Adicionando robustes semântica a sistemas de diálogo

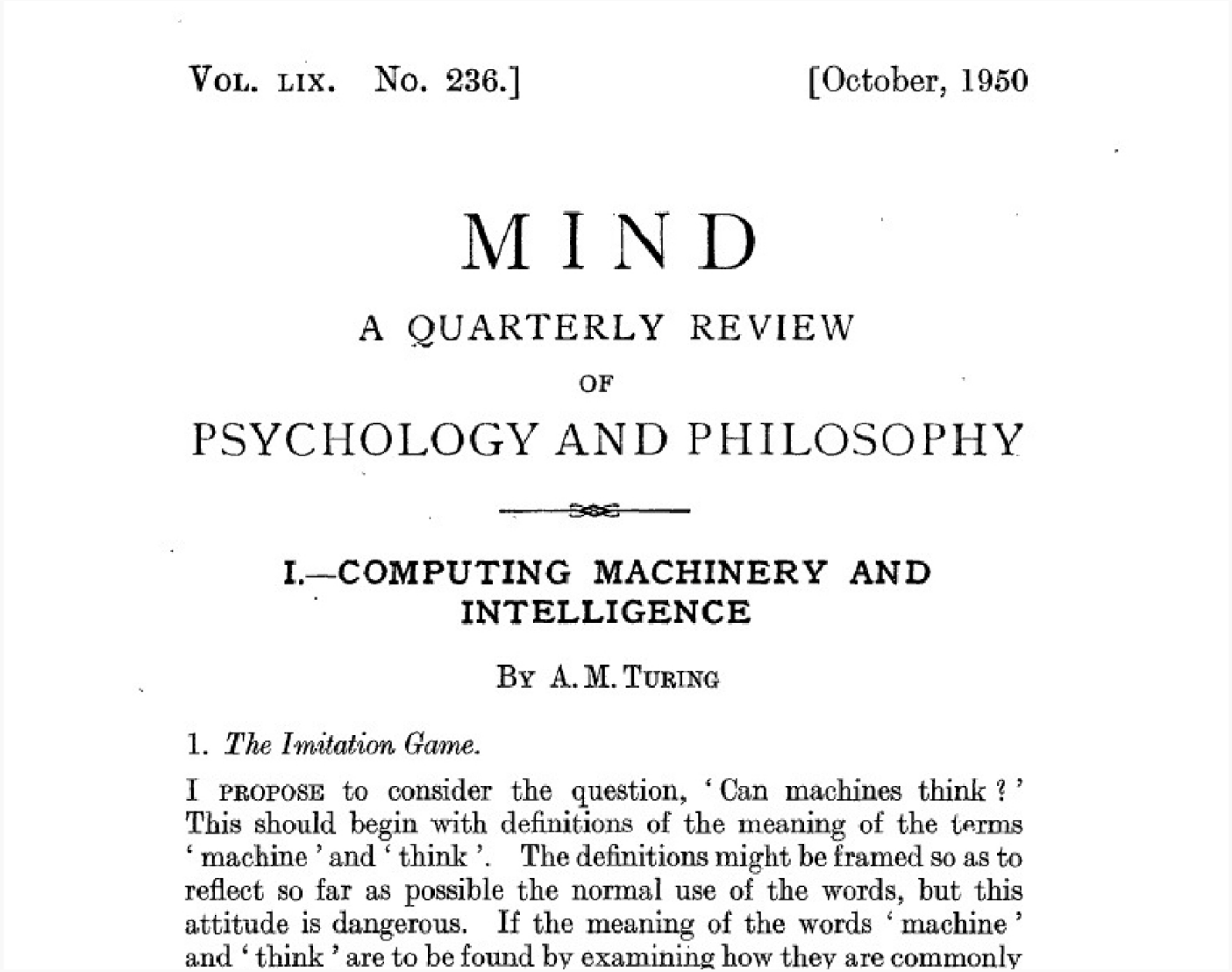
Felipe Salvatore
https://felipessalvatore.github.io/

