

Математический аппарат для оценивания некоторых показателей качества учебного вебинара посредством машинного обучения

В. В. Подколзин¹, А. Н. Полетайкин^{1,2}, М. Ю. Галкина²

¹ Кубанский государственный университет (КубГУ)

² Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: Данная статья является вторым этапом создания и исследования технологии оценивания качества контактной работы в вузе, реализуемой посредством вебинаров. Рассматривается математическое обеспечение интеллектуальной технологии анализа аудиовидеопотока записи учебного вебинара. Решается задача повышения эффективности и достоверности оценивания некоторых показателей качества контактной работы в институте заочного образования СибГУТИ, реализуемой посредством вебинаров, в ходе дистанционного обучения. Для анализа применяются методы машинного обучения и их реализация с использованием стандартных модулей и библиотек Python. Научная значимость работы состоит в разработке новых моделей и алгоритмов интеллектуального анализа аудиовидеопотока, параметризованных для обработки записи учебного мероприятия. Разработанные модели и алгоритмы позволяют повысить эффективность процесса оценивания качества контактной работы, реализуемой посредством вебинаров.

Ключевые слова: вебинар, область конференции, активность ведущего, критерии оценивания эффективности, интервалы времени, локализация текста, обработка аудиотрека, обработка изображений, машинное обучение.

Для цитирования: Подколзин В. В., Полетайкин А. Н., Галкина М. Ю. Математический аппарат для оценивания некоторых показателей качества учебного вебинара посредством машинного обучения // Вестник СибГУТИ. 2024. Т. 18, № 1. С. 121–143. <https://doi.org/10.55648/1998-6920-2024-18-1-121-143>.



Контент доступен под лицензией
Creative Commons Attribution 4.0
License

© Подколзин В. В., Полетайкин А. Н.,
Галкина М. Ю., 2024

Статья поступила в редакцию 15.12.2023;
принята к публикации 15.01.2024.

1. Введение

В Сибирском государственном университете телекоммуникаций и информатики (СибГУТИ) в рамках программы цифрового развития университета поставлена задача оценивания качества контактной работы, реализуемой посредством вебинаров, в ходе дистанционного обучения. Разработанная ранее математическая модель, базирующаяся на системе из 32 показателей качества вебинара и опубликованная в статье [1], работает преимущественно на методах экспертного оценивания и поэтому обладает высокой субъективностью и требует значительных затрат времени. Необходимо более простое и эффективное решение, характеризующееся приемлемой степенью объективности оценок.

Вебинар (от англ. webinar – web based seminar, онлайн-семинар, онлайн-конференция) – онлайн-трансляция с целью обучения или обеспечения корпоративной встречи [2]. В данной статье рассматриваются лишь обучающие вебинары. В вебинаре всегда участвуют две

стороны: докладчик-ведущий и слушатели-участники. Как правило, участники могут видеть ведущего, оставаясь при этом за кадром. Предполагается, что в течение вебинара все участники слушают ведущего и реагируют на его активности вовлечения слушателей в предмет обсуждения в основном в чате, реже – посредством выхода в эфир. Вебинар может проходить как в прямой трансляции, так и быть доступным в записи. К техническим возможностям вебинара можно отнести следующие:

- демонстрация презентаций;
- возможность быстро получать и сразу видеть точные результаты голосований и опросов среди участников вебинара;
- передача участникам нужных файлов и ссылок;
- обратная связь – в виде чата, куда можно написать свой вопрос или просто поделиться впечатлениями (если вебинар в прямой трансляции).

Для того чтобы оценить эффективность вебинаров, используются разнообразные критерии оценки. Чаще всего к ним относят такие общие критерии:

- количество зарегистрированных пользователей;
- распространение информации о предстоящем вебинаре;
- количество участников вебинара;
- активность участников.

Количество зарегистрированных пользователей показывает, насколько выбранная тема востребована у целевой аудитории. Распространение информации о предстоящем вебинаре, например, в социальных сетях, говорит о том, что есть пользователи, которые проявляют интерес к вебинару и готовы рекомендовать его другим людям. Количество участников на вебинаре – индикатор интереса слушателей к теме вебинара. Чем активнее участники вебинара будут взаимодействовать с ведущим, тем больше вероятность того, что они, например, оставят заявку, подпишутся на рассылку, посмотрят запись вебинара [3].

В 2018–2019 годах в СибГУТИ было проведено комплексное исследование [1], в результате которого разработана математическая модель, включающая систему из 32 показателей качества организации и проведения вебинара, позволяющих оценить с разных сторон качество вебинара с помощью 10-балльных экспертных оценок. Проблема заключается в существенных трудозатратах, которые несут эксперты при оценивании этих показателей. Кроме того, мнения экспертов субъективны. Целью дальнейших исследований в статье [1] заявлено сокращение трудоемкости работы экспертов и повышение объективности получаемых оценок качества вебинаров. Это достигается за счет сведения к минимуму влияния человеческого фактора в процедуре оценивания и применения для оценивания некоторых показателей качества вебинаров количественных математических методов, относящихся к категории Data Mining. Поэтому актуальной является задача разработки математических средств, которые позволят оценить большую часть показателей без участия человека в автоматическом режиме.

2. Постановка задачи

Предварительный анализ показателей из статьи [1], полный перечень которых представлен в приложении к данной статье, позволил выявить 17 таких (см. табл. 1), которые можно оценить без участия человека посредством компьютерного анализа аудиовидеозаписи вебинара. На рис. 1 показан стоп-кадр видеопотока аудиовидеозаписи вебинара на платформе Mirapolis (<https://www.mirapolis.ru>), где выделены характерные области, анализ которых позволяет определить величины показателей. Соответствие между показателями и областями кадра показано в табл. 1.

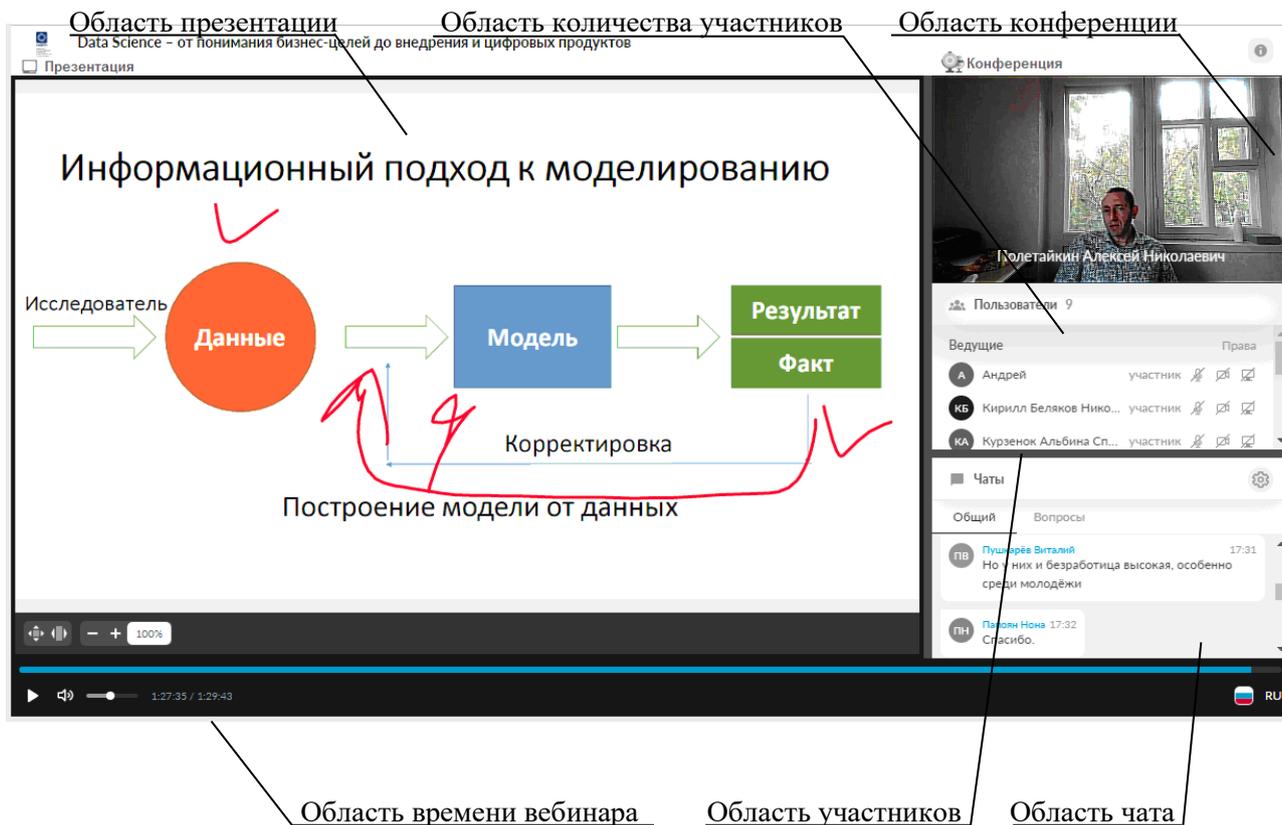


Рис. 1. Стоп-кадр аудиовидеопотока аудиовидеозаписи вебинара и его характерные области

