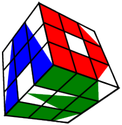


AD^3: A New Decoder for Structured Prediction

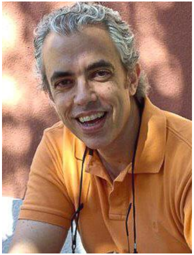
André Martins

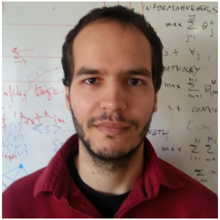
priberam

instituto de telecomunicações

Joint NAACL/ICML Symposium—Atlanta, 15/06/13

Collaborators





- Mário Figueiredo, Noah Smith, Pedro Aguiar, Eric Xing, Miguel Almeida.

Structured Prediction and NLP

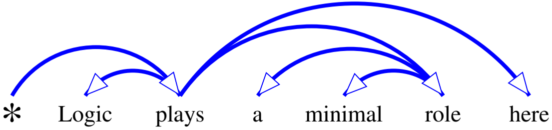
Structured prediction: a machine learning framework for predicting structured, constrained, and interdependent outputs

NLP deals with *structured* and *ambiguous* textual data:

- machine translation
- speech recognition
- syntactic parsing
- semantic parsing
- information extraction
- ...

Dependency Parsing

Map **sentences** to their **syntactic structure.**



- A lexicalized syntactic formalism
- Grammar functions represented as lexical relationships (dependencies)

(Eisner, 1996; McDonald et al., 2005; Nivre et al., 2006; Koo et al., 2007)

Multi-Document Summarization

Map a set of related **documents** to a brief **summary.**

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

Multi-Document Summarization

Map a set of related **documents** to a brief **summary.**

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations
- Statements in FOL

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations
- Statements in FOL
- Budget/knapsack constraints

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations
- Statements in FOL
- Budget/knapsack constraints
- Combination of structured models

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations
- Statements in FOL
- Budget/knapsack constraints
- Combination of structured models

We'll illustrate with two applications:

1. **Dependency parsing**
2. **Summarization**

This talk: AD^3

A new dual decomposition algorithm, suitable for many scenarios encountered in NLP and IR.

- Problems with factor graph representations
- Statements in FOL
- Budget/knapsack constraints
- Combination of structured models

We'll illustrate with two applications:

1. **Dependency parsing**
2. **Summarization**

Could be a great fit to many other applications!!

Outline

Outline

Structured Prediction

- For each $x \in \mathcal{X}$: a **large** set of candidate outputs $\mathcal{Y}(x)$
- **Decoding problem:**

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} F_{\mathbf{w}}(x, y)$$

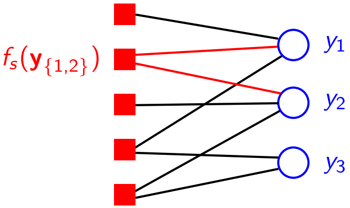
Structured Prediction

- For each $x \in \mathcal{X}$: a **large** set of candidate outputs $\mathcal{Y}(x)$
- **Decoding problem:**

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} F_{\mathbf{w}}(x, y)$$

- **Key assumption:** $F_{\mathbf{w}}$ decomposes into (overlapping) *parts*

$$\boxed{F_{\mathbf{w}}(x, y) := \sum_s f_s(\mathbf{y}_s)}$$



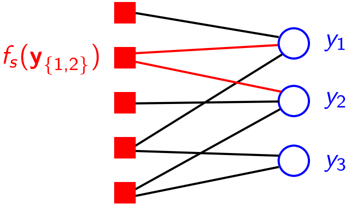
Structured Prediction

- For each $x \in \mathcal{X}$: a **large** set of candidate outputs $\mathcal{Y}(x)$
- **Decoding problem:**

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} F_{\mathbf{w}}(x, y)$$

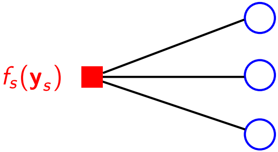
- **Key assumption:** $F_{\mathbf{w}}$ decomposes into (overlapping) *parts*

$$\boxed{F_{\mathbf{w}}(x, y) := \sum_{s} f_s(\mathbf{y}_s)}$$

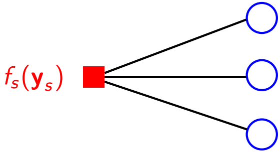


- **Examples:** HMMs, CRFs, PCFGs, general graphical models

What's in a Factor?



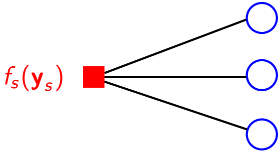
What's in a Factor?



Several kinds of factors:

1. **Dense:** $O(\exp(|N(s)|))$ degrees of freedom

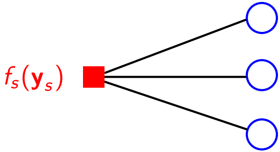
What's in a Factor?



Several kinds of factors:

1. **Dense:** $O(\exp(|N(s)|))$ degrees of freedom
2. **Structured factors:** $f_s(\mathbf{y}_s)$ is structured itself.
   E.g.: a sequence model, subtree, head automaton, PCFG (Smith and Eisner, 2008; Koo et al., 2010; Rush et al., 2010)

What's in a Factor?

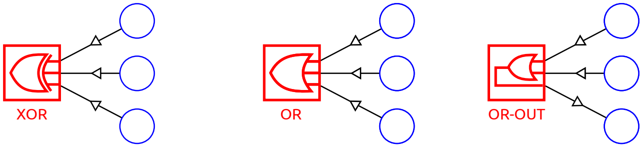


Several kinds of factors:

1. **Dense:** $O(\exp(|N(s)|))$ degrees of freedom
2. **Structured factors:** $f_s(\mathbf{y}_s)$ is structured itself.
   E.g.: a sequence model, subtree, head automaton, PCFG (Smith and Eisner, 2008; Koo et al., 2010; Rush et al., 2010)
3. **Hard constraint factors:**

$$f_s(\mathbf{y}_s) := \begin{cases} 0, & \text{if } \mathbf{y}_s \in \mathcal{Y}_s \\ -\infty, & \text{otherwise.} \end{cases}$$

Example: Hard Constraint Factors

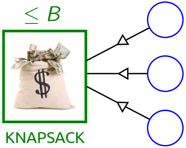


Logic factors: can express arbitrary FOL constraints

- Applications: Markov logic networks (Richardson and Domingos, 2006), constrained conditional models (Roth and Yih, 2004)

Knapsack factor: can express budget constraints

- Applications: summarization, diversity problems,...



(Martins et al., 2011b, 2012; Almeida and Martins, 2013)

Global/Local Decoding

Global decoding problem:

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} \sum_{s} f_s(\mathbf{y}_s)$$

Solvable with dynamic programming when the graph is a tree...

Global/Local Decoding

Global decoding problem:

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} \sum_s f_s(\mathbf{y}_s)$$

Solvable with dynamic programming when the graph is a tree...

