Statistical Models for Frame-Semantic Parsing

Dipanjan Das*
Google

Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore
June 27, 2014

*Thanks to Desai Chen, Kuzman Ganchev, Karl Moritz Hermann, André Martins, Nathan Schneider, Noah Smith and Jason Weston
I want to travel to Baltimore on Sunday
I want to travel to Baltimore on Sunday.
Frame-Semantic Parsing

I want to \textit{travel} to Baltimore on Sunday.

Participant or role for the frame
I want to travel to Baltimore on Sunday.

Frame-Semantic Parsing

Traveler

Desiring

Event

Time

Goal

Participant or role for the frame
Outline

• Why should we build statistical models for frame semantics?

• Overview of statistical methods

• Future directions
Outline

• Why should we build statistical models for frame semantics?

• Overview of statistical methods

• Future directions
Motivation
Motivation

Deeper understanding beyond syntax
Motivation

Deeper understanding beyond syntax

Grouping of predicates and arguments into clusters
Motivation

Deeper understanding beyond syntax

Grouping of predicates and arguments into clusters

Grounding of natural language to an ontology
Motivation

Bengal’s massive stock of food was reduced to nothing
Motivation

Bengal’s massive stock of food was reduced to nothing
Bengal’s massive stock of food was reduced to nothing.

Store or financial entity?
Bengal’s massive **stock** of food was reduced to nothing.
Store of what?
Of what size?
Whose store?

Motivation

Bengal’s massive stock of food was reduced to nothing.
Motivation

Bengal’s massive stock of food was reduced to nothing

What was reduced? To what?
Bengal's massive stock of food was reduced to nothing.

Motivation

What was reduced? To what?
Bengal’s massive stock of food was reduced to nothing.
Motivation
Motivation

BEHIND_THE_SCENES

Artist

Production

James Cameron directed Titanic
Motivation

James Cameron **directed** *Titanic*

**BEHIND_THE_SCENES**

**Artist**

**Production**

James Cameron **directed** *Titanic*

**BEHIND_THE_SCENES**

**Artist**

**Production**

James Cameron is *Titanic’s* **director**
Motivation

James Cameron directed *Titanic*

**BEHIND_THE_SCENES**

**Artist**

**Production**

Grouping across syntactic categories

James Cameron is *directed* Titanic’s *director*
Motivation

James Cameron **filmed** Titanic

Grouping across different lemmas

James Cameron is Titanic's **director**
Motivation

James Cameron filmed Titanic.

James Cameron is Titanic's director.

from PropBank

from NomBank
Motivation

James Cameron filmed Titanic. James Cameron is Titanic's director. Hard to align different lemmas across resources.
Motivation

James Cameron filmed Titanic.

James Cameron is Titanic's director.
Motivation

James Cameron filmed Titanic. 

James Cameron is Titanic's director.

Argument labels do not match up.
Motivation

B_{E}H_{I}N_{D}\_T_{H}_{E}\_S_{C}_{E}_{N}_{E}_{S}

Artist

James Cameron \textcolor{red}{\textit{filmed}}

Production

Titanic \quad \text{in 1997}
Motivation

BEHIND_THE_SCENE

Artist

James Cameron filmed Titanic in 1997

λe.filmed.arg1(e, James_Cameron)∧filmed.arg2(e, Titanic)∧filmed.in(e, 1997)
Motivation

BEHIND_THE_SCENES

Artist

James Cameron filmed

Production

Titanic

in 1997

\( \lambda e.\text{filmed.arg1}(e, \text{JamesCAMERON}) \land \text{filmed.arg2}(e, \text{Titanic}) \land \text{filmed.in}(e, 1997) \)

