

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

КУРСОВАЯ РАБОТА
ПО ДИСЦИПЛИНЕ «МЕТОДЫ ОПТИМИЗАЦИИ»
«СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РЕШЕНИЯ
ЗАДАЧИ КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ
ПРИ ИСПОЛЬЗОВАНИИ SVM В ЗАДАЧАХ
РАСПОЗНОВАНИЯ»

Выполнили
студенты группы 3630102/80201

Деркаченко А. О.
*Классический метод машины опорных векторов, Методы формирования вектора
признаков изображения лица для определения атрибутов личности,
Сравнительный анализ*
Хрипунков Д. В.
Введение, Постановка задачи
Войнова А. Н.
Модификации метода машины опорных векторов, Сравнительный анализ

Руководитель
к. ф.-м. н., доц.

Родионова Елена Александровна

Санкт-Петербург
2021

Содержание

1	Введение	2
2	Постановка задачи	3
2.1	Общая постановка задачи распознавания	3
2.2	Постановка задачи по определению атрибутов личности по изображению лица	5
3	Классический метод машины опорных векторов	5
3.1	Краткое описание метода	5
3.2	Множественная классификация по атрибуту	7
4	Методы формирования вектора признаков изображения лица для определения атрибутов личности	7
4.1	Группа методов Φ_t	7
4.1.1	Метод локальных бинарных шаблонов (LBP)	7
4.1.2	Метод построения гистограммы направленных градиентов (HOG) и гистограмм направления края изображения (ЕОН)	9
4.2	Группа методов Φ_a	9
4.2.1	Метод активной модели формы (ASM)	9
4.2.2	Метод активной модели внешности (AAM)	9
5	Модификации машины опорных векторов	10
5.1	1-norm SVM (LASSO SVM) и Doubly Regularized SVM (ElasticNet SVM)	11
5.2	Методы с регулируемой селективностью	13
5.2.1	Метод релевантных признаков с регулируемой селективностью	13
5.2.2	Метод опорных признаков с регулируемой селективностью	14
6	Сравнительный анализ	14
6.1	Метод опорных векторов (SVM)	14
6.2	Модификации метода опорных векторов	15
6.2.1	1-norm SVM (LASSO SVM)	15
6.2.2	Doubly Regularized SVM (ElasticNet SVM)	15
6.2.3	Метод релевантных признаков с регулируемой селективностью	15
6.2.4	Метод опорных признаков с регулируемой селективностью	16
6.2.5	Практическое сравнение модификаций SVM	16
6.3	Методы формирования вектора признаков изображения	17
6.3.1	Метод локальных бинарных шаблонов (LBP)	17
6.3.2	Метод построения гистограммы направленных градиентов (HOG)	18
6.3.3	Метод построения активной модели формы (ASM) и активной модели внешности (AAM)	18
7	Краткий вывод	18
8	Источники	20

1 Введение

Увеличение объёмов информации в современном мире привело к тому, что ручная обработка такой информации стала невозможной и возникла необходимость в создании систем и алгоритмов, которые автоматизируют эту работу. Одной из задач обработки информации оказалась задача распознавания - определение признаков, которые отличают один набор данных от других.

Обработка неструктурированных данных (фото, видео, аудио) является достаточно сложной, и самым популярным способом решения такой задачи является машинное обучение, которое пытается воссоздать процесс человеческого обучения на компьютере - группировку объектов в классы по некоторым признакам. Выделяются два вида машинного обучения: "с учителем" и "без учителя" [1, 19]:

- "Обучение с учителем" предполагает наличие некой исходной выборки, заведомо разделённой на классы учителем, а системе предлагается обнаружить общие признаки, которые будут описывать класс. В дальнейшем, система сможет на основе этих признаков распределять неразмеченные данные по классам. Такое обучение решает задачу классификации, когда количество классов заранее известно, и от системы требуется лишь отнести данные к одному из них [1, 19].
- "Обучение без учителя" может предполагать неизвестное число классов, а входные данные изначально не являются размеченными. Система должна сама определить правила, которые различают предложенные объекты, и на их основе создать классы. Классы могут быть как уже известные, так и созданные новые в процессе распознавания. Такое обучение уже решает задачу распознавания - формирование правила, которое разделяет объекты разных классов [1, 19].

