Resolução de Correferência utilizando Árvores Latentes Com Representação Contextual

Dissertação de Mestrado

Leonardo Oliveira

Novembro - 2020

Descrição do Problema

Resolução de Correferência

North Korea opened its doors to the U,S, today, welcoming Secretary of State Madeleine Albright. She says her visit is a good start. The U.S. remains concerned about North Korea missile development program and its exports of missiles to Iran.

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

Resolução de Correferência

North Korea opened its doors to the U.S. today, welcoming Secretary of State Madeleine Albright. She says her visit is a good start. The U.S. remains concerned about North Korea missile development program and its exports of missiles to Iran.

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

Resolução de Correferência

- Já existe como uma tarefa de NLP há muito tempo

- Apenas em 2011 na CoNLL foram definidos
  - Dataset de tamanho grande o suficiente
  - Métricas claras o suficiente (Score CoNLL)

- Na CoNLL de 2012
  - Introduzidos os datasets Chinês e Árabe
  - Primeiros modelos considerados bons o suficiente

Estado da Arte - Correferência

- CoNLL 2011/2012 metric - English only

Modelo	Ano	CoNLL Score
Joshi, et al. (30)	2019	79.6
Joshi, et al. (65)	2019	76.9
Kantor and Globerson. (116)	2019	76.6
Fei, et al. (36)	2019	73.8
Lee, et al. (29)	2018	73.0
Peters, et al. (63)	2018	70.4
Lee, et al. (28)	2017	68.8
Wiseman, et al. (55)	2016	64.2
Fernandes e Milidiú (6)	2012	63.4

Tabela 3.4: Estado da arte atual

Motivação

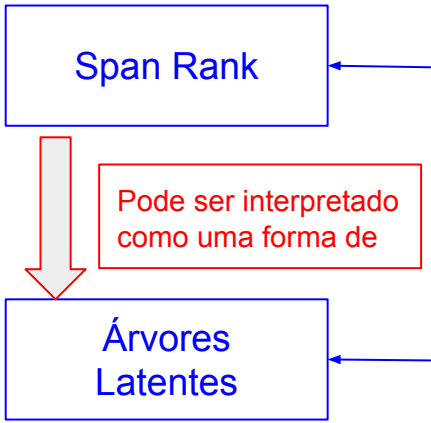
Span Rank

Árvores Latentes

Modelo	Ano	CoNLL Score
Joshi, et al. (30)	2019	79.6
Joshi, et al. (65)	2019	76.9
Kantor and Globerson. (116)	2019	76.6
Fei, et al. (36)	2019	73.8
Lee, et al. (29)	2018	73.0
Peters, et al. (63)	2018	70.4
Lee, et al. (28)	2017	68.8
Wiseman, et al. (55)	2016	64.2
Fernandes e Milidiú (6)	2012	63.4

Tabela 3.4: Estado da arte atual

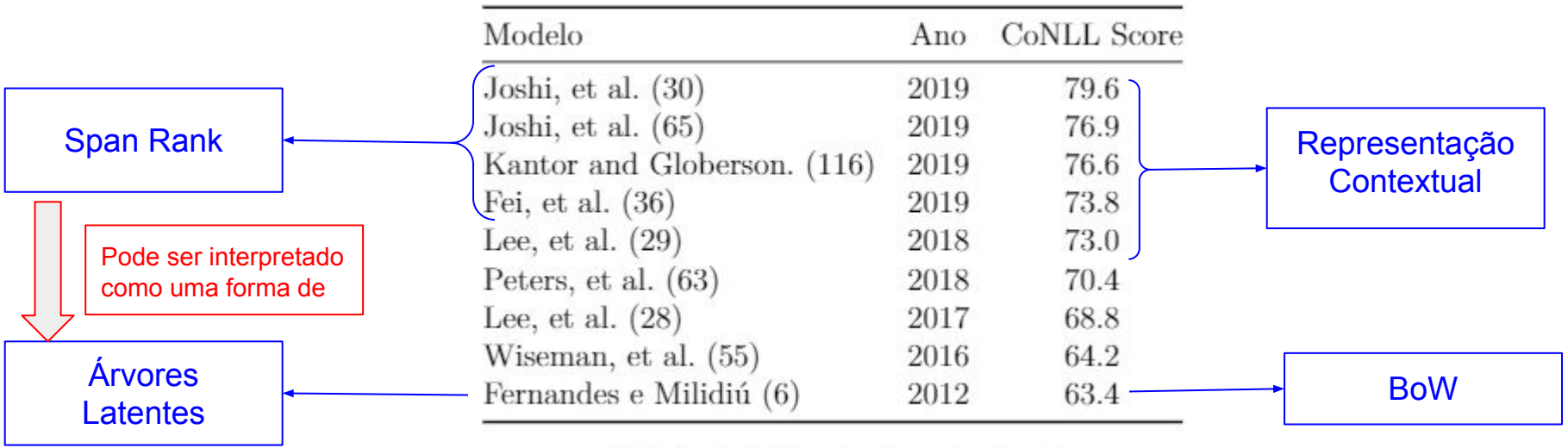
Motivação



Modelo	Ano	CoNLL Score
Joshi, et al. (30)	2019	79.6
Joshi, et al. (65)	2019	76.9
Kantor and Globerson. (116)	2019	76.6
Fei, et al. (36)	2019	73.8
Lee, et al. (29)	2018	73.0
Peters, et al. (63)	2018	70.4
Lee, et al. (28)	2017	68.8
Wiseman, et al. (55)	2016	64.2
Fernandes e Milidiú (6)	2012	63.4

Tabela 3.4: Estado da arte atual

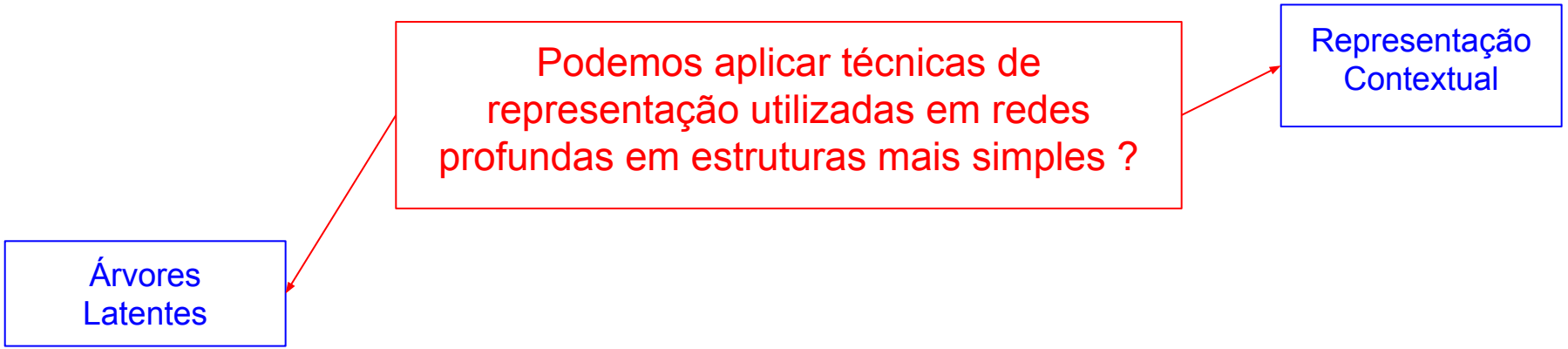
Motivação



Modelo	Ano	CoNLL Score
Joshi, et al. (30)	2019	79.6
Joshi, et al. (65)	2019	76.9
Kantor and Globerson. (116)	2019	76.6
Fei, et al. (36)	2019	73.8
Lee, et al. (29)	2018	73.0
Peters, et al. (63)	2018	70.4
Lee, et al. (28)	2017	68.8
Wiseman, et al. (55)	2016	64.2
Fernandes e Milidiú (6)	2012	63.4

Tabela 3.4: Estado da arte atual

Motivação


Podemos aplicar técnicas de representação utilizadas em redes profundas em estruturas mais simples ?
Representação Contextual
Árvores Latentes

Objetivo

Verificar a eficiência de

métodos automáticos de geração de features

das árvores latentes

para tarefa de correferência

Objetivo

Verificar a eficiência de

métodos automáticos de geração de features

das árvores latentes

para tarefa de correferência

Ou seja,
Gerar as features para uma árvore latente utilizando representação contextual

Resultados

Modelo	MUC	B-CUBED	CEAF-E	Score CoNLL
span_surface_2E6	61.72	52.04	45.89	53.2
span_surface	57.85	48.13	42.05	49.34
baseline	56.08	45.41	39.88	47.12
lexico_efi	48.89	37.01	31.07	38.99
glove_efi	44.23	33.95	28.78	35.65
bert_surface	24.99	30.88	25.51	27.13

Resultados no dataset de validação

Representação de palavras

Representação de palavras

This is a sample text. Nothing here is used.

