

Processamento lexical, morfologia e morfossintaxe

SCC5908 Introdução ao
Processamento de Língua Natural

Thiago A. S. Pardo

Tagging

▶ 2 principais abordagens

◦ Regras

- Por exemplo, uma palavra antecedida por um artigo é um substantivo

◦ Probabilidades

- Classe mais provável de uma palavra em função das palavras vizinhas, com aprendizado a partir de cópús

▶ Hibridismo também é possível

- Por exemplo, aprendizado de regras a partir de cópús

Tagging: probabilidades

- ▶ Abordagem antiga, desde a década de 60
- ▶ **Modelo de Markov Oculto** (HMM – *Hidden Markov Model*), um dos mais utilizados
 - Um tipo de inferência bayesiana
 - Tarefa de classificação: dadas algumas observações, quais as classes mais prováveis
 - Tagging: dada uma **sequência de palavras**, qual a **sequência de tags** mais provável

3

Tagging: probabilidades

- ▶ Tagging: dada uma sequência de palavras, qual a sequência de tags mais provável
- ▶ Queremos a sequência de tags **SeqTags** que **maximize** a probabilidade $P(\text{SeqTags}|\text{SeqPalavras})$

$$\hat{\text{SeqTags}} = \underset{\text{SeqTags}}{\text{argmax}} P(\text{SeqTags} | \text{SeqPalavras})$$

- ▶ Exemplo: *O menino prefere brincar do que estudar*
 - ▶ SeqTags 1: art subst verbo verbo contração pro verbo
 - ▶ SeqTags 2: art subst verbo verbo contração adv verbo
 - ▶ SeqTags 3: art subst verbo verbo contração conj verbo
 - ▶ SeqTags 4: pro subst verbo verbo contração pro verbo
 - ▶ SeqTags 5: pro subst verbo verbo contração adv verbo
 - ▶ SeqTags 6: pro subst verbo verbo contração conj verbo
- ▶ Qual a melhor sequência de tags, ou seja, qual destas sequências maximiza a probabilidade $P(\text{SeqTags}|\text{SeqPalavras})$?

4

Tagging: probabilidades

- ▶ Qual a maior probabilidade?

$$\hat{\text{SeqTags}} = \underset{\text{SeqTags}}{\text{argmax}} P(\text{SeqTags} | \text{SeqPalavras})$$

- ▶ P(art subst verbo verbo contração pro verbo | O menino prefere brincar do que estudar)
- ▶ P(art subst verbo verbo contração adv verbo | O menino prefere brincar do que estudar)
- ▶ P(art subst verbo verbo contração conj verbo | O menino prefere brincar do que estudar)
- ▶ P(pro subst verbo verbo contração pro verbo | O menino prefere brincar do que estudar)
- ▶ P(pro subst verbo verbo contração adv verbo | O menino prefere brincar do que estudar)
- ▶ P(pro subst verbo verbo contração conj verbo | O menino prefere brincar do que estudar)

5

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

6

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})}$$

7

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})} \longrightarrow \text{constante}$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})$$

8

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})} \longrightarrow \text{constante}$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})$$

- ▶ **Simplificações** para facilitar o cálculo
 - ▶ Cada palavra depende apenas de sua tag
 - ▶ Uma tag depende apenas da tag anterior na sentença

9

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})} \longrightarrow \text{constante}$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})}$$

- ▶ **Simplificações** para facilitar o cálculo
 - ▶ Cada palavra depende apenas de sua tag
 - ▶ Uma tag depende apenas da tag anterior na sentença

$$P(\text{SeqTags} | \text{SeqPalavras}) = \prod_{i=1}^{\text{número de palavras}} P(\text{palavra}_i | \text{tag}_i) \times P(\text{tag}_i | \text{tag}_{i-1})$$

10

Tagging: probabilidades

- ▶ Como calcular essas probabilidades?

