

## 5. 初歩の確率

内山将夫@NICT  
mutiyama@nict.go.jp

# 初歩の確率

- 確率の例と用語
- 連鎖規則 (Chain Rule)
- ベイズの定理

## 確率の例

- 硬貨を1回投げたとき，表がでる確率

$$\rightarrow \frac{1}{2}$$

- サイコロを投げたとき，1の目がでる確率

$$\rightarrow \frac{1}{6}$$

- 52枚のカードから1枚をひいたとき，エースがでる確率

$$\rightarrow \frac{4}{52} = \frac{1}{13}$$

## 確率の例 (言語モデル)

- 「今日は良い」という文字列の後に「天気」「日」「こと」が続く確率
- 検索エンジンで「今日は良い」を検索すると，426000 ヒット
- 「今日は良い天気」は，  
 $\rightarrow \frac{149000}{426000} = 0.35$
- 「今日は良い日」は，  
 $\rightarrow \frac{42200}{426000} = 0.1$
- 「今日は良いこと」は，  
 $\rightarrow \frac{19200}{426000} = 0.05$

任意の文字列(単語列)について，確率を割当てるモデルを「言語モデル」という。

## 確率の例（翻訳モデル）

「心」の訳語としては、「mind」「heart」「spirit」のどれが良く使われるか？

- 「心」と「mind」を両方含むページ → 1440000
- 「心」と「heart」を両方含むページ → 1570000
- 「心」と「spirit」を両方含むページ → 1360000

合計 = 4370000

3つが互いに排他的であるとして、

- 「心」と「mind」 → 0.33
- 「心」と「heart」 → 0.36
- 「心」と「spirit」 → 0.31

→ 日本語の単語列の翻訳確率を与えるモデルを「翻訳モデル」という

## 確率の例(確率による翻訳)

翻訳モデルと言語モデルを利用して訳を得る .

翻訳確率が

- 「脳」 → brain = 1.0 , 「と」 → and = 1.0 , 「心」 → mind = 1/3, heart = 1/3, spirit = 1/3

のとき ,

## 翻訳モデル

- 脳と心 → (A) brain and mind = 1/3, (B) brain and heart = 1/3, (C) brain and spirit = 1/3

## 言語モデル

- (A)brain and mind のヒット数 = 315000 → 0.48
- (B)brain and heart のヒット数 = 336000 → 0.52
- (C)brain and spirit のヒット数 = 800 → 0.00

より「翻訳モデル × 言語モデル」による推定では , (B) が優勢である .

## 直接数える

- (A) “brain and mind” & “脳と心” → 82
- (B) “brain and heart” & “脳と心” → 2
- (C) “brain and spirit” & “脳と心” → 0

より (A) が優勢である . (推定と実際での食い違い)

## 確率の用語：条件付き確率

条件付き確率  $P(B|A)$

- A を条件としたときの B の確率

たとえば，

- $P(\text{天気} | \text{今日は良い}) = 0.35$
- 「心」の訳語として「mind」「heart」「spirit」しか考  
えないとき

$$P(\text{mind} | \text{心}) + P(\text{heart} | \text{心}) + P(\text{spirit} | \text{心}) = 1.0$$

また，

- $P(\text{mind} | \text{心}) = 0.33$
- $P(\text{heart} | \text{心}) = 0.36$
- $P(\text{spirit} | \text{心}) = 0.31$

## 連鎖規則

$$\begin{aligned} P(X_1, X_2, X_3, \dots, X_N) \\ = P(X_1) \\ \times P(X_2 | X_1) \\ \times P(X_3 | X_1, X_2) \\ \times P(X_4 | X_1, X_2, X_3) \\ \dots \\ \times P(X_N | X_1, X_2, \dots, X_{N-1}) \end{aligned}$$

たとえば、

$$\begin{aligned} P(\text{今日, は, 良い, 天気, だ}) \\ = P(\text{今日}) \\ \times P(\text{は} | \text{今日}) \\ \times P(\text{良い} | \text{今日, は}) \\ \times P(\text{天気} | \text{今日, は, 良い}) \\ \times P(\text{だ} | \text{今日, は, 良い, 天気}) \end{aligned}$$

# 条件付き独立

## 1-gram モデル

$$\begin{aligned} P(X_1, X_2, X_3, \dots, X_N) \\ = P(X_1)P(X_2)P(X_3)P(X_4) \cdots P(X_N) \end{aligned}$$

## 2-gram モデル

$$\begin{aligned} P(X_1, X_2, X_3, \dots, X_N) \\ = P(X_1) \\ \times P(X_2 | X_1) \\ \times P(X_3 | X_2) \\ \times P(X_4 | X_3) \\ \cdots \\ \times P(X_N | X_{N-1}) \end{aligned}$$

## 3-gram モデル

$$\begin{aligned} P(X_1, X_2, X_3, \dots, X_N) \\ = P(X_1) \\ \times P(X_2 | X_1) \\ \times P(X_3 | X_1, X_2) \\ \times P(X_4 | X_2, X_3) \\ \cdots \\ \times P(X_N | X_{N-2}, X_{N-1}) \end{aligned}$$

## ベイズの定理

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

「脳と心」が入力、「brain and mind」が翻訳候補

$$P(\text{brain and mind}|\text{脳と心}) \\ = \frac{P(\text{脳と心}|\text{brain and mind})P(\text{brain and mind})}{P(\text{脳と心})}$$

言語モデルと翻訳モデルの組み合わせにより確率計算をする。

## 言語モデル

$$P(\text{brain and mind}) = P(\text{brain})P(\text{and}|\text{brain})P(\text{mind}|\text{and})$$

## 翻訳モデル

$$P(\text{脳と心}|\text{brain and mind}) \\ = P(\text{脳}|\text{brain}) \\ \times P(\text{と}|\text{and}) \\ \times P(\text{心}|\text{mind})$$

## まとめ

- 言語現象に確率を割当ることができる
- それを利用して，翻訳確率を計算できる