

Алгоритм распознавания эмоций на основе линейной регрессии

Е. С. Ковалёва¹

В работе предлагается метод распознавания эмоций на изображениях лиц при помощи линейной регрессии с использованием двух нейронных сетей (MediaPipe[1] и Dlib[2]), которые располагают ключевые точки на изображениях лиц людей. В данной статье будут показаны: создание базы данных для обучающей выборки, признаки, которые будут использоваться для обучения классификатора, построение линейного классификатора и результат его работы на тренировочной выборке. Целью работы является выявление ключевых признаков по ключевым точкам, по которым можно распознать ту или иную эмоцию. Предлагаемые подходы могут оказаться полезными, в частности, при обучении многослойных нейронных сетей, обучение которых повысит качество распознавания эмоций на изображениях.

Ключевые слова: ключевая точка, MediaPipe, Dlib, линейный классификатор.

1. Введение

Основной целью данной работы является разработка линейного классификатора для распознавания эмоций по фотографии человека.

Для её достижения были проделаны следующие этапы:

- создание базы данных для обучающей выборки из видео данных;
- применение нейронных сетей MediaPipe [1] и Dlib [2] для извлечения координат ключевых точек;
- создание векторов признаков, состоящих из расстояний и углов между ключевыми точками и характеризующих эмоции наилучшим образом;
- создание линейных классификаторов для отнесения каждого изображения к той или иной эмоции;
- обучение линейных классификаторов и поиск весовых коэффициентов;

¹ Ковалёва Елена Сергеевна — аспирантка каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: e.kovaleva.msu@yandex.ru.

Kovalyova Elena Sergeevna — graduate student, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intellectual Systems.

- тестирование классификаторов на тренировочной выборке;
- тестирование алгоритма на локальную и общую точность распознавания.

Последующей целью является выявление слабых и сильных сторон разработанного алгоритма и построение нейронной сети по найденным в работе векторам признаков.

2. Создание базы данных

Для создания обучающей выборки необходимо создать базу изображений лиц людей с различными эмоциями. В данной работе речь пойдет о распознавании семи эмоций: радость, грусть, удивление, гнев, страх, отвращение, презрение. Поскольку для выявления признаков впоследствии будут использоваться нейронные сети MediaPipe и Dlib, то для того, чтобы эти нейронные сети могли благополучно извлечь координаты ключевых точек, необходимы изображения с разрешением большим, чем 480×480 пикселей. Поэтому была поставлена цель создать базу данных с изображениями высокого качества. Так как такой базы данных нет в открытом доступе, появилась необходимость создавать её вручную.

Есть два способа, которыми можно было решить данную проблему:

- Скачивать отдельные изображения.
- Скачивать видео, которые позже можно будет нарезать на отдельные кадры, которые будут представлять собой изображения с лицами людей.

Было принято решение создать базу данных вторым способом, поскольку это оптимальнее по времени. Видео были собраны на ресурсе YouTube, так как данный сайт обладает не только безграничным количеством видеoinформации, но и качественной поисковой системой, позволяющей быстрее найти нужные для работы видео.

Видео были отобраны с учетом следующих параметров:

- Видео должно иметь качество 4k.
- Видео должно быть с участием людей.
- Лица людей на видео должны быть четкими и сняты крупным планом.
- В течение минимум минуты видео должно быть непрерывно по содержанию и повороту камеры.

- Видео должны содержать лица людей, выражающих все семь эмоций, которые необходимо будет впоследствии распознавать.

Количество изображений (кадров) из загруженных видео [3] для каждой эмоции было отрегулировано вручную. Для загрузки видео в высоком качестве был использован youtube-dl [4]. Итоговая база данных составила 523747 изображений.

3. Создание векторов признаков

3.1. Извлечение ключевых точек

Для создания вектора признаков для каждого изображения, применим к каждому изображению две нейронные сети MediaPipe и Dlib для извлечения координат ключевых точек. Эти две нейронные сети применяются для данной цели по причине большого числа преимуществ каждой из них.

MediaPipe содержит 468 ключевых точек и является самой подробной по количеству точек нейронной сетью. Эта нейронная сеть очень подробно описывает область носа, щёк и глаз. Пример работы данной нейронной сети изображен на рисунке 1, а маска с ключевыми точками - на рисунке 2.

