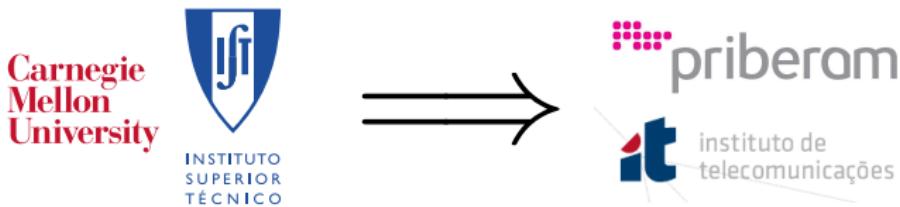


Turbo Parsers: Analisadores Sintácticos Estatísticos Baseados em Relaxações Lineares

André Martins



IST, Lisboa, 2 de Outubro de 2012

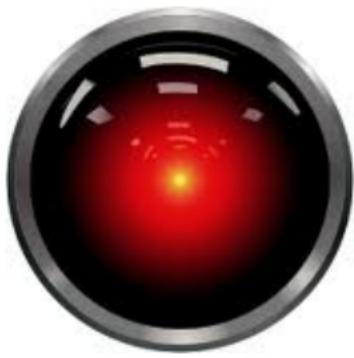
Colaboradores



■ Mário Figueiredo, Noah Smith, Pedro Aguiar e Eric Xing.

Processamento de Linguagem Natural

Objectivo: tornar as máquinas capazes de “entender” a linguagem dos humanos.



(HAL 9000)

Processamento de Linguagem Natural

Objectivo: tornar as máquinas capazes de “entender” a linguagem dos humanos.

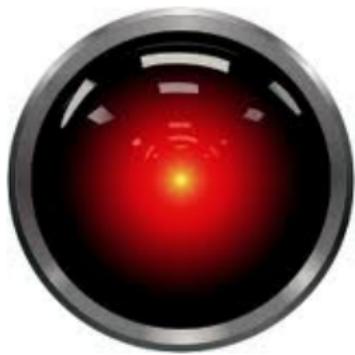


(HAL 9000)

- Pesquisa de Informação
- Tradução Automática
- Análise de Opiniões
- Reconhecimento de Entidades
- Análise Sintáctica
- Análise Semântica
- Reconhecimento de Fala
- Sumarização
- ...

Processamento de Linguagem Natural

Objectivo: tornar as máquinas capazes de “entender” a linguagem dos humanos.



(HAL 9000)

- Pesquisa de Informação
- Tradução Automática
- Análise de Opiniões
- Reconhecimento de Entidades
- **Análise Sintáctica**
- Análise Semântica
- Reconhecimento de Fala
- Sumarização
- ...

Exemplo: Sistemas de Pesquisa

The screenshot shows a Google search results page. The search bar at the top contains the query "processamento de linguagem natural". Below the search bar, there is a "Search" section showing "About 637,000 results (0.28 seconds)". A tip is displayed: "Tip: Search for English results only. You can specify your search language in Preferences". The search results list includes a link to the Wikipedia article on Natural Language Processing (pt.wikipedia.org...), followed by a snippet of text from the page: "Na teoria, processamento de linguagem natural é um método atrativo para interação homem-máquina. Sistemas mais antigos como SHRDLU, trabalhando com ...". Below this, another link is shown: "Aplicações e Limitações - Principais aplicações de PLN - Ferramentas - Ver também". On the left side of the search results, there is a sidebar with categories: Web, Images, Maps, Videos, News, and More.

Exemplo: Sistemas de Pesquisa

Google processamento de linguagem natural

Search About 637,000 results (0.28 seconds)

Web Tip: Search for English results only. You can specify preferences

Images

Maps

Videos

News

More

Processamento de linguagem natural
pt.wikipedia.org/.../Processamento_de_ling...
Na teoria, processamento de linguagem natural é a interação homem-máquina. Sistemas mais antigos...

Aplicações e Limitações - Principais aplicações de...

Pergunta

Search demo

Relacionados | Extracção de entidades

quem é jorge sampaio? Pesquisar em rtpnotícias Língua Português

Entre datas (Limpar Datas) De: 2002 Até: 2012 Resultados: 2850 Mostrar Resultados

Respostas

"Cidadãos começam a ficar dominados pelo medo"
Istanbul, 01 jun (Lusa) -- O alto representante do secretário-geral das Nações Unidas para a Aliança das Civilizações, Jorge Sampaio, disse à Lusa que o "medo, desconfiança e ressentimentos" provocados pela crise são uma "mistura explosiva" que é preciso resolver.

SNS passa por "tempos difíceis", afirma Jorge Sampaio
(31-03-2012)
Porto, 31 mar (Lusa) - O Serviço Nacional de Saúde (SNS) "passa por tempos difíceis", disse na sexta-feira à noite, no Porto, o antigo Presidente da República Jorge Sampaio na sessão final do ciclo Grandes Debates do Regime promovido pela autarquia portuguesa.

priberam

The figure shows a comparison between a standard Google search and a specialized search interface from Priberam. The Google search results page is on the left, showing a list of links and a bar chart. The Priberam search interface is on the right, featuring a search bar with a red oval highlighting the query 'quem é jorge sampaio?', a date range selector, and a results summary. Below these are sections for related entities and historical data. A red arrow points from the 'Pergunta' label to the Priberam search bar. A green arrow points from the 'Respostas' label to the text snippets in the Priberam interface.

(Priberam Search)

Exemplo: Análise de Opiniões



Canon PowerShot SX40 HS 12.1 MP Digital Camera
\$307 online ★★★★☆ 507 reviews
September 2011 - Canon - Point & Shoot - 12.1 megapixel - Electronic Viewfinder - Compact Sensor

Reviews

Summary - Based on 507 reviews

1 2 3 4 stars 5 stars

What people are saying

- pictures
- features
- zoom/lens
- design
- video
- value
- size

"Picture clarity is great."
"Great images, great features, easy to use."
"This is a great low end camera for Canon DSLR users."
"Another point is the overall camera speed."
"Great video quality."
"Amazing product with excellent price!!!"
"The SX 40's very easy to use.small learning curve."

