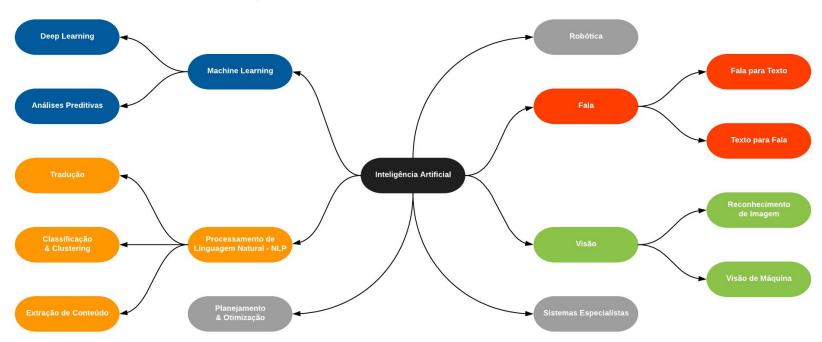
Natural Language Processing

NLP

O que é processamento de linguagem natural?

- O que é processamento de linguagem natural?
 - Uma subárea da Inteligência Artificial;



- O que é processamento de linguagem natural?
 - Uma subárea da Inteligência Artificial;
 - Estuda as capacidades e limitações em uma máquina gerar e compreender línguas humanas naturais.

Objetivos principais

- Objetivos principais
 - Através de um modelo computacional:

- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica;
 - Morfológica.

- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica;
 - Morfológica.
 - Resumo;

- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica;
 - Morfológica.
 - Resumo;
 - Extração de informação;

- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica;
 - Morfológica.
 - Resumo;
 - Extração de informação;
 - Interpretação;

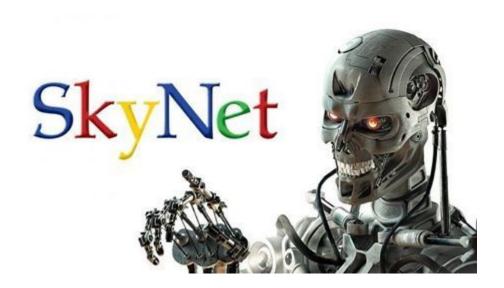
- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica;
 - Morfológica.
 - Resumo;
 - Extração de informação;
 - Interpretação;
 - Análise de sentimento;

- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica:
 - Morfológica.
 - Resumo;
 - Extração de informação;
 - Interpretação;
 - Análise de sentimento;
 - Reconhecimento de contexto;

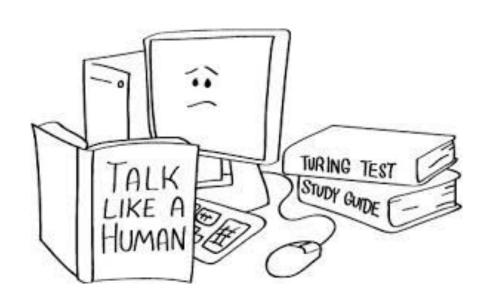
- Objetivos principais
 - Através de um modelo computacional:
 - Análise textual;
 - Sintática;
 - Semântica;
 - Léxica:
 - Morfológica.
 - Resumo;
 - Extração de informação;
 - Interpretação;
 - Análise de sentimento;
 - Reconhecimento de contexto;
 - Aprender termos e expressões.

Por que fazer a máquina entender nossa língua?

- Por que fazer a máquina entender nossa língua?
 - Melhorar a forma com que nós humanos interagimos com sistemas computacionais.



- Por que fazer a máquina entender nossa língua?
 - Melhorar a forma com que nós humanos interagimos com sistemas computacionais.
 - Teste de Turing

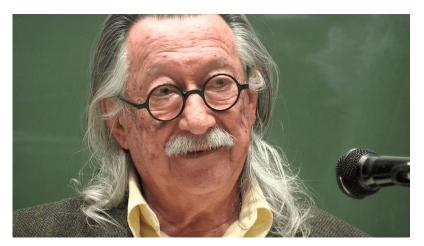




Alan Turing

- Por que fazer a máquina entender nossa língua?
 - Melhorar a forma com que nós humanos interagimos com sistemas computacionais.
 - Eliza
 - Desenvolvido entre 1964 e 1966 no laboratório de inteligência artificial do MIT.