Representação de palavras

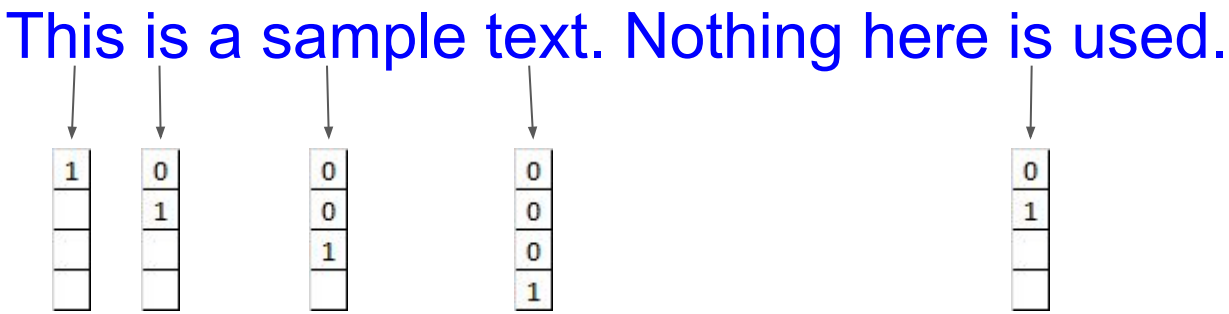
Bag of Words

This is a sample text. Nothing here is used.

Cada palavra é uma posição em um vetor binário estilo one-hot

Representação de palavras

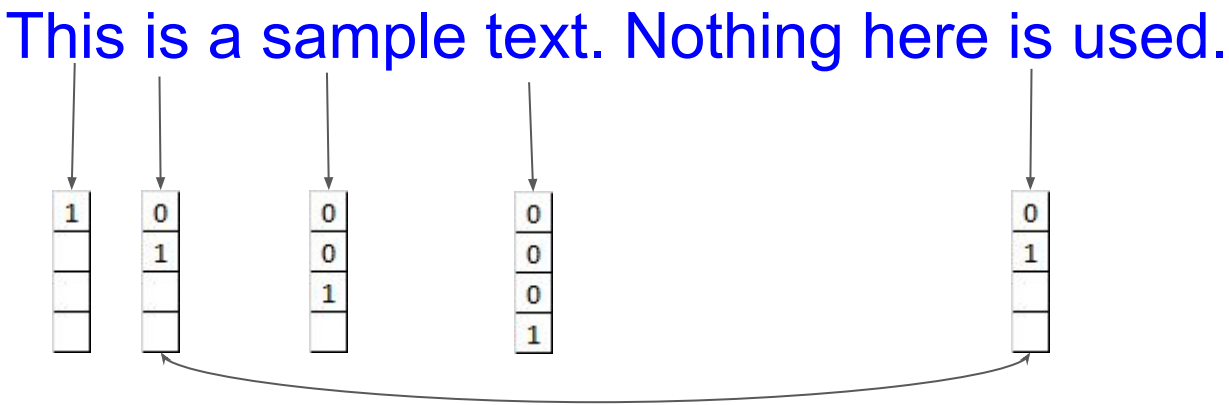
Bag of Words



Cada palavra é uma posição em um vetor binário estilo one-hot

Representação de palavras

Bag of Words



Cada palavra é uma posição em um vetor binário estilo one-hot

Conversão é trivial

Muitas dimensões
Cada dimensão é carrega pouco significado

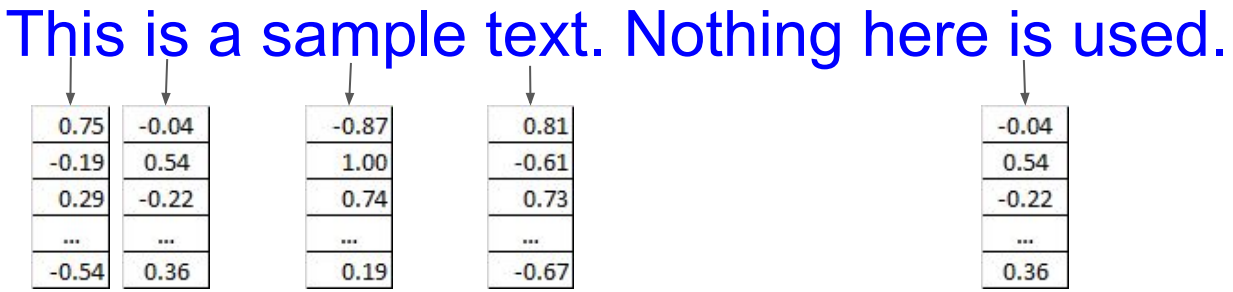
Representação de palavras

GloVe

This is a sample text. Nothing here is used.

Representação de palavras

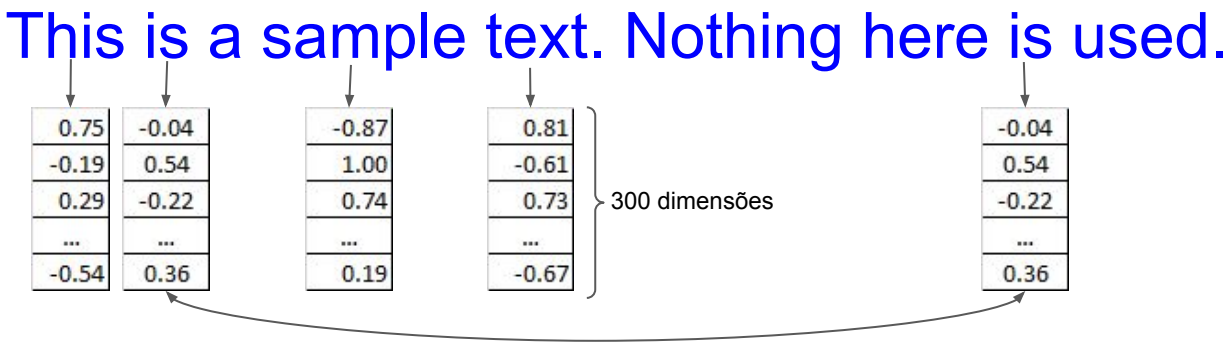
GloVe



Cada palavra recebe um vetor de números reais, construído de forma a representar um contexto global

Representação de palavras

GloVe



Cada palavra recebe um vetor de números reais, construído de forma a representar um contexto global

Aprendizado não é muito caro
Boa performance em muitas tarefas de NLP
Os vetores possuem alguma semântica

Todas as ocorrências da palavra possuem o mesmo vetor
Palavras muito frequentes não possuem uma representação com muito significado

Representação de palavras

GloVe

Os vetores possuem semântica:

king - man + woman = queen

Paris - France + Italy = Rome

Australian - Australia + Germany = German

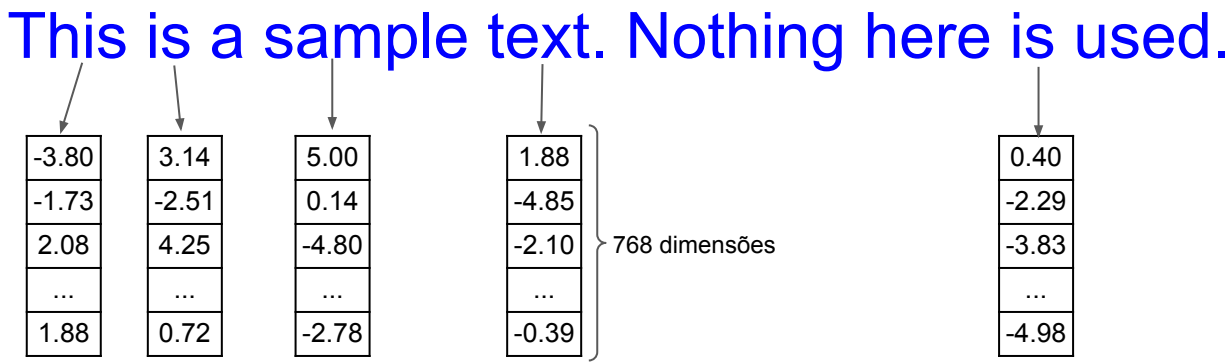
Representação de palavras

BERT - Representação Contextual

This is a sample text. Nothing here is used.