Самым распространённым видом информации в современном мире является фотография, пользователи социальных сетей активно делятся ими, а качество систем фото- и видеонаблюдения неуклонно растёт. Такое широкое распространение фотографии неуклонно приводит к тому, что задача распознавания ставится и в этом поле - распознавание информации с фотоснимков. Распознавание снимков чаще всего ставит цель в распознавании объектов, будь то буквы, цифры, дома, животные или люди. В этой работе мы рассмотрим частный случай - задачу распознавания лиц со снимков.

Тем не менее, задача распознавания может быть поставлена и в других сферах. Например, также крайне популярным и развивающимся полем является распознавание в речи, которое ставит сразу множество задач, таких как преобразование речи в текст, синтез речи из текста и определение дикторов и относящихся к ним фраз.

Распознавание объектов на изображениях можно разбить на ряд подзадач [1, 20]:

- Сопоставление
- Поиск
- Восстановление

- Классификация

На примере распознавания лиц эти пункты можно описать как:

- Анализ набора изображений для определения принадлежности к одному и тому же классу
- Поиск на изображении фрагмента для распознавания
- Восстановление пропущенных фрагментов по контексту
- Определение класса, к которому относится изображение

Для того чтобы распознавать человеческие лица, выделим три основных атрибута: возраст, расу и пол. Практически такое разделение может быть применено во многих сферах: поисковая выдача, оценка аудитории, реклама, обучение, возрастные и половые ограничения и многие другие [1, 22]. Существуют также и другие признаки, по которым возможно разделение личностей, но в данной работе будут рассматриваться указанные выше атрибуты.

Задача распознавания по этим признакам интересна ещё и тем, что каждый из них имеет разные категории: числовую, бинарную и множественную нечисловую. А задача определения возраста усложняется ещё и тем, что признаки старения у разных людей проявляются по-разному [1, 25].

2 Постановка задачи

2.1 Общая постановка задачи распознавания

Рассмотрим общую постановку задачи на примере распознавания символов. Предположим, система распознавания получила на вход некоторый символ (паттерн) X , который нужно распознать.

Система может считывать скорость изменения закрашенной поверхности как функцию от времени $X(t)$, называемую представлением символа X . Альтернативным вариантом можно считать сигнал в дискретные моменты времени, в результате чего получается вектор \bar{X} . Также возможны переходы из представления в виде функции в векторное [2, 11].

Предположим, что есть некоторое множество непересекающихся классов $\Omega = \{\bar{\omega}_1, \dots, \bar{\omega}_m\}$, где каждое ω_i отвечает некоторому символу. Системе распознавания нужно отнести входящий символ X к какому-то из классов ω_i . Для этого предпринимаются следующие шаги [2, 11]:

1. Символ X считывается в представление $X(t)$
2. Представление $X(t)$ преобразуется в векторную форму \bar{X}
3. Из вектора \bar{X} извлекаются информативные признаки и образуется вектор \hat{X}
4. Классификатор определяет, к какому классу относятся признаки \hat{X}

Для финального соотнесения к некоторому классу используется классификатор - набор правил для определения класса. После классификации символ может быть определён к одному из существующих классов или как неотносящийся ни к одному из них. Классификация может выполняться, например, посредством вычисления расстояния между классом и вектором признаков \bar{X} [2, 11].

Также в схеме распознавания может присутствовать блок обучения. Он выбирает учебные образы, которые заведомо распределены по классам. С их помощью можно сформировать правила классификации или определить наиболее информативные признаки [2, 11].

Декомпозируя задачу распознавания на подзадачи, получается следующий набор [2, 12]:

1. *Математически описать образ*

Такое описание удобнее всего проводить в векторной форме. Образцу X сопоставляется некоторый вектор $\bar{X} = (x_1, x_2, \dots)^T$, где каждое x_i - некоторый признак, а \bar{X} - элемент конечномерного метрического векторного пространства X

2. *Выбрать информативные признаки*

Не все признаки символа могут быть одинаково полезны при распознавании. Задача состоит в том, чтобы определить минимально необходимый набор признаков, достаточных для распознавания символа. Этот набор система должна определить сама.

3. *Описать классы*

Необходимо задать границы классов. Это может быть сделано на этапе разработки или самой системой в ходе её работы.

4. *Определить методы классификации*

Нужно определить методы, по которым образы будут соотноситься некоторым классам.

5. *Определить оценку достоверности распознавания*

Оценка нужна для того, чтобы иметь возможность определить величину потерь при неправильной классификации.

