



Eduardo Dias | Hugo Rocha | Simão Quintela | Tiago Monteiro

Análise de sentimento do discurso em transmissões televisivas de jogos de futebol

Índice

1	Objetivos do projeto
2	Base de dados SoccerNet e tratamento de dados 4-:
3	Benchmarking dos modelos Whisper
4	Construção dos nossos próprios datasets
5	Classificação Zero-Shot
	5.1 Vantagens e desvantagens
	5.2 Aplicação do método 1
	5.3 Métricas utilizadas 1
	5.4 Resultados
6	Classificação Few-shot
	6.1. Vantagens de modelos Set-Fit
	6.2. Estratégia e métricas utilizadas 1
	6.3. Resultados
7	Fine-tuning do modelo BERT
8	Conclusões e trabalho futuro

Objetivos do projeto

• Treinar e utilizar modelos de *deep learning* de classificação de texto para detetar momentos de destaque em transmissões TV de jogos de futebol.

Procedimentos realizados:

- Obtenção do *dataset* de vídeos de jogos de futebol;
- Extração do áudio desses vídeos;
- Fragmentação, tradução e transcrição do áudio resultante para texto;
- Classificação desses momentos dos jogos como relevantes ou não relevantes
- Análise de resultados;

Base de dados SoccerNet

- Composta por vídeos com relato, em várias línguas, de 550 jogos de futebol divididos em primeira e segunda parte;
- Numa fase inicial, considerámos apenas 4 jogos generalizando o trabalho feito para um qualquer número de jogos;
- Posteriormente, selecionamos 10 jogos da base de dados para trazer maior fiabilidade aos resultados;

Tratamento de dados

- Extração do áudio de cada um dos vídeos utilizando a biblioteca *ffmpeg* do *Python*, ficando com 2 ficheiros de áudio por jogo (um por cada parte);
- Fragmentação de cada um desses ficheiros de áudio em partições de cerca de 30 segundos cada (cerca de 90 a 100 fragmentos por parte);
- Garantia de continuidade no áudio
- Transcrição (com tradução para inglês) dos fragmentos utilizando o modelo Whisper;

Benchmarking dos modelos Whisper

jogo	model	run_time_per_game
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-tiny	7:52
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-base	13:34
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-small	34:52
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-medium	85:36
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large	173:11
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large-v2	187:16
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large-v3	176:02

jogo	model	run_time_per_game
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-tiny	2:55
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-base	3:24
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-small	4:23
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-medium	6:48
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large	10:49
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large-v2	10:05
2015-02-21 - 18-00 Chelsea 1 - 1 Burnley	openai/whisper-large-v3	9:42

Modelos testados em CPU

• Modelos testados em GPU do Google Colab

Construção dos nossos próprios datasets

De forma a consolidar todos os fragmentos dos vários jogos (transcrições) com a respetiva classificação do *SoccerNet* foram criados dois *datasets* com oito colunas cada, sendo que sete destas são:

- id
- league
- game
- fragment_start_time
- fragment_end_time
- commentary
- model_id

A oitava coluna contém as labels de classificação dos fragmentos.

No dataset principal essas labels são **Highlight** (correspondente às labels **Shots on target**, **Goal**, **Penalty e Red card** no SoccerNet) e **Non-Highligth** (correspondente a todas as outras labels do SoccerNet) e o dataset ficou com 2067 registos, dos quais 1915 estão associados à classe **Non-Highlight** e 152 à classe **Highlight**.

No dataset secundário a oitava coluna é uma string que contém as várias labels possíveis de classificação dos respetivos fragmentos (Exemplo: Corner/Shots on target/Ball out of play). Este dataset ficou com 1465 registos.

Classificação Zero-shot

- Tarefa de processamento de linguagem natural
- O modelo é pré-treinado num conjunto de dados rotulados
- O modelo é capaz de classificar novos exemplos de classes que não foram previamente vistas.

Zero-shot: Vantagens e desvantagens

Vantagens:

- Flexibilidade
- Aplicações práticas
- Generalização

Desvantagens:

- Desempenho inferior
- Dependência de boas representações semânticas
- Complexidade computacional

Zero-shot: Aplicação do Método

O modelo pré-treinado utilizado foi o bart-large-mnli.

Para a sua implementação usamos 3 estratégias:

- 2 labels Highlight e Non-Highlight;
- 2 labels Relativamente a Highlights e Non-Highlights, junção das labels vindas do SoccerNet numa só label. (Exemplo da label referente a Highlight: "Shots on target, Goal, Penalty, Red card");
- 6 labels três referentes a Highlights (Shots on target, Goal e Red card) e outras três referentes a Non-Highlights (Ball out of play, Clearance e Corner).

