Scaling Textual Inference to the Web

Stefan Schoenmackers
Oren Etzioni
Daniel S. Weld

Turing Center
University of Washington
What prevents osteoporosis?
How can we overcome this?

- Combine facts from multiple pages to infer the answer
  - kale is high in calcium (54 pages)
  - calcium prevents osteoporosis (440 pages)
  - ∴ Kale prevents osteoporosis (with high probability)

- Problems
  - Web text is noisy
  - **Inference is expensive**
Related work

- **Q/A**
  - Mulder (Kwok et al., 2001) AskMSR (Brill et al., 2002), (Harabagiu et al., 2000) and others.

- **Textual Entailment**
  - (Dagan et al., 2005), (Raina et al., 2005), (MacCartney and Manning, 2007), (Tatu and Moldovan, 2006), (Braz et al., 2005)

- **Differences:**

<table>
<thead>
<tr>
<th>Textual Entailment</th>
<th>Holmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sentence/page</td>
<td>100 Million pages</td>
</tr>
<tr>
<td>Complex representation</td>
<td>Restricted representation</td>
</tr>
</tbody>
</table>
Contributions

- Holmes
  - Inference over information extracted from hundreds of millions of Web pages
- Approximately Pseudo-Functional (APF) Property
  - Guarantees inference is $O(n)$
- Empirical Evaluation
  - Inference doubles recall on 20 test queries
Outline

- Motivation
- Holmes System
- Scalability
- Empirical Results
- Conclusions
Holmes

- Knowledge Based Model Construction (KBMC) (Wellman et al., 2005)
  - Horn clause inference rules ( \( A \land B \Rightarrow C \) )
  - Millions of relation instances extracted from text

- Focus on scalability!
  - Ignore syntactic problems (e.g. anaphora)
  - Ignore semantic challenges (e.g. quantification, temporal info)
Holmes

KBs
0.9 Prevents(Ca, osteo.)
0.8 Contains(kale, Ca)
...

Inf. Rules
Weighted Horn Clauses
1.5 \( P(X,Z) : -C(X,Y)^P(Y,Z) \)

Query
\( Q(X) : -\text{prevent}(X,\text{osteo.}) \)

Find Proof Trees

Construct Markov Network

Approximate Probabilistic Inference

Answers

Expand MN?
What vegetables prevent osteoporosis?

- **Query:** \( \text{IsA}(X, \text{“vegetable”}) \land \text{Prevents}(X, \text{“osteoporosis”}) \)

- **Facts:** SVO relations extracted from the web. E.g. \( \text{Prevents(“milk”, “osteoporosis”) \right) \)

- **Rule:** \( \text{Contains}(X,Y) \land \text{Prevents}(Y,Z) \Rightarrow \text{Prevents}(X,Z) \)
**What vegetables prevent osteoporosis?**

Kale matches the query
(Inferred: 0.6)

- \( \phi_4(x_4) \)
- kale
  - IsA vegetable
    - (WordNet: 0.9)
  - prevents osteoporosis
    - (Inferred: 0.7)
    - \( \phi_3(x_3) \)
    - kale
      - contains calcium
        - (TextRunner: 0.8)
      - prevents osteoporosis
        - (TextRunner: 0.9)

Cabbage matches the query
(Inferred: 0.8)

- \( \phi_1(x_1) \)
- cabbage
  - IsA vegetable
    - (WordNet: 0.9)
  - prevents osteoporosis
    - (Inferred: 0.9)
    - \( \phi_2(x_2) \)
    - cabbage
      - contains calcium
        - (TextRunner: 0.5)
Holmes’s scalability

Knowledge Bases

Inf. Rules

Query

Find Proof Trees

Construct Markov Network

Approximate Probabilistic Inference

Answers

KBMC only includes relevant facts. But Horn inference is polynomial

Trade off speed vs. time and accuracy of proof trees
Holmes’s scalability in practice
Why is inference expensive?

- $Q(X,Y,Z) \leq \text{Married}(X,Y) \land \text{LivedIn}(Y,Z)$

- Worst Case: Some person $y'$ married everyone, and lived in every place:
  - $|Q(X,y',Z)| = |\text{Married}| \times |\text{LivedIn}| = O(n^2)$
Why is inference expensive?

\[ Q(X,Y,Z) \leq\text{ Married}(X,Y) \land \text{ LivedIn}(Y,Z) \]

Common Case:
Relation like a function
A few spouses and a few locations.
Approximately Pseudo-Functional Relations

- E.g. Married$(X,Y)$ Most $Y$ have only 1 spouse mentioned
- People in $y_G$ have at most a constant $k_M$ spouses each
- People in $y_B$ have at most $k_M \log |y_G|$ spouses in total
Main Theorem

Under a few assumptions, Holmes’s inference over Approximately Pseudo-Functional (APF) relations scales linearly in the size of the corpus.
Prevalence of APF relations

APF degrees of 500 random relations extracted from text
Experimental setup

- 20 example queries in 3 domains
- KBs
  - WordNet (Fellbaum, 1998)
  - TextRunner - 180M SVO relations extracted from 117M web pages. (Banko, et al., 2007)
- 6 domain independent inference rules
  (MacCartney and Manning, 2007)
  + 2 domain specific rules, in one domain
- Compared with “no inference” baseline
Results

The diagram shows the increase in AuC (Area Under the Curve) for different categories: Geography, Business, and Nutrition. The y-axis represents the estimated recall, while the x-axis shows the estimated recall.

