

Часто для решения сложных задач или даже создания новых IT-продуктов используют искусственный интеллект. Но он появляется не сам по себе: обучением машин занимаются специалисты по Data Science. Именно такую роль вы сегодня и примерите на себя. Ваша основная задача: разработать модель машинного обучения для перевода аудиоконтента в текст.

Вы разберётесь, откуда вообще появляются алгоритмы и нейросети, на конкретной задаче обучите модель распознавания слов и проверите, насколько эффективно она работает. А ещё, несмотря на техническую сложность, вам может пригодиться знание русского языка — ведь надо будет указать алгоритму на его грамматические ошибки.

Часть 1. Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Это задание весьма непростое, хоть и не кажется таким на первый взгляд. Всё дело в том, что вам нужно **отличить задачи машинного обучения**, то есть буквально то, над чем будет работать алгоритм, **от бизнес-задач**, то есть задач всей компании, её отдельных департаментов или продуктов.

Обратите внимание: продуктовые задачи, или задачи бизнеса, обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей. Зачастую они даже имеют какой-то численный показатель, при котором задача считается успешно выполненной.

А вот задача машинного обучения всегда связана с **обработкой данных**, в результате которой получается прогноз или вывод о какой-то проблеме. Например, если вы генерируете изображение по текстовому запросу, то в результате получаете картинку, которая, по мнению алгоритма, лучше всего его отражает. Вопрос лишь в том, насколько точным оказалось «попадание» в этот запрос.

Как пройти это задание?

Поднять оценку «Удобный чат» от пользователей на 10% — это точно не задача машинного обучения, это так называемая продуктовая задача, которую можно решить разными способами. Новый функционал на основе машинного обучения может стать одним из инструментов для её достижения.

Зато чтобы пользователям было удобнее общаться, можно внедрить такой алгоритм, который люди бы оценили по достоинству. Задача такого алгоритма может звучать примерно так: распознавать данные в аудиодорожке и переводить её в текст. Именно такую задачу нужно решить специалисту по Data Science. И когда этот функционал позволит расшифровывать голосовые сообщения, чат точно станет удобнее!

Попробуйте распределить варианты по своим группам самостоятельно, но если вдруг возникают сложности, можете подсматривать в ответы ниже.

Задачи машинного обучения:

1. Распознавать данные в аудиодорожке и переводить её в текст
2. Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях

Продуктовые задачи:

1. Увеличить ежедневное количество пользователей чата на 15%
2. Чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения
3. Поднять оценку показателя «Удобный чат» на 10%
4. Добавить возможность расшифровки «капчи*» для слабовидящих

*Капча, или captcha — это специальный защитный код, по которому программа определяет, реальный ли вы человек или бот. Вы наверняка видели такую при входе в некий сервис: нужно ввести цифры на экране, определить объекты на картинке или ответить на вопрос. **1**

Часть 2. Сбор и подготовка данных

Самое важное для машинного обучения — данные. Данных нужно много, они должны быть разнообразны и качественно отобраны.

Если мы работаем над алгоритмом, который будет распознавать текст в аудио, то и данные для обучения нужны соответствующие. Чем больше у вас будет дорожек с записанной речью, тем лучше.

Как пройти это задание?

Посмотрите на предложенные варианты. Обратите внимание: они не воспроизводятся по нажатию на кнопку Play, поэтому под каждой звуковой дорожкой есть текст с расшифровкой.

Визуально оцените плотность данных на каждой дорожке, сравните её длительность и количество слов, а затем сделайте вывод: какие из аудиофайлов НЕ подходят для обучения модели.

Сравнили? Оценили? Нашли некачественные записи? Молодцы!

А если нет, то нажмите «Исключить» напротив записей №2, №5 и №6.

Часть 3. Препроцессинг

Сам по себе звук представляется в памяти компьютера как массив значений, показывающих колебания амплитуды по времени. Этот параметр обычно исчисляется в десятках тысяч точек в секунду (или кГц). Звучит сложно, но суть в том, что такая дорожка получается очень длинной и неудобной для работы.

Поэтому перед подачей в алгоритм звук предобрабатывают: в частности, переводят в спектрограмму — зависимость интенсивности колебания звука на конкретной частоте по времени. Ух, звучит ещё сложнее, да? Если коротко: компьютеру намного проще работать со значениями частот.