Zero-shot: Métricas utilizadas

Foram utilizadas as seguintes métricas:

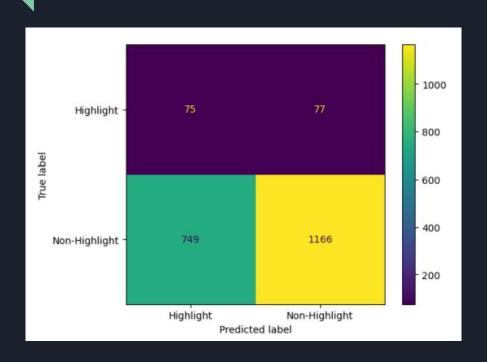
$$\frac{TP + TN}{TP + TN + FP + FN}$$

$$\frac{TP}{TP + FP}$$

$$\frac{TP}{TP + FN}$$

$$\frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Zero-shot: Primeira abordagem



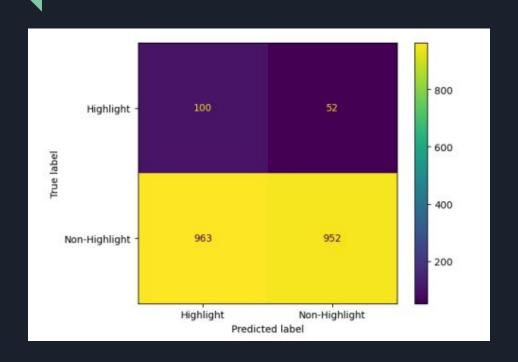
• Accuracy: 60.04%

• **Precision: 9.10%**

• Recall: 49.34%

• F1-Score: 15.37%

Zero-shot: Segunda abordagem



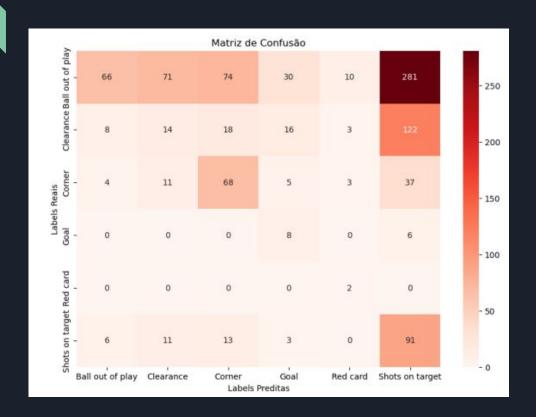
• *Accuracy*: 50.09%

• *Precision:* 9.41%

• Recall: 65.79%

F1-Score: 16.46%

Zero-shot: Terceira abordagem



Accuracy: 25.38%

• Precision: 16.37%

• Recall: 31.86%

• F1-Score: 21.63%

Classificação Few-Shot

- Tarefa de processamento de linguagem natural
- O modelo é treinado num conjunto de exemplos rotulados
- É fornecida uma certa quantidade de exemplos rotulados dos novos dados ao modelo de forma a que ele seja capaz de utilizar estes dados limitados e generalizá-los efetivamente de forma a classificar os novos dados.

Modelos Set-Fit: vantagens

Os modelos *SetFit* oferecem uma abordagem eficiente para o treino de modelos com poucos exemplos.

- Vantagens
- 1. Abordagem sem *prompt*;
- 2. Modelos menores e mais rápidos;
- 3. Desempenho em ambiente de alta performance;

Estratégia e métricas utilizadas

O modelo pré-treinado escolhido foi o *all-MiniLM-L6-v2*

Aplicamos o modelo sobre o *dataset* principal, que contém as *labels* binárias *Highlight* e *Non-Highlight*, para treino e teste deste modelo utilizando várias combinações dos mesmos.

