Vine Pruning
for Efficient Multi-Pass Dependency Parsing

Alexander M. Rush and Slav Petrov
As McGwire neared, fans went wild
Styles of Dependency Parsing

- Speed
  - Greedy: $O(n)$
  - K-best: $O(kn)$

- Accuracy
  - First-order: $O(n^3)$
  - Second-order: $O(n^3)$
  - Third-order: $O(n^4)$

- Transition-based parsers
  - (Nivre, 2004)

- Graph-based parsers
  - (Eisner, 2000), (McDonald, 2005)

This work
Styles of Dependency Parsing

- Greedy parsing: $O(n)$
- K-best parsing: $O(kn)$
- First-order parsing: $O(n^3)$
- Second-order parsing: $O(n^3)$
- Third-order parsing: $O(n^4)$

Transition-based parsers (Nivre, 2004)
Graph-based parsers (Eisner, 2000), (McDonald, 2005)

This work

Speed vs. Accuracy plot:
- Red square: Transition-based parsers
- Blue square: Graph-based parsers
- Red square: This work
As McGwire neared, fans went wild.
linear-size dependency representation
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
* As McGwire neared, fans went wild
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
* As McGwire neared, fans went wild
First-Order Feature Calculation

As McGwire neared, fans went wild
As McGwire neared, fans went wild.
Arc Length By Part-of-Speech

![Graph showing the arc length by part-of-speech with lines for different parts of speech: NOUN, ADP, DET, VERB, and ADJ. The x-axis represents length, and the y-axis represents counts. Each line shows a decreasing trend as length increases.]
Arc Length By Part-of-Speech
Arc Length By Part-of-Speech

![Graph showing arc length by part-of-speech. The x-axis represents length, and the y-axis represents counts. Different colors and lines represent different parts of speech: NOUN (dashed blue), ADP (dotted green), DET (dashed red), VERB (solid cyan), and ADJ (dotted magenta).]
The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.

But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian, " he said.

The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility. "

It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC.

We would have to wait until we have collected on those assets before we can move forward, " he said.

This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

The complicated language in the huge new law has muddied the fight.
This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

"That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian," he said.

"The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility," he said.

"It's a problem that clearly has to be resolved," said David Cooke, executive director of the RTC.

"We would have to wait until we have collected on those assets before we can move forward," he said.

The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.

The complicated language in the huge new law has muddied the fight.
This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian, " he said.

The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility.

It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC.

We would have to wait until we have collected on those assets before we can move forward, " he said.

The complicated language in the huge new law has muddied the fight.
It 's a problem that clearly has to be resolved, ' said David Cooke, executive director of the RTC.

The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility.

We would have to wait until we have collected on those assets before we can move forward, ' he said.

This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.

The complicated language in the huge new law has muddied the fight.

But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.
We would have to wait until we have collected on those assets before we can move forward, he said. "We would have to wait until we have collected on those assets before we can move forward," he said.

The RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian, " he said. "That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian, " he said.

The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility. "The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility."

It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC. "It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC.

This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit. "This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit."

The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization. "The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization."

The complicated language in the huge new law has muddied the fight. "The complicated language in the huge new law has muddied the fight."

The executive consulted separately the director. "The executive consulted separately the director."
But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

"The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility."

"It's a problem that clearly has to be resolved," said David Cooke, executive director of the RTC.

"We would have to wait until we have collected on those assets before we can move forward," he said.

This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.
But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately. The RTC is going to have to pay a price of prior consultation on the Hill if they want that kind of flexibility. " It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC. " We would have to wait until we have collected on those assets before we can move forward," he said. The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.
But the RTC also requires "working" capital to maintain the bad assets of thrifts that are sold, until the assets can be sold separately.

That secrecy leads to a proposal like the one from Ways and Means, which seems to me sort of draconian, " he said.

It's a problem that clearly has to be resolved, " said David Cooke, executive director of the RTC.

We would have to wait until we have collected on those assets before we can move forward, " he said.

This financing system was created in the new law in order to keep the bailout spending from swelling the budget deficit.

The bill intends to restrict the RTC to Treasury borrowings only, unless the agency receives specific congressional authorization.

