INSTITUTO DE INFORMÁTICA
Universidade Federal de Goiás

Deep Learning Brasil at ABSAPT 2022: Portuguese Transformer Ensemble Approaches

Juliana Resplande Sant'Anna Gomes
Eduardo Augusto Santos Garcia
Adalberto Ferreira Barbosa Junior
Ruan Chaves Rodrigues
Diogo Fernandes Costa Silva
Dyonnatan Ferreira Maia
Nádia Félix Felipe da Silva
Arlindo Rodrigues Galvão Filho
Anderson da Silva Soares

juliana.resplande@discente.ufg.br

INSTITUTO DE INFORMÁTICA
UFG

UFG
UNIVERSIDADE FEDERAL DE GOIÁS

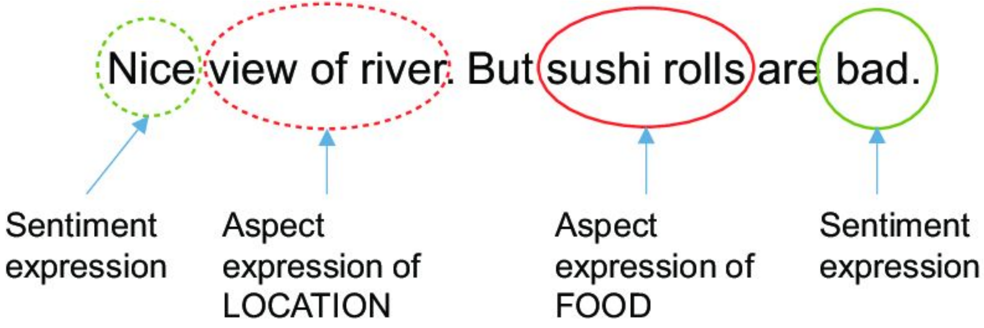
Schedule

- Introduction
- Related work
- Dataset
- **Methodology**
- **Experimental setup**
- **Results**
- Conclusion

Introduction

Aspect-based Sentiment Analysis (ABSA)

- **Fine-grained** approach to **sentiment analysis**
- **Instead** of classify an **entire sentence** under a single sentiment
- Classify the **individual sentiment** polarity of **all tokens** that make a significant polarity
- In the task terminology, these **tokens** are called the **aspect** terms of a sentence



ABSAPT 2022

- **Aspect-Based Sentiment Analysis in Portuguese 2022** at IberLEF 2022
- Inspired on the format of SemEval-2014 Task 4
- Two subtasks
  - **Aspect Term Extraction (ATE)**: given a set of sentences, the task is to **identify all aspect terms** present in each sentence
  - **Sentiment Orientation Extraction (SOE)**: given a set of sentences that have already been annotated for their aspect terms, the task is to **determine the sentiment polarity of each aspect term** (positive, negative or neutral).

Related work

ATE ABSITA

- ATE ABSITA
  - EVALITA 2020 shared task on **Aspect Term Extraction** and **Aspect-Based Sentiment Analysis**
- First and second-ranked team approach
  - **ATE** as a **Named Entity Recognition** task
  - Fine-tuning **SOTA Transformer models** on the training data for the task
  - We follow a **similar approach during our participation** at ABSAPT 2022

- The **first-ranked team** framed the second task as a problem of text classification, under the premise that the portion of the **text that surrounds each aspect should have the same overall sentiment as the aspect itself**.
- Although **we also experimented with this approach**, our best performing system at the SOE subtask framed it as a **text generation problem, similar to what was done by Zhang et al. [8] and Chebolu et al.** [9].

