

# Mimetização de estilos literarios a partir de redes neurais

## recorrentes

Tobias Veiga, João Secundino, João Pedro Mattos

ICMC - Universidade de São Paulo  
São Carlos, SP, Brasil

(tobias.veiga; jp.secundino; joao.pedro.mattos)@usp.br

### 1. Introdução e objetivos

O projeto proposto tem como objetivo realizar a mimetização de personalidades literárias através do uso de redes neurais recorrentes para a geração de sentenças curtas e naturais.

### 2. Redes neurais recorrentes

Redes neurais recorrentes são um tipo de rede neural que trabalha de maneira sequencial sobre uma entrada, tomando decisões que dependem não só do elemento atual processado como também de decisões tomadas em relação aos elementos processados anteriormente.

images/rrn2.png

Figura 1: Representação esquemática do funcionamento de uma Rede Neural Recorrente.

### 3. Método proposto

#### 3.1 Bases de dados

Duas obras foram utilizadas: Brás Cubas e a Bíblia. Essas foram escolhidas pois possuem licença livre, possuem traços textuais evidentes que marcam o estilo do texto e são longas o bastante para gerar um número suficiente de amostras para treinar um modelo de predição. Para treinar o modelo que mimetiza o estilo da Bíblia e de Brás Cubas, foram selecionadas aleatoriamente 500.000 e 30.000 sequências de caracteres, respectivamente. O comprimento das sequências foi definido em 40 e 80 caracteres para os dois livros.

#### 3.2 Modelo de predição

O sistema para recomendações de caracteres consiste de uma Rede Recorrente, utilizada da maneira descrita a seguir. Dada uma sentença de entrada, a rede percorre um número fixo de caracteres consecutivos a partir do início da frase, um a um, processando-os para predizer o último caractere dessa subsequência. Em seguida, a rede avança o intervalo da subsequência em um caractere, repetindo o processo até que um número definido de palavras completas seja predito. A Figura 2 apresenta uma demonstração gráfica do passos do algoritmo. Em cada iteração, a rede leva em conta a sequência de caracteres processados até ali por meio de um módulo de memória que salva uma representação da informação analisada anteriormente. Assim, são preditos novos caracteres que fazem sentido com o contexto do vocabulário/estilo literário de cada autor, mantendo assim a semântica da nova sequência a ser gerada.

images/1.pdf

Figura 2: Representação passo a passo do processamento sobre a entrada.

### 4. Resultados Obtidos

As Tabelas 1 e 2 exibem exemplos de recomendações geradas pelos modelos treinados em sequências de 40 caracteres.

Tabela 1: Exemplos de recomendações do modelo treinado na Bíblia

| Sequência inicial   | Sequência prevista   |
|---|--|
| companhia dos sacerdotes que matam no ca<br>se apanham com a rede maligna, e como os<br>a que o Senhor teu Deus te dá, não apren<br>tuas tendas; então levantarás o teu rost<br>us de Israel! Eis que trarei sobre este<br>odo o seu trabalho é dom de Deus. Eu sei<br>e me instruíam inclinei o meu ouvido! Qu<br>njo: Não estendas a mão sobre o mancebo,<br>sando bem cada oportunidade. A vossa pal | minho de terra de transgressão a<br>seus profetisse a sua casa<br>tas a sua casa de Deus<br>o e todos os seus profetisses<br>pergunto e a terra de Cristo,<br>o Senhor se alguma a<br>ando se fazer a sua casa<br>e todos os seus casas<br>avra do Senhor Deus de Israel |

Tabela 2: Exemplos de recomendações do modelo treinado em Brás Cubas

| Sequência inicial  | Sequência prevista  |
|--|---|
| ntar summariamente um galante episodio d<br>não importa a idade do adulado; a mulh<br>que ha, quando se não resolve um enigma,<br>der nada, a principio, depois entendendo<br>al é a materia do capitulo anterior. pod<br>enha cá jantar amanhã, disse-me o outra<br>ão sei se contava commigo na sala; mas e<br>ças de roupa, a batel-as, a ensaboal-as,<br>iro chinez, bojudo, destro, escanhoando<br>versos, e com elles ia tambem passando a<br>lhar interrogativo ou suspeito. pergun | a lha filha virgilia a talva<br>a trimagilia palama, tomantavamas a parta<br>em aguita filha pata pala<br>que tinha filha ficação, que<br>a tem filha pala com filha<br>uma nala. trava uma cala a<br>ma filha jatalha, a viva á<br>com a malha palação,— filha<br>com a cara, a fica, e<br>milha para para pala para<br>tava a lama a lama a |

Alguns padrões foram observados nas recomendações geradas pelos modelos:

- As palavras geradas quase sempre existem e, quando não existem, parecem com pala-  
vras já existentes.
- O texto sugerido normalmente não faz muito sentido, mas sua estrutura visual é convin-  
cente, ou seja, se alguém apenas ver o texto sem ler, pode ter a impressão de que trata-se  
de um texto comum.
- Algumas palavras aparecem com muito frequência, tais como 'terra', 'Israel', 'Deus', 'Se-  
nhor', além de outras preposições e pronomes, o que indica preferência do modelo de  
escolher termos que estatisticamente aparacem mais, para obter uma acurácia melhor  
durante o treino, em detrimento de realmente aprender a representar a linguagem.

Não conseguimos perceber diferença entre as predições geradas pelos modelos treinados em 40 e 80 caracteres. Além disso, aumentar a complexidade dos modelos permitiu atingir acurácias maiores, entretanto verificamos que atingir acurácias maiores não implicava em recomendar sentenças mais realistas. Isso indica que se tornar melhor em prever o próximo caractere não garante que o modelo será capaz de representar melhor a linguagem.

### 5. Conclusões

O trabalho desenvolvido teve resultados satisfatórios. Mesmo diante das limitações de ter apenas 40 (ou 80) caracteres de informação e ter que prever caracteres em vez de palavras pre determinadas, os modelos foram capazes de gerar palavras pertencentes a linguagem. Algumas direções para melhorar o trabalho incluem: treinar em múltiplas obras e tentar manter o estilo de um obra específica; prever palavras da linguagem em vez de caracteres e comparar essas abordagens.

Apoio

santander.jpg

prceu.png