



Литература

1. Soudry D., Hoffer E., Nacson M.S., Gunasekar S., Srebro N. The implicit bias of gradient descent on separable data//JMLR. – 2018. – 19. – pp. 1–57.
2. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения: 28.05.2019)
3. Kim H.S., Kang J.H., Park W.M., Ko S.H., Cho Y.H., Yu D.S., Song Y.S., Choi J.W. Convergence analysis of optimization algorithms. arXiv:1707.01647 (дата обращения: 28.05.2019)
4. Кузнецова К.А., Наумова Ю.А., Панова К.А. Исследования влияния верхней и нижней асимптот на скорость сходимости логарифмической функции потерь//Новые информационные технологии в научных исследованиях: материалы XXIII Всероссийской научно-технической конференции. Рязанский государственный радиотехнический университет. – С. 67-68.

Д.А. Лебедев, В.Г. Литвинов

АВТОМАТИЗИРОВАННАЯ СИСТЕМА КЛАССИФИКАЦИИ МУЗЫКАЛЬНЫХ ЗАПИСЕЙ ПО ЖАНРАМ

(Самарский университет)

Аннотация: в данной статье рассматривается способ классификации музыкальных записей по жанрам. Определены основные принципы, которым должно соответствовать приложение. Исследование показало, что классификация музыкальных записей по жанрам на основе распознанного текста может дать приемлемый по точности результат.

Ключевые слова: выборка, звук, частота звука, распознавание аудио, факторы точности распознавания.

Выборка или выборочная совокупность — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).

Характеристики выборки:

Качественная характеристика выборки — что именно мы выбираем и какие способы построения выборки мы для этого используем.

Количественная характеристика выборки — сколько случаев выбираем, другими словами объём выборки.

Существует необходимость в сборе вторичной информации. [1].

Звуком называется колебательное движение частиц упругой среды, распространяющееся в виде волн в газообразной, жидкой или твердой среде, которые, воздействуя на слуховой анализатор человека, вызывают слуховые ощущения. Источником звука является колеблющееся тело, например: колебания струны, вибрация камертона, движение диффузора громкоговорителя и др. [2].



Частота - звуки, воспринимаемые слуховым анализатором человека, образуют диапазон звуковых частот. Принято считать, что этот диапазон ограничен частотами от 16 до 20000 Гц. Эти границы весьма условны, что связано с индивидуальными особенностями слуха людей, возрастными изменениями чувствительности слухового анализатора и методом регистрации слуховых ощущений. Человек может различить изменение частоты на 0,3% на частоте порядка 1 кГц. [2].

Распознавание аудио происходит в три этапа:

Выделяются слова. Обычно существует несколько гипотез распознанного слова.

Гипотезы проверяются с помощью языковой модели. Модель проверяет, насколько согласуется новое слово со словами, распознанными ранее.

Обрабатывается распознанный текст — числительные преобразуются в цифры, расставляются некоторые знаки препинания (например, дефисы) и т. д. Этот преобразованный текст и является финальным результатом распознавания, который отправляется в теле ответа. [3].

На точность распознавания влияют [3]:

- 1 Качество исходного звука;
- 2 Качество кодирования аудио;
- 3 Разборчивость и темп речи;
- 4 Сложность фраз и их длина

Цель создания приложения – найти способ классификации музыкальных композиций по жанрам, на основе текста самих композиций.

Главным критерием такой классификации будет являться точность определения жанра. Для распознавания текста музыкальных записей используется новый веб сервис Яндекс.Облако, который доступен для бесплатного использования так как находится на этапе разработки. Для целей данной работы использован подсервис Yandex SpeechKit — Комплекс технологий распознавания и синтеза речи.

Композиция содержит музыкальное сопровождением помимо голоса исполнителя, что усложняет распознавание речи. Поэтому необходимо выделить участок частот, соответствующий голосу человека, остальные интервалы будут рассматриваться как не соответствующие условию эксперимента.

Голос типичного взрослого мужчины имеет фундаментальную частоту (нижнюю) от 85 до 155 Гц, типичной взрослой женщины от 165 до 255 Гц. Для эксперимента был выбран интервал от 85 до 255 Гц, так как исполнитель может обладать как низким, так и высоким голосом.

Для разработки использовалась среда разработки Java, для взаимодействия с Yandex SpeechKit был получен необходимый набор токенов и идентификаторов, которые позволили взаимодействовать с данным сервисом посредством клиента.

На вход приложения поступает музыкальная композиция. Затем все частоты, не входящие в обусловленный голосовой интервал, занижаются, чтобы выделить только необходимый участок. На последнем этапе массив байт аудио



файла преобразуется в бинарный и отправляется запросом к сервису. В ответе будет получен распознанный текст.

Затем возможно несколько способов классификации по жанрам. Одним из которых является — приведение соответствия отдельных слов из текста выборкам, которые содержат характерный набор слов для каждого музыкального жанра.

Недостатком данного способа является то, что он не позволяет корректно обработать данные, в случае нечеткого голоса исполнителя либо ярко выраженного музыкального сопровождения, на фоне которого невозможно распознать голос.

Разрабатываемый продукт сможет помочь людям с достаточной точностью определить жанр музыкальной композиции либо подобрать для себя песню, опираясь на набор слов, который соответствует их любимому жанру.

Литература

- 1 Теория статистики. – Учебник/ под редакцией проф. Г.Л. Громько.- 2-е изд., перераб. и доп. - М.: ИНФРА–М, 2016 г.- 476 с., гл. 1, стр. 28-37; стр. 149-193.
- 2 Соколович Ю.А., Богданова Г.С. Физика: справочник. – 2-е издание перераб. – Х.: Веста: издательство «Ранок», 2015. – 464 с.
- 3 Распознавание речи. URL: <http://cloud.yandex.ru/docs/speechkit/stt/> (дата обращения: 14.12.2017).

Д.А. Лебедев, В.Г. Литвинов

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО РАЗМЕРА НЕЙРОННОЙ СЕТИ

(Самарский университет)

Аннотация: в данной статье рассматривается способ определения оптимального размера нейронной сети через сопоставление весов синапса от размера нейросети. Эксперимент показывает, что локальные минимумы хорошо соответствуют свойствам традиционных кривых обучения — зависимостей ошибок обучения от размера нейросети. Такой подход может быть использован для определения оптимального размера сети при отсутствии тестовой выборки.

Ключевые слова: размер сети, кривая обучения.

При использовании искусственных нейронных сетей важной задачей является нахождение оптимального размера (структуры) сети — такого числа скрытых слоев нейронов и нейронов в слоях, которые дадут максимум обобщающих способностей, т.е. минимум ошибки обобщения, особенно в случае отсутствия тестовой выборки или невозможностью разделить выборку данных на обучающую и тестовую части из-за недостаточности общего объема данных.

Поэтому широко используется понятие "кривых обучения — зависимостей ошибок обучения и обобщения от размеров нейросети и обучающей вы-