Freebase

\( \lambda e.\text{filmed.arg1}(e, /m/03\_gd) \land \text{filmed.arg2}(e, /m/0dr\_4) \land \text{filmed.in}(e, 1997) \)
Motivation

\[ \lambda e. \text{filmed.arg1}(e, \text{James\_Cameron}) \land \text{filmed.arg2}(e, \text{Titanic}) \land \text{filmed.in}(e, 1997) \]
Motivation

\[ \lambda e. \text{filmed.arg1}(e, \text{James_Cameron}) \land \text{filmed.arg2}(e, \text{Titanic}) \land \text{filmed.in}(e, 1997) \]

FrameNet

\[ \lambda e. \text{BEHIND_THE_SCENES.Artist}(e, /m/03_gd) \land \text{BEHIND_THE_SCENES.Production}(e, /m/0dr_4) \land \text{BEHIND_THE_SCENES.Time}(e, 1997) \]
Applications of Frame-Semantic Parsing
Stance Classification

• Frame Semantics for Stance Classification
  Hasan and Ng (CoNLL 2013)

• Two sided debates in an online forum

• Classification of stance

• Improvement over a baseline that uses bag of words and dependencies
Dialog Systems

  Chen, Wang and Rudnicky (ASRU 2013)
- Annotation of dialog transcripts with frame-semantic structures
- Uses only a subset of frames
- Uses these annotations for slot induction
Stock Price Movement

- Semantic Frames to Predict Stock Price Movement
  Xie et al. (ACL 2013)
- Predict the change in stock price from financial news
- Lots of features along with features based on frames and roles
- Shows improvements over other features
Summarization

• Generating Automated Meeting Summaries
  Thomas Kleinbauer (PhD thesis, Saarland University)

• Part of a large system for generating meeting summaries
Outline

• Why should we build statistical models for frame semantics?

• Overview of statistical methods

• Future directions
Structure of Lexicon and Data

PLACING
Agent
Cause
Goal
Theme
Area
Time
Structure of Lexicon and Data

PLACING

Agent
Cause
Goal
Theme
Area
Time

frame
Structure of Lexicon and Data

.roles

.frame

PLACING

Agent

Cause

Goal

Theme

Area

Time
Structure of Lexicon and Data

- **core roles**: Agent, Cause, Goal, Theme
- **non-core roles**: Area, Time

**PLACING**

**frame**
Structure of Lexicon and Data

- **core roles**
  - Agent
  - Cause
  - Goal
  - Theme

- **non-core roles**
  - Area
  - Time

---

frame
excludes
relationship
Structure of Lexicon and Data

- **Core roles**
  - Agent
  - Cause
  - Goal
  - Theme

- **Non-core roles**
  - Area
  - Time

- **Frame excludes relationship**

- **Predicates**
  - archive.V,
  - arrange.V, bag.V,
  - bestow.V, bin.V
Structure of Lexicon and Data

**TRANSITIVE_ACTION**
- Agent
- Cause
- Patient
- Event
- Place
- Time

**PLACING**
- Agent
- Cause
- Goal
- Theme
- Area
- Time

**DISPERsal**
- Agent
- Cause
- Individuals
- Distance
- Time

**INSTALLING**
- Agent
- Component
- Fixed_location
- Area
- Time

**STORING**
- Agent
- Location
- Theme
- Area
- Time

**STORE**
- Possessor
- Resource
- Supply
- Descriptor
Structure of Lexicon and Data

**TRANSITIVE_ACTION**
- Agent
- Cause
- Patient
- Event
- Place
- Time

**INSTALLED**
- Agent
- Component
- Fixed_location
- Area
- Time

**PLACING**
- Agent
- Cause
- Goal
- Theme
- Area
- Time

**STORE**
- Possessor
- Resource
- Supply
- Descriptor

**DISPERSAL**
- Agent
- Cause
- Individuals
- Distance
- Time

**STORING**
- Agent
- Location
- Theme
- Area
- Time
Structure of Lexicon and Data

TRANSITIVE_ACTION
- Agent
- Cause
- Patient
- Event
- Place
- Time

PLACING
- Agent
- Cause
- Goal
- Theme
- Area
- Time

DISPERsal
- Agent
- Cause
- Individuals
- Distance
- Time

INSTALLING
- Agent
- Component
- Fixed_location
- Area
- Time

STORING
- Agent
- Location
- Theme
- Area
- Time

STORE
- Possessor
- Resource
- Supply
- Descriptor

inheritance

used by
Datasets

**Benchmark Dataset**
*(SemEval 2007)*

- 665 frames
- 720 role labels
- 8.4K unique predicate types

**Training set:**
- 2.2K sentences
- 11.2K predicate tokens

**Test set:**
- 120 sentences
- 1.1K predicate tokens
LTH Frame-Semantic Parser

Johansson and Nugues (2007)

