

Метод обґрунтування способу оцінювання якості моделі машинного навчання

Ольга Соловей, к.т.н., доцент ¹ (ORCID: 0000-0001-8774-7243)

¹ Київський національний університет будівництва і архітектури, 03037, м. Київ, проспект Повітряних Сил, 31, Україна

АНОТАЦІЯ

Методи машинного навчання застосовуються для створення аналітичних та прогнозних моделей для проєктів міського будівництва при цьому, ефективність такого підходу базується на якості побудованої моделі машинного навчання. Процес доведення ефективності запропонованої моделі машинного навчання значно ускладнюється, якщо не існує узгодженості серед дослідників, щодо прийнятних методів оцінки. Мета даної роботи є обґрунтування вибору способу оцінювання моделі машинного навчання. Для досягнення мети в роботі досліджено динаміку зміни різниці між значеннями оцінок моделі, які отримані різними методами залежно від значення порогу на ROC-кривій. Визначені функції, які визначають найбільшу можливу за модулем різницю між значеннями оцінок моделі, які отримані різними методами. Рекомендовано підхід для обґрунтування вибору способу оцінювання моделі на основі аналізу найбільшої можливої різниці.

Ключові слова: Youden J індекс, збалансована точність, F-міра, коефіцієнт кореляції Меттьюза, ROC-крива

1. ВСТУП

Моделі машинного навчання для набору даних, де всі спостереження належать або позитивному класу «1», або негативному – «0» можна оцінити за загально прийнятими оцінками, а саме: чутливість (SEN) і специфічність (SPE) – метрики для оцінки здатності моделі правильно ідентифікувати позитивні та негативні випадки відповідно. Позитивне прогнозне значення (PPV) і негативне прогнозне значення (NPV) – метрики оцінюють похибку моделі для позитивних і негативних випадків відповідно [1].

Значення інших методів оцінювання отримується функціями:

- від двох параметрів (SEN та SPE), наприклад збалансована точність (далі, BA) – метод враховує здатність моделі вірно визначати випадки, але не враховує помилку моделі;

- від трьох параметрів (SEN, SPE та PPV), наприклад F-міра (далі F₁) - метод враховує здатність моделі вірно визначати випадки і помилку моделі для позитивних випадків, залишаючи поза увагою помилку моделі для негативних випадків;

- від чотирьох параметрів (SEN, SPE, PPV, NPV), наприклад коефіцієнт кореляції Меттьюза (MCC) - оцінка враховує здатність моделі вірно визначати випадки й помилки моделі для позитивних та негативних випадків.

Значення SEN, SPE, PPV та NPV обчислюються за матрицею невідповідностей (табл. 1), яка фіксує кількість спостережень для яких належність класу була вірно визначена моделлю (TP, TN), та кількість спостережень, належність яких до класів визначена не вірно моделлю (FP, FN).

Таблиця 1. Матриця невідповідностей 2-х класової моделі

	Передбачений позитивний	Передбачений негативний	
Вірно позитивний	TP	FN	P _i =TP+FN
Вірно негативний	FP	TN	N _i =FP+TN
	P ₁ =TP+FP	N ₁ =FN+TN	

В роботах [2-3] проведені дослідження щодо об'єктивності оцінок отриманих методами: збалансована точність, F-міра, нормалізований коефіцієнт кореляції Меттьюза (nMCC). В таблиці 2 наведені два сценарії з зазначених в роботах.

Таблиця 2. Оцінки за методами nMCC, BA, F₁

№	TP	FN	FP	TN	nMCC	BA	F ₁
1	100	20	1000	30000	0.63	0.9	0.2
2	90000	10000	1	9	0.51	0.9	0.9

В рядку 1 (табл.2) модель невірно класифікувала 20 об'єктів класу «1» і 1000 об'єктів класу «0», в наслідок чого отримано низьку прогностичну значущість позитивного результату $PPV=100/(100+1000)=0.09$; здатність моделі вірно класифікувати позитивні та негативні випадки залишалась високою (SEN =0.87; SPE = 0.97) і високою залишалась прогностична значущість негативного результату $NPV=30000/(30000+20)=0.99$. Очікуваний результат – оцінки ефективності моделі покажуть її середню здатність вірно визначати клас спостереження. Фактичний результат – значення збалансованої точності BA невірно оцінює модель.

