Composição Automática de Músicas utilizando Redes Neurais Recorrentes

Nicolas Mathias Hahn
Orientador: Guilherme Pumi

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade Federal do Rio Grande do Sul (UFRGS)

Outubro, 2022
Porto Alegre - RS

Contextualização

- Composição algorítmica refere-se ao processo de criação de músicas com o mínimo de intervenção humana.
- A obra *Musikalisches Wurfelspiel* (*Dice Music*), de Wolfgang Amadeus Mozart (1756-1791), utilizou tal processo.

Musikalisches Würfelspiel

18th century

Rolls the dice 16 times (green row); Counts the points (orange column); Find the measure number (row/column intersection); Copy and paste in a score.

	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	15°	16°
2	96	22	141	41	105	122	11	30	70	121	26	9	112	49	109	14
3	32	6	128	63	146	46	134	81	117	39	126	56	174	18	116	83
4	69	95	158	13	153	55	110	24	66	139	15	132	73	58	145	79
5	40	17	113	85	161	2	159	100	90	176	7	34	67	160	52	170
6	148	74	163	45	80	97	36	107	25	143	64	125	76	136	1	93
7	104	157	27	167	154	68	118	91	138	71	150	29	101	162	23	151
8	152	60	171	53	99	133	21	127	16	155	57	175	43	168	89	172
9	119	84	114	50	140	86	169	94	120	88	48	166	51	115	72	111
10	98	142	42	156	75	129	62	123	65	77	19	82	137	38	149	8
11	3	87	165	61	135	47	147	33	102	4	31	164	144	59	173	78
12	54	130	10	103	28	37	106	5	35	20	108	92	12	124	44	131

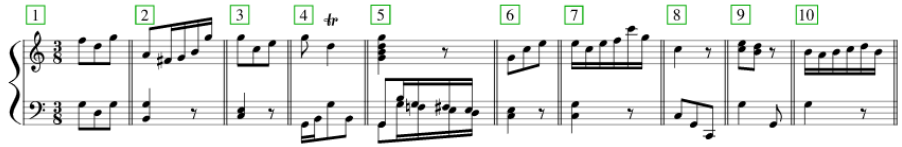


Figura: *Dice Music*

Contextualização

Figura: Tabuleiro de Xadrez com Foto Receptor

- A performance *Reunion*, de John Cage (1912-1992), também utilizou composição automática.
- David Cope (1941-), em 1981, criou o EMI (*Experiments in Musical Intelligence*), um sistema com descrições de estilos composicionais capaz de criar os próprios.

Literatura

- Via de regra, o problema de composição musical automática é explorado com foco na composição musical em si.
- Detalhes técnicos como os impactos que as modificações nos parâmetros têm na composição final, ainda, são amplamente desconhecidos.
- Exemplos: Agarwala et al. (2017), Kuang and Yang (2021).

Objetivos

- Estudar o quão sensível é um modelo de rede neural, baseado em processamento de linguagem natural, construído para composição musical.
- A mensuração será feita com a perplexidade, uma medida oriunda da teoria da informação.
- Avaliar, de forma subjetiva, as peças musicais obtidas em relação à musicalidade e à qualidade.

RNN - Redes Neurais Recorrentes

- RNNs são um tipo de redes neurais criadas para processar séries temporais e outros tipos de dados sequenciais (Fan et al., 2021).