$$P(\text{SeqTags} | \text{SeqPalavras})$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = \frac{P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})}{P(\text{SeqPalavras})} \longrightarrow \text{constante}$$

$$P(\text{SeqTags} | \text{SeqPalavras}) = P(\text{SeqPalavras} | \text{SeqTags}) \times P(\text{SeqTags})$$

- ▶ Simplificações para facilitar o cálculo
 - ▶ Cada palavra depende apenas de sua tag
 - ▶ Uma tag depende apenas da tag anterior na sentença

$$P(\text{SeqTags} | \text{SeqPalavras}) = \prod_{i=1}^{\text{número de palavras}} P(\text{palavra}_i | \text{tag}_i) \times P(\text{tag}_i | \text{tag}_{i-1})$$

Como calcular essas 2 probabilidades?

11

Tagging: probabilidades

- ▶ Exemplo
 - ▶ Supondo que se usa o Brown Corpus

$$P(\text{SeqTags} | \text{SeqPalavras}) = \prod_{i=1}^{\text{número de palavras}} P(\text{palavra}_i | \text{tag}_i) \times P(\text{tag}_i | \text{tag}_{i-1})$$

$$P(\text{palavra}_i | \text{tag}_i) = P(\text{"is"} | \text{VBZ})$$

tag=VBZ → 21.627 ocorrências no cópuz
 palavra="is" com tag=VBZ → 10.073 ocorrências no cópuz

$P(\text{"is"} | \text{VBZ}) = \text{número de vezes de "is" com VBZ} / \text{número de VBZ}$
 $P(\text{"is"} | \text{VBZ}) = 10.073 / 21.627 = 0.47$ ou 47%

12

Tagging: probabilidades

▶ Exemplo

- ▶ Supondo que se usa o Brown Corpus

$$P(\text{SeqTags} | \text{SeqPalavras}) = \prod_{i=1}^{\text{número de palavras}} P(\text{palavra}_i | \text{tag}_i) \times P(\text{tag}_i | \text{tag}_{i-1})$$

$$P(\text{tag}_i | \text{tag}_{i-1}) = P(\text{NN} | \text{DT})$$

tag_{i-1}=DT → 116.454 ocorrências no cópús
tag_i=NN com tag_{i-1}=DT → 56.509 ocorrências no cópús

P(NN|DT) = número de vezes de NN precedido por DT / número de DT
P(NN|DT) = 56.509 / 116.454 = 0.49 ou 49%

13

Tagging: probabilidades

▶ Exemplo

- ▶ P(art subst verbo verbo contração pro verbo | O menino prefere brincar do que estudar)

$$P(\text{SeqTags} | \text{SeqPalavras}) = \prod_{i=1}^{\text{número de palavras}} P(\text{palavra}_i | \text{tag}_i) \times P(\text{tag}_i | \text{tag}_{i-1})$$

- ▶ = P(O|art) x P(art|<início da sentença>) x
P(menino|subst) x P(subst|art) x
P(prefere|verbo) x P(verbo|subst) x
P(brincar|verbo) x P(verbo|verbo) x
P(do|contração) x P(contração|verbo) x
P(que|pro) x P(pro|contração) x
P(estudar|verbo) x P(verbo|pro)

- ▶ Faz-se isso para **todas as possíveis seqüências de tags** (com probabilidades aprendidas de cópús)

- ▶ **A seqüência com maior probabilidade é escolhida**

14

Tagging: probabilidades

▶ Considerando a probabilidade abaixo

- ▶ $P(\text{art subst verbo verbo contração pro verbo} \mid \text{O menino prefere brincar do que estudar}) =$
 $P(\text{O}|\text{art}) \times P(\text{art}|\langle \text{início da sentença} \rangle) \times$
 $P(\text{menino}|\text{subst}) \times P(\text{subst}|\text{art}) \times$
 $P(\text{prefere}|\text{verbo}) \times P(\text{verbo}|\text{subst}) \times$
 $P(\text{brincar}|\text{verbo}) \times P(\text{verbo}|\text{verbo}) \times$
 $P(\text{do}|\text{contração}) \times P(\text{contração}|\text{verbo}) \times$
 $P(\text{que}|\text{pro}) \times P(\text{pro}|\text{contração}) \times$
 $P(\text{estudar}|\text{verbo}) \times P(\text{verbo}|\text{pro})$

▶ Por que ela é **provavelmente menor** do que a abaixo – que é a interpretação correta?