Математически задачу распознавания можно поставить так:

Дано множество образов U , отдельный образ обозначим $x \in U$. Из (возможно, несчётного) множества признаков образов x нужно выбрать конечное подмножество - пространство признаков. Пространство признаков конечномерное, линейное или метрическое, обозначим как X . Каждому образцу соответствует элемент $\bar{x} \in U$ и оператор $P : U \rightarrow X$ отображения образа в пространство признаков [2, 15].

Введём конечное множество классов $\Omega = \{\bar{\omega}_1, \dots, \bar{\omega}_m\}$, для которого верно, что $\cup_{i=1}^m \omega_i = U$ и $\omega_i \cap \omega_j = \emptyset, \forall i \neq j$. Для классификации образа $x \in U$ по классам из Ω нужно найти индикаторную функцию $g : U \rightarrow Y, Y = \{y_1, \dots, y_m\}$, где Y - множество меток класса. То есть $g(x) = y_i$, если $x \in \omega_i$. Поскольку в реальности мы работаем не с самими образами, а их признаками, то нужно найти решающую функцию $\tilde{g} : X \rightarrow Y$ для $\bar{x} = P * x \in X$, то есть $\tilde{g}(\bar{x}) = y_i$, если $\bar{x} = P * x \in \omega_i$ [2, 15].

Поскольку множество $P^{-1}\bar{x}$, $\bar{x} \in X$ может иметь непустые пересечения с разными классами ω_i , то функция $\tilde{g}(x)$ будет неоднозначной, тогда из неё нужно выделить однозначную ветвь, удовлетворяющую условиям оптимальности, например, минимальность ошибки неправильной классификации. На этапе обучения система известна некоторыми парами (\bar{x}_j, y_j) , называемые прецедентами и множество $H = \{\bar{x}_1, \dots, \bar{x}_N\}$, называемое обучающей выборкой. По множеству прецедентов (H, Y) нужно построить решающую функцию $\tilde{g}(x)$, которая будет осуществлять классификацию [2, 16].

2.2 Постановка задачи по определению атрибутов личности по изображению лица

Задача определения атрибутов личности по изображению лица - частный случай задачи обучения по прецедентам, описанной выше, где множество U - множество нормализованных изображений [1, 28].

Для каждого из выбранных нами атрибутов будет выбрана соответствующая задача классификации в зависимости от множества меток классов Y [1, 31]:

1. Для атрибута "пол" будет бинарная классификация $Y = \{-1, 1\}$
2. Для атрибута "раса" множественная классификация $Y = \{-1, 0, 1\}$ для рас "европеоидная", "монголоидная" и "негроидная"
3. Для атрибута "возраст" восстановление регрессии $Y = [5, 100]$

Эти задачи связаны друг с другом, так, например, задача множественной классификации может быть разложена на несколько задач бинарной классификации, а задача восстановления регрессии формулируется исходя из решения задачи множественной классификации [1, 31].

Задача определения атрибутов личности по изображению лица сводится к тому, что нужно определить метод формирования вектора признаков $P : U \rightarrow X$ и решающую функцию $\tilde{g} : X \rightarrow Y$ так, чтобы оценка достоверности распознавания была максимальной. Решение такой задачи получается сравнением решений из конечного множества частных оптимизационных задач, в которых выбраны разные способы формирования вектора признаков и решающей функции [1, 32].

3 Классический метод машины опорных векторов

3.1 Краткое описание метода

Метод опорных векторов (SVM) разработан в 60-е годы коллективом советских математиков под руководством В.Н.Валника и рассчитан на классифицирование объектов по двум классам.

Пусть имеется обучающая выборка $G^l, |G^l| = n$, заданная множеством пар прецедентов $(\bar{x}_i, y_i), i = \overline{1, n}, \bar{x}_i \in \mathbb{R}^m, y_i \in \{-1, +1\}$. Множество F_{SVM} , из которого выбираются решающие функции по методу опорных векторов, образовано функциями вида:

$$f(\bar{x}) = \text{sign}(\langle \bar{w}, \bar{x} \rangle + w_0), \quad (1)$$

где \langle, \rangle - скалярное произведение векторов, \bar{w} - ортонормированный вектор к разделяющей гиперплоскости, w_0 - вспомогательный параметр (сдвиг гиперплоскости).