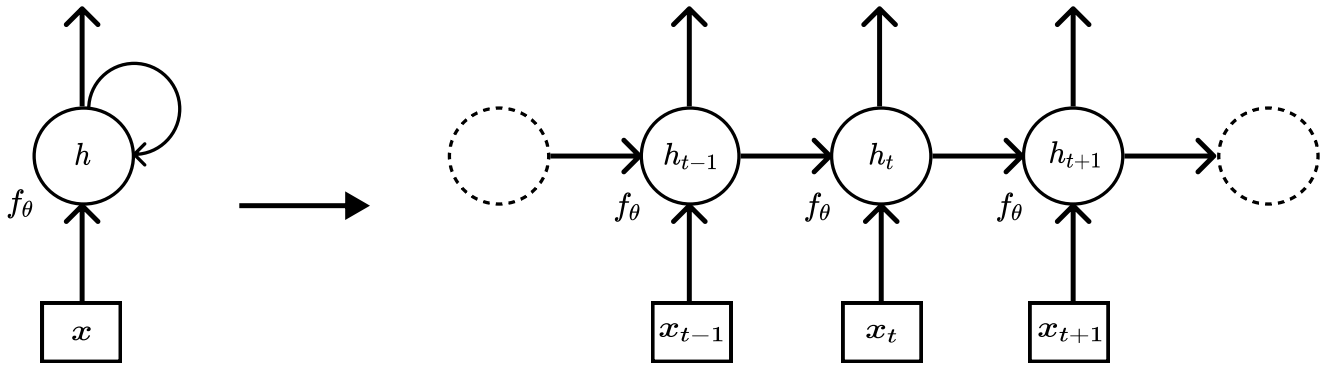


Figura: Diagrama de uma RNN *Vanilla* (adaptado de Goodfellow et al., 2016; Kamath et al., 2019)

LSTM - *Long Short-Term Memory*

- De acordo com Goodfellow et al. (2016), as LSTM fazem parte de uma classe de modelos chamada de RNN fechadas (*gated RNN*).
- Os portões (*gates*), que também são camadas da rede neural, controlam o fluxo de informação, mantendo ou descartando o estado oculto $\boldsymbol{h}_t$ a cada passo temporal (Kamath et al., 2019).
- É uma das variantes de RNN desenvolvidas para contornar o problema de dissipação (ou explosão) do gradiente, que ocorre no ajuste da rede.

LSTM - *Long Short-Term Memory*

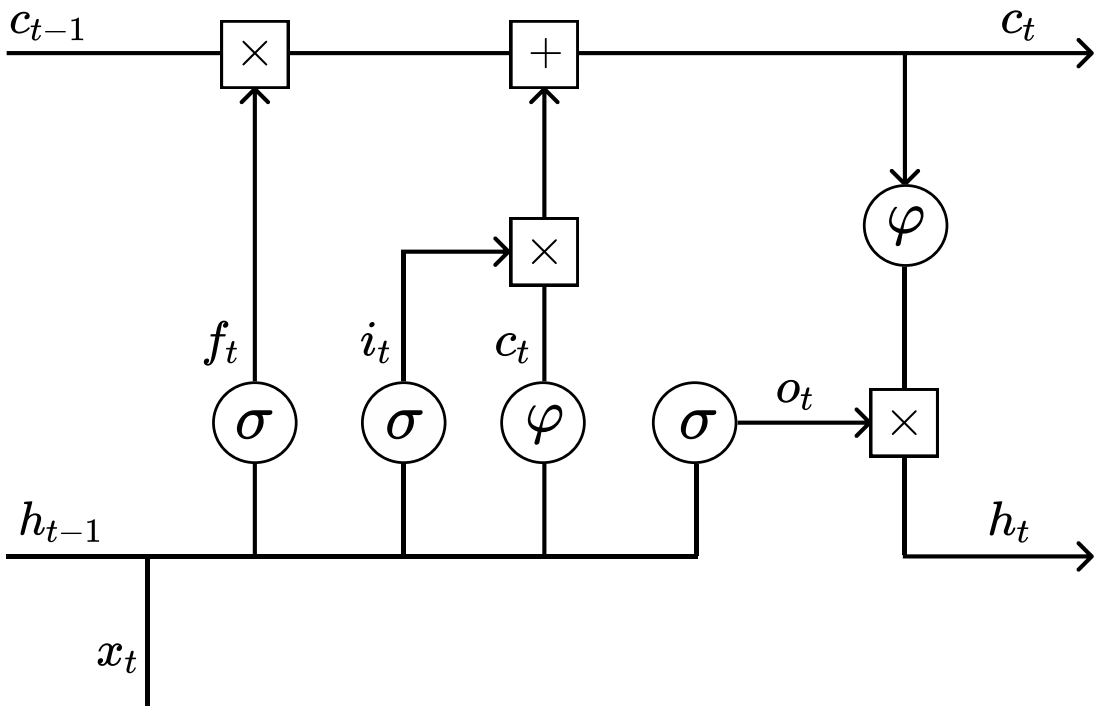
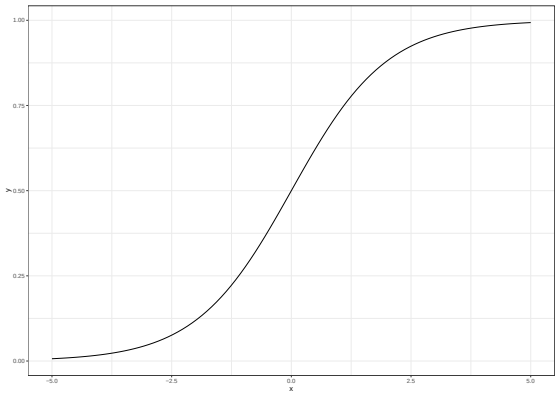


Figura: Diagrama de uma LSTM. Considere σ como a função de ativação *logit* (adaptado de Kamath et al., 2019).

