DOI: 10.25558/VOSTNII.2025.47.21.003
УДК 622.831

А. В. ДЯГИЛЕВА
канд. техн. наук, доцент
доцент кафедры
КузГТУ, г. Кемерово
e-mail: dyagileva1952@mail.ru

П. А. ПЫЛОВ
аспирант
КузГТУ, г. Кемерово
e-mail: gedrosten@mail.ru

Р. В. МАЙТАК
магистр
КузГТУ, г. Кемерово
e-mail: maytak.roman@mail.ru

Е. А. НИКОЛАЕВА
канд. физ.-мат. наук, доцент
заведующий кафедрой
КузГТУ, г. Кемерово
e-mail: nikolaevaea@yandex.ru

МЕТРИКИ И КРИТЕРИИ КАЧЕСТВА МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ОЦЕНКИ ТОЧНОСТИ ПРОГНОЗИРОВАНИЯ СЕЙСМИЧЕСКОГО УДАРА ВЫСОКОЙ ЭНЕРГИИ

Сейсмическая опасность является одним из факторов, который ставит под угрозу жизнь и здоровье человека. В случае сотрудников горнодобывающих предприятий, этот фактор является одним из ключевых, так как в толще земли сейсмические толчки могут привести к разрушению горных пород, поломке и обрушению шахтных конструкций, а также к блокировке путей эвакуации. Сейсмическая опасность по своему типу относится к природной сфере опасных факторов, поэтому спрогнозировать её появление является творческой и очень трудоемкой задачей. Превентивные меры, предложенные человечеством ещё несколько сотен лет назад, не всегда позволяют с достаточной точностью определить катастрофу и предупредить её последствия. С развитием информационных технологий (прикладного искусственного интеллекта, в частности), а также развитым программно-аппаратным комплексом, стало возможно применение интеллектуальных систем для автоматизации различных технологических задач. Не исключением является и предметная область горнодобывающего дела — разнообразные модификации глубоких нейронных сетей, а также самообучающиеся модели машинного обучения могут быть успешно интегрированы для автоматизации процесса прогнозирования сейсмических ударов высокой энергии. Однако, для того чтобы успешно имплементировать модель прикладного искусственного интеллекта, прежде всего необходимо формализовать метрику качества, на основе которой можно будет судить о том, какая модель функционирует лучше, а какая хуже. В рамках настоящей статьи авторами были представлены результаты исследования

по применению различных метрик для оценки качества прогноза сейсмических ударов. Авторами представлены и детально исследованы критерии, обобщение точности которых наиболее полно коррелирует с эффективностью решения задачи предметной области горнодобывающего дела. Ценность научно-исследовательской работы состоит в том, что в дальнейшем исследователям данных не потребуется затрачивать временные, человеческие, материальные и иные ресурсы для выбора оптимальной метрики качества, которая будет использоваться в качестве показателя точности автоматизирующей интеллектуальной модели. Немаловажным фактом является и то, что в рамках статьи отдельное внимание было сфокусировано на математических основах метрик качества, поэтому выбор языка программирования, на котором будет в дальнейшем реализована модель машинного обучения, не является критичным.

Ключевые слова: ШАХТА, УГОЛЬНАЯ ОТРАСЛЬ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, ИНТЕЛЛЕКТУАЛЬНЫЕ ЯЗЫКОВЫЕ МОДЕЛИ, ОБРАБОТКА СЕЙСМИЧЕСКИХ ДАННЫХ МОДЕЛЯМИ ГЛУБОКОГО И МАШИННОГО ОБУЧЕНИЯ.

Горнодобывающая промышленность всегда была и остается неразрывно связанной с потенциальными опасностями, которые обычно именуются опасностями при добыче полезных ископаемых [1–4]. Частным случаем такой угрозы является сейсмическая опасность, которая сохраняет высокую вероятность возникновения во многих подземных промышленных комплексах. Сейсмическую опасность наиболее трудно обнаружить и предсказать в сравнении со всеми остальными типами аварий и стихийных бедствий [1]. Системы сейсмического и сейсмоакустического мониторинга позволяют распознавать процессы в массиве горных пород и определять методы прогнозирования сейсмической опасности. Однако точность реализованных методов по состоянию на февраль 2024 года ещё далека от ожидаемого идеала [2].

Сложность сейсмических процессов и колоссальный дисбаланс между количеством низкоэнергетических сейсмических событий ($\approx 10^{0-1}0^{1}$) и величиной высокоэнергетических явлений ($\approx 10^{4}$) неминуемо приводят к тому, что статистические методы оказываются малоэффективными для прогнозирования сейсмической опасности. По этой причине крайне важно детерминировать новые возможности для результативного прогнозирования опасности, в том числе с использованием методов прикладного искусственного интеллекта и машинного обучения, в частности. Для оценки сейсмической опасности могут быть использованы методы кластеризации данных [1], а для прогнозирования сейсмических толчков эффективно задействуются искусственные нейронные сети [3].

В большинстве случаев итоговые результаты, полученные с помощью упомянутых методов, представлены в виде двух состояний, которые интерпретируются как «опасные условия» и «неопасные условия». Несбалансированное распределение положительных («опасные условия») и отрицательных («неопасные условия») состояний целевой функции является серьезной проблематикой при прогнозировании сейсмической опасности. Применяемые в настоящее время методы все еще недостаточны для достижения результатов с низкой величиной ложноположительной и ложноотрицательной ошибки.

В научном исследовании М. Буковска [4] был предложен ряд факторов, влияющих на возникновение сейсмической опасности. Среди основных факторов было перечислено возникновение подземных толчков с энергией, превышающей 10^{4} Дж.

