Эквивалентность порога метода Оцу решающему правилу MAP-классификатора в задаче обнаружения событий информационной безопасности

А.Я. Бучаев, И.И. Комаров

***Аннотация* — Исследована связь эмпирического порога классического метода Оцу с решающим правилом байесовского классификатора с максимальной апостериорной вероятностью. Установлено, что строгая эквивалентность двух порогов имеет место при выполнении двух условий: равенства дисперсий классов и равенства априорных вероятностей. При снятии условия равенства априорных вероятностей (при сохранении гомоскедастичности) пороги расходятся на аналитически выписываемую величину, пропорциональную логарифму отношения априорных вероятностей. В общем гетероскедастическом случае MAP-классификатор задает квадратичную границу, не представимую единственным скалярным порогом, и метод Оцу остается ее линейным приближением. Таким образом, совпадение порогов реализуется только в частном случае, а за его пределами получено аналитическое выражение величины отклонения. Полученный результат позволяет рассматривать применение метода Оцу в задаче обнаружения событий информационной безопасности как теоретически обоснованную процедуру, а не эвристику, и связывает эмпирический критерий с вероятностной моделью наблюдаемых признаков.**

***Ключевые слова* — апостериорная вероятность, байесовский классификатор, информационная безопасность, коэффициент разделимости, межклассовая дисперсия, метод Оцу, обнаружение аномалий, сетевой трафик.**

I. Введение

Задача автоматического выделения событий информационной безопасности (ИБ) в потоке событий системы мониторинга сводится к бинарной классификации одномерной или многомерной статистики на два класса, отвечающие нормальному и аномальному состояниям наблюдаемого процесса. Среди непараметрических методов построения порога наиболее распространен критерий Оцу, предложенный в 1979 году для сегментации полутоновых изображений [1]. Критерий выбирает порог, максимизирующий межклассовую дисперсию гистограммы признака, и не требует предположений о виде распределения наблюдений. Это свойство послужило основанием для переноса метода в задачи обнаружения событий ИБ, где априорная параметрическая модель распределения признаков, как правило, отсутствует.

Вместе с тем вероятностно оптимальным в смысле минимума среднего риска остается байесовский классификатор, принимающий решение по максимуму апостериорной вероятности (далее – MAP) [2]. Между двумя подходами существует связь, впервые явно отмеченная в работе Куриты, Оцу и Абдельмалека [3], где показано, что метод Оцу соответствует оценке параметров смеси двух гауссовских распределений методом условного максимума правдоподобия при равных дисперсиях. В работе Киттлера и Иллингворта был предложен критерий минимальной ошибки [4], в котором порог определяется прямо из модели смеси и в гетероскедастическом случае уже не совпадает с порогом Оцу. Дальнейшие исследования уточнили соотношение этих критериев [5], [6]. В отечественной литературе теоретическая связь порога Оцу с решающим правилом MAP в приложении к задачам ИБ до сих пор систематически не рассматривалась.

Настоящая работа продолжает цикл публикаций, посвященных автоматической классификации событий ИБ [7], [8], и посвящена теоретическому обоснованию эмпирического порога метода Оцу. Показано, что в условиях, выполняющихся для большинства практических задач обнаружения событий ИБ в компьютерных сетях, этот порог совпадает с байесовским оптимальным. Отличие предлагаемого рассмотрения от классических работ состоит в явном последовательном усложнении модели: от случая равных дисперсий и равных априорных вероятностей к случаю различающихся априорных вероятностей и далее к общему гетероскедастическому случаю, а также в интерпретации возникающих поправок в терминах задачи обнаружения событий ИБ, в которой априорная вероятность аномального класса существенно меньше априорной вероятности нормального класса.

Бучаев Абдулхамид Яхьяевич, аспирант факультета безопасности информационных технологий, Университет ИТМО, Санкт-Петербург, Россия. https://orcid.org/0009-0001-1058-9125 E-mail: abdulhamid0055@yandex.ru

Комаров Игорь Иванович, кандидат физико-математических наук, доцент, доцент факультета безопасности информационных технологий, Университет ИТМО, Санкт-Петербург, Россия. E-mail: i_krov@mail.ru

II. ПОСТАНОВКА ЗАДАЧИ

Пусть наблюдается скалярная статистика x, получаемая из вектора признаков сетевого соединения путем агрегирования и нормировки. Принимается, что x порождается одним из двух классов: нормальным состоянием (класс 0) и аномальным состоянием (класс 1). Априорные вероятности классов обозначаются P_0 и P_1, причем $P_0 + P_1 = 1$. Условные плотности распределения обозначаются $p(x|0)$ и $p(x|1)$, математические ожидания – μ_0 и μ_1, дисперсии – σ_0^2 и σ_1^2. Без ограничения общности принимается $\mu_0 < \mu_1$.