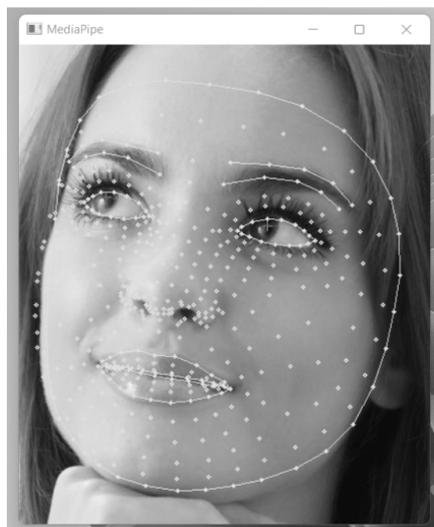


Рис. 1. Пример работы MediaPipe.

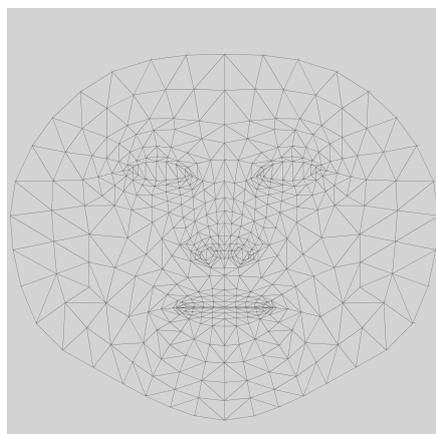


Рис. 2. Ключевые точки MediaPipe.

Преимущества Dlib в сравнении с MediaPipe заключаются в точности расположения ключевых точек на лицах. Визуальный анализ показал, что Dlib очень точно располагает на изображениях ключевые точки глаз, бровей, носа и рта. Однако, в отличие от MediaPipe он не умеет располагать ключевые точки на щеках, в области носа и в области лба. Именно по этой причине (по причине наличия преимуществ в каждой из этих двух нейронных сетей) было принято решение извлекать координаты ключевых точек, используя и MediaPipe и Dlib.

Пример работы Dlib на изображении показан на рисунке 3. На рисунке 4 продемонстрированы непосредственно сами ключевые точки.



Рис. 3. Пример работы Dlib.

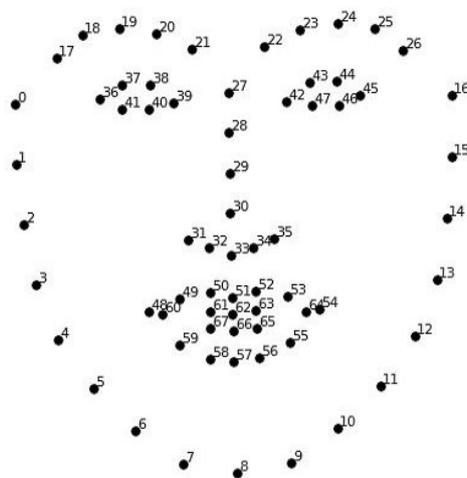


Рис. 4. Ключевые точки Dlib.

3.2. Вектор признаков

Визуальный анализ результатов работы MediaPipe и Dlib на изображениях показал, что положение точек в области глаз, рта, переносицы и щек влияют на показываемую человеком эмоцию. Однако для более точного

вывода необходимо понять, какие именно точки, расстояния и углы между ними влияют на изображение той или иной эмоции. Для этой цели необходимо создать векторы признаков, которые впоследствии будучи приняты в классификатор покажут насколько то или иное расстояние или тот или иной угол между ключевыми точками влияют на показанную на изображении эмоцию.

Для начала условно обозначим все точки от 0 до 467 из MediaPipe как M_0, M_1, \dots, M_{467} соответственно, а все точки от 0 до 67 из Dlib как D_0, D_1, \dots, D_{67} соответственно. Для входных значений классификатора понадобятся признаки в виде нормированных расстояний и углов (в радианах) между ключевыми точками и отношений этих углов друг к другу, изображенных на рисунке 5. Для того, чтобы нормировать расстояния, их необходимо разделить на длину отрезка $M_{127}M_{356}$. Нормировать расстояния необходимо, чтобы данные исследования не зависели от размера изображения и размера лица на изображении.

