

練習問題2

Mr. O、Ms. H、Mr. Tの3氏の、最近2週間のつぶやきを、解析対象を10語に絞って調べたところ、以下の出現があった。

O氏(つぶやき数10): USA great Democrats care mother

H氏(つぶやき数15): mother love Democrats care mother USA

T氏(つぶやき数25): Russia fake USA Mexico great Mexico haters

「USA Democrats mother」を含む新たなつぶやきが発信された。上のデータと単純ベイズ分類器の考え方をもとに、この新たなつぶやきが、3氏のうちの誰のものと考えられるか予測せよ。なお、各氏の最近2週間のつぶやき数を、事前確率とすること。また、全度数の初期値を1とするLaplace smoothingを用いよ。

練習問題2改

Mr. O、Ms. H、Mr. Tの3氏の、最近2週間のつぶやきを、解析対象を10語に絞って調べたところ、以下の出現があった。

O氏(つぶやき数10): USA great Democrats care mother

H氏(つぶやき数15): mother love Democrats care mother USA

T氏(つぶやき数25): Russia fake USA Mexico great Mexico haters

「USA Democrats mother **mother**」ならどのように計算？

練習問題2改の解答例

dm-04-practice2-ans-suppl.ipynb

1. 事前確率

- $P(O) = 10 / (10+15+25) = 1 / 5 = 0.2$
- $P(H) = 15 / (10+15+25) = 3 / 10 = 0.3$
- $P(T) = 25 / (10+15+25) = 1 / 2 = 0.5$

2. 尤度

- $P(\text{USA}|O) \times P(\text{Democrats}|O) \times P(\text{mother}|O) \times P(\text{mother}|O) = 2/15 \times 2/15 \times 2/15 \times 2/15$
- $P(\text{USA}|H) \times P(\text{Democrats}|H) \times P(\text{mother}|H) \times P(\text{mother}|H) = 2/16 \times 2/16 \times 3/16 \times 3/16$
- $P(\text{USA}|T) \times P(\text{Democrats}|T) \times P(\text{mother}|T) \times P(\text{mother}|T) = 2/17 \times 1/17 \times 1/17 \times 1/17$

練習問題2改の解答例

3. 事後確率 = 尤度 x 事前確率

$$\blacksquare P(O|\text{単語群}) = 2/15 \times 2/15 \times 2/15 \times 2/15 \times 0.2 = 0.000063$$

$$\blacksquare P(H|\text{単語群}) = 2/16 \times 2/16 \times 3/16 \times 3/16 \times 0.3 = 0.000165$$

$$\blacksquare P(T|\text{単語群}) = 2/17 \times 1/17 \times 1/17 \times 1/17 \times 0.5 = 0.000012$$

4. 割合

$$\blacksquare P(O|\text{単語群}) / (P(O|\text{単語群}) + P(H|\text{単語群}) + P(T|\text{単語群})) = 0.2633\cdots$$

$$\blacksquare P(H|\text{単語群}) / (P(O|\text{単語群}) + P(H|\text{単語群}) + P(T|\text{単語群})) = 0.6867\cdots$$

$$\blacksquare P(T|\text{単語群}) / (P(O|\text{単語群}) + P(H|\text{単語群}) + P(T|\text{単語群})) = 0.0498\cdots$$

5. 最大の事後確率に対応するのは Ms. H

※この計算は、DM-05 の naive_bayes_small.ipynb を使うことでも確認できる。