Решающее правило представляет собой скалярный порог θ: значению $x \le \theta$ присваивается метка класса 0, значению $x > \theta$ – метка класса 1. Сравниваются два способа выбора порога: непараметрический критерий Оцу, использующий только эмпирическое распределение x, и байесовский критерий MAP, использующий параметрическую модель $p(x|k)$ и априорные вероятности P_k. Цель работы – установить условия, при которых оба порога совпадают, и выразить разность между ними в условиях, когда совпадение отсутствует.

.

III. СВЯЗЬ ПОРОГА ОЦУ С РЕШАЮЩИМ ПРАВИЛОМ MAP-КЛАССИФИКАТОРА

A. MAP-классификатор для бинарной задачи

Решающее правило максимума апостериорной вероятности имеет вид: объект, породивший наблюдение x, относится к тому классу k, для которого произведение $P_k p(x|k)$ максимально. По формуле Байеса это эквивалентно максимизации апостериорной вероятности:

$$k_{\text{MAP}} = \arg\max_k P(k|x) = \arg\max_k P_k p(x|k), \qquad (1)$$

где $k \in \{0,1\}$ – метка класса; $P(k|x)$ – апостериорная вероятность класса k при наблюдении x; P_k – априорная вероятность класса k; $p(x|k)$ – условная плотность распределения значений x в классе k.

Граница решения в бинарной задаче задается условием равенства произведений $P_0 p(x|0) = P_1 p(x|1)$. После логарифмирования условие принимает форму равенства логарифма отношения правдоподобий логарифму отношения априорных вероятностей:

$$L(x) = \ln\left(\frac{p(x|1)}{p(x|0)}\right) = \ln\left(\frac{P_0}{P_1}\right), \qquad (2)$$

где $L(x)$ – логарифм отношения правдоподобий; правая часть отражает априорный дисбаланс классов и не зависит от x. При равных априорных вероятностях правая часть обращается в ноль, и граница определяется только соотношением правдоподобий.

Для гауссовских классов логарифм отношения правдоподобий при равных дисперсиях линеен по x, а при различных дисперсиях – квадратичен; эта особенность далее определяет характер получающегося порога.

B. Критерий Оцу как оптимизационная задача

В исходной формулировке Оцу [1] порог определяется как точка, в которой достигается максимум межклассовой дисперсии распределения. Обозначим через $w_0(\theta)$ и $w_1(\theta)$ вероятностные веса классов, получаемые при разбиении области значений случайной величины x пороговой точкой θ, а через $\mu_0(\theta)$ и $\mu_1(\theta)$ – условные математические ожидания в получившихся подмножествах. Межклассовая дисперсия при пороге θ задается формулой:

$$\sigma_B^2(\theta) = w_0(\theta) w_1(\theta) (\mu_0(\theta) - \mu_1(\theta))^2, \qquad (3)$$

где $w_0(\theta)$ – вероятность попадания x в интервал $(-\infty, \theta]$; $w_1(\theta) = 1 - w_0(\theta)$; $\mu_0(\theta)$ и $\mu_1(\theta)$ вычисляются как условные средние по соответствующим подмножествам.

Справедливо классическое тождество разложения дисперсии $\sigma_T^2 = \sigma_W^2(\theta) + \sigma_B^2(\theta)$, где σ_T^2 – полная дисперсия, не зависящая от θ, $\sigma_W^2(\theta)$ – внутриклассовая дисперсия, $\sigma_B^2(\theta)$ – межклассовая дисперсия. Следовательно, максимизация межклассовой дисперсии эквивалентна минимизации внутриклассовой. В качестве нормированной меры разделимости вводится коэффициент:

$$\eta(\theta) = \frac{\sigma_B^2(\theta)}{\sigma_T^2}, \qquad (4)$$

где η принимает значения в отрезке [0, 1] и совпадает с эта-квадратом однофакторного дисперсионного анализа.

Оптимальное значение порога θ_O определяется из необходимого условия стационарности $\frac{d\sigma_B^2}{d\theta} = 0$. Применяя тождество баланса средних $w_0\mu_0 + w_1\mu_1 = \mu_T$, где μ_T – полное среднее распределения, и дифференцируя σ_B^2, можно показать [9], что стационарная точка удовлетворяет уравнению:

$$\theta_O = \frac{\mu_0(\theta_O) + \mu_1(\theta_O)}{2}, \qquad (5)$$

где $\mu_0(\theta_O)$ и $\mu_1(\theta_O)$ – условные средние классов при оптимальном пороге. Это тождество указывает, что порог Оцу всегда лежит в середине между условными средними двух классов, вычисленными при том же пороге. Следствием является известное смещение порога в сторону класса с большей дисперсией [9].