Função de Ativação

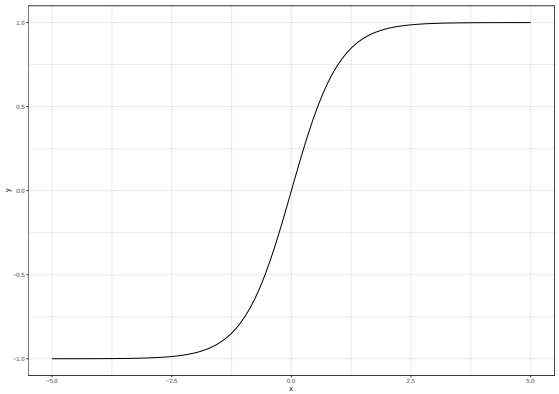
Sigmóide (*logit*)



$f : \mathbb{R} \to (0, 1)$ dada por

$$f(x) := \frac{1}{1 + e^{-x}}$$

Tangente Hiperbólica (*tanh*)



$f : \mathbb{R} \to (-1, 1)$ dada por

$$f(x) := \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Ajustando uma RNA

- Os pesos da rede são ajustados por meio do método do gradiente descendente estocástico (SGD).
- Utiliza-se uma partição aleatória dos dados de treino, denominada lote.
- A taxa de aprendizagem ϵ define o tamanho do passo em direção ao gradiente negativo.
- Uma passagem pelo conjunto de dados de treinamento é denominada época.

PLN - Processamento de Linguagem Natural

- Aplicação de métodos estatísticos e computacionais para modelar e extrair informações da linguagem humana (Kamath et al., 2019).
- É importante testar os algoritmos em mais de uma linguagem, especialmente em linguagens com diferentes propriedades (Jurafsky and Martin, 2021).
- Um modelo estatístico de linguagem é aquele que atribui probabilidades para uma sequência de *tokens*.

PLN - Perplexidade

- A perplexidade $\mathcal{P}(W) = e^{H(W)}$ é uma medida para a avaliação intrínseca de um modelo de linguagem.
- Medidas menores de perplexidade são indicativas de uma predição melhor.
- A perplexidade de dois modelos de linguagem apenas podem ser comparadas se ambos utilizam o mesmo vocabulário.

Notação ABC

- Notação ABC é um sistema popular de notação musical baseada em texto para transcrever, publicar e compartilhar músicas, particularmente de forma *online*.

```
X:1
T:Speed the Plough
M:4/4
C:Trad.
K:G
|:GABc dedB|dedB dedB|c2ec B2dB|c2A2 A2BA|
  GABc dedB|dedB dedB|c2ec B2dB|A2F2 G4:|
|:g2gf gdBd|g2f2 e2d2|c2ec B2dB|c2A2 A2df|
  g2gf g2Bd|g2f2 e2d2|c2ec B2dB|A2F2 G4:|
```