Typically challenging otherwise:

- combination of dynamic programs **(blow-up the number of states)**
- non-projective parsing with higher-order features **(NP-hard)**
- summarization as relevance/redundancy trade-off **(NP-hard)**
- summarization as max-cover with knapsack constraints **(NP-hard)**

Global/Local Decoding

Global decoding problem:

$$\widehat{y} = \arg\max_{y \in \mathcal{Y}(x)} \sum_s f_s(\mathbf{y}_s)$$

Solvable with dynamic programming when the graph is a tree...

Typically challenging otherwise:

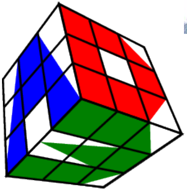
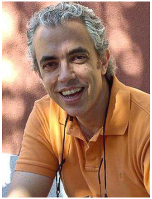
- combination of dynamic programs **(blow-up the number of states)**
- non-projective parsing with higher-order features **(NP-hard)**
- summarization as relevance/redundancy trade-off **(NP-hard)**
- summarization as max-cover with knapsack constraints **(NP-hard)**

We'll build (approximate) global decoders given only local decoders.

Outline

1. Structured Prediction and Factor Graphs
2. AD^3: Alternating Directions Dual Decomposition
3. Turbo Parsers
4. Compressive Summarization
5. Conclusions

AD^3: Alternating Directions Dual Decomposition


- A. Martins, M. Figueiredo, P. Aguiar, N. Smith, E. Xing.
  "An Augmented Lagrangian Approach to Constrained MAP Inference."
  ICML, Bellevue, USA, 2011.

Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

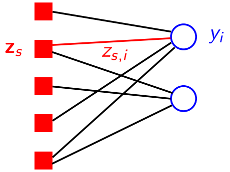
- Define *local* variables $\langle \mathbf{z}_s \rangle_{s=1}^{S}$ ("views")

Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

- Define *local* variables $\langle \mathbf{z}_s \rangle_{s=1}^{S}$ ("views")
- Reformulate the problem by constraining views to agree on overlaps:

$$\begin{aligned} \text{maximize} \quad & \sum_s f_s(\mathbf{z}_s) \\ \text{w.r.t.} \quad & \mathbf{y} \in \mathbb{R}^{|V|},\ \mathbf{z}_s \in \mathcal{Y}_s,\ \forall s \\ \text{s.t.} \quad & y_i = z_{s,i},\ \forall s, i \end{aligned}$$

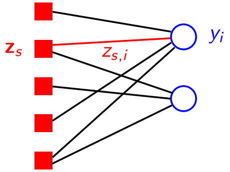


Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

- Define *local* variables $\langle \mathbf{z}_s \rangle_{s=1}^{S}$ ("views")
- Reformulate the problem by constraining views to agree on overlaps:

$$\begin{array}{rl} \text{maximize} & \sum_s f_s(\mathbf{z}_s) \\ \text{w.r.t.} & \mathbf{y} \in \mathbb{R}^{|V|},\ \mathbf{z}_s \in \mathcal{Y}_s,\ \forall s \\ \text{s.t.} & y_i = z_{s,i},\ \forall s, i \end{array}$$



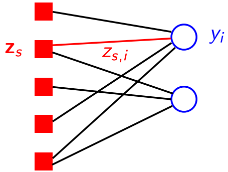
1. **Relax:** replace combinatorial set $\mathcal{Y}_s$ by $\mathcal{Z}_s := \text{conv}(\mathcal{Y}_s)$

Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

- Define *local* variables $\langle \mathbf{z}_s \rangle_{s=1}^{S}$ ("views")
- Reformulate the problem by constraining views to agree on overlaps:

$$\begin{array}{rl} \text{maximize} & \sum_s f_s(\mathbf{z}_s) \\ \text{w.r.t.} & \mathbf{y} \in \mathbb{R}^{|V|},\ \mathbf{z}_s \in \mathcal{Z}_s,\ \forall s \\ \text{s.t.} & y_i = z_{s,i},\ \forall s, i \end{array}$$



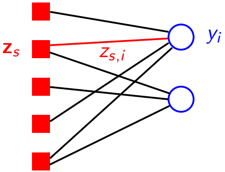
1. **Relax:** replace combinatorial set $\mathcal{Y}_s$ by $\mathcal{Z}_s := \text{conv}(\mathcal{Y}_s)$

Dual Decomposition (Komodakis et al., 2007; Rush et al., 2010)

Solve a relaxation of the original problem.

- Define *local* variables $\langle \mathbf{z}_s \rangle_{s=1}^{S}$ ("views")
- Reformulate the problem by constraining views to agree on overlaps:

$$\begin{array}{rl} \text{maximize} & \sum_s f_s(\mathbf{z}_s) \\ \text{w.r.t.} & \mathbf{y} \in \mathbb{R}^{|V|}, \ \mathbf{z}_s \in \mathcal{Z}_s, \ \forall s \\ \text{s.t.} & y_i = z_{s,i}, \ \forall s, i \end{array}$$



1. **Relax:** replace combinatorial set $\mathcal{Y}_s$ by $\mathcal{Z}_s := \text{conv}(\mathcal{Y}_s)$
2. Dualize out the equality constraints via **Lagrange multipliers** $\lambda_{s,i}$

Projected Subgradient (Komodakis et al., 2007)

initialize $\lambda = \mathbf{0}$

Projected Subgradient (Komodakis et al., 2007)

initialize $\lambda = \mathbf{0}$
repeat

until certificate found or maximum number of iterations reached

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$
repeat
  for each factor $s = 1, \ldots, S$ **do**

  end for

until certificate found or maximum number of iterations reached

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$
repeat
  for each factor $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \underbrace{\arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \left(f_s(\mathbf{z}'_s) + \sum_{i \in N(s)} \lambda_{s,i} z'_{s,i} \right)}_{\text{local maximization}}$$

  end for

until certificate found or maximum number of iterations reached

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$

repeat

for each factor $s = 1, \ldots, S$ **do**

z-step:

$$\mathbf{z}_s := \underbrace{\arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \left(f_s(\mathbf{z}'_s) + \sum_{i \in N(s)} \lambda_{s,i} z'_{s,i} \right)}_{\text{local maximization}}$$

end for

y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$ $\longleftarrow$ **averaging**

until certificate found or maximum number of iterations reached

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$
repeat
for each factor $s = 1, \ldots, S$ **do**
z-step:

$$\mathbf{z}_s := \underbrace{\arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \left(f_s(\mathbf{z}'_s) + \sum_{i \in N(s)} \lambda_{s,i} z'_{s,i} \right)}_{\text{local maximization}}$$

end for
y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$ $\longleftarrow$ **averaging**
$\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta\left(z_{s,i} - y_i\right)$ $\longleftarrow$ **price adjustment**
until certificate found or maximum number of iterations reached

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$
repeat
  for each factor $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \underbrace{\arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \left(f_s(\mathbf{z}'_s) + \sum_{i \in N(s)} \lambda_{s,i} z'_{s,i} \right)}_{\text{local maximization}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$ $\longleftarrow$ **averaging**
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta \left(z_{s,i} - y_i \right)$ $\longleftarrow$ **price adjustment**
until certificate found or maximum number of iterations reached

- **certificate of optimality:** $z_{s,i} = y_i$, $\forall s$ (only if relaxation is tight)

Projected Subgradient (Komodakis et al., 2007)

initialize $\boldsymbol{\lambda} = \mathbf{0}$
repeat
  for each factor $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \underbrace{\arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \left(f_s(\mathbf{z}'_s) + \sum_{i \in N(s)} \lambda_{s,i} z'_{s,i} \right)}_{\text{local maximization}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$ $\longleftarrow$ **averaging**
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta \left(z_{s,i} - y_i \right)$ $\longleftarrow$ **price adjustment**
until certificate found or maximum number of iterations reached

- **certificate of optimality:** $z_{s,i} = y_i, \ \forall s$ (only if relaxation is tight)
- **problem:** convergence is slow when there are many factors

Accelerating Consensus

How can dual decomposition be accelerated?

