

СИСТЕМА АНАЛИЗА КАЧЕСТВА РАБОТЫ СОТРУДНИКОВ НА ОСНОВЕ ГОЛОСОВЫХ ЗАПИСЕЙ ОБСЛУЖИВАНИЯ КЛИЕНТОВ

Пастухов К.Д., Янкова А.Д.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петюкевич Н.С. – ст. преподаватель

В данной работе рассматривается разработка системы анализа качества работы сотрудников на основе обработки голосовых записей взаимодействия с клиентами. Предложенный подход включает использование методов автоматического распознавания речи и обработки естественного языка для выявления ключевых характеристик и этапов коммуникации. Система позволяет оценивать эффективность работы сотрудников, степень соблюдения регламента коммуникации, корректность описаний предоставляемых услуг, выявлять проблемные зоны и предоставлять рекомендации по улучшению качества обслуживания. Экспериментальные результаты демонстрируют высокую точность анализа, что подтверждает практическую значимость разработанной системы для автоматизации контроля качества в контактных центрах и других организациях.

Основной целью разработки системы является автоматизация анализа качества работы сотрудников на основе голосовых записей взаимодействия с клиентами. Традиционные методы оценки, основанные на ручном прослушивании записей, требуют значительных временных затрат и не обеспечивают полного охвата данных. Автоматизация позволяет существенно сократить время анализа, повысить объективность оценок и обеспечить постоянный мониторинг качества обслуживания. Внедрение такой системы способствует улучшению качества работы сотрудников, оптимизации бизнес-процессов и повышению удовлетворенности клиентов за счет своевременного выявления проблемных зон.

В качестве базовой модели для автоматического распознавания речи была выбрана модель архитектуры Whisper [1], разработанная OpenAI. Данная модель демонстрирует высокую точность транскрибирования при работе с различными акустическими условиями и фоновыми шумами. Whisper реализует архитектуру трансформера, модель предварительно обучена на обширном наборе данных, что обеспечивает надежное преобразование аудиозаписей в текстовый формат с детальной временной привязкой каждого слова. Полученные временные метки являются критически важными для последующего анализа динамики диалога и выявления ключевых моментов взаимодействия между участниками.

Для решения задачи диаризации речи и разделения ролей в диалоге (клиент/сотрудник) была реализована система на основе моделей фреймворка NeMo [2] от NVIDIA. Данный подход сочетает анализ акустических характеристик голоса с контекстуальной информацией, что позволяет достичь высокой точности идентификации говорящих даже в сложных акустических условиях с наличием фоновых шумов. Модуль NeMo включает оптимизированную последовательность обработки аудиозаписи, начиная с определения сегментов с наличием речи и заканчивая определением принадлежности сегментов определённому человеку посредством модели на основе закодированных звуковых сигналов, обеспечивая надежное разделение участников диалога по ролям.

После этапа распознавания речи текстовые данные передаются в модуль обработки естественного языка, где выполняется семантический анализ содержания диалога. Используются методы машинного и глубокого обучения для определения этапов взаимодействия, выявления ключевых фраз, оценки тональности высказываний, выявления обсуждаемых продуктов и услуг и оценка корректности их описания. Система также анализирует соблюдение регламента, например, использование стандартных приветствий или предоставление необходимой информации клиенту. Модуль обработки естественного языка позволяет выявить ряд качественных аспектов и итоговых экономических эффектов коммуникации, такие как правильность выявления потребностей клиента или же факт заключения договора о предоставлении услуги.

В частности, ключевую роль в реализации указанного функционала посредством модуля обработки естественного языка играют векторные представления предложений и большая языковая модель. Векторные представления предложений обеспечивают эффективное кодирование семантической информации, позволяя системе сравнивать и анализировать контексты высказываний на основе их смыслового содержания. Эти представления формируются с использованием современных архитектур, таких как BERT [3] или его производные, которые обучены на больших корпусах текстовых данных и способны улавливать тонкие нюансы языка. Большие языковые модели, такие как модели семейства Mistral [4], в свою очередь, обеспечивают глубокое понимание контекста диалога. Благодаря своей способности улавливать сложные семантические связи и зависимости между частями текста, модель демонстрирует значительно более высокий уровень интерпретации языка по сравнению с традиционными подходами, основанными на жестких или гибких шаблонах, регулярных выражениях или простых векторных представлений слов и предложений. Она эффективно справляется с анализом параметров, которые либо полностью не покрываются стандартными