Задача сейсмического прогнозирования может быть детерминирована по-разному, но основной целью всех методик оценки сейсмической опасности является прогнозирование (с заданной точностью относительно времени и даты) повышенной сейсмической активности, которая потенциально может вызвать детонацию газа (метан) и взрыв породы [2, 4]. В исследуемом наборе данных каждая строка содержит сводную информацию о сейсмической активности в массиве горных пород

в течение одной смены (8 часов). Если атрибут решения имеет значение 1, то в следующую смену был зарегистрирован любой сейсмический удар с выделяемой энергией выше 10^4Дж.

Задача прогнозирования опасностей основана на взаимосвязи между энергией зарегистрированных подземных толчков и сейсмоакустической активностью с потенциальной возможностью возникновения камнепада. Следовательно, такой прогноз опасности не связан с точным прогнозированием горного взрыва. Более того, располагая информацией о возможности возникновения опасной ситуации, соответствующая служба надзора может самостоятельно снизить риск горного взрыва или заранее вывести работников из угрожаемой зоны. Таким образом, прогнозирование повышенной сейсмической активности имеет высокую практическую ценность.

Представленный набор данных характеризуется несбалансированным распределением положительных и отрицательных примеров. В наборе данных содержится только 170 положительных кортежей объектов, представляющих класс «1» (7 % от общего числа наблюдений). Подавляющее большинство других кортежей, соответственно, относится к классу «0» (93 % измерений). Таким образом, дисбаланс данных накладывает серьезные математические ограничения на условия разбиения набора на тренировочные и тестовые данные [5].

Информация об атрибутах данных (прецедентах) удобнее всего может быть представлена в формате сводной результирующей таблицы. В таблице 1 зафиксировано наименование признака (таким, как оно представлено в наборе данных и в дальнейшем проекте решения), его русскоязычное представление и поясняющая семантику термина описательная характеристика.

Таблица 1

Признаки набора данных

Наименование признака	Русскоязычное представление	Описание семантики прецедента
id	Идентификатор записи	Представляет собой итеративное значение счетчика, которое является уникальным для каждого кортежа записи набора
seismic	Сейсмический показатель	Результат оценки сейсмической опасности сдвига в горной выработке, полученный сейсмическим методом (a — отсутствие опасности, b — низкая опасность, c — высокая опасность, d — состояние опасности)
seismoacoustic	Сейсмоакустический показатель	Результат оценки сейсмической опасности сдвига в горной выработке, полученный сейсмоакустическим методом (a — отсутствие опасности, b — низкая опасность, c — высокая опасность, d — состояние опасности)
shift	Тип смены	Информация о типе смены (W — добыча угля, N — подготовительная смена)
genergy	Сейсмическая энергия	Сейсмическая энергия, зафиксированная в течение предыдущей смены наиболее активным геофоном (GMax) из геофонов, контролирующих наиболее длинную стену
gpuls	Суммарная величина импульсов	Количество импульсов, записанных в течение предыдущей смены с помощью GMax
gdenergy	Отклонение текущей величины энергии от значения средней	Отклонение энергии, зарегистрированной в течение предыдущей смены по GMax, от средней энергии, зарегистрированной в течение восьми предыдущих смен

Наименование признака	Русскоязычное представление	Описание семантики прецедента
genergy	Сейсмическая энергия	Сейсмическая энергия, зафиксированная в течение предыдущей смены наиболее активным геофоном (GMax) из геофонов, контролирующих наиболее длинную стену
gpuls	Суммарная величина импульсов	Количество импульсов, записанных в течение предыдущей смены с помощью GMax
gdenergy	Отклонение текущей величины энергии от значения средней	Отклонение энергии, зарегистрированной в течение предыдущей смены по GMax, от средней энергии, зарегистрированной в течение восьми предыдущих смен
gdpuls	Отклонение текущего количества импульсов от средней величины этого параметра	Отклонение количества импульсов, зарегистрированных в течение предыдущей смены, по GMax от среднего количества импульсов, зарегистрированных в течение восьми предыдущих смен
ghazard	Оценка опасности сдвига	Результат оценки сейсмической опасности сдвига в горной выработке, полученный сейсмоакустическим методом на основе регистрации, поступающей только из формы GMax
nbumps	Суммарное количество толчков в предыдущей смене	Количество сейсмических толчков, зарегистрированных в течение предыдущей смены
nbumps2	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^2; 10^3)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^2; 10^3)$, зарегистрированных в течение предыдущей смены
nbumps3	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^3; 10^4)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^3; 10^4)$, зарегистрированных в течение предыдущей смены
nbumps4	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^4; 10^5)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^4; 10^5)$, зарегистрированных в течение предыдущей смены
nbumps5	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^5; 10^6)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^5; 10^6)$, зарегистрированных в течение предыдущей смены
nbumps6	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^6; 10^7)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^6; 10^7)$, зарегистрированных в течение предыдущей смены
nbumps7	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^7; 10^8)$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^7; 10^8)$, зарегистрированных в течение предыдущей смены
nbumps89	Суммарное количество толчков в предыдущей смене диапазона значений энергий $[10^8; 10^{10})$	Количество сейсмических толчков (ударов) в диапазоне энергий $[10^8; 10^{10})$, зарегистрированных в течение предыдущей смены

Наименование признака	Русскоязычное представление	Описание семантики прецедента
energy	Суммарная энергия сейсмических толчков за предыдущую смену	Общая суммарная энергия сейсмических толчков, зарегистрированных в течение предыдущей смены
maxenergy	Максимальное значение энергии зарегистрированного сейсмического толчка	Максимальная энергия сейсмических толчков, зарегистрированных в течение предыдущей смены
class	Целевой класс	Целевая переменная задачи — «1» означает, что в следующую смену произойдёт сейсмический удар высокой энергии («опасное состояние»), «0» означает, что в следующую смену не произойдет сейсмических ударов высокой энергии («неопасное состояние»)

Таблица 1 в описательной форме предоставляет информацию о прецедентах набора данных. По этой информации уже можно выполнить детальную математическую аналитику всех признаков датасета для того, чтобы уже на начальном шаге определить параметры устойчивости будущей модели машинного обучения, а также избрать наиболее оптимальный для решения задачи алгоритм прикладного искусственного интеллекта.