Таблица 1. Некоторые показатели качества вебинара, которые могут быть рассчитаны на основе компьютерного анализа аудиовидеозаписи вебинара (нумерация показателей соответствует таблице в приложении, сложность оценена по шкале из табл. 2)

№	Название показателя	Сложность	Источник информации о показателе
7	Лаконичность	2	Область презентации
9	Контактные данные	1	Область презентации
10	Энергичность и эмоциональность	2	Область конференции / звуковой ряд
11	Темп изложения	2	Звуковой ряд
13	Внешний вид	1	Область конференции / звуковой ряд
14	Количество и качество информации на слайдах	1	Область презентации
15	Количество и качество иллюстраций	1	Область презентации
16	Аккуратность	1	Область презентации
17	Использование указки	3	Область презентации
18	Разнообразие презентационных инструментов	3	Область презентации / область чата
20	Проведение опросов	3	Область презентации
23	Частота взаимодействия	3	Область чата
24	Наличие вопросов слушателей	1	Область чата
25	Реактивность на вопросы	4	Звуковой ряд
26	Качество ответов на вопросы	4	Звуковой ряд
31	Точность заявленной длительности мероприятия	1	Область времени вебинара
32	Показатели посещаемости	2	Область количества участников

Таблица 2. Шкала ранжирования показателей качества вебинара по временной сложности процедуры экспертного оценивания [1]

Балл	Описание трудозатрат	Затраты времени, мин.
0	Не требует внимания экспертов	0
1	Минимальные затраты (чат, материалы, ведущий)	2–5
2	Беглый эпизодический просмотр вебинара	5–10
3	Внимательный просмотр нескольких эпизодов вебинара	10–30
4	Беглый просмотр вебинара целиком	20–40

В данной статье будут рассмотрены показатели 7, 9 – 11, 14 – 17, 31 и 32 и выработаны научно обоснованные математические решения для их эффективного оценивания. Для этого необходим в основном компьютерный анализ изображений слайдов из области презентации. Для оценивания показателей 10 и 13 необходим анализ изображений ведущего из области конференции. Оценивание показателей 31 и 32 требует распознавания чисел из областей времени вебинара и области количества участников соответственно.

Анализ области презентации сводится к анализу изображений каждого слайда презентации. Факт того, что демонстрируется именно презентация, а не видеоролик или рабочий стол ведущего, удостоверяется наличием заголовка «Презентация», а также инструментов масштабирования (— +) в нижней части области.

Показатель 7 *Лаконичность* предполагает в наиболее простом варианте идентификацию рисунков, схем, диаграмм и других изображений, а также тестовых блоков по объёму, в т.ч. в таблицах (как правило, чем меньше объём блока текста, тем больше его лаконичность). Результатом анализа должно быть число и классификация текстовых блоков по категориям лаконичности, а также общее число изображений. На основе этих данных будет рассчитано значение показателя 7, а также показатель 15 *Количество и качество иллюстраций*.

Для оценивания показателя 14 *Количество и качество информации на слайдах* необходимо для каждого слайда презентации рассчитать отношение к общей площади пространства слайда площади, занимаемой на слайде: 1) изображениями и 2) текстовыми блоками.

Показатель 16 *Аккуратность* предполагает идентификацию текстовых блоков по размеру шрифта и цвету, списков по размеру шрифта и маркерам, изображений по размеру и положению на слайде (по координатам левого верхнего угла).

Расчёт показателя 17 *Использование указки*: необходимо идентифицировать указку и сформировать массив, элементы которого соответствуют сеансам использования указки (под сеансами понимаются интервалы между паузами в использовании указки длительностью более 10 секунд) и содержат суммарную длину траектории ее перемещения (криволинейную траекторию можно рассматривать как ломаную).

Показатель 9 *Контактные данные* предполагает идентификацию на слайдах телефонных номеров и электронных адресов по их характерным признакам. Здесь, однако, существует проблема идентификации посторонних контактных данных, например, в изображениях оконного интерфейса с контактными данными. Поэтому необходимо предусмотреть этот случай и такие данные игнорировать. Результатом анализа должен быть факт наличия в презентации контактных данных ведущего.

Распознавание числовых данных требуется для анализа таких показателей, как 31 *Точность заявленной длительности мероприятия* и 32 *Показатели посещаемости*. В частности, необходимо идентифицировать в области времени вебинара момент времени:

- 1) первого включения изображения ведущего в области конференции,
- 2) последнего отключения изображения ведущего в области конференции,

а также идентифицировать в области числа участников максимальное число участников между моментами времени 1 и 2 и число участников в момент времени 2.

Показатель 10 *Энергичность и эмоциональность* требует анализа динамичности изображения ведущего в области конференции – по крайней мере, мимики лица ведущего и его

положения в области конференции.

Таким образом, для оценивания указанных показателей необходимо осуществлять распознавание изображений в видеопотоке вебинара. Для решения такого рода задач используются искусственные нейронные сети и программное обеспечение на их основе.

Показатель *11 Темп изложения* требует анализа аудиотрека вебинара на предмет вычисления скважности его звукового ряда, для этого применяются инструменты работы с аудиоданными на основе дискретного преобразования Фурье.

Целью исследования является повышение оперативности и достоверности оценивания некоторых показателей качества вебинара. Объектом исследования является категория качества учебного вебинара, представленного аудиовидеопотоком его записи. Предмет исследования – нейросетевые математические модели и специальное программное обеспечение на их основе при решении задачи оценивания качества вебинара.

Решением задачи должна стать программная среда, разработанная на языке Python и реализующая разработанный математический аппарат в отношении заданной аудиовидеозаписи вебинара.

3. Выбор инструментария для решения задачи

Для программной реализации выбран язык программирования Python, к главным чертам которого относятся динамическая типизация, понятный синтаксис, поддержка объектно-ориентированного программирования, а также наличие обширной библиотечной базы (в том числе для задач машинного обучения) и возможность легкой установки библиотечных модулей, что значительно расширяет функционал программы и упрощает её создание. Важным отличием Python является наличие множества свободно распространяемых библиотек. В соответствии с постановкой задачи к основным выбранным модулям и библиотекам можно отнести:

- NumPy – для работы с многомерными массивами и использования некоторых функций линейной алгебры, например, для расчёта траектории движения указки или карандаша ведущего при оценивании показателя 17;
- LibROSA – для построения и анализа спектрограмм аудиопотока вебинара, например, для оценивания энергичности и эмоциональной насыщенности вебинара по показателю 10, а также темпа изложения (показатель 11);
- SciPy и Scikit-learn – для решения задач класса Data Mining, таких как кластеризация, распознавание изображений и работа с ними;
- PyQt5 – для работы с компонентами графического интерфейса пользователя, виджетами (библиотека написана на языке C++, поэтому требует установки средств разработки Visual C++);
- Cv2 – для реализации методов машинного обучения OpenCV.

4. Алгоритм работы с аудиоданными

Обработка аудиотрека вебинара используется для определения трёх видов интервалов времени:

- интервалы времени тишины;
- интервалы времени разговора ведущего;
- остальные интервалы времени.

Для извлечения аудиотрека из записи вебинара применяется модуль `moviepy.editor` и функция `VideoFileClip(filename = pathVideo + nameVideo)`, принимающая параметр `filename`, которая определяет путь к файлу аудиовидеозаписи вебинара:

```
import moviepy.editor as mp
video_clip = mp.VideoFileClip(filename = pathVideo + nameVideo)
audio_clip = video_clip.audio
```

В представленном примере с помощью библиотеки Python `moviepy.editor` сначала загружается видеотрек вебинара, затем извлекается из него аудиотрек в формате `wav`.

Первым шагом в работе с аудиопотоком является фильтрация. Для данной задачи был выбран полосовой фильтр Баттерворта [4], АЧХ которого является достаточно гладкой на частотах полосы пропускания, соответствующих спектру голоса человека. Используется фильтр с бесконечной импульсной характеристикой (БИХ-фильтр), которая вычисляется по рекурсивной формуле:

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{k=1}^M a_k y(n-k),$$

где N – порядок фильтра прямой связи; M – порядок фильтра обратной связи ($M \leq N$); a_k и b_k – коэффициенты фильтра. Для вычисления коэффициентов применяется функция `butter(order, wn, btype)`, где `order` – порядок фильтра, `wn` – диапазон частот (частота среза), `btype` – тип фильтра. Для применения полученных коэффициентов используется функция `lfilter: filtered = lfilter(b, a, x)`, где `b, a` – полученные с помощью функции `butter` коэффициенты свертки; `x` – данные, которые нужно отфильтровать.

4.1. Извлечение интервалов времени тишины

К самым неинформативным частям аудиопотока относятся не только участки шума, но и участки тишины (участки, где нет никаких голосов). Речевой сигнал, содержащий одно слово, как минимум содержит паузу до и после произнесения слова. В случае с голосовым сигналом, содержащим высказывание из нескольких слов, участками такой информации выступают паузы между словами.

Среди алгоритмов выделения границ речи наибольшую популярность получил метод, основанный на энергии и количестве переходов сигнала через нулевое значение в короткие промежутки времени, и алгоритм, основанный на плотности распределения значения отсчетов паузы. Однако методы на основе кратковременной энергии речевого сигнала и числа нулей интенсивности обладают менее высоким качеством. Чаще используется метод выделения речи на основе нормального распределения.