Opiniões

Aspectos

Awesome
★★★★★ By Moeman - Feb 14, 2012 - B&H
... it for photography and it works very well in all situations. The SX 40 is very easy to use.small learning curve. Have not found any problems yet all is great.... [Read full review](#)

Exemplo: Tradução Automática

The screenshot shows the Google Translate interface. At the top, it says "Andre Martins 0 + Share". Below that, the "Translate" button is visible along with language selection dropdowns: "From: Portuguese" and "To: English". The main area displays a sentence in Portuguese: "A bateria da câmera que comprei ontem durou pouquíssimo!" To its right is the English translation: "The camera battery lasted very little that I bought yesterday!". A tooltip above the English sentence says "Click to edit and see alternate translations". Below the sentences are various input and output options like microphone, keyboard, and document icons.

Exemplo: Tradução Automática

A bateria da câmera que comprei ontem durou pouquíssimo!

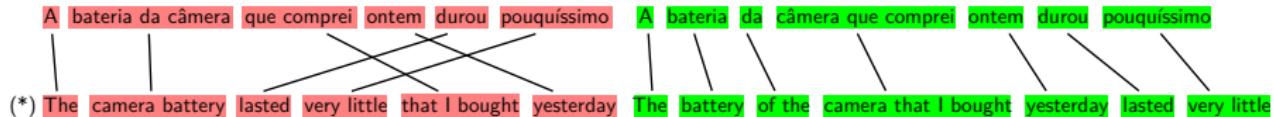
The camera battery lasted very little that I bought yesterday!

A bateria da câmera que comprei ontem durou pouquíssimo

(*) The camera battery lasted very little that I bought yesterday

Exemplo: Tradução Automática

The screenshot shows the Google Translate interface. The input text is "A bateria da câmera que comprei ontem durou pouquíssimo!" and the output is "The camera battery lasted very little that I bought yesterday!". The "To: English" dropdown is selected. A tooltip "Click to edit and see alternate translations" is visible above the English sentence. Below the sentences are various interface elements like microphones, keyboards, and sharing options.



Análise Sintáctica (“parsing”¹)

(Chomsky, 1965; Magerman, 1995; Charniak, 1996; Collins, 1999; Klein and Manning, 2003)

¹Do Latim *pars* = “parte”.

Análise Sintáctica (“parsing”¹)

(Chomsky, 1965; Magerman, 1995; Charniak, 1996; Collins, 1999; Klein and Manning, 2003)

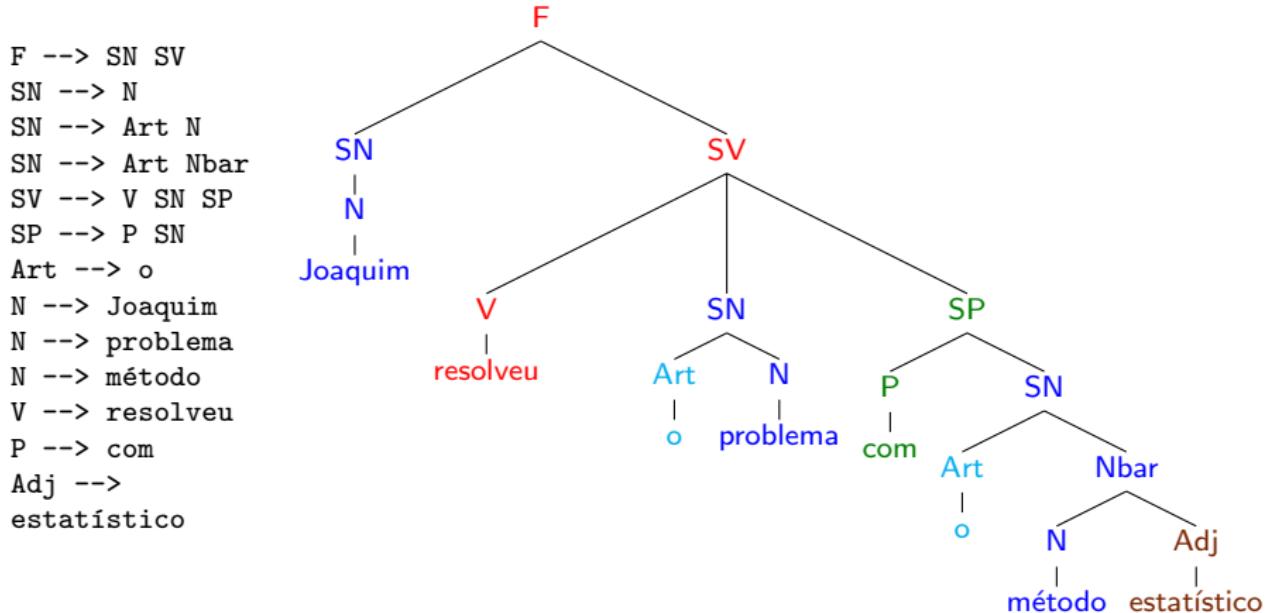
Joaquim resolveu o problema com o método estatístico.

¹Do Latim *pars* = “parte”.