Representação de palavras

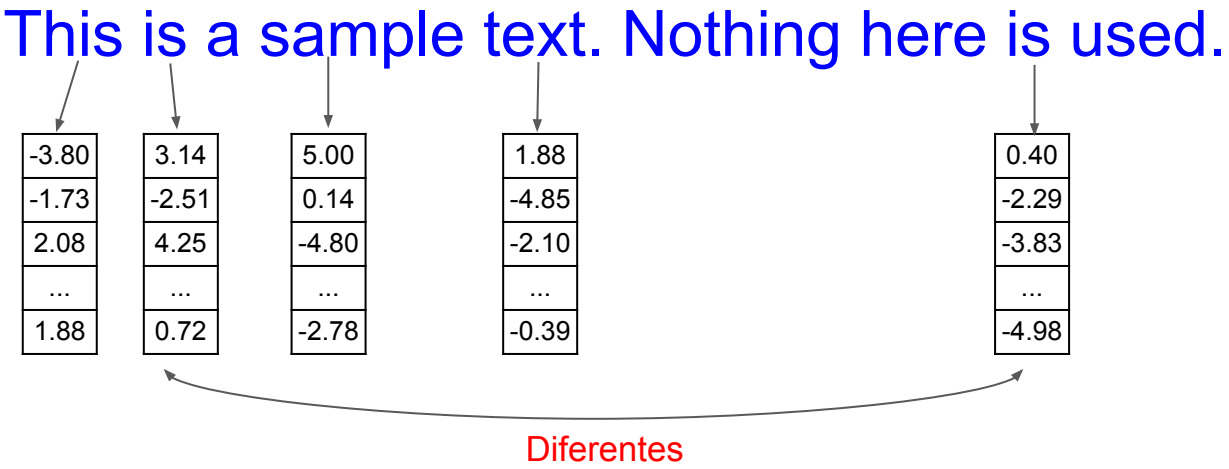
BERT - Representação Contextual



Cada palavra recebe um vetor de números reais, construído de forma a representar um contexto global

Representação de palavras

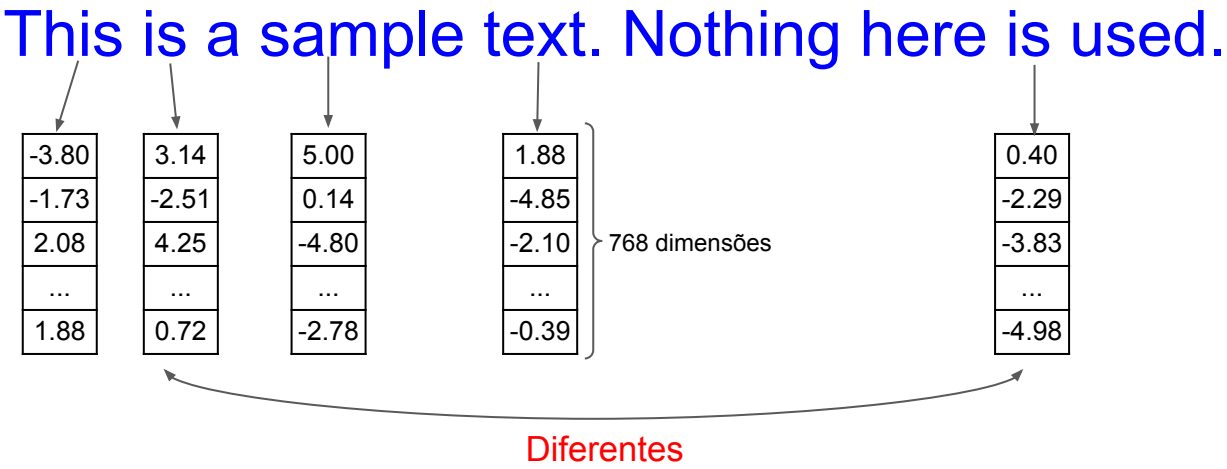
BERT - Representação Contextual



Cada palavra recebe um vetor de números reais, construído de forma a representar um contexto global

Representação de palavras

BERT - Representação Contextual



Cada palavra recebe um vetor de números reais, construído de forma a representar um contexto global

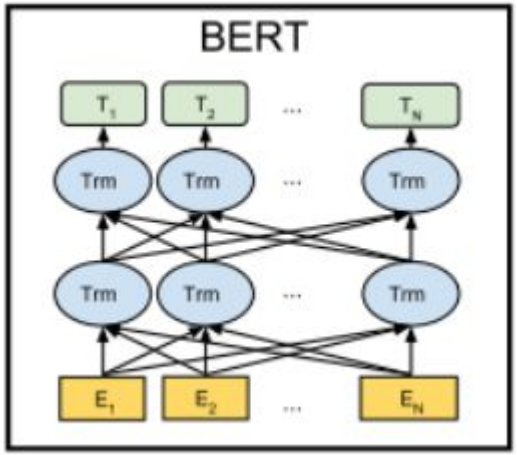
Transfer Learning
Cada ocorrência possui uma representação dependente do contexto
Aprendizado caro

Diferença Representação Contextual

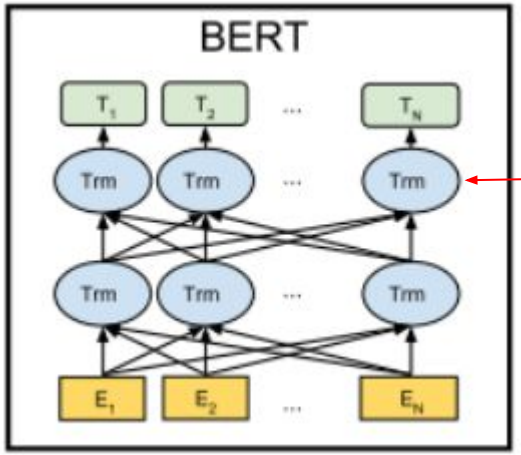
	Source	Nearest Neighbors
GloVe	play	playing, game, games, played, players, plays, player, Play, football, multiplayer
biLM	Chico Ruiz made a spectacular play on Alusik 's grounder {...}	Kieffer , the only junior in the group , was commended for his ability to hit in the clutch , as well as his all-round excellent play .
	Olivia De Havilland signed to do a Broadway play for Garson {...}	{...} they were actors who had been handed fat roles in a successful play , and had talent enough to fill the roles competently , with nice understatement .

