

# 言語処理のための機械学習入門

第六章 実験の仕方など

6. 3 ~ 6. 4

10NM733X 林華

## 6. 3 評価指標

- 6.3.1 分類正解率
- 6.3.2 精度と再現率
- 6.3.3 精度と再現率の統合
- 6.3.4 多クラスデータを用いる場合の実験設定
- 6.3.5 評価指標の平均

## 6.3.1 分類正解率

分類正解率：最も直感的な評価指標であり、正解した割合を表す。  
単に正解率とも呼ばれる。

$$\text{分類正解率} = \frac{\text{正解した評価事例数}}{\text{評価事例数}}$$

## 6.3.2 精度と再現率

分割表（分類結果を示す）

	Cに属する	Cに属しない
Cであると予測	a	b
Cでないと予測	c	d

注：精度と再現率は、ある特定のクラスに対して定義された値であり、そのクラスであるか否かという二値分類の評価指標である。

精度：分類器がCであると判断された結果の正確率

再現率：実際にCであるものを分類器が正しく判断できた正解率

$$\text{精度} = \frac{a}{a+b}, \quad \text{再現率} = \frac{a}{a+c}$$

### 6.3.3 精度と再現率の統合

問題点：精度と再現率だけでは比較しにくい — 両者が反比例関係。

解決策：なるべく両者の平均を取る。

➤ F値：精度と再現率の調和平均

$$F\text{値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}}$$

➤ 再現率/精度break-evenポイント：精度と再現率等しくなる点

➤ 11点平均精度：再現率がそれぞれ0.0, 0.1, ..., 1.0となる11点における精度の平均

## 6.3.4 多クラスデータを用いる場合の実験設定

- ◆ 一つの多値分類問題として実験・評価する。分類正解率を用いる。
  - 各評価事例の複数のラベルが全て予測できたときに正解
  - 各評価事例の複数のラベルのうち一つでも予測できれば正解
- ◆ 各クラスについて、二値分類問題として考える。この場合、クラス数の分類器を構築する必要があり、各評価事例がクラス数回分類される。

## 6.3.5 評価指標の平均

クラスC1			複合した分類表		
C1であると予測	a1	b1	各クラスに属 する	属さない	各クラスである と予測
C1でないと予測	c1	d1			
クラスC2					
C1であると予測	a2	b2	各クラスである と予測	属さない	各クラスである と予測
C1でないと予測	c2	d2			
クラスC3					
C1であると予測	a3	b3	各クラスである と予測	属さない	各クラスである と予測
C1でないと予測	c3	d3			

- マクロ平均：各クラスについて評価指標を計算し、それらの平均
- マイクロ平均：全てクラスで実験結果を一つの分割表にまとめ、評価指標を計算

$$\begin{aligned}
 Prec_{mi} &= \frac{a_1 + a_2 + a_3}{(a_1 + a_2 + a_3) + (b_1 + b_2 + b_3)} & Rec_{mi} &= \frac{a_1 + a_2 + a_3}{(a_1 + a_2 + a_3) + (c_1 + c_2 + c_3)} \\
 Ac_{ma} &= \frac{1}{3} \left( \frac{a_1 + d_1}{a_1 + b_1 + c_1 + d_1} + \frac{a_2 + d_2}{a_2 + b_2 + c_2 + d_2} + \frac{a_3 + d_3}{a_3 + b_3 + c_3 + d_3} \right) \\
 Ac_{mi} &= \frac{(a_1 + d_1) + (a_2 + d_2) + (a_3 + d_3)}{(a_1 + b_1 + c_1 + d_1) + (a_2 + b_2 + c_2 + d_2) + (a_3 + b_3 + c_3 + d_3)}
 \end{aligned}$$

## 6.4 検定

統計的検定：実験結果をどの程度信頼してよいかを測るもの。

	既存手法	提案手法
文書1	○	○
文書2	×	○
文書3	○	×
文書4	○	○
文書5	×	○
分類正解率	0.6	0.8

### 検定プロセス

- ① 「二つの手法の評価指標が同じ確立分布に従う」と仮説し、最後に否定一帰無仮説
- ② 仮定の基に、この実験結果が出る確率を計算－p値
- ③  $p = 0.5$ を有意水準とし、 $p$ 値 $< 0.05$ なら、この実験結果がまれで、仮説が間違っていると判断－この実験結果が偶然ではない－統計的に有意

### p値の計算法

- a. 符号検定
  - b. ウィルコクソンの符号付順位和検定
- t－検定