Система распознавания пользователей по извлекаемым признакам голоса с применением фильтра Калмана

Н.Д. Лушников

Уфимский университет науки и технологий, Уфа

Аннотация: Рассматривается задача распознавания личности по голосу с применением адаптивного фильтра Калмана. В качестве признаков биометрической аутентификации личности использованы извлеченные признаки акустического сигнала. Приведена сравнительная таблица ошибок разделения дикторов и оценки системы разделения дикторов с применением фильтра Калмана.
Ключевые слова: биометрическая аутентификация, голос, компиляция нейронной сети, адаптивный фильтр Калмана.

Введение

Индивидуальным идентификатором для каждого пользователя являются извлекаемые биометрические признаки. К числу основных биометрических признаков следует отнести характеристики голоса. Исследования в данном сегменте направлены на получение требуемых показателей эффективности, которые определяют качество и производительность выполняемых операций. Для достижения данных показателей следует применить те методы и алгоритмы, которые обработают входные данные и позволяют получить результат без нежелательных шумовых эффектов. Одним из таких методов является адаптивный фильтр Калмана, который предназначен для фильтрации помех и шумоподавления. Фильтр Калмана позволяет оценивать состояние системы, удовлетворяющей «зашумленному» линейному стохастическому дифференциальному уравнению, на основе использованного ряда зашумленных наблюдений [1]. Ранее фильтрация Калмана была представлена в трудах Липцера Р.Ш. и Ширяева А.Н., являющаяся неотъемлемой частью теории оптимальной нелинейной фильтрации [2], Браммера К. и Зиффлинга Г., которые исследовали задачу стохастической и линейной фильтрации в интерпретации случайных векторных процессов [3], Агафонова В.Ю. при детектировании и

для экстраполяции позиции объекта [4], Тележкина В.Ф. для обнаружения и фильтрации шума в системах обработки показаний датчиков [5].

Целью рассматриваемого исследования является усовершенствование процессов распознавания пользователей по набору извлекаемых признаков голоса посредством модифицированной фильтрации Калмана. Результатом исследования является разработанная система распознавания пользователей по извлекаемым признакам голоса с применением адаптивного фильтра Калмана.

Распознавание пользователей по извлекаемым признакам голоса

Выявление признаков начинается с разбиения входного акустического сигнала на временные окна небольшой длины с фиксированным шагом смещения. Для каждого полученного кадра применяются следующие преобразования:

Предварительная фильтрация с конечной импульсной характеристикой (КИХ-фильтр):

$$y_t = x_t - b \times x_{t-1}$$

где y_t – акустический сигнал после фильтрации, x_t – входной акустический сигнал, t – количество кадров, b – коэффициент фильтрации.

Дискретное преобразование Фурье (ДПФ) [6]:

$$F_k = \sum_{t=0}^{T-1} w_t \times y_t e^{\frac{-2\pi i}{T} kt}$$

где T – отсчеты в кадре, w_t – весовая оконная функция, k – индекс частоты.

Весовая оконная функция (окно Хэмминга и окно Ханна) применяется с целью уменьшения краевых эффектов, возникающих в результате разбиения сигнала на кадры:

$$w_t^{hamm} = 0.54 - 0.46cos\left(\frac{2\pi t}{T-1}\right)$$

$$w_t^{hann} = 0.5\left(1 - cos\left(\frac{2\pi t}{T-1}\right)\right)$$

Дискретное косинусное преобразование для значений энергий фильтров E_s [6]:

$$C_l = \sum_{s=0}^{M-1} E_s cos\left(l\left(s+\frac{1}{2}\right)\frac{\pi}{M}\right)$$

где C_l – коэффициент под номером l, M – количество фильтров.

Алгоритм распознавания пользователей по извлекаемым признакам голоса представлен на рис. 1.