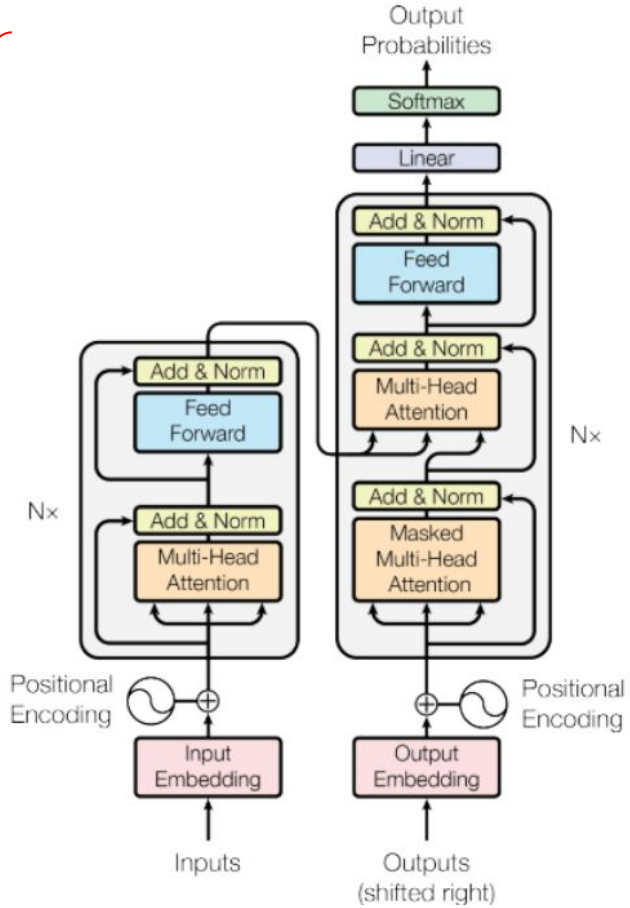
BERT - OverView

Bidirectional **E**ncoder **R**epresentations from **T**ransformers



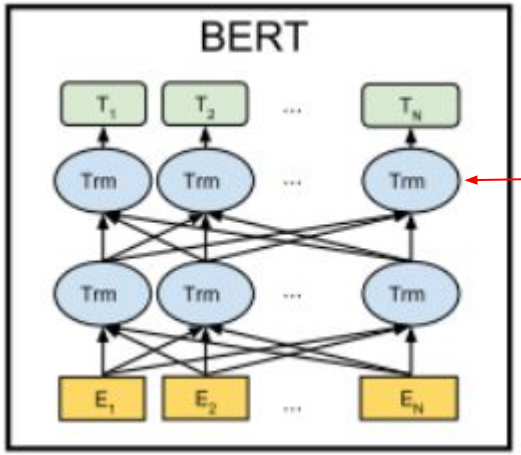
BERT - OverView

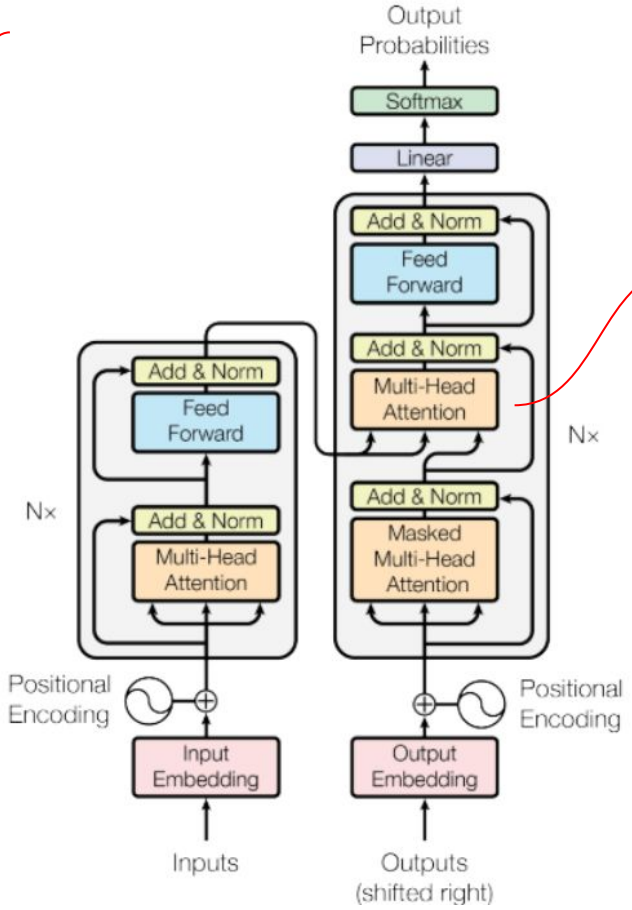


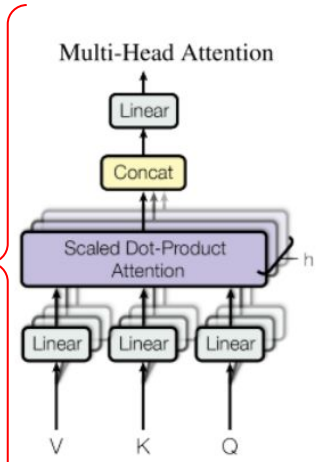


Transformer

BERT - OverView

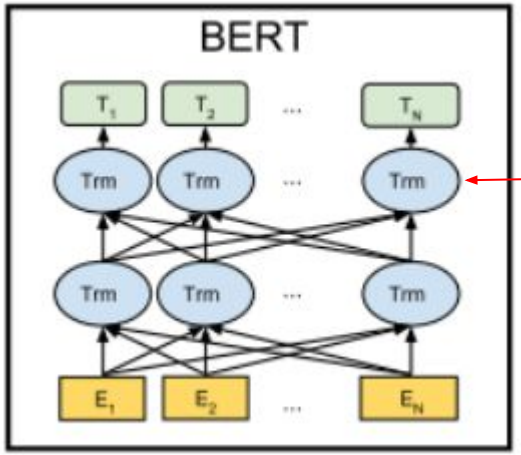


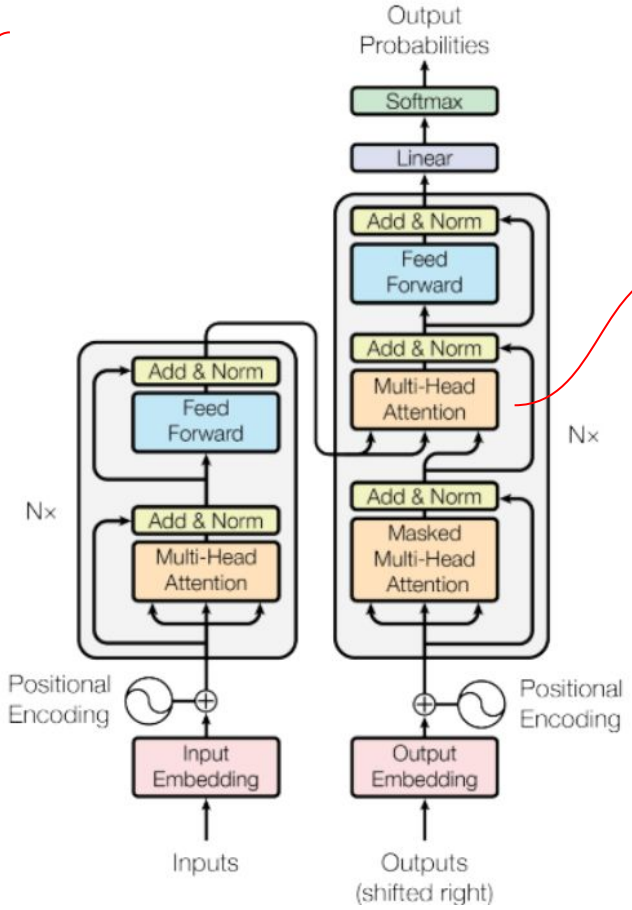


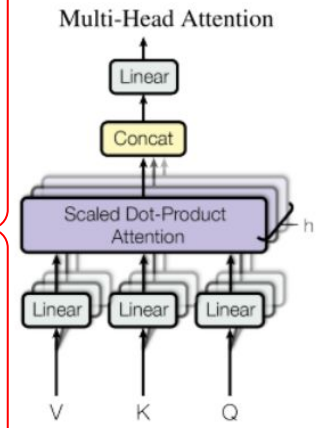


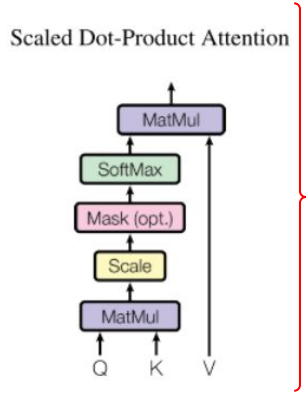
Transformer

BERT - OverView









Transformer

BERT - OverView

- Tarefas auxiliares (Self-supervised learning)
  - Criar tarefas com resposta conhecidas a partir de bases grandes
  - Treinar o modelo como se fosse supervisionado
  - Utilizar parte encoder do modelo treinado em outras tarefas

BERT - OverView

- Bert Original:
  - Next Sentence
    - A segunda frase vem depois da primeira ?
  - Masked Language Model
    - Deduzir valor de um token mascarado

BERT - OverView

- Bert Original:
  - Next Sentence
    - A segunda frase vem depois da primeira ?
  - Masked Language Model
    - Deduzir valor de um token mascarado

- SpanBert:
  - Span Boundary Objective
    - Deduzir todos os tokens de um trecho a partir das extremidades
  - Masked Language Model
    - Deduzir valor de vários tokens mascados em sequência

SpanBERT - Vantagens

- Criado para resolver correferência
- Tarefas originais eram “fáceis”
- Atinge estado da arte em diversas tarefas

	CoLA	SST-2	MRPC	STS-B	QQP	MNLI	QNLI	RTE	(Avg)
Google BERT	59.3	**95.2**	88.5/84.3	86.4/88.0	71.2/89.0	86.1/85.7	93.0	71.1	80.4
Our BERT	58.6	93.9	90.1/86.6	88.4/89.1	71.8/89.3	87.2/86.6	93.0	74.7	81.1
Our BERT-1seq	63.5	94.8	**91.2**/87.8	89.0/88.4	**72.1/89.5**	88.0/87.4	93.0	72.1	81.7
SpanBERT	**64.3**	94.8	90.9/**87.9**	**89.9/89.1**	71.9/**89.5**	**88.1/87.7**	**94.3**	**79.0**	**82.8**

Árvores Latentes

Correferência - Overview

North Korea opened its doors to the U.S. today, welcoming Secretary of State Madeleine Albright. She says her visit is a good start. The U.S. remains concerned about North Korea missile development program and its exports of missiles to Iran.

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

Correferência - Menções

North Korea opened its doors to the U.S. today, welcoming Secretary of State Madeleine Albright. She says her visit is a good start. The U.S. remains concerned about North Korea missile development program and its exports of missiles to Iran.

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

Correferência - Clusters

North Korea opened its doors to the U.S. today, welcoming Secretary of State Madeleine Albright. She says her visit is a good start. The U.S. remains concerned about North Korea missile development program and its exports of missiles to Iran.

