1. Motivation

- Much of the information conveyed in natural language text must be inferred

Example - “Barack Obama is the President of the United States of America.”
Query - “Barack Obama is the citizen of what country?”

- Human readers read between the lines to infer implicit facts using
  - Explicitly stated facts
  - Common-sense knowledge
- Information extraction systems cannot perform deeper inference to extract implicit facts

2. Our Approach

- Extract explicitly stated facts using an off-the-shelf information extraction (IE) system
- Learn common sense knowledge in the form of first-order rules
- Use Bayesian Logic Programs [Kersting and De Raedt, 2007] for inference of additional facts

3. Online Rule Learner

Example text in training
“Barack Obama is the 44th and the current President of USA. Obama, citizen of USA was born on August 4, 1961 in Hawaii, USA.”

IE extractions
nationState(USA)
person(BarackObama)
isLedBy(USA,BarackObama)
hasBirthPlace(BarackObama,USA)
hasCitizenship(BarackObama,USA)

Frequency counts for relation predicates
isLedBy: 30, hasBirthPlace: 23, hasCitizenship: 20

First-order rules constructed by ORL
isLedBy(X,Y) ^ person(Y) ^ nationState(X) \rightarrow hasBirthPlace(Y,X)

hasBirthPlace(X,Y) ^ person(X) ^ nationState(Y) \rightarrow hasCitizenship(X,Y)

4. Data and Experiments

- DARPA’s intelligence community (IC) data set from the Machine Reading Project (MRP)
  - News articles on politics, terrorism, and other international events
- Learned rules for 10 target relations
- Manually evaluated inferences and sorted based on their marginal probabilities

5. Results

Running time
ORL – 3.8 mins, LIME -11.23 hours