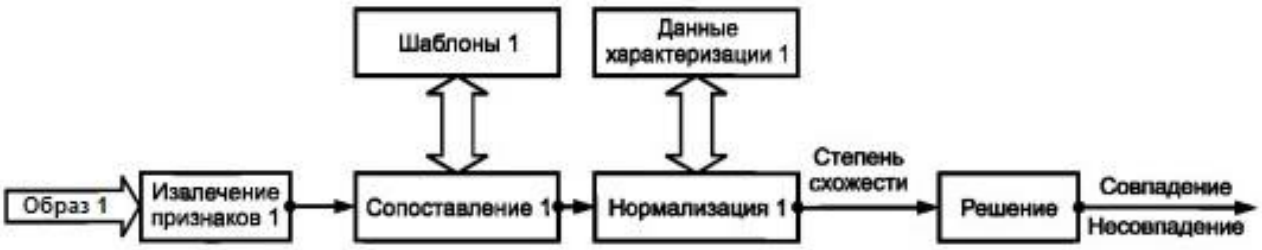

Рис. 1. – Схема принятия решения при обработке голоса

В качестве итоговых значений берутся первые несколько коэффициентов дискретного косинусного преобразования.

При создании нейронных сетей на начальном этапе были сформированы датасеты аудиозаписей на основе категориальной кросс-энтропии. Для формирования датасета были созданы папки с тренировочной (train), валидационной (val) и тестовой (test) обучающими выборками [7].

Объем обучающей выборки с акустическими признаками составляет 450 аудиозаписей от двух пользователей. Длительность каждой аудиозаписи составляет 8, 15 и 25 секунд, соответственно. Аудиозапись длительностью 8 секунд представлена в виде файла с записанным голосом, в котором

производится счет от одного до пяти. В аудиозаписи длительностью 15 секунд производится счет от одного до десяти, а в аудиозаписи длительностью 25 секунд – от одного до двадцати [7].

В программном коде скомпилированы модели на основе функций оптимизации Adam с количеством указанных эпох обучения. В данном исследовании выбранное количество эпох является оптимальным объемом для достижения необходимого уровня качества проводимого обучения нейронных сетей. Сгенерированы папки обучающих выборок с помощью функции datagen.

При проведении данного исследования также была скомпилирована и протестирована модель архитектуры нейронной сети Wav2vec с использованием таких датасетов аудиозаписей, как TIMIT и VoxCeleb [7].

Данное программное обеспечение разработано на основе принципа Zero Trust («Нулевое доверие»), концепция которого заключается в отсутствии доверенных или проверенных пользователей [8]. Разработана система хэширования биометрических персональных данных, которая направлена на предотвращение несанкционированного доступа и фальсификации биометрических данных.

Применение адаптивного фильтра Калмана

При извлечении акустических признаков в процессе распознавания личности по голосу в режиме онлайн для шумоподавления применялся адаптивный фильтр Калмана с увеличением веса центрального значения многомерных массивов аудиозаписей:

$$M_2^{low} = 0{,}1\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Непрерывный алгоритм фильтрации Калмана во времени при распознавании личности по голосу основывается на системе дифференциальных уравнений:

$$\frac{dx_0}{dt} = \Phi(t)x_0(t) + B(t)U(t) + D(t)F(t) + \sum_{i=1}^{N} K_i(t)(z_i(t) - H(t)x_0(t))$$

$$\frac{dP(t)}{dt} = V_w(t) + \Phi(t)P(t) + P(t)\Phi^T(t) - P(t)H^T(t)V_v^{-1} \times (t)H(t)P(t)$$

где $z_i(t)$ – вектор наблюдений, $z_i(t) = H(t)x_0(t)$ – вектор оценок наблюдений, $x_0(t)$ – оценка вектора состояния, $\Phi(t)$ – переходная матрица, $P(t)$ – корреляционная матрица, $H(t)$ – матрица наблюдения, $K_i(t)$ – матрица коэффициентов, $U(t)$ – вектор управления, $F(t)$ – вектор измеренных сигналов с выхода объекта, $B(t)$ – матрица коэффициентов управления, $D(t)$ – матрица коэффициентов измерения, $K_i(t) = S_i(t)P(t)H^T(t)V_v^{-1}(t)$ – матрица коэффициентов.