Определив и обобщенно исследовав состав признаков в наборе данных, можно приступать к статистическому анализу данных. С точки зрения математического описания данных в задачах прикладного искусственного интеллекта, классифицируют три основных типа прецедентов [6]:

1. Количественные (представляют собой измеренные значения некоторого признака);

a. Непрерывные (могут принимать абсолютно любое число из заданного промежутка, например, рост людей.);

b. Дискретные (принимают только определенные значения из заданного диапазона, например, количество детей в семье, как минимум, может быть только целочисленным значением).

2. Номинативные (разделяют объекты выборки на группы, например, все люди мужского пола будут обозначены классом «1», а женского пола — «0». Цифры не несут в себе математического числа, но позволяют работать с данными как с числами, а не текстом);

3. Ранговые (используются для ранжирования наблюдений, например, для сравнения объектов по определенному критерию — тогда, чем больше (или меньше) будет значение, тем большее (или меньшее) преобладание характеристики будет выражено у объекта выборки в сравнении его с другими экземплярами генеральной совокупности данных).

Сравним все прецеденты набора данных (отраженных в таблице 1) по их статистическому типу. Результаты зафиксируем в новой сводной таблице 2.

Таблица 2

Статистический тип признаков

Признак	Статистический тип данных признака	Диапазон возможных значений
id	Количественный (дискретный)	Начинается с 1 и ограничен финальным кортежем набора данных — 2584.
seismic	Номинативный	Четыре возможных типа значения: a — отсутствие опасности, b — низкая опасность, c — высокая опасность, d — состояние опасности

Признак	Статистический тип данных признака	Диапазон возможных значений
seismoa-coustic	Номинативный	Четыре возможных типа значения: a — отсутствие опасности, b — низкая опасность, c — высокая опасность, d — состояние опасности
shift	Номинативный	Два возможных типа значения: W — добыча угля, N — подготовительная смена
genergy	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 100, максимальное — 2595650, среднее значение — 90242,523
gpuls	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 2, максимальное — 4518, среднее значение — 538,579
gdenergy	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — –96, максимальное — 1245, среднее значение — 12,375
gdpuls	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — –96, максимальное — 838, среднее значение — 4,509
ghazard	Номинативный	Три возможных типа значения: a — отсутствие опасности, b — низкая опасность, c — высокая опасность
nbumps	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 9, среднее значение — 0,859
nbumps2	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 8, среднее значение — 0,394
nbumps3	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 7, среднее значение — 0,393
nbumps4	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 3, среднее значение — 0,068
nbumps5	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 1, среднее значение — 0,005
nbumps6	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 0, среднее значение — 0,0
nbumps7	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 0, среднее значение — 0,0
nbumps89	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 0, среднее значение — 0,0
energy	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 402000, среднее значение — 4975,271
maxenergy	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 400000, среднее значение — 4278,851
class	Количественный (дискретный)	Целочисленный тип данных, минимальное значение из диапазона — 0, максимальное — 1, среднее значение — 0,065

Из таблицы 2 следует, что все признаки обозначены либо количественными (непрерывными и дискретными) значениями, либо номинативными. С точки зрения построения моделей машинного обучения наиболее практично, что в наборе данных отсутствует текстовые поля и информация, которую было бы трудно/невозможно заменить номинативными переменными [7]. Это также обосновано ещё и тем, что расширяется спектр применимости моделей машинного обучения для решения данной задачи, так как многие алгоритмы относятся к вероятностным моделям, которые работают только с числовыми значениями [8].

Отметим также, что числовые значения признаков имеют разные порядки, то есть разнятся в значительном диапазоне значений. Таким образом, для правильного функционирования большинства моделей машинного обучения потребуется применение стандартизации и нормализации прецедентов набора данных [9]. Номинативные признаки, в свою очередь, в некоторых моделях машинного обучения (более всего это требуется для моделей вероятностного типа) будут перекодированы в числовые уникальные значения для того, чтобы алгоритмическая функция модели машинного обучения могла работать с признаками как с числами. Несмотря на то, что, как было сказано ранее, номинативные признаки не несут в себе коренного математического смысла, однако такая концепция позволяет внести коррективы в формирование правильной обобщающей способности итоговой модели прикладного искусственного интеллекта [10]. Что касается оценки алгоритма по результирующей функции потерь, то стоит сразу учесть, что в наборах данных, которые характеризуются высоким дисбалансом целевых классов, оценка по такой методике всегда неэффективна [8]. Это объясняется высоким перевесом одного из классов, в то время как функция потерь старается равнозначно учесть верную классификацию как на одном целевом классе, так и на другом. Из-за первоначального несбалансированного разделения набора данных такая оценка будет выдавать пессимистически заниженные результаты функционирования модели машинного обучения на валидационных данных и отложенной выборке прецедентов.

Оптимальные метрики оценивания точности моделей машинного обучения для решения прикладной задачи дихотомической классификации

Начиная исследование применимости алгоритмов машинного обучения для решения прикладной задачи превентивного определения сейсмоакустической опасности, следует сразу определить, что данная задача относится к типу задач классификации.

В противовес ей, в задаче регрессии выходной переменной является не класс целевой переменной («1» или «2»), а величина значения с плавающей точкой (как правило, вероятность отнесения объектов к заданному классу или вероятность наступления конкретного события из нескольких взаимоисключающих событий).

Начнем исследование моделей машинного обучения с главной задачи — определения качества в задаче классификации.

Для определения ошибок в задаче классификации определим некоторые математические ограничения. Существует задача классификации на два класса, при этом $y_i \in \{-1; +1\}$. Для решения задачи применяется некий алгоритм (1).

$$a(x_i) \in \{-1;+1\} \qquad (1)$$

Объекты могут принадлежать одному из двух классов, поэтому было бы рациональнее всего обозначить их как ±1.

