Syntagmatic Relation Discovery: Conditional Entropy

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Syntagmatic Relation Discovery: Conditional Entropy

5. Text-based prediction

3. Topic mining and analysis

Real World

Observed World

Text Data

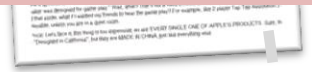
Perceive

(Perspective)

Express

(English)

1. Natural language processing and text representation



4. Opinion mining and sentiment analysis

2. Word association mining and analysis

What If We Know More About a Text Segment?

Prediction question: Is "**meat**" present (or absent) in this segment?

Does presence of "**eats**" help predict the presence of "**meat**"?
Does it **reduce** the uncertainty about "meat", i.e., **H(X_{meat})?**

What if we know of the absence of "eats"? Does it also help?

Conditional Entropy

Know nothing about the segment → **Know "eats" is present (X_{eats} =1)**

$$p(X_{meat} = 1) \dashrightarrow p(X_{meat} = 1 \mid X_{eats} = 1)$$

$$p(X_{meat} = 0) \dashrightarrow p(X_{meat} = 0 \mid X_{eats} = 1)$$

$$H(X_{meat}) = -p(X_{meat} = 0)\log_2 p(X_{meat} = 0) - p(X_{meat} = 1)\log_2 p(X_{meat} = 1)$$

$$H(X_{meat} \mid X_{eats} = 1) = -p(X_{meat} = 0 \mid X_{eats} = 1)\log_2 p(X_{meat} = 0 \mid X_{eats} = 1)$$
$$- p(X_{meat} = 1 \mid X_{eats} = 1)\log_2 p(X_{meat} = 1 \mid X_{eats} = 1)$$

$H(X_{meat} \mid X_{eats} = 0)$ **can be defined similarly**

Conditional Entropy: Complete Definition

$$H(X_{meat} \mid X_{eats}) = \sum_{u \in \{0,1\}} [p(X_{eats} = u)\ H(X_{meat} = v \mid X_{eats} = u)]$$

$$= \sum_{u \in \{0,1\}} [p(X_{eats} = u) \sum_{v \in \{0,1\}} [-p(X_{meat} = v \mid X_{eats} = u) \log_2 p(X_{meat} = v \mid X_{eats} = u)]]$$

In general, for any discrete random variables X and Y, we have H(X) ≥ H(X|Y)

What's the minimum possible value of H(X|Y)?

Conditional Entropy to Capture Syntagmatic Relation

$$H(X_{meat} \mid X_{eats}) = \sum_{u \in \{0,1\}} [p(X_{eats} = u)\ H(X_{meat} = v \mid X_{eats} = u)]$$

$$H(X_{meat} \mid X_{meat}) = ?$$

Which is smaller? $H(X_{meat} \mid X_{the})$ or $H(X_{meat} \mid X_{eats})$?

For which word w, does $H(X_{meat} \mid X_w)$ reach its minimum (i.e., 0)?

For which word w, does $H(X_{meat} \mid X_w)$ reach its maximum, $H(X_{meat})$?

Conditional Entropy for Mining Syntagmatic Relations

- For each word W1
  - For every other word W2, compute conditional entropy $H(X_{W1}|X_{W2})$
  - Sort all the candidate words in ascending order of $H(X_{W1}|X_{W2})$
  - Take the top-ranked candidate words as words that have potential syntagmatic relations with W1
  - Need to use a threshold for each W1
- However, while $H(X_{W1}|X_{W2})$ and $H(X_{W1}|X_{W3})$ are comparable, $H(X_{W1}|X_{W2})$ and $H(X_{W3}|X_{W2})$ aren't!

How can we mine the strongest K syntagmatic relations from a collection?