При синтезе фильтра Калмана формируется вектор коэффициентов оптимального фильтра [9]:

$$K(k) = P(k-1)C^T(k) \times (C(k)P(k-1)C^T(k) + Q_M(k))^{(-1)}$$

где $K(k)$ – коэффициент оптимального фильтра Калмана, $Q_M(k)$ – ковариационная матрица некоторой случайной величины, $P(k-1)$ – дисперсионная матрица вектора состояния.

Фильтр Калмана позволяет минимизировать дисперсию оценки векторного случайного процесса, изменяющегося во времени следующим образом [10]:

$$x(k+1) = \Phi(k)x(k) + v(k)$$

где $x(k)$ – векторный случайный процесс, $\Phi(k)$ – матрица перехода, $v(k)$ – случайный вектор (шум).

Результаты применения адаптивного фильтра Калмана в процессе распознавания личности по изображению лица представлены на рис. 2.

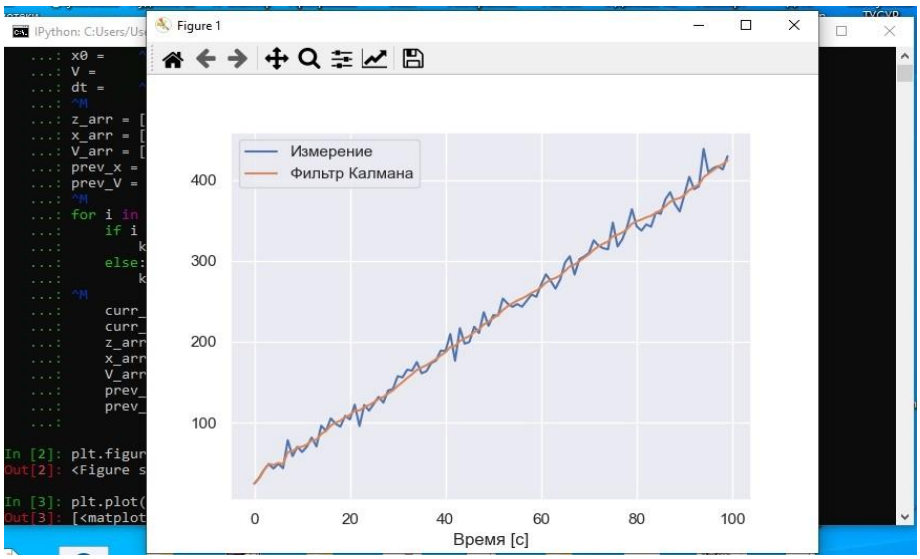

Рис. 2. – Значения адаптивного фильтра Калмана при обработке аудиозаписей

Фильтр Калмана предназначен для оптимизации и нормализации звукового сигнала. Полученные значения сигнала являются обработанными, которые представляют собой обучающую выборку архитектур нейронных сетей.

Показатели системы распознавания пользователей по извлекаемым признакам голоса

Для оценивания системы распознавания речи пользователей необходимо учесть показатели ошибок разделения дикторов, средней чистоты кластеров (содержание в аудиозаписи только речи диктора) и средней чистоты дикторов (содержание в аудиозаписи речи пользователей по отдельности), которые указаны в таблице № 1 [7]. Также приведена сравнительная таблица № 2 данных показателей с применением модифицированного фильтра Калмана. Данные в таблицах были подсчитаны

в соответствии с формулой, по которой вычисляется ошибка разделения дикторов:

$$E_{spkr} = \frac{\sum_{seg}(T(seg)\times \min(N_{ref}(seg), N_{sys}(seg)) - N_{correct}(seg))}{\sum_{seg} T(seg)\times N_{ref}(seg)},$$

где E_{spkr} – ошибка разделения дикторов, $T(seg)$ – длительность речевого сегмента seg, $N_{ref}(seg)$ – количество дикторов (эталонная разметка), $N_{sys}(seg)$ – количество дикторов (оцениваемая система), $N_{correct}(seg)$ – количество верно отнесенных дикторов.