Few-shot: Resultados

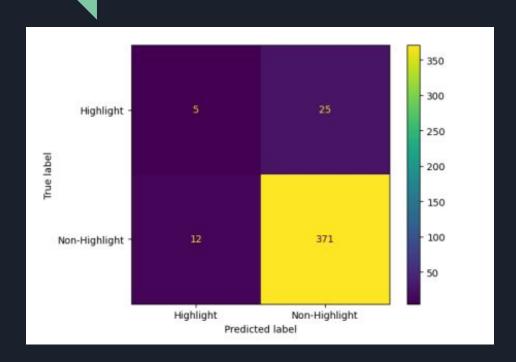
	COMBINAÇÃO 1	COMBINAÇÃO 2	COMBINAÇÃO 3	COMBINAÇÃO 4
Treino	78	270	400	1654
Teste	1987	1797	1667	413

Few-shot: Resultados

	COMBINAÇÃO 1	COMBINAÇÃO 2	COMBINAÇÃO 3	COMBINAÇÃO 4
Highlight	15.5%	7.66%	14.71%	21.28%
Non-Highlight	72.90%	80.48%	91.6%	95.26%
Overall Accuracy	58.93%	67.67%	84.7%	91.04%

Few-shot: Resultados da última combinação

teste: 30



<u>Higniignt</u> :	<u>Non-Hignlight</u> :
Número de exemplos para treino: 122	Número de exemplos para treino: 1532
Número de exemplos para	Número de exemplos para

teste: 383

	Precision	Recall	F1-Score	Accuracy
Highlight	29.41%	16.67%	21.28%	
Non-Highlight	93.70%	96.87%	95.26%	
Overall				91.04%

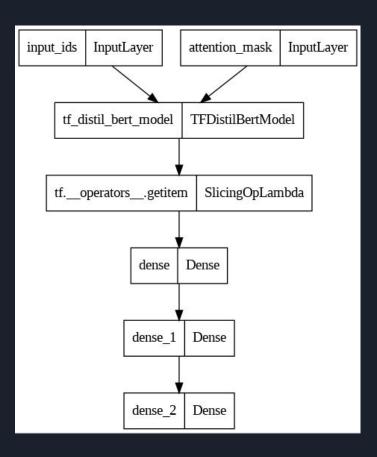
Fine-Tuning de um classificador

- Ajuste de um modelo pré-treinado para realizar tarefas específicas
- Redução substantiva no tempo de treino
- BERT Bidirectional Encoder Representation from Transformers
- Distilbert Versão destilada do BERT

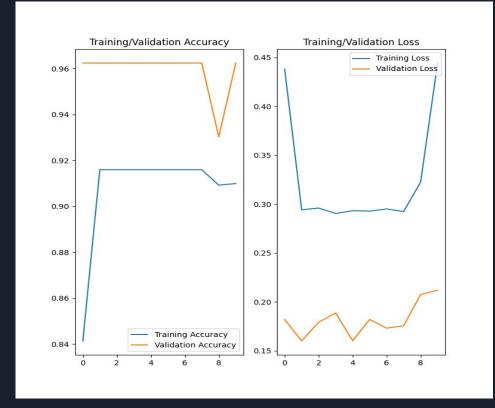
Fine-Tuning do Distilbert

- Obter output do Transformer
- Adicionar MLP no topo do Transformer
- Realizar treino supervisionado
- Obter e analisar resultados

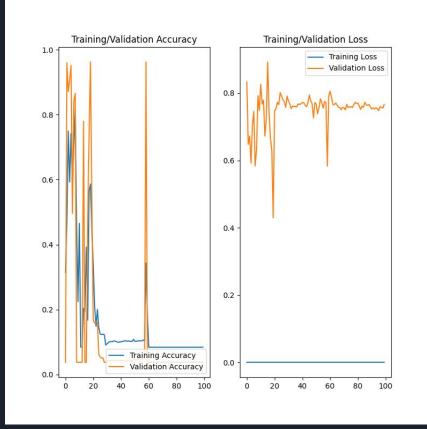
Arquitetura da rede



- Épocas: 10
- Batch size: 64
- Loss: Categorical Cross Entropy
- Otimizador: ADAM
- 20% dados para validação



- Épocas: 100Batch Size: 64
- Loss: Categorical Cross Entropy
- Otimizador: ADAM
- 20% dados para validação
- Com pesos de classes



Conclusões e trabalho futuro

- Consideramos os resultados obtidos através dos classificadores zero-shot e few-shot relativamente bons
- Os resultados do fine-tuning de um classificador tradicional ficaram um pouco aquém das nossas expectativas
- O projeto permitiu-nos aplicar e desenvolver os conhecimentos adquiridos ao longo do mestrado e obter novas competências

- Aplicação de um modelo diferente na tarefa de transcrição
- Aprofundar o fine-tuning realizado
- Utilizar a totalidade da base de dados



Projeto Integrado em Matemática e Computação

Eduardo Dias | Hugo Rocha | Simão Quintela | Tiago Monteiro

Análise de sentimento do discurso em transmissões televisivas de jogos de futebol