May 9, 2018

IME-USP: Instituto de Matemática e Estatística - Universidade de São Paulo

Sistemas de diálogo

criar um programa capaz de dialogar com ser humano



VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly

Sistemas de diálogo

goal -driven vs non-goal driven

Sistemas de diálogo baseados em redes neurais

Modelos de linguagem baseados em redes neurais

Nos chamamos de modelo de linguagem uma distribuição de probabildiade sobre uma sequencia de tokens em uma lingua natural.

$$P(x_1, x_2, x_3, x_4) = p$$

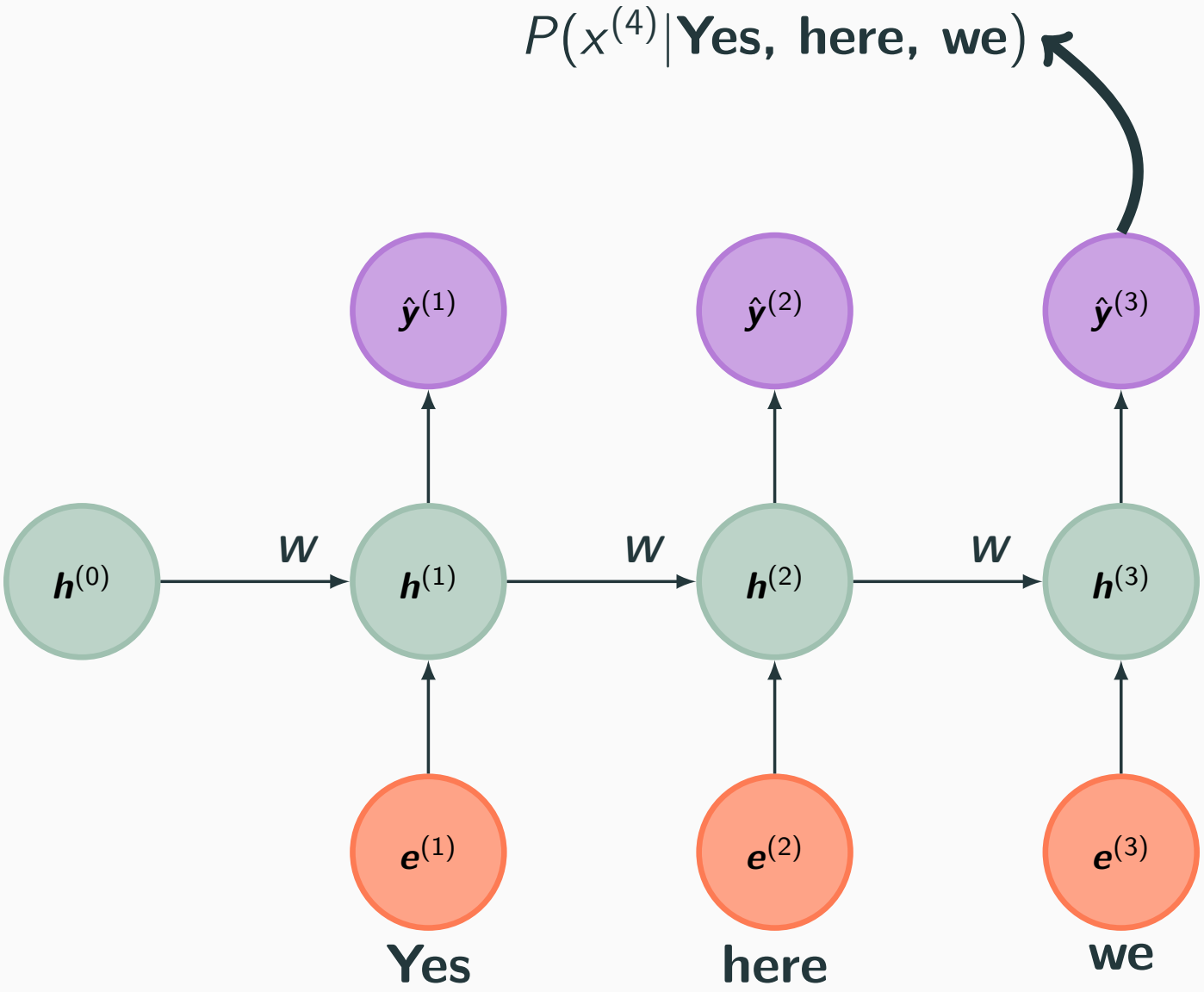
Em vez de usar uma abordagem que seja específica para o domínio da linguagem natural, podemos usar um modelo para predição de dados sequencias: **uma rede recorrente (RNN)**.