Análise Sintáctica (“parsing”¹)

(Chomsky, 1965; Magerman, 1995; Charniak, 1996; Collins, 1999; Klein and Manning, 2003)

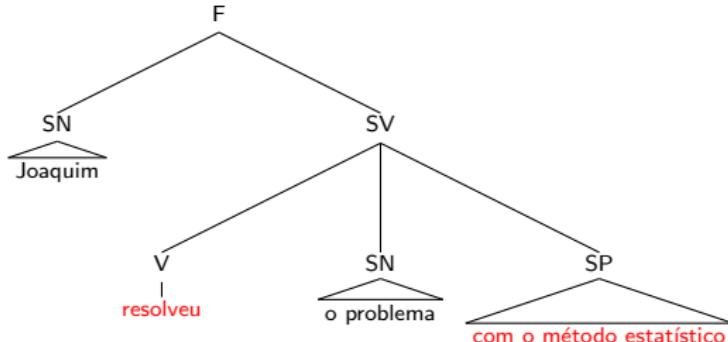
Joaquim resolveu o problema com o método estatístico.



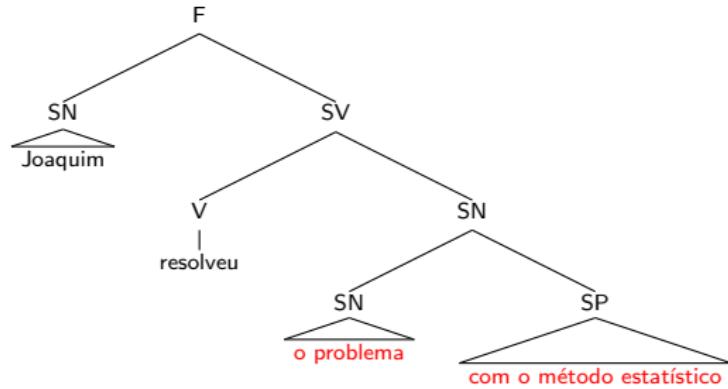
¹Do Latim *pars* = “parte”.

Ambiguidade Sintáctica

1 Joaquim *empregou* o método estatístico:

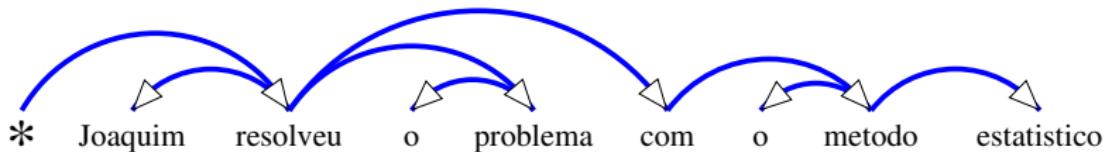


2 O método estatístico *tinha* um problema:



Sintaxe de Dependências

(Pānini, séc. IV a.C., Tesnière 1959; Hudson 1984; Mel'čuk 1988; Eisner 1996; McDonald et al. 2005; Nivre et al. 2006; Koo et al. 2007)



Árvore obtida “lexicalizando” a árvore sintáctica anterior.

- Um formalismo sintáctico mais simples, sem constituintes
- Funções gramaticais representadas como relações lexicais

Principais Contribuições

- 1 Novos modelos estatísticos *mais expressivos* para a análise sintáctica de dependências
- 2 Nova formulação: programa inteiro com um número *polinomial* de restrições, baseado em fluxos multi-mercadoria
- 3 **Turbo Parsers**:² relaxação linear equivale a inferência aproximada num modelo gráfico, em que se desprezam efeitos cíclicos
- 4 **AD**³: novo algoritmo que tira partido da *estrutura* do grafo
- 5 Validação experimental em 14 línguas diferentes

²Inspirado nos turbo-códigos (Berrou et al., 1993; McEliece et al., 1998)

Aprendizagem Estatística



“Unreasonable effectiveness of data”: um novo paradigma que coloca ênfase no processamento em grande escala dos dados (Halevy et al., 2009)

- Conjunto de observáveis \mathcal{X} e classes \mathcal{Y}
- **Objectivo:** aprender uma função de classificação $h : \mathcal{X} \rightarrow \mathcal{Y}$ a partir de exemplos $\{(x_i, y_i)\}_{i=1}^N \subseteq \mathcal{X} \times \mathcal{Y}$

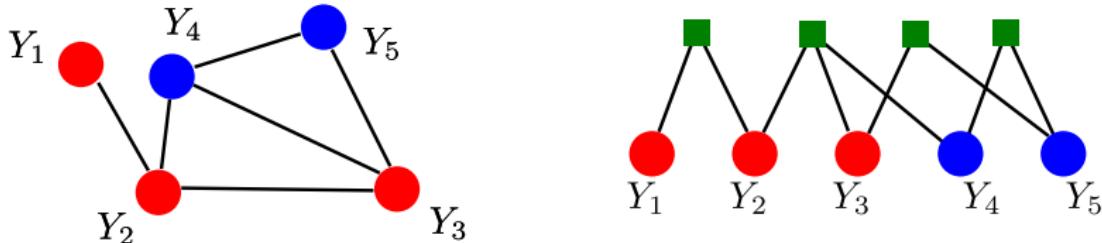
Aprendizagem Estatística



“Unreasonable effectiveness of data”: um novo paradigma que coloca ênfase no processamento em grande escala dos dados (Halevy et al., 2009)

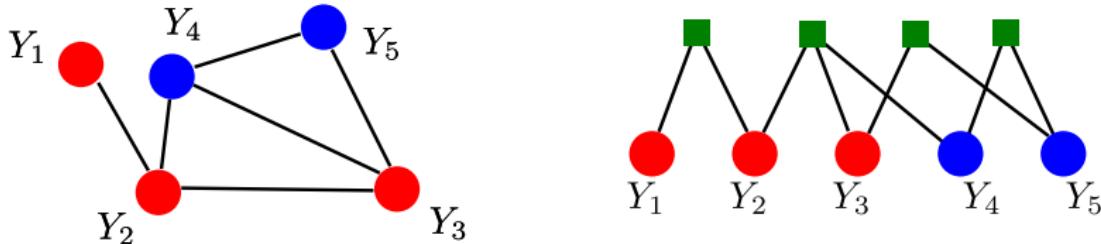
- Conjunto de observáveis \mathcal{X} e classes \mathcal{Y}
- **Objectivo:** aprender uma função de classificação $h : \mathcal{X} \rightarrow \mathcal{Y}$ a partir de exemplos $\{(x_i, y_i)\}_{i=1}^N \subseteq \mathcal{X} \times \mathcal{Y}$
- Corpora anotado: *Penn Treebank*, *Floresta Sintática*, ...
- **Classificação Estatística Estruturada:** variáveis possuem estrutura (Lafferty et al., 2001; Collins, 2002; Taskar et al., 2003)

