Markov Logic in Machine Reading

Hoifung Poon
Dept. of Computer Science & Eng.
University of Washington
“Drowning in Information, Starved for Knowledge”
Example: Biomedical Research

- PubMed contains 18 millions abstracts
- Adds more than 2000 every day
Machine Reading: Text → Knowledge

\[ \text{INDUCE}(e_1) \land \text{IL-4}(e_2) \land \text{CD11B}(e_3) \land \text{INDUCER}(e_1,e_2) \land \text{INDUCED}(e_1,e_3) \]
Example: Literature-Based Discovery

Propose new hypotheses by assembling knowledge across subfields [Swanson & Smalheiser, 1997]

Machine reading can revolutionize literature-based discovery methods
Machine Reading: Challenges

- Complexity
- Uncertainty
- Language variations

E.g., same meaning, different expressions

*Microsoft buys Powerset*

*Microsoft acquires semantic search engine Powerset*

*Powerset is acquired by Microsoft Corporation*

*The Redmond software giant buys Powerset*

*Microsoft’s purchase of Powerset, ...*

……..
This Talk

- Statistical relational learning offers promising solutions to machine reading
- Markov logic is a leading unifying framework
- USP: End-to-End Machine Reading
  - Read text, extract knowledge, answer questions, all without any training examples
  - Substantially outperformed state of the art
    - Extracted five times as many correct answers
    - Raised accuracy from below 60% to 91%
Interestingly, the DEX-mediated IkappaBalpha induction was completely inhibited by IL-2, but not IL-4, in Th1 cells, while the reverse profile was seen in Th2 cells.

Q: What does IL-2 control?
Interestingly, the DEX-mediated IkappaBalpha induction was completely inhibited by IL-2, but not IL-4, in Th1 cells, while the reverse profile was seen in Th2 cells.

Q: What does IL-2 control?
Interestingly, the DEX-mediated IkappaBalpha induction was completely inhibited by IL-2, but not IL-4, in Th1 cells, while the reverse profile was seen in Th2 cells.

Q: What does IL-2 control?
A: The DEX-mediated IkappaBalpha induction

\[
\begin{align*}
\text{REGULATE} & \quad \text{regulate, control, govern, modulate} \\
\text{INDUCE} & \quad \text{induce, enhance, trigger, augment, up-regulate} \\
\text{INHIBIT} & \quad \text{inhibit, block, suppress, prevent, abolish, abrogate, down-regulate} \\
\text{ACTIVATE} & \quad \text{activate}
\end{align*}
\]
Statistical Relational Learning

- Emerging direction in machine learning
- **Combines logic and probability**
- **Joint inference:**
  - Models complex interdependencies
  - Propagates information from more certain decisions to resolve uncertainty in others
Markov Logic

- **Intuition:** Soften logical constraints
- **Syntax:** Weighted first-order formulas
- **Semantics:** Feature templates for Markov networks

A Markov Logic Network (MLN) is a set of pairs \((F_i, w_i)\) where

- \(F_i\) is a formula in first-order logic
- \(w_i\) is a real number

\[
P(x) = \frac{1}{Z} \exp \left( \sum_i w_i \cdot n_i(x) \right)
\]
Markov Logic

- Unified inference and learning algorithms
  → Can handle millions of variables, billions of features, ten of thousands of parameters

- Easy-to-use software: Alchemy

- Many successful applications
  E.g.: Information extraction, coreference resolution, semantic parsing, ontology induction, etc.
Semantic Parsing

Goal  Microsoft buys Powerset  →  BUY (MICROSOFT, POWERSET)

Challenge  Microsoft buys Powerset
Microsoft acquires semantic search engine Powerset
Powerset is acquired by Microsoft Corporation
The Redmond software giant buys Powerset
Microsoft’s purchase of Powerset, ...
Unsupervised Semantic Parsing

- **USP** [Poon & Domingos, EMNLP-09]
  - First unsupervised approach for semantic parsing
  - End-to-end machine reading system
  - Read text, answer questions

- **OntoUSP = USP + Ontology Induction** [Poon & Domingos, ACL-10]
  - Encoded in a few Markov logic formulas

**Best Paper Award**
USP: Key Idea # 1

- Target predicates and objects can be learned
- Viewed as clusters of syntactic or lexical variations of the same meaning

\[
\text{BUY} (\cdot, \cdot, \cdot) = \{\text{buys, acquires, 's purchase of, ...} \}
\]

\[
\text{MICROSOFT} = \{\text{Microsoft, the Redmond software giant, ...} \}
\]

= Cluster of various expressions for acquisition

= Cluster of various mentions of Microsoft
USP: Key Idea # 2

- **Relational clustering** = Cluster relations with same objects
- **USP** = **Recursively** cluster arbitrary expressions with similar subexpressions

*Microsoft buys Powerset*
*Microsoft acquires semantic search engine Powerset*
*Powerset is acquired by Microsoft Corporation*
*The Redmond software giant buys Powerset*
*Microsoft’s purchase of Powerset, ...*
USP: Key Idea # 2

- **Relational clustering** = Cluster relations with same objects
- **USP** = *Recursively* cluster arbitrary expressions with similar subexpressions

*Microsoft buys Powerset*

*Microsoft acquires semantic search engine Powerset*

*Powerset is acquired by Microsoft Corporation*

*The Redmond software giant buys Powerset*

*Microsoft’s purchase of Powerset, ...*

Cluster same forms at the atom level
USP: Key Idea # 2

- **Relational clustering** = Cluster relations with same objects
- **USP** = **Recursively** cluster **arbitrary** expressions with similar subexpressions

Microsoft *buys* Powerset

Microsoft *acquires* semantic search engine Powerset

Powerset *is acquired by* Microsoft Corporation

The Redmond software giant *buys* Powerset

Microsoft’s *purchase of* Powerset, ...