Таким образом, входная база данных для каждого изображения i состоит из вектора признаков x_i (включающего в себя нормированные расстояния отрезков и размер углов в радианах, изображенных на рисунке 5) и вектора вероятностей y_i ($y_{ij} = 0$, если на изображении i нет эмоции j , и $y_{ij} = 1$ в противном случае).

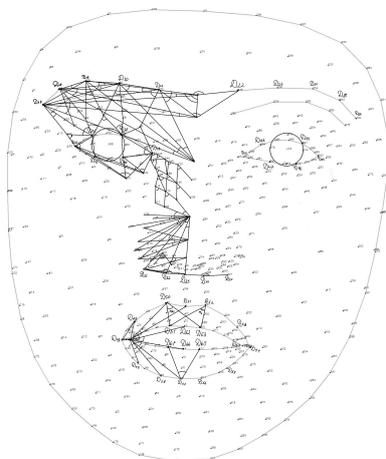


Рис. 5. Расстояния и углы для вектора признаков.

4. Линейная регрессия для распознавания эмоций

4.1. Математическая формулировка задачи

Для идентификации одной из семи эмоций необходимо построить семь классификаторов, определяющих вероятность изображения эмоций на фото, следующим образом:

- 1) Разделяем случайным образом базу данных изображений (из параграфа 2) на тестовую и тренировочную выборки по следующим правилам:
 - Размер тренировочной выборки $m_{train} = 366625$ (по 52375 на каждую эмоцию).
 - Размер тестовой выборки, соответственно, $m_{test} = 157122$ (по 22446 на каждую эмоцию).
- 2) Для каждого из m изображений:
 - Извлекаем признаки x_{ij} (согласно алгоритму, изложенному в параграфе 3.2), где i - номер признака, j - номер изображения. Строим матрицу характеристик $X = \{x_{ij}\}_{m_{train} \times n+1}$ с первым столбцом из единиц и с величинами признаков во всех остальных столбцах.
 - Строим вектор вероятностей y_j для каждого изображения j размера 7 (по количеству эмоций) таким образом, чтобы $y_{jt} = 1$, если на изображении есть эмоция под номером t , и $y_{jt} = 0$ в противоположном случае.
Пример: если на изображении вторая эмоция, то вектор будет выглядеть следующим образом: $y_j = (0, 1, 0, 0, 0, 0, 0)$.
- 3) Для каждого изображения i зададим уравнения следующего вида:
$$y_{it} = w_{0t} + w_{1t}x_{1t} + w_{2t}x_{2t} + \dots + w_{nt}x_{nt} + e_{it}$$
, где
 $w_{0t}, w_{1t}, \dots, w_{nt}$ - веса признаков, которые необходимо найти,
 e_{it} - погрешность модели,
 y_{it} - вероятность того, что на фото изображена эмоция под номером t ,
 t - номер эмоции.
- 4) Таким образом наша задача сводится к поиску решения следующих СЛАУ для каждой из 7 эмоций:
$$X_{m_{train} \times n} w_{n \times 1} + e_{m_{train} \times 1} = y_{m_{train} \times 1}$$
,
где w - вектор весов, e - вектор погрешностей, y - вектор вероятностей.

4.2. Поиск вектора весов методом наименьших квадратов

Построим решение для поиска вектора весов в классификаторе эмоции с номером t в тренировочной выборке. Для этой цели:

- 1) Строим СЛАУ $X_{m_{train} \times n} w_t + e_t = y_t$ для векторов y_{it} , где i - номер изображения.
- 2) Для решения данной СЛАУ применяем метод наименьших квадратов [5] и получаем решение для вектора весов:
$$w_t = (X^T X)^{-1} X^T y_t. \quad (1.1)$$
- 3) Решаем систему (1.1) для $t = 0, 1, \dots, 6$.
- 4) Находим значения весов признаков w_{it} для $i = 0, 1, \dots, m_{train} - 1$ и $t = 0, 1, \dots, 6$.

