

多値分類問題における F_1 score の信頼区間の提案

大阪市立大学 高橋佳苗
国立がん研究センター 口羽文
横浜市立大学 山本紘司

F_1 score は分類問題におけるモデルの性能を評価する指標の 1 つである。情報検索分野で用いられることが多いが、医療におけるスクリーニング検査の性能評価など、他の様々な分野でも利用されている。 F_1 score は同じく分類問題のモデル評価指標である再現率 (recall: R)、適合率 (precision: P) から求めることができる¹。一方、データを r カテゴリ ($r \geq 2$) に分類する場合、分類結果は行変数を予測結果、列変数を真の結果とした $r \times r$ 混同行列で表される。これを各カテゴリ i ごとにまとめると、表 1 の形の混同行列で表すことができる。

表 1: カテゴリ i の分類結果

		真の結果	
		class i	other
予測結果	class i	True Positive(TP_i)	False Positive(FP_i)
	other	False Negative(FN_i)	True Negative(TN_i)

表 1 をもとに、全カテゴリを併合した再現率 (R_{micro})、適合率 (P_{micro}) から micro-averaged F_1 score (F_{1micro}) を求めることができる²。

$$R_{micro} = \frac{\sum_{i=1}^r TP_i}{\sum_{i=1}^r TP_i + FN_i}, \quad P_{micro} = \frac{\sum_{i=1}^r TP_i}{\sum_{i=1}^r TP_i + FP_i}$$
$$F_{1micro} = \frac{2R_{micro}P_{micro}}{R_{micro} + P_{micro}}$$

また、各カテゴリの再現率、適合率の平均 (R_{macro} , P_{macro}) を用いた macro-averaged F_1 score (F_{1macro}) も以下の式より求めることができる²。

$$R_{macro} = \frac{1}{r} \sum_{i=1}^r \frac{TP_i}{TP_i + FN_i}, \quad P_{macro} = \frac{1}{r} \sum_{i=1}^r \frac{TP_i}{TP_i + FP_i}$$
$$F_{1macro} = \frac{2R_{macro}P_{macro}}{R_{macro} + P_{macro}}$$

F_{1micro} および F_{1macro} は多値分類問題のモデル評価指標として用いられているが、その推定精度についてはあまり着目されていない。そこで本研究では、 F_{1micro} および F_{1macro} の信頼区間を新たに提案することとした。

参考文献

1. C. J. van Rijsbergen. Information Retrieval, 2nd edition. Butterworth-Heinemann, 1979.
2. C. D. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.