

Определение жанра музыкальной композиции

Федор Петряйкин

Николай Серов

Алексей Кулагин

Москва, 2017

Цели и задачи

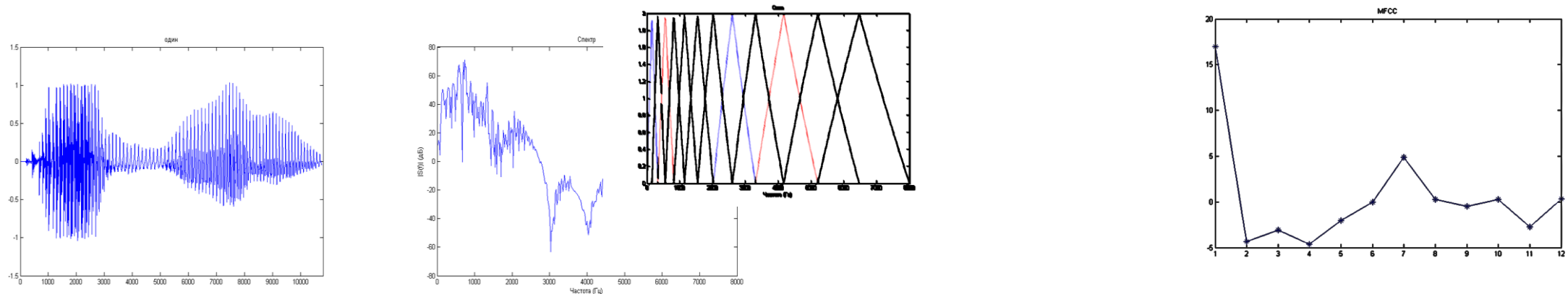
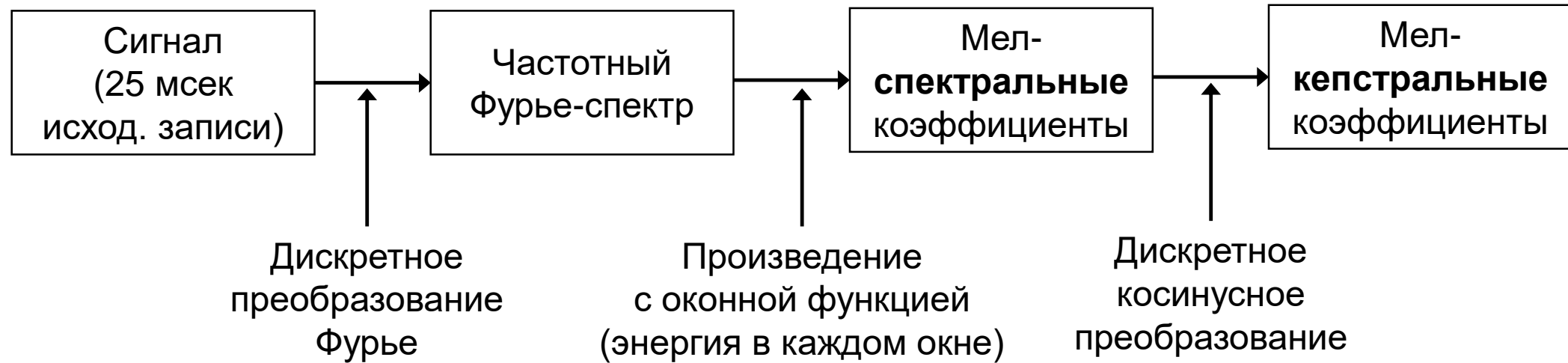
Цель: Разработать систему определения жанра представленной музыкальной композиции

Задачи:

- Разработать алгоритм сбора численных характеристик музыкальной композиции
- Подготовить **датасет**, выбрать **метрику**, провести на нем обучение нескольких моделей, на основании кросс-валидации выбрать **оптимальную модель** и ее **гиперпараметры**
- Оценить качество предсказания на **тестовой** выборке
- Реализовать прототип веб-интерфейса для работы с системой

Численное представление музыки

Были использован метод mfcc – метод мел-кепстральных коэффициентов



Далее рассчитывались статистики распределения этих коэффициентов по файлу

Используемые алгоритмы

Проблема: много фич (26 коэффициентов+статистики: mean, std, ...)