Cluster forms in composition with same forms
USP: Key Idea # 2

- **Relational clustering** = Cluster relations with same objects
- **USP** = *Recursively* cluster *arbitrary* expressions with similar subexpressions

*Microsoft buys Powerset*
*Microsoft acquires *semantic search engine Powerset*
*Powerset is acquired by Microsoft Corporation*
*The Redmond software giant *buys Powerset*
*Microsoft’s purchase of Powerset, ...*

Cluster forms in composition with same forms
USP: Key Idea # 2

- **Relational clustering** = Cluster relations with same objects
- **USP** = *Recursively* cluster *arbitrary* expressions with similar subexpressions

*Microsoft buys Powerset*
*Microsoft acquires semantic search engine Powerset*
*Powerset is acquired by Microsoft Corporation*
*The Redmond software giant buys Powerset*
*Microsoft’s purchase of Powerset, …*

Cluster forms in composition with same forms
Generating Quasi-Logical Forms

Convert each node into an unary atom

Diagram:

buys

nsubj Microsoft

doobj Powerset
Generating Quasi-Logical Forms

buys\( (n_1) \)

nsubj \quad dobj

Microsoft\( (n_2) \) \quad Powerset\( (n_3) \)

\( n_1, n_2, n_3 \) are Skolem constants
Generating Quasi-Logical Forms

Convert each edge into a binary atom
Generating Quasi-Logical Forms

- \text{buys}(n_1)
- \text{nsubj}(n_1, n_2)
- \text{dobj}(n_1, n_3)
- \text{Microsoft}(n_2)
- \text{Powerset}(n_3)

Convert each edge into a binary atom
A Semantic Parse

Partition QLF into subformulas

buys($n_1$)

nsubj($n_1,n_2$) dobj($n_1,n_3$)

Microsoft($n_2$) Powerset($n_3$)
A Semantic Parse

buys(n₁)

nsubj(n₁,n₂) dobj(n₁,n₃)

Microsoft(n₂) ∈ MICROSOFT

Powerset(n₃) ∈ POWERSET

Assign subformula to object cluster
Probabilistic Model

- Exponential prior on number of parameters
- Cluster mixtures:

Object Cluster: BUY

- buys 0.1
- acquires 0.4

Property Cluster: BUYER

- nsubj 0.5
- agent 0.4

- MICROSOFT 0.2
- GOOGLE 0.1

- Zero 0.1
- One 0.8
ISA → Can Inherit Parameters

REGULATE

REGULATED

... ...

INDUCE

...

INHIBIT

...

ISA $\rightarrow$ Can Inherit Parameters
Learning: Greedily Maximize Posterior

Initialize

Search Operators

MERGE

induces 1.0 enhances 1.0

induces 0.2 enhances 0.8

COMPOSE

amino 1.0 acid 1.0

amino acid 1.0
Experiments

- **Apply to machine reading:** Extract knowledge from text and answer questions
- **Evaluation:** Number of answers and accuracy
- **GENIA dataset:** 1999 Pubmed abstracts
- Used factoid questions, e.g.:
  - *What does anti-STAT1 inhibit?*
  - *What regulates MIP-1 alpha?*
- Sampled two thousand questions by frequency
USP extracted five times as many correct answers as TextRunner.

Highest precision of 91%
Why USP Did Better

- **Resolve many nontrivial variations**
- Argument forms that mean the same, e.g.,
  - expression of $X = X$ expression
  - $X$ stimulates $Y = Y$ is stimulated with $X$
- Active vs. passive voices
- Synonymous expressions
- Etc.
Clusters And Compositions

- Clusters in core forms
  - { investigate, examine, evaluate, analyze, study, assay }
  - { diminish, reduce, decrease, attenuate }
  - { synthesis, production, secretion, release }
  - { dramatically, substantially, significantly }

- Compositions
  - amino acid, t cell, immune response, transcription factor, initiation site, binding site …
Future Work

- Scale joint inference to the Web
- Add more joint inference to USP
- Harness social computing
- Apply machine reading to biomed research
Summary

- **Statistical relational learning** offers promising solutions for machine reading

- **Markov logic** provides a language for this
  - **Syntax**: Weighted first-order logical formulas
  - **Semantics**: Feature templates of Markov nets

- **Open-source software**: Alchemy
  - alchemy.cs.washington.edu

- **A success story**: USP
  - alchemy.cs.washington.edu/papers/poon09

- **Four key research directions**