Joseph Weizenbaum

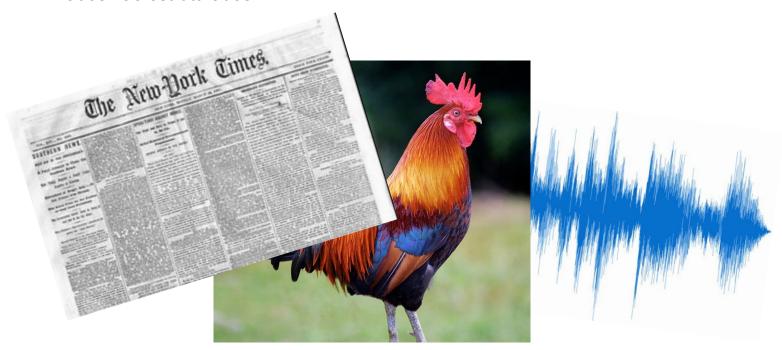


```
IIIIIII ZZZZZZZ
            EEEEEEE
            EEEEE
            EEEEEEEE LLLLLLL IIIIIII ZZZZZZ
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
   ZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
T > CAN I TALK ABOUT MY PROBLEM ?
            TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
           EASE RELAX A LITTLE. SLEEP WELL.
        THANKS FOR YOUR ADVICE.
 LIZA > BYE AND KEEP IN TOUCH...
```



Desafio em trabalhar com texto

- Desafio em trabalhar com texto
 - Dados não estruturados



- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;
 - Letras maiúsculas para minúsculas;

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;
 - Letras maiúsculas para minúsculas;
 - Remover stopwords;
 - que, de, do, o, a, e...

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;
 - Letras maiúsculas para minúsculas;
 - Remover stopwords;
 - que, de, do, o, a, e...
 - Remover números e unidades;

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;
 - Letras maiúsculas para minúsculas;
 - Remover stopwords;
 - que, de, do, o, a, e...
 - Remover números e unidades;
 - Correção ortográfica.

- Desafio em trabalhar com texto
 - Dados não estruturados;
 - Aplicações de filtros;
 - Remover de caracteres especiais;
 - Letras maiúsculas para minúsculas;
 - Remover stopwords;
 - que, de, do, o, a, e...
 - Remover números e unidades;
 - Correção ortográfica.
 - Converter o texto para uma representação numérica (encoding).

Encoding (One-Hot Encoding)

- Encoding (One-Hot Encoding)
 - Vocabulário:

- Encoding (One-Hot Encoding)
 - Vocabulário:
 - piloto;

- Encoding (One-Hot Encoding)
 - Vocabulário:
 - piloto;
 - carro;

- Encoding (One-Hot Encoding)
 - Vocabulário:
 - piloto;
 - carro;
 - veloz.

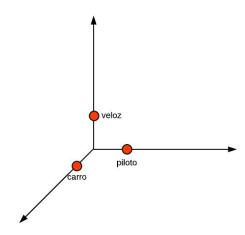
- Encoding (One-Hot Encoding)
 - Vocabulário:
 - piloto;
 - carro;
 - veloz.

	piloto	carro	veloz
piloto	1	0	0
carro	0	1	0
veloz	0	0	1

- Encoding (One-Hot Encoding)
 - Vocabulário:
 - piloto;
 - carro;
 - veloz.

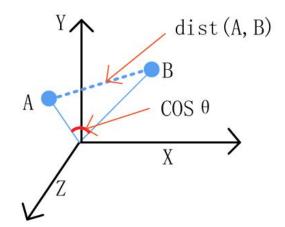
Palavra	Representação Vetorial
piloto	[1, 0, 0]
carro	[0, 1, 0]
veloz	[0, 0, 1]

- Encoding (One-Hot Encoding)
 - Cosine similarity = 0, ângulo de 90°

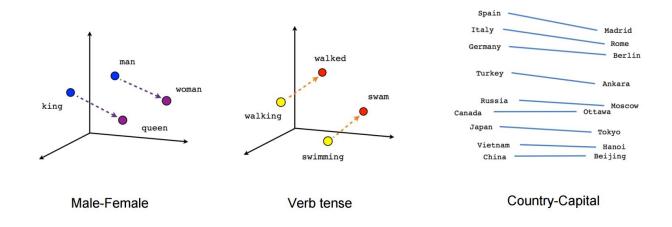


Embedding (Word2Vec)

