplos rotulados de classificações. Sendo uma Recomendada e outra Não Recomendada e solicitamos que o modelo realizasse a classificação de outros 7 comentários.

O modelo acertou todas as classes originais dos exemplos apresentados.

Resultados e conclusões:

Métodos de aprendizado profundo com transferência de aprendizagem foram capazes de obter resultados muito melhores do que outros métodos e alcançar um bom desempenho de forma consistente usando tanto o BERT quanto o GPT-2 e GPT-4 (prompt), o que era um resultado esperado, já que esses modelos pré-treinados têm muito mais parâmetros e são treinados em um número muito maior de tokens para obter um melhor desempenho.