Frame Identification  Argument Filtering  Argument Labeling
LTH Frame-Semantic Parser

Johansson and Nugues (2007)

Frame Identification  Argument Filtering  Argument Labeling
Bengal’s massive *stock* of food was reduced to nothing.
Bengal’s massive *stock* of food was reduced to nothing

Find the best among all frames
Bengal’s massive stock of food was reduced to nothing.

$$\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score} (\text{frame, predicate, sentence})$$

Taken from an SVM classifier trained on ambiguous predicates.
To increase coverage, potential predicates were extracted from WordNet and automatically frame labels were selected for them.
LTH Frame-Semantic Parser
Johansson and Nugues (2007)

Frame Identification Accuracy

F-Measure

LTH

50.0
61.3
72.5
83.8
95.0

57.3


LTH Frame-Semantic Parser

Johansson and Nugues (2007)

Frame Identification

Argument Filtering

Argument Labeling
Argument Filtering

Bengal’s massive stock of food was reduced to nothing
Argument Filtering

Bengal’s massive stock of food was reduced to nothing

Potential Arguments

- Bengal’s
- Bengal
- massive stock
- of food
- food
- massive
- Bengal’s massive
- to nothing
- to
- of
- ’s
Bengal’s massive stock of food was reduced to nothing.
LTH Frame-Semantic Parser

Johansson and Nugues (2007)

- Frame Identification
- Argument Filtering
- Argument Labeling
Argument Filtering

Bengal’s massive stock of food was reduced to nothing

Multiclass SVM Classification
Argument Filtering

Bengal's massive stock of food was reduced to nothing.

Potential Arguments
Multiclass SVM Classification

Bengal's massive stock of food
food massive to nothing
Bengal’s massive stock of food was reduced to nothing.
LTH Frame-Semantic Parser

Johansson and Nugues (2007)

Full Frame-Semantic Structure Prediction
SEMAFOR
Das, Chen, Martins, Schneider, Smith (2014)

Frame Identification

Argument Identification
SEMAFOR
Das, Chen, Martins, Schneider, Smith (2014)

Frame Identification

Argument Identification
Bengal’s massive *stock* of food was reduced to nothing

$$\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score(frame, predicate, sentence)}$$
Bengal's massive **stock** of food was reduced to nothing

\[
\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \; \text{score}(\text{frame, predicate, sentence}) \\
= \log p(\text{frame} \mid \text{predicate, sentence})
\]
SEMAFOR: Frame Identification

\[
\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score}(\text{frame, predicate, sentence})
\]

\[
= \log p(\text{frame} | \text{predicate, sentence})
\]

Logistic regression with a latent variable
SEMAFOR: Frame Identification

$$\text{best frame} = \arg \max_{\text{all frames}} \text{score} \left( \text{frame}, \text{predicate}, \text{sentence} \right)$$

$$= \log p \left( \text{frame} \mid \text{predicate}, \text{sentence} \right)$$

Logistic regression with a latent variable

$$p \left( \text{frame} \mid \text{predicate}, \text{sentence} \right) = \frac{1}{\mathcal{Z}} \sum_{\text{proto-predicates evoking frame}} \exp \mathbf{w} \cdot \mathbf{f} \left( \text{frame}, \text{proto-predicate}, \text{sentence}, \text{lexical-semantic relations between predicate and proto-predicate} \right)$$
**SEMAFOR: Frame Identification**

\[
\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score(frame, predicate, sentence)} = \log p(\text{frame} \mid \text{predicate, sentence})
\]

\[
p(\text{frame} \mid \text{predicate, sentence}) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp \mathbf{w} \cdot \mathbf{f} \left( \text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate} \right)
\]