- Embedding (Word2Vec)
 - Modelo capaz de gerar um vetor de similaridade entre palavras vizinhas



- Embedding (Word2Vec)
 - Modelo capaz de gerar um vetor de similaridade entre palavras vizinhas



Entradas

- Entradas
 - "O Rei é um homem de bravura"

Entradas

- "O Rei é um homem de bravura"
- "A Rainha é uma mulher de sabedoria"

- Entradas
 - "O Rei é um homem de bravura"
 - "A Rainha é uma mulher de sabedoria"
- Aplicando filtros

- Entradas
 - "O Rei é um homem de bravura"
 - "A Rainha é uma mulher de sabedoria"
- Aplicando filtros
 - "rei homem bravura"

Entradas

- "O Rei é um homem de bravura"
- "A Rainha é uma mulher de sabedoria"

Aplicando filtros

- "rei homem bravura"
- "rainha mulher sabedoria"

Word2Vec (Skip-gram)

Palavra	Vizinho	
rei	homem	
rei	bravura	
homem	rei	
homem	bravura	
bravura	rei	
bravura	homem	
rainha	mulher	
rainha	sabedoria	
mulher	rainha	
mulher	sabedoria	
sabedoria	rainha	
sabedoria	mulher	

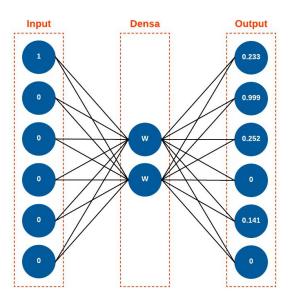
Word2Vec (Skip-gram)

Palavra	One-Hot	Vizinho	One-Hot
rei	[1, 0, 0, 0, 0, 0]	homem	[0, 1, 0, 0, 0, 0]
rei	[1, 0, 0, 0, 0, 0]	bravura	[0, 0, 1, 0, 0, 0]
homem	[0, 1, 0, 0, 0, 0]	rei	[1, 0, 0, 0, 0, 0]
homem	[0, 1, 0, 0, 0, 0]	bravura	[0, 0, 1, 0, 0, 0]
bravura	[0, 0, 1, 0, 0, 0]	rei	[1,0,0,0,0,0]
bravura	[0, 0, 1, 0, 0, 0]	homem	[0, 1, 0, 0, 0, 0]
rainha	[0, 0, 0, 1, 0, 0]	mulher	[0, 0, 0, 0, 1, 0]
rainha	[0, 0, 0, 1, 0, 0]	sabedoria	[0, 0, 0, 0, 0, 1]
mulher	[0, 0, 0, 0, 1, 0]	rainha	[0, 0, 0, 1, 0, 0]
mulher	[0, 0, 0, 0, 1, 0]	sabedoria	[0, 0, 0, 0, 0, 1]
sabedoria	[0, 0, 0, 0, 0, 1]	rainha	[0, 0, 0, 1, 0, 0]
sabedoria	[0, 0, 0, 0, 0, 1]	mulher	[0, 0, 0, 0, 1, 0]

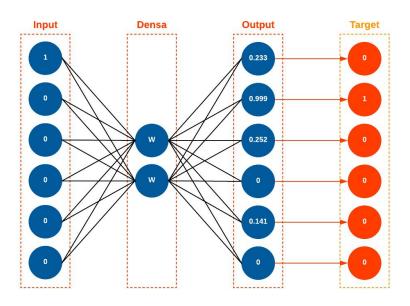
Word2Vec (Input e meta)

Input	Target
[1, 0, 0, 0, 0, 0]	[0, 1, 0, 0, 0, 0]
[1,0,0,0,0,0]	[0,0,1,0,0,0]
[0, 1, 0, 0, 0, 0]	[1,0,0,0,0,0]
[0, 1, 0, 0, 0, 0]	[0, 0, 1, 0, 0, 0]
[0,0,1,0,0,0]	[1,0,0,0,0,0]
[0, 0, 1, 0, 0, 0]	[0, 1, 0, 0, 0, 0]
[0, 0, 0, 1, 0, 0]	[0, 0, 0, 0, 1, 0]
[0, 0, 0, 1, 0, 0]	[0, 0, 0, 0, 0, 1]
[0, 0, 0, 0, 1, 0]	[0, 0, 0, 1, 0, 0]
[0,0,0,0,1,0]	[0, 0, 0, 0, 0, 1]
[0, 0, 0, 0, 0, 1]	[0, 0, 0, 1, 0, 0]
[0, 0, 0, 0, 0, 1]	[0, 0, 0, 0, 1, 0]