Accelerating Consensus

How can dual decomposition be accelerated?

Alternating Direction Method of Multipliers (ADMM):

- old method in optimization inspired by augmented Lagrangians (Gabay and Mercier, 1976; Glowinski and Marroco, 1975)
- a natural fit to consensus problems
- lot of recent applications in ML (Boyd et al., 2011)

Accelerating Consensus

How can dual decomposition be accelerated?

Alternating Direction Method of Multipliers (ADMM):

- old method in optimization inspired by augmented Lagrangians (Gabay and Mercier, 1976; Glowinski and Marroco, 1975)
- a natural fit to consensus problems
- lot of recent applications in ML (Boyd et al., 2011)

AD^3 = *ADMM applied to graphical models* (Martins et al., 2010a, 2011a)

From Subgradient to AD^3 (Martins et al., 2011a)

initialize $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{y}$ uniformly
repeat
  for each component $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \; f_s(\mathbf{z}_s) + \sum_{i \in N(s)} \lambda_{s,i} z_{s,i}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta \left(z_{s,i} - y_i \right)$
until certificate found or maximum number of iterations reached

From Subgradient to AD^3 (Martins et al., 2011a)

initialize $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{y}$ uniformly
repeat
  for each component $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} f_s(\mathbf{z}_s) + \sum_{i \in N(s)} \lambda_{s,i} z_{s,i} - \underbrace{\frac{\eta}{2} \sum_{i \in N(s)} (z_{s,i} - y_i)^2}_{\text{penalty term}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta\left(z_{s,i} - y_i\right)$
until certificate found or maximum number of iterations reached

From Subgradient to AD^3 (Martins et al., 2011a)

initialize $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{y}$ uniformly
repeat
  for each component $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} f_s(\mathbf{z}_s) + \sum_{i \in N(s)} \lambda_{s,i} z_{s,i} - \underbrace{\frac{\eta}{2} \sum_{i \in N(s)} (z_{s,i} - y_i)^2}_{\textbf{penalty term}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta \left(z_{s,i} - y_i \right)$
until certificate found or maximum number of iterations reached

- **faster consensus:** regularize $\mathbf{z}$-step towards average votes in $\mathbf{y}$

From Subgradient to AD^3 (Martins et al., 2011a)

initialize $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{y}$ uniformly
repeat
  for each component $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} \; f_s(\mathbf{z}_s) + \sum_{i \in N(s)} \lambda_{s,i} z_{s,i} - \underbrace{\frac{\eta}{2} \sum_{i \in N(s)} (z_{s,i} - y_i)^2}_{\text{penalty term}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta \left(z_{s,i} - y_i \right)$
until certificate found or maximum number of iterations reached

- **faster consensus:** regularize **z**-step towards average votes in **y**
- **better stopping conditions:** keeps track of primal and dual residuals

From Subgradient to AD^3 (Martins et al., 2011a)

initialize $\boldsymbol{\lambda} = \mathbf{0}$ and $\mathbf{y}$ uniformly
repeat
  for each component $s = 1, \ldots, S$ **do**
    z-step:

$$\mathbf{z}_s := \arg\max_{\mathbf{z}'_s \in \mathcal{Z}_s} f_s(\mathbf{z}_s) + \sum_{i \in N(s)} \lambda_{s,i} z_{s,i} - \underbrace{\frac{\eta}{2} \sum_{i \in N(s)} (z_{s,i} - y_i)^2}_{\text{penalty term}}$$

  end for
  y-step: $y_i := |N(i)|^{-1} \sum_{s \in N(i)} z_{s,i}$
  $\boldsymbol{\lambda}$-step: $\lambda_{s,i} := \lambda_{s,i} - \eta\,(z_{s,i} - y_i)$
until certificate found or **primal/dual residuals below threshold**

- **faster consensus:** regularize **z**-step towards average votes in **y**
- **better stopping conditions:** keeps track of primal and dual residuals

Theoretical Guarantees of AD^3

Convergent in primal and dual (Glowinski and Le Tallec, 1989)

Iteration bound: $O(1/\epsilon)$ (cf. $O(1/\epsilon^2)$ for projected subgradient)

Inexact AD^3 subproblems: still convergent if residuals are summable (Eckstein and Bertsekas, 1992)

Always dual feasible: can compute upper bounds

Theoretical Guarantees of AD^3

Convergent in primal and dual (Glowinski and Le Tallec, 1989)

Iteration bound: $O(1/\epsilon)$ (cf. $O(1/\epsilon^2)$ for projected subgradient)

Inexact AD^3 subproblems: still convergent if residuals are summable (Eckstein and Bertsekas, 1992)

Always dual feasible: can compute upper bounds

May embed AD^3 in a branch-and-bound procedure for *exact* decoding:

- D. Das, A. Martins, N. Smith.
  "An Exact DD Algorithm for Shallow Semantic Parsing with Constraints."
  *SEM Workshop, 2012.

AD^3 Subproblems

$$
\begin{array}{rl}
\textbf{maximize} & f_s(\mathbf{z}_s) + \sum_i \lambda_{s,i} z_{s,i} - \overbrace{\frac{\eta}{2}\sum_i (z_{s,i} - y_i)^2}^{\textbf{penalty term}} \\
\textbf{w.r.t.} & \mathbf{z}_s \in \mathcal{Z}_s
\end{array}
$$

AD^3 Subproblems

$$\begin{array}{rl} \textbf{maximize} & f_s(\mathbf{z}_s) + \sum_i \lambda_{s,i} z_{s,i} - \overbrace{\frac{\eta}{2} \sum_i (z_{s,i} - y_i)^2}^{\text{penalty term}} \\ \textbf{w.r.t.} & \mathbf{z}_s \in \mathcal{Z}_s \end{array}$$

- **More involved than in projected subgradient**

AD^3 Subproblems

$$\begin{array}{rl} \textbf{maximize} & f_s(\mathbf{z}_s) + \sum_i \lambda_{s,i} z_{s,i} - \overbrace{\frac{\eta}{2}\sum_i (z_{s,i} - y_i)^2}^{\text{penalty term}} \\ \textbf{w.r.t.} & \mathbf{z}_s \in \mathcal{Z}_s \end{array}$$

- More involved than in projected subgradient
- Closed form solution for pairwise binary factors (Ising models)

AD^3 Subproblems

$$\begin{array}{rl} \textbf{maximize} & f_s(\mathbf{z}_s) + \sum_i \lambda_{s,i} z_{s,i} - \overbrace{\frac{\eta}{2} \sum_i (z_{s,i} - y_i)^2}^{\text{penalty term}} \\ \textbf{w.r.t.} & \mathbf{z}_s \in \mathcal{Z}_s \end{array}$$