Predicates evoking a frame in supervised data, e.g.
evoke Store
SEMAFOR: Frame Identification

\[
p(frame \mid predicate, sentence) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp w \cdot f(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate})
\]

frame = STORE

predicate = stock.N

proto-predicate = stockpile.N

sentence = Bengal's massive stock of food was reduced to nothing
\[ p(\text{frame} \mid \text{predicate}, \text{sentence}) = \frac{1}{\mathcal{Z}} \sum_{\text{proto-predicates evoking frame}} \exp \mathbf{w} \cdot \mathbf{f}(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate}) \]

frame = \text{STORE}

predicate = \text{stock.N}

proto-predicate = \text{stockpile.N}

sentence = \text{Bengal's massive stock of food was reduced to nothing}
\[
p(frame \mid \text{predicate, sentence}) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp w \cdot f(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate})
\]

\[
\text{frame} = \text{STORE}
\]

\[
\text{predicate} = \text{stock.N}
\]

\[
\text{proto-predicate} = \text{stockpile.N}
\]

\[
\text{sentence} = \text{Bengal's massive stock of food was reduced to nothing}
\]

\[
f_{10245} = 1 \quad \text{If} \quad \text{frame} = \text{STORE} \wedge \text{proto-predicate} = \text{stockpile.N} \quad \text{synonym} \in \text{LexSem}
\]
SEMAFOR: Frame Identification

\[ p(\text{frame} \mid \text{predicate}, \text{sentence}) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp w \cdot f(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate}) \]

\[
\begin{align*}
\text{frame} &= \text{STORE} \\
\text{predicate} &= \text{stock}.N \\
\text{proto-predicate} &= \text{stockpile}.N \\
\text{sentence} &= \text{Bengal's massive stock of food was reduced to nothing}
\end{align*}
\]

LexSem = \{synonym\} (comes from WordNet!)

\[ f_{10245} = 1 \quad \text{if} \quad \begin{align*}
\text{frame} &= \text{STORE} \land \\
\text{proto-predicate} &= \text{stockpile}.N \\
synonym &\in \text{LexSem}
\end{align*} \]
Datasets

**Benchmark Dataset**
*(SemEval 2007)*

- 665 frames
- 720 role labels
- 8.4K unique predicate types

**Training set:**
- 2.2K sentences
- 11.2K predicate tokens

**Test set:**
- 120 sentences
- 1.1K predicate tokens

**New Data**
*(FrameNet 1.5, 2010)*

- 877 frames
- 1068 role labels
- 9.3K unique predicate types

**Training set:**
- 3.3K sentences
- 19.6K predicate tokens

**Test set:**
- 2420 sentences
- 4.5K predicate tokens
Datasets

**Benchmark Dataset**  
*(SemEval 2007)*

- 665 frames
- 720 role labels
- 8.4K unique predicate types

**Training set:**
- 2.2K sentences
- 11.2K predicate tokens

**Test set:**
- 120 sentences
- 1.1K predicate tokens

**New Data**  
*(FrameNet 1.5, 2010)*

- 877 frames
- 1068 role labels
- 9.3K unique predicate types

**Training set:**
- 3.3K sentences
- 19.6K predicate tokens

**Test set:**
- 2420 sentences
- 4.5K predicate tokens
SEMAFOR: Frame Identification

Results

Benchmark

New Data

F-Measure

95.0
83.8
72.5
61.3
50.0

LTH

SEMAFOR log-linear

57.3
61
SEM AF OR: Frame Identification

Results

Benchmark

- F-Measure

New Data

- auto predicates

LTH

- 57.3

SEMAFOR

- log-linear

- 61
Results

SEMAFOR: Frame Identification

Benchmark

<table>
<thead>
<tr>
<th></th>
<th>LTH</th>
<th>SEMAFOR log-linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>auto predicates</td>
<td>57.3</td>
<td>61.3</td>
</tr>
</tbody>
</table>