Nossa tarefa de aprendizado é estimar a distribuição de probabilidade

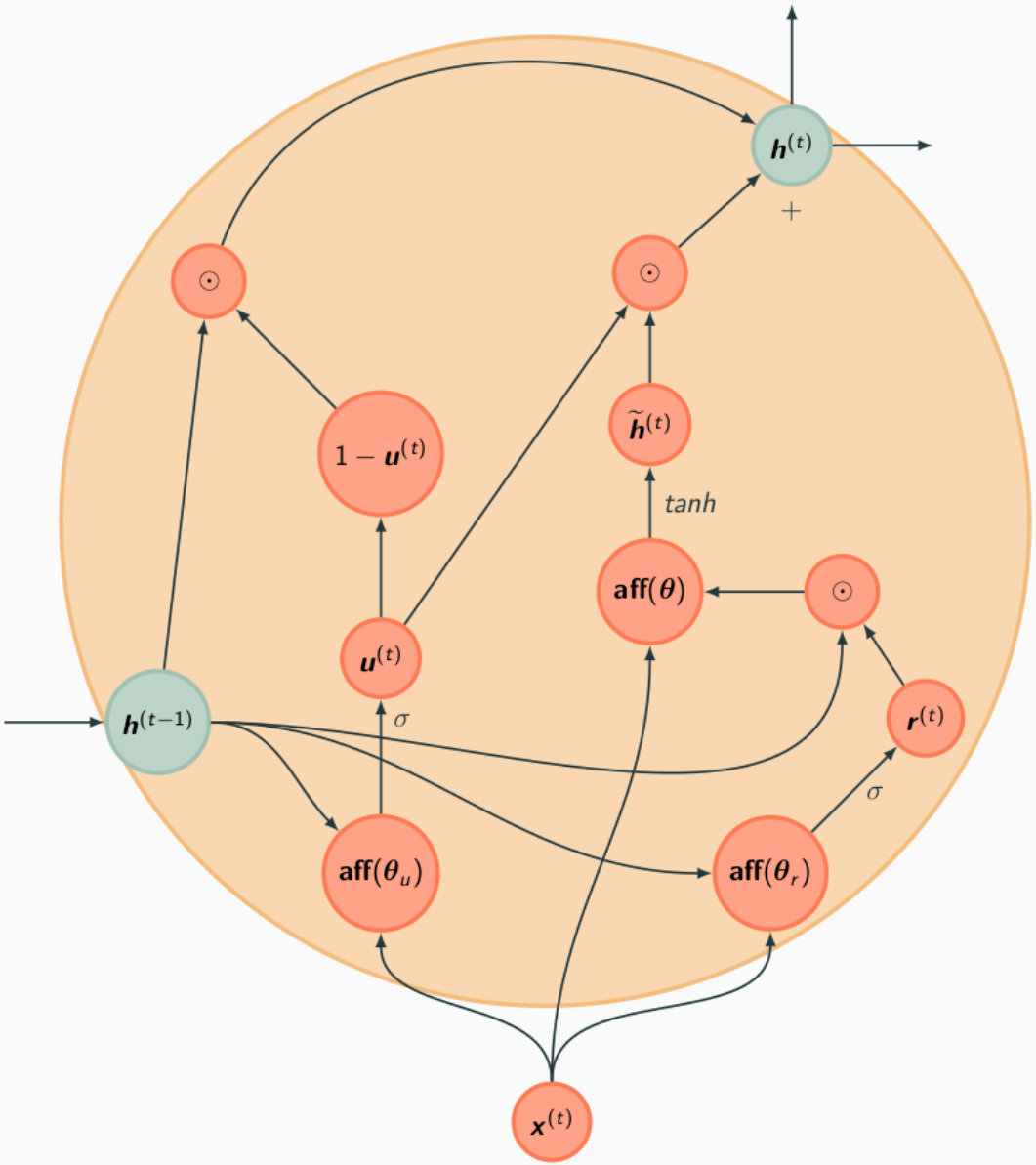
$$P(x_n = \text{palavra}_{j^*} | x_1, \ldots, x_{n-1})$$

para qualquer $(n-1)$ sequencia de palavras $x_1, \ldots, x_{n-1}$.