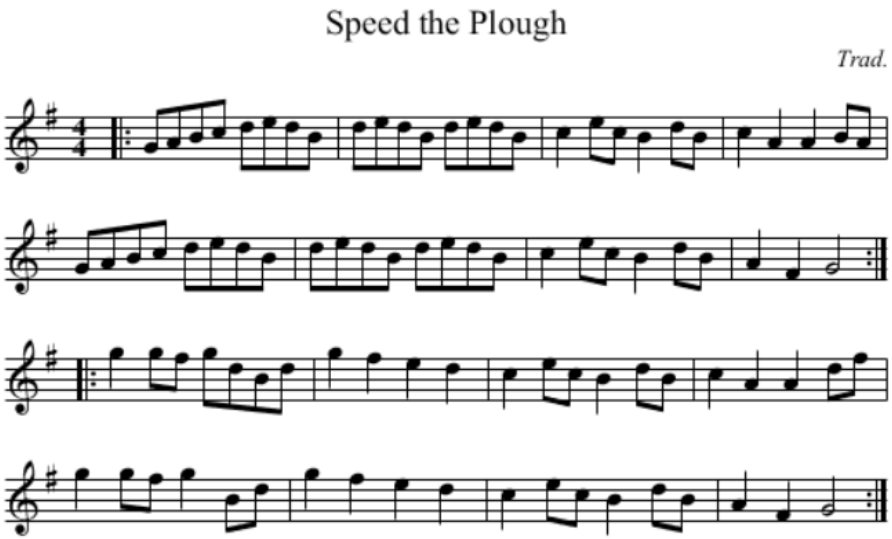


Figura: Exemplo de notação ABC convertendo em música.

Web Scraping

- De acordo com Lawson (2015); Patil and Patil (2016), envolve dois programas:
  - *crawler*: sistematicamente coleta os dados da Internet;
  - *scraper*: extrai a informação relevante e armazena em uma base de dados.

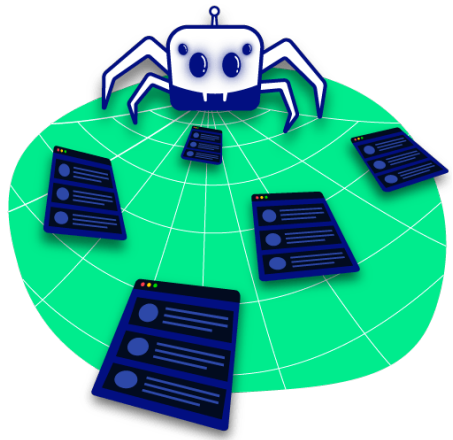


Figura: *Crawler* coletando páginas *web*.

Bases de Dados - Fontes

Irish

- contém 817 músicas folclóricas irlandesas no formato *.abc*;
- versão disponibilizada pelo Instituto de Tecnologia de Massachusetts (MIT).

ABC Notation

- coleta, via *web scraping*, do site abcnotation.com;
- foram coletados 184.900 músicas;
- selecionada uma amostra de 5.000.

Bases de Dados - Tratamentos

- **tratamento**: via expressões regulares, removeu-se caracteres que não afetam diretamente as músicas (título, letra de música);
- **união**: todas as músicas foram "coladas", como se fizessem parte de um único texto.
- **tokenização**: para cada caractere (*token*) presente, foi criado um único índice.

Modelo - Arquitetura

- Utilizou-se um modelo de RNN-LSTM, que foi ajustado com ambas as bases de dados de forma independente;
- Quatro camadas: entrada, *Embedding*, LSTM e *Dense*;
- Foi feita uma divisão dos dados em 80% treino e 20% teste.

Modelo - Parâmetros (inicial)

- 2000 épocas no ajuste do modelo;
- função perda: entropia cruzada;
- métrica de avaliação: perplexidade;
- segmentou-se em duas etapas.

Modelo - Parâmetros (1ª etapa)

- fixou-se a função de ativação *tanh* na camada LSTM;
- foram alterados os seguintes parâmetros e hiperparâmetros:
  - $vocab\_size \in \{64, 125\}$;
  - $lstm\_units \in \{256, 1024\}$;
  - $embedding\_dim \in \{256, 512\}$;
  - $learning\_rate \in \{10^{-3}, 10^{-5}\}$;
  - $seq\_length \in \{50, 200\}$;
  - $batch\_size \in \{4, 16\}$.
- foram ajustados 64 modelos, sendo 32 para cada base de dados.

Modelo - Parâmetros (2ª etapa)

- selecionou-se os dois melhores e os dois piores modelos da 1ª etapa;
- critério: perplexidade;
- trocou-se função de ativação para *logit*.

Processo Gerador de Música

1. Construir um modelo com os devidos parâmetros;
2. Fixado o modelo, carregam-se os pesos de um modelo similar ajustado previamente;
3. Fornecer uma sequência de caracteres inicial, no caso “X:”;
4. Iterativamente, o modelo estima um novo elemento para compor a sequência até atingir um comprimento definido;
5. Extrair da sequência blocos de texto candidatos a músicas e, ao serem convertidos com sucesso, resultam em músicas.