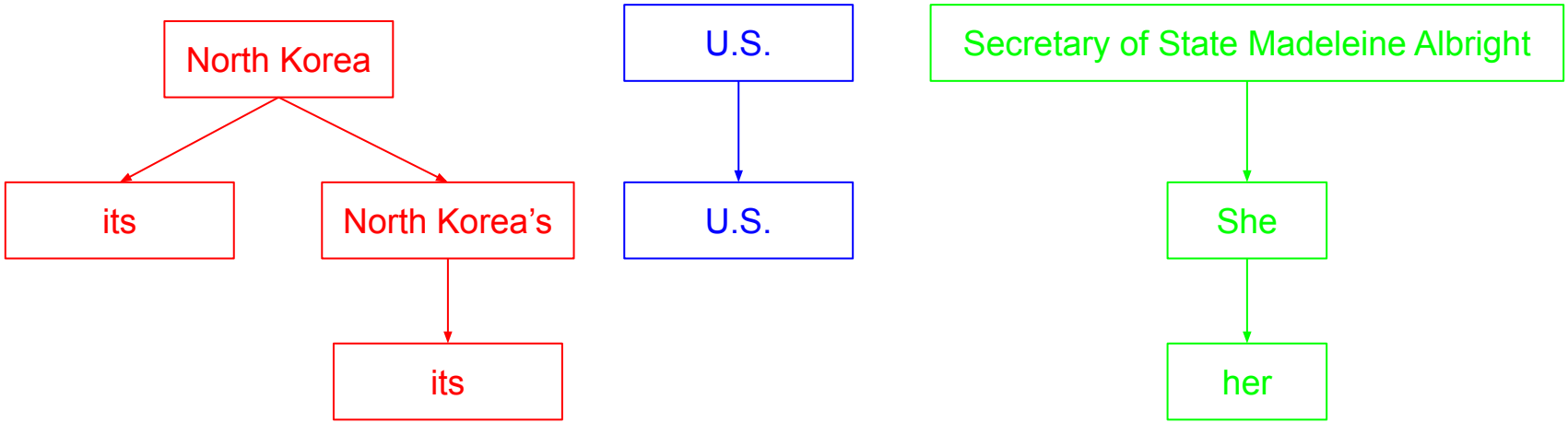
FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

Correferência - Representação

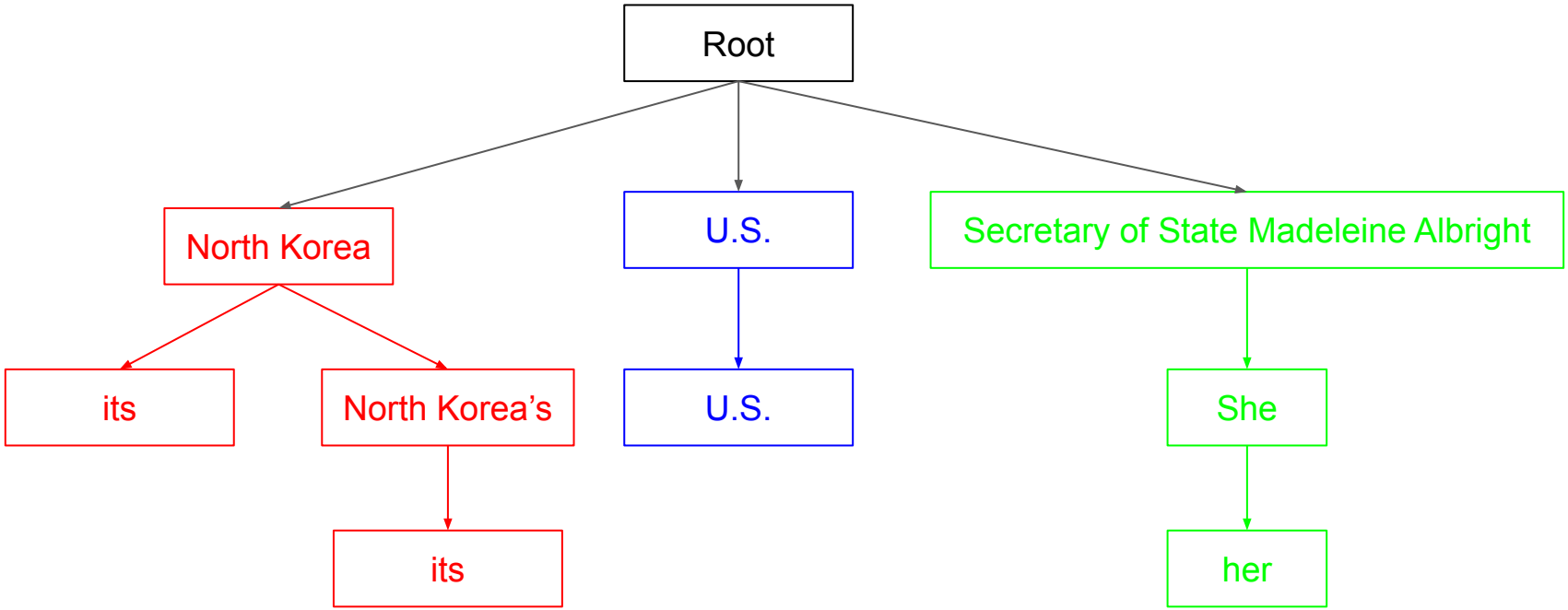
North Korea its U.S.
Secretary of State Madeleine Albright.
She her U.S.
North Korea
its Iran.