При этом, учитывая (1), возможно наличие четырех ситуаций. Для удобства все ситуации отображены на рисунке 1.

	ответ классификатора	правильный ответ
TP, True Positive	$a(x_i) = +1$	$y_i = +1$
TN, True Negative	$a(x_i) = -1$	$y_i = -1$
FP, False Positive	$a(x_i) = +1$	$y_i = -1$
FN, False Negative	$a(x_i) = -1$	$y_i = +1$

Рис. 1. Все возможные ответы алгоритма классификации

На рисунке 1 используются общепринятые международные обозначения (True Positive, True Negative, False Positive и False Negative) [11, 12]. Positive и Negative обозначают соответственно метку класса (положительный +1 или отрицательный класс –1), а характер классификации True и False означает ошибку и верное определение значения алгоритмом классификации.

Доля правильных классификаций (на основе данных рисунка 1) может быть рассчитана по несложной формуле отношения верных классификаций алгоритма ко всем классифицируемым объектам (2).

$$\text{Точнсоть} = Accuracy = \frac{1}{l}\sum_{i=1}^{l}[a(x_i) = y_i] =$$

$$= \frac{TP + TN}{FP + FN + TP + TN} \qquad (2)$$

Измерение доли правильных классификаций по формуле (2) позволило бы определять точность алгоритма классификации, так как в числители дроби отнесены все верно классифицированные объекты, а в знаменателе фактически находится полный объем выборки, по которой и организуется критерий точности модели машинного обучения. Недостаток критерия состоит в том, что данная метрика (2) не учитывает дисбаланс целевых классов. То есть этот недостаток очень точно указывает на рассматриваемую прикладную задачу. Использование данной метрики чревато серьезными, неочевидными на первый взгляд ошибками: в нашем случае объектов класса «1» более, чем в 13 раз меньше, чем объектов противоположного класса «2». Тогда, если, например, на всех объектах класса «1» будет произведена неверная классификация, то процент ошибки составит всего лишь 7 %. Разумеется, что на первый взгляд ошибка в 7 % (точность в 93 %) кажется не очень существенной, однако задача осталась полностью нерешенной.

Таким образом, для исследуемой прикладной задачи метрика *Accuracy* не подходит, так как она не учитывает ни численность (дисбаланс) классов, ни цену ошибки на объектах разных классов. Рассмотрим методы, позволяющие учитывать дисбаланс между целевыми классами. Исследуем теперь модель классификации с дискриминантной функцией (3) взамен стандартному алгоритму (1).

$$a(x;w;\, w_0) = sign(g(x,w) - w_0) \qquad (3)$$

где $g(x,w)$ — это дискриминантная функция.

В формуле (3) чем больше (в определенных границах) w_0 , тем больше x_i таких, что $a(x_i) = -1$. Таким образом, классификатор (3) можно довести до крайнего состояния, в котором он:

1. Станет вырожденным, то есть классификатор на всех объектах будет выдавать –1.

2. Уменьшить до предела, когда классификатор будет выдавать на всех объектах значение +1.

В зависимости от того, как будет перемещаться параметр свободного члена классификатора (3), будет зависеть и то, насколько много объектов выборки относится к объектам определенного класса. Это означает, что баланс классов зависит от параметра свободного члена w_0.

Остается только определиться с тем, как будут учитываться различающиеся стоимости классов. Под стоимостью может пониматься как мощность классов, так и значимость одного класса относительно другого [10]. Для этого определим штраф за ошибку на объекте класса y как λ_y. Тогда смещенная функция потерь (4) теперь зависит от параметра λ_y .

$$\mathcal{L}(a,y) = \lambda_{y_i}[a(x_i; w, w_0) \neq y_i] =$$

$$= \lambda_{y_i}[(g(x_i,w) - w_0)y_i < 0] \qquad (4)$$

На практике штрафы $\{\lambda_y\}$ могут быть неоднократно пересмотрены, в связи с чем:

- Требуется более явный способ подбора свободного члена w_0 в зависимости от текущей установки штрафов $\{\lambda_y\}$.
- Требуется характеристика качества модели $g(x,w)$, которая бы не зависела от штрафов $\{\lambda_y\}$ и численности классов.

Правильная математическая характеристика для измерения дискриминантной функции, которая [метрика] не зависит от соотношения классов, называется кривой ошибок (ROC-кривая) [9]. ROC-кривая представляет собой график зависимости, где по оси абсцисс откладывается доля ошибочных положительных классификаций (False Positive Rate) (5), а по оси ординат — доля правильных положительных классификаций (True Positive Rate) (6).

$$FPR(a, X^l) = \frac{\sum_{i=1}^{l}[y_i = -1][a(x_i; w, w_0) = +1]}{\sum_{i=1}^{l}[y_i = -1]}; \quad (5)$$

$$TPR(a, X^l) = \frac{\sum_{i=1}^{l}[y_i = +1][a(x_i; w, w_0) = +1]}{\sum_{i=1}^{l}[y_i = +1]}; \quad (6)$$

При этом величина $1 - FPR(a)$ называется специфичностью алгоритма a, то есть характеризует точность, минимизирующую ошибку ложноположительного типа. Величина $TPR(a)$ называется чувствительностью алгоритма a, характеризуя собой минимизацию ошибки ложноотрицательного типа.

Таким образом, чем меньше значение FPR (5), тем лучше. С величиной TPR (6) всё обстоит ровно наоборот: чем больше это значение, тем выше точность [11].

Рассмотрим графическое представление ROC-кривой (рисунок 2).