Modelos Gráficos Probabilísticos



- Inspirados na Mecânica Estatística (Ising, 1925; Potts, 1952)
- Aplicações em teoria de códigos, visão, biologia computacional, ...
(Tanner, 1981; Pearl, 1988; Kschischang et al., 2001; Koller and Friedman, 2009)

Modelos Gráficos Probabilísticos

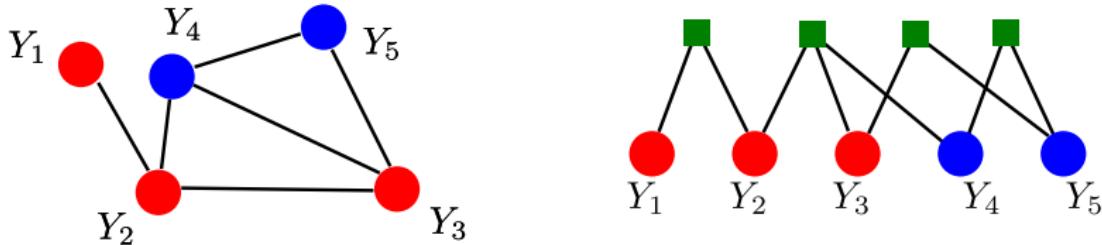


- Inspirados na Mecânica Estatística (Ising, 1925; Potts, 1952)
- Aplicações em teoria de códigos, visão, biologia computacional, ...
(Tanner, 1981; Pearl, 1988; Kschischang et al., 2001; Koller and Friedman, 2009)

Objectivo: obter a configuração de energia mínima:

$$E(y_1, y_2, y_3, y_4, y_5) = f_{12}(y_1, y_2) + f_{234}(y_2, y_3, y_4) + f_{35}(y_3, y_5) + f_{45}(y_4, y_5)$$

Modelos Gráficos Probabilísticos



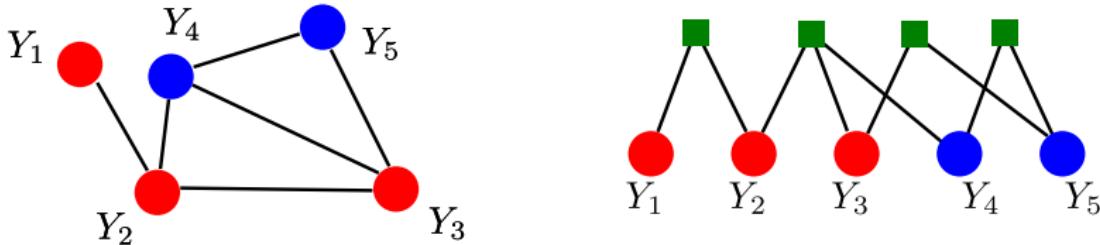
- Inspirados na Mecânica Estatística (Ising, 1925; Potts, 1952)
- Aplicações em teoria de códigos, visão, biologia computacional, ...
(Tanner, 1981; Pearl, 1988; Kschischang et al., 2001; Koller and Friedman, 2009)

Objectivo: obter a configuração de energia mínima:

$$E(y_1, y_2, y_3, y_4, y_5) = f_{12}(y_1, y_2) + f_{234}(y_2, y_3, y_4) + f_{35}(y_3, y_5) + f_{45}(y_4, y_5)$$

- Grafos *sem ciclos*: fácil com programação dinâmica (Viterbi, 1967)

Modelos Gráficos Probabilísticos



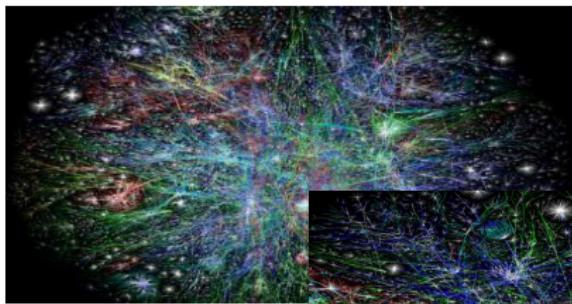
- Inspirados na Mecânica Estatística (Ising, 1925; Potts, 1952)
- Aplicações em teoria de códigos, visão, biologia computacional, ...
(Tanner, 1981; Pearl, 1988; Kschischang et al., 2001; Koller and Friedman, 2009)

Objectivo: obter a configuração de energia mínima:

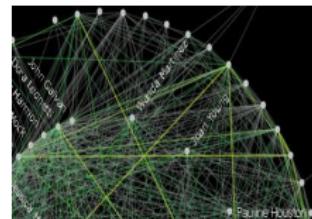
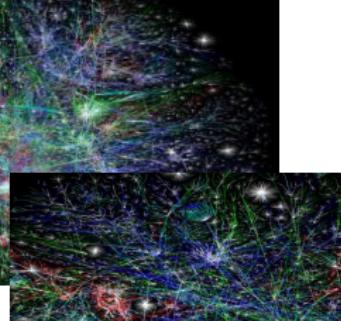
$$E(y_1, y_2, y_3, y_4, y_5) = f_{12}(y_1, y_2) + f_{234}(y_2, y_3, y_4) + f_{35}(y_3, y_5) + f_{45}(y_4, y_5)$$