Так как любая гиперплоскость может быть задана в виде $\langle \bar{w}, \bar{x} \rangle + w_0 = 0$, то объекты с $f(\bar{x}) \leq -1$ попадут в один класс, а объекты с $f(\bar{x}) \geq +1$ - в другой.

Базовая идея метода: найти такие \bar{w}, w_0 , которые максимизируют расстояние между классами, что приводит к более уверенной классификации объектов.

То есть условие $-1 < \langle \bar{w}, \bar{x} \rangle + w_0 < +1$ задает полосу, разделяющую классы. При этом ни одна точка из множества X^l не должна лежать внутри полосы, а границами полосы являются две параллельные гиперплоскости, проходящие через точки (объекты), ближайšie к разделяющей гиперплоскости, которая находится по середине данной полосы. И объекты, через которые проходят границы полосы, называются *опорными векторами*.

Проблема нахождения максимума расстояния сводится к нахождению минимума $\|\bar{w}\|^2$, которая является стандартной задачей квадратичного программирования:

$$\begin{cases} \|\bar{w}\|^2 \rightarrow \min \\ y_i \langle \bar{w}, \bar{x}_i \rangle + w_0 \geq 1 \end{cases} \quad (2)$$

и решается методом множителей Лагранжа.

Задача квадратичного программирования, содержащая только двойственные переменные метода множителей Лагранжа λ_i , имеет вид:

$$\begin{cases} -\sum_{i=1}^n \lambda_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \langle \bar{x}_i, \bar{x}_j \rangle \rightarrow \min \\ \sum_{i=1}^n \lambda_i y_i = 0 \\ \lambda_i \geq 0 \end{cases} \quad (3)$$

В результате решающая функция приобретает вид:

$$f(\bar{x}) = \text{sign}\left(\sum_{i=1}^n \lambda_i y_i \langle \bar{x}_i, \bar{x} \rangle + w_0\right), \quad (4)$$

где параметр $w_0 = \text{med}\{\langle \bar{w}, \bar{x}_i \rangle - y_i\}$, $\lambda_i \neq 0$.

Вышеуказанные рассуждения справедливы для линейно разделимой обучающей выборки. Но на практике встречаются случаи линейной неразделимости и решающей функции позволяют допускать ошибки на обучающей выборке, но эти ошибки минимизируют и используют управляющую константу C как компромисс между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки. Тогда вводят ограничение сверху $0 \leq \lambda_i \leq C$ и такой алгоритм называют SVM с "мягким зазором" (soft-margin SVM), иначе имеется "жесткий" зазор.

Если признаки x_i заданы в виде функции $\theta(x_i)$, то решающая функция строится аналогично. Тогда функция $K(u, v) = \langle \theta(u), \theta(v) \rangle$ - *ядро*, если она симметрична и положительно определена. Для решения практических задач классификации изображений по атрибуту "пол" используют RFB-ядро вида:

$$K(\bar{x}_i, \bar{x}) = \exp(-\gamma \|\bar{x}_i - \bar{x}\|^2), \quad (5)$$

вычисляющее оценку близости вектора \bar{x} к опорному вектору \bar{x}_i , где γ - некоторый параметр [1, 57-63].

3.2 Множественная классификация по атрибуту

Классификация по атрибуту "раса" является типичной задачей множественной классификации. Для решения такой задачи используется подход *"один против всех"*, реализующий сведение задачи множественной классификации к последовательному применению бинарных классификаторов. В рамках данного подхода строится бинарное дерево решающих функций $f \in F$, каждая из которых выделяет только один класс объектов [1, 68-69].

В случае классификации по атрибуту "раса" на первом шаге объекты разделяются решающей функцией f_1 на два класса: "европеоиды" и "все остальные". Если объект не попал в класс "европеоиды" то на втором шаге другая решающая функция f_2 производит разделение на класс "монголоиды" и "все остальные" ("негроиды").

Данный подход позволяет использовать большую часть разработок в области бинарной классификации для решения задач множественной классификации.

4 Методы формирования вектора признаков изображения лица для определения атрибутов личности

Различают две группы методов Φ формирования вектора признаков X' [1, 34]:

- группа методов Φ_t , основанная на использовании текстуры (значений интенсивности пикселей) изображения
- группа методов Φ_a , основанная на выделении антропометрических точек на изображении лица человека с последующим выделением информации о расстоянии между этими точками и их взаимном расположении

4.1 Группа методов Φ_t

К этой группе относятся методы, в которых используются либо непосредственно значения интенсивности пикселей нормализованных изображений, либо результаты применения некоторых операторов к значениям интенсивностей. При этом прямое использование пикселей изображения подходит только для изображений малой размерности и на практике используется для выделения лиц на общем изображении.