4.3. Применение классификатора на тестовой выборке

Теперь, когда веса признаков для каждой эмоции известны, можем провести эксперимент по работе классификатора на тренировочной выборке. Для этого на вход каждого классификатора подаются следующие данные:

- Матрица признаков с единичным первым столбцом:
 $X = \{x_{ij}\}_{m_{test} \times n+1}$, где $i = 0, 1, \dots, m_{test} - 1$ (номер изображения), $j = 0, 1, \dots, n$ (номер признака).
- Значения весов для весового вектора: w_{it} , где $i = 0, 1, \dots, m_{test} - 1$ (номер изображения), а $t = 0, 1, \dots, 6$ (номер классификатора).

Затем для каждого классификатора вычисляется произведение матрицы признаков и его вектора весов:

$f_t = X w_t$ - вектор вероятностей, показывающий, с какой вероятностью на фото изображена эмоция под номером t .

Например, если величина $f_{it} = 0.75$, то это означает, что на фото с номером i с вероятностью 0.75 изображена эмоция под номером t .

Таким образом, применив все 7 классификаторов ко всем изображениям тестовой выборки, на выходе алгоритма получаем матрицу вероятностей $F = \{f_{ij}\}_{m_{test} \times 7}$, где $i = 0, 1, \dots, m_{test} - 1$ (номер изображения), $t = 0, 1, \dots, 6$ (номер классификатора). После этого для каждого изображения $i = 0, 1, \dots, m_{test} - 1$ вычисляем:

$k_i = \arg t(\max_{t \in [0,6]}(f_{it}))$ - номер эмоции, которая изображена на фото с номером i с наибольшей вероятностью;

$p_i = \max_{t \in [0,6]}(f_{it})$ вероятность того, что на фото с номером i изображена эмоция k_i .

Таким образом вычисляется эмоция, показанная на изображении.

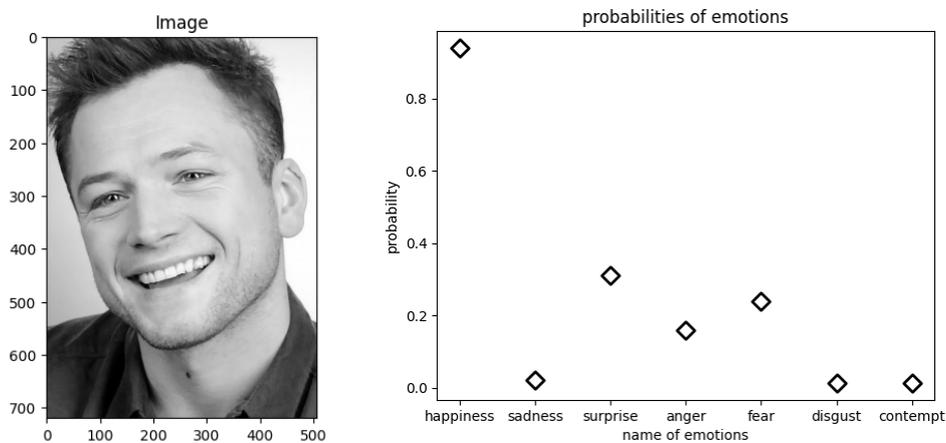


Рис. 6. Пример работы алгоритма на одном изображении. График вероятностей.

5. Примеры работы алгоритма и результаты

На рисунке 6 показан пример работы алгоритма на одном изображении из тестовой выборки. График отображает вычисленную каждым из семи классификаторов вероятность изображения на фотографии эмоции.

На данном примере можно заметить, что первый классификатор, отвечающий за вероятность эмоции радости дал наибольший результат среди всех классификаторов, что является правильным ответом.

На рисунке 7 приведен пример работы первого классификатора, отвечающего за вероятности эмоции радости, примененного ко всем изображениям тестовой выборки. На оси x расположены номера изображений. На оси y расположены вероятность идентификации эмоции радости на изображениях.

Вертикальные полосы на графике отделяют изображения, относящиеся к одной группе. Например, во втором столбце на оси x расположены номера изображений с эмоцией грусти.

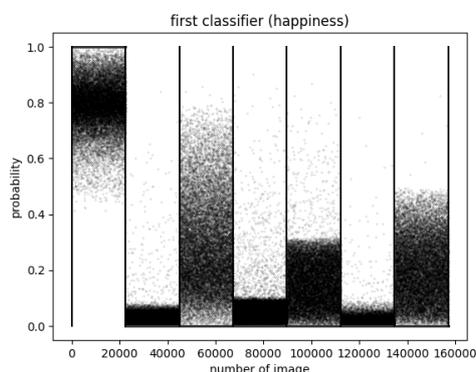


Рис. 7. Результат работы первого классификатора на тестовой выборке.