The complicated language in the huge new law has muddied the fight.
Arc Length Heat Map
Arc Length Heat Map
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.

Banded Matrix
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
dynamic programs for parsing
Inference Questions

questions:

• How do we reduce inference time to $O(n)$?

• How do we decide which arcs to prune?

Vine Parsing (Eisner and Smith, 2005)
Eisner First-Order Rules

\[ h \rightarrow h + m + 1 \]
\[ h \leftarrow h + m + e \]
First-Order Parsing

As McGwire neared, fans went wild.
First-Order Parsing

* As McGwire neared, fans went wild
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
First-Order Parsing

As McGwire neared, fans went wild.
First-Order Parsing

As McGwire neared, fans went wild
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
Vine Parsing Rules

0 e ← 0 e −1 + e −1 e

0 e ← 0 m + e m

0 e ← 0 e −1 + e −1 e

0 e ← 0 m + m e

0 e ← 0 e

0 e ← 0 m + m e

0 e ← 0 e −1 + e −1 e
As McGwire neared, fans went wild.
Vine Parsing

As McGwire neared, fans went wild
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
As McGwire neared, fans went wild.
Arc Pruning

• Prune arcs based on max-marginals.

\[ \text{maxmarginal}(a) = \max_{y:a \in y} (y \cdot w) \]

• Can compute using inside-outside algorithm.

• Generic algorithm using hypergraph parsing.
As McGwire neared, fans went wild
Max-Marginals for Outer Arcs

\[
\text{maxmarginal}(\text{LEFT } \rightarrow \text{ fans}) > \text{threshold}?
\]

* As McGwire neared, fans went wild
pruning and training
Max-Marginal Pruning

**goal:** Define a threshold on max-marginal score.

- Validation parameter $\alpha$ trades off between speed and accuracy.
  
  $$t_{\alpha}(w) = \alpha \max_y (y \cdot w) + (1 - \alpha) \frac{1}{|A|} \sum_{a \in A} \text{maxmarginal}(a, w)$$

- Highest scoring parse upper bounds any max-marginal.

- Assume average of max-marginals is lower than gold.
Pruning Threshold

feature two

max

average

max-marginal

feature one

w

max

average

max-marginal

α
Pruning Threshold

feature one
feature two

max
average
max-marginal

α

feature two

feature one

W
Pruning Threshold

feature one
feature two
w
max
average
max-marginal
α
Pruning Threshold

feature one

feature two

max

average

max-marginal

$\alpha$
Pruning Threshold

feature one

feature two

w

max

average

max-marginal

α
Pruning Threshold

- Feature one
- Feature two

- Max
- Average
- Max-marginal

\(\alpha\)
Pruning Threshold

feature one

feature two

w

max

average

max-marginal

α
Pruning Threshold

feature one
feature two

$w$

max

average

max-marginal

$\alpha$
Pruning Threshold

feature one

feature two

max

average

max-marginal

α
Pruning Threshold
Pruning Threshold

feature one
feature two
w
max
average
max-marginal

max

average
max-marginal

α

feature one
feature two
Structured Cascade Training (Weiss and Taskar, 2011)

- Train a linear model with a loss function for pruning.
- Regularized risk minimization with loss based on threshold

\[
\min_w \lambda \|w\|^2 + \frac{1}{P} \sum_{p=1}^{P} [1 - y^{(p)} \cdot w + t^{(p)}_{\alpha}(w)]_+
\]

- Can use a simple variant of perceptron/pegasos to train.
Structured Cascade Training

![Diagram](image-url)
Structured Cascade Training

- Feature one
- Feature two
- \( w \)
- Max
- Gold
- Average
- Max-marginal

\( \text{feature one} \)
\( \text{feature two} \)
\( w \)
\( \text{max} \)
\( \text{gold} \)
\( \text{average} \)
\( \text{max-marginal} \)

\( \text{feature one} \)
\( \text{feature two} \)
\( w \)
\( \text{max} \)
\( \text{gold} \)
\( \text{average} \)
\( \text{max-marginal} \)

\( \text{feature one} \)
\( \text{feature two} \)
\( w \)
\( \text{max} \)
\( \text{gold} \)
\( \text{average} \)
\( \text{max-marginal} \)
Structured Cascade Training