В рядку 2 (табл.2) модель невірно класифікувала 10000 об'єктів класу «1», які було віднесено до класу «0» в наслідок чого отримали низьку прогностичну значущість негативного результату $NPV=9/(9+10000)=0.0008$, при цьому здатність моделі вірно класифікувати позитивні та негативні випадки і прогностична значущість позитивного результату залишалась високою (SEN =0.9; SPE = 0.99, PPV =0.99). Очікуваний результат – оцінки ефективності моделі покаже її нездатність визначать клас спостереження. Фактичний результат – значення збалансованої точності BA та F-міра невірно оцінює модель.

Порівнюючи фактичні та очікувані результати оцінок для сценаріїв (табл. 2) в роботах [2-3], запропоновано оцінювати моделі за методом коефіцієнта кореляції Меттьюза (MCC), і не використовувати збалансовану точність та F-міра.

В дані роботі, запропонуємо підхід для обґрунтування вибору методів оцінки моделі машинного навчання для двох класового набору даних, визначивши, що метод MCC завжди повертає об'єктивний результат.

2. ОСНОВНА ЧАСТИНА

Дослідження [2-3] були проведені для моделей, коли прийняття рішення щодо класу спостереження приймається за правилом: коли ймовірність з якою спостереження належить до класу більше або дорівнює порогу 0.5, тоді спостереження належить класу «1», в іншому випадку – класу «0». Це правило є розповсюдженим, але не оптимальним для всіх типів задач аналізу з застосуванням машинного навчання [4].

Youden J індекс (далі J) використовується для визначення оптимального порогу для алгоритмів прийняття рішень моделей машинного навчання [5]. Якщо SEN і SPE мають однаковий пріоритет, тоді Youden J індекс розраховується як максимальна різниця між TPR і FPR (1).

$$J = \max_{\tau} \{TRP - FPR\}, \quad (1)$$

де TPR, FPR – доли спостережень для яких модель вірно визначила належність класу (значення на осі ординат, рис. 1) та доля спостережень з невірно визначеним класом (значення на осі абсцис, рис. 1).

Верхня точка пунктирної лінії на рисунку 1, визначає оптимальний поріг (далі t_{opt}), нижня точка, фіксує значення FPR при якому Youden J індекс отримує максимальне значення відповідно до виразу (1).

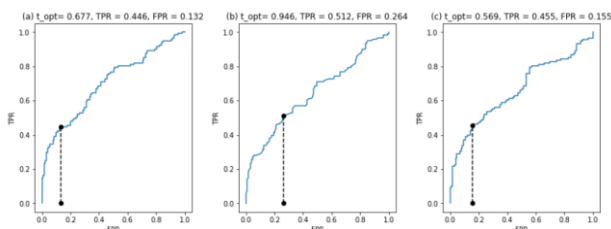


Рисунок 1. ROC- криві трьох різних наборів даних з оптимальними порогоми: (a) - 0.7, (b)-0.9, (c)-0.6.

Дослідження динаміки зміни різниці між значеннями метрик $nMCC$ та BA (рис. 2) і $nMCC$ та F_1 (рис. 3) залежно від значення порогу показує, що максимальна різниця в значеннях відповідає порогу, який більше порогів зі значенням 0.5 та зі значенням t_{opt} . Значить, якщо визначити найбільші значення оцінок $nMCC$, BA , F_1 , тоді можна отримати найбільшу абсолютну різницю (далі $\max \Delta$) між їх значеннями і порівнявши $\max \Delta$ з заданим допустимим рівнем θ на цій основі обґрунтувати вибір метрики незалежно від значення порогу.

