Unsupervised Induction of Shallow Semantic Representations with Feature-Rich Models

Diego Marcheggiani, Ivan Titov, and Ehsan Khoddam

University of Amsterdam
Natural language processing (NLP)

The key bottleneck: the lack of accurate methods for producing meaning representations of texts and reasoning with these representations.
Modern frame-semantic parsers rely on supervised learning.

**Challenge #1**

It is impossible to annotate enough data to estimate an effective broad-coverage semantic parser.
Lansky left Australia to study the piano at the Royal College of Music.

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?
Lansky left Australia to study the piano at the Royal College of Music.

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?
CMU's SEMAFOR [Das et al., 2012] trained on 100,000 sentences (FrameNet)

Lansky left Australia to study the piano at the Royal College of Music.

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?

The parser's output does not let us answer even this simple question.
Lansky left Australia to study the piano at the Royal College of Music.

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?
"Correct" semantics as imposed by linguists

1. Where did Lansky get his diploma? Trinity or RCM ????

Challenge #2

Representations defined by linguists are not appropriate for reasoning
"Correct" semantics as imposed by linguists

1. Where did Lansky get his diploma?

| Trinity or RCM ???? |

**Challenge #2**

Representations defined by linguists are **not always appropriate for reasoning helpful for solving specific tasks (i.e., QA)**.
The challenges motivated research in unsupervised role / frame induction:

- **Role induction** [Swier and Stevenson '04; Grenager and Manning '06; Lang and Lapata '10, '11, '14; Titov and Klementiev '12; Garg and Henderson '12;...]

- **Frame induction** [Titov and Klementiev '11; O'Connor '12; Modi et al.'12; Materna '12; Kawahara et al. '13; Cheung et al.'13; Chambers et al., 14;...]

These models rely on **very restricted sets of features**

Not (quite) appropriate for **inference** (i.e., QA task)
Contributions

- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - language-independent

- The framework naturally supports:
  - Integration of prior linguistic knowledge
  - Semi-supervised learning
Outline

- **Framework:** reconstruction error minimization for semantics
- **Special case:** inferring missing arguments
- **Empirical evaluation:** role induction, frame induction
General framework

- **Left-out facts**
  - **Reconstruction**
  - **Semantic representations**
    - **Encoding**
  - **Text(s)**

- Not observable in the data – need to be induced
Instead of using annotated data, induce representations beneficial for inferring left-out facts.
General framework

- **Left-out facts**
  - ideas from statistical relational learning e.g., [Yilmaz et al., '11]

- **Semantic representations**

- **Text(s)**

- **Inference model:** tensor factorization

- **Encoding**
General framework

- **Left-out facts**
  - Inference model: tensor factorization
  - Semantic representations
    - Ideas from supervised parsing
  - Semantic parser: expressive model
    - E.g., [Das et al., '10, Titov et al., '09]

Inference model and semantic parser are **jointly** estimated from unlabeled data.
Lansky left Australia to study the piano at the Royal College of Music.

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?  Trinity

The learning objective can ensure that the representations are informative for reasoning
Outline

- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: role induction, frame induction
Feature-rich models of semantic frames

Consider a frame realization

The police charged the demonstrators with their batons

\[
\mathbf{a} = (a_1, \ldots, a_n) \quad \text{- arguments (police, the demonstrators, their batons)}
\]

\[
\mathbf{r} = (r_1, \ldots, r_n) \quad \text{- roles (Perpetrator, Victim, Instrument)}
\]

\[
f \quad \text{- frame (Assault)}
\]
Consider a frame realization

\[ \mathbf{a} = (a_1, \ldots, a_n) \quad \text{- arguments (police, the demonstrators, their batons)} \]
\[ \mathbf{r} = (r_1, \ldots, r_n) \quad \text{- roles (Perpetrator, Victim, Instrument)} \]
\[ f \quad \text{- frame (Assault)} \]
Argument reconstruction

Consider a frame realization

The police charged the demonstrators with their batons

"Argument prediction" model

\[ p(a_i | a_{-i}, r, f, \theta) \]

Assault(Agent: police, Patient: demonstrator, Instrument: baton)
Argument reconstruction

Consider a frame realization

The police charged the demonstrators with their batons

Hypothesis: semantic roles and frames are the latent representation which helps to reconstruct arguments
Argument reconstruction

Consider a frame realization

The police charged the demonstrators with their batons

"Argument prediction" model
\[ p(a_i | \mathbf{a}_{-i}, \mathbf{r}, f, \theta) \]

Feature-rich model
\[ p(\mathbf{r}, f | x, \mathbf{w}) \]

Feature representation of "The police charged...