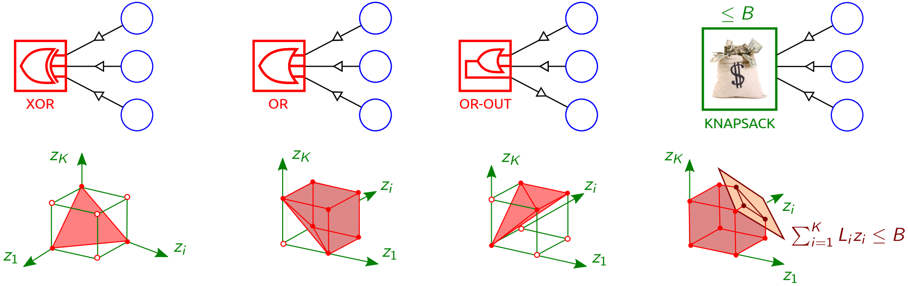
- **More involved than in projected subgradient**
- **Closed form solution for pairwise binary factors (Ising models)**
- **Hard constraint factors:** projection onto the marginal polytope $\mathcal{Z}_s$

AD^3 Subproblems

$$\begin{array}{rl} \textbf{maximize} & f_s(\mathbf{z}_s) + \sum_i \lambda_{s,i} z_{s,i} - \overbrace{\frac{\eta}{2} \sum_i (z_{s,i} - y_i)^2}^{\text{penalty term}} \\ \textbf{w.r.t.} & \mathbf{z}_s \in \mathcal{Z}_s \end{array}$$

- **More involved than in projected subgradient**
- **Closed form solution for pairwise binary factors (Ising models)**
- **Hard constraint factors:** projection onto the marginal polytope $\mathcal{Z}_s$
- **Very easy and efficient for logic and knapsack factors!**

Projecting onto Hard Constraint Polytopes



- Martins et al. (2012): logic factors can be solved in $O(K)$ time
- **Almeida and Martins (2013): same for knapsack factors!**

Structured Factors

What about structured factors?

Structured Factors

What about structured factors?

Projected subgradient handles these quite well

- combinatorial machinery (Viterbi, Chu-Liu-Edmonds, Fulkerson-Ford, Floyd-Warshall,...)

We cannot solve the AD^3 subproblems with that machinery...

Structured Factors

What about structured factors?

Projected subgradient handles these quite well

- combinatorial machinery (Viterbi, Chu-Liu-Edmonds, Fulkerson-Ford, Floyd-Warshall,...)

We cannot solve the AD^3 subproblems with that machinery...

Or can we?

Structured Factors

What about structured factors?

Projected subgradient handles these quite well

- combinatorial machinery (Viterbi, Chu-Liu-Edmonds, Fulkerson-Ford, Floyd-Warshall,...)

We cannot solve the AD^3 subproblems with that machinery...

Or can we?

Active set method: seek the support of the solution by adding/removing components; very suitable for warm-starting (Nocedal and Wright, 1999)

Only requirement is a local-max oracle (as in projected subgradient)

Structured Factors

What about structured factors?

Projected subgradient handles these quite well

- combinatorial machinery (Viterbi, Chu-Liu-Edmonds, Fulkerson-Ford, Floyd-Warshall,...)

We cannot solve the AD^3 subproblems with that machinery...

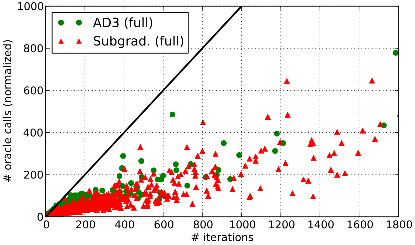
Or can we?

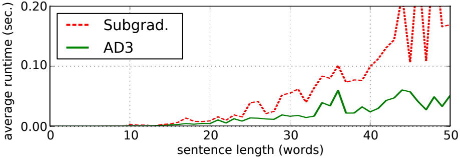
Active set method: seek the support of the solution by adding/removing components; very suitable for warm-starting (Nocedal and Wright, 1999)

Only requirement is a local-max oracle (as in projected subgradient)

More info: Martins et al. (2012)

Runtime of AD^3 vs Subgradient (Parsing)

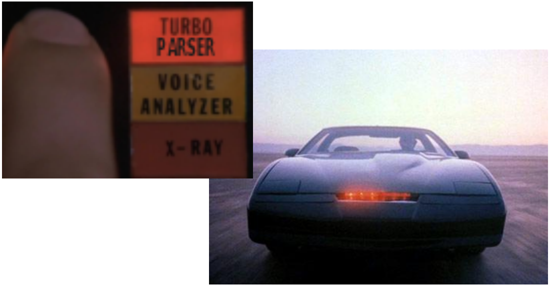




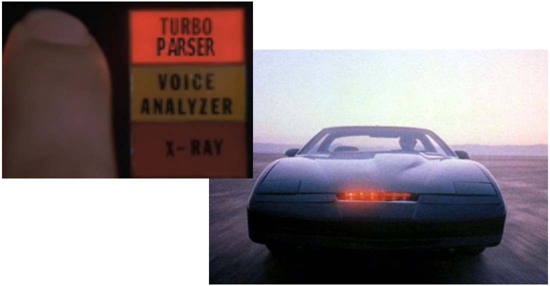
- **Caching** and **warm-starting** the subproblems reduces drastically the number of oracle calls—**huge speed-ups!!**
- **AD^3 faster to achieve consensus** (due to the quadratic penalty)

Outline

What is a Turbo Parser?

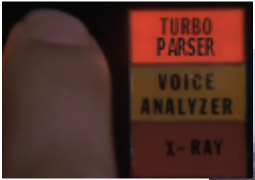


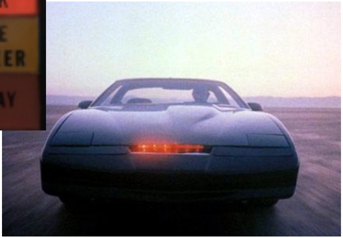
What is a Turbo Parser?



- **A parser that runs inference in factor graphs, ignoring global effects caused by loops (Martins et al., 2010b)**
- name inspired from *turbo* decoders (Berrou et al., 1993)

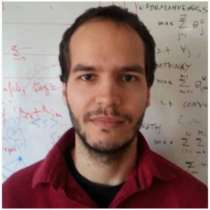
What is a Turbo Parser?





- **A parser that runs inference in factor graphs, ignoring global effects caused by loops (Martins et al., 2010b)**
- name inspired from *turbo* decoders (Berrou et al., 1993)
- **This talk:** we speed up turbo parsers via AD^3 w/ active set

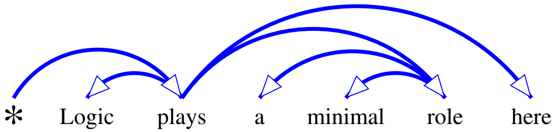
Forthcoming Paper

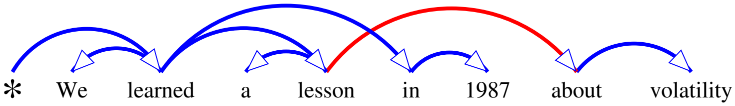

- André F. T. Martins, Miguel B. Almeida, Noah A. Smith.
  "Turning on the Turbo: Fast Third-Order Non-Projective Turbo Parsers."
  To appear at ACL 2013.

