

Aprendizagem Guiada para Análise Morfossintática usando Redes Neurais Recursivas

Marcos Vinícius Treviso
marcosvtreviso@gmail.com

Orientador: Fábio Natanael Kepler
Trabalho de Conclusão de Curso I

8 de julho de 2015

Universidade Federal do Pampa

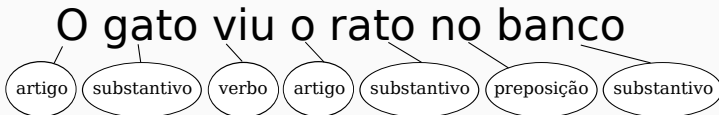
- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Cópus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Cópus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

POS Tagging

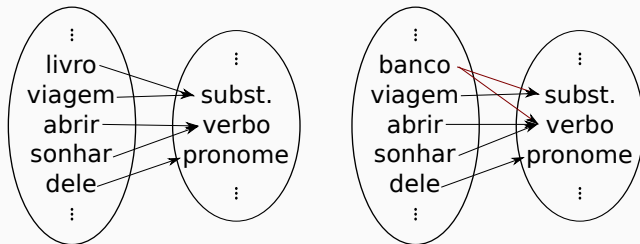
É conhecido em Processamento de Linguagem Natural (PLN) como o ato de classificar uma palavra pertencente a um conjunto de textos em uma classe gramatical.

- Qual a medida de eficiência?
 - Acurácia
 - Atualmente cerca de 97%
- Quais são as aplicações?
 - Tradução automática
 - Sumarização



O problema

- Linguagens naturais são ambíguas
- Estratégia trivial não é eficaz
- Necessário analisar o contexto
- Aprendizado de máquina

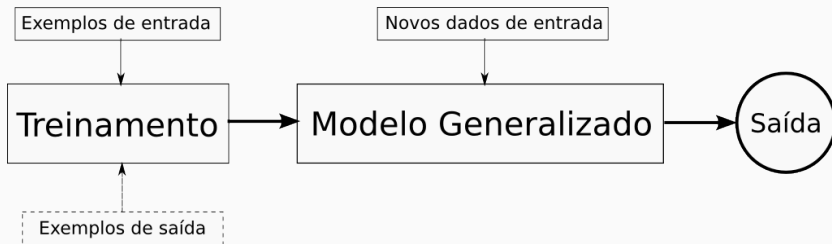


- Desenvolver novo modelo para POS Tagging
 - A princípio para o português brasileiro
- Alcançar estado da arte
 - Combinar abordagens existentes
- Analisar a eficiência
 - Acurácia
 - Tempo de treinamento

- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Cópus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

Aprendizado de máquina

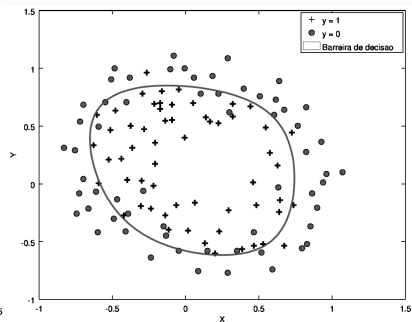
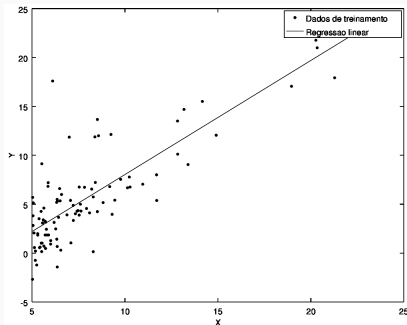
- Aprendizado supervisionado
 - Regressão
 - Classificação
- Aprendizado não supervisionado



$$h_{\theta}(x) = \theta_0 + \theta_1 f(x_1) + \theta_2 f(x_2) + \dots + \theta_n f(x_n)$$

Aprendizado de máquina

- Aprendizado supervisionado
 - Regressão
 - Classificação
- Aprendizado não supervisionado



- Coleções de textos agrupados
- Anotação gramatical manual
- *Córpus* para o português brasileiro:

Córpus	Sentenças	Palavras	Classes gramaticais
Mac-Morpho original	53,374	1,221,465	41
Mac-Morpho revisado ¹	49,932	945,958	26
Tycho Brahe	55,932	1,541,654	265

- Por que não combiná-los?

