

ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ МОДЕЛЕЙ КЛАССИЧЕСКОГО МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ТОНАЛЬНОСТИ РЕЧЕВЫХ СИГНАЛОВ ЧЕЛОВЕКА

Талецкий А.И.

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

Калугина М.А. – канд. физ-мат. наук

Одно из наиболее популярных направлений в современном машинном обучении – методы классификации тональности различных данных, таких как фотографии лиц, тексты постов в интернете или записи речевых сигналов человека. Оценка тональности речевых сигналов отстает от других направлений и, в связи с этим, представляет особый интерес для различных исследований.

В работе используется датасет, загруженный с сайта стэндфордского университета и распространяемый по свободной некоммерческой лицензии. Датасет состоит из коротких аудиофайлов, каждый по 4 секунды. 12 мужчин и 12 женщин зачитывают 2 фразы в 8 тональных классах:

- нейтральность;
- спокойствие;
- счастье;
- грусть, печаль;
- злость, ненависть;
- испуг;
- удивление.

В результате варьирования различных параметров, не влияющих на общую тональность записи, получается датасет содержащий 2800 примеров.

Для оценки качества работы алгоритмов используется кросс-валидационная оценка [1]. Подсчет производится по следующим метрикам:

- precision;
- recall;
- f1-мера.

Результирующим считалось значение f1-меры.

Для упрощения разработки и последующего расширения программного средства было принято решение строить его по принципу слабосвязанных модулей на основе принципов проектирования SOLID. А именно, программное средство было разделено на три модуля:

- модуль извлечения признаков;
- модуль обучения модели;
- модуль использования модели.

Модуль извлечения признаков извлекает признаки из тонально размеченного набора аудио файлов и выполняет предварительную обработку, стандартизируя извлеченные признаки. В качестве извлекаемых из аудио сигналов признаков выбраны мел-кепстральные коэффициенты, получаемые в результате двойного преобразования Фурье. В статье [2] обосновывается выбор мел-кепстральных коэффициентов в качестве признаков.

Задача модуля обучения модели состоит в обучении выбранной на текущей итерации модели. Обучается модель на основе подготовленных на предыдущем шаге и, далее, сериализуется для последующей оценки качества её работы.

Модуль использования модели поочередно загружает обученные модели и выполняет кросс-валидационную оценку их качества работы.

В результате получаются следующие оценки качества работы алгоритмов:

- наивный Байес: 0.2 (или 20%);
- метод опорных векторов 0.3 (или 30%)
- линейная регрессия 0.4 (или 40%)
- решающие деревья 0.56 (или 56%)
- ближайшие соседи 0.67 (или 67%)
- композиция градиентного бустинга и ближайших соседей 0.71 (или 71%)

В дополнение стоит сказать, что весьма эффективными для решения этой задачи могут оказаться рекуррентные нейронные сети и это может быть темой отдельного исследования.

Список использованных источников:

1. A Study of CrossValidation and Bootstrap for Accuracy Estimation and Model Selection, Ron Kohavi, Computer Science Department, Stanford University, Stanford, CA, 9405.
2. D. Wu, T. D. Parsons, S.S. Narayanan, "Acoustic feature analysis in speech emotion primitives estimation", 2010.