An Important Distinction

- A projective tree:

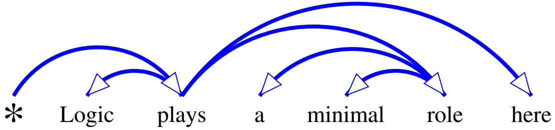


- A non-projective tree:

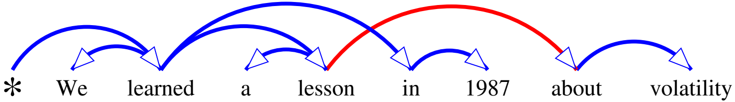


An Important Distinction

- A projective tree:



- A non-projective tree:



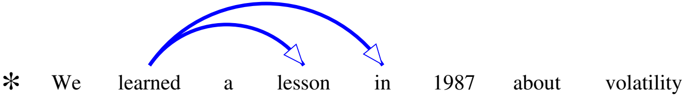
This talk: we allow non-projective trees.

Suitable for languages with flexible word order (Dutch, German, Czech,...)

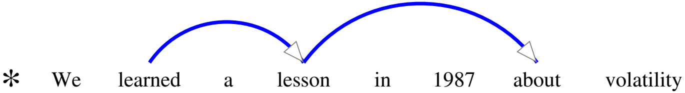
First-Order Scores for Arcs

* We learned a lesson in 1987 about volatility

Second-Order Scores for Consecutive Siblings



Second-Order Scores for Grandparents



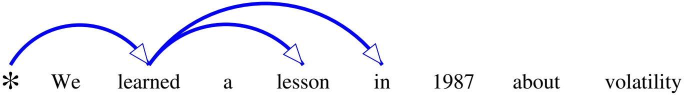
Scores for Arbitrary Siblings

* We learned a lesson in 1987 about volatility

Scores for Head Bigrams

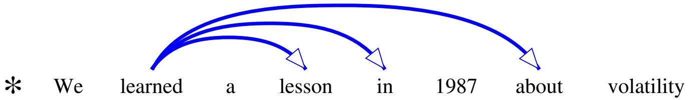
* We <u>learned</u> <u>a</u> lesson in 1987 about volatility

Third-Order Scores for Grand-siblings



Used by Koo and Collins (2010) for *projective* parsing.

Third-Order Scores for Tri-siblings



Used by Koo and Collins (2010) for *projective* parsing.

Decoding

How to deal with all these parts?

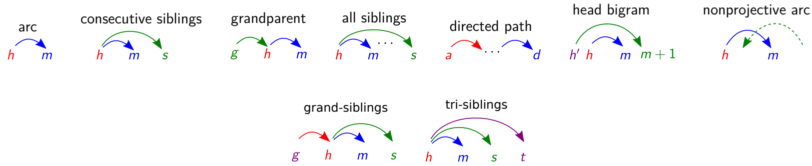
- Dynamic programming only available for the *projective* case...

Decoding

How to deal with all these parts?

- Dynamic programming only available for the *projective* case...
- Beyond arc-factored models, non-projective parsing is **NP-hard** (McDonald and Satta, 2007)
- **Need to embrace approximations!**

Approximate Dependency Parsers



	parser	AF	CS	G	AS	DP	HB	NPA	GS	TS
McDonald et al. (2006)	projective + greedy	✓	✓							
Smith et al. (2008)	loopy BP	✓	✓	✓	✓					
Martins et al. (2010)	LP solver	✓		✓	✓			✓		
Koo et al. (2010)	dual decomp.	✓	✓							
Martins et al. (2011)	AD^3	✓	✓	✓	✓	✓	✓	✓		
This work	AD^3 & active set	✓	✓	✓	✓		✓		✓	✓

Factor Graph Representation

- Variables nodes for **dependency arcs**, linked to a tree constraint
- Head automata for **consecutive siblings and grandparents** (as in Smith and Eisner (2008); Koo et al. (2010))
- Pairwise factors for **arbitrary siblings** (as Martins et al. (2011b))

Factor Graph Representation

- Variables nodes for **dependency arcs**, linked to a tree constraint
- Head automata for **consecutive siblings and grandparents** (as in Smith and Eisner (2008); Koo et al. (2010))
- Pairwise factors for **arbitrary siblings** (as Martins et al. (2011b))
- Third-order head automata for **grand-siblings and tri-siblings**
- Sequence model for **head bigrams**

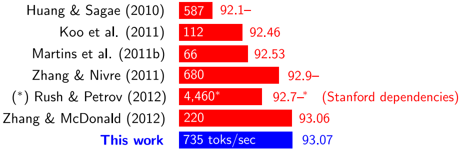
Factor Graph Representation

- Variables nodes for **dependency arcs**, linked to a tree constraint
- Head automata for **consecutive siblings and grandparents** (as in Smith and Eisner (2008); Koo et al. (2010))
- Pairwise factors for **arbitrary siblings** (as Martins et al. (2011b))
- Third-order head automata for **grand-siblings and tri-siblings**
- Sequence model for **head bigrams**

We solve an LP relaxation with AD^3.

Parsing Accuracies/Runtimes

- Projective dataset (English PTB):



Parsing Accuracies/Runtimes

- Projective dataset (English PTB):

	Runtime	Accuracy	
Huang & Sagae (2010)	587	92.1–	
Koo et al. (2011)	112	92.46	
Martins et al. (2011b)	66	92.53	
Zhang & Nivre (2011)	680	92.9–	
(*) Rush & Petrov (2012)	4,460*	92.7–*	(Stanford dependencies)
Zhang & McDonald (2012)	220	93.06	
This work	735 toks/sec	93.07	

- Non-projective datasets (CoNLL-2006 and CoNLL-2008):

EN

	Runtime	Accuracy
Koo et al. (2011)	131	92.57
Martins et al. (2011b)	–	92.68
This work	785 toks/sec	**93.22**

DE

	Runtime	Accuracy
Martins et al. (2011b)	–	91.89
Rush & Petrov (2012)	2,880	90.8–
Zhang & McDonald (2012)	–	91.35
This work	965 toks/sec	**92.41**

NL

	Runtime	Accuracy
Koo et al. (2011)	121	85.81
Martins et al. (2011b)	–	85.53
This work	599 toks/sec	**86.19**

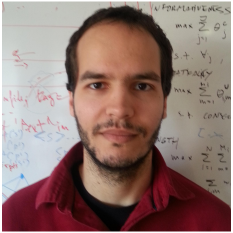
CZ

	Runtime	Accuracy
Martins et al. (2010)	–	88.78
Martins et al. (2011b)	–	89.46
This work	501 toks/sec	**90.32**

We get SOTA accuracies for the largest non-projective datasets!

Outline

Forthcoming Paper



- Miguel B. Almeida and André F. T. Martins.
  "Fast and Robust Compressive Summarization with Dual Decomposition and Multi-Task Learning."
  To appear at ACL 2013.