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

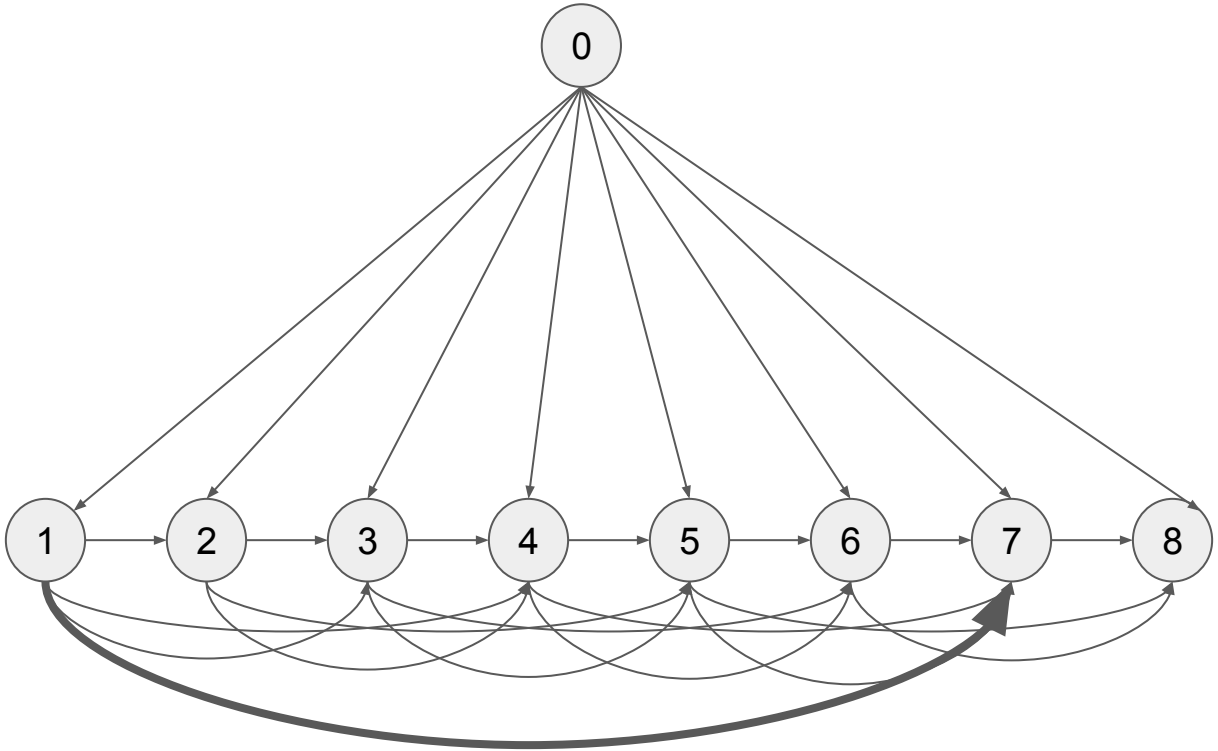
Árvores Latentes



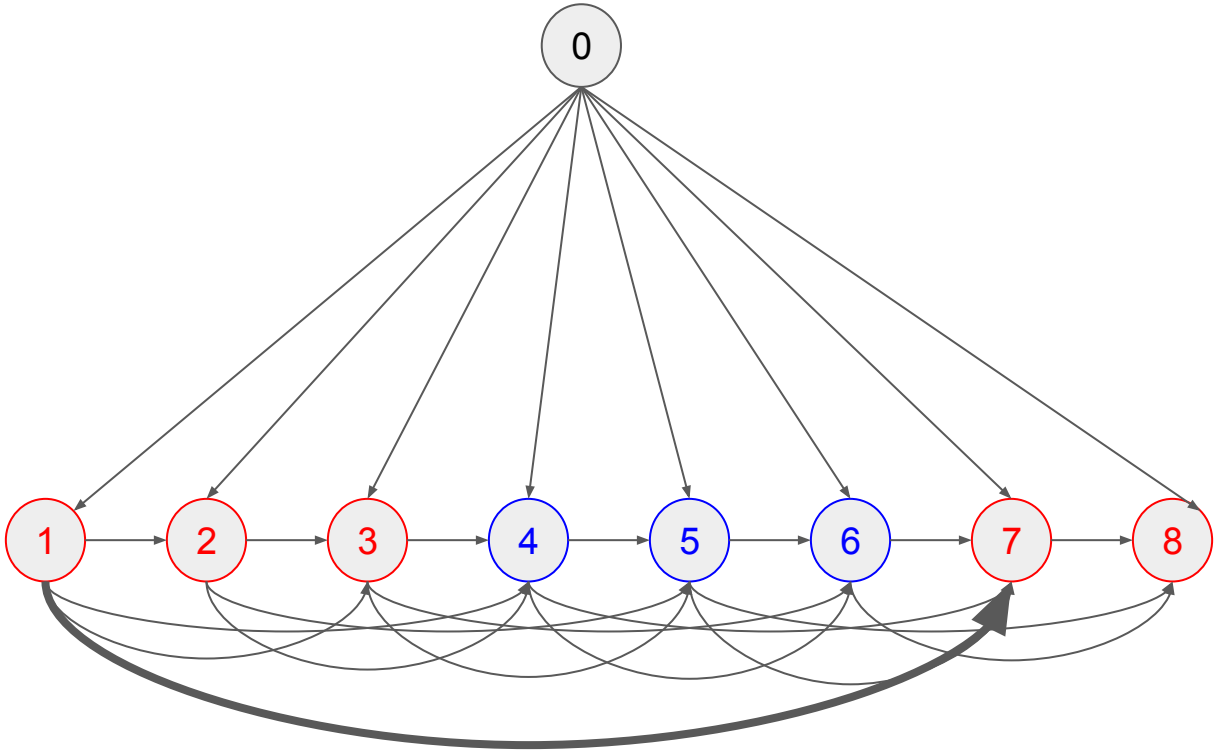
Árvores Latentes - Documento



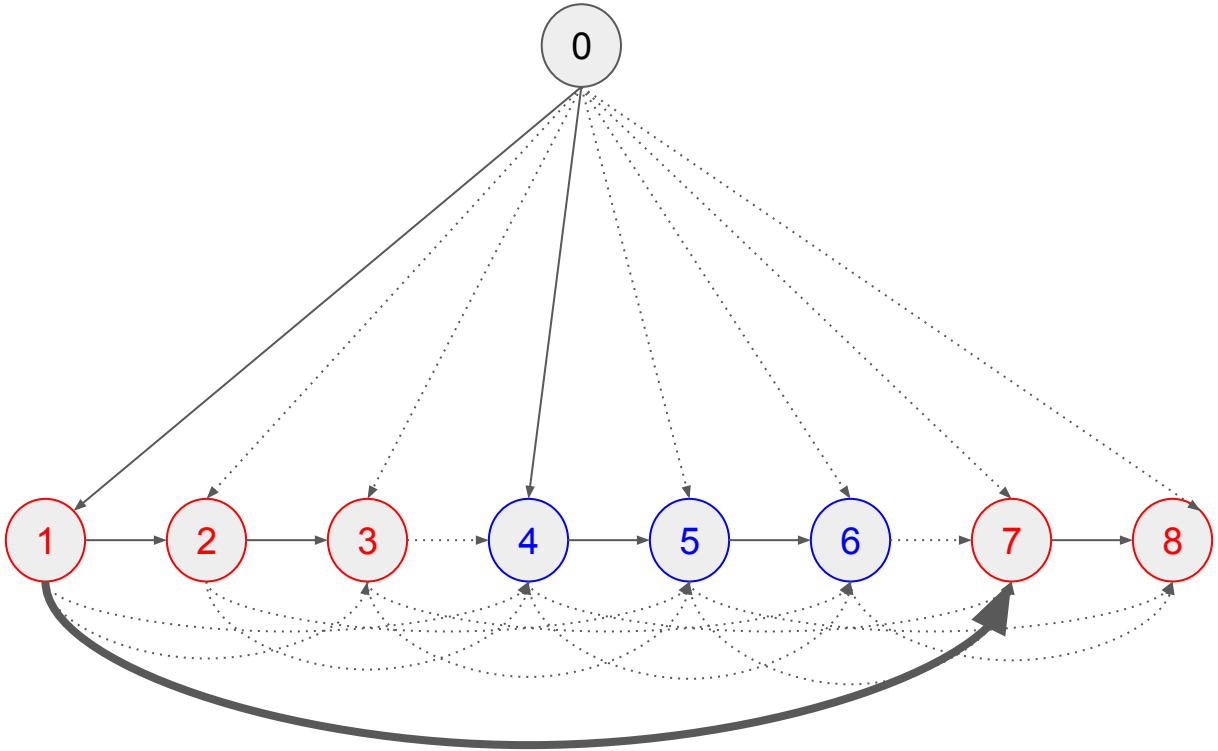
Latent Trees - Candidate Arcs



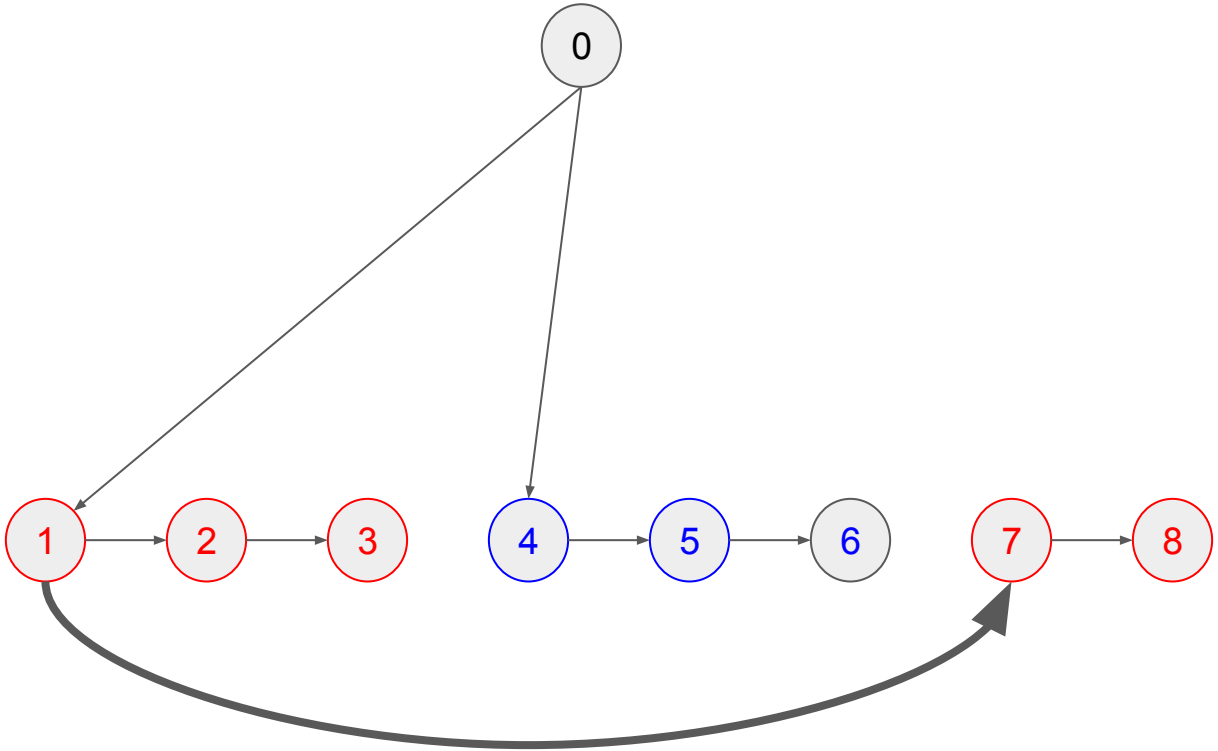
Latent Trees - Mention Clusters



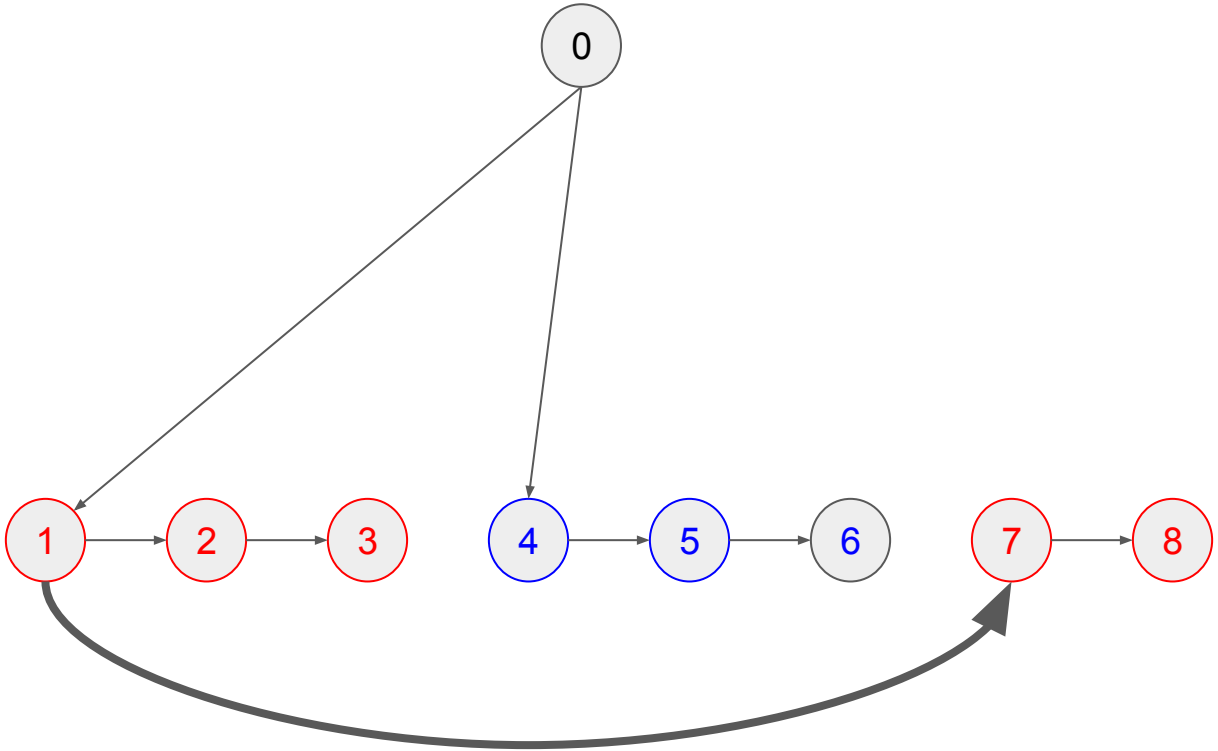
Latent Trees - Intercluster arcs



Latent Trees - Intercluster arcs



Latent Trees - Silver Tree



Features

Structural

- Speaker gender
- Mention Number (Singular, Plural)
- Semantic class (Person, Object, Numeric)
- Named Entity tag of mention head
- Number of tokens of a mention
- Tokens of both mentions match exactly
- Tokens of both mention heads match exactly
- Both mentions have the same speaker
- ...

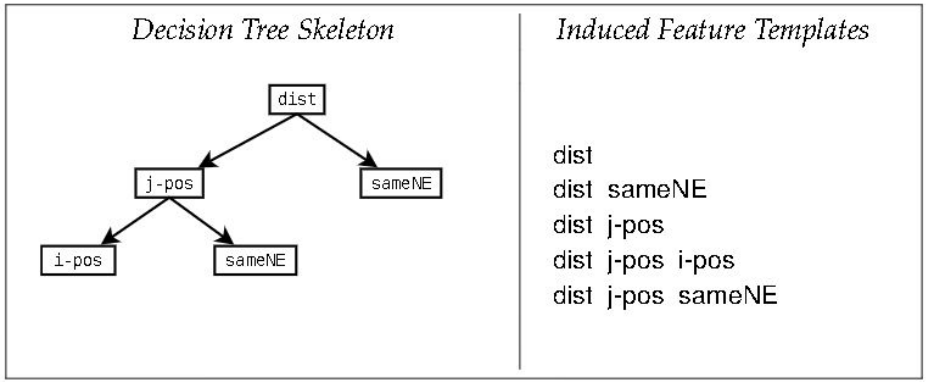
Lexical

- Tokens of the mention head
- First token of the mention
- Last token of the mention
- Token preceding the mention
- Token after the mention
- Token for governor of mention

MARTSCHAT, S.; STRUBE, M..Latent Structures for CoreferenceResolution. Transactions of the Association for Computational Linguis-tics, 2018
DOS SANTOS, C.; CARVALHO, D..Rule and tree ensembles forunrestricted coreference resolution.

Entropy Feature Induction

- Uses a pruned decision tree to create templates
- No duplicated templates
- Uses only templates that occurs on training
