Fine Grained Weight Learning in Markov Logic Networks

Happy Mittal : IIT Delhi
Shubhankar Singh : IIT Delhi
Vibhav Gogate : UT Dallas
Parag Singla : IIT Delhi
Markov Logic Networks

1.5 \( \forall x \text{ Smokes}(x) \Rightarrow \text{Cancer}(x) \)

1.1 \( \forall x, y \text{ Friends}(x, y) \land \text{Smokes}(x) \Rightarrow \text{Smokes}(y) \)

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

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

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

\[
P(y \mid x; w) = \frac{1}{Z_x} \exp \left( \sum_i w_in_i(x, y) \right)
\]

- \( y \) : query atoms
- \( X \) : evidence atoms
- Weight of formula \( i \)
- No. of true groundings of formula \( i \) in \( x \)
Motivation

\[ \forall x \ Smokes(x) \Rightarrow Cancer(x) \]
Motivation

\[ \forall x \, \text{Smokes}(x) \Rightarrow \text{Cancer}(x) \]

Per Constant Learning

\[ w_1 : \text{Smokes}(P_1) \Rightarrow \text{Cancer}(P_1) \]
\[ w_2 : \text{Smokes}(P_2) \Rightarrow \text{Cancer}(P_2) \]
\[ \vdots \]
\[ w_n : \text{Smokes}(P_n) \Rightarrow \text{Cancer}(P_n) \]

High Variance
Motivation

\[ \forall x \text{ Smokes}(x) \Rightarrow \text{Cancer}(x) \]

Per Constant Learning

\[ w_1 : \text{Smokes}(P_1) \Rightarrow \text{Cancer}(P_1) \]
\[ w_2 : \text{Smokes}(P_2) \Rightarrow \text{Cancer}(P_2) \]
\[ \vdots \]
\[ w_n : \text{Smokes}(P_n) \Rightarrow \text{Cancer}(P_n) \]

High Variance

First Order Learning

\[ w : \text{Smokes}(x) \Rightarrow \text{Cancer}(x) \]
\[ \Delta x = \{P_1, P_2, \ldots, P_n\} \]

High Variance
Motivation

\[ \forall x \text{ Smokes}(x) \Rightarrow \text{Cancer}(x) \]

**Per Constant Learning**
- \( w_1 : \text{Smokes}(P_1) \Rightarrow \text{Cancer}(P_1) \)
- \( w_2 : \text{Smokes}(P_2) \Rightarrow \text{Cancer}(P_2) \)
- \( \vdots \)
- \( w_n : \text{Smokes}(P_n) \Rightarrow \text{Cancer}(P_n) \)

**Fine Grained Learning**
- \( w_1 : \text{Smokes}(P_1) \Rightarrow \text{Cancer}(P_1) \)
- \( \Delta s_1 = \{P_1, P_2, \ldots, P_i\} \)
- \( w_2 : \text{Smokes}(s_2) \Rightarrow \text{Cancer}(s_2) \)
- \( \Delta s_2 = \{P_{i+1}, \ldots, P_{i+k}\} \)
- \( \vdots \)
- \( w_k : \text{Smokes}(s_k) \Rightarrow \text{Cancer}(s_k) \)
- \( \Delta s_k = \{P_{k+1}, \ldots, P_{k+l}\} \)

**First Order Learning**
- \( w : \text{Smokes}(x) \Rightarrow \text{Cancer}(x) \)
- \( \Delta x = \{P_1, P_2, \ldots, P_n\} \)

**High Variance**

**Right tradeoff b/w bias and variance**

**High Variance**
Objective

- Automatically finding the hidden attributes.
- Partition the ground formulas.
- Learn separate weight for each partition.
Partitioning the Ground Formulas

• Partition the underlying constants into subtypes (hidden attribute).
• Partition ground formulas with same subtype signatures.

\[ \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]

\[ \Delta x = \Delta y = \{A, B, C\} \]
Partitioning the Ground Formulas

- Partition the underlying constants into subtypes (hidden attribute).
- Partition ground formulas with same subtype signatures.

\[
\text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y)
\]
\[
\Delta x = \Delta y = \{A, B, C\}
\]

Partitioning of constants:

\[
\text{sub}(A) = \text{sub}(B) = s_1
\]
\[
\text{sub}(C) = s_2
\]
Partitioning the Ground Formulas

- Partition the underlying constants into subtypes (hidden attribute).
- Partition ground formulas with same subtype signatures.

\[ \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]
\[ \Delta x = \Delta y = \{A, B, C\} \]

\[ \text{sub}(A) = \text{sub}(B) = s_1 \]
\[ \text{sub}(C) = s_2 \]

\[ w_1 : \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]
\[ \Delta x = \Delta y = \{A, B\} \]

\[ w_2 : \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]
\[ \Delta x = \{A, B\}, \Delta y = \{C\} \]

\[ w_3 : \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]
\[ \Delta x = \{C\}, \Delta y = \{A, B\} \]

\[ w_4 : \text{Smokes}(x) \land \text{Friends}(x, y) \Rightarrow \text{Smokes}(y) \]
\[ \Delta x = \Delta y = \{C\} \]
Weight Learning

- Probability Distribution of modified MLN

\[ P(y, s \mid x; w) = \frac{1}{Z_x} \exp \left( \sum_{i=1}^{k} \sum_{j=1}^{g_{fi}} w_i^j n_i^j (x, y, s) \right) - (1) \]

- Two cases:
  - Case 1: Subtypes are known
    - Learn discriminatively by (1)
  - Case 2: Subtypes are not known
    - Model the subtypes as hidden predicates.
    - Two methods to deal with this case.
Method 1 : K-means clustering

• Cluster the constants.
• Feature vector : Avg. weight of ground formulas constant appears in.
Method 2: Joint Learning of Subtypes

- Maximize log-likelihood by summing over hidden subtypes.

\[
P(y \mid x; w) = \sum_s \frac{1}{Z_x} \exp \left( \sum_{i=1}^{k} \sum_{j=1}^{g_f_i} w_{i,j} n_{i,j} (x, y, s) \right)
\]

- Use EM algorithm.
Experiments & Results

• Experimented on IMDB dataset.
THANK YOU