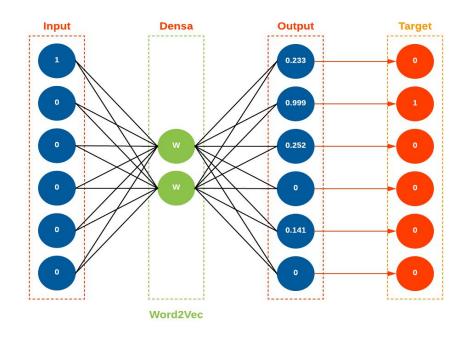
Word2Vec treinamento



Word2Vec treinamento



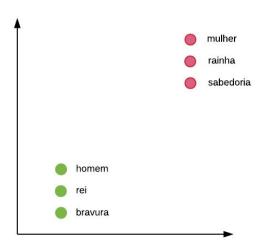
Word2Vec treinamento



Word2Vec embedding

Palavra	Embedding
rei	[10, 10]
homem	[10,15]
bravura	[10, 5]
rainha	[50, 50]
mulher	[50, 55]
sabedoria	[50, 45]

Word2Vec embedding



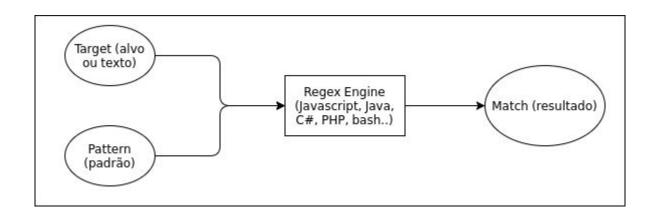
Atividade 1

Expressões Regulares REGEX

• O que são expressões regulares?

- O que são expressões regulares?
 - Uma linguagem de busca de padrões;

- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.



- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?

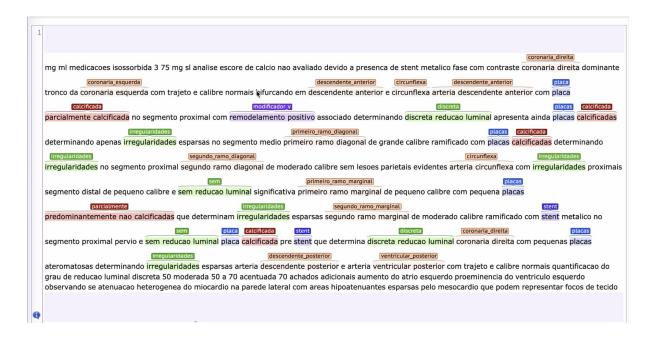
- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?
 - Identificar qualquer padrão conhecido em um texto livre;
 - Ex. Números de telefone, documentos, CRM, datas em arquivos.

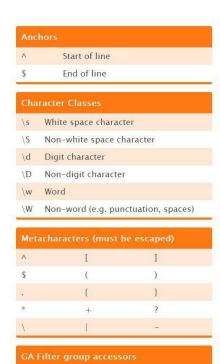
- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?
 - Identificar qualquer padrão conhecido em um texto livre;
 - Ex. Números de telefone, documentos, CRM, datas em arquivos.
 - Limpeza de documentos;
 - Ex. Remover cabeçalhos e rodapés de documentos onde conheço o padrão.

- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?
 - Identificar qualquer padrão conhecido em um texto livre;
 - Ex. Números de telefone, documentos, CRM, datas em arquivos.
 - Limpeza de documentos;
 - Ex. Remover cabeçalhos e rodapés de documentos onde conheço o padrão.
 - Extração de entidades textuais;

- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?
 - Identificar qualquer padrão conhecido em um texto livre;
 - Ex. Números de telefone, documentos, CRM, datas em arquivos.
 - Limpeza de documentos;
 - Ex. Remover cabeçalhos e rodapés de documentos onde conheço o padrão.
 - Extração de entidades textuais;
 - Formatação de documentos;

- O que são expressões regulares?
 - Uma linguagem de busca de padrões;
 - Percorre o texto buscando por patterns definidas.
- Qual a aplicação?
 - Identificar qualquer padrão conhecido em um texto livre;
 - Ex. Números de telefone, documentos, CRM, datas em arquivos.
 - Limpeza de documentos;
 - Ex. Remover cabeçalhos e rodapés de documentos onde conheço o padrão.
 - Extração de entidades textuais;
 - Formatação de documentos;
 - o Etc...