![Graph showing feature one vs. feature two with points labeled as max, average, and gold. The graph includes a dashed line and a vector labeled W.]
Structured Cascade Training

feature one
feature two
w
max
gold
average
max-marginal

W

feature one

feature two
Structured Cascade Training
Structured Cascade Training
Structured Cascade Training

feature one
feature two
max
average
max-marginal

feature one
feature two
max

gold
Structured Cascade Training
Structured Cascade Training

feature one
feature two
w
max
gold
average
max-marginal

feature one
feature two
w
gold
max
gold

feature one
feature two
Structured Cascade Training

- feature one
- feature two

- max
- gold
- average
- max-marginal
Structured Cascade Training

Structured Cascade Training
experiments
Implementation

Inference

• Experiments use a highly-optimized C++ implementation.

• Baseline first-order parser processes 2000 tokens/sec.

• Hypergraph parsing framework with shared inference.

Model

• Final models trained with hamming-loss MIRA.

• Full collection of dependency parsing features (Koo, 2010).

• First-, second-, and third-order models match state-of-the-art.
## Baselines

| Baseline          | Description                                                                                     |
|-------------------|-------------------------------------------------------------------------------------------------|
| NoPrune           | exhaustive parsing model with no pruning                                                       |
| LocalShort        | unstructured classifier over $O(n)$ short arcs (Bergsma and Cherry, 2010)                      |
| Local             | unstructured classifier over $O(n^2)$ arcs (Bergsma and Cherry, 2010)                          |
| FirstOnly         | structured first-order model in cascade (Koo, 2010)                                             |
| VinePosterior     | posterior pruning cascade trained with L-BFGS                                                  |
| ZhangNivre        | reimplementaton of state-of-the-art, k-best, transition-based parser (Zhang and Nivre, 2011). |
Speed/Accuracy Experiments: First-Order Parsing

Accuracy

Accuracy

Relative Speed

Relative Speed

NoPrune

Local

FirstOnly

VinePosterior

VineCascade

ZhangNivre(8)

ZhangNivre(16)

ZhangNivre(64)
Speed/Accuracy Experiments: Second-Order Parsing

![Graph 1: Relative Speed vs. Algorithms]

- NoPrune
- Local
- FirstOnly
- VinePosterior
- VineCascade
- ZhangNivre(16)

![Graph 2: Accuracy vs. Algorithms]

- NoPrune
- Local
- FirstOnly
- VinePosterior
- VineCascade
- ZhangNivre(16)
Empirical Complexity: First-Order Parsing

![Diagram showing sentence length vs time for NoPrune and VineCascade]

- NoPrune: [2.8] time units
- VineCascade: [1.4] time units
Empirical Complexity: Second-Order Parsing

![Graph showing sentence length vs time with two lines representing NoPrune and VineCascade]

- NoPrune [2.8]
- VineCascade [1.8]
Empirical Complexity: Third-Order Parsing

![Graph showing time versus sentence length for NoPrune and VineCascade methods. The graph indicates that NoPrune has a higher time complexity compared to VineCascade.]

- **NoPrune**: [3.8]
- **VineCascade**: [1.9]
Multilingual Experiments: First-Order Parsing

![Graph showing relative speed of different languages](image)

- **En**: Relative Speed
- **Bg**: Relative Speed
- **De**: Relative Speed
- **Pt**: Relative Speed
- **Sw**: Relative Speed
- **Zh**: Relative Speed

Legend:
- **NoPrune**
- **VineCascade**
Multilingual Experiments: Second-Order Parsing

Relative Speed

- En
- Bg
- De
- Pt
- Sw
- Zh

Legend:
- NoPrune
- VineCascade
Multilingual Experiments: Third-Order Parsing

![Bar chart showing relative speed for different languages and models: En, Bg, De, Pt, Sw, Zh (NoPrune, VineCascade).]
Special thanks to:
Ryan McDonald, Hao Zhang, Michael Ringgaard, Terry Koo, Keith Hall, Kuzman Ganchev, Yoav Goldberg, Andre Martins, and the rest of the Google NLP team