New Data

<table>
<thead>
<tr>
<th></th>
<th>SEMAFOR log-linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>gold predicates</td>
<td>83.0</td>
</tr>
</tbody>
</table>
SEMAFOR: Frame Identification

Frame Identification

Accuracy

All Predicates:
- 83.0

Unknown Predicates:
- 23.1
SEMAFOR: Handling Unknown Predicates

Knowledge of only 9,263 predicates in supervised data
Knowledge of only 9,263 predicates in supervised data

However, English has lot more potential predicates (~65,000 in newswire English)
Knowledge of only 9,263 predicates in supervised data

However, English has lot more potential predicates (~65,000 in newswire English)

Lexicon expansion using graph-based semi-supervised learning
How can label propagation help?

- Build a graph over potential predicates as vertices
  - compute similarity matrix using co-occurrence statistics
- Label distribution at each vertex
  \[ \approx \] distribution over frames that the predicate can evoke
Example Graph

Seed predicates:
- powerlessness
- deprivation
- destitution
- unemployment
- joblessness
- unemployment_rate
- poverty
- rich
- wealthy
- inequality
- homelessness
- resemblance
- difference
- divergence

Unseen predicates:
- complexity
- wealth
- destitution
- unemployment
- joblessness
- unemployment_rate
- poverty
- rich
- wealthy
- inequality
- homelessness
- resemblance
- difference
- divergence

Similarity:
- similarity
- variant
- resemblance
- difference
- divergence
Example Graph

Seed predicates:
- powerlessness
- deprivation
- destitution
- unemployment rate
- unemployment
- employment

Unseen predicates:
- rich
- wealthy
- poverty
- inequality
- resemblance
- similarity
- variant
- resemble
- disparity
- discrepancy
- divergence

Graph Propagation
Example Graph

Seed predicates
- powerlessness
- deprivation
- destitution
- unemployment
- rate
- UNEMPLOYMENT_RATE

Unseen predicates
- richness
- wealth
- poverty
- inequality
- homelessness
- joblessness
- disparity
- discrepancy

Graph Propagation
Example Graph

- Seed predicates
- Unseen predicates

Graph Propagation

Continues till convergence...
SEMAFOR: Unknown Predicates

Frame Identification

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>23.1</td>
</tr>
<tr>
<td>Self-Training</td>
<td>18.9</td>
</tr>
<tr>
<td>Graph-Based</td>
<td>42.7</td>
</tr>
</tbody>
</table>
SEMAFOR
Das, Chen, Martins, Schneider, Smith (2014)

Frame Identification

Argument Identification
Bengal’s massive stock of food was reduced to nothing.
Bengal’s massive stock of food was reduced to nothing.
Bengal's massive stock of food

Possessor: Bengal's
Resource: stock
Descriptor: massive
Use: stock
Supply: Ø
Bengal's massive stock of food
Bengal’s massive stock of food violates overlap constraints.
SEMAFOR: Argument Identification

Other types of structural constraints

Mutual exclusion constraint

Other types of structural constraints

If an **agent** places something, there cannot be a **cause** role in the sentence.

Mutual exclusion constraint

Other types of structural constraints

The waiter placed food on the table.

In Kabul, hauling water put food on the table.
Other types of structural constraints

**SEMAFOR: Argument Identification**

**SIMILARITY**
- Dimension
- Differentiating_fact
- Entity_1
- Entity_2
- Degree

Requires constraint

Other types of structural constraints

A mulberry resembles a loganberry.

