**Introduction**

**Goal:** Learn general knowledge rules from natural texts

**Challenges:**
- With a stunning display of offensive powers between MiamiDolphins and the Eagles, 39-35 routed Eagles back home
- Radically incomplete: Only a very small portion of the “whole truth” is actually mentioned in the documents.
- Systematically biased: Information is biased towards newsworthiness

**Mention Model**

Shared Domain Knowledge, K

Writer Communicates Reader: \( F \land \neg G \)

ReaderWillInfer(\( G \), \( F \), \( K \)):

Reader will infer \( G \), when told \( F \) using \( K \)

Efficient Communication Model/Mention Model:

\[ \text{Mention}(F) \land \text{ReaderWillInfer}(G, F, K) \land G \Rightarrow \neg \text{Mention}(G) \]

**Mention Observations** (mention-mention rules)

\[ \text{mention}_\text{GameAwayTeam}(g, t_2) \land \text{mention}_\text{TeamInGame}(g, t_1) \land \text{mention}_\text{TeamInGame}(g, t_2) \Rightarrow \neg \text{mention}_\text{GameHomeTeam}(g, t_1) \]

**Facts** (fact-fact rules)

\[ \text{fact}_\text{GameAwayTeam}(g, t_2) \land \text{fact}_\text{TeamInGame}(g, t_1) \land \text{fact}_\text{TeamInGame}(g, t_2) \Rightarrow \text{fact}_\text{GameHomeTeam}(g, t_1) \]

**Mention Model contd.**

Similarly, Writer Communicates Reader: \( F \land \neg G \)

\[ \text{Mention}(F) \land \text{ReaderWillInfer}(G, F, K) \land \neg G \Rightarrow \neg \text{Mention}(G) \]

**Components:**
- **Fact-to-Fact Rules:**
  - Domain Knowledge rules; we learn candidates from observations; soft rules
- **Mention-to-Fact Rules:**
  - Anything mentioned is true; hard rules
- **Fact-to-Mention Rules:**
  - Facts are likely to be mentioned; soft rules
- **Mention-to-Mention Rules:**
  - Captures the mention model; soft rules

**Idea:**

- Learn candidate fact-to-fact rule from observations
- Use MLN to represent facts and mention rules
- Use Probabilistic Inference to predict missing facts
- Apply EM to estimate parameters of mention model

**Learn Rules**

**Given:** \( P \), a set of predicates; \( D \), mention observations; \( \tau \), support threshold

- For each Head,
  - Exhaustive Search over other predicates
  - Construct candidate rules
- FOIL like pruning idea, use \( \tau \) to grow rules set
- Unlike FOIL, allow multiples rules per head
- Cross-validation to limit size of the rules set

**Latent Variable Model:**

\[ \text{Latent: facts Observed: mentions} \]

**Experiments**

**Learn Weights**

- Use generative learning in MLN to learn weights
- Use Lazy MC-SAT to infer facts
- Use EM to estimate latent variable model

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.6</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>19.3</td>
<td>1.0</td>
<td>0.99 0.97 0.96 0.90 0.85</td>
</tr>
<tr>
<td>31.3</td>
<td>1.0</td>
<td>0.99 0.98 0.97 0.93 0.87</td>
</tr>
<tr>
<td>36.5</td>
<td>1.0</td>
<td>0.98 0.92 0.92 0.82 0.66</td>
</tr>
<tr>
<td>48.3</td>
<td>0.99 0.98 0.72 0.71 0.61 0.54</td>
<td></td>
</tr>
<tr>
<td>57.9</td>
<td>0.91 0.81 0.72 0.68 0.56 0.41</td>
<td></td>
</tr>
</tbody>
</table>

**Experiments on Real Data (NFL):**

- At high missingness, - performance decreases
- learns noisy rules

**Extractions in D1 is extremely incomplete (co-ref err) and only a few (<3%) home/away mentions**

- Extracts in D1 is extremely incomplete (co-ref err) and only a few (<3%) home/away mentions

**OSU Oregon State University**

**ERUDITE**