Multi-Document Summarization

Map a set of related **documents** to a brief **summary.**

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

Multi-Document Summarization

Map a set of related **documents** to a brief **summary.**

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

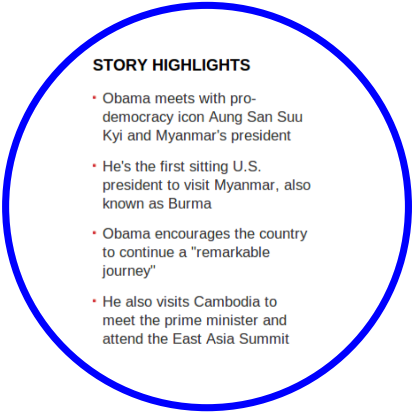
BBC NEWS

What Makes a Good Summary?

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

What Makes a Good Summary?



STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

1. conciseness

What Makes a Good Summary?

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

1. conciseness
2. informativeness

What Makes a Good Summary?

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

1. conciseness
2. informativeness
3. grammaticality

Extractive Summarization

Just **extract** the most salient sentences.

Extractive Summarization

Just **extract** the most salient sentences.

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

Extractive Summarization

Just **extract** the most salient sentences.

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

What We Do: Compressive Summarization

Jointly **extract** and **compress** sentences.

What We Do: Compressive Summarization

Jointly **extract** and **compress** sentences.

CNN

Obama hopes for 'continued progress' in Myanmar

STORY HIGHLIGHTS

- Obama meets with pro-democracy icon Aung San Suu Kyi and Myanmar's president
- He's the first sitting U.S. president to visit Myanmar, also known as Burma
- Obama encourages the country to continue a "remarkable journey"
- He also visits Cambodia to meet the prime minister and attend the East Asia Summit

(CNN) -- Barack Obama met with Nobel Peace Prize winner Aung San Suu Kyi at her home in Myanmar on Monday, praising her "courage and determination" during a historic visit to the once repressive and secretive country.

The first sitting U.S. president to visit Myanmar, Obama urged its leaders, who have embarked on a series of far-reaching political and economic reforms since 2011, not to extinguish the "flickers of progress that we have seen."

Obama said that his visit to the lakeside villa where the pro-democracy icon spent years under house arrest marked a new chapter between the two countries.

"Here, through so many difficult years, is where she has displayed such unbreakable courage and determination," Obama told reporters, standing next to his fellow Nobel peace laureate. "It is here where she showed that human freedom and human dignity cannot be denied."

Myanmar Obama visit

The country, which is also known as Burma, was ruled by military leaders until early 2011 and for decades was politically and economically cut off from the rest of the world.

Suu Kyi acknowledged that Myanmar's opening up would be difficult.

The New York Times

YANGON, Myanmar — President Obama journeyed to this storied tropical outpost of pagodas and jungles on Monday to "extend the hand of friendship" as a land long tormented by repression and poverty begins to throw off military rule and emerge from decades of isolation.

The visit was intended to show support for the reforms put in place by Thein Sein's government since the end of military rule in November 2010.

Activists have warned that the visit may be too hasty - political prisoners remain behind bars and ethnic conflicts in border areas are unresolved.

BBC NEWS

Compressive Summarization as Global Optimization

- Indicator variables for every word of the nth sentence, $\mathbf{z}_n := \langle z_{n,\ell} \rangle_{\ell=1}^{L_n}$

Compressive Summarization as Global Optimization

- Indicator variables for every word of the *n*th sentence, $\mathbf{z}_n := \langle z_{n,\ell} \rangle_{\ell=1}^{L_n}$
- Summary length must not exceed the **budget** (B words)
- Quality function rewards *global* **informativeness** (through $g(\mathbf{z})$)...
- ... but also *local* **grammaticality** (through $h_n(\mathbf{z}_n)$):

$$\begin{aligned} \text{maximize} \quad & g(\mathbf{z}) + \sum_{n=1}^{N} h_n(\mathbf{z}_n) \\ \text{s.t.} \quad & \sum_{n=1}^{N} \sum_{\ell=1}^{L_n} z_{n,\ell} \leq B. \end{aligned}$$

Grammaticality: Sentence Compression Model

Inspired by Knight and Marcu (2000)'s word deletion model

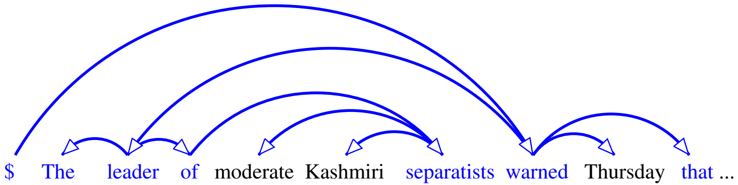
Other relevant work: McDonald (2006); Clarke and Lapata (2008); Martins and Smith (2009); Berg-Kirkpatrick et al. (2011)

Grammaticality: Sentence Compression Model

Inspired by Knight and Marcu (2000)'s word deletion model

Other relevant work: McDonald (2006); Clarke and Lapata (2008); Martins and Smith (2009); Berg-Kirkpatrick et al. (2011)

Our model factors over dependency arcs:

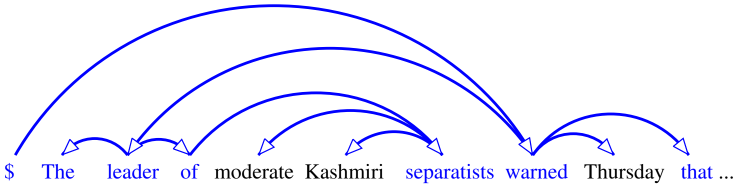


Grammaticality: Sentence Compression Model

Inspired by Knight and Marcu (2000)'s word deletion model

Other relevant work: McDonald (2006); Clarke and Lapata (2008); Martins and Smith (2009); Berg-Kirkpatrick et al. (2011)

Our model factors over dependency arcs:



A structured factor, locally decodable with dynamic programming.

Informativeness: Coverage Model

Inspired by *extractive* max-coverage models (Filatova and Hatzivassiloglou, 2004; Yih et al., 2007; Gillick et al., 2008; Lin and Bilmes, 2010)

Informativeness: Coverage Model

Inspired by *extractive* max-coverage models (Filatova and Hatzivassiloglou, 2004; Yih et al., 2007; Gillick et al., 2008; Lin and Bilmes, 2010)

- Extract a list of **concept types** $\langle c_m \rangle_{m=1}^{M}$ from the original documents
- Define concept **relevance scores** $\langle \sigma_m \rangle_{m=1}^{M}$

Informativeness: Coverage Model

Inspired by *extractive* max-coverage models (Filatova and Hatzivassiloglou, 2004; Yih et al., 2007; Gillick et al., 2008; Lin and Bilmes, 2010)

- Extract a list of **concept types** $\langle c_m \rangle_{m=1}^{M}$ from the original documents
- Define concept **relevance scores** $\langle \sigma_m \rangle_{m=1}^{M}$
- Define $g(\mathbf{z})$ as the global concept coverage:

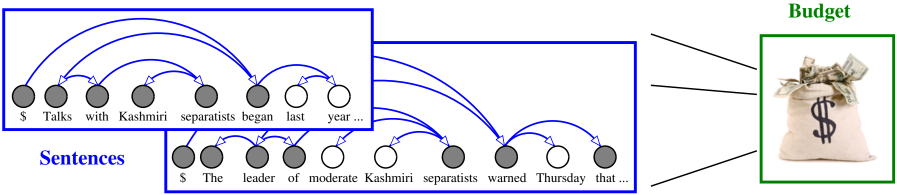
$$g(\mathbf{z}) := \sum_{m=1}^{M} \sigma_m u_m, \quad \text{where } u_m := \bigvee_{\langle n, \ell_s, \ell_e \rangle \in \mathcal{T}_m} \left(\bigwedge_{\ell=\ell_s}^{\ell_e} z_{n,\ell} \right)$$

- Intuitively: a **concept type** occurs if some **concept token** occurs.

