

Fabício Veloso de Jesus

Análise e classificação de comentários

Brasil

2018, v1.0

Fabício Velôso de Jesus

Análise e classificação de comentários

Trabalho monografico apresentado para obtenção do grau de bacharel em ciências exatas e tecnológicas.

Universidade Federal do Recôncavo da Bahia - UFRB Bacharelado em Ciências Exatas e Tecnológicas

Orientador: Tiago Palma Pagano

Brasil

2018, v1.0

Resumo

Palavras-chave:

Abstract

Keywords:

Lista de ilustrações

Figura 1 – Rede Neural Multicamada	20
--	----

Lista de quadros

Lista de tabelas

Lista de abreviaturas e siglas

IA	Sigla para Inteligência Artificial
SOM	<i>Self-Organizing Map</i> , em português Mapas auto organizáveis
RNAs	Sigla para Redes Neurais Artificiais

Lista de símbolos

Γ Letra grega Gama

Sumário

1	INTRODUÇÃO	17
1.1	Objetivo	17
1.2	Objetivos específicos	17
1.3	Justificativa	17
1.4	Metodologia	17
1.5	Problematização	17
2	REFERÊNCIAL TEÓRICO	19
2.1	Misoginia	19
2.1.1	Cassificação	19
2.2	Mineração de Textos	19
2.3	Inteligência Artificial	19
2.3.1	Redes Neurais Artificiais	19
2.3.1.1	Motivação para as RNAs: redes biológicas	20
2.3.2	Mapas Auto Organizáveis de Kohonen	21
3	DESENVOLVIMENTO	23
4	TESTES E ANÁLISE DE RESULTADOS	25
5	CONCLUSÃO	27
	REFERÊNCIAS	29
	Appendices	31

1 Introdução

1.1 Objetivo

Analisar e classificar comentários de twitter segundo seu caráter misógino.

1.2 Objetivos específicos

Utilizar métodos capazes de classificar os comentários segundo seu caráter misógino. Dentro deste comportamento de aversão às mulheres existem subcategorias, que devem ser declaradas e evidenciadas na classificação.

Analisar características comuns as frases que pertecem ao mesmo grupo e determinar a ocorrência e relevância de determinadas palavras para a identificação.

Determinar se tal comportamento possui direcionamento a um usuário em específico, ou é realizado de forma a generalizar todas as mulheres.

1.3 Justificativa

Como consequência, a análise dos resultados obtidos neste trabalho poderá prover um padrão específico referente ao comportamento de usuários misóginos no twitter.

1.4 Metodologia

Aplicar métodos de mineração de dados em textos para realizar o ajuste dos dados existentes na base.

Utilizar aprendizado de máquina nos dados ajustados para criar uma rotina de classificação das frases. A proposta aqui é com o auxílio de redes neurais, evidenciar dados específicos encontrados em comentários que refletem um cunho misógino, no qual destacamos o método de mapas auto organizáveis com o intuito de evidenciar características comuns em frases que possuem a mesma classificação.

1.5 Problematização

Com auxílio de métodos inerentes a inteligência artificial é possível determinar a existência de misoginia em um comentário?

Através do agrupamento de características é praticável a classificação das frases misóginas em subcategorias?

Existe um padrão para comentários que apresentam cunho misógino?

2 Referencial Teórico

Neste capítulo as referências conceituais e conceitos envolvidos neste trabalho serão descritos. Partindo da definição de misoginia, passando pelas técnicas envolvidas, e arrematando com as concepções de análise dos dados.

2.1 Misoginia

2.1.1 Classificação

2.2 Mineração de Textos

2.3 Inteligência Artificial

2.3.1 Redes Neurais Artificiais

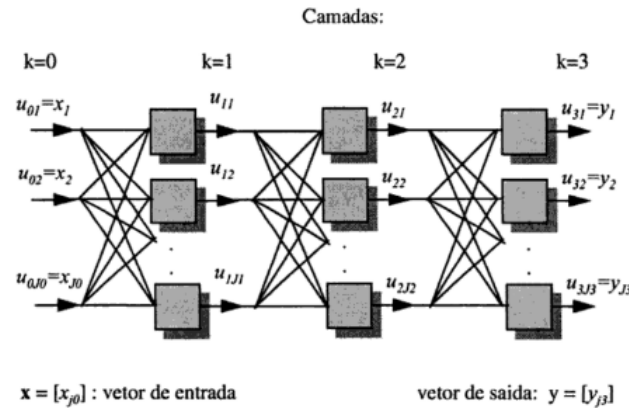
Segundo [Braga, Carvalho e Ludermir \(2000\)](#) RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas (normalmente não-lineares). Essas unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Estes modelos de conexões normalmente estão associados a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O funcionamento destas redes é inspirado em uma estrutura física natural: o cérebro humano.