Используем линейные модели, допускающие **снижение размерности**:

- Линейный (квадратичный) дискриминантный анализ

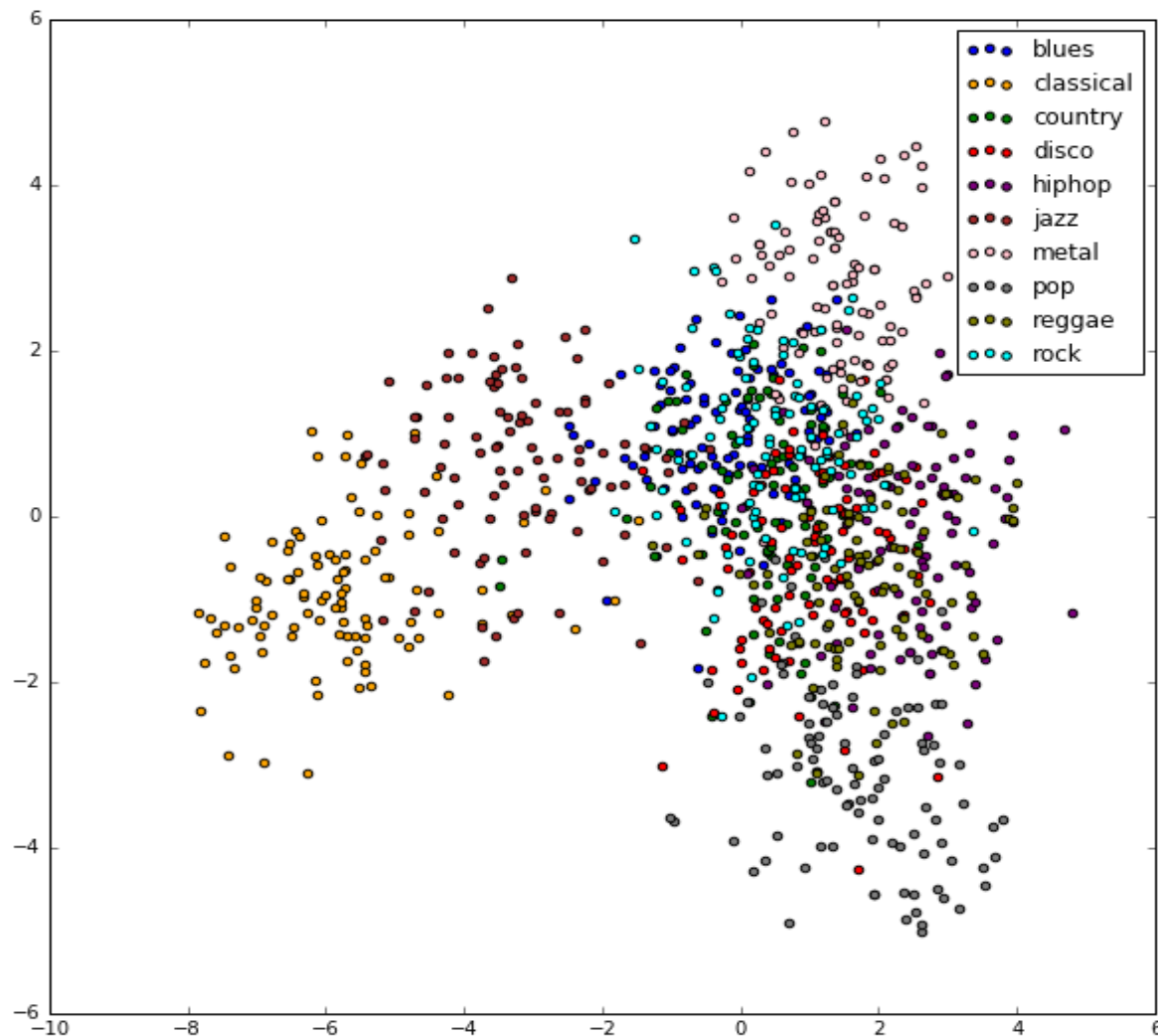
$$P(y = k|X) = \frac{P(X|y = k)P(y = k)}{P(X)} = \frac{P(X|y = k)P(y = k)}{\sum_l P(X|y = l) \cdot P(y = l)}$$

$$p(X|y = k) = \frac{1}{(2\pi)^n |\Sigma_k|^{1/2}} \exp \left(-\frac{1}{2} (X - \mu_k)^t \Sigma_k^{-1} (X - \mu_k) \right)$$

Размерность снижается при оценке матрицы ковариаций

- PCA+SVM

Датасет и визуализация данных



Визуализация датасета при помощи LDA

GTZAN dataset - один из самых часто используемых в задачах распознавания музыки.