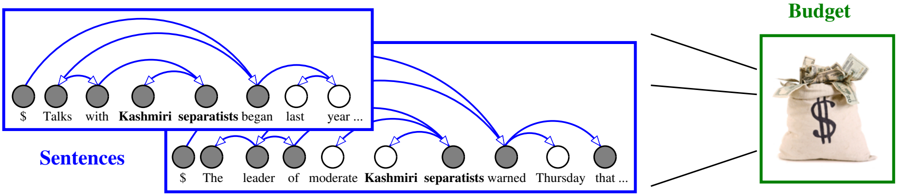
Graphical Model for Our Compressive Summarizer

Budget

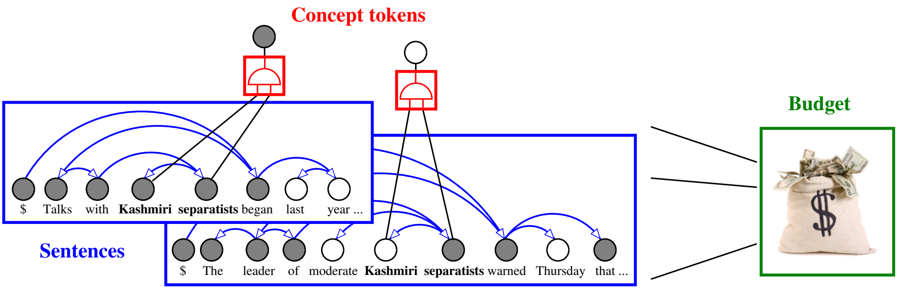
Graphical Model for Our Compressive Summarizer



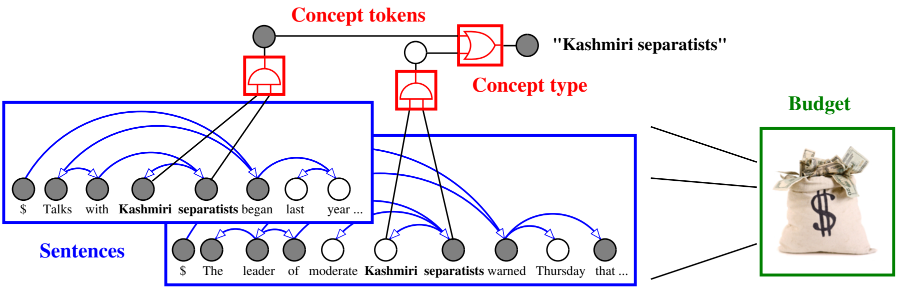
Graphical Model for Our Compressive Summarizer



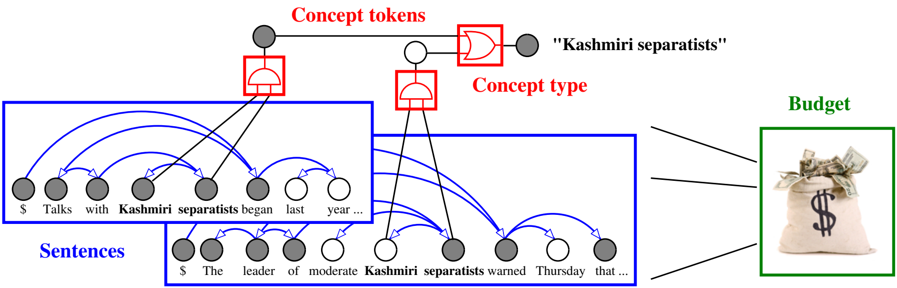
Graphical Model for Our Compressive Summarizer



Graphical Model for Our Compressive Summarizer

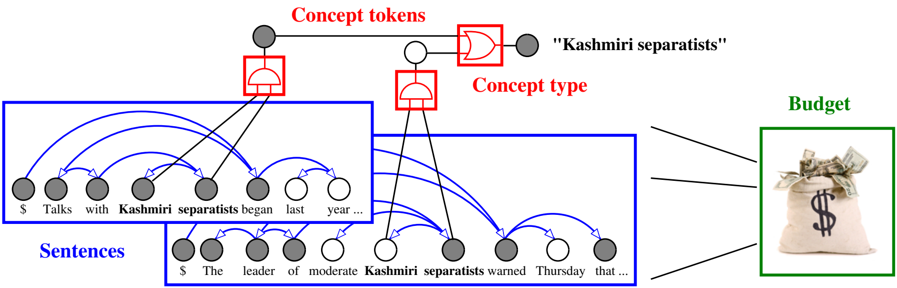


Graphical Model for Our Compressive Summarizer



1. We use dual decomposition (AD^3) for solving a linear relaxation
2. We apply a fast rounding procedure to obtain a valid summary

Graphical Model for Our Compressive Summarizer



1. We use dual decomposition (AD^3) for solving a linear relaxation
2. We apply a fast rounding procedure to obtain a valid summary

Multi-task learning: user-generated data (Simple English Wikipedia) along with manual abstracts and compressive summaries

Results on TAC-2008 Dataset

- Quality scores:

	ROUGE-2	Pyramid	LQ (1–5)
ICSI-1 (Gillick et al., 2008)	11.03	–	–
Berg-Kirkpatrick et al. (2011)	11.71	–	–
Woodsend and Lapata (2012)	11.37	–	–
Extractive, ILP	11.16	36.0	**4.6**
Single-task, AD^3	11.88	41.0	3.8
Multi-task, AD^3	**12.30**	**42.6**	4.2

Results on TAC-2008 Dataset

- Quality scores:

	ROUGE-2	Pyramid	LQ (1–5)
ICSI-1 (Gillick et al., 2008)	11.03	–	–
Berg-Kirkpatrick et al. (2011)	11.71	–	–
Woodsend and Lapata (2012)	11.37	–	–
Extractive, ILP	11.16	36.0	**4.6**
Single-task, AD^3	11.88	41.0	3.8
Multi-task, AD^3	**12.30**	**42.6**	4.2

- Averaged runtimes per summarization problem (10 documents):

Solver	Runtime (sec.)	ROUGE-2
ILP Exact, GLPK	10.394	12.40
LP-Relax., GLPK	2.265	12.38
AD^3 (1,000 its.)	**0.406**	**12.30**
Extractive (ILP)	0.265	11.16

Outline

Conclusions

- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)

Conclusions

- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **$\mathbf{AD}^3$** is a new decoder based on dual decomposition and ADMM