- Grafos *sem ciclos*: fácil com programação dinâmica (Viterbi, 1967)
- **Em geral NP-completo!**

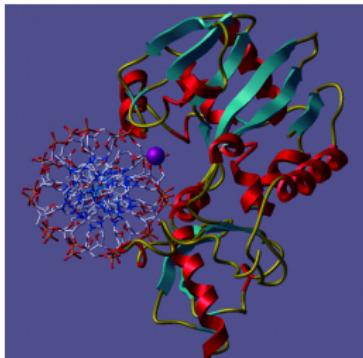
Exemplos



WWW graph



Facebook graph



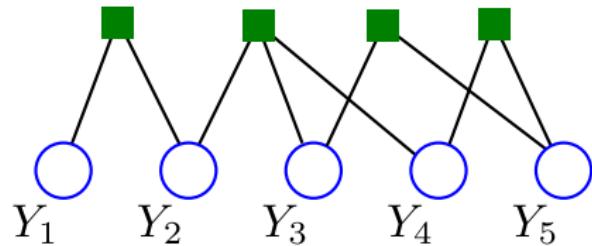
Protein folding



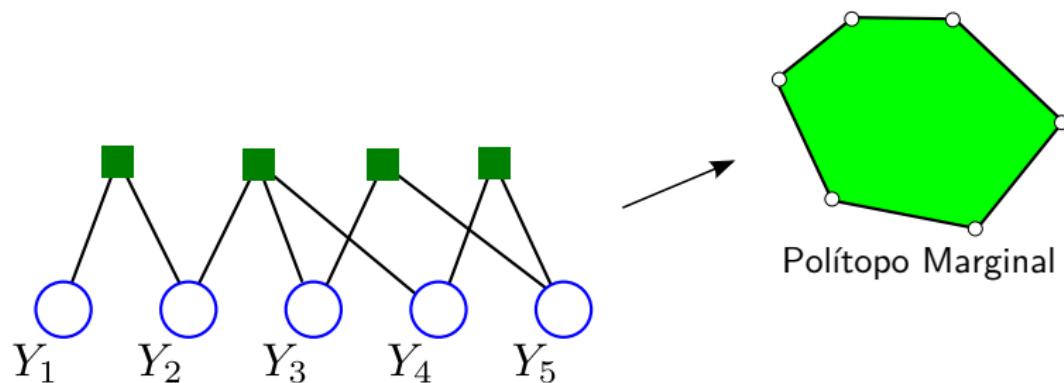
Image Segmentation



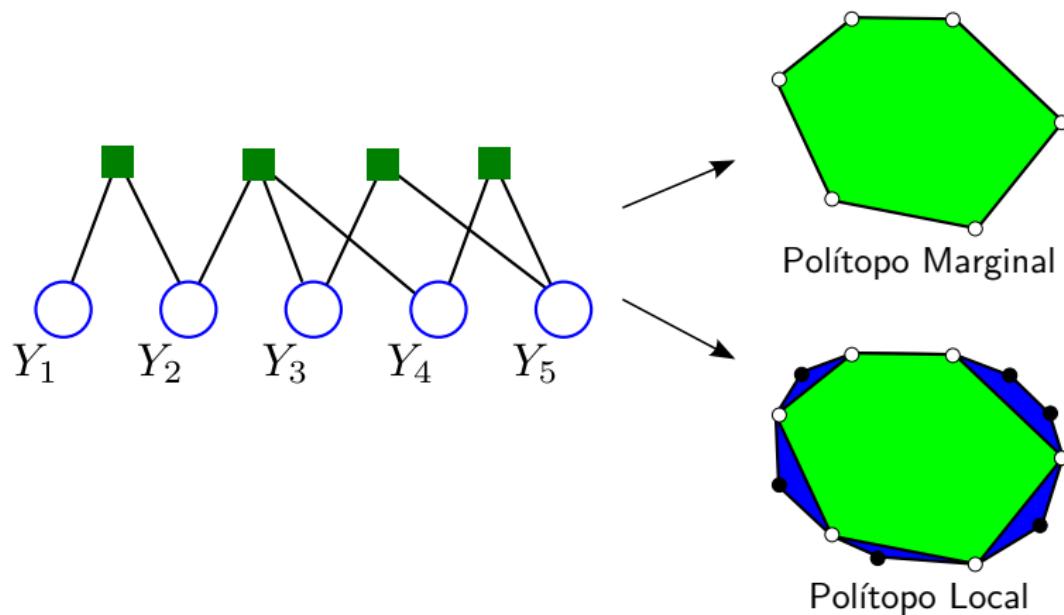
Polítopo Marginal e Polítopo Local



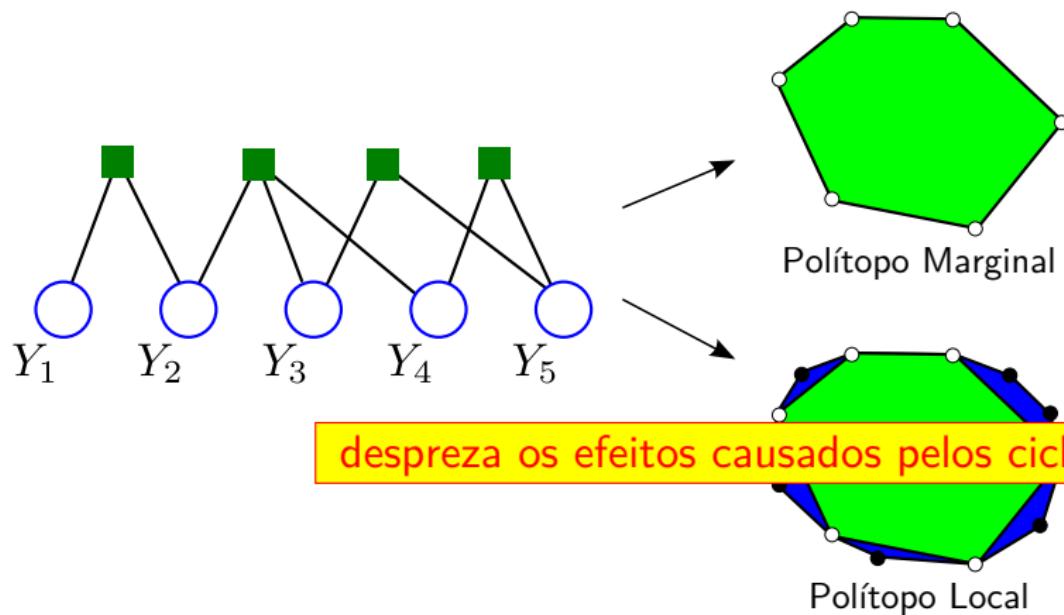
Polítopo Marginal e Polítopo Local



Polítopo Marginal e Polítopo Local



Polítopo Marginal e Polítopo Local

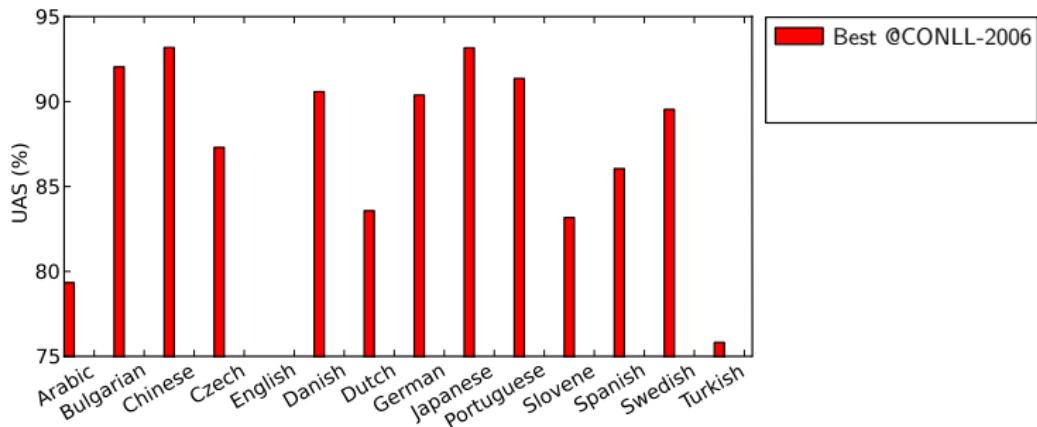


Algoritmo AD³

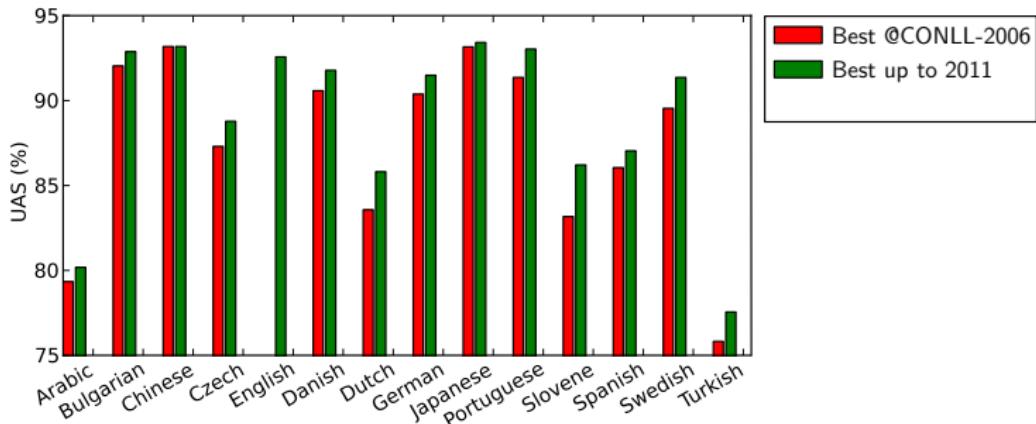
