

Қазақстан Республикасы Тәуелсіздігінің 30 жылдығына арналған «Сейфуллин оқулары – 17: «Қазіргі аграрлық ғылым: цифрлық трансформация» атты халықаралық ғылыми – тәжірибелік конференцияға материалдар = Материалы международной научно – теоретической конференции «Сейфуллинские чтения – 17: «Современная аграрная наука: цифровая трансформация», посвященной 30 – летию Независимости Республики Казахстан.- 2021.- Т.1, Ч.4 - С.237-241

ОЦЕНКА КАЧЕСТВА РАБОТЫ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В РЕШЕНИИ ЗАДАЧ КЛАССИФИКАЦИИ

Тулегенов Т.Н.

Рост объема и разнообразия данных (структурированных и неструктурированных) актуализировал проблему извлечения полезной информации из данных самой разной природы. Не исключением является и медицинская сфера, в которой на сегодняшний день уже накоплены огромные массивы разнородных данных и с внедрением систем автоматизации в медицинских учреждениях скорость их накопления постоянно растет.

Разделы науки о данных, которые освещают различные методы изучения структур больших данных с целью нахождения в них взаимосвязей и закономерностей, называются машинное обучение и анализ данных. При этом, несмотря на высокую популярность и широкое распространение методов машинного обучения в решении самого широкого класса задач, на сегодняшний день отсутствуют единые подходы в оценке качества используемых алгоритмов [1]. Одним из основных классов задач, решаемых алгоритмами машинного обучения, является отнесение наблюдаемого объекта к тому или иному классу для последующего принятия решения автоматически или человеком.

Созданию модели машинного обучения, позволяющей в зависимости от заданного критерия эффективно выявлять и описывать скрытые зависимости между признаками при помощи различных функций, сопутствующей является задача определения качества выбранного алгоритма.

Существует множество различных метрик, используемых для оценки эффективности алгоритмов машинного обучения в решении задач классификации. Необходимость тщательного выбора метрик для оценки эффективности алгоритмов машинного обучения обусловлена следующими факторами:

- Зависимость способов измерения и сравнения производительности алгоритмов машинного обучения от выбранной метрики.
- Оценка важности различных характеристик в результате также будет зависеть от выбранной метрики.

Основными метриками эффективности для задач классификации являются [2]:

1. Безошибочность (ассигасу) - это наиболее интуитивно понятный показатель эффективности, являющийся отношением правильно спрогнозированных наблюдений к их общему количеству. Показатель безошибочности способен сразу показать, правильно ли обучается модель и работоспособна ли она в целом. Однако он не дает подробной информации о применимости модели к решению конкретной проблемы.

Высокая безошибочность позволяет сделать первоначальный вывод о выборе наилучшего алгоритма, но только в тех случаях, когда у нас есть симметричные наборы данных, в которых значения положительных и отрицательных классов практически равны. В остальных случаях безошибочность не лучший показатель, поскольку не учитывает оценку вероятности.

2. Матрица ошибок (confusion matrix) – также является интуитивно понятной и простой метрикой, используемой для определения корректности и точности алгоритмов в задачах классификации, где на выходе могут быть два или более типов классов. Однако так же, как и показатель безошибочности, матрица ошибок не учитывает оценку вероятности.

Матрица ошибок представляет собой матрицу размером $N \times N$, где N - количество прогнозируемых классов. Для случая простейшей задачи классификации $N = 2$, следовательно, мы получаем матрицу 2×2 с двумя измерениями («фактическое» и «прогнозируемое») и наборами «классов» в обоих измерениях. Где фактический класс - это строки, а прогнозируемый – столбцы (табл. 1)

Таблица 1- Матрица ошибок

	Прогнозируемый 0	Прогнозируемый 1
Фактический 0	TN	FP
Фактический 1	FN	TP

Истинно-отрицательный (TN - True Negatives) - когда и фактический, и прогнозируемый класс данных равны 0. Пример: случай, когда показатель фактически отрицательный (0) и модель классифицировала этот показатель как отрицательный (0).

Ложноположительный (FP - False Positives) - случай, когда фактический класс показателя данных равен 0, а прогнозируемый класс данных равен 1. Пример: случай, когда показатель фактически отрицательный (0), а модель классифицировала этот показатель как положительный (1).

Ложноотрицательный (FN - False Negatives) - случай, когда фактический класс данных равен 1, а прогнозируемый класс равен 0. Пример: случай, когда

показатель фактически положительный (1), а модель классифицировала этот показатель как отрицательный (0).

Истинно-положительный (TP - True Positives) - когда и фактический, и прогнозируемый класс данных равны 1. Пример: случай, когда показатель фактически положительный (1) и модель классифицировала этот показатель как положительный (1).

Сама по себе матрица ошибок как таковая не является метрикой производительности, но почти все метрики эффективности основаны на значениях внутри нее. Обозначим общее количество отрицательных срабатываний модели как N, а общее количество позитивных срабатываний как P.

В контексте матрицы ошибок и применительно к задачам классификации, безошибочность – это количество правильных прогнозов, сделанных алгоритмом по всем видам сделанных прогнозов.

Примеры. Ложное срабатывание (FP) системы защиты от спама переместит полезное электронное письмо в ящик со спамом. Ложноотрицательный результат (FN) при медицинском обследовании может неверно указывать на отсутствие болезни, хотя на самом деле диагноз является положительным.

Точность (precision) сообщает обо всех прогнозируемых положительных срабатываниях алгоритма, а также о том, какой процент данных на самом деле положителен.