O modelo de linguagem com RNN



GRU: Gated Recurrent Units



Exemplo: TrumpBot

`https://github.com/felipessalvatore/MyTwitterBot`

Felipe Salvatore
@Felipessalvador

Hillary can make america great again. @greta @MarkBurnettTV #DinheiroNãoCompra #SecretBallot #خسوف_القمر

Traduzir do inglês

15:10 - 7 de ago de 2017

Felipe Salvatore
@Felipessalvador

Obama is all beautiful. I agree with people attacking me. Amazing. @CLewandowski_ #SecretBallot @garyplayer @greta

Traduzir do inglês

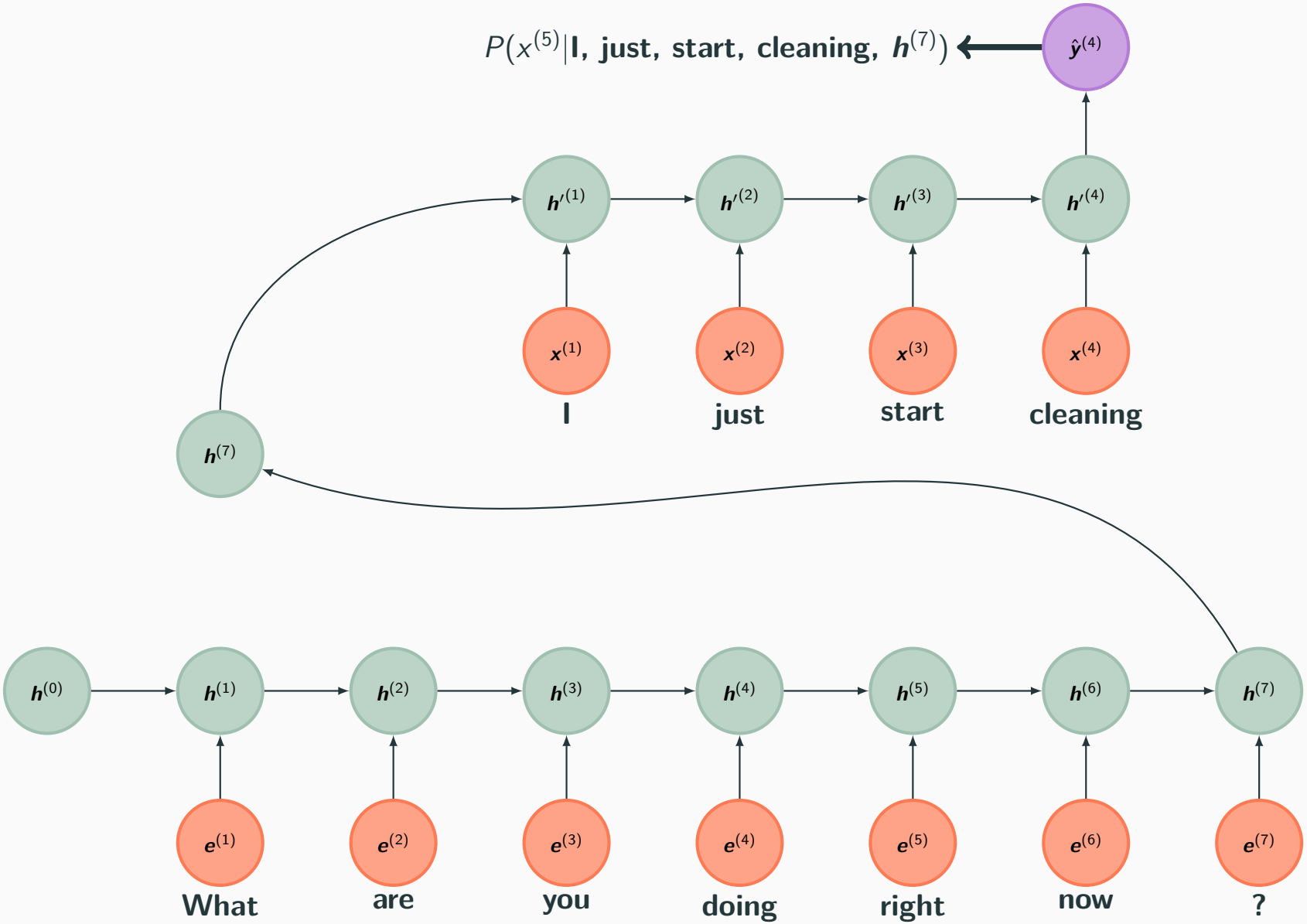
14:40 - 7 de ago de 2017

Exemplo: Funk Generator

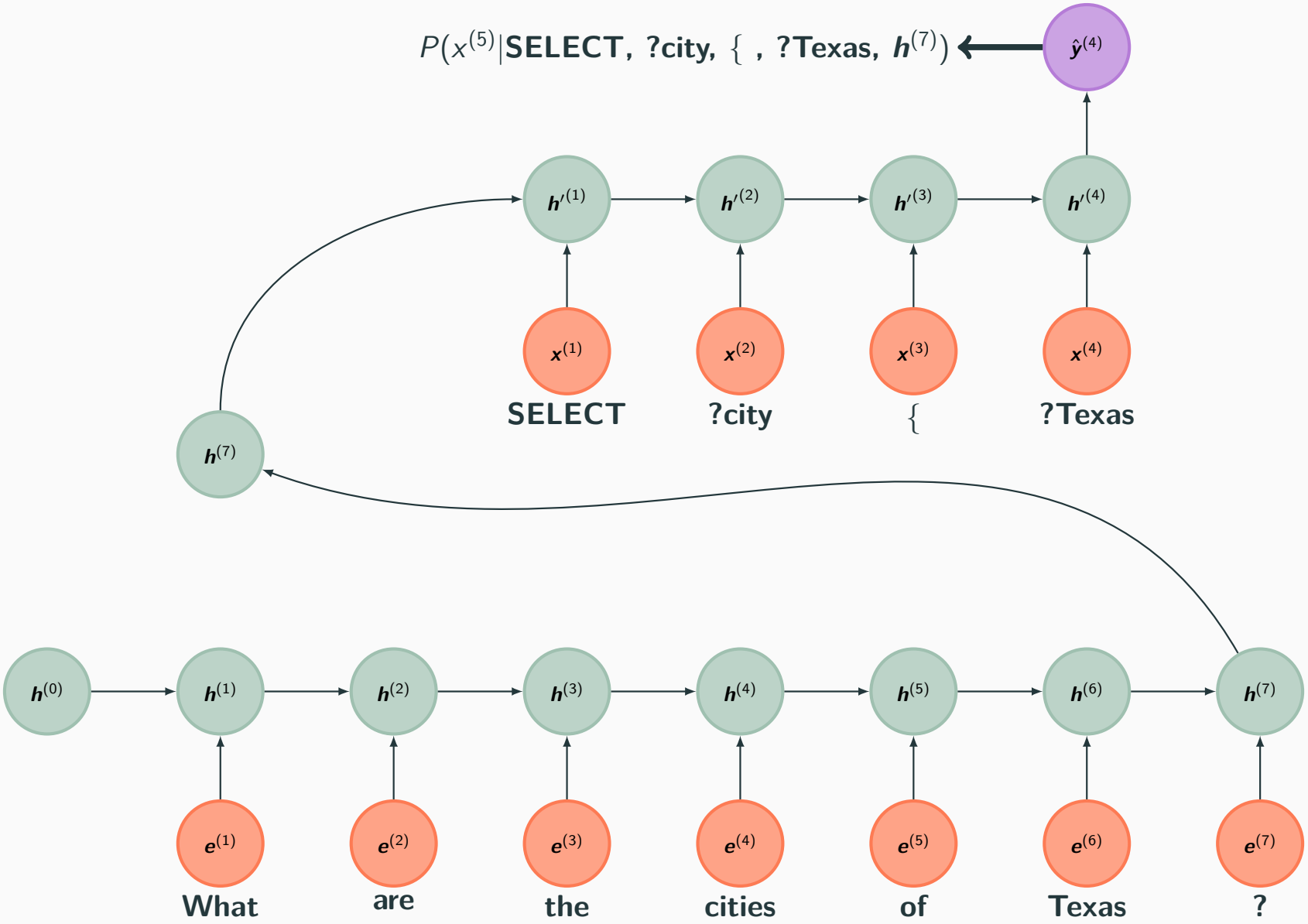
`https://github.com/lucasmoura/funk_generator`