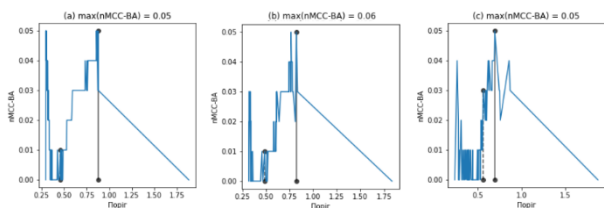


Рисунок 2. Динаміка зміни різниці оцінок методів $nMCC$ та BA залежно від порогу

Максимальні значення оцінок $nMCC$, BA , F_1 можна отримати, якщо максимізувати J індекс і описати функції оцінок $nMCC$, BA , F_1 , як такі, що залежать від максимізованого J індекса.

Максимальне значення J індексу отримаємо прийнявши $FP=0$, тобто припустивши, що модель не помилятиметься для спостережень, які вірно належать класу «0».

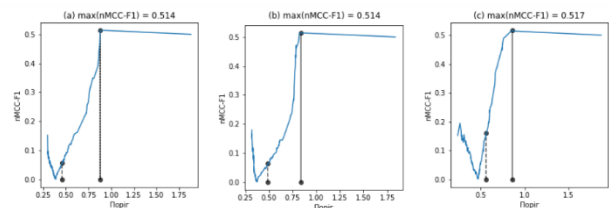


Рисунок 3. Динаміка зміни різниці оцінок методів $nMCC$ та F_1 залежно від порогу

Вираз (1) для $FP=0$ прийме вигляд (2). Функції оцінок $nMCC$, BA , F_1 [1] визначені від параметра $J_{fp \rightarrow 0}$ опишемо виразами (3-5).

$$J_{fp \rightarrow 0} = \frac{P_j}{P_i}, \quad (2)$$

$$MCC_{fp \rightarrow 0} = \frac{P_j \cdot N_i}{\sqrt{P_j \cdot N_i \cdot P_i \cdot N_j}} = \sqrt{\frac{P_j}{P_i}} \cdot \sqrt{\frac{N_i}{N_j}} = \sqrt{J_{fp \rightarrow 0}} \cdot \sqrt{\frac{N_i}{N_j}}, \quad (3)$$

$$BA_{fp \rightarrow 0} = \frac{1}{2} \left(\frac{P_j}{P_i} + 1 \right) = \frac{J_{fp \rightarrow 0} + 1}{2}, \quad (4)$$

$$F_{1fp \rightarrow 0} = \frac{2P_j}{2P_j + P_i - P_j} = \frac{2P_j}{P_j + P_i} = \frac{2}{1 + \frac{1}{J_{fp \rightarrow 0}}}, \quad (5)$$

В таблиці 3, показані значення ймовірностей для кривих з рисунка 3 і обчислені найбільші абсолютні різниці $\max \Delta 1 = |nMCC - BA|$; $\max \Delta 2 = |nMCC - F_1|$.

Таблиця 3. Найбільші абсолютні різниці оцінок моделі за методами $nMCC$, BA , F_1

№	TP	FP	FN	TN	Поріг	$\max \Delta 1$	$\max \Delta 2$
1	5	0	116	129	0.9	0.05	0.5
2	3	0	118	129	0.8	0	0.5
3	2	0	119	129	0.8	0.05	0.5

При заданому допустимому значенні різниці $\theta=0.5$ зробимо висновок, що оцінювання моделі за методом F_1 не є доцільним для заданого набору даних.

Список літератури

- [1] Tharwat, A.: Classification assessment methods. Applied computing and informatics, 17(1),2020. 168-192p.
- [2] Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1),2020. 1-13p.
- [3] Chicco, D., Tötsch, N. and Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData mining, 14(1),2021, 1-22p.
- [4] Honcharenko T, Solovei O. Optimal bin number for histogram binning method to calibrate binary probabilities. CEUR Workshop Proceedings, vol-3628,2023. 126–135p.
- [5] Attwood, K. and Tian, L. Confidence interval estimation of the Youden index and corresponding cut-point for a combination of biomarkers under normality. Communications in Statistics-Theory and Methods, 51(2),2022. 501-518p.