Вектор признаков в таком случае представляет собой гистограмму значений интенсивностей или их модификаций.

4.1.1 Метод локальных бинарных шаблонов (LBP)

Локальный бинарный шаблон - определенный вид признака, представляющий собой описание окрестности пикселя изображения в двоичном представлении.

Базовый оператор LBP, применяемый к пикселю изображения, использует восемь пикселей окрестности и принимает значение интенсивности центрального пикселя в качестве порога. Пиксели, имеющие значение, большее или равное указанному, принимают значение "1" остальные - "0". Таким образом, результатом применения базового оператора LBP к пикселю изображения является восьмиразрядный бинарный

код, описывающий окрестность пикселя по часовой стрелке, начиная с верхнего левого пикселя окрестности. Далее этот код рассматривается, как двоичная запись некоторого числа, сопоставленная данному пикселю.

После получения указанных значений строится их гистограмма, то есть формируется набор $\Phi_{LBP}(i) = \bar{\mathbf{x}}_i = (\bar{0}_i, \bar{1}_i, \dots, \bar{x}_i, \dots, \bar{255}_i)$ частот \bar{x}_i появления бинарных шаблонов, имеющих десятичные числовые эквиваленты $x, x \in \{0, 1, \dots, 255\}$ в изображении $i : \bar{x}_i \in \mathbb{R}, 0 \leq \bar{x}_i < 1$. При этом коллекция I' из m элементов обучающей и k элементов тестовой выборок представляются матрицами размером $m \times 256$ и $k \times 256$:

$$\Phi_{LBP}(I^l) = X^l = \begin{pmatrix} \bar{0}_1 & \bar{1}_1 & \dots & \bar{x}_1 & \dots & \bar{255}_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_i & \bar{1}_i & \dots & \bar{x}_i & \dots & \bar{255}_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_m & \bar{1}_m & \dots & \bar{x}_m & \dots & \bar{255}_m \end{pmatrix} \quad (6)$$

$$\Phi_{LBP}(I^t) = X^t = \begin{pmatrix} \bar{0}_1 & \bar{1}_1 & \dots & \bar{x}_1 & \dots & \bar{255}_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_i & \bar{1}_i & \dots & \bar{x}_i & \dots & \bar{255}_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{0}_k & \bar{1}_k & \dots & \bar{x}_k & \dots & \bar{255}_k \end{pmatrix} \quad (7)$$

Соответствующими векторами заменяются элементы i в описаниях прецедентов (i, y_i) в обучающей \tilde{G}^l и тестовой \tilde{G}^t выборках.

Локальные бинарные шаблоны характеризуют локальные особенности изображения, а частоты этих особенностей можно рассматривать как обобщенную модель лица. При этом шаблоны не характеризуют расположение данных особенностей на изображении. Для этих целей нормализованное изображение разбивается на s регионов, для каждого из которых вычисляется своя гистограмма $\Phi_{LBP}(i_j)$. Тогда результирующим описанием изображения i является вектор значений \bar{x}_i , сформированный как конкатенация s гистограмм, полученных по s регионам исходного изображения:

$$\Phi_{LBP}(i) = \bar{\mathbf{x}}_i = (\bar{0}_{i_1}, \bar{1}_{i_1}, \dots, \bar{x}_{i_1}, \dots, \bar{255}_{i_1}, \dots, \bar{0}_{i_j}, \bar{1}_{i_j}, \dots, \bar{x}_{i_j}, \dots, \bar{255}_{i_j}, \dots, \bar{0}_{i_k}, \bar{1}_{i_k}, \dots, \bar{x}_{i_s}, \dots, \bar{255}_{i_s}) \quad (8)$$

Данное описание уже не является гистограммой, так как сумма значений его элементов равна s .

Исходя из соотношений, характерных для элементов лица человека, нормализованное изображение традиционно разбивается на $k = 6 \times 7 = 42$ региона. При более крупном разбиении теряется информация о локальном расположении особенностей, а при более мелком - увеличивается размерность вектора признаков, что усложняет процесс классификации.