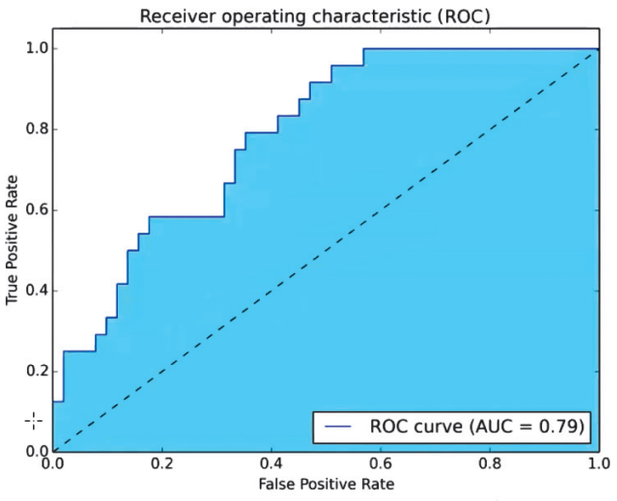

Рис. 2. ROC-кривая

На рисунке 2 выделенной пунктирной диагональю является плохой классификатор, так как он всегда допускает половину ошибок [12, 13]. Идеальным вариантом классификатора является модель, у которой весь квадрат 1.0×1.0 будет заполнен голубой заливкой. Ступенчатая кривая также имеет своё обоснование. Когда в формулах (5) и (6) начинает попеременно меняться параметр свободного члена, то точка данных начинает перемещаться по кривой, принимая конкретные значения из допустимого диапазона. Не все точки сразу же стремятся вдоль оси ординат, так как зачастую нельзя одновременно минимизировать ошибку первого и второго рода одновременно (ложноотрицательную и ложноположительную). Однако, чтобы учесть ошибки первого и второго рода, требуются соответствующие метрики.

Для этого выделено подразделение:

- Минимизация ошибки второго рода (ложноположительная) достигается путем поиска как можно большей доли релевантных объектов среди найденных (7).

$$\text{Чувствительность, } Precision = \frac{TP}{TP + FP} \quad (7)$$

- Минимизация ошибки первого рода (ложноотрицательной) достигается за счет определения доли найденных среди релевантных объектов (8).

$$\text{Специфичность, } Recall = \frac{TP}{TP + FN} \quad (8)$$

Формулы (7) и (8) удобнее представить на схематичном отображении релевантных и найденных объектов (рисунок 3).

При измерении метрик (7) и (8), как уже следует из их названия, необходимо стремиться к их максимизации, то есть приближению значений каждой метрики к величине 1,00. При обобщении на многоклассовую классификацию возникает дилемма выбора важности точности на отдельных объектах или точности на классах. Метрики точности (7) и (8) обобщаются до (9) и (10).

$$Precision = \frac{\sum_y TP_y}{\sum_y (TP_y + FP_y)} \quad (9)$$

$$Recall = \frac{\sum_y TP_y}{\sum_y (TP_y + FN_y)} \quad (10)$$

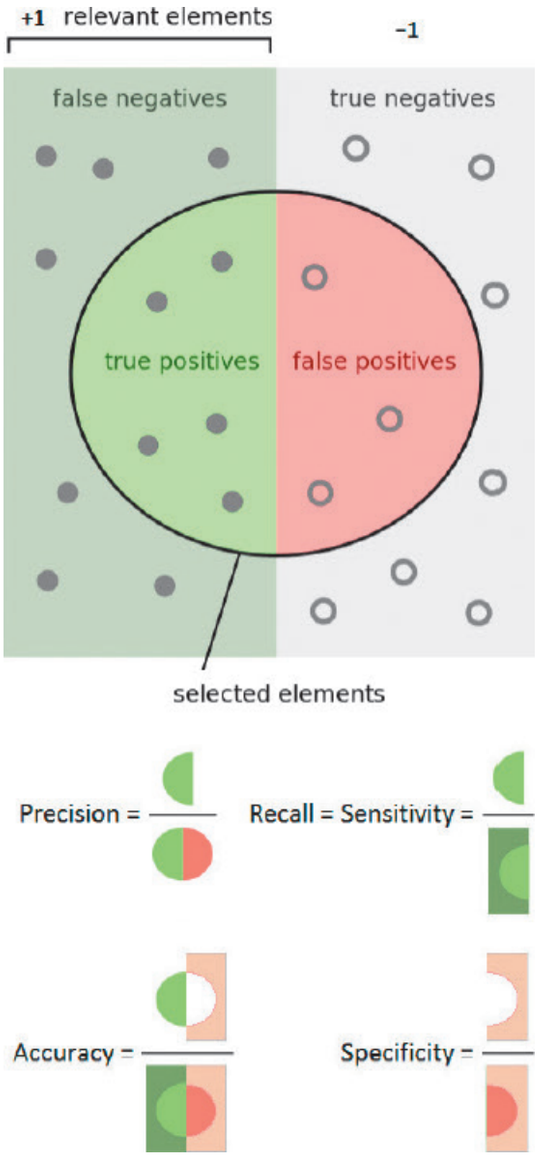

Рис 3. Представление метрик
для контроля ошибок разного рода.

Из более обобщенных формул (7) и (8) возникает две возможности настройки алгоритма для многоклассовой классификации. В первом рассмотренном случае (9) и (10) отражена суть микроусреднения. Сумма по классам, которая организована в числителе и знаменателе дроби. Микроусреднение не чувствительно к ошибкам на малочисленных классах, поэтому его нельзя использовать для несбалансированных наборов данных многоклассовой классификации. Второй способ настройки алгоритма позволяет применять метрики (11) и (12) для макроусреднения.

$$Precision = \frac{1}{|Y|}\sum_y \frac{TP_y}{TP_y + FP_y} \quad (11)$$

$$Recall = \frac{1}{|Y|}\sum_y \frac{TP_y}{TP_y + FN_y} \quad (12)$$

В макроусреднении значения метрик (11) и (12) усредняются по классам. По этой причине макроусреднение чувствительно к ошибкам на малочисленных классах. У кооперации метрик *Precision* и *Recall* существует и свое собственное графическое представление в виде кривой, которая называется *Precision-Recall Curve*. Пример такой кривой для задачи многоклассовой классификации представлен на рисунке 4.