В нем содержится 30-ти секундные фрагменты 1000 аудиозаписей следующих жанров: (blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock), по 100 аудио на каждый жанр.

Недостатки датасета: некоторые ошибки в заданной классификации

+: практически нет аномалий

Подбор гиперпараметров моделей

Основные подбираемые параметры:

LDA: Алгоритм построения матриц ковариации, регуляризация

QDA: Регуляризация

PCA: n_components

SVC: C (регуляризация), ядро (тип, степень, coef0)

$$y = (y_1, y_2, \dots, y_N)$$

$$\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)$$

Метрика: accuracy_score

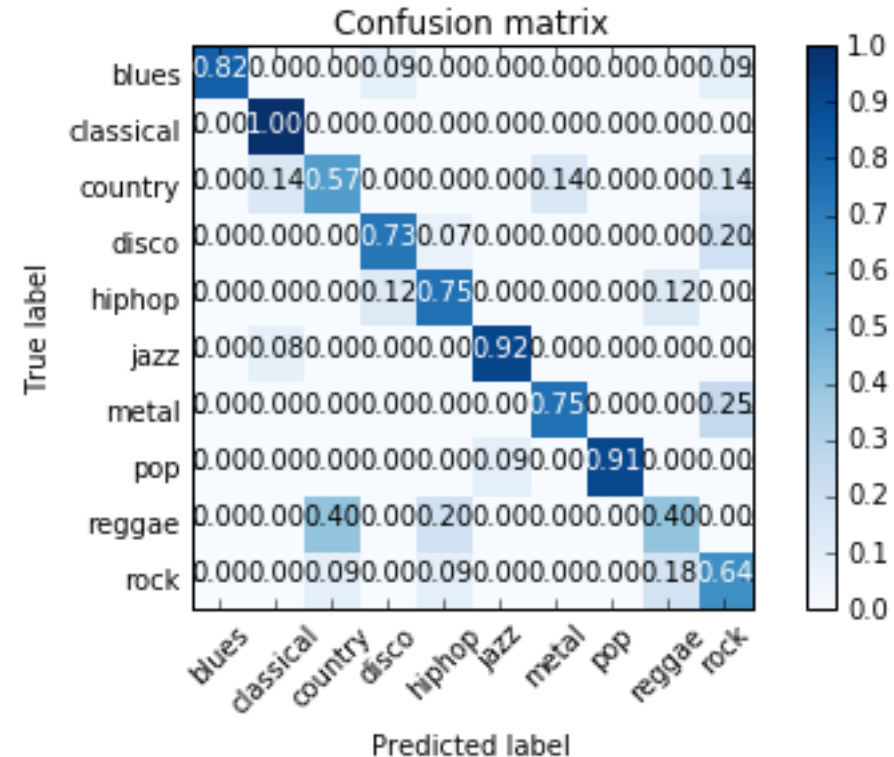
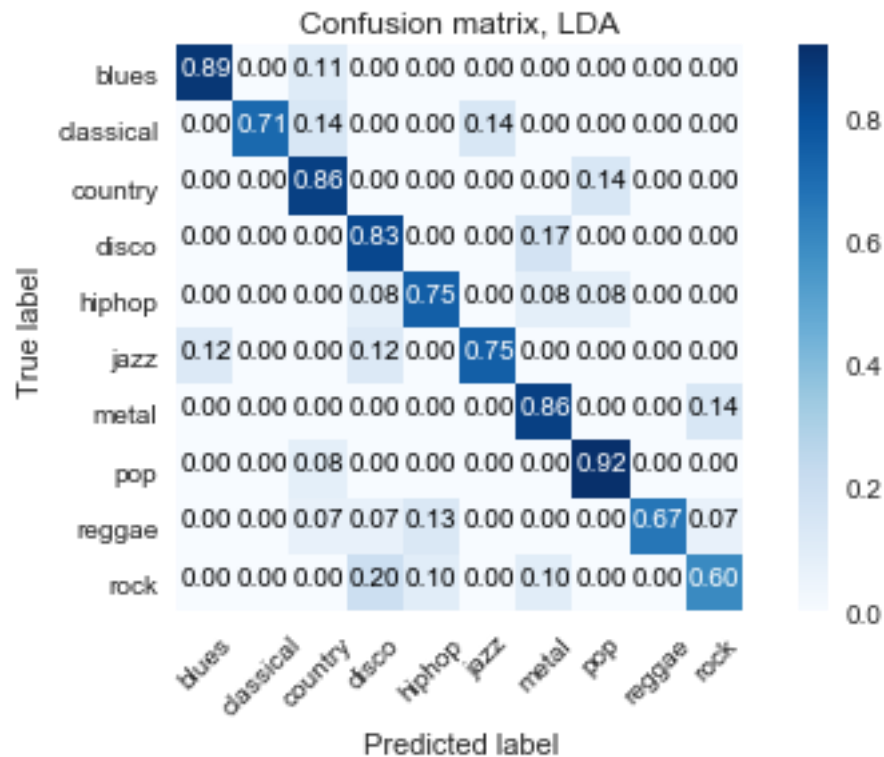
$$comp(y_i, \hat{y}_i) = \begin{cases} 1, & \text{если } y_i = \hat{y}_i \\ 0, & \text{если } y_i \neq \hat{y}_i \end{cases}$$