C. Случай равных априорных вероятностей и равной дисперсии

Рассматривается базовый случай: $p(x|k) = N(\mu_k, \sigma^2)$, $k \in \{0,1\}$; $P_0 = P_1 = 1/2$. Логарифм отношения правдоподобий принимает форму:

$$L(x) = \frac{\mu_1 - \mu_0}{\sigma^2}\left(x - \frac{\mu_0 + \mu_1}{2}\right), \qquad (6)$$

где σ^2 – общая дисперсия классов, μ_0 и μ_1 – их

математические ожидания. Подставляя условие MAP $L(x) = \ln\left(\frac{P_0}{P_1}\right) = 0$, получаем линейное уравнение, единственный корень которого:

$$\theta_{\text{MAP}} = \frac{\mu_0 + \mu_1}{2}. \quad (7)$$

Для критерия Оцу рассмотрим полную плотность распределения x, представляющую собой взвешенную сумму условных плотностей классов – плотность смеси двух гауссовских компонент: $p(x) = \frac{1}{2}N(x;\, \mu_0, \sigma^2) + \frac{1}{2}N(x;\, \mu_1, \sigma^2)$. В силу симметрии двух одинаковых по форме гауссовских компонент относительно точки $m = \frac{\mu_0+\mu_1}{2}$ плотность $p(x)$ удовлетворяет тождеству $p(m - t) = p(m + t)$ для любого t. Отсюда следует, что при $\theta = m$ выполняются равенства $w_0(m) = w_1(m) = 1/2$ и $\mu_T = m$, а условие (5) удовлетворяется автоматически, что дает:

$$\theta_O = \frac{\mu_0 + \mu_1}{2} = \theta_{\text{MAP}}. \quad (8)$$

Таким образом, в рассматриваемом базовом случае эмпирический порог Оцу точно совпадает с порогом MAP-классификатора. Это совпадение не является случайным, оно отражает тот факт, что в гомоскедастической модели с равными априорными вероятностями симметрия правдоподобий и симметрия межклассовой дисперсии задают одну и ту же точку.

D. Случай различающихся априорных вероятностей и равной дисперсии

Сохраняется предположение о равенстве дисперсий $\sigma_0^2 = \sigma_1^2 = \sigma^2$, но снимается условие равенства априорных вероятностей. Логарифм отношения правдоподобий остается линейным по x. Условие MAP принимает вид:

$$\frac{\mu_1 - \mu_0}{\sigma^2}\left(x - \frac{\mu_0 + \mu_1}{2}\right) = \ln\left(\frac{P_0}{P_1}\right). \quad (9)$$

Решение относительно x дает порог MAP-классификатора:

$$\theta_{\text{MAP}} = \frac{\mu_0 + \mu_1}{2} + \frac{\sigma^2}{\mu_1 - \mu_0}\ln\left(\frac{P_0}{P_1}\right), \quad (10)$$

где второе слагаемое представляет собой свободный член, сдвигающий порог относительно среднего арифметического математических ожиданий. Свободный член положителен, если $P_0 > P_1$ (класс 0 более вероятен), и сдвигает границу в сторону μ_1, уменьшая область принятия решения о более редком классе. При $P_0 = P_1$ свободный член обращается в ноль, и формула сводится к результату предыдущего подраздела.

Классический порог Оцу в этом случае, вообще говоря, отличен от θ_{MAP}. В пределе хорошей разделимости классов, когда расстояние между средними существенно превосходит стандартное отклонение ($|\mu_1 - \mu_0| \gg \sigma$), хвосты гауссовских компонент в области между средними практически не перекрываются, и условные средние $\mu_0(\theta)$ и $\mu_1(\theta)$ при любом θ из интервала (μ_0, μ_1) остаются близкими к μ_0 и μ_1 соответственно. В этих условиях из тождества (5) следует $\theta_O \approx \frac{\mu_0+\mu_1}{2}$, то есть классический Оцу при хорошей разделимости не видит априорного дисбаланса и дает порог, соответствующий случаю равных априорных вероятностей. Отсюда разность порогов равняется:

$$\Delta\theta = \theta_{\text{MAP}} - \theta_O \approx \frac{\sigma^2}{\mu_1 - \mu_0}\ln\left(\frac{P_0}{P_1}\right). \quad (11)$$