Requires constraint

SIMILARITY

Dimension
Differentiating_fact
Entity_1
Entity_2
Degree

first entity

second entity
Other types of structural constraints

SEMMAFAR: Argument Identification

Similarity
- Dimension
- Differentiating_fact
- Entity_1
- Entity_2
- Degree

Requires constraint


A mulberry resembles.
Bengal’s massive stock of food

A constrained optimization problem

SEMAFOR: Argument Identification
Bengal's massive stock of food
SEMAFOR: Argument Identification

score(role ↔ span) = w · g(role, span, frame)

Possessor
- Stock

Resource
- Bengal's massive stock
- Food

Descriptor
- Massive

Use
- Bengal's massive massive stock

Supply
- Ø
A constrained optimization problem

\[ \mathcal{Z}_{\text{role} \leftrightarrow \text{span}} \]
A constrained optimization problem

\[ \exists_{\text{role} \leftrightarrow \text{span}} \]

a binary variable for each role, span tuple
A constrained optimization problem

$$z = \langle \tilde{z}_{\text{role} \leftrightarrow \text{span}} \rangle$$

a binary vector for all role, span tuples
SEMAFOR: Argument Identification

A constrained optimization problem

\[ \sum_{\text{roles, spans}} \bar{z}_{\text{role} \leftrightarrow \text{span}} \cdot score(\text{role} \leftrightarrow \text{span}) \]

subject to \( \sum_{\text{roles, spans}} \bar{z}_{\text{role} \leftrightarrow \text{span}} \leq 1 \)
A constrained optimization problem

\[ z = \langle z_{\text{role}\leftrightarrow\text{span}} \rangle \]

maximize \[ \sum_{\text{roles,spans}} z_{\text{role}\leftrightarrow\text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \]

w.r.t. \[ z \]

s.t \[ \forall \text{ roles, } \sum_{\text{spans}} z_{\text{role}\leftrightarrow\text{span}} = 1 \]
A constrained optimization problem

\[ \hat{z}_{\text{role} \leftrightarrow \text{span}} \]

\[ z = \langle \hat{z}_{\text{role} \leftrightarrow \text{span}} \rangle \]

maximize \[ \sum_{\text{roles, spans}} \hat{z}_{\text{role} \leftrightarrow \text{span}} \cdot \text{score(role} \leftrightarrow \text{span)} \]

w.r.t. \[ z \]

s.t \[ \forall \text{ roles}, \sum_{\text{spans}} \hat{z}_{\text{role} \leftrightarrow \text{span}} = 1 \]

Uniqueness
A constrained optimization problem

$$z_{\text{role} \leftrightarrow \text{span}}$$

$$z = \langle z_{\text{role} \leftrightarrow \text{span}} \rangle$$

maximize \[ \sum_{\text{roles, spans}} z_{\text{role} \leftrightarrow \text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \]

w.r.t. \[ z \]

s.t \[ \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role} \leftrightarrow \text{span}} = 1 \]

\[ \forall \text{ sentence positions}, \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role} \leftrightarrow \text{span}} \leq 1 \]

Prevents overlap
A constrained optimization problem

\[ z = \langle z_{\text{role} \leftrightarrow \text{span}} \rangle \]

An integer linear program (ILP)

\[
\begin{align*}
\text{maximize} & \quad \sum_{\text{roles, spans}} z_{\text{role} \leftrightarrow \text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \\
\text{w.r.t.} & \quad z \\
\text{s.t} & \quad \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role} \leftrightarrow \text{span}} = 1 \\
& \quad \forall \text{ sentence positions}, \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role} \leftrightarrow \text{span}} \leq 1
\end{align*}
\]
A constrained optimization problem

\[ z_{\text{role} \leftrightarrow \text{span}} \]

\[ z = \langle z_{\text{role} \leftrightarrow \text{span}} \rangle \]

An integer linear program (ILP)

\[
\text{maximize } \sum_{\text{roles}, \text{spans}} z_{\text{role} \leftrightarrow \text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \\
\text{w.r.t. } z
\]

\[
\text{s.t } \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role} \leftrightarrow \text{span}} = 1
\]

\[
\forall \text{ sentence positions}, \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role} \leftrightarrow \text{span}} \leq 1
\]