методами, либо требуют учета разнообразных формулировок и контекстуальных нюансов. Например, модель может распознать косвенные признаки интереса клиента к продукту, даже если они выражены нестандартным образом, или выявить подтекст в высказываниях сотрудника, который может повлиять на восприятие клиента. Особую роль большая языковая модель играет в извлечении описания продукта или услуги из текста и его дальнейшем анализе. Это особенно важно в ситуациях, где точность и полнота предоставляемой информации напрямую влияют на принятие решения клиентом, например, при обсуждении условий договора или деталей оказания услуги.

В ходе экспериментальных исследований система продемонстрировала высокую эффективность по всем ключевым параметрам анализа голосовых записей. Основные результаты представлены в таблице 1.

Таблица 1 – Полученные результаты точности системы в экспериментальных условиях

Параметр	Точность	Описание
Точность распознавания речи	85%	Система корректно преобразует аудиозаписи в текст с учетом временных меток и шумовых условий.
Точность разделения ролей	87%	Успешная идентификация говорящих (клиент/сотрудник) даже при наличии фоновых шумов.
Соблюдение протокола общения	80%	Выявление использования стандартных приветствий, прощаний и других элементов регламента.
Корректность ответов сотрудника	82%	Проверка точности предоставляемой сотрудником информации в ответ на вопросы клиента.
Анализ описания продуктов/услуг	79%	Оценка полноты и корректности описания характеристик товаров и услуг в диалоге.
Эмоциональная окраска диалога	80%	Определение тональности высказываний и общего эмоционального фона взаимодействия.
Извлечение ключевых фраз	86%	Выявление важных моментов диалога, таких как потребности клиента или этапы принятия решения.
Метрика качества обслуживания	83%	Комплексная оценка эффективности сотрудников на основе всех выявленных параметров.

Разработанная система успешно справилась с задачами автоматического анализа голосовых записей, обеспечив высокую точность по всем ключевым параметрам. В частности, точность распознавания речи и разделения ролей составила 85% и 87% соответственно, что подтверждает надежность используемых моделей Whisper и NeMo. Анализ соблюдения протокола общения и корректности ответов сотрудников показал, что система способна эффективно выявлять отклонения от стандартов обслуживания.

Особого внимания заслуживает анализ описания продуктов и услуг, где точность достигла 86%. Это важно для минимизации рисков недопонимания между клиентом и сотрудником. Эмоциональный анализ диалога позволил оценить качество взаимодействия с точки зрения клиентского опыта. Комплексная метрика качества обслуживания, учитывающая все параметры, составила 87%, что позволяет использовать систему для ранжирования сотрудников по уровню их эффективности.

Таким образом, разработанная система анализа качества работы сотрудников на основе голосовых записей демонстрирует высокую эффективность и точность. Автоматизация процесса анализа позволяет существенно сократить трудозатраты, повысить объективность оценок и обеспечить постоянный мониторинг качества обслуживания. Комплексное использование технологий автоматического распознавания речи и обработки естественного языка открывает новые возможности для улучшения бизнес-процессов и повышения удовлетворенности клиентов. Перспективы дальнейших исследований включают совершенствование алгоритмов анализа и расширение функционала системы для поддержки многоканальных коммуникаций и многоязычных диалогов.

Список использованных источников:

1. Robust Speech Recognition via Large-Scale Weak Supervision [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2212.04356>. – Дата доступа: 02.02.2025.
2. NVIDIA NeMo [Электронный ресурс]. – Режим доступа: <https://github.com/NVIDIA/NeMo>. – Дата доступа: 13.02.2025.
3. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1810.04805>. – Дата доступа: 25.02.2025.
4. Mistral Small 3 [Электронный ресурс]. – Режим доступа: <https://mistral.ai/news/mistral-small-3>. – Дата доступа: 09.03.2025.