Зависимость разности от параметров задачи имеет следующую структуру. Она линейна по логарифму отношения априорных вероятностей, пропорциональна дисперсии класса и обратно пропорциональна расстоянию между средними. Иначе говоря, сдвиг MAP-порога относительно порога Оцу значителен при слабой разделимости классов и большой асимметрии априорных вероятностей и пренебрежимо мал при сильной разделимости или близких априорных вероятностях. Для численной иллюстрации при $\mu_0 = 0, \mu_1 = 5, \sigma = 1, P_0 = 0{,}95, P_1 = 0{,}05$ получается $\ln\left(\frac{P_0}{P_1}\right) = \ln 19 \approx 2{,}944$, $\Delta\theta \approx 0{,}2 \cdot 2{,}944 \approx 0{,}589$. Таким образом, MAP-порог сдвинут приблизительно на 0,59 единицы вправо относительно порога Оцу, что в задаче обнаружения событий ИБ соответствует расширению области принятия решения о нормальном классе.

E. Общий случай и границы применимости

Снимается предположение о равенстве дисперсий. Логарифм отношения правдоподобий двух гауссовских плотностей становится квадратичным по x, и условие MAP порождает квадратное уравнение $Ax^2 + Bx + C = 0$ с коэффициентами:

$$A = \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}, \quad (12)$$

$$B = -2\left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right), \quad (13)$$

$$C = \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_1^2}{\sigma_1^2} - 2\ln\left(\frac{P_0\sigma_1}{P_1\sigma_0}\right), \quad (14)$$

где A – квадратичный коэффициент, обращающийся в ноль при равных дисперсиях; B – линейный коэффициент; C – свободный член, включающий логарифмическую поправку, зависящую от отношения априорных вероятностей и отношения стандартных отклонений классов.

Уравнение имеет, вообще говоря, два действительных корня. Практический выбор одного из них состоит в предпочтении корня, лежащего в промежутке между μ_0 и μ_1, тогда как второй корень обычно расположен далеко в хвосте распределения с большей дисперсией и соответствует реклассификациям при крайне малых

плотностях, несущественным для задачи обнаружения событий ИБ. При $\sigma_0 \to \sigma_1$ коэффициент A обращается в ноль, и квадратное уравнение вырождается в линейное, из которого восстанавливается формула предыдущего подраздела. Это подтверждает непрерывный переход между случаями и согласованность полученных результатов.

В общем гетероскедастическом случае порог Оцу, оставаясь по построению скалярным и порождающим линейное решающее правило, не может совпасть с MAP, задающим квадратичную границу. Оцу в этой ситуации дает наилучшее линейное приближение байесовского классификатора и, как показано в [9], систематически смещается в сторону класса с большей внутриклассовой дисперсией.

Границы, в которых отклонение пренебрежимо, задаются тремя условиями: отношение дисперсий близко к единице, расстояние между средними превосходит несколько стандартных отклонений (традиционно не менее трех), и модуль логарифма отношения априорных вероятностей умерен. При выполнении этих условий поправка к порогу Оцу либо обращается в ноль, либо оказывается существенно меньше самой величины порога, и метод Оцу сохраняет статус практического эквивалента MAP-классификатору

IV. Практические следствия для обнаружения событий ИБ в компьютерных сетях

В задаче обнаружения событий ИБ практические условия, при которых возможна эквивалентность порога Оцу порогу MAP, встречаются систематически. После нормировки агрегированного признака распределение значений нормального класса в большинстве случаев удовлетворительно аппроксимируется гауссовским, а распределение аномального класса в пространстве отобранных признаков сохраняет унимодальность. Это соответствует условиям случая с равными дисперсиями, при сбалансированных априорных вероятностях реализуется также случай равных весов классов, в котором пороги Оцу и MAP совпадают строго.

Скалярное расстояние вектора состояния устройства до центроида скользящего окна является суммой вкладов множества независимых компонент признакового пространства: частот пакетов по протоколам, размеров полезной нагрузки, временных интервалов между соединениями, распределений по портам. Центральная предельная теорема обеспечивает сходимость такой суммы к нормальному закону при штатной работе устройства, когда каждая компонента представляет собой стационарный случайный процесс с ограниченной дисперсией. Эмпирические наблюдения в наборах сетевых данных CIC-IDS [10], NSL-KDD [11], UNSW-NB15 [12] подтверждают, что после z-преобразования или логарифмической нормировки распределение расстояний от центроида для нормальных состояний хорошо описывается гауссовской моделью с коэффициентом асимметрии в пределах 0,3-0,5 и эксцессом, близким к трем.