Не переживайте, специалистам по Data Science в этой задаче помогает специальное программное обеспечение. А чтобы вам стало немного понятнее, мы предлагаем пройти интерактивное задание.

Как его пройти?

Вам нужно перевести звук из вида осциллограммы в спектрограмму. Нажимайте на все пиковые значения на «зелёном» графике, чтобы они отразились на спектрограмме.

После этого надо убедиться, что точно все значения сохранились, и вновь проставить недостающие точки.

Часть 3. Акустическая модель

После того как сырая дорожка переведена в удобное для нейросети представление, мы готовы распознавать речь: получать из звука распределение вероятностей букв по времени.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени. То есть модель «смотрит» на спектрограмму, видит там набор звуков и пытается спрогнозировать, где какая фонема расположена.

Как пройти это задание?

Чаще всего такие модели работают по принципу «что слышится, то и пишется». Поэтому вам нужно подобрать в каждом слове не букву, соответствующую словарному написанию, а звук — как если бы вы делали транскрипцию.

Вы можете посмотреть на визуальное обозначение фонемы и подобрать ей соответствующую пару из предложенных сверху вариантов, а затем выбрать его в выпадающем списке.

Часть 3. Лингвистическая модель

Ура! Вы составили фонетические транскрипции, а теперь с помощью отдельной языковой модели «причешите» результат — пора исправить грамматические и орфографические ошибки в расшифровках.

Так как модель предсказывает, какой набор символов вероятнее всего встретить в языке, её обучают на классических текстах, словарях и статьях, сленге, разговорных сообщениях и неологизмах.

Как его пройти?

Сейчас вы на месте алгоритма: нужно исправить все ошибки в словах! Выбирайте подходящие варианты и двигайтесь дальше.

Это, конечно же, не экзамен по русскому языку, поэтому вы можете ошибаться, и ничего вам за это не будет.

Варианты ответов не очень запутанные, но если вдруг не удаётся распознать ошибку, то вот правильный набор слов:

Пойдем вечером в восемь в кино там вышел новый классный фильм

Часть 3. Пунктуационная модель

После того, как мы получили читаемую последовательность слов, можем расставить знаки препинания. Это особенно полезно, когда текст длинный: с точками и запятыми читать его точно будет проще.

Подход в генерации данных для обучения тут похожий: мы берём тексты с расставленными знаками препинания и искусственно «портим» эти данные, убирая пунктуацию. А затем учимся восстанавливать.

Как пройти это задание?

Попробуйте самостоятельно отметить места, в которых должны стоять знаки препинания. Не волнуйтесь, здесь тоже можно ошибаться. Профпроба — это не контрольная работа.

Чтобы не было совсем просто, даём такую подсказку:

- В примере №1 пропущены два знака.
- В примере №2 — два знака.
- В примере №3 — пять знаков.
- В примере №4 — два знака.

Отлично! Теперь нужно отметить на дорожке места, в которых вероятнее всего должны стоять знаки препинания. Ориентируйтесь на паузы между словами и нажимайте на кликабельные зоны звуковой дорожки.

Часть 4. Оценка результата с помощью метрик

Метрика — это какой-то рассчитываемый показатель, который используется для оценки эффективности. Метрики помогают понять, насколько хорошо работает модель, и сравнить её с другими моделями. В нашем случае используется метрика word error rate — чем меньше ошибок делает алгоритм, тем выше её значение.

Представьте аналогию: ваша успеваемость тоже может быть представлена в виде метрики. Допустим, у вас 10 предметов, и за четверть вы получили все пятерки. Получается, что ваш средний балл — 5,0.

А что если сюда затесались тройка по физкультуре и четвёрка по литературе? Тогда средний балл будет уже 4,7.

Примерно так работают метрики — они позволяют оценить самые разные показатели.

Как пройти это задание?

Сейчас на аудиодорожке можно заметить пять отдельных кусочков с какими-то словами. Нажмите поочерёдно на каждый из них, чтобы расшифровать.

Спойлер! Ваша модель работает так круто, что метрика WER равна 100%. В жизни, конечно, так почти не бывает — даже самый крутой алгоритм иногда ошибается. А уж на стадии обучения ошибки — обычное дело, и специалист по Data Science ищет способ их решения и улучшения модели.