Punyakanok, Roth and Yih (2008)
A constrained optimization problem

An integer linear program (ILP)

maximize \( \sum_{\text{roles,spans}} z_{\text{role} \leftrightarrow \text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \)

w.r.t. \( z \)

s.t \( \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role} \leftrightarrow \text{span}} = 1 \)

\( \forall \text{ sentence positions, } \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role} \leftrightarrow \text{span}} \leq 1 \)

Often, very slow solutions
A constrained optimization problem

**An integer linear program (ILP)**

maximize \( \sum_{\text{roles,spans}} z_{\text{role} \leftrightarrow \text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span}) \)

w.r.t. \( z \)

s.t \( \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role} \leftrightarrow \text{span}} = 1 \)

\( \forall \text{ sentence positions}, \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role} \leftrightarrow \text{span}} \leq 1 \)

More structural constraints

Fast ILP solvers proprietary
A constrained optimization problem

An integer linear program (ILP)

\[
\text{maximize} \quad \sum_{\text{roles,spans}} z_{\text{role}\leftrightarrow\text{span}} \cdot \text{score}(\text{role} \leftrightarrow \text{span})
\]

w.r.t. \( z \)

s.t \( \forall \text{ roles}, \sum_{\text{spans}} z_{\text{role}\leftrightarrow\text{span}} = 1 \)

\( \forall \text{ sentence positions}, \sum_{\text{span covers position}} \sum_{\text{roles}} z_{\text{role}\leftrightarrow\text{span}} \leq 1 \)

Dual Decomposition
SEMAFOR: Frame-Semantic Parsing

Final Results

Benchmark

New Data

F-Measure

LTH

SEMAFOR
SEMAFOR: Frame-Semantic Parsing

Final Results

Benchmark

<table>
<thead>
<tr>
<th></th>
<th>LTH</th>
<th>SEMAFOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Measure</td>
<td>42.0</td>
<td>46.5</td>
</tr>
</tbody>
</table>

New Data

auto predicates
SEMAFOR: Frame-Semantic Parsing

Final Results

Benchmark

<table>
<thead>
<tr>
<th></th>
<th>LTH</th>
<th>SEMAFOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Measure</td>
<td>42.0</td>
<td>46.5</td>
</tr>
</tbody>
</table>

New Data

<table>
<thead>
<tr>
<th></th>
<th>SEMAFOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Measure</td>
<td>64.5</td>
</tr>
</tbody>
</table>

auto predicates

gold predicates
Hermann, Das, Weston and Ganchev
ACL 2014
Frame Identification with Embeddings

Bengal’s massive **stock** of food was reduced to nothing.

\[
\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score}(\text{frame}, \text{predicate, sentence})
\]

\[
= \log p(\text{frame} | \text{predicate, sentence})
\]
Bengal’s massive stock of food was reduced to nothing.

The best frame is selected as the one that maximizes the score:

$$\text{best \ frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score}(\text{frame}, \text{predicate}, \text{sentence})$$

$$= \log p(\text{frame} | \text{predicate, sentence})$$

The probability of the frame given the predicate and sentence is calculated as:

$$p(\text{frame} | \text{predicate, sentence}) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp w \cdot f(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate})$$
Frame Identification with Embeddings

Discrete lexical features

Bengal’s massive **stock** of food was reduced to nothing

\[
\text{best frame} = \arg \max_{\text{frame} \in \text{all frames}} \text{score}(\text{frame, predicate, sentence})
\]

\[
= \log p(\text{frame} | \text{predicate, sentence})
\]

\[
p(\text{frame} | \text{predicate, sentence}) = \frac{1}{Z} \sum_{\text{proto-predicates evoking frame}} \exp w \cdot f(\text{frame, proto-predicate, sentence, lexical-semantic relations between predicate and proto-predicate})
\]
Bengal’s massive stock of food was reduced to nothing.
Bengal’s massive stock of food was reduced to nothing.