- Each arc has few “on” positions
- Each basic feature is considered categorical
- Words are embedded as a dictionary list



$$\phi_m(\boldsymbol{x}, e) = \begin{cases} 1 & \text{if } \textit{dist}{=}2 \text{ and } \textit{j-pos}{=}\textit{Noun} \text{ and } \textit{sameNE}{=}N, \\ 0 & \text{otherwise.} \end{cases}$$

FERNANDES, E. R.; DOS SANTOS, C. N. ; MILIDIÚ, R. L..Latent trees for coreference resolution. Computational Linguistics, 40(4):801–835, 12 2014. 1, 2.1, 2.2

SURFACE

	Antecedente	Anáfora
head	Vicente	him
POS	NOUN	PRON

Vicente
him
NOUN
PRON

SURFACE

	Antecedente	Anáfora	Ante + Ana
head	Vicente	him	Vicente + him
POS	NOUN	PRON	NOUN + PRON

Vicente
him
NOUN
PRON
Vicente + him
NOUN + PRON

SURFACE

	Antecedente	Anáfora	Ante + Ana
head	Vicente	him	Vicente + him
POS	NOUN	PRON	NOUN + PRON

	Antecedente	Anáfora
Fine Type	Vicente	he

Vicente
him
NOUN
PRON
Vicente + him
NOUN + PRON
Vicente + Vicente
Vicente + him
Vicente + NOUN
Vicente + PRON
Vicente + Vicente + him
Vicente + NOUN + PRON
he + Vicente
he + him
he + NOUN
he + PRON
he + Vicente + him
he + NOUN + PRON

O trabalho

Features Automáticas

- Do ponto de vista das árvores latentes:
  - Features ⇔ Vetores

- Substituir features manuais por representação contextual
  - Utilizar a representação de cada menção como as features básicas
  - Manter forma de aumentar o número de features