Por volta do fim da década de 1950, na Universidade de Cornell, Rosenblatt deu continuidade às idéias de McCulloch. Criando uma genuína rede de múltiplos neurônios do tipo *discriminadores lineares* esta rede foi descrita como rede de *perceptron*. Um perceptron é uma rede com a seguinte topologia, os neurônios são dispostos em várias *camadas*. Os que recebem das entradas diretamente formam o que é chamada de *camada de entrada*. A camada que recebe a saída da camada de entrada como entrada constituem a segunda camada e assim consecutivamente até a última camada que é chamada de *camada de saída*. Camadas que ficam entre as de entrada e saída são comumente referidas como *camadas ocultas*.

Com referência à [Figura 1](#). Uma rede neural multicamada de K camadas, terá como entrada um vetor \mathbf{x} de dimensão J_0 de componentes x_{j_0} , $j_0 = 1, 2, \dots, J_0$. Estas conectam-se às entradas dos J_1 neurônios numa primeira camada. As saídas u_{lj_1} , $j_1 = 1, 2, \dots, J_1$ destes, formando as componentes de um novo vetor \mathbf{u}_1 de dimensão J_1 , conectam-se às entradas

dos J_2 neurônios da camada seguinte e assim sucessivamente até a última camada que consistirá de J_K neurônios fornecendo como saída da rede um vetor $\mathbf{y} = \mathbf{u}_K$ de dimensão J_K . Genéricamente, u_{kj_k} denota a saída do j_k -ésima entrada da rede, e para $k = K$ a j_k -ésima saída da rede. (KOVÁCS, 2002, p. 39–40)

Figura 1 – Rede Neural Multicamada



Fonte: Kovács (2002, p. 40)

Ainda segundo Kovács (2002) e Braga, Carvalho e Ludermir (2000) o problema que Rosenblatt propôs a resolver foi o de casos simples com implementação de funções booleanas **E** e **OU** de duas variáveis, que são problemas linearmente separáveis, isto é, problemas cuja solução pode ser obtida ao dividir o espaço de entrada em duas regiões através de uma reta. O perceptron, não consegue detectar conectividade, paridade e simetria, que são problemas não-linearmente separáveis. Estes são exemplos de *hard learning problems* (problemas difíceis de aprender).

A abordagem conexionista ficou adormecida durante os anos 70, porém alguns pesquisadores continuaram desenvolvendo trabalhos na área. Dentre eles podem ser citados Igor Aleksander (redes sem pesos) na Inglaterra, Kunihiko Fukushima (cognitron e neocognitron) no Japão, Steven Grossberg (sistemas auto-adaptativos) nos EUA, e Teuvo Kohonen (memórias associativas e auto-organizadas) na Finlândia.

2.3.1.1 Motivação para as RNAs: redes biológicas

O cérebro humano é um imenso e complexo bosque de células e conexões intercelulares. Esse bosque emaranhado é composto de aproximadamente 100 bilhões de neurônios ($1 * 10^{11}$) de formas e tamanhos diferentes. Considera-se que apenas no córtex cerebral, que contém quase a metade desse número, isto é, cerca de 50 bilhões, existam mais de 500 tipos de neurônios morfológicamente diferentes, distribuídos em 52 áreas. (MORA, 2016, p.18)

A estrutura dos nodos, a topologia dessas conexões e o comportamento conjunto dos neurônios naturais constroem a base de estudo das RNAs. As RNAs tendem a reproduzir as funções das redes biológicas, buscando colocar em prática a sua dinâmica e seu comportamento básico.

Conforme [Braga, Carvalho e Ludermir \(2000\)](#), como características comuns, ambos os sistemas são baseados em unidades de computação paralela e distribuída que se comunicam por meio de conexões sinápticas, possuem detetores de características, redundância e modularização das conexões. Apesar de pouca similaridade entre os dois sistemas do ponto de vista biológico, estas características semelhantes permitem às RNAs reproduzirem com fidelidade várias funções inerentes dos seres humanos

2.3.2 Mapas Auto Organizáveis de Kohonen

3 Desenvolvimento

4 Testes e Análise de Resultados

5 Conclusão

Referências

BRAGA, A. d. P.; CARVALHO, A.; LUDERMIR, T. B. *Redes neurais artificiais: teoria e aplicações*. [S.l.]: Livros Técnicos e Científicos Rio de Janeiro, 2000. Citado 3 vezes nas páginas 19, 20 e 21.

KOVÁCS, Z. L. *Redes neurais artificiais*. [S.l.]: Editora Livraria da Física, 2002. Citado na página 20.

MORA, F. *Continuum: Como Funciona o Cérebro?* [S.l.]: Artmed Editora, 2016. Citado na página 20.

Appendices