Conclusions

- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **AD^3** is a new decoder based on dual decomposition and ADMM
- Faster than the subgradient algorithm both in theory and in practice

Conclusions

- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **AD^3** is a new decoder based on dual decomposition and ADMM
- Faster than the subgradient algorithm both in theory and in practice
- AD^3 subproblems can be solved in linear time for **logic**, **budget**, and **knapsack constraints**

Conclusions

- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **AD^3** is a new decoder based on dual decomposition and ADMM
- Faster than the subgradient algorithm both in theory and in practice
- AD^3 subproblems can be solved in linear time for **logic**, **budget**, and **knapsack constraints**
- An **active set method** to tackle structured factors assuming only a local decoder

Conclusions

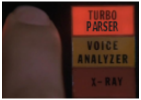
- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **AD^3** is a new decoder based on dual decomposition and ADMM
- Faster than the subgradient algorithm both in theory and in practice
- AD^3 subproblems can be solved in linear time for **logic**, **budget**, and **knapsack constraints**
- An **active set method** to tackle structured factors assuming only a local decoder
- Two applications with SOTA results: **turbo parsing** and **compressive summarization**

Conclusions

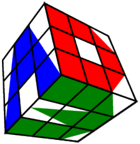
- Many structured problems in NLP are NP-hard or expensive (constrained models, diversity, combination of structured models)
- **AD^3** is a new decoder based on dual decomposition and ADMM
- Faster than the subgradient algorithm both in theory and in practice
- AD^3 subproblems can be solved in linear time for **logic**, **budget**, and **knapsack constraints**
- An **active set method** to tackle structured factors assuming only a local decoder
- Two applications with SOTA results: **turbo parsing** and **compressive summarization**
- **Could be a great fit to many other applications!!**

Thank you!

The parser and AD^3 are freely available at:



http://www.ark.cs.cmu.edu/TurboParser
http://www.ark.cs.cmu.edu/AD3



ARK lti Carnegie Mellon

Acknowledgments

- Fundação para a Ciência e Tecnologia, grant PTDC/EEI-SII/2312/2012.
- Priberam: QREN/POR Lisboa (Portugal), EU/FEDER programme, Discooperio project, contract 2011/18501.
- Priberam: QREN/POR Lisboa (Portugal), EU/FEDER programme, Intelligo project, contract 2012/24803.

References I

Almeida, M. B. and Martins, A. F. T. (2013). Fast and robust compressive summarization with dual decomposition and multi-task learning. In *Proc. of the Annual Meeting of the Association for Computational Linguistics*.

Berg-Kirkpatrick, T., Gillick, D., and Klein, D. (2011). Jointly learning to extract and compress. In *Proc. of Annual Meeting of the Association for Computational Linguistics*.

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.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Now Publishers.

Clarke, J. and Lapata, M. (2008). Global Inference for Sentence Compression An Integer Linear Programming Approach. *Journal of Artificial Intelligence Research*, 31:399–429.

Das, D., Martins, A. F. T., and Smith, N. A. (2012). An Exact Dual Decomposition Algorithm for Shallow Semantic Parsing with Constraints. In *Proc. of First Joint Conference on Lexical and Computational Semantics (*SEM)*.

Eckstein, J. and Bertsekas, D. (1992). On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318.

Eisner, J. M. (1996). Three new probabilistic models for dependency parsing: An exploration. In *Proc. of International Conference on Computational Linguistics*, pages 340–345.

Filatova, E. and Hatzivassiloglou, V. (2004). A formal model for information selection in multi-sentence text extraction. In *Proc. of International Conference on Computational Linguistics*.

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.

Gillick, D., Favre, B., and Hakkani-Tur, D. (2008). The icsi summarization system at tac 2008. In *Proc. of Text Understanding Conference*.

Glowinski, R. and Le Tallec, P. (1989). *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*. Society for Industrial Mathematics.

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.

References II

Knight, K. and Marcu, D. (2000). Statistics-based summarization—step one: Sentence compression. In *AAAI/IAAI*.

Komodakis, N., Paragios, N., and Tziritas, G. (2007). MRF optimization via dual decomposition: Message-passing revisited. In *Proc. of International Conference on Computer Vision*.

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.

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*.

Lin, H. and Bilmes, J. (2010). Multi-document summarization via budgeted maximization of submodular functions. In *Proc. of Annual Meeting of the North American chapter of the Association for Computational Linguistics*.

Martins, A. F. T., Figueiredo, M. A. T., Aguiar, P. M. Q., Smith, N. A., and Xing, E. P. (2011a). An Augmented Lagrangian Approach to Constrained MAP Inference. In *Proc. of International Conference on Machine Learning*.

Martins, A. F. T., Figueiredo, M. A. T., Aguiar, P. M. Q., Smith, N. A., and Xing, E. P. (2012). Alternating directions dual decomposition. Arxiv preprint arXiv:1212.6550.

Martins, A. F. T. and Smith, N. A. (2009). Summarization with a Joint Model for Sentence Extraction and Compression. In *North American Chapter of the Association for Computational Linguistics: Workshop on Integer Linear Programming for NLP*.

Martins, A. F. T., Smith, N. A., Aguiar, P. M. Q., and Figueiredo, M. A. T. (2011b). Dual decomposition with many overlapping components. In *Proc. of Empirical Methods for Natural Language Processing*.

Martins, A. F. T., Smith, N. A., Xing, E. P., Aguiar, P. M. Q., and Figueiredo, M. A. T. (2010a). Augmented Dual Decomposition for MAP Inference. In *Neural Information Processing Systems: Workshop in Optimization for Machine Learning*.

Martins, A. F. T., Smith, N. A., Xing, E. P., Figueiredo, M. A. T., and Aguiar, P. M. Q. (2010b). Turbo parsers: Dependency parsing by approximate variational inference. In *Proc. of Empirical Methods for Natural Language Processing*.

McDonald, R. (2006). Discriminative sentence compression with soft syntactic constraints. In *Proc. of Annual Meeting of the European Chapter of the Association for Computational Linguistics*.

References III

McDonald, R. and Satta, G. (2007). On the complexity of non-projective data-driven dependency parsing. In *Proc. of International Conference on Parsing Technologies*.

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*.

Nivre, J., Hall, J., Nilsson, J., Eryiğit, G., and Marinov, S. (2006). Labeled pseudo-projective dependency parsing with support vector machines. In *Procs. of International Conference on Natural Language Learning*.

Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer-Verlag.

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*.

Rush, A., Sontag, D., Collins, M., and Jaakkola, T. (2010). On dual decomposition and linear programming relaxations for natural language processing. In *Proc. of Empirical Methods for Natural Language Processing*.

Smith, D. and Eisner, J. (2008). Dependency parsing by belief propagation. In *Proc. of Empirical Methods for Natural Language Processing*.

Woodsend, K. and Lapata, M. (2012). Multiple aspect summarization using integer linear programming. In *Proc. of Empirical Methods in Natural Language Processing*.

Yih, W.-t., Goodman, J., Vanderwende, L., and Suzuki, H. (2007). Multi-document summarization by maximizing informative content-words. In *Proc. of International Joint Conference on Artifical Intelligence*.