В алгоритме на основе нормального распределения случайной величины в качестве таковой выступают значения отсчетов паузы (фонового шума). Таким образом, некоторые свойства функций нормального распределения, которые имеются в сигнале паузы (тишины, шума), будут являться критериями и использоваться для сегментации звука на сигнал и паузу.

Для выделения интервалов времени тишины используется понятие нормального распределения и одномерного расстояния Махаланобиса. При этом алгоритм извлечения паузы состоит из следующих шагов:

- 1) вычисление статистических характеристик шума;
- 2) выполнение нормализации сигнала по амплитуде;
- 3) выполнение фрагментирования всего аудиосигнала;
- 4) вычисление расстояния Махаланобиса и выполнение проверки.

Вначале происходит вычисление статистических характеристик распределения точек c_k на множестве (мощности m) точек выделенного аудиотрека: МО \bar{c} и СКО σ первых 200 мс аудиотрека:

$$\bar{c} = \frac{1}{m} \sum_{k=1}^m c_k, \quad \sigma = \sqrt{\frac{1}{m} \sum_{k=1}^m (c_k - \bar{c})^2}.$$

Этот участок аудиопотока считается паузой. После чего происходит нормализация сигнала по амплитуде. Для этого используется фильтр с конечной импульсной характеристикой (КИХ-фильтр):

$$\bar{s}(n) = s(n) - 0.95 \cdot s(n-1),$$

где $\bar{s}(n)$ – отсчет n в выходном отрезке \bar{s} ; $s(n)$ – отсчет n во входном отрезке s . Нормализованный аудиосигнал разбивается на отрезки по 25 мс. Для каждого такого отрезка от точки x вычисляется одномерное расстояние Махаланобиса:

$$d_M(x) = \frac{|x - \bar{c}|}{\sigma}.$$

Если это расстояние больше 4, то отрезок считается сигналом (голосом), иначе – паузой.

4.2. Извлечение интервалов времени разговора ведущего

Для извлечения интервалов времени разговора ведущего из полученных интервалов времени голоса (см. подраздел 4.1) реализованы ручной и полуавтоматический способы. В ручном режиме пользователь вручную на основе уже выделенных интервалов голоса вводит интервалы времени, в которые ведущий не говорит. При работе полуавтоматического режима решается задача идентификации личности по голосу, которая основана на парном сравнении биометрических признаков (содержащихся в голосе индивидуальных значимых признаков личности), выделенных из речи в фонограмме. Структура процесса распознавания личности по голосу в данной работе изображена на рис. 2.



Рис. 2. Структура процесса распознавания личности по голосу

Среди методов обучения текстонезависимого (обучаемого и тестируемого на произвольных речевых данных) идентификатора наиболее используемым является метод моделей гауссовых смесей (Gaussian mixture models, GMM). Создается GMM с помощью функции `gmm = GaussianMixture(n_components, max_iter, covariance_type, n_init)`, где `gmm` – выходная GMM; `n_components` – количество компонент GMM; `max_iter` – количество итераций обучения GMM; `covariance_type` – тип ковариационной матрицы; `n_init` – количество начальных инициализаций при обучении (K-means).

Обучение GMM проводится с помощью метода `fit` построенной GMM. Её синтаксис: `gmm.fit(X)`, где `X` – обучающие данные.

В роли речевых признаков в большинстве систем идентификации личности по голосу используются мел-частотные кепстральные коэффициенты (mel-frequency cepstral coefficients, MFCC). MFCC извлекаются с помощью библиотечной функции `python_speech_features.mfcc(data, sr, winlen, winstep, numcep)`, где `data` – данные аудиопотока, из которых извлекаются коэффициенты; `sr` – частота дискретизации аудиосигнала; `winlen` – длина окна в секундах, на которые будут разбиты `data`; `winstep` – длина шага по `data` (определяет, будут ли перекрываться окна); `numcep` – количество получаемых коэффициентов на каждом окне.

Таким образом, в полуавтоматическом режиме пользователь вручную вводит один небольшой интервал времени, в котором говорит ведущий. На основе введенного интервала идет обучение GMM-модели голоса ведущего. Затем идет разбиение всего аудиопотока на отрезки по 25 мс и проверка принадлежности обученной модели каждого отрезка. После разбиения отрезков на 2 части пользователю предлагается вручную откорректировать полученные интервалы времени, в которых не говорит ведущий. Исходя из того, что в вебинаре большую часть времени говорит ведущий, корректировка таких интервалов более целесообразна.

5. Алгоритм работы с изображениями

Как отмечалось в разделе 2, большая часть показателей рассчитывается на основе анализа видеопотока вебинара. Для этого использованы средства стандартной библиотеки OpenCV, которая предоставляет множество алгоритмов для обработки изображений и распознавания текста [5].

Рассмотрим работу с областью конференции вебинара. Для начала эту область необходимо выделить из видеопотока. Для этого интервал времени между двумя последовательно взятыми кадрами вебинара выбран равным значению параметра длины окна разбиения аудиосигнала (25 мс, см. раздел 4.1). Предполагается, что все области в вебинаре имеют стандартное относительное положение. В частности, область конференции расположена в правой верхней части кадра (см. рис. 1).

5.1. Предобработка данных

Прежде чем начать работать со слайдом видеопотока вебинара, необходимо произвести его предобработку. Алгоритм предобработки слайда согласно [6]:

- слайд переводится с помощью порогового преобразования `Threshold` в бинарное изображение (`CV_THRESH_BINARY`);
- получившееся изображение переводится в цветовое пространство HSV с помощью функции `cvtColor`.

После предобработки на слайде находятся контуры. Результатом будет бинарное изображение, в котором содержатся лишь контуры. Эту операцию проводит функция `canny = cv2.Canny(image, threshold1, threshold2)`, где `image` – одноканальное изображение для обработки (в градации серого, в данном случае оно бинарное); `threshold1` – порог минимума; `threshold2` – порог максимума.

Для того чтобы оперировать контурами и хранить их, необходимо заполнить матрицу точек, которая будет хранить двумерные координаты контуров изображения. Для этого используется функция `contours, hierarchy = cv2.findContours(image, mode, method, offset)`, где `image` – исходное одноканальное изображение (ненулевые пиксели обрабатываются как 1, а нулевые как 0, для получения такого изображения из градаций серого можно, например, использовать функции `Threshold` или `Canny`); `contours` – координаты

контуров на выходе; `hierarchy` – иерархия контуров (при использовании соответствующих режимов поиска); `mode` – режим поиска; `method` – метод аппроксимации. Так как все области вебинара имеют прямоугольную форму, то в качестве режима поиска используется режим `RETR_EXTERNAL`, обеспечивающий поиск только крайних внешних контуров, а в качестве метода аппроксимации и хранения в программе используется метод `CHAIN_APPROX_SIMPLE`, так как в этом режиме они будут кодироваться лишь с помощью четырёх угловых точек. При этом анализируются не сами контуры, а минимальные прямоугольники, которые их покрывают, что исключает возможность работы с неровными фигурами контуров. Для получения минимального прямоугольника, покрывающего выделенный почти прямоугольный контур, используется стандартная функция: `minAreaRect = cv2.minAreaRect(points)`, где `points` – последовательность или массив точек (в нашем случае – контур, который необходимо обвести).

Далее необходимо произвести фильтрацию выделенных областей. Из предположения, что границы областей вебинара являются самыми большими прямоугольниками, первичная фильтрация проходит по размеру полученных прямоугольников. Вторичная фильтрация проводится по заголовкам (по надписи над полученными прямоугольниками, если она есть). Для проведения вторичной фильтрации необходимо выделить в отдельное изображение и рассмотреть небольшие прямоугольники над полученными ранее, которые должны содержать в себе название областей вебинара (на рис. 1 – презентация, конференция, пользователи, чаты). После их получения необходимо провести их предобработку, прежде чем будет произведена процедура поиска и распознавания в нём текста. Сначала необходимо увеличить размер этих изображений при помощи функции `out_im = cv2.resize(in_im, size)`, где `in_im` – входное изображение; `out_im` – выходное изображение; `size` – размер выходного изображения. Далее изображение с помощью порогового преобразования `Threshold` переводится в бинарное изображение (`CV_THRESH_BINARY`) и в цветовое пространство `Gray` с помощью функции `cvtColor`.

5.2. Обнаружение лица ведущего и дальнейшая его обработка

Для обнаружения лица ведущего была использована функция `faces = dlib.get_frontal_face_detector().detector(image)`, где `image` – входное изображение области конференции; `faces` – лист координат всех лиц на входном изображении (в нашем случае – координаты одного лица ведущего вебинара). В результате получены координаты лица ведущего в положении анфас (лицом к смотрящему) на слайде. Данная функция использует метод НОГ, который эффективно идентифицирует лицо ведущего, представленное в цветовом пространстве `Gray` [7], так как данные о цвете не нужны для решаемой задачи. Если хотя бы на одном из двух анализируемых слайдов лицо ведущего не было найдено, то считается, что ведущий отвлекся и ничего не говорит. Дальнейшие действия будут происходить в прямоугольнике, в котором находится лицо ведущего вебинара.

