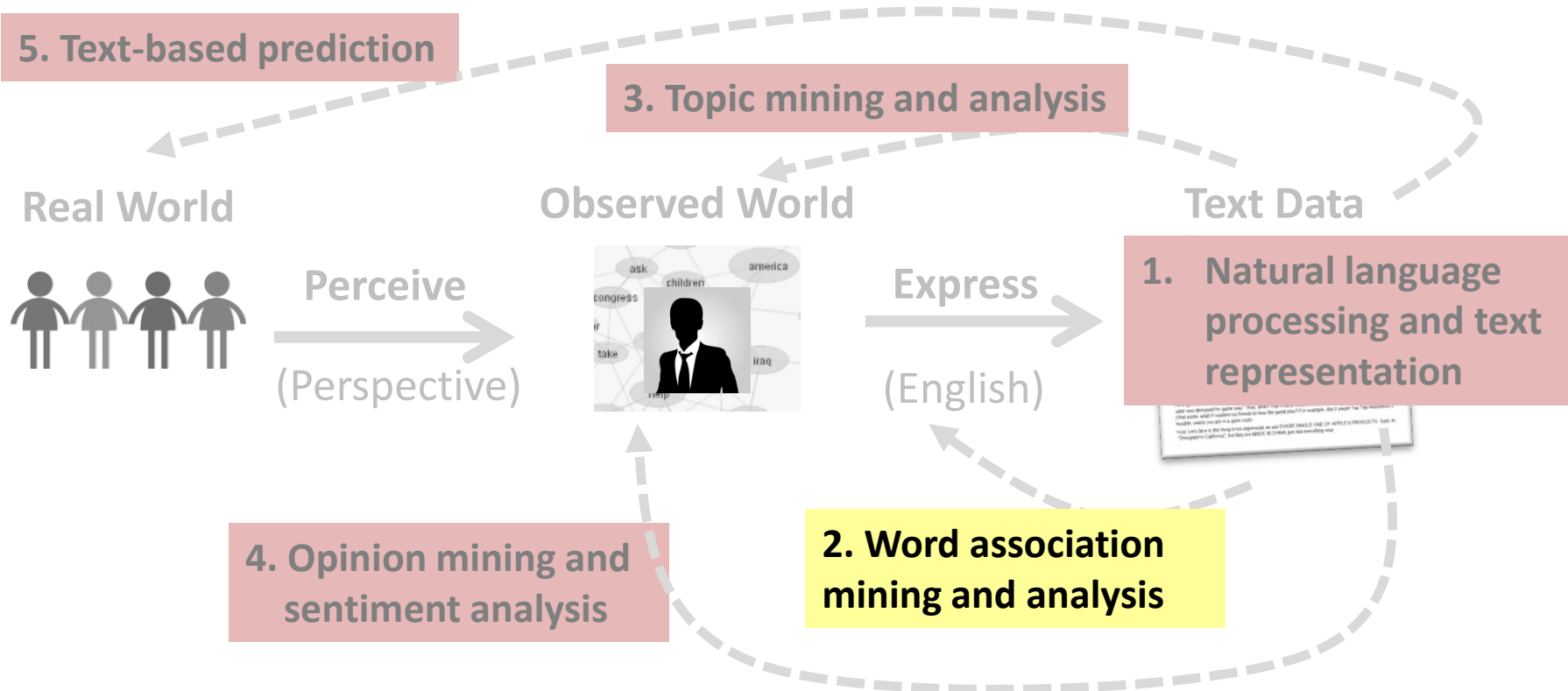


Syntagmatic Relation Discovery: Conditional Entropy

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

Syntagmatic Relation Discovery: Conditional Entropy



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., $H(X_{\text{meat}})$?

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

Conditional Entropy 条件熵

Know nothing about the segment

Know "eats" is present ($X_{eats} = 1$)

$$p(X_{meat} = 1) \quad \text{-----} \rightarrow \quad p(X_{meat} = 1 | X_{eats} = 1)$$

$$p(X_{meat} = 0) \quad \text{-----} \rightarrow \quad p(X_{meat} = 0 | X_{eats} = 1)$$

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



在熵中加入条件概率

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

$H(X_{meat} / X_{eats} = 0)$ can be defined similarly

Conditional Entropy: Complete Definition

$$\begin{aligned} H(X_{meat} / X_{eats}) &= \sum_{u \in \{0,1\}} [p(X_{eats} = u) H(X_{meat} | X_{eats} = u)] \\ &= \sum_{u \in \{0,1\}} [p(X_{eats} = u) \sum_{v \in \{0,1\}} [-p(X_{meat} = v | X_{eats} = u) \log_2 p(X_{meat} = v | X_{eats} = u)]] \end{aligned}$$

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

知道的附加条件越多，熵越小；越难预测

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

Conditional Entropy to Capture Syntagmatic Relation

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

$$H(X_{meat} | X_{meat}) = ? \quad 0 \quad \text{已知 } X_{meat}, \text{ predict } X_{meat}, \text{ 熵为0.}$$

知道 the, 并不会
知道 meat 很多可能性

Which is smaller? $H(X_{meat} | X_{the})$ or $H(X_{meat} | X_{eats})$? 后者 熵更小, 更易预测

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

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

$w =$ 跟 meat 不相关的词, 比如 the.

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 $H(X_{W1} | X_{W2})$ 倒序排列
 - Take the top-ranked candidate words as words that have potential syntagmatic relations with $W1$ 词频最小, 最相关
 - 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!
无法比较! 它们的 upperbound 不同.

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