Математические ожидания μ_0 и μ_1 и дисперсии σ^2 в реальных сетях не являются фиксированными константами, они меняются в зависимости от времени суток, дня недели, сезонных шаблонов нагрузки, изменения состава устройств сети. Суточные колебания среднего числа активных соединений достигают двух-трех раз, распределение по протоколам заметно смещается между рабочими и нерабочими часами. Для корректного применения поправки $\Delta\theta$ необходимо оценивать параметры распределения по скользящему окну соответствующей длительности, согласованной с характерным временем нестационарности. Практическая рекомендация: при интервале дискретизации τ порядка десяти секунд длина окна ретроспективы должна составлять 60-120 интервалов, что соответствует 10-20 минутам, этого достаточно для стабилизации оценок параметров и одновременно меньше характерного времени суточной нестационарности.

Решающее значение имеет априорный дисбаланс классов. В типичных наборах сетевого трафика доля редких классов атак составляет доли процента, и априорная вероятность аномального класса существенно меньше априорной вероятности нормального. При таких соотношениях свободный член $\frac{\sigma^2}{\mu_1-\mu_0}\ln\left(\frac{P_0}{P_1}\right)$ играет роль систематической поправки, смещающей байесовский оптимальный порог в сторону большего математического ожидания, то есть в сторону аномального класса. Игнорирование этой поправки приводит к завышенному числу ложных тревог при редких аномалиях. Использование поправки на практике осложняется тем, что априорные вероятности классов в реальной сети неизвестны и меняются во времени вместе с профилем эксплуатации. Возможные способы получения рабочих оценок P_0 и P_1 рассмотрены далее.

A. Оценка априорных вероятностей в режиме мониторинга

Оценка априорных вероятностей для вычисления поправки может быть построена тремя способами, различающимися источником информации и точностью.

Первый способ – априорная экспертная оценка. Для сети с устоявшимся профилем эксплуатации доля временных интервалов с аномальной активностью оценивается по статистике предшествующих инцидентов, отчетам центров реагирования на инциденты ИБ и данным эталонных наборов сетевого трафика. Для типичных корпоративных сетевых инфраструктур принимается значение P_1 в диапазоне от 10^{-4} до 10^{-2}, что соответствует логарифму отношения $\ln\left(\frac{P_0}{P_1}\right)$ в диапазоне от 4,6 до 9,2. Такая оценка груба, но устойчива во времени и обеспечивает корректный знак и порядок величины поправки. Способ применим, если характеристики сети достаточно стационарны, а изменения профиля эксплуатации происходят медленно.

Второй способ – рекурсивная оценка по вспомогательному скользящему окну. Пусть оперативное окно размера N используется для формирования порога Оцу по описанной схеме, а вспомогательное окно размера M, существенно

превосходящего N (с отношением M/N порядка 50-100), служит для подсчета доли наблюдений, получивших решение об отнесении к аномальному классу на предыдущих шагах. Эта доля принимается в качестве текущей оценки P_1 и подставляется в формулу поправки на следующем шаге. Схема является адаптивной и позволяет системе самонастраиваться при изменении профиля сети, однако требует разделения двух временных масштабов. Оперативное окно N отражает текущее локальное поведение устройства и определяет чувствительность метода к быстрым переходам. Вспомогательное окно M отражает усредненную статистику частоты аномалий и должно быть достаточно большим, чтобы случайные флуктуации доли не приводили к неустойчивости порога. Сходимость схемы обеспечивается при условии, что скорость изменения истинной априорной вероятности во времени существенно меньше обратной длительности вспомогательного окна, и что доля ошибочных решений, попадающих в M, мала по сравнению с долей корректных. Дополнительным условием является разнесение масштабов N и M, препятствующее положительной обратной связи между оперативным и вспомогательным контурами: ошибочное срабатывание в оперативном окне не должно заметно сдвигать оценку P_1, поскольку его вклад в окне M разбавляется большим числом штатных наблюдений.

Третий способ – робастная интервальная оценка. При полной неопределенности P_1 поправка вычисляется для двух предельных значений, соответствующих пессимистическому и оптимистическому сценариям — например, $P_1 = 10^{-4}$ и $P_1 = 10^{-1}$ соответственно. Для каждого значения получается свой скорректированный порог, а итоговое решение о наличии аномалии принимается, только если оно сохраняется при обоих предельных значениях поправки. В противном случае выдается сигнал неопределенности, требующий дополнительной проверки или вмешательства аналитика ИБ. Такой подход избавляет от необходимости точной оценки P_1 в большинстве практических ситуаций и переводит неопределенность априорной вероятности в интервальную форму, более подходящую реальным условиям мониторинга ИБ.