- ▶ $P(\text{art subst verbo verbo contração conj verbo} \mid \text{O menino prefere brincar do que estudar}) =$
 $P(\text{O}|\text{art}) \times P(\text{art}|\langle \text{início da sentença} \rangle) \times$
 $P(\text{menino}|\text{subst}) \times P(\text{subst}|\text{art}) \times$
 $P(\text{prefere}|\text{verbo}) \times P(\text{verbo}|\text{subst}) \times$
 $P(\text{brincar}|\text{verbo}) \times P(\text{verbo}|\text{verbo}) \times$
 $P(\text{do}|\text{contração}) \times P(\text{contração}|\text{verbo}) \times$
 $P(\text{que}|\text{conj}) \times P(\text{conj}|\text{contração}) \times$
 $P(\text{estudar}|\text{verbo}) \times P(\text{verbo}|\text{conj})$

15

Tagging: probabilidades

▶ Considerando a probabilidade abaixo

- ▶ $P(\text{art subst verbo verbo contração pro verbo} \mid \text{O menino prefere brincar do que estudar}) =$
 $P(\text{O}|\text{art}) \times P(\text{art}|\langle \text{início da sentença} \rangle) \times$
 $P(\text{menino}|\text{subst}) \times P(\text{subst}|\text{art}) \times$
 $P(\text{prefere}|\text{verbo}) \times P(\text{verbo}|\text{subst}) \times$
 $P(\text{brincar}|\text{verbo}) \times P(\text{verbo}|\text{verbo}) \times$
 $P(\text{do}|\text{contração}) \times P(\text{contração}|\text{verbo}) \times$
 $P(\text{que}|\text{pro}) \times P(\text{pro}|\text{contração}) \times$
 $P(\text{estudar}|\text{verbo}) \times P(\text{verbo}|\text{pro})$

▶ Por que ela é **provavelmente menor** do que a abaixo – que é a interpretação correta?

- ▶ $P(\text{art subst verbo verbo contração conj verbo} \mid \text{O menino prefere brincar do que estudar}) =$
 $P(\text{O}|\text{art}) \times P(\text{art}|\langle \text{início da sentença} \rangle) \times$
 $P(\text{menino}|\text{subst}) \times P(\text{subst}|\text{art}) \times$
 $P(\text{prefere}|\text{verbo}) \times P(\text{verbo}|\text{subst}) \times$
 $P(\text{brincar}|\text{verbo}) \times P(\text{verbo}|\text{verbo}) \times$
 $P(\text{do}|\text{contração}) \times P(\text{contração}|\text{verbo}) \times$
 $P(\text{que}|\text{conj}) \times P(\text{conj}|\text{contração}) \times$
 $P(\text{estudar}|\text{verbo}) \times P(\text{verbo}|\text{conj})$

Esses termos devem ser
mais prováveis do que os
correspondentes na
interpretação errada

