

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРОЦЕССА  
ИДЕНТИФИКАЦИИ ЛИЧНОСТИ В СИСТЕМАХ МОНИТОРИНГА  
ПОСЕЩАЕМОСТИ НА ОСНОВЕ ПОРОГОВОГО АНАЛИЗА  
БИОМЕТРИЧЕСКИХ ДЕСКРИПТОРОВ.

Абдуллаев Дилавар Бахтиярович

студент группы 304-25

направления «Компьютерные сети и их программное обеспечение»

Ташкентского университета

информационных технологий имени Мухаммада аль-Хорезми

**Аннотация.** Данная статья посвящена разработке и обоснованию математической модели процесса идентификации личности в автоматизированных системах мониторинга посещаемости, использующих биометрические дескрипторы (эмбеддинги). В условиях, когда точность и скорость идентификации критически важны для оперативной работы, актуальность формализации процесса принятия решения становится ключевой. Модель базируется на пороговом анализе метрики сходства, вычисляемой между текущим биометрическим дескриптором и эталонными образцами, хранящимися в базе данных. Основной фокус сделан на обосновании выбора метрики сходства (например, косинусное сходство) как наиболее эффективной для высокоразмерных векторных данных, а также на анализе методов определения оптимального порогового значения (Threshold), которое минимизирует вероятность ошибок FAR (False Acceptance Rate) и FRR (False Rejection Rate).

**Ключевые слова:** Идентификация личности (Identity Verification), Математическая модель (Mathematical Model), Биометрические дескрипторы (Biometric Descriptors), Пороговый анализ (Threshold Analysis), Косинусное сходство (Cosine Similarity), ROC-кривая.

## ВСТУПЛЕНИЕ

Системы мониторинга посещаемости, основанные на биометрии (например, распознавании лиц), требуют высокоточного и быстрого механизма идентификации (1:N сравнение) или верификации (1:1 сравнение). В основе этих механизмов лежит сравнение биометрического дескриптора (вектора признаков, полученного глубокой нейронной сетью) с эталонными дескрипторами в базе данных.

Процесс принятия решения об идентификации сводится к пороговому анализу: если мера сходства между текущим и эталонным дескриптором превышает некоторый заранее заданный порог ( $T$ ), личность считается идентифицированной. Неправильный выбор метрики сходства или неверно установленный порог ведут к неприемлемо высокому уровню ошибок: ошибки ложного допуска (FAR), когда система принимает "чужого" за "своего", и ошибки ложного отказа (FRR), когда система отказывает в доступе зарегистрированному пользователю.

Цель данной статьи — построить математическую модель процесса идентификации, формализовать выбор метрики сходства для биометрических дескрипторов и предложить метод оптимизации порогового значения для достижения максимальной эффективности системы учета посещаемости.

## МОДЕЛИРОВАНИЕ ПРОЦЕССА ИДЕНТИФИКАЦИИ

Модель биометрического дескриптора. Пусть  $x \in R^D$  — это биометрический дескриптор (эмбеддинг), полученный из изображения лица с помощью глубокой нейронной сети (например, ArcFace), где  $D$  — размерность вектора (например, 512). Пусть  $DB = \{y_1, y_2, \dots, y_N\}$  — база данных эталонных дескрипторов  $N$  зарегистрированных пользователей.

Выбор метрики сходства. Для сравнения высокоразмерных дескрипторов  $x$  и  $y_i$  необходима метрика сходства и  $S(x, y_i)$  которая количественно оценивает их близость.

Обоснование выбора косинусного сходства: В современных алгоритмах генерации эмбеддингов (например, ArcFace, SphereFace) обучение нейронной сети нацелено на то, чтобы дескрипторы одного человека формировали плотный кластер на гиперсфере. Таким образом, угловое расстояние между векторами становится наиболее информативным показателем их сходства, а не Евклидово расстояние.

