CS 6364: Advanced Topics

Markov Logic
Natural Language Processing
Computer Vision

Vibhav Gogate
(slides borrowed from Pedro Domingos and Dan Klein)
Markov Logic: Putting the Pieces Together

- Markov Logic
  - Probability
  - First-Order Logic
  - Propositional Logic
Markov Logic

- A logical KB is a set of **hard constraints** on the set of possible worlds
- Let’s make them **soft constraints**: When a world violates a formula, it becomes less probable, not impossible
- Give each formula a **weight** (Higher weight \( \Rightarrow \) Stronger constraint)

\[
P(\text{world}) \propto \exp\left(\sum \text{weights of formulas it satisfies}\right)
\]
Definition

- A Markov Logic Network (MLN) is a set of pairs \((F, w)\) where
  - \(F\) is a formula in first-order logic
  - \(w\) is a real number
- Together with a set of constants, it defines a Markov network with
  - One node for each grounding of each predicate in the MLN
  - One feature for each grounding of each formula \(F\) in the MLN, with the corresponding weight \(w\)
Example: Friends & Smokers

Smoking causes cancer.
Friends have similar smoking habits.
Example: Friends & Smokers

∀x \ Smokes(x) \Rightarrow \ Cancer(x)
∀x, y \ Friends(x, y) \Rightarrow (\ Smokes(x) \Leftrightarrow \ Smokes(y))\)
Example: Friends & Smokers

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>$\forall x \ Smokes(x) \Rightarrow Cancer(x)$</td>
</tr>
<tr>
<td>1.1</td>
<td>$\forall x, y \ Friends(x, y) \Rightarrow (Smokes(x) \Leftrightarrow Smokes(y))$</td>
</tr>
</tbody>
</table>
Example: Friends & Smokers

| 1.5 | \( \forall x \ Smokes(x) \Rightarrow Cancer(x) \) |
| 1.1 | \( \forall x, y \ Friends(x, y) \Rightarrow (Smokes(x) \iff Smokes(y)) \) |

Two constants: Anna (A) and Bob (B)
Example: Friends & Smokers

1.5 \( \forall x \ Smokes(x) \Rightarrow Cancer(x) \)

1.1 \( \forall x, y \ Friends(x, y) \Rightarrow (Smokes(x) \iff Smokes(y)) \)

Two constants: Anna (A) and Bob (B)
Example: Friends & Smokers

Two constants: Anna (A) and Bob (B)
### Example: Friends & Smokers

| 1.5 | $\forall x \ Smokes(x) \Rightarrow Cancer(x)$ |
| 1.1 | $\forall x, y \ Friends(x, y) \Rightarrow (Smokes(x) \Leftrightarrow Smokes(y))$ |

Two constants: **Anna** (A) and **Bob** (B)
Example: Friends & Smokers

1.5 $\forall x \ Smokes(x) \Rightarrow Cancer(x)$
1.1 $\forall x, y \ Friends(x, y) \Rightarrow (Smokes(x) \Leftrightarrow Smokes(y))$

Two constants: Anna (A) and Bob (B)
Markov Logic Networks

- **MLN is template** for ground Markov nets
- Probability of a world $x$:
  \[
  P(x) = \frac{1}{Z} \exp \left( \sum_i w_i n_i(x) \right)
  \]
  - Weight of formula $i$
  - No. of true groundings of formula $i$ in $x$

- **Typed** variables and constants greatly reduce size of ground Markov net
- Functions, existential quantifiers, etc.
- Open question: Infinite domains
Relation to First-Order Logic

- Infinite weights $\Rightarrow$ First-order logic
- Satisfiable KB, positive weights $\Rightarrow$
  Satisfying assignments = Modes of distribution
- Markov logic allows contradictions between formulas
MAP/MPE Inference

- **Problem:** Find most likely state of world given evidence

\[
\max_y P(y \mid x)
\]

Query \hspace{2cm} Evidence
MAP/MPE Inference

Problem: Find most likely state of world given evidence

$$\max_y \frac{1}{Z_x} \exp \left( \sum_i w_i n_i(x, y) \right)$$
MAP/MPE Inference

- **Problem:** Find most likely state of world given evidence

\[ \max_y \sum_{i} w_i n_i(x, y) \]
MAP/MPE Inference

- **Problem:** Find most likely state of world given evidence

\[
\max_y \sum_i w_i n_i(x, y)
\]

- This is just the weighted MaxSAT problem
- Use weighted SAT solver
  (e.g., MaxWalkSAT [Kautz et al., 1997])
- Potentially faster than logical inference (!)
for $i \leftarrow 1$ to $\text{max-tries}$ do
  solution = random truth assignment
  for $j \leftarrow 1$ to $\text{max-flips}$ do
    if $\sum \text{weights(sat. clauses)} > \text{threshold}$ then
      return solution
    $c \leftarrow \text{random unsatisfied clause}$
    with probability $\rho$
      flip a random variable in $c$
    else
      flip variable in $c$ that maximizes
      $\sum \text{weights(sat. clauses)}$
  return failure, best solution found
But … Memory Explosion