Для анализа динамики в изображении лица ведущего используется оценка положения 68 лицевых ориентиров (ключевых точек) лица человека (см. рис. 3), предложенная Вахидом Кэземи и Джозефином Салливаном [8]. Она основана на ансамбле деревьев регрессии (деревьев решений), обучение которых осуществляется с помощью алгоритма градиентного бустинга деревьев с суммой квадратичных потерь. Для нахождения координат лицевых точек используются функции `predictor=dlib.shape_predictor(shapes_dat_file)` и `shape = predictor(image, face)`, где `shapes_dat_file` – путь к отдельному файлу формата `dat`, который содержит уже обученный на поиск лицевых точек каскад деревьев; `predictor` – загруженный детектор лицевых точек; `image` – входное изображение области конференции; `face` – координаты лица ведущего на `image`; `shape` – координаты лицевых точек на слайде.

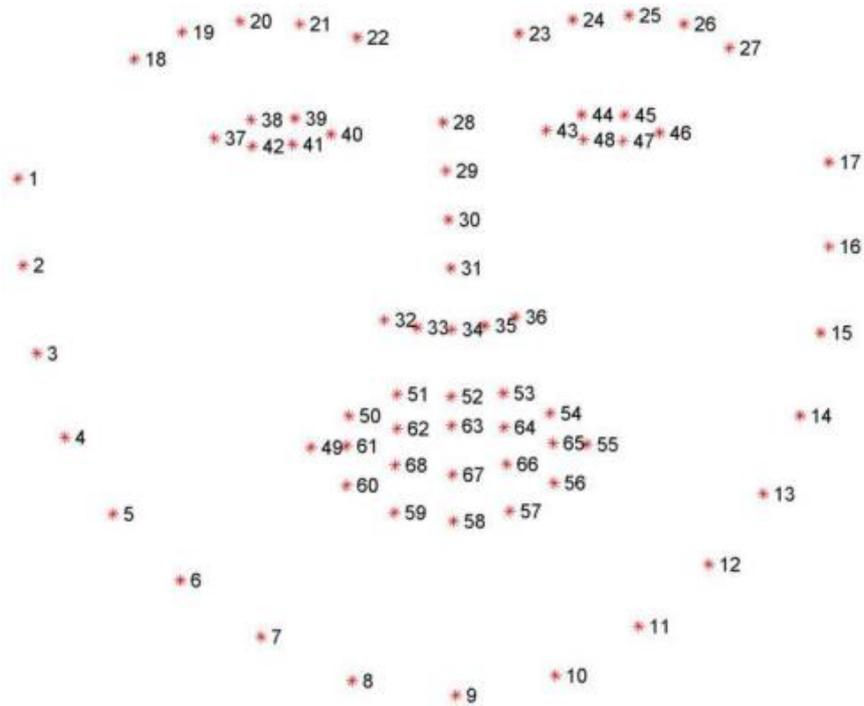


Рис. 3. Локализация антропометрических точек локализации лицевых ориентиров [8]

На рис. 3 видно, что для проведения анализа необходимо лишь 20 точек лица в области губ ведущего (с 49 по 68). Чтобы зафиксировать движение, считается расстояние Евклида между соответствующими точками в двух рассматриваемых слайдах. Если количество точек, чье расстояние попало в заданный интервал, больше половины списка анализируемых точек, то считается, что движение произошло. Интервал допустимых расстояний задается с целью фильтрации слишком больших или незримых амплитудных движений. Таким образом, были выделены все интервалы времени для решения поставленной задачи.

5.3. Анализ движения головы ведущего

Для фиксирования движения ведущего всеми остальными частями тела, кроме головы, используется оставшая часть области конференции вебинара. Чтобы понять, как изменился кадр по сравнению с предыдущим, разработана функция вычитания изображений `absdiff(im1, im2, res)`, где `im1`, `im2` – входные изображения, которые вычитаются; `res` – выходное изображение. Полученное изображение перед анализом проходит предобработку до бинарного. Для получения более точного результата и более качественного контроля выполняется морфологическая операция сужения изображения – увеличение темных областей прямоугольным ядром 5×5 . Изображение, полученное в результате, показано на рис. 4.

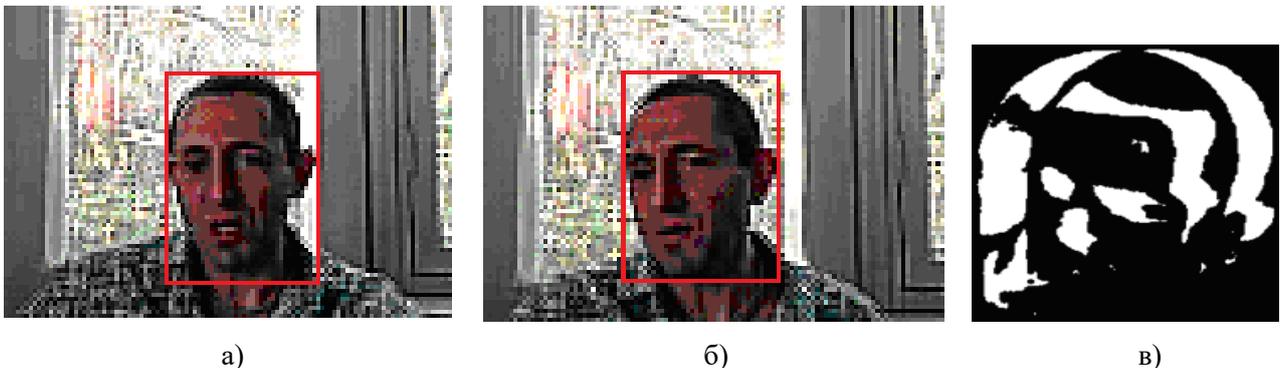


Рис. 4. Результат разности двух кадров изображения ведущего в области конференции
а) входное изображение `im1`; б) входное изображение `im2`; в) выходное изображение `res`

После завершения предобработки всё пространство, которое находится вокруг прямоугольника, описывающего голову ведущего, разделяется на 4 части: северная, южная, западная и восточная. Если хотя бы в одной из них количество ненулевых пикселей будет отличаться от нуля (это означает, что ведущий совершил движение), то счетчик движения увеличивается на 1.

Проведенный анализ области конференции видеопотока анализируемого вебинара при установленных амплитудах 10, 10, 10 по осям X, Y, Z соответственно и интервале 1000 мс между анализируемыми кадрами показал более точный результат по сравнению с интервалами времени 100 или 3000 мс. Это объясняется тем, что при установлении малого интервала времени движения проходят более плавно. Это означает, что амплитуда движения будет также малой, что приводит к потере (неучитыванию) некоторых совершённых ведущим движений. При установлении большого интервала времени также возможна потеря некоторых движений, так как они могли быть совершены в нерассматриваемых кадрах. Таким образом, наиболее подходящие интервалы времени находятся примерно на отрезке [500, 1500] мс. Так, при интервале 100 (3000) мс в течение всего вебинара ведущим было совершено 632 (1048), 360 (978), 462 (698) движений головы по осям X, Y, Z соответственно и 545 (899) движений остальных частей тела. При интервале 1000 мс в течение всего вебинара ведущим было совершено 2286, 1878, 1414 движений головы по осям X, Y, Z соответственно и 1827 движений остальных частей тела. На рис. 5 представлена зависимость точности результата от выбранного интервала времени.

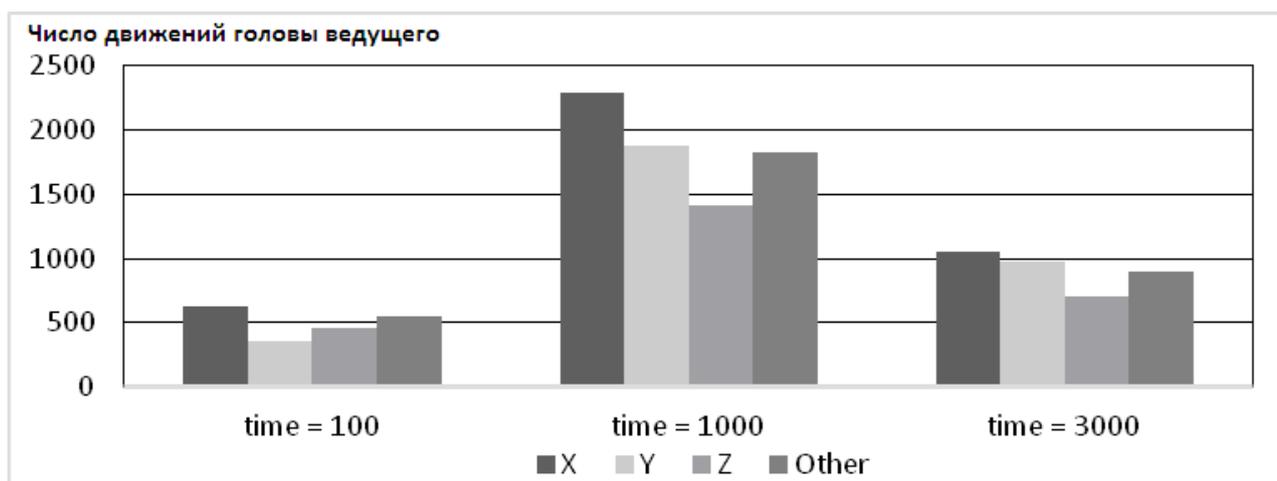


Рис. 5. Диаграмма зависимости зафиксированного числа движений головы ведущего вебинара от выбранного интервала времени в миллисекундах

5.4. Локализация текста на изображениях

На сегодняшний день алгоритмы локализации текста развиты достаточно хорошо. Тем не менее они не могут дать абсолютно точный результат, так как входные данные могут быть подвержены искажениям. Основные из них: низкое разрешение, сложный фон, неравномерное освещение, геометрические искажения, дисторсия, шумы и др. В отношении анализа слайдов презентации вебинара актуальны лишь первые два недостатка. Поэтому достоверность результатов локализации текста (а именно локализация необходима для оценивания показателей 14–16, см. табл. 1) во многом зависит от качества исполнения слайдов презентации, которые демонстрируются на кадрах видеопотока вебинара.