Dataset

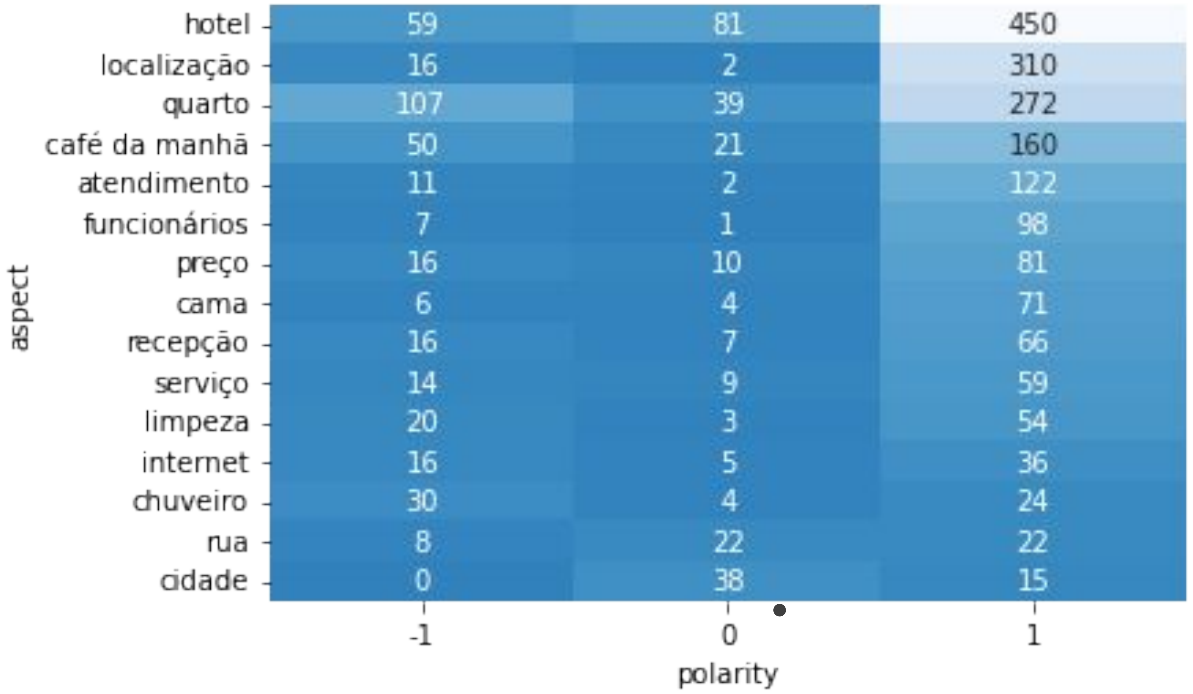
The dataset

- The dataset was taken from TripAdvisor
- The data was in the following format:

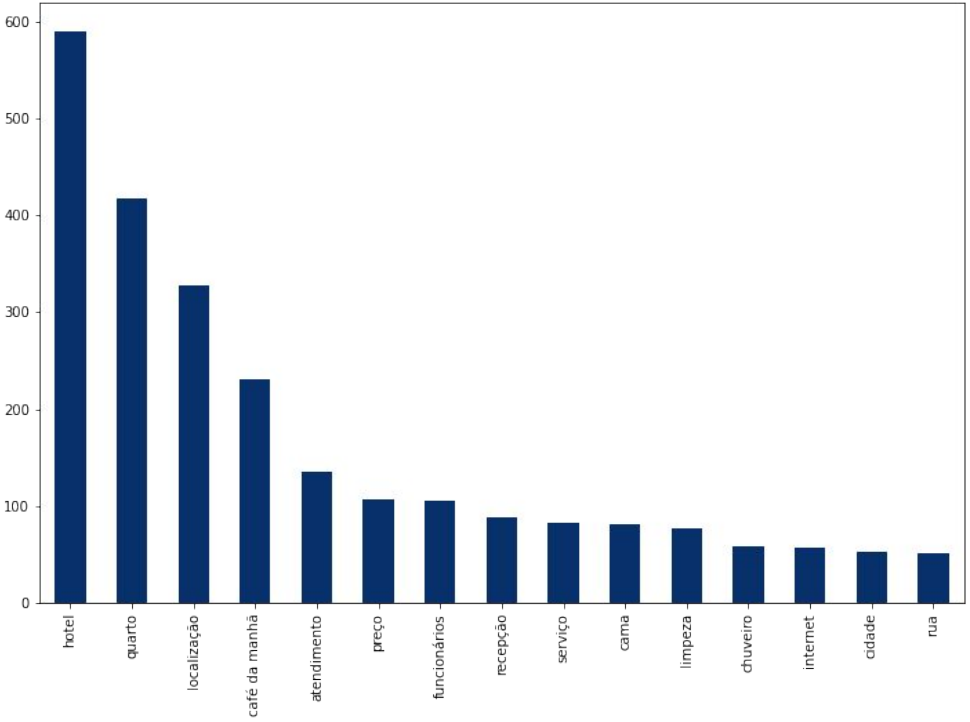
Id	Review	Polarity	Aspect	Start pos	End pos
2414	Hospedei-me em maio nesse hotel pela terceira vez ...	1	hotel	26	31

Aspect polarity heat map from top 15 occurrences.