Нестационарность априорных вероятностей является самостоятельным осложнением, не устранимым подбором схемы оценки. Суточные и недельные циклы работы сети, плановые технические окна, периоды усиленной активности перед отчетными датами изменяют базовое значение P_1 в пределах одного-двух порядков. Прямое применение константного значения P_1 без учета цикличности приводит к систематической ошибке поправки в периоды пониженной или повышенной активности. Для учета нестационарности оценка P_1 во втором способе должна обновляться с характерным временем, согласованным с доминирующим циклом бизнес-процессов: при суточной цикличности вспомогательное окно выбирается порядка одного часа, при недельной порядка суток. Для систем с выраженной многомасштабной нестационарностью возможно введение нескольких вспомогательных окон разной длительности с объединением их оценок взвешенным усреднением.

B. Поведение поправки на различных этапах атаки

На фазе разведки (network reconnaissance) аномальный трафик характеризуется низкой интенсивностью и малым отличием от штатной работы: $(\mu_1 - \mu_0)$ снижается, поправка $\Delta\theta = \frac{\sigma^2}{\mu_1 - \mu_0} \ln\left(\frac{P_0}{P_1}\right)$ принимает большие значения, что снижает чувствительность к ранним признакам атаки. Это отражает естественную трудность обнаружения разведки и согласуется с известными эмпирическими наблюдениями. На фазе активного воздействия (отравление таблицы MAC-адресов, разрыв TCP-соединения) различие средних велико, поправка мала в относительном выражении [13], порог Оцу дает результат, практически совпадающий с MAP-порогом. На фазе закрепления и латерального перемещения, характеризующейся длительными низкоинтенсивными аномалиями с дисбалансом порядка $P_1 \approx 0{,}01$, поправка снова становится значимой и должна явно вводиться для сохранения качества обнаружения. Эта фазовая зависимость предоставляет аналитику ИБ инструмент для адаптивной настройки, величина применяемой поправки должна зависеть от предполагаемого этапа атаки или, при отсутствии такого предположения или от текущей оценки $(\mu_1 - \mu_0)$ по наблюдаемым данным.

Методы на основе глубоких автокодировщиков (Kitsune) [14], статистических потоковых признаков (NFStream, CICFlowMeter) [15], [16] требуют репрезентативной обучающей выборки и не допускают прямой интерпретации порога в терминах вероятности ошибки. При смене условий эксплуатации, например, обновлении конфигурации, вводе новых устройств, изменении профиля нагрузки, эти методы требуют переобучения на новой размеченной выборке, что в корпоративной среде затруднено отсутствием своевременной и корректной разметки. Предложенный теоретический аппарат позволяет применять непараметрический метод Оцу с аналитически обоснованной поправкой, адаптируясь к текущему дисбалансу классов без переобучения. Дополнительное преимущество выражается в возможности декомпозиции решения, положительный вердикт о наличии события ИБ сопровождается явными значениями σ^2, $(\mu_1 - \mu_0)$, $\ln\left(\frac{P_0}{P_1}\right)$, позволяющими оценить основания решения и исключить его при необходимости, что невозможно в нейросетевых классификаторах.

Полученный теоретический результат позволяет сформулировать практическую процедуру: применять критерий Оцу для формирования порога по наблюдаемой смеси, а при наличии оценки априорного дисбаланса классов корректировать полученный порог аналитической поправкой, соответствующей свободному члену. Такая корректировка приближает эмпирический порог к байесовскому оптимальному без перехода к полной параметрической модели и без необходимости

оценки параметров отдельных классов.

V. Заключение

Строгая эквивалентность порога Оцу порогу MAP-классификатора имеет место лишь при одновременном выполнении двух условий: равенства дисперсий классов и равенства априорных вероятностей. При нарушении условия равенства априорных вероятностей пороги расходятся на величину, пропорциональную логарифму отношения априорных вероятностей и отражающую смещение границы в сторону класса с большей априорной вероятностью. В гетероскедастическом случае линейный порог Оцу принципиально не может совпасть с квадратичной границей MAP и представляет собой ее линейное приближение с систематическим смещением в сторону класса с большей дисперсией. Таким образом, область строгой эквивалентности узка, однако за ее пределами величина отклонения выражена аналитически, и условия ее пренебрежимости сформулированы явно.

Новизна проведенного рассмотрения состоит в том, что эмпирический порог, традиционно обосновываемый соображениями максимальной разделимости, получает прямую вероятностную интерпретацию через правило максимума апостериорной вероятности, а поправка, возникающая при несбалансированных классах, выписывается явной аналитической формулой через параметры модели и непосредственно применима в расчетах. Это позволяет рассматривать применение метода Оцу в задаче обнаружения событий информационной безопасности не как эвристику, а как теоретически обоснованное приближение байесовского классификатора. Полученный результат создает основание для дальнейшего уточнения методов автоматической классификации событий информационной безопасности: перспективным направлением развития является учет априорного дисбаланса классов при оценке отклонения путем добавления явно выписанной поправки к порогу, формируемому критерием Оцу, без перехода к полностью параметрической постановке.