1. Revisado por: Fonseca, Rosa e Aluísio (2015).

Representação de palavras

- Vetores reais valorados em um espaço multidimensional (*word embeddings*)
- Mais desempenho de aplicações em PLN e menos engenharia de *features*
- Conseguem capturar informações sintáticas e semânticas
- Geradas de maneiras diferentes dependendo da técnica utilizada
- Palavras fora do vocabulário de treinamento também tem seu próprio vetor

Técnicas para geração de *word embeddings*

- Palavras similares estão próximas

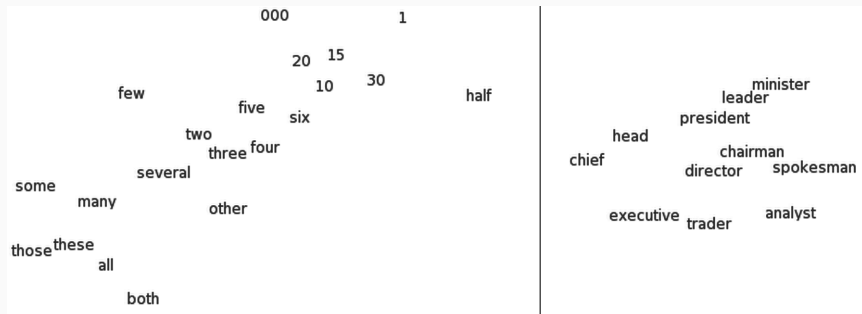


Imagem criada pelo t-SNE
Fonte: Turian, Ratinov e Bengio (2010)

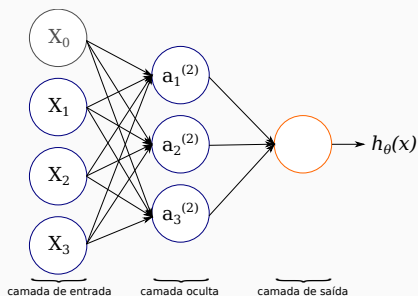
Técnicas para geração de *word embeddings*

- Matriz de coocorrência
- *Neural Language Model* (NLM): Através de redes neurais
- *Hyperspace Analogue to Language* (HAL): Matriz de coocorrência com um método de decomposição (Escalamento Multidimensional)
- Modelação *Skip-Gram* (SG): Previsão de palavras vizinhas num conjunto de tamanho finito: *word2vec*
- *Global Vectors* (GloVe): Razão das probabilidades na matriz de coocorrência em relação ao contexto de uma outra palavra do vocabulário

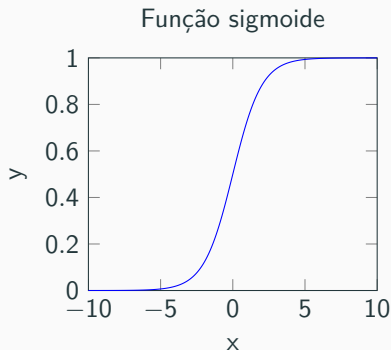
$$W : \text{word} \rightarrow \mathbb{R}^n$$

Redes neurais

- Simulação do cérebro humano
- Unidades de ativação: $a_i^{(j)}$
- Pesos: $\theta^{(j)}$
- Parâmetros: $z^{(j+1)} = \theta^{(j)} a^{(j)}$
- Função de ativação: $g(z)$
- $a^{(j+1)} = g(z^{(j+1)})$

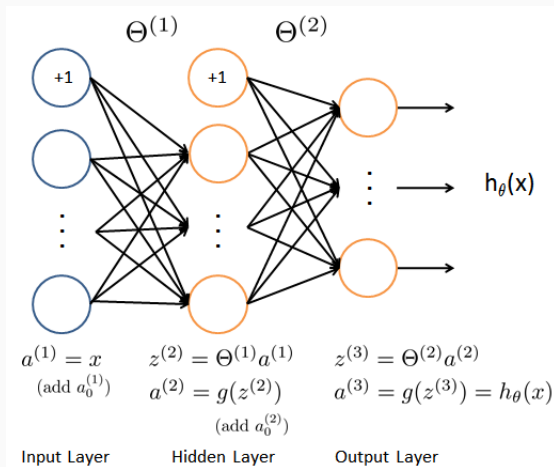


- Simulação do cérebro humano
- Unidades de ativação: $a_i^{(j)}$
- Pesos: $\theta^{(j)}$
- Parâmetros: $z^{(j+1)} = \theta^{(j)} a^{(j)}$
- Função de ativação: $g(z)$
- $a^{(j+1)} = g(z^{(j+1)})$