16

Tagging: probabilidades

▶ Modelo de Markov oculto

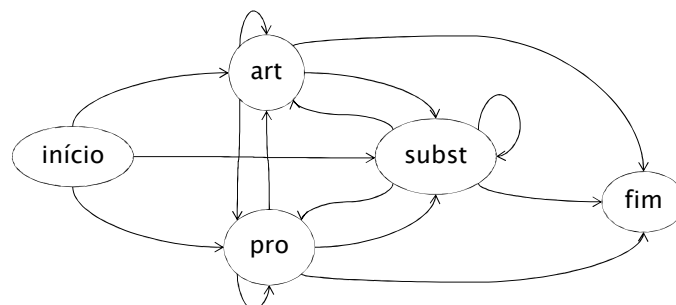
- Modelam-se eventos observados (palavras) e eventos não observados, ou seja, ocultos (tags)
- Como dito anteriormente, tipo especial de autômato
 - Probabilidades nos arcos (transições): $P(\text{tag}_i | \text{tag}_{i-1})$
 - Probabilidades nos nós: $P(\text{palavra}_i | \text{tag}_i)$

17

Tagging: probabilidades

▶ Modelo de Markov oculto

- Exemplo hipotético



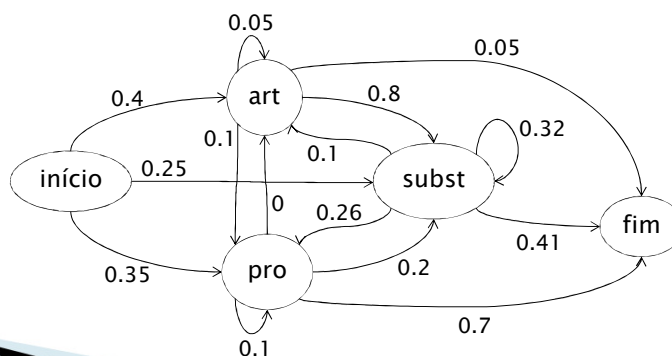
18

Tagging: probabilidades

► Modelo de Markov oculto

◦ Exemplo hipotético

A soma das probabilidades dos arcos que saem de cada nó devem somar 1



19

Tagging: probabilidades

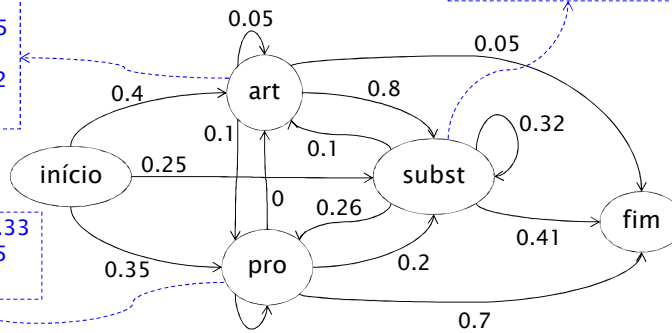
► Modelo de Markov oculto

◦ Exemplo hipotético

$P(o|art)=0.2$
 $P(os|art)=0.25$
 $P(a|art)=0.2$
 $P(as|art)=0.12$
 ...

$P(menino|subst)=0.1$
 $P(casa|subst)=0.04$
 ...

$P(que|pro)=0.33$
 $P(ele|pro)=0.5$
 ...



A soma das probabilidades associadas a cada nó devem somar 1

Tagging: probabilidades

- ▶ **Taggers atuais** usam normalmente **2 tags anteriores** como contexto
 - $P(\text{tag}_i | \text{tag}_{i-1}, \text{tag}_{i-2})$
- ▶ É preciso **otimizar a busca por sequências de tags** mais prováveis, senão pode haver explosão combinatória
 - **Programação dinâmica** (estudaremos logo)
- ▶ É preciso **lidar com probabilidades muito baixas ou próximas de zero**
 - Por exemplo, situações em que $P(\text{tag}_i | \text{tag}_{i-1}, \text{tag}_{i-2}) \approx 0$
 - Solução usual: combinar probabilidades ponderadas
 - $P(\text{tag}_i | \text{tag}_{i-1}, \text{tag}_{i-2}) = p_1 * P(\text{tag}_i | \text{tag}_{i-1}, \text{tag}_{i-2}) + p_2 * P(\text{tag}_i | \text{tag}_{i-1}) + p_3 * P(\text{tag}_i)$, com $p_1 + p_2 + p_3 = 1$