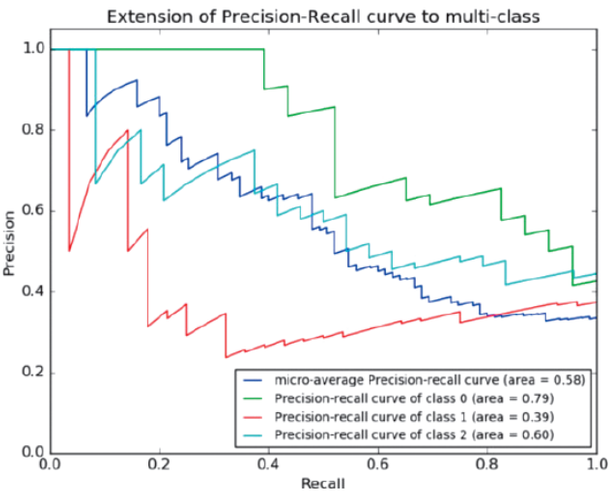

Рис. 4. Представление Precision-Recall-кривой

Кривая взаимосвязи *Precision-Recall* позволяет выбирать между минимизациями ошибок первого и второго рода. В зависимости от выбора, точки данных будут отстраиваться на графике аналогично графику ROC-кривой. Однако, в случае кривой *Precision-Recall* главной целью является попадание в правый верхний угол квадрата точности 1.0×1.0. Перейдем к задаче выбора модели и метода обучения. Сразу отметим, что выбор модели по обучающей выборки невозможен, так как данный критерий будет всегда оптимистично смещенным [14]. Для решения задачи оптимального выбора модели, необходимо найти метод μ_t с наилучшей обобщающей способностью.

Частные случаи задачи выбора модели:

- Выбор лучшей модели A_t (например, когда из десяти построенных моделей машинного обучения выбирается одна наилучшая по заданным параметрам);
- Выбор метода обучения μ_t для заданной модели A_t (например, оптимизация гиперпараметров модели, пересчет весов модели, степень полинома при полиномиальной аппроксимации, число слоев в нейронной сети и так далее);
- Отбор признаков:

$F = \{f_j : X \to D_y : j = 1, \ldots, n\}$ – *множество признаков; метод обучения* μ_t *использует только признаки* $J \subseteq F$.

Оценка качества моделей обучения по прецедентам может быть выполнена и наиболее простым способом — тестированием алгоритма на отложенной выборке. Однако от способа формирования этой выборки зависит конечный результат точности модели машинного обучения, что является ненадежным показателем (13) для оценки модели, так как каждый раз при формировании отложенной выборки необходимо учитывать её репрезентативность.

$$X^L = X^l \sqcup X^k \qquad (13)$$

Формула (13) отражает формализованную математическую зависимость оценки от способа разбиения выборки. При формировании отложенной выборки также необходимо следить и за оптимумом сложности (рисунок 5). На рисунке 5 представлены характеристики зависимости внешнего и внутреннего критерия.

Внутренний критерий монотонно убывает с ростом сложности модели машинного обучения (например, от возрастания количества признаков).

Внешний критерий имеет характерную точку глобального экстремума (минимума), при этом данное значение является оптимальной сложностью модели.

На рисунке 5 в точке n = 21 заметен характерный минимум. После него ошибка вновь начинает монотонно возрастать. С точки зрения интерпретируемости результатов, точка n = 21 сообщает о том, что невозможно по слишком малочисленной выборке (для возросшего значения сложности модели до n = 21) настраивать большое количество параметров модели.

Когда модель становится избыточно сложной, возникает эффект переобучения. По этой причине определять точность всегда необходимо по внешнему критерию. Внешний критерий в выборке данных — это критерий, у которого зависимость от сложности имеет

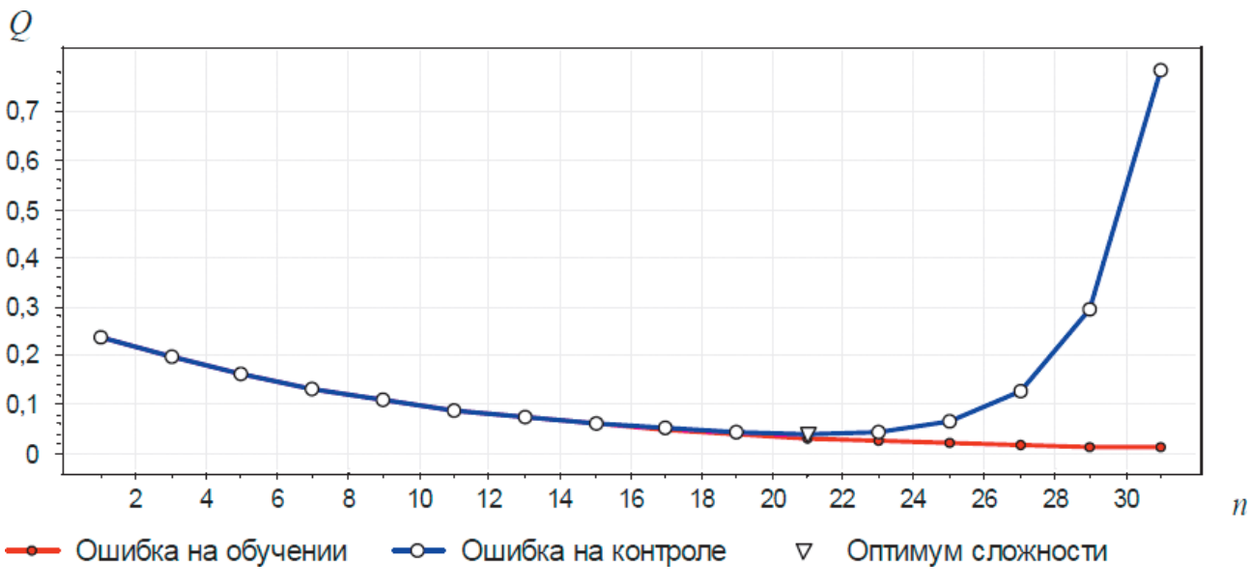

Рис. 5. Зависимость внешнего и внутреннего критерия контроля точности модели машинного обучения