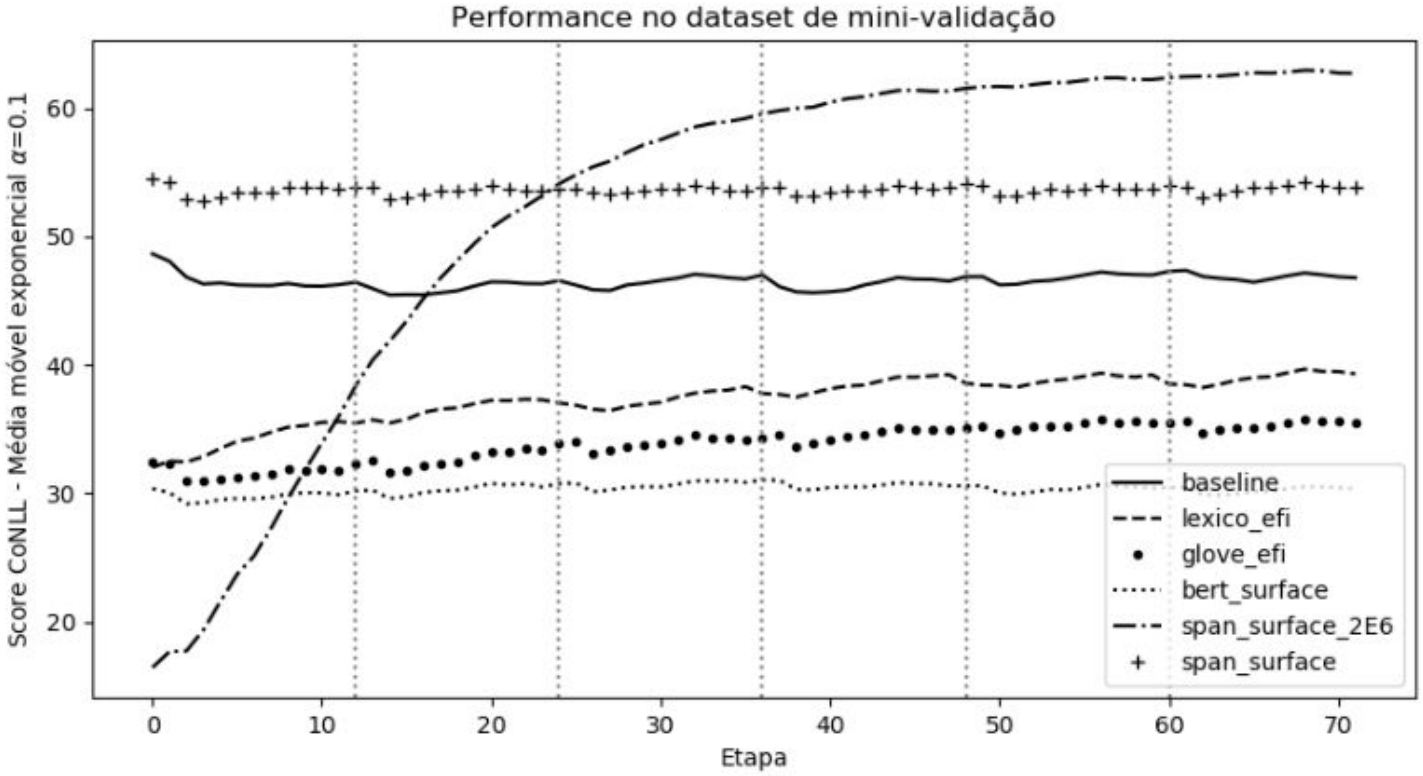
Resultados

Resultados

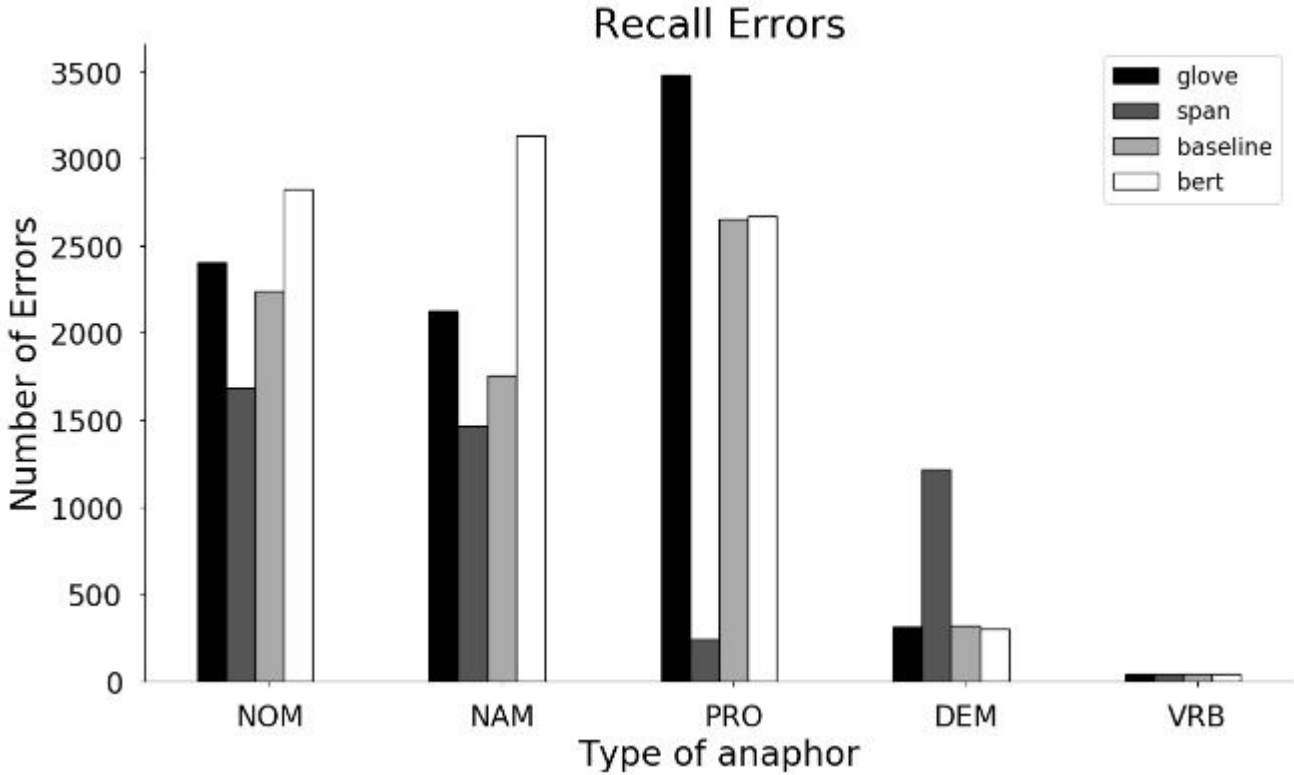
Modelo	MUC	B-CUBED	CEAF-E	Score CoNLL
span_surface_2E6	61.72	52.04	45.89	53.2
span_surface	57.85	48.13	42.05	49.34
baseline	56.08	45.41	39.88	47.12
lexico_efi	48.89	37.01	31.07	38.99
glove_efi	44.23	33.95	28.78	35.65
bert_surface	24.99	30.88	25.51	27.13

Resultados no dataset de validação

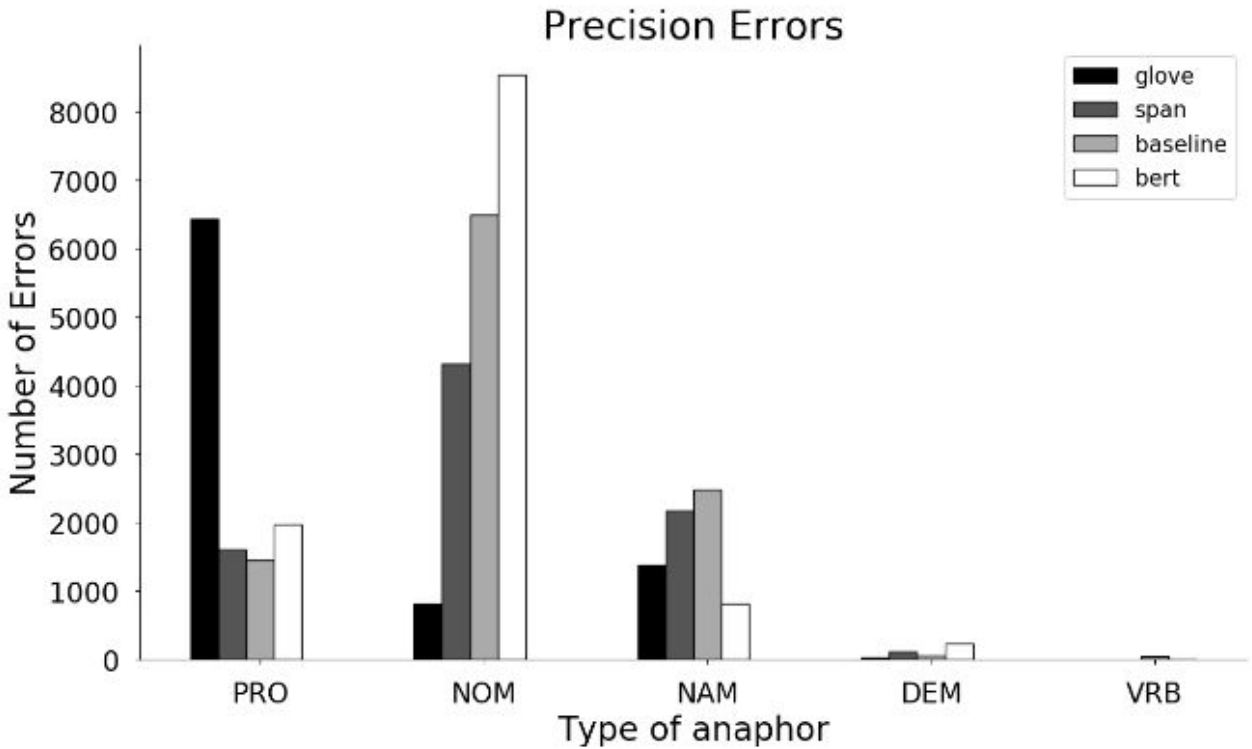
Resultados


Performance no dataset de mini-validação
Score CoNLL - Média móvel exponencial α=0.1
Etapa
0
10
20
30
40
50
60
70
20
30
40
50
60
baseline
lexico_efi
glove_efi
bert_surface
span_surface_2E6
span_surface

Análise de Erros



Análise de Erros



Conclusão

Conclusão

- Representações contextuais são capazes de criar features automáticas para a tarefa de correferência

- É possível utilizar modelos já treinados em uma arquitetura simples para atingir boa performance na tarefa de correferência

Trabalhos Futuros

- Otimizar os hiperparâmetros dos modelos apresentados
  - Exemplo: SpanBERT margem muito larga

- Automatizar escolha de como as features das menções são criadas
  - Incorporar a seleção das features em algum tipo de rede

Obrigado