\$Ax Access group x in field A (e.g. \$A1)

\$Bx Access group x in field B (e.g. \$B1)

Quant	ifiers
÷	Zero or more (greedy)
*?	Zero or more (lazy)
+	One or more (greedy)
+?	One or more (lazy)
?	Zero or one (greedy)
??	Zero or one (lazy)
{X}	Exactly X (e.g. 3)
{X,}	X or more, (e.g. 3)
{X, Y}	Between X and Y (e.g. 3 and 5) (lazy)
Range	s and Groups
	Any character
(a b)	a or b (case sensitive)

Group, e.g. (keyword)

Range (a or b or c)

Digit between 0 and 7

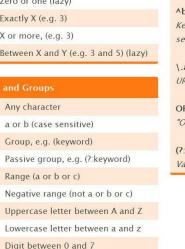
(...) (?:...)

[abc]

[^abc] [A-Z]

[a-z]

[0-7]



Sample Patterns
^/directory/(.*)
Any page URLs starting with /directory/
(brand\s*?term)
Brand term with or without whitespace
between words
^brand\s+[^cf]
Key phrases beginning with 'brand' and the
second word not starting with c or f
\.aspx\$
URLs ending in '.aspx'
ORDER\-\d{6}
"ORDER-" followed by a six digit ID
(?:\? &)utm=([^&\$]+)
Value of 'utm' querystring parameter

https://www.rexegg.com/regex-quickstart.html

Atividade 2

Redes Neurais Recorrentes

O que são RNN's?

São redes voltadas para processamento de dados sequenciais;

O que são RNN's?

- São redes voltadas para processamento de dados sequenciais;
- Possui uma estrutura capaz de guardar estados anteriores;

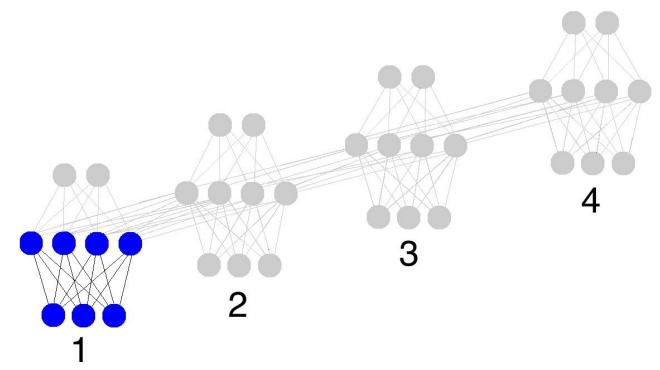
O que são RNN's?

- São redes voltadas para processamento de dados sequenciais;
- Possui uma estrutura capaz de guardar estados anteriores;
- Ao final de uma iteração, o valor de output da rede é novamente inserido como input na próxima iteração;

O que são RNN's?

- São redes voltadas para processamento de dados sequenciais;
- Possui uma estrutura capaz de guardar estados anteriores;
- Ao final de uma iteração, o valor de output da rede é novamente inserido como input na próxima iteração;
- Sua arquitetura é muito similar às redes feed-forward.

O que são RNN's?



Processamento de Linguagem Natural (NLP);

- Processamento de Linguagem Natural (NLP);
 - Extração de informações;

- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;

- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;
 - Resumo textual;

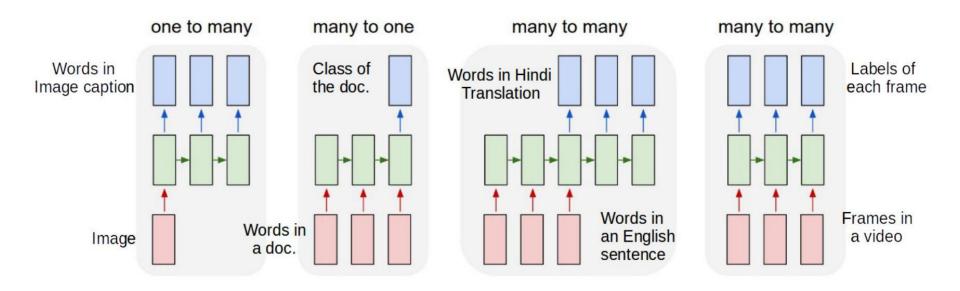
- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;
 - Resumo textual;
 - Chat-bots.

- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;
 - Resumo textual;
 - Chat-bots.
- Processamento de vídeos

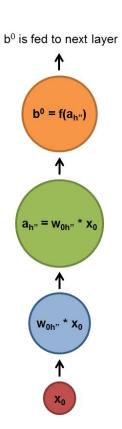
- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;
 - Resumo textual;
 - Chat-bots.
- Processamento de vídeos
- Processamento de imagens

- Processamento de Linguagem Natural (NLP);
 - Extração de informações;
 - Tradução;
 - Resumo textual;
 - Chat-bots.
- Processamento de vídeos
- Processamento de imagens
- Séries temporais

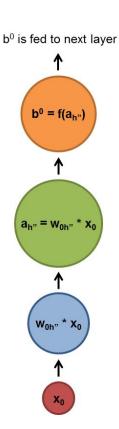
- Processamento de Linguagem Natural (NLP)
 - Extração de informações;
 - Tradução;
 - Resumo textual;
 - Chat-bots.
- Processamento de vídeos
- Processamento de imagens
- Séries temporais
 - Previsão de valores em um delta(tempo).

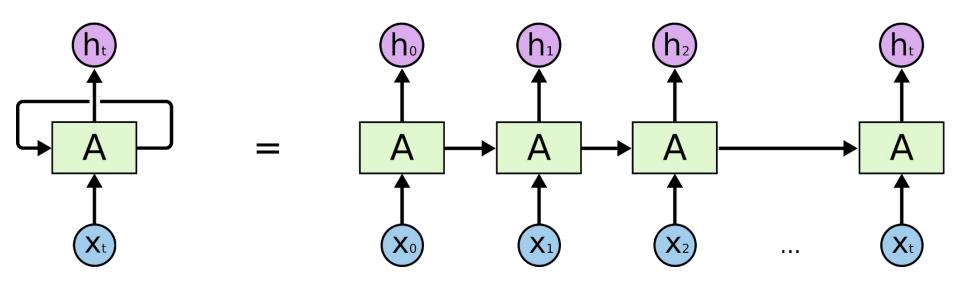


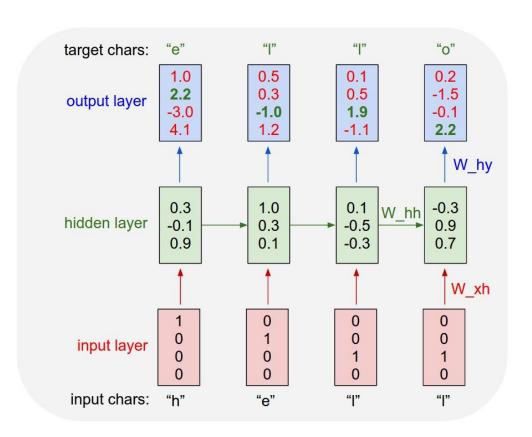
- Na arquitetura de uma RNN existe um componente conhecido como célula de memória;
- A célula de memória recebe como entrada o valor de output da rede em seu t-1;



- Este tipo de célula permite que uma rede neural consiga "lembrar" de comportamentos;
- Tornando possível gerar respostas para perguntas como em uma solução de chat-bot, por exemplo.







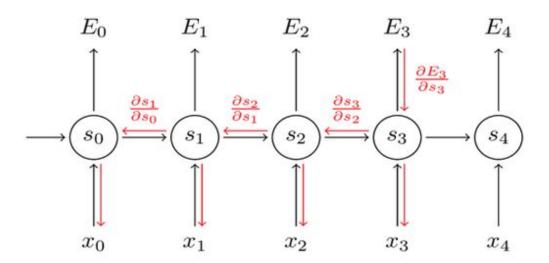
BTT?

 RNN's utilizam um conceito de backpropagation through time (BTT) para realizar a correção dos pesos de suas camadas;

BTT?