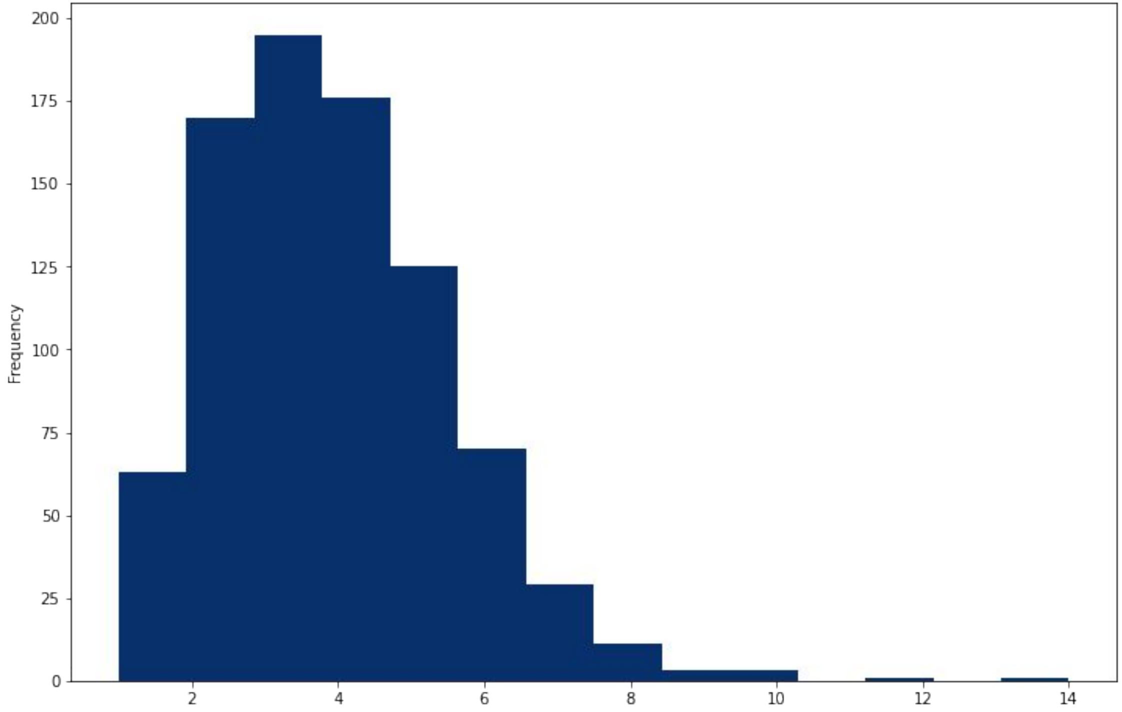
- The polarity is imbalanced towards 1



Top 15 aspect occurrences



Aspect frequency distribution



Methodology

Task 1 - ATE: Review Preprocessing

- For the Aspect Term Extraction task, we build a NER dataset converting the original train data to the BIO/IOB format (Inside, Outside, Beginning), a common tagging format where each token can be classified with the prefixes B, I and O.

A	estrutura	do	hotel	é	muito	boa,	mas	o	café	da	manhã	deixa	a	desejar.
O	O	O	B-ASPECT	O	O	O	O	O	B-ASPECT	I-ASPECT	I-ASPECT	O	O	O

Task 1 - ATE: Training Strategy

- Review preprocessing
  - Transform in to a Named Entity Recognition task
  - Each sentence can have more than one Aspect
- Transformer model
  - Bertimbau
  - mDeBERTav3 base
  - RoBERTa base Portuguese
- Dataset configuration
  - Whether to **use external data** using others ABSA datasets
  - Evalita (Italian), MAMS (English), Semeval (English, Spanish, Arabic, Chinese, …)

Task 2 - SOE: Training Strategy

- Review preprocessing
  - Whether **only the relevant portion** of the review to the aspect term
- Transformer model
  - Transformer models **with Portuguese support**.
  - Bertimbau (base and large), PPT5 (base and large)
  - mDeBERTa base, XLM-RoBERTa base
  - LaBSE, Canine-c, RemBERT
- Dataset configuration
  - Whether to **use external data** using B2W dataset and ABSA datasets

Review preprocessing example

- Aspect: Quarto
- Review

O hotel é perto de todos os pontos principais de Paris, cercado de estações de metrô, dá para fazer tudo a pé. O Hotel é charmoso, aconselho ficar nos quartos no piso superior, perto de telhado, típica arquitetura parisiense. O café da manhã é simples, mas na medida certa, tudo muito gostoso, croissant sensacional! O preço é um pouco salgado, mas se puder vale muito a pena.

Two approaches

1. **Hole review** and **aspect term** as input: **pair classification**

   ‘O hotel é perto …. vale muito a pena.’ **+** “quarto”

- Seq2seq

  “Review: O hotel é perto …. vale muito a pena. Aspect: quarto“

- Encoder

  “O hotel é perto …. vale muito a pena [SEP] quarto“

2. Select the **relevant part in review** for the **aspect term** as input: single sentence classification

aconselho ficar nos quartos no piso superior

Experimental Setup

Experimental setup

- Two V100 GPUs (32 GB) were employed, one for each task.
- Training code is available on https://github.com/ju-resplande/dlb_absapt2022

Task 1 - ATE: training subsets

- In order to **assess the best training configuration** explained in Methodology, we **divide the original training set** in (70% / 30%)

- **Including the external data,** we create two subsets for the training:
  - **Portuguese subset:** 70% random split of the original training-set of the competition dataset.
  - **Multilingual subset**: All external data from Evalita, MAMS, Semeval, plus the *Portuguese subset*.