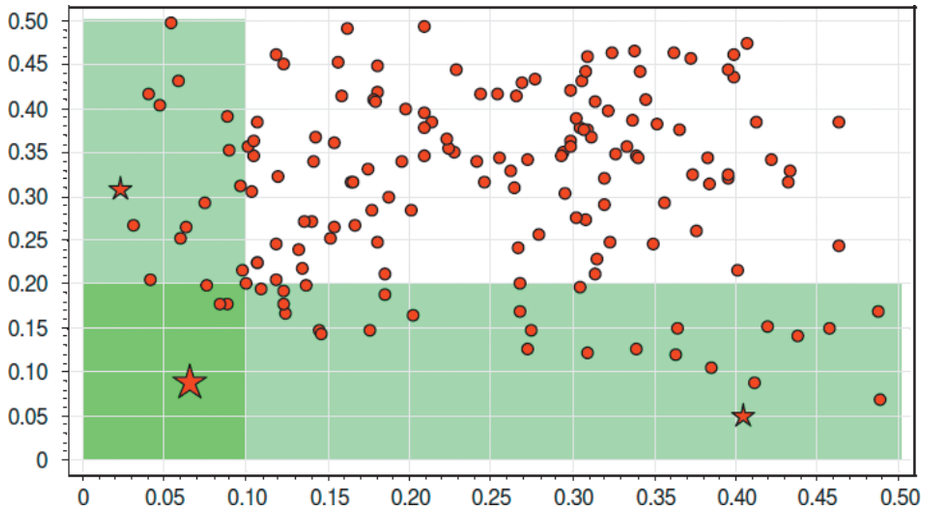

Рис. 6. Перевес «неоптимальности» модели по другому параметру за счет сильного перевеса в сторону одного оптимального критерия

унимодальный характер. В формализованном математическом языке тезис может быть оформлен в виде формулы (14).

$Q(J) = Q(\mu_j, X^l)$ – выбранный внешний критерий

$$Q(J) \to \min \qquad (14)$$

В некоторых случаях может существовать более одного критерия (особенно в задачах многоклассовой классификации) [12]. В таком случае необходимо понимать, что модель, немного неоптимальная по обоим критериям, чаще всего оказывается лучше, чем модель, оптимальная по одному критерию и сильно неоптимальная по другому критерию. Пример сильной «неоптимальности» представлен на рисунке 6.

Из рисунка 6 отчетливо следует, что идеально оптимизированная (с полосой устойчивости) модель по критерию «Звездочка» многократно теряет в точности классификации на критерии «Точка».

Величина результирующей ошибки дихотомической классификации приведет модель к заметному снижению точности по метрикам **(7)** и **(8)**. При этом теряется весь смысл оптимальности по одному критерию, так как общая точность оказывается существенно меньшей точности оптимальности модели по совокупности критериев.

СПИСОК ЛИТЕРАТУРЫ

1. Приказ Минтруда России от 31.01.2022 N 36 «Об утверждении Рекомендаций по классификации, обнаружению, распознаванию и описанию опасностей» [Электронный ресурс]. Режим доступа: Электронный ресурс http;//www/.consultant.ru/document/cons_doc_LAW_408713/a5a7ef19e6ef97eb94e86607fd0f0db8077f77fa/ (дата обращения: 15.02.2024).

2. Киселев А. Промышленная безопасность опасных производственных объектов. Альфа-Пресс. 2017. 240 с.

3. Dmitriy Serdyuk, Otavio Braga, and Olivier Siohan. Audio-visual speech recognition is worth 32x32x8 voxels. 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021.

4. Yi-Lin Sung, Jaemin Cho, and Mohit Bansal. Vl-adapler: Parameter-efficient transfer learning for vision and language tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 5227–5237,

5. Tejas Srinivasan, Ramon Sanabria, and Florian Metze. Looking enhances listening: Recovering missing speech using im- ages. 1CASSP, 2020.

6. Gabriel Synnaeve, Qiantong Xu, Jacob Kalin, Tatiana Likhomanenko, Edouard Grave, Vineel Pratap, Anuroop Sri- ram, Vitaliy Liptchinsky, and Ronan Collobert. End-to-end asr: from supervised to semi-supervised learning with modem architectures. arXiv preprint ar Xiv: 1911.08460, 2019.

7. C. M. Bishop. Pattern Recognition and Machine Learning. Springer. 2006. 738 с.

8. Солтис М. Введение в анализ алгоритмов. СПб.: ДМК. 2019. 269 с.

9. Домингос П. Верховный алгоритм. Манн, Иванов и Фербер. 2016. 315 с.

10. Гифт Н. Программный ИИ. СПб.: Питер. 2019. 304 с.

11. Кормен Т., Лейзерсон Ч. Алгоритмы: построение и анализ, 3-е издание. М.: ООО «И.Д. Вильямс». 2013. 1328 с.

12. Qiantong Xu, Alexei Baevski, Tatiana Likhomanenko, Paden Tomasello, Alexis Conneau, Ronan Collobert, Gabriel Synnaeve, and Michael Auli. Self-training and pre-training are complementary for speech recognition. In ICASSP, 2021.

13. Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loi'c Barrault, Lucia Specia. and Florian Metze. How2: a large-scale dataset for multimodal language un- derstanding. In Proceedings of the Workshop on Visually Grounded Interaction and Language (ViGIL), 2018.