“Alternating Directions Dual Decomposition.”

- Resolve a relaxação tirando partido da estrutura do grafo
- **Ideia fundamental:** decompõe o grafo em *partes*, em cada iteração t resolve *subproblemas locais* e agrega as várias soluções tentando promover um *consenso*
- Taxa de convergência $O(1/t)$
- Semelhante a outros algoritmos de passagem de mensagens, mas com melhores propriedades de convergência
- Baseado em algoritmos clássicos de optimização (Glowinski and Marroco, 1975; Gabay and Mercier, 1976)

Experiências em Análise Sintáctica Multi-lingue



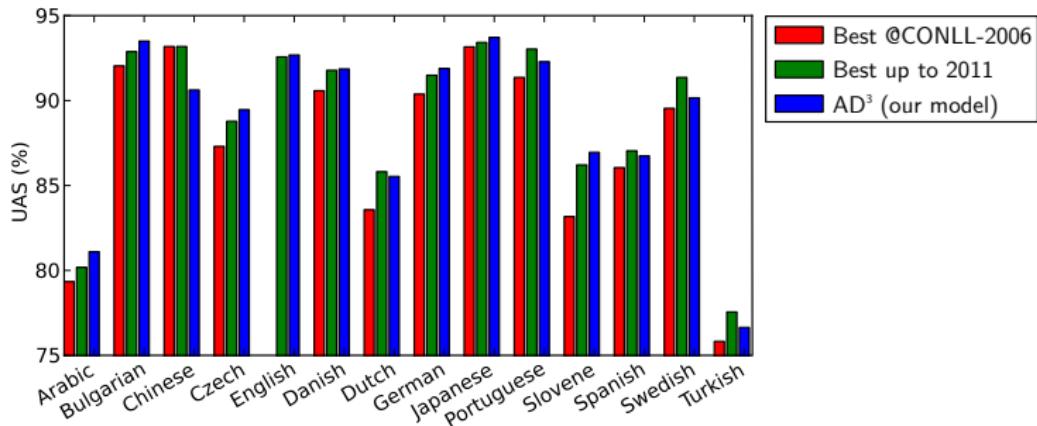
Experiências em Análise Sintáctica Multi-lingue