Task 2 - SOE: training subsets

- In order to **asses the best training configuration** explained in Methodology, we **divide the original training set** in (70% / 30%)
- **Using three approaches**, generating different training subsets:
    - **subset 1**: **Random (naive)** approach. The training set for the shared task is arbitrarily split into a new training and validation set.
    - **subset 2**: We are careful to keep the same **proportion of reviews of each polarity** on each
    - **subset 3**: Besides polarity, we **also** try to keep the same **ratio of aspect terms** between the splits.

Task 2 - SOE: training details

1. **Best model in subset assessment:** PTT5 Large without cutting the original review and without external data;
2. Fine-tune it under **four** distinct combinations of **learning rate** and **random seed**:
   - {3e-4, 7}, {1e-4, 5}, {5e-5, 10}, {3e-5, 8}
3. Produce the final submission through a **majority voting ensemble**

Results

Task 1 - ATE

Table 3
Best results on the ATE task. The symbol {MAMs, Evalita, Semeval}* refers to mDeBERTa previously trained on external datasets, explained in 4.1.2.

model	external data	acc.	precision	recall	f1
BERTimbau base	-	98.2	78.1	87.8	82.6
RoBERTa PT base	-	98.4	80.8	90.7	85.5
mDeBERTa base	MAMs, Evalita, Semeval	98.4	79.1	**94.0**	**85.9**
mDeBERTa base	{MAMs, Evalita, Semeval}*	**98.5**	**81.4**	90.1	85.5

Table 4
Competition final results for the Task 1 (ATE).

team_name	acc
TeamDeepLearningBrasil	**67.1448**
Teampiln	65.4974
TeamUFSCAR	59.3715
TeamPeAm	33.8243
TeamMachadoPardo	22.1050
TeamUFPR	17.1908
TeamOwl	2.6265

Best results on training subsets. We also include zero-shot results obtained with GPT-3 [15] . The symbol {MNLI, XNLI}* refers to mDeBERTa previously trained on MNLI and XNLI, explained in 4.2.4.

subset	model	external data	acc.	f1	f1(pos)	f1(neu)	f1(neg)
subset 1	PTT5 large	-	**86.8**	**78.8**	**92.6**	62	**81.7**
	PTT5 large	target swap	86.3	78.3	92.4	63	79.6
	PTT5 base	target swap	86.3	78.2	92.6	**63.1**	78.8
	GPT-3	-	80.0	66.0	90.0	36.0	73.0
subset 2	mDeBERTa base	{MNLI, XNLI}*	**85.1**	75.9	**92.4**	**57.1**	78.2
	mDeBERTa base	MAMs, Evalita, Semeval	84.6	**76.6**	91.6	55.2	**83**
	mDeBERTa base	-	83.3	73.5	91.2	52.1	77.1
subset 3	PTT5 large	-	**77.4**	**75.6**	**82.3**	60.6	**83.8**
	mDeBERTa base	MAMs, Evalita, Semeval	76.9	**75.6**	82	**62**	82.8
	mDeBERTa base	{MNLI, XNLI}*	74.2	72.1	81.6	58.1	76.7

Task 2 - SOE

Table 6
Competition final results for the Task 2 (SOE).

team_name	bacc	f1	precison	recall
TeamDeepLearningBrasil	**82.3756**	**81.7988**	**81.3144**	**82.3756**
Teampiln	78.8619	77.4794	76.5911	78.8619
TeamUFSCAR	62.8992	61.2248	65.5697	62.8992
TeamPeAm	62.8992	61.2248	65.5697	62.8992
TeamUFPR	62.8992	61.2248	65.5697	62.8992
TeamOwl	53.5995	57.2803	68.9396	53.5996

Conclusion

- **Best performing system** on ABSAPT 2022 at IberLEF 2022
    - new SOTA results on **both** ATE and SOE tasks.
- System
    - For ATE, an ensemble of RoBERTa and mDeBERTa’s trained in Portuguese and multilingual datasets
    - For SOE, a voting ensemble of PTT5 large without external data.
- In future work
    - Experiment with a **coreference resolution** to improve ATE
    - Fine-tuning Transformer models for **both subtasks simultaneously** through a **multi-task learning framework**

Thank you

CAPES

INSTITUTO DE INFORMÁTICA UFG

UFG UNIVERSIDADE FEDERAL DE GOIÁS