14. Пылов П. А., Майтак Р. В., Зайцева Е. Г. Применение мультимодального трансформера для прогнозирования выходных параметров насыщенных углеводородных соединений из состава тяжелой нефти в присутствии катализаторов // Труды Института системного программирования РАН. 2023. Т. 35. № 5. С. 229–244.

DOI: 10.25558/VOSTNII.2025.47.21.003
UDC 622.831

A. V. DYAGILEVA
Candidate of Engineering Sciences
Associate Professor
KuzSTU, Kemerovo
e-mail: dyagileva1952@mail.ru

P. A. PYLOV
Graduate Student
KuzSTU, Kemerovo
e-mail: gedrosten@mail.ru

R. V. MAJTAK
Master
KuzSTU, Kemerovo
e-mail: maytak.roman@mail.ru

E. A. NIKOLAEVA
Candidate of Physical
and Mathematical Sciences, Associate Professor,
Head of Department
KuzSTU, Kemerovo
e-mail: nikolaevaea@yandex.ru

METRICS AND QUALITY CRITERIA OF DEEP LEARNING MODELS FOR EVALUATING THE ACCURACY OF HIGH-ENERGY SEISMIC SHOCK PREDICTION

Seismic hazard is one of the factors that endanger human life and health. In the case of mining workers, it is one of the key factors because seismic shocks in the ground can lead to the destruction of rocks, breakage and collapse of mine structures, and blockage of evacuation routes. Seismic hazard is a natural hazard and predicting its occurrence is a creative and very labor-intensive task. The preventive measures proposed by mankind several hundred years ago do not always make it possible to identify a disaster with sufficient accuracy and to prevent its consequences. With the development of information technologies (in particular applied artificial intelligence) and the development of hardware and software,

it has become possible to use intelligent systems to automate various technological tasks. Mining is no exception - various modifications of deep neural networks and self-learning machine learning models can be successfully integrated to automate the process of predicting high-energy seismic shocks. However, in order to successfully implement a model of applied artificial intelligence, it is first necessary to formalize a quality metric by which it will be possible to judge which models perform better and which perform worse. In this paper, the authors have presented the results of a study on the application of different metrics to assess the quality of seismic shock prediction. The authors have presented and studied in detail the criteria whose generalization of accuracy best correlates with the effectiveness of the mining problem. The value of the research is that in the future researchers will not have to spend time, human, material and other resources to select the optimal quality metric to be used as an indicator of the accuracy of the automated intelligent model. It is also an important fact that within the framework of the article a separate attention was focused on the mathematical foundations of quality metrics, so the choice of programming language in which the machine learning model will be further implemented is not critical.

Keywords: MINE, COAL INDUSTRY, DATA MINING, INTELLIGENT DATA ANALYSIS, INTELLIGENT LANGUAGE MODELS, SEISMIC DATA PROCESSING BY DEEP AND MACHINE LEARNING MODELS.

REFERENCES

1. Order of the Ministry of Labor of the Russian Federation No. 36 dated 01/31/2022 «On approval of Recommendations on classification, detection, recognition and description of hazards» [Electronic resource]:http;//www/.consultant.ru/document/cons_doc_LAW_408713/a5a7ef19e6ef97eb94e86607fd0f0db8077f77fa/ (date of request: 15.02.2024). [In Russ].

2. Kiselev A. Industrial safety of hazardous production facilities. Alfa-Press. 2017. 240 p. [In Russ].

3. Dmitriy Serdyuk, Otavio Braga, and Olivier Siohan. Audio-visual speech recognition is worth 32x32x8 voxels. 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021.

4. Yi-Lin Sung, Jaemin Cho, and Mohit Bansal. Vl-adapler: Parameter-efficient transfer learning for vision and language tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. P. 5227–5237,

5. Tejas Srinivasan, Ramon Sanabria, and Florian Metze. Looking enhances listening: Recovering missing speech using im- ages. 1CASSP, 2020.

6. Gabriel Synnaeve, Qiantong Xu, Jacob Kalin, Tatiana Likhomanenko, Edouard Grave, Vineel Pratap, Anuroop Sri- ram, Vitaliy Liptchinsky, and Ronan Collobert. End-to-end asr: from supervised to semi-supervised learning with modem architectures. arXiv preprint ar Xiv: 1911.08460, 2019.

7. C. M. Bishop. Pattern Recognition and Machine Learning. Springer. 2006. 738 p.

8. Soltis M. Introduction to the analysis of algorithms. St. Petersburg: DMK. 2019. 269 p. [In Russ].

9. Domingos P. The Supreme algorithm. Mann, Ivanov and Ferber. 2016. 315 p. [In Russ].

10. Gift N. Program AI. St. Petersburg: Peter. 2019. 304 p. [In Russ].

11. Kormen T., Lazerson C. Algorithms: construction and analysis, 3rd edition. Moscow: I.D. Williams LLC. 2013. 1328 p. [In Russ].

12. Qiantong Xu, Alexei Baevski, Tatiana Likhomanenko, Paden Tomasello, Alexis Conneau, Ronan Collobert, Gabriel Synnaeve, and Michael Auli. Self-training and pre-training are complementary for speech recognition. In ICASSP, 2021.

13. Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Desmond Elliott, Loi'c Barrault, Lucia Specia. and Florian Metze. How2: a large-scale dataset for multimodal language un- derstanding. In Proceedings of the Workshop on Visually Grounded Interaction and Language (ViGIL), 2018.

14. Pylov P. A., Maitak R. V., Zaitseva E. G. Application of a multimodal transformer for predicting the output parameters of saturated hydrocarbon compounds from heavy oil in the presence of catalysts // Proceedings of the Institute of System Programming of the Russian Academy of Sciences. 2023. Vol. 35. No. 5. P. 229–244. [In Russ].