Redes neurais

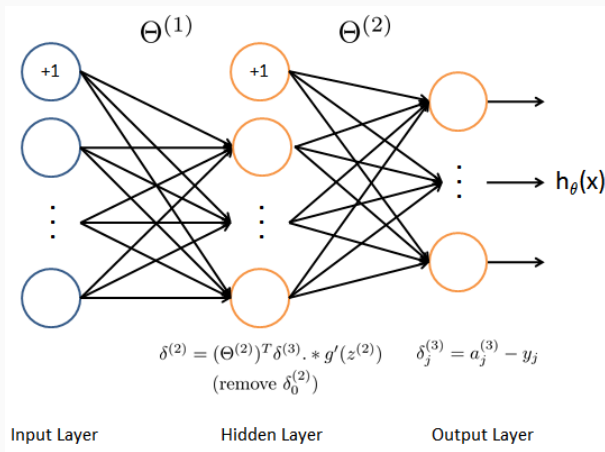
- Processo de aprendizagem (*Forward Propagation*):



Fonte: Ng (2015)

Redes neurais

- Processo de aprendizagem (*Backpropagation*):



Fonte: Ng (2015)

- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Córpus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

- Escopo do português brasileiro

Autores	Modelo	Rep. palavras	Cópus	Acurácia
Kepler (2005)	VLMC	Seq. de carac.	Tycho Brahe	95,51%
Santos e Zadrozny (2014)	Redes neurais profundas	Vetores	Tycho Brahe; Mac-Morpho	97,47%
Fonseca, Rosa e Aluísio (2015)	Redes neurais	Vetores	Tycho Brahe; Mac-Morpho	97,57%
Este trabalho	Redes neurais recursivas	Vetores	Tycho Brahe; Mac-Morpho	-

- Estado da arte com 97,57% de acurácia para todas palavras. Treinado com o Mac-Morpho original (FONSECA; ROSA; ALUÍSIO, 2015)
- Estado da arte com 94,34% de acurácia para palavras fora do vocabulário. Treinado com o Mac-Morpho revisado (FONSECA; ROSA; ALUÍSIO, 2015)

- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Córpus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

Representação de palavras

- Técnicas:
 - NLM
 - SG
 - GloVe

$$w_i \in \omega \rightarrow v_i \in \mathbb{R}^d$$

$$c_i \in \gamma \rightarrow z_i \in \mathbb{R}^d$$

- Capitalização
- Prefixos

Pontuações para classes gramaticais

- Janela de palavras com tamanho t :

$$V_n = \{v_{n-(t-1)/2}, \dots, v_n, \dots, v_{n+(t-1)/2}\}$$

- Pontuações para classes gramaticais:

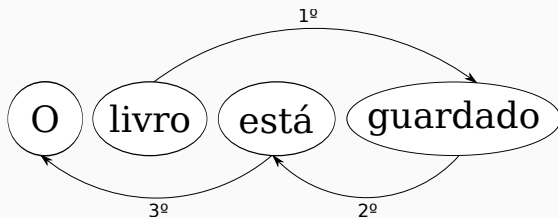
$$s_c(V_n)$$

$$A_{c,d,e}$$

- Pontuação final para w_i^t dado c_1^t :

$$S(w_1^t, c_1^t) = \sum_{k=1}^t \left(\arg \max_{1 \leq i \leq t, i \notin Q} (s_{c_i}(V_i) + A_{c_{i-1}, c_i, c_{i+1}}) \right)$$

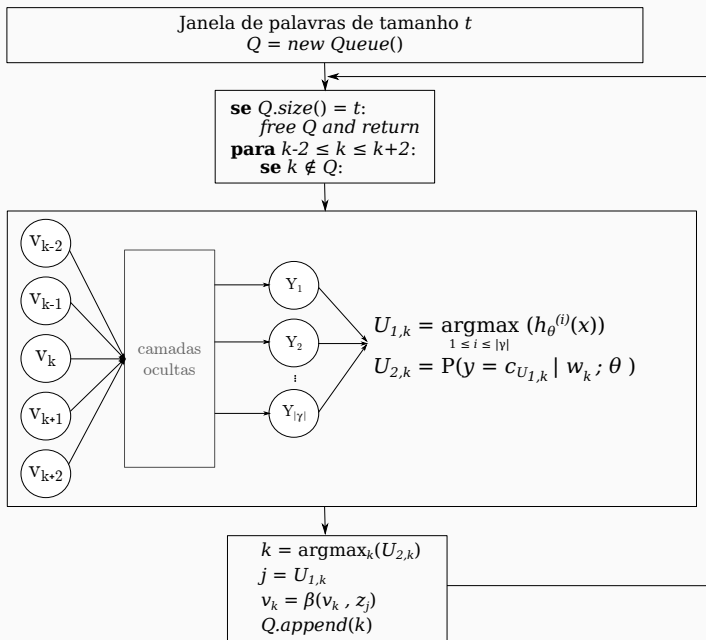
- Treinamento supervisionado
- Rede neural recursiva
- Aprendizagem guiada por palavras mais fáceis (SHEN; SATTA; JOSHI, 2007)