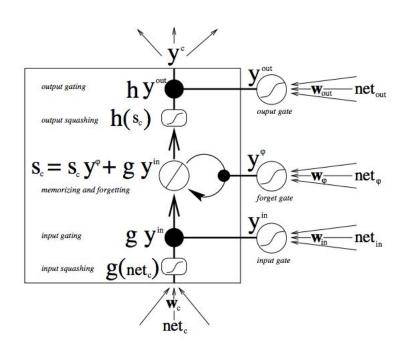
- RNN's utilizam um conceito de backpropagation through time (BTT) para realizar a correção dos pesos de suas camadas;
- A diferença para o método tradicional é que no BTT existe uma ordem bem definida das séries geradas por cada iteração da rede, criando uma conexão e assim permitindo o cálculo para correção dos pesos.

BTT?

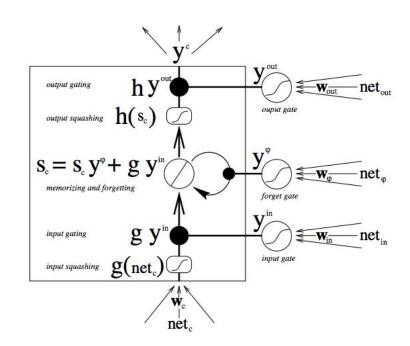


Backpropagation Through Time

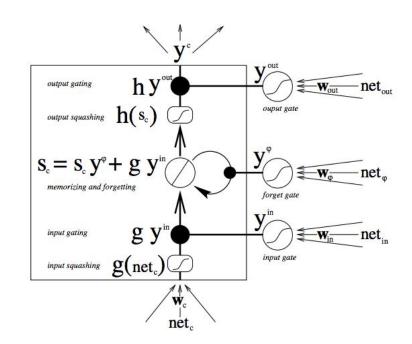
 Long Short-Term Memory Unit (LSTM);



- Long Short-Term Memory Unit (LSTM);
- Possui um input gate, um output gate e um forget gate;



- Long Short-Term Memory Unit (LSTM);
- Possui um input gate, um output gate e um forget gate;
- Tem a propriedade de aprender uma parcela de informação por iteração realizada.



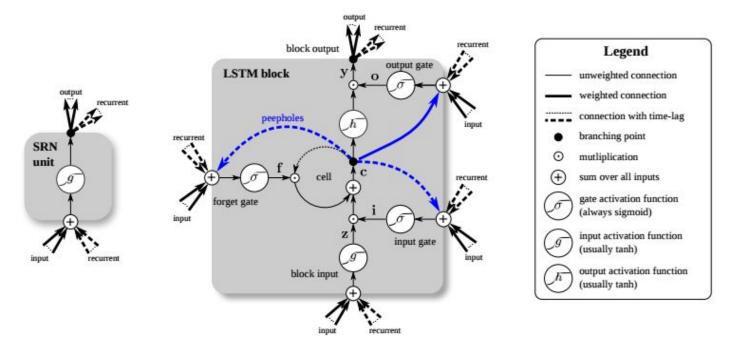
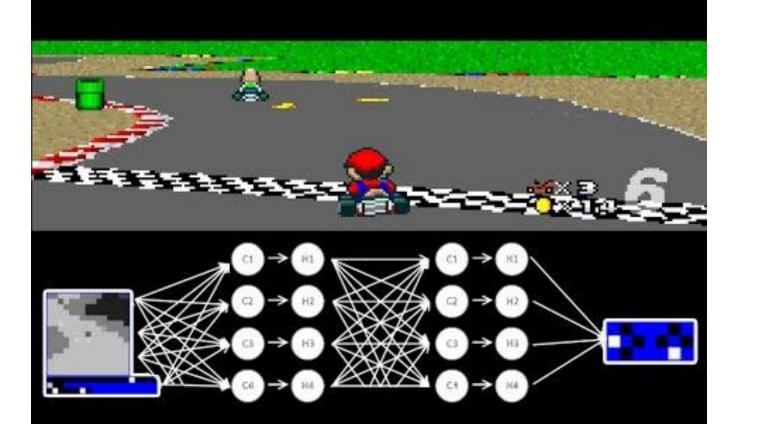


Figure 1. Detailed schematic of the Simple Recurrent Network (SRN) unit (left) and a Long Short-Term Memory block (right) as used in the hidden layers of a recurrent neural network.

THE RECURRENT NEURAL NETWORK MODEL





RNN's

Atividade 3