Irish - 1ª Etapa

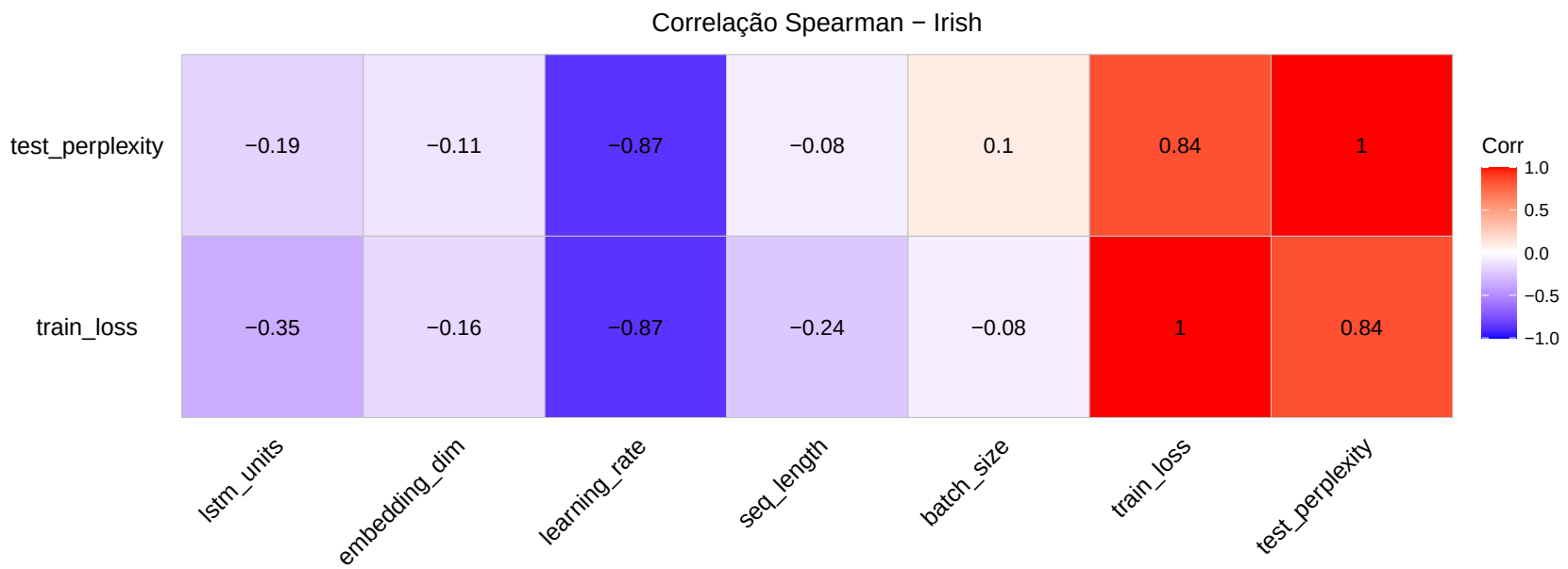


Figura: Correlação entre parâmetros e métricas para Irish

Irish - 2ª Etapa

idx	*train_loss*		*test_perplexity*	
	tanh	*logit*	*tanh*	*logit*
5	2.964	3.021	24.709	27.576
7	2.963	3.135	20.401	24.726
11	0.982	1.195	2.835	3.549
25	1.171	1.389	2.752	3.263