É o dj que tá tocando e não sabe de nada
Eu já tô no clima e já tô no meu nome
Cordão de ouro no pescoço eu tô na moda
Com a camisa da
Louis
Vuitton
Pulo da morena que elas gosta
E se eu te pego no baile
De captiva de citroen ou de hayabusa
Tu viu a 1100 cilindradas
Se eu tô no litoral de cordão de ouro
De cordão de ouro no pescoço

Seq2seq: diálogo [5]



Seq2seq: tradução [3]



Exemplo de diálogo [5]

Human: *what is moral ?*
Machine: *what empowered humanity , what intellectual the essence is .*
Human: *what is immoral ?*
Machine: *the fact that you have a child .*
Human: *what is morality ?*
Machine: *what is altruism ?*
Human: *what is the definition of altruism ?*
Machine: *if you don 't believe in god , then you don 't know .*
Human: *ok ... so what is the definition of morality ?*
Machine: *well , the truth is , you 're not a believer in god almighty .*
Human: *tell me the definition of morality , i am quite upset now !*
Machine: *i 'm not ashamed of being a philosopher !*

Métricas

Avaliação humana [2]

In the first trial, we asked the following questions to the users, for each response:

1. How appropriate is the response overall? (overall, scale of 1-5)
2. How on-topic is the response? (topicality, scale of 1-5)
3. How specific is the response to some context? (specificity, scale of 1-5)
4. How much background information is required to understand the context? (background, scale of 1-5)

Avaliação automática: BLEU [4]

Essa métrica compara n-gramas (até 4) da resposta candidata com os n-gramas da refência da tradução e conta o numero de acertos. Essa métrica também penaliza traudções muito curtas:

$$BLUE(r, \hat{r}) = min\left(1, \frac{len(\hat{r})}{len(r)}\right)\left(\prod_{n=1}^{4} precision_n(r, \hat{r})\right)^{\frac{1}{4}} \quad (1)$$

em que $precision_n(r, \hat{r})$ é o número de overlap de n gramas de r e $\hat{r}$ dividido pelo número de todos os n-gramas de $\hat{r}$.

$BLUE(r, \hat{r}) \in [0, 1]$

Avaliação automática: problemas

Context of Conversation	**Context of Conversation**
A: dearest! question. how many thousands of people can panaad occupy? B: @user panaad has <number> k seat capacity while rizal has <number> k thats why they choose rizal i think .	A: never felt more sad than i am now B: @user aww why ? A: @user @user its a long story ! sure you wanna know it ? bahaha and thanks for caring btw <heart>
Ground Truth Response	**Ground Truth Response**
A: now i know about the siting capacity . thanks for the info @user great evening.	A: @user i don 't mind to hear it i 've got all day and youre welcome <number>
Proposed Response	**Proposed Response**
A: @user makes sense. thanks!	A: @user i know , i 'm just so happy for you !

"In particular, we show that these metrics (BLEU, METEOR, ROUGE) have only a small positive correlation on the chitchat oriented Twitter dataset, and no correlation at all on the technical Ubuntu Dialogue Corpus." [1]

De diálogos abertos para pequenas tarefas

bAbI

Entailment-QA

SICK

quora

first results

Referências I

C. Liu, R. Lowe, I. V. Serban, M. Noseworthy, L. Charlin, and J. Pineau.
How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation.
CoRR, abs/1603.08023, 2016.

R. Lowe, M. Noseworthy, I. V. Serban, N. Angelard-Gontier, Y. Bengio, and J. Pineau.
Towards an automatic turing test: Learning to evaluate dialogue responses.
CoRR, abs/1708.07149, 2017.

Referências II


F. F. Luz and M. Finger.
Semantic parsing natural language into sparql: an lstm enconder- decoder neural net approach.
2017.


K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu.
Bleu: A method for automatic evaluation of machine translation.
In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318. Association for Computational Linguistics, 2001.


O. Vinyals and Q. V. Le.
A neural conversational model.
CoRR, abs/1506.05869, 2015.