Также в таблице приведены значения средней чистоты кластеров:

$$ACP_c = \frac{\sum_{s=1}^{S} n_{sc}^2}{(N_c^{cluster})^2},$$

$$ACP = \frac{1}{N}\sum_{c=1}^{M} ACP_c \times N_c^{cluster},$$

И значения средней чистоты дикторов:

$$ASP_s = \frac{\sum_{c=1}^{M} n_{sc}^2}{(N_c^{cluster})^2},$$

$$ASP = \frac{1}{N}\sum_{s=1}^{S} ASP_s \times N_s^{speaker},$$

где ACP – средняя чистота кластеров, ASP – средняя чистота дикторов, S – количество дикторов (эталонная разметка), M – полученное количество кластеров, n_{sc} – количество данных в кластере c, которые принадлежат диктору s, $N_c^{cluster} = \sum_{s=1}^{S} n_{sc}$ – количество данных в кластере c, $N_s^{speaker} = \sum_{c=1}^{M} n_{sc}$ – количество данных, принадлежащих диктору s, $N = \sum_{s=1}^{S}\sum_{c=1}^{M} n_{sc}$ – количество всех данных.

В таблице в качестве итоговой совокупной оценки системы разделения пользователей используется среднее геометрическое значений ASP и ACP, обозначается как K:

$$K = \sqrt{ACP \times ASP}$$

Таблица № 1

Показатели ошибок разделения дикторов и оценка системы разделения дикторов

Набор обучающей выборки	MFCC		LPC		PLP		CQCC		SCF	
	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K
DataSet 100	8,26	0,862	7,60	0,873	9,42	0,845	7,52	0,793	8,34	0,853
DataSet 300	8,01	0,821	7,99	0,832	8,87	0,839	7,49	0,769	8,15	0,821
DataSet 450	7,67	0,794	7,81	0,801	8,34	0,821	7,21	0,742	7,99	0,800

Таблица № 2

Таблица ошибок разделения дикторов и оценки системы разделения дикторов с применением фильтра Калмана

Набор обучающей выборки	MFCC		LPC		PLP		CQCC		SCF	
	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K	E_{spkr} (%)	K
DataSet 100	8,01	0,786	7,51	0,851	9,01	0,819	7,34	0,781	8,33	0,832
DataSet 300	7,99	0,793	7,79	0,822	8,65	0,825	7,33	0,755	8,12	0,823
DataSet 450	7,42	0,722	7,43	0,800	8,19	0,802	7,11	0,741	7,99	0,797

Мел-частотные кепстральные коэффициенты (MFCC) представляют собой набор признаков, которые описывают общую форму спектральной огибающей. Данный набор признаков моделирует характеристики человеческого голоса на основе частотного распределения (по размеру окна).

Кодирование речи коэффициентами линейного предсказания (LPC) опирается на теорию статистического анализа временных рядов.

Перцепционные коэффициенты линейного предсказания (PLP) составлены на основе асимптотической оценки рекурсивных соотношений.

Постоянное Q-преобразование (CQT) констант Q-кепстральных коэффициентов (CQCC) характеризует разрешение в области низких частот и временное разрешение в области высоких частот.

Частота спектрального центроида (SCF) представляет собой средневзвешенную частоту для поддиапазона, где весовые коэффициенты представляют собой нормированную энергию каждого частотного компонента в этом поддиапазоне.

Согласно проведенному анализу таблиц № 1 и № 2, применение фильтра Калмана позволяет уменьшить значения показателей ошибок разделения дикторов в информационной системе.

Результаты

При прохождении обучения искусственной нейронной сети с применением адаптивного фильтра Калмана при обработке входного аудиосигнала точность работы (accuracy) составила 93,0%. Точность обучения валидационного набора (val accuracy) составляет 93,9%. Соответственно показатели потерь (loss) равны 7,0%. Потери валидационного набора (val loss) равны 6,1%.

Таким образом, с применением фильтра Калмана уровень точности обучения имеет положительную динамику и расположенность к достижению максимальных показателей при увеличении входных параметров (от 92 к 93 %).

Литература

1. Мисюра В.В., Мисюра И.В. Обработка и фильтрация сигналов. Современное состояние проблемы // Инженерный вестник Дона, 2013, №4. URL: ivdon.ru/magazine/archive/n4y2013/2130.