Библиография

[1] Otsu N. A Threshold Selection Method from Gray-Level Histograms // IEEE Transactions on Systems, Man, and Cybernetics. 1979. Vol. 9, No. 1. P. 62–66. doi: 10.1109/TSMC.1979.4310076

[2] Duda R.O., Hart P.E., Stork D.G. Pattern Classification. 2nd ed. New York: Wiley, 2001. 654 p.

[3] Kurita T., Otsu N., Abdelmalek N. Maximum likelihood thresholding based on population mixture models // Pattern Recognition. 1992. Vol. 25, No. 10. P. 1231–1240. doi: 10.1016/0031-3203(92)90024-D

[4] Kittler J., Illingworth J. Minimum error thresholding // Pattern Recognition. 1986. Vol. 19, No. 1. P. 41–47. doi: 10.1016/0031-3203(86)90030-0

[5] Sahoo P.K., Soltani S., Wong A.K.C. A survey of thresholding techniques // Computer Vision, Graphics, and Image Processing. 1988. Vol. 41, No. 2. P. 233–260. doi: 10.1016/0734-189X(88)90022-9

[6] Sezgin M., Sankur B. Survey over image thresholding techniques and quantitative performance evaluation // Journal of Electronic Imaging. 2004. Vol. 13, No. 1. P. 146–168. doi: 10.1117/1.1631315

[7] Бучаев А.Я., Бегаев А.Н., Комаров И.И. Метод автоматического обнаружения аномалий в пространстве событий информационной безопасности // Промышленные АСУ и контроллеры. 2024. № 2. С. 31–41.

[8] Бучаев А.Я. Метод автоматического формирования информативного пространства для выявления событий информационной безопасности в корпоративных компьютерных сетях // Научно-технический вестник информационных технологий, механики и оптики [Scientific and Technical Journal of Information Technologies, Mechanics and Optics] -2026. - Т. 26. - № 2. - С. 287–294

[9] Xu X., Xu S., Jin L., Song E. Characteristic analysis of Otsu threshold and its applications // Pattern Recognition Letters. 2011. Vol. 32, No. 7. P. 956–961. doi: 10.1016/j.patrec.2011.01.021

[10] Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization // Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018). Funchal, Madeira, Portugal: SciTePress, 2018. P. 108–116. doi: 10.5220/0006639801080116.

[11] Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A Detailed Analysis of the KDD CUP 99 Data Set // 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). Ottawa, ON, Canada: IEEE, 2009. P. 1–6. doi: 10.1109/CISDA.2009.5356528.

[12] Moustafa N., Slay J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set) // 2015 Military Communications and Information Systems Conference (MilCIS). Canberra, ACT, Australia: IEEE, 2015. P. 1–6. doi: 10.1109/MilCIS.2015.7348942.

[13] Cabaj K., Plamowski S., Chaber P., Ławryńczuk M., Marusak P., Nebeluk R., Wojtulewicz A., Zarzycki K. Cyber4OT dataset: Network traces for cyber-security vulnerability evaluation in industrial control systems // SoftwareX. 2025. Vol. 31, article 102196. doi: 10.1016/j.softx.2025.102196

[14] Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection // NDSS. 2018. doi: 10.14722/ndss.2018.23204

[15] NFStream: a Flexible Network Data Analysis Framework [Электронный ресурс]. URL: https://github.com/nfstream/nfstream (дата обращения: 15.03.2026)

[16] Lashkari A.H., Draper-Gil G., Mamun M.S.I., Ghorbani A.A. Characterization of Tor Traffic Using Time Based Features // ICISSP. 2017. P. 253–262. doi: 10.5220/0006105602530262

Equivalence of the Otsu Method Threshold to the Decision Rule of the MAP Classifier in the Task of Information Security Event Detection

A.Ya. Buchaev, I.I. Komarov

***Abstract* — The relationship between the empirical threshold of the classical Otsu method and the decision rule of the Bayesian maximum a posteriori (MAP) classifier is investigated. It is established that strict equivalence of the two thresholds holds under two conditions: equal class variances and equal prior probabilities. When the equal-priors condition is relaxed (with homoscedasticity preserved), the thresholds diverge by an analytically expressed quantity proportional to the logarithm of the ratio of prior probabilities. In the general heteroscedastic case, the MAP classifier defines a quadratic boundary that cannot be represented by a single scalar threshold, and the Otsu method remains its linear approximation. Thus, the thresholds coincide only in the special case; outside of it, an analytical expression for the magnitude of the deviation has been obtained. The obtained result allows the application of the Otsu method to the task of information security event detection to be considered a theoretically grounded procedure rather than a heuristic, and establishes the connection between the empirical criterion and the probabilistic model of the observed features.**