Первый этап большинства алгоритмов распознавания текста – его локализация. Главное отличие текста от фона и изображений – высокий контраст. Также предполагается, что текст имеет определенный цвет и яркость, что и отделяет его от фона. Кроме того, текст презентации чаще всего выполнен одинаковым шрифтом постоянного размера. Для таких задач хорошо работают градиентные методы локализации текста [9]. Идентификация блоков текста

осуществляется аналогично выделению области конференции в режиме RETR_EXTERNAL (см. подраздел 5.2). На последнем этапе алгоритм выделения текста требует выполнения процедуры фильтрации контуров, для чего используется несколько работающих последовательно фильтров. Первый из них удаляет все вложенные друг в друга контуры и те, чья площадь была меньше допустимой. Второй фильтр пытается, кроме этого, отличить текстовый блок от других изображений, основываясь на том факте, что текстовый блок можно разбить на строки, строки можно разбить на слова, а слова разбить на символы. Формы символов, в свою очередь, имеют определенные ограничения. В реализации для каждой строки вычисляется величина: *высота строки / среднее арифметическое ширины символа в строке*. Далее вычисляется медиана ряда полученных выше величин. Эта медиана t проверяется неравенством: $1 \leq t \leq 3.5$. Если неравенство выполняется, то данный блок считается текстовым, иначе отбрасывается. Для выделения строк текста для заданной области вычисляется вертикальная гистограмма как вектор, длина которого равна высоте выделенной области, а каждый элемент равен количеству небелых пикселей в соответствующей строке. Каждая непрерывная последовательность ненулевых элементов вектора определяется границей строки текста.

Пример работы алгоритма выделения текста представлен на рис. 6. С использованием полученного методом `findContours` массива точек (см. подраздел 5.1) рассчитывается площадь, занимаемая выделенными на слайде блоками. В данном примере текстовые блоки занимают 0.719 пространства слайда.

5.5. Распознавание текста на изображениях

Алгоритм распознавания текста на изображении состоит из следующих шагов:

1. Предобработка изображения посредством масштабирования, поворота и сегментации текста.
2. Применение сверточных слоёв для извлечения признаков из изображения.
3. Объединение извлеченных признаков и преобразование в единый вектор признаков.
4. Применение полносвязных слоёв для классификации полученного вектора признаков.
5. Получение результата распознавания текста.

Предобработка изображения рассмотрена в подразделе 5.1. Наряду с этим использован метод билинейной интерполяции, который позволяет уменьшить размерность изображения, сгладив пиксели, сохраняя при этом основные детали. Уменьшение размерности изображения позволяет снизить объём данных, с которыми необходимо работать, что повышает эффективность работы алгоритмов распознавания текста [10].

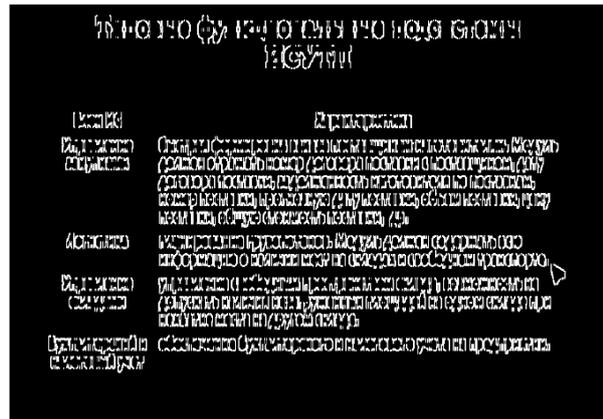
Шаги 2–4 реализуют процедуру поиска и распознавания текста на изображении с помощью библиотеки Tesseract OCR, в которой реализованы нейросетевые алгоритмы машинного обучения для распознавания текста на изображениях. Это открытое ПО, которое позволяет автоматически распознавать как единичную букву, так и сразу текст. Его существенным недостатком является плохая работа с замыленным, битым, грязным и деформированным текстом, однако авторы исходят из того, что область кадра вебинара не содержит таких повреждений текста.

Для получения текста на изображении (в нашем случае – заголовка области вебинара) используется функция `text = pytesseract.image_to_string(image, lang)`, где `image` – входное изображение; `lang` – обозначение используемого для распознавания языка в виде строки («rus», «eng» и др.); `text` – текст с входного изображения в виде строки.

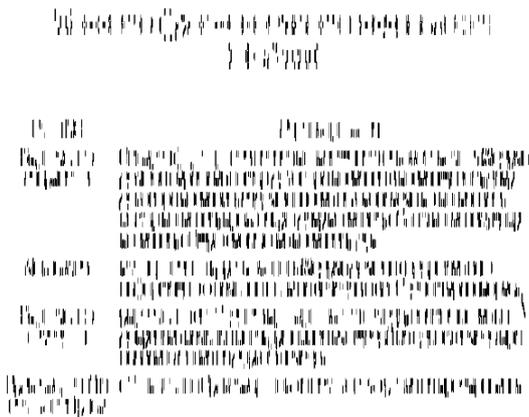
Результатом работы Tesseract являются строки, содержащие текст на входных изображениях. Прямоугольная область, в процессе обработки которой было получено, что заголовком является строка «Конференция», содержит область конференции. Информация о расположении области конференции вебинара обновляется в течение всего вебинара (установлен интервал времени обновления 5 минут). Соответственно, по строкам «Презентация», «Пользователи» и «Чаты» выполняется идентификация соответствующих областей.

Типовые функциональные подсистемы ИСУТП	
Блок ИС	Характеристика
Управление закупками	быстрая формировка заказа поставщикам-изготовителям. Модуль должен отражать номер договора поставки с поставщиком, дату договора поставки, задолженность изготовителя по поставкам, номер поставки, прогнозную дату поставки, объем поставки, цену поставки, общую стоимость поставки, др.
Логистика	планирование грузопотоков. Модуль должен содержать всю информацию о наличии мест на складах и свободном транспорте.
Управление складами	управление свободным пространством склада, возможность не допускать наличия незагруженных площадей на одном складе при нехватке места на другом складе.
Бухгалтерский и налоговый учет	обеспечение бухгалтерского и налогового учета на предприятии.

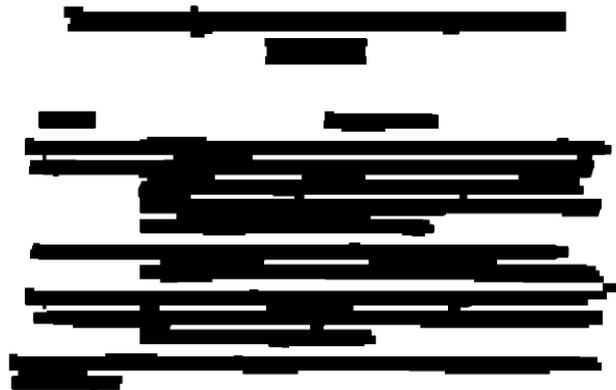
а)



б)



в)



г)

Типовые функциональные подсистемы ИСУТП	
Блок ИС	Характеристика
Управление закупками	быстрая формировка заказа поставщикам-изготовителям. Модуль должен отражать номер договора поставки с поставщиком, дату договора поставки, задолженность изготовителя по поставкам, номер поставки, прогнозную дату поставки, объем поставки, цену поставки, общую стоимость поставки, др.
Логистика	планирование грузопотоков. Модуль должен содержать всю информацию о наличии мест на складах и свободном транспорте.
Управление складами	управление свободным пространством склада, возможность не допускать наличия незагруженных площадей на одном складе при нехватке места на другом складе.
Бухгалтерский и налоговый учет	обеспечение бухгалтерского и налогового учета на предприятии.

д)

Типовые функциональные подсистемы ИСУТП	
Блок ИС	Характеристика
Управление закупками	быстрая формировка заказа поставщикам-изготовителям. Модуль должен отражать номер договора поставки с поставщиком, дату договора поставки, задолженность изготовителя по поставкам, номер поставки, прогнозную дату поставки, объем поставки, цену поставки, общую стоимость поставки, др.
Логистика	планирование грузопотоков. Модуль должен содержать всю информацию о наличии мест на складах и свободном транспорте.
Управление складами	управление свободным пространством склада, возможность не допускать наличия незагруженных площадей на одном складе при нехватке места на другом складе.
Бухгалтерский и налоговый учет	обеспечение бухгалтерского и налогового учета на предприятии.