ABC Notation - 1ª Etapa

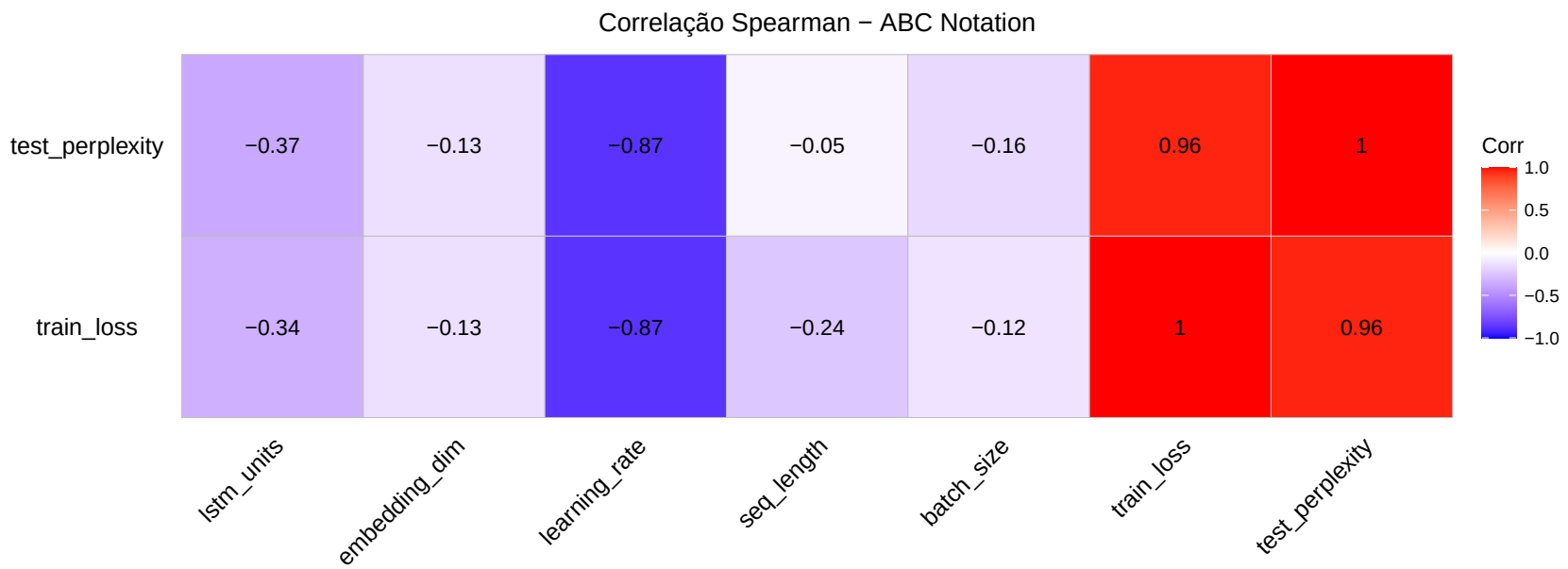


Figura: Correlação entre parâmetros e métricas para ABC Notation

ABC Notation - 2ª Etapa

idx	*train_loss*		*test_perplexity*	
	tanh	*logit*	*tanh*	*logit*
8	3.628	3.670	31.707	33.038
6	3.522	3.571	35.991	37.849
20	1.350	1.616	3.971	5.142
28	1.305	1.585	3.940	4.960

Geração de Músicas

Irish: 24 músicas

- $idx = 11$: 10 músicas
  - *logit*: 4 músicas
  - *tanh*: 6 músicas
- $idx = 25$: 14 músicas
  - *logit*: 6 músicas
  - *tanh*: 8 músicas

ABC Notation: 14 músicas

- $idx = 20$: 8 músicas
  - *logit*: 4 músicas
  - *tanh*: 4 músicas
- $idx = 28$: 6 músicas
  - *logit*: 1 músicas
  - *tanh*: 5 músicas

Percepções sobre as Músicas

Irish

- trechos similares
- voz única
- musicalmente plausível

ABC Notation

- mais variabilidade
- voz múltipla
- musicalmente plausível

Os comentários feitos referente às composições geradas são as percepções do autor com seu limitado conhecimento musical.

Resumo

- Explorou-se o problema de composição automática de músicas.
- Investigou-se a sensibilidade dos modelos via modificação dos parâmetros.
- Mediu-se impactos das mudanças via perplexidade.
- Gerou-se novos arquivos *.abc* que, quando possível, foram convertidos em músicas.

Extensões

- Modificar a arquitetura contemplando o aumento do número de camadas;
- Explorar outros perfis de redes, como a *transformer* (Vaswani et al., 2017);
- Analisar detalhadamente características musicais das composições.

Referências I

Agarwala, N., Inoue, Y., and Sly, A. (2017). Music composition using recurrent neural networks. *CS 224n: Natural Language Processing with Deep Learning, Spring*, 1:1–10.

Fan, J., Ma, C., and Zhong, Y. (2021). A selective overview of deep learning. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 36(2):264.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Jurafsky, D. and Martin, J. H. (2021). Speech and language processing. *US: Prentice Hall*, 3.

Kamath, U., Liu, J., and Whitaker, J. (2019). *Deep learning for NLP and speech recognition*, volume 84. Springer.

Referências II

Kuang, J. and Yang, T. (2021). Popular song composition based on deep learning and neural network. *Journal of Mathematics*, 2021.

Lawson, R. (2015). *Web scraping with Python*. Packt Publishing Ltd.

Patil, Y. and Patil, S. (2016). Review of web crawlers with specification and working. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(1):220–223.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.