На рисунке 8 приведена таблица с долями изображений, в которых алгоритм семи классификаторов успешно и правильно распознал эмоции.

Emotion	Probability
happiness	0.87
sadness	0.93
surprise	0.86
anger	0.91
fear	0.94
disgust	0.98
contempt	0.95

Рис. 8. Доли правильно распознанных эмоций в тестовой выборке.

Общая доля ошибок составила во всей тестовой выборке составила 8%.

По результатам вероятностей успешного распознавания, приведенного в таблице на рисунке 8, можно заметить, что хуже всего распознаются эмоции радости и удивления, а эмоции отвращения и грусти распознаются лучше всего.

Данная аномалия связана с тем, что у некоторых пар эмоций существует большее количество общих признаков, чем у других пар эмоций. Например, на одном изображении могут быть одновременно эмоция удивления и эмоция радости. Также на одном изображении одновременно могут быть эмоции удивления и страха. В некоторых случаях трудно отличить

эмоцию презрения и эмоцию радости из-за наличия улыбки в обоих случаях. По этой причине лучше всего распознаются эмоции с более уникальным набором признаков, чем у других.

6. Выводы

В данной работе были проделаны следующие этапы:

- Создана база видеоданных.
- На её основе создана база данных с 523747 изображениями.
- Создана база входных данных с признаками на основе двух нейронных сетей (MediaPipe и Dlib).
- Математически поставлена задача распознавания эмоций с помощью классификаторов на основе линейной регрессии.
- Разработан алгоритм распознавания эмоций на основе линейной регрессии и метода наименьших квадратов.
- Алгоритм был обучен на тренировочной выборке изображений.
- Алгоритм был протестирован на тестовой выборке, включающей в себя 157122 изображения.
- Были вычислены вероятности правильных ответов и общая доля ошибки алгоритма.
- Был вычислен успех работы алгоритма, который составил 92%.

В дальнейшем планируется построение многослойной нейронной сети для достижения более высокого результата правильно распознанных изображений.

Список литературы

- [1] Руководство по использованию MediaPipe,
https://github.com/googlesamples/mediapipe/tree/main/examples/face_landmarker/python.
- [2] Руководство по использованию Dlib,
<http://dlib.net/python/index.html>.

- [3] Плейлист с видео для обучающей выборки, <https://www.youtube.com/playlist?list=PLbvXU2Yespkgvzwk400999CKrVV1PZYCI>.
- [4] Руководство по использованию youtube-dl, <https://github.com/ytdl-org/youtube-dl/blob/master/README.md#readme>.
- [5] Линник Ю.В., “Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений”, *Государственное издательство Физико-математической литературы*, 1958.

Emotion recognition algorithm based on linear regression Kovalyova E.S.

The paper proposes a method for recognizing emotions in facial images using linear regression using two neural networks (MediaPipe[1] and Dlib[2]), which locate key points in images of people’s faces. This article will show: creating a database for the training set, features that will be used to train the classifier, building a linear classifier and the result of its work on the training set. The goal of the work is to identify key features at key points by which one or another emotion can be recognized. The proposed approaches may be useful, in particular, when training multilayer neural networks, the training of which will improve the quality of recognition of emotions in images.

Keywords: keypoint, MediaPipe, Dlib, linear classifier.

References

- [1] MediaPipe user manual, https://github.com/googlesamples/mediapipe/tree/main/examples/face_landmarker/python.
- [2] Dlib user manual, <http://dlib.net/python/index.html>.
- [3] Playlist with videos for the training set, <https://www.youtube.com/playlist?list=PLbvXU2Yespkgvzwk400999CKrVV1PZYCI>.
- [4] Youtube-dl user manual, <https://github.com/ytdl-org/youtube-dl/blob/master/README.md#readme>.
- [5] Linnik Yu.V., “The least squares method and the fundamentals of the mathematical and statistical theory of observation processing (In Russian)”, *State Publishing House of Physical and Mathematical Literature*, 1958.