Replace context words with off-the-shelf word embeddings.
Frame Identification with Embeddings

Input sparse embedding vector with context blocks
Frame Identification with Embeddings

Frame instance space

\[ N \times \mathbb{R}^d \]
Frame Identification with Embeddings

Joint Space $\mathbb{R}^m$

Frame instance space $N \times \mathbb{R}^d$
Frame Identification with Embeddings

Frame Instance Space

\[ M : \mathbb{R}^d \rightarrow \mathbb{R}^m \]

Joint Space

\[ \mathbb{R}^m \]

Frame Instance Map

\[ N \times \mathbb{R}^d \]

Frame instance space
Frame Identification with Embeddings

Frame Instance Map

\[ M : \mathbb{R}^d \rightarrow \mathbb{R}^m \]

Joint Space

\[ \mathbb{R}^m \]

Frame instance space

\[ N \times \mathbb{R}^d \]

Set of FrameNet labels

... STINGINESS STORE STORING ...

117
Frame Identification with Embeddings

Frame Instance Map

$$M : \mathbb{R}^d \rightarrow \mathbb{R}^m$$

Joint Space

$$\mathbb{R}^m$$

Label matrix

$$Y \in \mathbb{R}^{F \times m}$$

Set of FrameNet labels

... STINGINESS STORE STORING ...

Frame instance space

$$N \times \mathbb{R}^d$$
Frame Identification with Embeddings

Frame Instance Map: $M : \mathbb{R}^d \rightarrow \mathbb{R}^m$

Frame instance space

Joint Space: $\mathbb{R}^m$

Label matrix: $Y \in \mathbb{R}^{F \times m}$

Set of FrameNet labels

- STORE
- STORING
- STINGINESS

$N \times \mathbb{R}^d$
Frame Identification with Embeddings
Frame Identification with Embeddings

frame instances  ○
Frame Identification with Embeddings

frame instances

frame labels
Frame Identification with Embeddings

Frame instances

Frame labels

Test predicate
Frame Identification with Embeddings

Frame instances

frame labels

test predicate
Frame Identification with Embeddings

Frame instances

Frame labels

Test predicate
Frame Identification with Embeddings

frame instances  
frame labels
Frame Identification with Embeddings

frame instances
frame labels
Results on Unknown Predicates

Frame Identification

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>23.1</td>
</tr>
<tr>
<td>Self-Training</td>
<td>18.9</td>
</tr>
<tr>
<td>Graph-Based</td>
<td>42.7</td>
</tr>
<tr>
<td>Embeddings</td>
<td>46.15</td>
</tr>
</tbody>
</table>

SEMAFOR
Results on All Predicates

Frame Identification

Accuracy

Supervised: 83.0
Self-Training: 82.3
Graph-Based: 83.6
Embeddings: 86.49

SEMAFOR
Final Results
Frame-Semantic Parsing

F-Score

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>64.1</td>
</tr>
<tr>
<td>Graph-Based</td>
<td>64.5</td>
</tr>
<tr>
<td>Embeddings</td>
<td>68.7</td>
</tr>
</tbody>
</table>

SEMAFOR
Outline

• Why should we build statistical models for frame semantics?

• Overview of statistical methods

• Future directions
Better Models

• Very little research on argument identification
  • More non-local features
  • Using distributed representations
  • Generalization using PropBank resources
Data

• Number of FrameNet annotated sentences ~30 times less than PropBank/Ontonotes.

• Number of argument labels ~30 times more

• To make systems usable, we need annotations
  • inter-annotator agreement studies
  • annotation guidelines
Custom FrameNets

• Is a general FrameNet lexicon useful?
  • often, FrameNet frames are too general or too specific
  • is it possible to quickly build customized FrameNet lexicons for applications?
  • is it possible to use PropBank-style frames to induce FrameNet-style frames?
Conclusions

- Lot of exciting work in predicate-argument structure prediction
- Semi-supervised methods improve coverage
- Systems trained on small amounts of FrameNet-style data shown to be useful
- More annotations will result in usable systems
Thank You