- Composição dos vetores:

$$\beta(v_n, z_c) \mapsto v_n + z_c$$

Modelo Proposto



- Ajustes feitos para maximizar:

$$\sum_{(w_1^t, c_1^t) \in \phi} P(c_1^t | w_1^t; \theta)$$

- Função de custo:

$$J(\theta) = \log \left(\sum_{u_1^t \in \gamma^t} e^{S(w_1^t, u_1^t)} \right) - S(w_1^t, c_1^t)$$

- Gradiente Descendente, Gradiente Descendente Estocástico, Adagrad, Adadelta, etc. (BENGIO; GOODFELLOW; COURVILLE, 2015)

- Introdução
 - Part-of-speech* (POS) Tagging
 - O problema
 - Objetivos
- Fundamentação
 - Aprendizado de máquina
 - Cópus
 - Representação de palavras
 - Redes neurais
 - Aprendizagem profunda
- Trabalhos relacionados
- Metodologia
 - Representação de palavras
 - Pontuações para classes gramaticais
 - Treinamento
- Cronograma

Cronograma

A1 - Implementação do modelo neural recursivo

A2 - Treinamento do modelo

A3 - Avaliação dos resultados obtidos

A4 - Escrita da monografia

	Agosto	Setembro	Outubro	Novembro	Dezembro
A1	X	X	X		
A2		X	X	X	
A3			X	X	X
A4	X	X	X	X	X

BENGIO, Y.; GOODFELLOW, I. J.; COURVILLE, A. Deep learning. Book in preparation for MIT Press. 2015. Disponível em:

<<http://www.iro.umontreal.ca/~bengioy/dlbook>>.

FONSECA, E. R.; ROSA, J. L. G.; ALUÍSIO, S. M. Evaluating word embeddings and a revised corpus for part-of-speech tagging in portuguese. *Journal of the Brazilian Computer Society*, Springer, v. 21, n. 1, p. 1–14, 2015.

KEPLER, F. N. *Um etiquetador morfo-sintático baseado em cadeias de Markov de tamanho variável*. Tese (Doutorado) — Instituto de Matemática e Estatística da Universidade de São Paulo, 12/04/2005., 2005.

NG, A. *Course of Machine Learning*. [S.l.], 2015. Disponível em: <<https://www.coursera.org/learn/machine-learning/>>.

- SANTOS, C. N. dos; ZADROZNY, B. Training state-of-the-art portuguese pos taggers without handcrafted features. In: *Computational Processing of the Portuguese Language*. [S.l.]: Springer, 2014. p. 82–93.
- SHEN, L.; SATTA, G.; JOSHI, A. Guided learning for bidirectional sequence classification. In: CITESEER. *ACL*. [S.l.], 2007. v. 7, p. 760–767.
- TURIAN, J.; RATINOV, L.; BENGIO, Y. Word representations: a simple and general method for semi-supervised learning. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 48th annual meeting of the association for computational linguistics*. [S.l.], 2010. p. 384–394.

Aprendizagem Guiada para Análise Morfossintática usando Redes Neurais Recursivas

Marcos Vinícius Treviso

`marcosvtreviso@gmail.com`

Orientador: Fábio Natanael Kepler

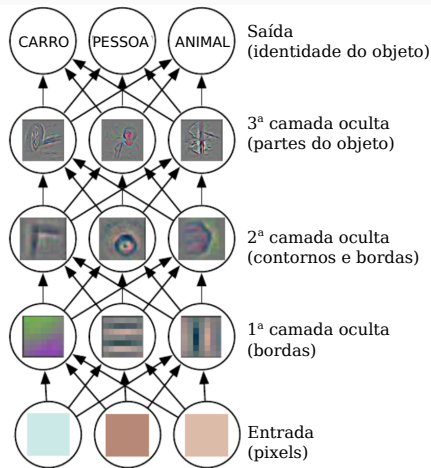
Trabalho de Conclusão de Curso I

8 de julho de 2015

Universidade Federal do Pampa

Aprendizagem profunda

- Muitas transformações não lineares
- Objetivo de aprender automaticamente boas *features*
- Crescimento do desempenho computacional e criação de novos algoritmos
- Redes neurais com múltiplas camadas



Adaptado de: Bengio, Goodfellow e Courville (2015)

Redes neurais recursivas

- Grafo computacional parece como uma árvore
- Aplica-se transformações recursivamente
- Composição da saída com entrada