е)

Рис. 6. Результат работы алгоритма выделения текста из изображения слайда презентации:
 а) исходное изображение, б) результат работы алгоритма поиска границ,
 в) результат инверсии и первого этапа морфологических операций,
 г) результат второго этапа морфологических операций – объединения границ символов в блоки,
 д) результат операции выделения контуров текстовых блоков,
 е) результат фильтрации контуров из предыдущего этапа

В задаче анализа качества вебинара по показателям из табл. 1 распознавание текста используется для:

- 1) точной идентификации областей в кадре видеопотока (область конференции, область презентации, область чата, область количества участников);
- 2) идентификации контактных данных ведущего и оценивания показателя 9 *Контактные данные*;

3) считывания количества участников в области количества участников для расчёта показателя *32 Показатели посещаемости*;

4) считывания текущего и общего времени вебинара из области времени вебинара для расчёта показателя *31 Точность заявленной длительности мероприятия*.

Пример работы алгоритма распознавания текста представлен на рис. 7. Функция `pytesseract.image_to_string` возвращает строку "1:27:35/1:29:43", которая при помощи функции `strptime(str, format)` библиотеки `datetime` определяет оперативное время вебинара T_0 (левая часть результирующей строки) и фактическую длительность вебинара T_f (правая часть результирующей строки).

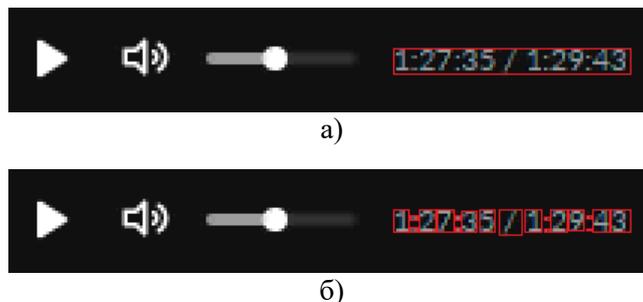


Рис. 7. Результат работы алгоритма выделения текста из изображения слайда презентации:
а) результат фильтрации контура, б) результат детализации контура для распознавания текста

Оперативное время используется для параметризации модели анализа звукового ряда (см. раздел 4), а также для расчёта некоторых показателей, отражающих динамику вебинара, например, *17 Использование указки* и *23 Частота взаимодействия*.

Фактическое время используется для расчёта точности соблюдения заявленной длительности вебинара T_B и его завершения в указанный в регламенте момент времени в соответствии с рисунком в приложении к данной статье. В рассмотренном примере при заявленном времени вебинара 1.5 часа значение показателя *31 Точность заявленной длительности мероприятия* составляет $9.96 \approx 10$ баллов.

6. Результат работы программы анализа аудиовидеопотока

Для компьютерной реализации разработанных моделей и алгоритмов написана программа на Python.

На первом этапе необходимо загрузить видеопоток вебинара. После загрузки видеопотока появится окно установки параметров (см. табл. 3). Значения этих параметров не следует устанавливать слишком малыми, так как при анализе выполняется большое число операций и расходуется много времени, а результаты излишне подробны и сложно анализируемы в дальнейшем. Также значения этих параметров не следует устанавливать слишком большими, так как возможно слияние разных типов интервалов времени, а их редактирование будет более затратно по времени. Все действия и инструкции программы на данном этапе будут написаны в текстовом окне «Статус процессов».

Таблица 3. Параметры алгоритма первичного анализа аудиовидеопотока вебинара

№	Наименование параметра	Значение по умолчанию
1	Длина одного фрейма, образованного при разбиении полученного аудиопотока, мс	100
2	Допустимая пауза между интервалами разных типов, мс	500
3	Минимальная длина фрейма после объединения соседних однотипных интервалов, мс	1000

После того как параметры будут установлены, выполняется фильтрация аудиосигнала и извлечение из него интервалов тишины. Исходя из того, что в интервалы голоса могли попасть посторонние звуки (стук, скрип, фоновый шум и др.), а в интервалы тишины могли попасть тихие голоса людей, имеется возможность редактировать образованные интервалы тишины. Из них потом будут образованы интервалы голосов. Для редактирования интервалов разработан специальный инструмент, обеспечивающий возможность изменять, удалять, склеивать имеющиеся в списке интервалы, а также добавлять новые интервалы. Все интервалы представлены в миллисекундах. Для улучшения анализа интервалов соответствующие им части аудиопотока записываются в отдельную директорию. Количество отредактированных интервалов и значения параметров зависят от оборудования записи вебинара, влияющего на качество звука и видео, а также голоса ведущего, частоты смены личности говорящего, фонового шума.

В рассматриваемом примере было выделено 226 интервалов тишины. На их основе было выделено 225 интервалов голосов. Так как в интервалы голоса могли попасть посторонние звуки (стук, скрип, фоновый шум и др.), а в интервалы тишины могли попасть тихие голоса людей, эти интервалы были отредактированы (всего на этом этапе внесено 19 изменений). После внесения изменений было получено 207 интервалов голосов. Остальные интервалы отнесены к интервалам тишины.

На втором этапе пользователь выбирает режим работы извлечения голоса ведущего из интервалов времени голосов (ручной или полуавтоматический).

При выборе ручного режима инициируется описанная выше процедура редактирования интервалов. Исходя из того, что большую часть аудиопотока занимает голос ведущего, на редактирование в качестве основного списка интервалов подаётся список интервалов не ведущего. При выборе ручного режима этот список интервалов будет пустым. Вспомогательным списком является выделенный ранее список голосов. На основе вспомогательного списка и заполняется основной список интервалов.

При выборе полуавтоматического режима работы сначала появится окно для ввода одного небольшого интервала (больше двух секунд), в котором говорит один ведущий. Этот интервал нужен для получения GMM голоса ведущего. После тренировки GMM автоматически будут выделены предположительные интервалы времени голосов не ведущего, которые затем будут доступны для редактирования.

В общем случае, если запустить программу на одном и том же видеопотоке несколько раз, то результаты получатся похожими, но разными. Связано это с тем, что среди недостатков EM-алгоритма обучения GMM есть неустойчивость к выбору начальных значений параметров. От них зависит как скорость сходимости, так и результат работы, то есть алгоритм может сойтись к локальному экстремуму функции правдоподобия, который может быть существенно ниже глобального. В качестве начальных параметров обучения в программе установлены обобщенные результаты работы модели кластеризации (методом k-means) некоторого количества данных из начальной выборки. Одним из недостатков этого метода является чувствительность к выбору начальных центров кластеров. Если они выбираются случайно, то это может являться источником погрешности. Поэтому целесообразно выполнять начальную кластеризацию несколько раз для получения более точного результата. После того как интервалы голосов не ведущего были откорректированы, остальные интервалы голосов автоматически будут считаться интервалами голоса ведущего.

В рассматриваемом примере был выбран полуавтоматический режим. Исходя из введённого интервала времени голоса ведущего, на котором обучалась GMM, было выделено 433 интервала времени разговора не ведущего, что не соответствует действительности. Для уменьшения разницы между полученными и реальными интервалами происходит их корректировка (всего на этом этапе внесено 217 изменений). Также здесь происходит фильтрация интервалов по длине. Интервалы с меньшей длиной попадают в интервалы голоса ведущего, исходя из того, что ведущий в вебинаре говорит большую часть времени. После корректировки было выделено 207 интервалов голоса ведущего.

На третьем этапе выполняется анализ изображений вебинара из интервалов времени голосов не ведущего с параметрами согласно табл. 4. После установки параметров выполняется анализ слайдов вебинара. Рассчитанные интервалы доступны для редактирования пользователем.

Таблица 4. Параметры алгоритма вторичного анализа аудиовидеопотока вебинара

№	Наименование параметра	Значение по умолчанию
1	Время между анализируемыми слайдами вебинара, мс	100
2	Минимальная и максимальная амплитуды движения губ ведущего (амплитуда, находящаяся в этом интервале, свидетельствует о том, что ведущий что-то говорит)	[2, 3]
3	Требуется ли показ области вебинара во время анализа слайдов	Нет
4	Требуется ли сохранять все полученные типы интервалов в файл	Да

Значения параметров 1 и 2 не следует задавать слишком малыми и слишком большими. При слишком малых значениях разница положения ведущего в анализируемых слайдах является почти незаметной. Значит, будет пропущено много движений. При слишком больших значениях также будут пропущены некоторые движения, которые произошли между рассматриваемыми слайдами. Чтобы резкие движения головы или тела ведущего не считались движением губ, в качестве параметра 2 используется интервал.

После завершения работы программы в директории с видеофайлом вебинара появляется директория «Result» с двумя текстовыми файлами с интервалами тишины (209 интервалов) и интервалами голоса ведущего (207 интервалов).

Таким образом, на протяжении работы системы на данном примере выделено 659 интервалов. После 236 исправлений было сохранено 416 интервалов.

Следует заметить, что для увеличения достоверности полученных результатов разметки аудиовидеопотока и уменьшения затрат на корректировку в конечной реализации модели необходимо реализовать композицию анализа аудио- и видеопотоков. Однако это относится к следующему, заключительному, этапу исследований.

7. Заключение

В данной статье на основе методов Data Mining разработано и исследовано математическое обеспечение для интеллектуального анализа аудиовидеопотоков вебинаров в части идентификации и оценивания статических и динамических характеристик вебинара с целью упрощения и объективизации их оценивания. Это вторая часть исследования, имеющая целью повышение оперативности и достоверности оценивания некоторых показателей качества вебинара, допускающих интеллектуальный анализ аудиовидеопотока вебинара без активного участия эксперта и непосредственного применения экспертных методов оценивания.