- **Problem:**
  If there are $n$ constants and the highest clause arity is $c$, the ground network requires $O(n^c)$ memory

- **Solution:**
  Exploit sparseness; ground clauses lazily
  $\rightarrow$ LazySAT algorithm [Singla & Domingos, 2006]
Computing Probabilities

- $P(\text{Formula}|\text{MLN,C}) = ?$
- MCMC: Sample worlds, check formula holds
- $P(\text{Formula}_1|\text{Formula}_2,\text{MLN,C}) = ?$
- If $\text{Formula}_2 = \text{Conjunction of ground atoms}$
  - First construct min subset of network necessary to answer query (generalization of KBMC)
  - Then apply MCMC (or other)
- Can also do lifted inference [Braz et al, 2005]
Ground Network Construction

\[
\begin{align*}
\text{network} & \leftarrow \emptyset \\
\text{queue} & \leftarrow \text{query nodes} \\
\text{repeat} \\
& \quad \text{node} \leftarrow \text{front(queue)} \\
& \quad \text{remove node from queue} \\
& \quad \text{add node to network} \\
& \quad \text{if node not in evidence then} \\
& \quad & \quad \text{add neighbors(node) to queue} \\
\text{until} & \quad \text{queue} = \emptyset
\end{align*}
\]
Lifted Inference

- Take Advantage of symmetries in the network
- Bring the power of resolution (first-order inference) to probabilistic inference
Natural Language Processing

Hello, I am Eliza.

Hi, my name is Watson.
What is NLP?

- Fundamental goal: analyze and process human language, broadly, robustly, accurately…
- End systems that we want to build:
  - Ambitious: speech recognition, machine translation, information extraction, dialog interfaces, question answering…
  - Modest: spelling correction, text categorization…
Problem: Ambiguities

- Headlines:
  - Enraged Cow Injures Farmer With Ax
  - Hospitals Are Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Local HS Dropouts Cut in Half
  - Juvenile Court to Try Shooting Defendant
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks

- Why are these funny?
Parsing as Search

S
  └── NP ── VP
     └── N ── V ── NP
          └── protest

Hershey bars protest

S
  └── NP ── VP
     └── N ── N ── V
          └── protest
Grammar: PCFGs

- Natural language grammars are very ambiguous!
- PCFGs are a formal probabilistic model of trees
  - Each “rule” has a conditional probability (like an HMM)
  - Tree’s probability is the product of all rules used
- Parsing: Given a sentence, find the best tree – search!

ROOT → S 375/420
S → NP VP . 320/392
NP → PRP 127/539
VP → VBD ADJP 32/401
…..
Hurricane Emily howled toward Mexico's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.

[Demo: Berkeley NLP Group Parser http://tomato.banatao.berkeley.edu:8080/parser/parser.html]
Dialog Systems

1. Hello, I am Eliza.

2. Hi, my name is Watson.
ELIZA

- A “psychotherapist” agent (Weizenbaum, ~1964)
- Led to a long line of chatterbots
- How does it work:
  - Trivial NLP: string match and substitution
  - Trivial knowledge: tiny script / response database
  - Example: matching “I remember __” results in “Do you often think of __”? 
- Can fool some people some of the time?

[Demo: http://nlp-addiction.com/eliza]
Watson

"a camel is a horse designed by a committee"

Wikipedia

a camel is a horse designed by a committee

Contents [hide]
1 English
1.1 Alternative forms

The Phrase Finder

A camel is a horse designed by committee

Posted by Ruben P. Mendez on April 16, 2004

Does anyone know the origin of this maxim? I heard it way back at the United Nations, which is chockfull of committees. It may have originated there, but I'd like an authoritative explanation. Thanks

- Re: A camel is a horse designed by committee [SR 16/April/04]
- Re: A camel is a horse designed by committee [Henry 18/April/04]
What’s in Watson?

- A question-answering system (IBM, 2011)
- Designed for the game of Jeopardy
- How does it work:
  - Sophisticated NLP: deep analysis of questions, noisy matching of questions to potential answers
  - Lots of data: onboard storage contains a huge collection of documents (e.g. Wikipedia, etc.), exploits redundancy
  - Lots of computation: 90+ servers
- Can beat all of the people all of the time?
Machine Translation
Machine Translation

- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
  - What fragments? [learning to translate]
  - How to make efficient? [fast translation search]
Computer Vision
Object Detection
Object Detection Approach 1: HOG + SVM
Features and Generalization

[Dalal and Triggs, 2005]
Features and Generalization

Image

HoG
Training

- **Round 1**
  - Training set =
    - Positive examples: from labeling
    - Negative examples: random patches
  → preliminary SVM

- **Round 2** ("bootstrapping" or "mining hard negatives")
  - Training set =
    - Positive examples: from labeling
    - Negative examples: patches that have score $\geq -1$
  → final SVM
State-of-the-art Results

[Girschik, Felzenszwalb, McAllester]
State-of-the-art Results

[Girschik, Felzenszwalb, McAllester]

person

car

horse