21

Tagging: Transformation-Based

- ▶ *Transformation-Based tagging*
 - Aplicação de TBL (*Transformation-Based Learning*), da linha de aprendizado de máquina
 - **Regras são aprendidas e aprimoradas** automaticamente
 - Várias iterações
 - A cada iteração, o processo melhora
 - Ao estabilizar, fim do processo de aprendizado

22

Tagging: Transformation-Based

▶ Transformation-Based tagging

- **Processo básico** com base em um **cópus anotado** manualmente
 - Inicialmente, **anota-se um cópus automaticamente**, assumindo-se que a tag de uma palavra é a sua **tag mais frequente** (segundo um cópus/léxico)
 - Verificam-se os **erros cometidos** (comparando-se com a anotação humana correspondente) e, dentre todas as possibilidades de correção, **monta-se uma regra de correção** com maior precisão
 - Aplica-se essa **regra nova em todo o cópus**
 - Verificam-se novamente os **erros cometidos** e monta-se uma segunda **regra de correção** com maior precisão
 - E assim por diante, **até não se obter mais melhora de performance**

23

Tagging: Transformation-Based

▶ Exemplo

1. Inicialmente, anotação com base em frequência

... is/VBZ expected/VBN to/TO **race/NN** tomorrow/NN
 ... the/DT race/NN for/IN outer/JJ space/NN
 Book/VB the/DT **flight/VB** to/TO...

2. Verificando-se os erros, aprende-se uma nova regra

Troque NN para VB quando a tag anterior é TO

3. Corrige-se a etiquetagem anterior

... is/VBZ expected/VBN to/TO **race/VB** tomorrow/NN
 ... the/DT race/NN for/IN outer/JJ space/NN
 Book/VB the/DT **flight/VB** to/TO...

24

Tagging: Transformation-Based

▶ Exemplo

4. Aprende-se uma nova regra com base nos erros existentes

Troque VB para NN quando a tag anterior é DT e a posterior é TO

5. Corrige-se a etiquetação anterior

... is/VBZ expected/VBN to/TO **race/VB** tomorrow/NN
 ... the/DT race/NN for/IN outer/JJ space/NN
 Book/VB the/DT **flight/NN** to/TO...

6. E assim por diante

25

Tagging: Transformation-Based

▶ Exemplo

- **Resultado** do processo

Regra 1: etiqüete as palavras com suas tags mais frequentes

Regra 2: troque NN para VB quando a tag anterior é TO

Regra 3: troque VB para NN quando a tag anterior é DT e a posterior é TO

...

26

Tagging: Transformation-Based

▶ Transformation-Based tagging

- Ao final do processo, há um **conjunto de regras ordenadas** que devem ser aplicadas sequencialmente para etiquetar um novo texto
- Como há muitas regras possíveis de serem aprendidas, costuma-se **limitar as estruturas aceitas de regras**
 - Troque a tag da palavra corrente de A para B se a palavra anterior tem a tag X
 - Troque a tag da palavra corrente de A para B se a palavra anterior tem a tag X e a posterior tem a tag Y
 - Etc.

Caso contrário, o que acontece?

27

Humanos e máquinas

▶ Palavras/morfologia e léxico mental

- Nem a listagem exaustiva, nem todas as regras de flexão/derivação
- Há indícios de que humanos armazenam em seu léxico mental os lemas das palavras e também algumas formas plenas
 - Stanners et al. (1979): são mantidas separadamente as palavras *happy* e *happiness*, mas somente o verbo *pour*, sem suas flexões

▶ Morfossintaxe

- Experimentos mostram que humanos discordam em 3–4% das tags
 - Melhores taggers com 97%, normalmente com problemas justamente onde os humanos discordam
- Voutilainen (1995) mostra que humanos atingem 100% se se permite que discutam as tags com problemas

28