2. Липцер Р.Ш., Ширяев А.Н. Статистика случайных процессов. М.: Наука, 1974. 696 с.

3. Браммер К., Зиффлинг Г. Фильтр Калмана-Бьюси. Детерминированное наблюдение и стохастическая фильтрация. М.: Наука, 1982. 199 с.

4. Агафонов В.Ю., Розалиев В.Л., Заболеева-Зотова А.В. Использование фильтра Калмана в задачах трекинга объектов // Интеллектуальные системы. Теория и приложения, 2016, №4, С. 11-15.

5. Тележкин В.Ф., Саидов Б.Б. Обработка информации с использованием фильтра Калмана в Matlab Simulink // Системы анализа и обработки данных, 2021, №4, С. 49-63.

6. Щербань И.В., Доброходский В.В., Ефименко А.А. Online–программа аутентификации, основанная на оконном преобразовании Фурье речевых фраз пользователя // Символ науки, 2016, №1. URL: cyberleninka.ru/article/n/online–programma–autentifikatsii–osnovannaya–na–okonnom–preobrazovanii–furie–rechevyh–fraz–polzovatelya.

7. Исмагилова А.С., Лушников Н.Д. Комплексная биометрическая аутентификация пользователей информационной системы с применением нейронных сетей // Инженерный вестник Дона, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8961.

8. Валеев С.С., Кондратьева Н.В., Гузаиров М.Б., Исмагилова А.С. Иерархическая динамическая система управления информационной безопасностью информационной системы предприятия // Инженерный вестник Дона, 2023, №11. URL: ivdon.ru/ru/magazine/archive/n11y2023/8802.

9. C.Chui and Chen Guanrong. Kalman Filtering with Real-Time Applications. Springer Series in Information Sciences, 2017. 245 p.

10. Девицына С.Н., Елецкая Т.А., Балабанова Т.Н., Гахова Н.Н. Разработка интеллектуальной системы биометрической идентификации пользователя // Научные ведомости. Серия: Экономика. Информатика, 2019, № 1, С.148-160.

References

1. Misyura V.V., Misyura I.V. Inzhenernyj vestnik Dona, 2013, №4. URL: ivdon.ru/magazine/archive/n4y2013/2130.

2. Liptser R.Sh., Shiryayev A.N. Statistika sluchaynykh protsessov [Statistics of random processes]. M.: Nauka, 1974. 696 p.

3. Brammer K., Ziffling G. Fil′tr Kalmana-B′yusi. Determinirovannoye nablyudeniye i stokhasticheskaya fil′tratsiya [Deterministic observation and stochastic filtering]. M.: Nauka, 1982. 199 p.

4. Agafonov V.Yu., Rozaliyev V.L., Zaboleyeva-Zotova A.V. Intellektual′nyye sistemy. Teoriya i prilozheniya, 2016, №4, pp. 11-15.

5. Telezhkin V.F., Saidov B.B. Sistemy analiza i obrabotki dannykh, 2021, №4, pp. 49-63.

6. Shcherban′ I.V., Dobrokhodskiy V.V., Yefimenko A.A. Simvol nauki, 2016, №1 URL: cyberleninka.ru/article/n/online–programma–autentifikatsii–osnovannaya–na–okonnom–preobrazovanii–furie–rechevyh–fraz–polzovatelya.

7. Ismagilova A.S., Lushnikov N.D. Inzhenernyj vestnik Dona, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8961.

8. Valeyev S.S., Kondrat′yeva N.V., Guzairov M.B., Ismagilova A.S. Inzhenernyj vestnik Dona, 2023, №11 URL: ivdon.ru/ru/magazine/archive/n11y2023/8802.

9. C.Chui and Chen Guanrong. Kalman Filtering with Real-Time Applications. Springer Series in Information Sciences, 2017. 245 p.

10. Devitsyna S.N., Yeletskaya T.A., Balabanova T.N., Gakhova N.N. Nauchnyye vedomosti. Seriya: Ekonomika. Informatika, 2019, № 1, pp. 148-160.

Дата поступления: 31.12.2023

Дата публикации: 9.02.2024