Не все шаблоны обладают одинаковой информативностью. Поэтому выделяют *равномерные шаблоны* - бинарные комбинации, содержащие не более трех серий "0" и "1". Равномерные LBP определяют только важные локальные особенности изображения: концы линий, грани, углы и пятна. Использование только равномерных шаблонов приводит к существенному сокращению размерности вектора признаков, в этом случае используется только $p(p-1) + 2$ шаблона [1, 34-39].

4.1.2 Метод построения гистограммы направленных градиентов (HOG) и гистограмм направления края изображения (ЕОН)

Метод построения гистограммы направленных градиентов (HOG) основан на вычислении градиента изменения интенсивности для каждого пикселя изображения и формировании гистограммы данных градиентов для различных участков изображения. При этом выделяются восемь базовых направлений: "север", "северо-восток" и т.д. В итоге формируется пространство признаков, близкое по своим характеристикам к методу LBP [1, 39].

Данный метод имеет свой аналог в виде метода формирования гистограмм направления края изображения ЕОН. Эти методы нашли основное применение при решении задач выделения заданного объекта на изображении, а не при определении атрибутов личности.

4.2 Группа методов Φ_a

Методы данной группы основаны на воспроизведении процесса описания лица, характерном для человека: фиксируются особые точки и их расположение на лице (границы глаз, носа, рта, бровей, подбородка и т.д.)

4.2.1 Метод активной модели формы (ASM)

Главной предпосылкой для построения ASM является наблюдение, что между расположением антропометрических точек есть зависимости. Для моделирования этих зависимостей по обучающей выборке \tilde{G}^l строится статическая модель положения важных антропометрических точек на изображении лица. При этом в качестве признакового описания изображения используются координаты выделенных точек на нормализованных изображениях, то есть формируется набор $\Phi_{ASM}(i) = \bar{\mathbf{x}}_i = (x_1, x_2, \dots, x_u, y_1, y_2, \dots, y_u)^T$, а вся совокупность из n прецедентов обучающей выборки \tilde{G}^l образует матрицу $\Phi_{ASM}(\tilde{G}^l) = X = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_i, \dots, \bar{\mathbf{x}}_n)$.

После выделения главных компонент указанной матрицы получаем выражение для синтезированной формы:

$$\begin{aligned} X &= \bar{X} + P_{sh} B_{sh} \\ B_{sh} &= P_{sh}^T (X - \bar{X}) \end{aligned} \quad (9)$$

где \bar{X} - форма изображения, усредненная по всем реализациям обучающей выборки (базовая форма), P_{sh} - матрица собственных векторов, B_{sh} - вектор параметров формы.

Приведенное выражение означает, что форма любого изображения X может быть выражена как сумма базовой формы \bar{X} и линейной комбинации собственных форм, содержащихся в матрице P_{sh} . Меняя значение B_{sh} можно синтезировать различные формы изображений [1, 40-43].

4.2.2 Метод активной модели внешности (AAM)

Данная модель является развитием активной модели формы (ASM), путем добавления в нее модели текстуры.

На первом этапе происходит построение активной модели формы (ASM) путем разметки нормализованного изображения лица - выделения на нем некоторых пронумерованных характерных антропометрических точек. Качественное описание изображения достигается при наличии $u = 60 - 70$ таких точек.

Автоматическое определение важных антропометрических точек основано на выделении характерных особенностей изображения: Т-образные пересечения краев, углы, легко различимые биологические особенности, границы изображения и пр.

На втором этапе построения модели ААМ учитывается информация о текстуре (значениях интенсивности пикселей) изображения. Для этого изображение разбивается на непересекающиеся треугольники, вершинами которых являются выделенные антропометрические точки активной модели формы. Далее для каждого треугольника вычисляется среднее значение интенсивности, то есть изображение описывается вектором:

$$\Phi_{AAM}(i) = \bar{\mathbf{z}} = (z_1, z_2, \dots, z_v), \quad (10)$$

где v - количество треугольников в изображении. А вся совокупность из n прецедентов обучающей выборки \tilde{G}^l образует матрицу:

$$\Phi_{AAM}(\tilde{G}^l) = Z = (\bar{\mathbf{z}}_1, \bar{\mathbf{z}}_2, \dots, \bar{\mathbf{z}}_i, \dots, \bar{\mathbf{z}}_n) \quad (11)$$

Аналогично построению активной модели формы по методу главных компонент строится приближенная модель текстуры изображения:

$$\begin{aligned} Z &= \bar{Z} + P_t B_t \\ B_t &= P_t^T (Z - \bar{Z}) \end{aligned} \quad (12)$$

Как и в модели формы, текстура любого изображения Z может быть синтезирована суммой базовой текстуры \bar{Z} и линейной комбинации собственных форм, содержащихся в матрице P_t . Меняя значение B_t можно синтезировать различные текстуры изображений.