$$accuracy_score = \frac{\sum_{i=1}^N comp(y_i, \hat{y}_i)}{N}$$

Результаты кросс-валидации на оптимальных параметрах

accuracy_score при кросс-валидации:

LDA: 0.736, QDA: 0.70, PCA+SVC: 0.76



Результаты работы алгоритмов при выделении train (900) и test (100) из датасета

Тест на вновь собранных данных

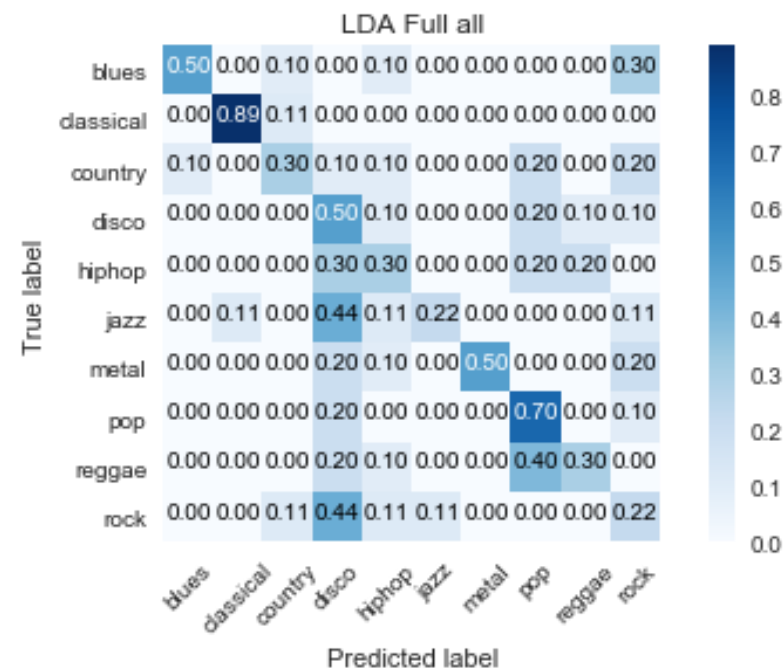
В качестве финальной модели был выбран LDA, поскольку при практически одинаковом качестве классификации, LDA позволяет предсказывать **список вероятностей** отдельных классов

Был подготовлен дополнительный датасет (по 10 каждого жанра) песен

Accuracy: 0.443

Необходимый жанр среди

3х с максимальной вероятностью: **0.670**



Спасибо за внимание!

Пример представления



"Уверенность" от позиции истинного ответа в
отсортированном списке вер-тей