■ Os melhores resultados reportados até 2011 incluem:

- Nivre et al. (2006); McDonald et al. (2006); Nivre and McDonald (2008); Martins et al. (2008, 2010); Huang and Sagae (2010); Koo and Collins (2010); Koo et al. (2010)

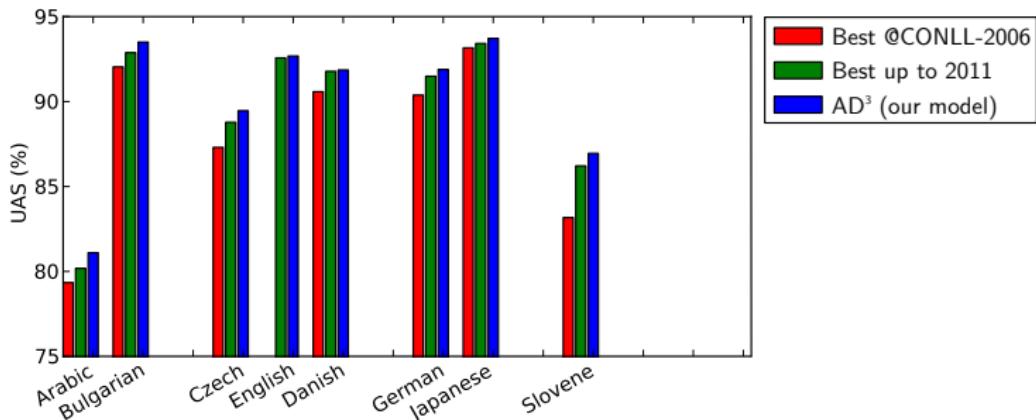
Experiências em Análise Sintáctica Multi-lingue



■ Os melhores resultados reportados até 2011 incluem:

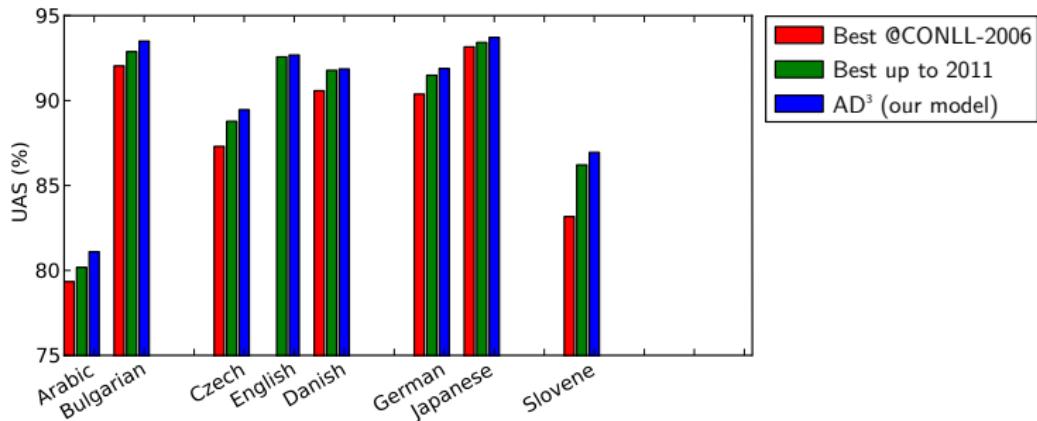
- Nivre et al. (2006); McDonald et al. (2006); Nivre and McDonald (2008); Martins et al. (2008, 2010); Huang and Sagae (2010); Koo and Collins (2010); Koo et al. (2010)

Experiências em Análise Sintáctica Multi-lingue



- Os melhores resultados reportados até 2011 incluem:
 - Nivre et al. (2006); McDonald et al. (2006); Nivre and McDonald (2008); Martins et al. (2008, 2010); Huang and Sagae (2010); Koo and Collins (2010); Koo et al. (2010)
- **AD³: estado-da-arte para 8 línguas em 14**

Experiências em Análise Sintáctica Multi-lingue



- Os melhores resultados reportados até 2011 incluem:
 - Nivre et al. (2006); McDonald et al. (2006); Nivre and McDonald (2008); Martins et al. (2008, 2010); Huang and Sagae (2010); Koo and Collins (2010); Koo et al. (2010)
- **AD³: estado-da-arte para 8 línguas em 14**
- **velocidade ≈ 1200 palavras por segundo.**

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente
- A *análise sintáctica* tem uma importância decisiva na qualidade de sistemas de pesquisa, mineração e tradução

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente
- A *análise sintáctica* tem uma importância decisiva na qualidade de sistemas de pesquisa, mineração e tradução
- **Este trabalho:** novos analisadores sintáticos estatísticos com grande capacidade expressiva

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente
- A *análise sintáctica* tem uma importância decisiva na qualidade de sistemas de pesquisa, mineração e tradução
- **Este trabalho:** novos analisadores sintáticos estatísticos com grande capacidade expressiva
- Nova interpretação do problema como minimização de energia num modelo gráfico com restrições

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente
- A *análise sintáctica* tem uma importância decisiva na qualidade de sistemas de pesquisa, mineração e tradução
- **Este trabalho:** novos analisadores sintáticos estatísticos com grande capacidade expressiva
- Nova interpretação do problema como minimização de energia num modelo gráfico com restrições
- **Turbo parsers:** relaxação que consiste em desprezar ciclos do grafo

Conclusões

- O processamento de *linguagem natural* é uma disciplina com interesse tecnológico crescente
- A *análise sintáctica* tem uma importância decisiva na qualidade de sistemas de pesquisa, mineração e tradução
- **Este trabalho:** novos analisadores sintáticos estatísticos com grande capacidade expressiva
- Nova interpretação do problema como minimização de energia num modelo gráfico com restrições
- **Turbo parsers:** relaxação que consiste em desprezar ciclos do grafo
- **AD³:** novo algoritmo que tira partido da estrutura do grafo

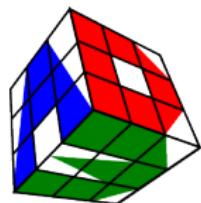
Obrigado!