Общее признаковое описание изображения i по методу ААМ представляет собой конкатенацию векторов \bar{x}_i и \bar{z}_i .

Метод ААМ ориентирован на решение другой задачи - идентификации личности по изображению лица, так как в этом случае разметку эталонного изображения осуществляет человек, выделяя наиболее важные антропометрические точки. Построенная "вручную" модель внешности затем используется для поиска требуемой личности на других изображениях или определения выражения эмоций [1, 40-45].

5 Модификации машины опорных векторов

Метод SVM имеет следующие недостатки:

- На выходы у SVM мы имеем решение, а апостериорные вероятности, которые в некоторых задачах оказываются более значимыми получить из чистого SVM не предоставляется возможным
- SVM используется для двух классов, обобщить на несколько проблематично.

- Есть параметр C (или ν , или ещё вдобавок ϵ), подбор которых сопровождается рядом проблем
- Предсказания – линейные комбинации ядер, которым необходимо быть положительно определёнными и которые центрированы на точках из датасета.

Поэтому существует огромное количество модификаций SVM алгоритма, которые стараются решить недостатки SVM.

На основе анализа мировой литературы будем рассматривать четыре критерия для оценивания «качества» селективных свойств конкретного метода обучения [3, 5]:

1. эффективное подавление полностью шумовых попарно независимых признаков;
2. эффективное подавление полностью шумовых признаков, имеющих значительную линейную зависимость между собой;
3. эффективное выделение группы информативных линейно независимых признаков;
4. эффективное выделение группы информативных признаков, только совместное участие которых в решении может обеспечить приемлемую точность распознавания.

Рассмотрим критерий, который представляет собой обобщение классического метода опорных векторов. Критерий обучения эквивалентен оптимизационной задаче минимизации целевой функции $J(a, b, \delta|c)$ на выпуклом множестве, заданном набором линейных ограничений-неравенств для объектов обучающей совокупности [3, 11]:

$$\begin{cases} J(a, b, \delta|c) = -\ln \Psi(a) + \sum_{j=1}^N (\delta_j) \rightarrow \min(a, b, \delta) \\ y_i(\sum_{j=1}^n (a_i x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N \end{cases} \quad (13)$$

Наиболее популярными модификации метода опорных векторов, наделяющие его свойством отбора признаков являются **Lasso SVM (1-norm SVM)** и **Elastic Net SVM (Doubly Regularized SVM)**. Оба эти метода несколькими средствами отбирают подмножество информативных признаков, число которых определяется структурными параметрами.

5.1 1-norm SVM (LASSO SVM) и Doubly Regularized SVM (ElasticNet SVM)

Известные методы автоматического сокращения размерности признакового описания объектов в рамках метода опорных векторов, получаются из обобщенного критерия (13) специальным выбором априорной плотности распределения направляющего вектора $=\Psi(a)$. Метод Elastic Net SVM, как более общий из них, получается из (13) при предположении, что случайный направляющий вектор образован независимыми компонентами $a = (\alpha_1, \dots, \alpha_n)$ с одинаковыми плотностями распределения

$$\Psi(\alpha_i|\beta, \mu) = \text{Exp}[-(\beta\alpha_i^2 + \mu|\alpha_i|)]$$

которые определяются одинаковыми значениями параметров (β, μ) . Здесь нормирующая константа определяется обоими параметрами:

$$D = [2\sqrt{\frac{\pi}{\beta}} \exp(\frac{\mu^2}{4\beta}) \Phi(\frac{\mu}{\sqrt{2\beta}})] \text{ где } \Phi(u) = \frac{1}{2\pi} (\int_u^\infty \exp(-\frac{z^2}{2}) dz) - \text{это функция Лапласа}$$

В этом случае совместная априорная плотность распределения направляющего вектора определяется выражением

$$\Psi(\alpha_i | \beta, \mu) = D^n \exp(-\sum_{i=1}^n (\beta * \alpha_i^2 + \mu |\alpha_i|))$$

$$\ln \Psi(\alpha_i | \beta, \mu) = \text{const} - (\beta \sum_{i=1}^n (\alpha_i^2) + \mu \sum_{i=1}^n (|\alpha_i|))$$