- **Geography**: Increase in AuC is +102% with a total inference time of 55 seconds.
- **Business**: Increase in AuC is +2,634% with a total inference time of 145 seconds.
- **Nutrition**: Increase in AuC is +5,595% with a total inference time of 64 seconds.

The line graph compares baseline and Holmes approaches, indicating improvement in recall with the Holmes method.
Future work

- Overcome extractor errors
  - “Raspberries contain ellagic acid” -> contain(raspberries, acid)
  - “Folic acid prevents anemia” -> prevents(acid, anemia)
  - :: prevents(raspberries, anemia)

- Polysemous words
  - Florida developed a program -> Florida developed software

- More expressive representation

- Learn inference rules
Conclusions

- Introduced Holmes
  - Textual inference over 117 million Web pages

- Approximately Pseudo-Functional Property
  - Guarantees scalable inference

- Empirical results
  - APF property common in text relations
  - Inference can significantly improve recall
Thanks! Questions?

*See me for a live demo of HOLMES*
Inference Rules

- Similar to rules of (MacCartney & Manning, 2007)
- Observed relations are probably true
  - \( R(X,Y) :- \) Observed_In_Corpus(X, R, Y)
- Synonym substitution preserves meaning
  - \( R(X',Y) :- R(X,Y) \land \text{Synonym}(X,X') \)
  - \( R(X,Y') :- R(X,Y) \land \text{Synonym}(Y,Y') \)
- Generalizations preserve meaning
  - \( R(X',Y) :- R(X,Y) \land \text{IsA}(X,X') \)
  - \( R(X,Y') :- R(X,Y) \land \text{IsA}(Y,Y') \)
- Transitivity of Part Meryonyms
  - \( R(X,Y') :- R(X,Y) \land \text{PartOf}(Y,Y') \) (if \( R \) matches “* in”)

25
Queries

- **Geography**
  - E.g. ‘Who was born in Germany?’

- **Business**
  - E.g. ‘What companies are acquiring software companies?’

- **Nutrition**
  - E.g. ‘What vegetables prevent osteoporosis?’
Join Size

\[ Q(X,Y,Z) \leftarrow \text{Married}(X,Y) \bowtie \text{LivedIn}(Y,Z) \]

Each person \( Y \) falls into either \( Y_G \) or \( Y_B \) for each relation

\[ |\text{Married}(X,Y_{BM}) \bowtie \text{LivedIn}(Y_{BL},Z)| \leq k_M k_L \sqrt{|Y_{GM}| \sqrt{|Y_{GL}|}} = O(|Y_{Gmax}|) \]

\[ |\text{Married}(X,Y_{GM}) \bowtie \text{LivedIn}(Y_{BL},Z)| \leq k_M k_L \sqrt{|Y_{GM}|} = O(|Y_{Gmax}|) \]

\[ |\text{Married}(X,Y_{BM}) \bowtie \text{LivedIn}(Y_{GL},Z)| \leq k_M k_L \sqrt{|Y_{GM}|} = O(|Y_{Gmax}|) \]

\[ |\text{Married}(X,Y_{GM}) \bowtie \text{LivedIn}(Y_{GL},Z)| \leq k_M k_L \max(|Y_{GM}|, |Y_{GL}|) = O(|Y_{Gmax}|) \]

\[ \therefore |Q| \text{ is linear in the size of the relations} \]
Approximately Pseudo-Functional

- E.g. \texttt{Married}(X,Y) has 1 X for each Y
- People in \( y_G \) have at most a constant \( k_M \) spouses each
- People in \( y_B \) have at most \( k_M^*\sqrt{|y_G|} \) spouses in total

![Graph showing Good Set (\( y_G \)) and Bad Set (\( y_B \))](image-url)
A relation $R = \{(x, y)\} \subseteq \mathcal{X} \times \mathcal{Y}$

A relation $R$ is **Approximately Pseudo-Functional** in $\mathcal{X}$ with degree $k$ if $k$ is a constant such that $\mathcal{X}$ can be partitioned into two sets, $\mathcal{X}_G$ and $\mathcal{X}_B$, such that:

\[ \forall x \in \mathcal{X}_G : |\{ y | (x, y) \in R \}| \leq k \]

\[ \sum_{x \in \mathcal{X}_B} |\{ y | (x, y) \in R \}| \leq k \log |R| \]

**Theorem**: For a fixed query, set of inference rules, and bounded proof tree size, if all relations involved are APF, then Horn inference scales linearly in the corpus size.
Scaling Inference: Assumptions

- Number of extractions scales linearly with $|\text{Corpus}|$
  - $|R|$ is $O(|\text{Corpus}|)$
- Query and inference rules are fixed
- All proof trees are bounded by some maximum size $M$