How do the components look like and how do we estimate them jointly?
Component 1: argument reconstruction

*`demonstrator`*

Argument prediction ( = Reconstruction)

"Argument prediction" model

\[ p(\mathbf{a}_i | \mathbf{a}_{-i}, \mathbf{r}, f, \theta) \]

Assault (Agent: *police*, Patient: *demonstrator*, Instrument: *baton*)

Distributed vectors:

\[ \mathbf{u}_a \in \mathbb{R}^d \] - encode semantic properties of argument \( a \)
Component 1: argument reconstruction

Assault(Agent: police, Patient: demonstrator, Instrument: baton)

"Argument prediction" model

\[ p(a_i | a_{-i}, r, f, \theta) \]

Distributed vectors:

\[ u_a \in \mathbb{R}^d \]  - encode semantic properties of argument \( a \)

\[ C_{f,r} u_a \in \mathbb{R}^k \]  - encode expectations about other arguments

A role-specific projection matrix

If Agent of Assault is the police, then Patient can be demonstrators or protestors
Component 1: argument reconstruction

"Argument prediction" model

\[ p(a_i | a_{-i}, r, f, \theta) \]

Distributed vectors:

\[ u_a \in \mathbb{R}^d \quad - \text{encode semantic properties of argument } a \]

\[ C_{f, r} u_a \in \mathbb{R}^k \quad - \text{encode expectations about other arguments} \]

A role-specific projection matrix

\[ p(a_i | a_{-i}, r, f, C, u) = \frac{\exp(u_{a_i}^T C_{f, r}^T \sum_{j \neq i} C_{f, r_j} u_{a_j})}{Z(r, f, i)} \]

If Agent of Assault is the police, then Patient can be demonstrators or protestors.
Component 1: argument reconstruction

**demonstrator**

Argument prediction ( = Reconstruction)

"Argument prediction" model

\[ p(a_i|a_{-i}, r, f, \theta) \]

Assault(Agent: *police*, Patient: *demonstrator*, Instrument: *baton*)

Distributed vectors:

- **\( u_a \in \mathbb{R}^d \)** - encode semantic properties of argument \( a \)
- **\( C_{f,r} u_a \in \mathbb{R}^k \)** - encode expectations about other arguments

A role-specific projection matrix

If Agent of Assault is the *police*, then Patient can be *demonstrators* or *protestors*

\[
p(a_i|a_{-i}, r, f, C, u) = \frac{\exp(u^T_{a_i} C^T_{f,r_i} \sum_{j \neq i} C_{f,r_j} u_{a_j})}{Z(r, f, i)}
\]

Factorization function: scoring arguments tuples for a given frame and role assignment
Component 2: frame + role prediction

The role and frame labeling model:

\[ p(r, f|x, w) \propto \exp(w^T g(x, f, r)) \]

It can be any model as long as role and frame posteriors \( p(r_i|x, w) \) and \( p(f|x, w) \) can be computed (or approximated)
For every structure, we aim to optimize the argument prediction quality given roles and frames:

\[
\sum_{i=1}^{N} \sum_{r,f} q(r, f | x, w) \log p(a_i | a_{-i}, r, f, C, u)
\]
Joint learning

For every structure, we aim to optimize the argument prediction quality given roles and frames:

\[
\sum_{i=1}^{N} \sum_{r,f} q(r, f|x, w) \log p(a_i|a_{-i}, r, f, C, u)
\]

reconstruction
For every structure, we aim to optimize the argument prediction quality given roles and frames:

$$\sum_{i=1}^{N} \sum_{r,f} q(r, f|x, w) \log p(a_i|a_{-i}, r, f, C, u)$$
For every structure, we aim to optimize the argument prediction quality given roles and frames:

$$\sum_{i=1}^{N} \sum_{r,f} q(r,f|x,w) \log p(a_i|a_{-i}, r, f, C, u)$$
Outline

- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- **Empirical evaluation:** role induction, frame induction
Task: induce for each argument the appropriate semantic role

Evaluate on a dataset annotated with roles (PropBank for En, SALSA for De)

Compare against previous models evaluated in this set-up
  - use clustering evaluation measures (purity, collocation, F1)
Results English (F1)

Optimal deterministic mapping from syntactic relations

Performs on par with best methods (without language-specific priors)

Logistic: Lang and Lapata ('10)
GraphP: Lang and Lapata ('11a)
Linking: Fürstenau and Rambow ('12)
Aggl: Lang and Lapata ('11b)
Order: Garg and Henderson ('12)
Aggl+: Lang and Lapata ('14)
Bayes: Titov and Klementiev ('12)

Our model
Results German (F1)

Bayes: Titov and Klementiev ('12a)
Bayes (De): Titov and Klementiev ('12b)

Performs on par with the best method without language-specific engineering

Bayes modified for German

The feature rich model (the same as for En)
Relation discovery (Frame induction)

- **Task**: Induce semantic relations between two given arguments
- **Data**: New York Times corpus
- **Evaluation against**: Freebase
Results (F1)

- DIRT: 0.283
- Rel-LDA: 0.296
- Our model: 0.358

[Marcheggiani and Titov, TACL '16]

Clustering baseline
Generative baseline
Results (F1)

- DIRT: 0.283
- Rel-LDA: 0.296
- Our model: 0.358

Our model is 6.2% more accurate than the Rel-LDA baseline.

[Marcheggiani and Titov, TACL '16]
Conclusion

- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - expressive features
  - supports semi-supervised learning
Conclusion

- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - expressive features
  - supports semi-supervised learning

Thank you!

Funding:
- NWO VIDI grant
- Google Focused Award on Natural Language Understanding