На основе первой части исследования [1] была поставлена задача разработки математического аппарата, рассмотрена теоретическая часть для решения поставленной задачи, определены средства разработки для создания специального программного обеспечения. Для создания ПО задействованы функциональные возможности языка Python для машинного обучения и подключаемых к нему модулей и библиотек (OpenCV, DLib, LibROSA, PyQt5, NumPy и др.). Полученные компоненты математического обеспечения исследованы на контрольном аудиовидеопотоке вебинара и показали свою эффективность.

Третьей и завершающей частью исследования должна стать разработка формального аппарата для вычисления 10-балльных оценок показателей вебинара, а также формирования

модельно-инструментального комплекса, включающего в себя а) компоненты математической модели [1], отвечающие за формирование и отладку системы показателей качества вебинара, б) компоненты математического обеспечения на основе методов Data Mining, разработанные в данной статье, и в) формализмы для вычисления иерархической системы оценок качества вебинара. Построенный модельно-инструментальный комплекс потребует исследования адекватности и эффективности как гибридной модели в целом, так и отдельных её подмоделей, в том числе посредством сравнения результатов оценивания с полученными ранее результатами в рамках исследования [1] оценок качества вебинаров.

Приложение

В приложении содержится таблица, отражающая систему показателей качества вебинара, разработанная в ходе исследования [1]. Представлены шкала и критерии оценивания показателей, а также временная сложность процедуры их оценивания согласно табл. 2 основной части статьи. В таблице не отражены веса показателей, которые не играют существенной роли для материала статьи. В конце приложения представлен рисунок для оценивания показателя 31 *Точность заявленной длительности мероприятия*.

Таблица. Показатели качества вебинара, критерии и экспертные оценки сложности их оценивания

№	Показатель	Критерии оценивания		Оценка	Сложность
I Сценарий: начало вебинара					
1.	Самопрезентация	Нет	<i>Ведущий не представился</i>	0	1
		Очень краткая		1...2	
		Краткая Достаточная	<i>Ведущий представился</i>	3...5 6...10	
2.	Обозначение регламента, в т.ч. правил	Нет Краткое Достаточное (обозначены правила, в т.ч. дисциплина ответов на вопросы)		0 3...5 6...10	1
3.	Обозначение цели вебинара	Нет Краткое Достаточное		0 3...5 6...10	1
4.	Обозначение плана вебинара	Нет Краткое Достаточное (обозначена ценность)		0 3...5 6...10	1
II Сценарий: ход вебинара					
5.	Проблематика	Не обозначена Не обозначена. Эпизодические обозначения «болевых» точек Обозначена в начале вебинара. Регулярные обозначения «болевых» точек		0 3...5 6...10	4
6.	Ясность	Понятность, логичность и последовательность изложения		0...10	2
7.	Лаконичность	«Вода»		0	2
		Разжиженный, преимущественно текстовый, материал		1...4	
		Более-менее компактный материал с некоторым количеством «воды»		5...7	
		Концентрированный материал со схемами и иллюстрациями		8...10	

№	Показатель	Критерии оценивания	Оценка	Сложность
8.	Резюмирование	В конце нет резюме В конце подводятся итоги, есть резюме Резюмирование происходило как в конце, так и несколько раз внутри вебинара	0 3...5 6...10	4
9.	Контактные данные	Не представлены Представлены	0 10	1
III Качества ведущего				
10.	Энергичность и эмоциональность ¹	Вялый, заторможенный, пессимистичный, эмоции выражаются неуместно Напряженный, агрессивный, торопливый, неуместное выражение эмоций Бодрый, раскованный, оптимистичный, эмоции выражаются уместно	0...2 3...7 8...10	2
11.	Темп изложения	Очень медленно, скучно Очень быстро Монотонно, заметные паузы Размеренно, ритмично, «как песня льется»	0...1 2...3 4...6 7...10	2
12.	Компетентность	Компетентность невысокая, опыта нет Ведущий приводит некоторые абстрактные примеры из практики Опыт ведущего налицо, приводимые примеры привязаны к реальности	0...2 3...5 6...10	3
13.	Внешний вид ¹	Опрятность, корректность визуального фона, шумы звукового фона	0...10	1
IV Качество презентационных материалов				
14.	Количество и качество информации на слайдах	Слайды сильно перегружены, на них слишком много монотонного текста Слайды слегка перегружены или недогружены, местами требуется больше текста, иллюстраций там, где есть только голос и видео Слайды использованы оптимально	0...2 3...7 8...10	1
15.	Количество и качество иллюстраций	Скучные и однообразные текстовые слайды Более-менее разнообразные слайды, мало иллюстраций, много текста Интересные и разнообразные слайды, иллюстрации преобладают над текстом	0...2 3...7 8...10	1
16.	Аккуратность	– одинаковые списки – единая цветовая гамма для всех элементов – отсутствие неаккуратностей (обрезанные края, лишние линии и т.п.) – примерно одинаковые размеры изображений и их размещение на слайдах – одинаковые размеры шрифтов (не более 2–4 размеров на всю презентацию) из диапазона 20–30 пт – одинаковые расстояния между элементами	0...10	1
17.	Использование указки	Нет Эпизодическое Регулярное и уместное Чрезмерное	0 5 10 3	3
18.	Разнообразие презентационных инструментов	Ничего Указка, рисование, доска, поднятие рук, эмоции, демонстрация рабочего стола	0 1...10	3

№	Показатель	Критерии оценивания	Оценка	Сложность
V Интерактив по инициативе ведущего				
19.	Знакомство	Нет Ведущий знакомится со слушателями, выясняет их степень знакомства с предметом изучения (обратная связь через чат или виртуальные руки, опрос)	0 3...10	2
20.	Проведение опросов	Нет 1 опрос 2 и более опросов	0 5 10	3
21.	Формат вопросов ведущего	Открытые, предполагающие длинные ответы Открытые, предполагающие короткие ответы, или закрытые, предполагающие знак в чате, виртуальную эмоцию или виртуальное поднятие руки	0...5 5...10	4
22.	Разнообразие интерактива	– вывод участников в эфир – вызов участников к доске – дискуссии – конкурсы и т.д.	0...10	3
23.	Частота взаимодействия	Нет Интервалы более 10 минут Интервалы менее 5 минут Интервалы от 5 до 10 минут	0 1...3 4...5 6...10	2
VI Интерактив по инициативе слушателей				
24.	Наличие вопросов слушателей	Нет 1–2 3–5 6 и более	0 2 5 10	1
25.	Реактивность на вопросы ²	Нет Промежуточный вариант Моментальная и в соответствии с регламентом (правилами)	0 1...9 10	4
26.	Качество ответов на вопросы ²	– озвучивание имени задающего вопрос – повторение вопроса вслух – вербальная оценка вопроса – лаконичные ответы	0...10	4
VII Оценки слушателей				
27.	Понятность ³	Оценка слушателей	0...10	
28.	Интересность ³	Оценка слушателей	0...10	
29.	Полезность ³	Оценка слушателей	0...10	
VIII Параметры результативности вебинара				
30.	Точность начала мероприятия	Пунктуальность ведущего, начало мероприятия в заявленное время T_0 с погрешностью 5 минут	0...10	1
31.	Точность заявленной длительности мероприятия	Завершение мероприятия в заявленное время (T_B – заявленная ведущим длительность вебинара ⁴)	0...10 (см. рис. ниже)	1
32.	Коэффициент результативной посещаемости вебинаров	Отношение числа слушателей, дослушавших вебинар до конца, к числу слушателей, посетивших вебинар	0...10	1
Средняя сложность экспертного оценивания:				2.87

¹ При отсутствии видеопотока число баллов по этим показателям не может быть больше 5.

² При отсутствии вопросов слушателей данный показатель принимается равным нулю.

³ При отсутствии оценок слушателей данные показатели принимаются равными нулю.

⁴ В случае, если ведущий не обозначил регламент в начале вебинара, принимается во внимание время, заявленное при планировании вебинара (в анонсе).

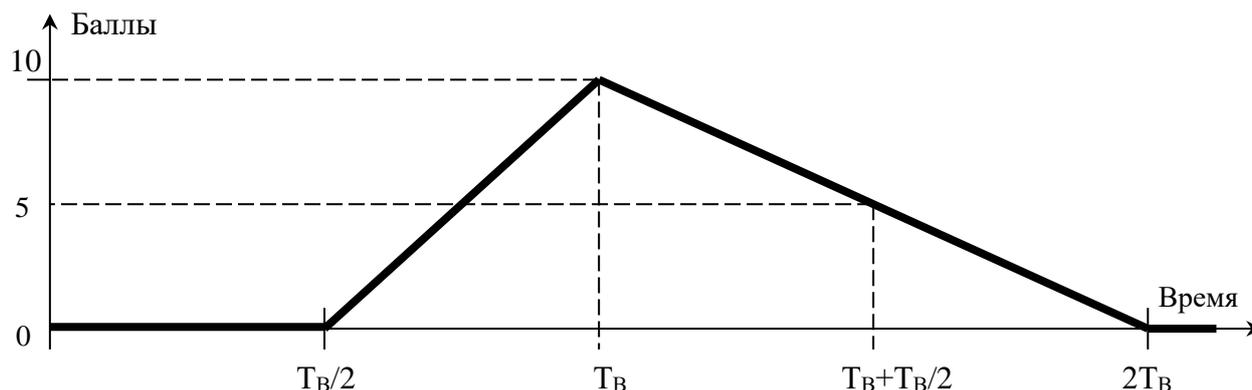


Рис. 8. График определения оценок показателя
31. Точность заявленной длительности мероприятия [1]
(T_в – заявленная длительность вебинара, мин.)