Таким образом, к примеру, даже если удалось зафиксировать только одно совпадение, и оно зафиксировано верно, точность будет являться 100%.

Чувствительность (recall) отражает отношение количества истинно-положительных срабатываний к общему количеству срабатываний, которые должны были быть предсказаны как положительные, т.е. показывает, какая доля положительных объектов была выделена классификатором как положительная.

В случаях приоритетности минимизации ложноотрицательных результатов, необходимо стремиться к показателю чувствительности как можно ближе к 100% без слишком плохой точности. Чувствительность дает информацию о производительности алгоритма в отношении ложных срабатываний, в то время как точность дает нам информацию о его производительности по отношению к ложным срабатываниям.

Для того, чтобы каждый раз при решении задач классификации не учитывать по отдельности показатели точности и чувствительности была выведена отдельная комплексная метрика **F1-Score**, учитывающая оба этих показателя. Математически F1-Score представляет собой средневзвешенное значение точности и чувствительности. Лучшее значение F1-Score равно 1, а худшее – 0. F1-Score имеет равный относительный вклад точности и чувствительности:

$$F1-Score = 2*(precision*recall)/(precision+recall) \quad (1)$$

3. Рабочие характеристики приемника (Receiver-operating characteristic – ROC) были первоначально разработаны во время Второй мировой войны для анализа точности классификации при дифференцировании сигналов от шума в радиолокации. В последнее время эта методология была адаптирована к различным клиническим областям, в значительной степени зависящим от скрининговых и диагностических тестов, таким как лабораторные исследования, эпидемиология, радиология и биоинформатика [3].

Кривая рабочих характеристик приемника (ROC Curve) – это график зависимости истинно-положительного уровня от ложноположительного.

Площадь под кривой (Area Under Curve – AUC) это показатель производительности, основанный на различных пороговых значениях для задач классификации. Как следует из названия, ROC – это кривая вероятности, а AUC измеряет разделимость (рис.3).

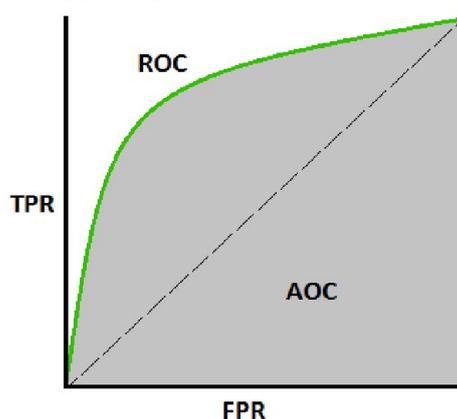


Рисунок 3 - Взаимосвязь показателей ROC и AUC

Кривая ROC является графиком TPR и FPR при разных значениях порогов классификации. Снижение порога классификации классифицирует больше примеров как позитивные, при этом увеличиваются и ложноположительные и истинно-положительные.

AUC предоставляет совокупное измерение производительности при всех возможных значениях классификационного порога. Один из путей интерпретировать AUC - рассматривать ее как вероятность, что модель ранжирует ложноположительный пример выше, чем ложнонегативный.

4. Логарифм потери (Log Loss) – одна из важных метрик классификации основанная на вероятностях. Если алгоритм дает нам оценку вероятности, то Log-loss является лучшим показателем производительности как для задач двоичной, так и для множественной классификации. Цель алгоритмов машинного обучения минимизировать значение логарифма потери.

Одним из недостатков данной метрики является трудность интерпретации необработанных значений, однако логарифм потери является хорошим показателем для сравнения алгоритмов. Для любой задачи классификации более низкое значение логарифма потери означает лучшие прогнозы. Вычисление логарифма потери для бинарной классификации:

$$\text{LogLoss} = \frac{1}{N} \sum_{i=1}^N (y_i (\log p_i) + (1 - y_i) \log(1 - p_i)) \quad (2)$$

Логарифм потери количественно определяет среднюю разницу между прогнозируемым и ожидаемым распределениями вероятностей [4].

Заключение. Рассмотренные метрики являются лишь одними из основных для задач классификации. Как было отмечено в начале статьи, каждая задача имеет свою специфику и приоритеты, в связи с чем невозможно гарантированно определить, какая из метрик является наилучшей и универсальной. Приведенные примеры оценки качества алгоритмов машинного обучения будут использованы на практике при анализе медицинских лабораторных данных в рамках магистерской диссертации.

Список литературы

1. Р.И. Мухамедиев, Е.Л. Мухамедиева, Я.И. Кучин 2021. Таксономия методов машинного обучения и оценка качества классификации и обучаемости. [online] Available at: <https://cyberleninka.ru/article/n/taksonomiya-metodov-mashinnogo-obucheniya-i-otsenka-kachestva-klassifikatsii-i-obuchaemosti>
2. Studme. 2021. Оценка качества работы алгоритма машинного обучения. [online] Available at: https://studme.org/139975/informatika/otsenka_kachestva_raboty_algoritma_mashinnogo_obucheniya
3. Fawcett T. (2005). An introduction to ROC analysis. Institute for the Study of Learning and Expertise. Institute for the Study of Learning and Expertise, Palo Alto, USA. [online] Available at: <https://ccrma.stanford.edu/workshops/mir2009/references/ROCintro.pdf>
4. Loginom.ru. 2021. Логистическая регрессия и ROC-анализ - математический аппарат. [online] Available at: <https://loginom.ru/blog/logistic-regression-roc-auc>

Научный руководитель Исмаилова А.А., PhD