***Keywords* — a posteriori probability, anomaly detection, Bayesian classifier, between-class variance, information security, network traffic, Otsu method, separability coefficient.**

REFERENCES

[1]Otsu N. A Threshold Selection Method from Gray-Level Histograms // IEEE Transactions on Systems, Man, and Cybernetics. 1979. Vol. 9, No. 1. P. 62–66. doi: 10.1109/TSMC.1979.4310076

[2]Duda R.O., Hart P.E., Stork D.G. Pattern Classification. 2nd ed. New York: Wiley, 2001. 654 p.

[3]Kurita T., Otsu N., Abdelmalek N. Maximum likelihood thresholding based on population mixture models // Pattern Recognition. 1992. Vol. 25, No. 10. P. 1231–1240. doi: 10.1016/0031-3203(92)90024-D

[4]Kittler J., Illingworth J. Minimum error thresholding // Pattern Recognition. 1986. Vol. 19, No. 1. P. 41–47. doi: 10.1016/0031-3203(86)90030-0

[5]Sahoo P.K., Soltani S., Wong A.K.C. A survey of thresholding techniques // Computer Vision, Graphics, and Image Processing. 1988. Vol. 41, No. 2. P. 233–260. doi: 10.1016/0734-189X(88)90022-9

[6]Sezgin M., Sankur B. Survey over image thresholding techniques and quantitative performance evaluation // Journal of Electronic Imaging. 2004. Vol. 13, No. 1. P. 146–168. doi: 10.1117/1.1631315

[7]Buchaev A.Ja., Begaev A.N., Komarov I.I. Metod avtomaticheskogo obnaruzhenija anomalij v prostranstve sobytij informacionnoj bezopasnosti // Promyshlennye ASU i kontrollery. 2024. # 2. S. 31–41.

[8]Buchaev A.Ja. Metod avtomaticheskogo formirovanija informativnogo prostranstva dlja vyjavlenija sobytij informacionnoj bezopasnosti v korporativnyh komp'juternyh setjah // Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki [Scientific and Technical Journal of Information Technologies, Mechanics and Optics] -2026. - T. 26. - # 2. - S. 287–294

[9]Xu X., Xu S., Jin L., Song E. Characteristic analysis of Otsu threshold and its applications // Pattern Recognition Letters. 2011. Vol. 32, No. 7. P. 956–961. doi: 10.1016/j.patrec.2011.01.021

[10] Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization // Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018). Funchal, Madeira, Portugal: SciTePress, 2018. P. 108–116. doi: 10.5220/0006639801080116.

[11] Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A Detailed Analysis of the KDD CUP 99 Data Set // 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). Ottawa, ON, Canada: IEEE, 2009. P. 1–6. doi: 10.1109/CISDA.2009.5356528.

[12] Moustafa N., Slay J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set) // 2015 Military Communications and Information Systems Conference (MilCIS). Canberra, ACT, Australia: IEEE, 2015. P. 1–6. doi: 10.1109/MilCIS.2015.7348942.

[13] Cabaj K., Plamowski S., Chaber P., Ławryńczuk M., Marusak P., Nebeluk R., Wojtulewicz A., Zarzycki K. Cyber4OT dataset: Network traces for cyber-security vulnerability evaluation in industrial control systems // SoftwareX. 2025. Vol. 31, article 102196. doi: 10.1016/j.softx.2025.102196

[14] Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection // NDSS. 2018. doi: 10.14722/ndss.2018.23204

[15] NFStream: a Flexible Network Data Analysis Framework [Jelektronnyj resurs]. URL: https://github.com/nfstream/nfstream (data obrashhenija: 15.03.2026)

[16] Lashkari A.H., Draper-Gil G., Mamun M.S.I., Ghorbani A.A. Characterization of Tor Traffic Using Time Based Features // ICISSP. 2017. P. 253–262. doi: 10.5220/0006105602530262

Buchaev Abdulkhamid Yakhyaevich, PhD student, Faculty of Information Technology Security, ITMO University, St. Petersburg, Russia. https://orcid.org/0009-0001-1058-9125 E-mail: abdulhamid0055@yandex.ru

Komarov Igor Ivanovich, Ph.D., (in Maht.), Associate Professor, Faculty of Information Technology Security, ITMO University, St. Petersburg, Russia. E-mail: i_krov@mail.ru