Литература

1. Полетайкин А. Н., Шевцова Ю. В., Подколзин В. В., Струкова Е. Г. Математическая модель оценивания качества контактной работы, реализуемой посредством вебинаров в ходе дистанционного обучения // Информатика и образование. 2019. № 7 (306). С. 42–53. DOI: 10.32517/0234-0453-2019-34-7-42-53.
2. Сизов Л. А. Вебинары – передовая технология непрерывного сетевого образования // Материалы XIV Международной научно-практической конференции «Проблемы экономики и информатизации образования». Тула: ТИЭИ. 2017. С. 92–95.
3. Цаллагов Н. А., Астахова Т. А. Проведение эффективных вебинаров от А до Я. ООО «Мираполис», 2015. 174 с. / [Электронный ресурс]. URL: <http://indo.kg/wp-content/uploads/2018/11/Проведение-эффективных-вебинаров-от-А-до-Я.pdf> (дата обращения: 31.12.2023).
4. Сорокин Г. А. Фильтры нижних частот // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. 2015. № 1. С. 100–107.
5. Попов А. Н., Шелупанов А. А. Машинное обучение. Эволюция методов. М.: ДМК Пресс, 2019. 304 с.
6. Bradski G. and Kaehler A. Learning OpenCV: Computer Vision with the OpenCV Library. First Edition. USA: O'Reilly Media, 2008. 543 p.
7. Андрянов Н. А., Дементьев В. Е., Ташлинский А. Г. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet // Компьютерная оптика. 2022. Т. 46, № 1. С. 139–159. DOI: 10.18287/2412-6179-CO-922.
8. Kazemi V. and Sullivan J. One millisecond face alignment with an ensemble of regression trees // Proc. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014. P. 1867–1874. DOI: 10.1109/CVPR.2014.241.
9. Liang J., Doermann D., and Li H. Camera-based analysis of text and documents: a survey // IJDAR. 2005. V. 7. P. 84–104. DOI: 10.1007/s10032-004-0138-z.
10. Дворянкин Д. К., Кулешов Д. А., Кадурын А. В. Нейронные сети и глубокое обучение. СПб.: Питер, 2019. 352 с.

Подколзин Вадим Владиславович

к.ф.-м.н., заведующий кафедрой информационных технологий, Кубанский государственный университет (350040, Краснодар, ул. Ставропольская, 149), e-mail: vvp_kubsu@mail.ru, ORCID: 0000-0002-4491-1493.

Полетайкин Алексей Николаевич

к.т.н., доцент, доцент кафедры информационных технологий Кубанского государственного университета (350040, Краснодар, ул. Ставропольская, 149), тел. +7 861 2199-577;

доцент кафедры математического моделирования и цифрового развития бизнес-систем СибГУТИ (630102, Новосибирск, ул. Кирова, 86);

e-mail: alex.poletaykin@gmail.com, ORCID ID: 0000-0002-5128-1952.

Галкина Марина Юрьевна

доцент, доцент кафедры прикладной математики и кибернетики, Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ, 630102, Новосибирск, ул. Кирова, д. 86), тел. +7 383 2698 272, e-mail: gmur7@sibguti.ru, ORCID ID: 0009-0002-4515-9923.

Авторы прочитали и одобрили окончательный вариант рукописи.

Авторы заявляют об отсутствии конфликта интересов.

Вклад соавторов: Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.

Mathematical Apparatus for Some Quality Indicators Assessment of a Training Webinar Using Machine Learning

Vadim V. Podkolzin¹, Aleksey N. Poletaikin^{1,2}, Marina Yu. Galkina²

¹ Kuban State University (KubSU)

² Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: This article is the second stage of creating and researching a technology for assessing the quality of contact work at a university implemented through webinars. The mathematical support of the intelligent technology for analyzing the audio-video stream of an educational webinar recording is considered. The problem of increasing the efficiency and reliability of assessing some indicators of the quality of contact work at the Faculty of Distance Learning of SibSUTIS implemented through distance learning webinars is being solved. Machine learning methods and their implementation using standard Python modules and libraries are used for analysis. The scientific significance of the work lies in the development of new models and algorithms for intelligent analysis of audio-video streams parameterized for processing the recording of an educational event. The developed models and algorithms make it possible to increase the efficiency of the process of assessing the quality of contact work implemented through webinars.

Keywords: webinar, conference area, presenter activity, performance evaluation criteria, time intervals, text localization, audio track processing, image processing, machine learning.

For citation: Podkolzin V. V., Poletaikin A. N., Galkina M. Yu. Mathematical apparatus for some quality indicators assessment of a training webinar using machine learning (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 1, pp. 121-143. <https://doi.org/10.55648/1998-6920-2024-18-1-121-143>.



Content is available under the license
Creative Commons Attribution 4.0
License

© Podkolzin V. V., Poletaikin A. N.,
Galkina M. Yu., 2024

The article was submitted: 15.12.2023;
accepted for publication 15.01.2024.

References

1. Poletaikin A. N., Shevtsova Yu. V., Podkolzin V. V., Strukova E. G. Matematicheskaya model' otsenivaniya kachestva kontaknoi raboty, realizuemoi posredstvom vebinarov v khode distantsionnogo obucheniya [Mathematical model for assessing the quality of contact work implemented through webinars during distance learning]. *Informatika i obrazovanie*, 2019, no. 7 (306), pp. 42-53, DOI: 10.32517/0234-0453-2019-34-7-42-53.
2. Sizov L. A. Vebinary – peredovaya tekhnologiya nepreryvnogo setevogo obrazovaniya [Webinars - advanced technology for continuous online education]. *Problemy ekonomiki i informatizatsii obrazovaniya. Materialy XIV Mezhdunarodnoi nauchno-prakticheskoi konferentsii*, Tula, TIEI, 2017. pp. 92-95.
3. Tsallagov N. A., Astakhova T. A. *Provedenie effektivnykh vebinarov ot A do Ya* [Conducting effective webinars from A to Z]. ООО «Mirapolis», 2015, 174 p., available at: <http://indo.kg/wp-content/uploads/2018/11/Provedenie-effektivnykh-vebinarov-ot-A-do-Ya.pdf> (accessed: 31.12.2023).
4. Sorokin G.A. Fil'try nizhnikh chastot [Low pass filters]. *Vestnik YuUrGU, Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika*, 2015, no. 1, pp. 100-107.
5. Popov A. N., Shelupanov A. A. *Mashinnoe obuchenie. Evolyutsiya metodov*. [Machine learning. Evolution of methods]. Moscow, DMK Press, 2019. 304 p.
6. Gary Bradski, Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*, First Edition, USA, O'Reilly Media, 2008. 543 p.
7. Andriyanov N. A., Dement'ev V. E., Tashlinskii A. G. Obnaruzhenie ob"ektov na izobrazhenii: ot kriteriev Baiesa i Neimana–Pirsona k detektoram na baze neyronnykh setei EfficientDet [Detection of objects in an image: from Bayes and Neyman-Pearson criteria to detectors based on EfficientDet neural networks]. *Komp'yuternaya optika*, 2022, vol. 46, no. 1, pp. 139-159. DOI: 10.18287/2412-6179-CO-922.
8. Kazemi V., Sullivan J. One millisecond face alignment with an ensemble of regression trees. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 1867-1874, DOI: 10.1109/CVPR.2014.241.
9. Liang J., Doermann D., Li H. Camera-based analysis of text and documents: a survey. *IJDAR*, 2005, vol. 7, pp. 84-104, DOI: 10.1007/s10032-004-0138-z.
10. Dvoryankin D. K., Kuleshov D. A., Kadurin A. V. Neironnyye seti i glubokoe obuchenie [Neural networks and deep learning]. Saint Petersburg, Piter, 2019. 352 p.

Vadim V. Podkolzin

Cand. of Sci. (Physical and Mathematical), Head at the Department of Information Technologies, (KubSU, Krasnodar, Russia). e-mail: vvp_kubsu@mail.ru, ORCID: 0000-0002-4491-1493, Scopus AuthorID: 57215326416, Researcher ID: AAU-8042-2021.

Aleksey N. Poletaikin

Cand. of Sci. (Engineering), Assistant Professor at the Information Technologies Department, Kuban State University (KubSU, Krasnodar, Russia).

Assistant professor at the Mathematical Modeling and Digital Development of Business System Department, Siberian State University of Telecommunications and Information Science (SibSUTIS, Russia, 630102, Novosibirsk, Kirov St. 86). e-mail: alex.poletaykin@gmail.com, ORCID ID: 0000-0002-5128-1952, Scopus AuthorID: 57213829361, ResearcherID: ABF-6799-2020.

Marina Yu. Galkina

Assistant Professor of the Department of Applied Mathematics and Cybernetics, Siberian State University of Telecommunications and Information Science (SibSUTIS, Russia, 630102, Novosibirsk, Kirov St. 86). e-mail: gmur7@sibguti.ru, ORCID ID: 0009-0002-4515-9923.