

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Deteccção de sarcasmo em redes sociais  
utilizando DeBERTa**

Lucas Paiolla Forastiere

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE  
FORMATURA SUPERVISIONADO

Supervisor: Prof. Dr. Ricardo Marcondes Marcacini

São Paulo  
2017

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0  
(Creative Commons Attribution 4.0 International License)*

## Resumo

Lucas Paiolla Forastiere. **Detecção de sarcasmo em redes sociais utilizando DeBERTa**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

[illegible]

**Palavras-chave:** Palavra-chave1. Palavra-chave2. Palavra-chave3.



# Abstract

Lucas Paiolla Forastiere. **Sarcasm detection in social media using decoding-enhanced BERT with disentangled attention**. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2017.

[illegible]

**Keywords:** Keyword1. Keyword2. Keyword3.



## Lista de Abreviaturas

CFT	Transformada contínua de Fourier ( <i>Continuous Fourier Transform</i> )
DFT	Transformada discreta de Fourier ( <i>Discrete Fourier Transform</i> )
EIIP	Potencial de interação elétron-íon ( <i>Electron-Ion Interaction Potentials</i> )
STFT	Transformada de Fourier de tempo reduzido ( <i>Short-Time Fourier Transform</i> )
ABNT	Associação Brasileira de Normas Técnicas
URL	Localizador Uniforme de Recursos ( <i>Uniform Resource Locator</i> )
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo

## Lista de Símbolos

$\omega$	Frequência angular
$\psi$	Função de análise <i>wavelet</i>
$\Psi$	Transformada de Fourier de $\psi$

## **Lista de Figuras**

## **Lista de Tabelas**

## **Lista de Programas**



# Sumário

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Fundamentos e Trabalhos Relacionados</b>	<b>3</b>
2.1	Detecção de Sarcasmo . . . . .	3
2.2	Trabalhos Relacionados . . . . .	4



# Capítulo 1

## Introduction

Sarcasm detection is an important aspect of many natural language processing (NLP) systems, with many implications in natural language understanding, dialog systems, and data mining. However, sarcasm detection is difficult because it is infrequent in many conversations and, many times, it is difficult even for humans to discern.

Many studies have been made in the area and many datasets have been proposed with either *balanced* or *unbalanced* data. Also many of these datasets use humans to annotate sarcastic statements.

In this paper, we use the Self-Annotated Reddit Corpus (SARC), which is a large corpus for sarcasm detection created using Reddit posts to get labels automatically, to train, evaluate, and compare many NLP models.



## Capítulo 2

# Fundamentos e Trabalhos Relacionados

### 2.1 Detecção de Sarcasmo

De acordo com o dicionário Dicio, sarcasmo é uma zombaria que busca ofender, enquanto ironia é a ação de dizer o oposto do que se deseja expressar. Ainda segundo ele, a diferença entre esses dois termos se dá no fato de que sarcasmo é um dito ácido que pode ou não ser expresso por meio de uma ironia e essa, por sua vez, pode ou não ser utilizada para ofender.**dicio\_sarcdicio\_irony**

(H. Paul Grice. 1975. Logic and Conversation)

(Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcin)

O termo detecção de sarcasmo refere-se à determinação de se há ou não sarcasmo em uma porção de texto verbal. E o termo detecção automática de sarcasmo refere-se a métodos computacionais de se resolver o problema acima mencionado. Computacionalmente, podemos definir esse problema como uma **classificação binária**, termo explicado mais a frente.

Entretanto, na literatura é bastante comum também se definir detecção de sarcasmo como a determinação de se há ou não sarcasmo ou ironia verbal em uma porção de texto verbal. Portanto, em geral, ao se falar de sarcasmo, a literatura engloba tanto sarcasmo quanto ironia como se fossem a mesma coisa. Este texto também não fará discriminação entre sarcasmo ou ironia.

Em geral, esse problema é difícil, pois, por vezes, nem humanos conseguem perceber essa figura de linguagem (e.g. “*Oba, hoje está tão ensolarado, que vontade de ir para a escola.*”). Além disso, o contexto tende a importar muito. Muitas características externas ao texto podem servir para discriminar se uma pessoa está ou não sendo sarcástica. Entre alguns exemplos estão intonação, locutor, interlocutor, conhecimento prévio sobre a fala, tempo e espaço em que se fala e elementos não verbais.

(Humans Require Context to Infer Ironic Intent (so Computers Probably do, too))

O problema de detecção de sarcasmo pode ser tratado como uma classificação binária. Esse tipo de problema é muito estudado no campo do aprendizado de máquina e envolve classificar os elementos de um conjunto em dois grupos chamados de classes. Por classificar entende-se a determinação de um elemento do conjunto entre pertencente à primeira ou segunda classe.

No caso da detecção de sarcasmo, os elementos são os pedaços de texto e as duas classes são *ser sarcástico* e *não ser sarcástico*. Em termos matemáticos, dizemos que temos um conjunto de exemplos denotado pela matriz  $X$  de dimensões  $n \times m$  onde  $n$  é o número de exemplos e  $m$  é o número de características que usamos para descrever cada um desses exemplos. Cada linha de  $X$  é denotada por  $x^{(i)}$  e chamada de *exemplo*  $i$  ou *instância*  $i$ .

Cada instância  $i$  pertence ou à primeira classe ou à segunda, modelamos isso por um valor  $y^{(i)}$  chamado de  $i$ -ésimo rótulo (do inglês, *label*). E dizemos que  $y^{(i)} \in \{0, 1\}$ , onde cada valor representa uma das classes. No caso de sarcasmo, podemos representar por *não sarcasmo* o valor 0 e *sarcasmo*, o valor 1. Com esses valores  $y^{(i)}$  construímos um vetor  $y$  onde  $y_i = y^{(i)}$ .

Ao final, o problema é descobrir uma função  $f$  que mapeie bem  $X$  em  $y$ . Note que  $n$ , o número de exemplos, é possivelmente infinito, pois sempre se pode achar novos exemplos. Ou seja, tenta-se achar essa função  $f$  conhecendo parcialmente o conjunto das instâncias.

## 2.2 Trabalhos Relacionados