O analisador sintáctico e o AD³ estão disponíveis em código aberto:



<http://www.ark.cs.cmu.edu/TurboParser>

<http://www.ark.cs.cmu.edu/AD3>



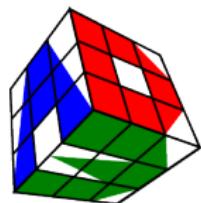
Obrigado!

O analisador sintáctico e o AD³ estão disponíveis em código aberto:



<http://www.ark.cs.cmu.edu/TurboParser>

<http://www.ark.cs.cmu.edu/AD3>



Mais sobre aprendizagem estatística:

Priberam Machine Learning Lunch Seminars, quinzenalmente, no IST.



References I

- Berrou, C., Glavieux, A., and Thitimajshima, P. (1993). Near Shannon limit error-correcting coding and decoding. In *Proc. of International Conference on Communications*, volume 93, pages 1064–1070.
- Charniak, E. (1996). Tree-bank grammars. In *Proc. of the National Conference on Artificial Intelligence*, pages 1031–1036.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*, volume 119. The MIT press.
- Collins, M. (1999). *Head-driven statistical models for natural language parsing*. PhD thesis, University of Pennsylvania.
- Collins, M. (2002). Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms. In *Proc. of Empirical Methods for Natural Language Processing*.
- Eisner, J. (1996). Three new probabilistic models for dependency parsing: An exploration. In *Proc. of International Conference on Computational Linguistics*, pages 340–345.
- Gabay, D. and Mercier, B. (1976). A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40.
- Glowinski, R. and Marroco, A. (1975). Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Rev. Franc. Automat. Inform. Rech. Operat.*, 9:41–76.
- Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24(2):8–12.
- Huang, L. and Sagae, K. (2010). Dynamic programming for linear-time incremental parsing. In *Proc. of Annual Meeting of the Association for Computational Linguistics*, pages 1077–1086.
- Hudson, R. (1984). *Word grammar*. Blackwell Oxford.
- Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258.
- Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics*, 24(4):613–632.
- Klein, D. and Manning, C. (2003). Accurate unlexicalized parsing. In *Proc. of Annual Meeting on Association for Computational Linguistics*, pages 423–430.
- Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press.
- Koo, T. and Collins, M. (2010). Efficient third-order dependency parsers. In *Proc. of Annual Meeting of the Association for Computational Linguistics*, pages 1–11.

References II

- Koo, T., Globerson, A., Carreras, X., and Collins, M. (2007). Structured prediction models via the matrix-tree theorem. In *Empirical Methods for Natural Language Processing*.
- Koo, T., Rush, A. M., Collins, M., Jaakkola, T., and Sontag, D. (2010). Dual decomposition for parsing with non-projective head automata. In *Proc. of Empirical Methods for Natural Language Processing*.
- Kschischang, F. R., Frey, B. J., and Loeliger, H. A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47.
- Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of International Conference of Machine Learning*.
- Magerman, D. (1995). Statistical decision-tree models for parsing. In *Proc. of Annual Meeting on Association for Computational Linguistics*, pages 276–283.
- Martins, A. F. T., Das, D., Smith, N. A., and Xing, E. P. (2008). Stacking Dependency Parsers. In *Proc. of Empirical Methods for Natural Language Processing*.
- Martins, A. F. T., Smith, N. A., Xing, E. P., Figueiredo, M. A. T., and Aguiar, P. M. Q. (2010). Turbo Parsers: Dependency Parsing by Approximate Variational Inference. In *Proc. of Empirical Methods for Natural Language Processing*.
- McDonald, R., Lerman, K., and Pereira, F. (2006). Multilingual dependency analysis with a two-stage discriminative parser. In *Proc. of International Conference on Natural Language Learning*.
- McDonald, R. T., Pereira, F., Ribarov, K., and Hajic, J. (2005). Non-projective dependency parsing using spanning tree algorithms. In *Proc. of Empirical Methods for Natural Language Processing*.
- McElice, R. J., MacKay, D. J. C., and Cheng, J. F. (1998). Turbo decoding as an instance of Pearl's "belief propagation" algorithm. *IEEE Journal on Selected Areas in Communications*, 16(2).
- Mel'čuk, I. (1988). *Dependency syntax: theory and practice*. State University of New York Press.
- Nivre, J., Hall, J., Nilsson, J., Eryiğit, G., and Marinov, S. (2006). Labeled pseudo-projective dependency parsing with support vector machines. In *Proc. of International Conference on Natural Language Learning*.
- Nivre, J. and McDonald, R. (2008). Integrating graph-based and transition-based dependency parsers. In *Proc. of Annual Meeting of the Association for Computational Linguistics*.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

References III

- Potts, R. (1952). Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society*, volume 48, pages 106–109. Cambridge Univ Press.
- Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine Learning*, 62(1):107–136.
- Roth, D. and Yih, W. (2004). A linear programming formulation for global inference in natural language tasks. In *International Conference on Natural Language Learning*.
- Tanner, R. (1981). A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547.
- Taskar, B., Guestrin, C., and Koller, D. (2003). Max-margin Markov networks. In *Proc. of Neural Information Processing Systems*.
- Tesnière, L. (1959). *Eléments de syntaxe structurale*. Librairie C. Klincksieck.
- Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269.