Косинусное сходство  $S_C(x, y_i)$  определяется как косинус угла между двумя векторами:

$$S_C(x, y_i) = \frac{x \cdot y_i}{\|x\| \cdot \|y_i\|}$$

- Преимущества: Если векторы  $x$  и  $y_i$  были нормализованы до единичной длины ( $\|x\| = \|y_i\| = 1$ ), то  $S_C$  совпадает со скалярным произведением. Диапазон значений  $S_C$  находится в пределах  $[-1, 1]$ . Чем ближе значение к 1, тем выше сходство

Модель процесса идентификации и пороговый анализ. Процесс идентификации (1:N) в системе посещаемости формулируется как поиск эталонного дескриптора  $y^*$  в  $DB$ , который максимизирует сходство с входным дескриптором  $x$ :

$$y^* = \arg_{y_i \in DB} \max(S_C(x, y_i))$$

Решение об идентификации принимается на основе сравнения максимального сходства  $S_{max} = S_C(x, y^*)$  с порогом  $T$ :

$$\text{Идентификация} = \begin{cases} y^*, \text{ если } S_{max} \geq T \\ \text{Отказано, если } S_{max} < T \end{cases}$$

Определение оптимального порогового значения. Выбор порога  $T$  напрямую влияет на коэффициенты ошибок:

- FAR (False Acceptance Rate): Вероятность того, что чужой будет принят ( $S_{max} \geq T$  для несовпадающего лица).

- FRR (False Rejection Rate): Вероятность того, что свой будет отвергнут ( $S_{max} < T$  для совпадающего лица).

Метод определения оптимального порога. Оптимальный порог  $T_{opt}$  определяется на основе анализа распределения оценок сходства "своих" (Intra-class scores) и оценок сходства "чужих" (Inter-class scores) на тестовой выборке.

- Построение ROC-кривой (Receiver Operating Characteristic): На основе тестового набора данных строится график зависимости True Positive Rate (TPR) от False Positive Rate (FPR) при вариации порога  $T$ .

• Equal Error Rate (EER): Порог  $T_{ERR}$ , при котором  $FAR = FRR$ . Это наиболее распространенная метрика, обеспечивающая баланс между двумя типами ошибок.

• Zero-FAR: В системах с высокими требованиями к безопасности (где ложный допуск критичен) выбирается максимальный порог  $T_{Z FAR}$ , при котором  $FAR = 0$  (ценой увеличения FRR).

• Точка операционной потребности (Operating Point): В системах посещаемости часто требуется низкий FAR (для предотвращения "отметок" за других), но и приемлемый FRR. Оптимальный порог  $T_{opt}$  выбирается как точка на ROC-кривой, соответствующая заданному максимально допустимому значению FAR (например,  $FAR \leq 0.01\%$ ).

Математическое обоснование:  $T_{opt}$  — это значение, которое минимизирует функцию стоимости, взвешивающую последствия FAR и FRR. Для систем посещаемости, где FAR ведет к финансовым или дисциплинарным нарушениям, часто выбирают  $T_{opt} \approx T_{ERR}$  или точку, где FAR немного ниже FRR.

## ВЫВОД

Математическая модель процесса идентификации личности, основанная на пороговом анализе, является фундаментальной для систем мониторинга посещаемости. Выбор косинусного сходства как метрики является наиболее обоснованным для биометрических дескрипторов,

поскольку он отражает угловую кластеризацию, достижимую современными глубокими сетями.

Оптимизация порога  $T$  — это не только математическая, но и прикладная задача. Рекомендуемый подход заключается в использовании анализа ROC-кривой для определения точки равных ошибок (EER) или операционной точки, соответствующей строгому требованию к максимально допустимому FAR. Эта модель обеспечивает формальный, воспроизводимый и измеримый подход к настройке системы, гарантируя ее высокую эффективность и надежность в реальных условиях эксплуатации.

## **ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА**

1. Наумушкин В. А. и др. КОМПОЗИТНАЯ МЕТРИКА ЭМОЦИОНАЛЬНОЙ КОГЕРЕНТНОСТИ В СИСТЕМАХ ИЗВЛЕЧЕНИЯ КОНТЕКСТА.
2. Бикбавлеев А. А. и др. Автоматизированный прогноз краткосрочного движения стоимости ценных бумаг на основе нейронных сетей: выпускная бакалаврская работа по направлению подготовки: 01.03. 02- Прикладная математика и информатика. – 2022.
3. Девицына С. Н. и др. Разработка интеллектуальной системы биометрической идентификации пользователя //Экономика. Информатика. – 2019. – Т. 46. – №. 1. – С. 148-160.
4. Луценко Е. В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос», системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине //Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2016. – №. 116. – С. 1-60.
5. Носков С. И., Базилевский М. П. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования. – 2018.