и (13) примет вид, эквивалентный методу Elastic Net SVM [3, 13]:

$$\begin{cases} J_{DrSVM}(a, b, \delta | c, \beta, \mu) = \beta \sum_{i=1}^n (\alpha_i^2) + \mu \sum_{i=1}^n (|\alpha_i|) + c \sum_{j=1}^N (\delta_j) - > \min(a, b, \delta) \\ y_i (\sum_{i=1}^n (a_i x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N \end{cases}$$

В еще более частном случае, когда $\beta = 0$, получается метод обучения Lasso SVM [3, 13]:

$$\begin{cases} J_{1nSVM}(a, b, \delta | c, r) = \sum_{i=1}^n |\alpha_i| + c(r/2)^{\frac{1}{2}} \sum_{j=1}^N (\delta_j) - > \min(a, b, \delta) \\ y_i (\sum_{i=1}^n (a_i x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N \end{cases}$$

Этот критерий эквивалентен (13) при интерпретации независимых компонент случайного направляющего вектора $a = (\alpha_1, \dots, \alpha_n)$ как распределенных по закону Лапласа $\psi(\alpha_i | r) = (2r)^{-\frac{1}{2}} |\alpha_i|$.

В силу специфики принятых априорных распределений чрезвычайно важным свойством беспереборного отбора признаков непосредственно в ходе обучения. Это свойство является результатом склонности этих критериев в точке минимума присваивать нулевые значения большинству компонент направляющего вектора $\alpha_i = 0$, реализуя при этом, вообще говоря, разумный отбор подмножества «полезных» признаков.

Суждение о классическом SVM тривиально - все направления одинаковы по предпочтительности, и свойство селективности полностью отсутствует. Что же касается методов Elastic Net SVM и Lasso SVM, то оказалось, что они идентичны по ориентации предпочтительных направлений - наилучшим является направление, оставляющее единственный признак, а роль наихудшего играет направление, оценивающее все признаки как равнозначные. Это означает, что соответствующие априорные распределения направляющего вектора выражают предположение, что среди признаков есть только один признак, полезный для распознавания класса объекта. Следовательно, в ситуациях, когда в исходных данных таких признаков более одного, обучение методами Elastic Net SVM и Lasso SVM может приводить к снижению обобщающей способности полученных решающих функций.

Известно также, что Lasso SVM в группе линейно зависимых признаков полностью игнорирует факт их зависимости - отбирает лишь один такой признак, отбрасывая остальные. В то же время Elastic Net SVM стремится оставлять всю линейно зависимую группу целиком, даже если она малоинформативна.

Таким образом, необходима разработка новых способов регуляризации SVM, способных, во-первых, допускать существование независимых одинаково полезных признаков, и, во-вторых, удалять шумовые признаки без учета их линейной зависимости.

Для разрешения этой проблемной ситуации разработаны два новых класса априорных распределений направляющего вектора дискриминантной гиперплоскости и, следовательно, две новые модификации метода опорных векторов.

5.2 Методы с регулируемой селективностью

Один из новых методов, названный метод релевантных признаков (Relevance Feature Machine - RFM), не выделяя строгого подмножества информативных признаков, наделяет все признаки неотрицательными весами [3, 14]. Чем больше значение структурного параметра селективности, тем большее число весов приближаются к нулевым значениям, фактически исключая соответствующие признаки из принятия решения о классе объекта.

Другой предложенный метод, названный методом опорных признаков (Support Feature Machine - SFM), разбивает все множество признаков на три группы - полностью активные признаки, взвешенные признаки и полностью подавленные признаки. Можно считать, что метод SFM снабжает все признаки весами, но, в отличие от метода RFM, часть весов оказываются строгими единицами, часть принимает значения между единицей и нулем, а часть строго равны нулю.

5.2.1 Метод релевантных признаков с регулируемой селективностью

В дополнение к уже рассмотренным методам регуляризации, предложены два новых класса априорных распределений направляющего вектора, наделяющих обобщенный критерий обучения (13) свойством автоматического отбора признаков.

Метод релевантных признаков с регулируемой селективностью использует естественное обобщение классического метода опорных векторов за счет введения в его вероятностную постановку дополнительного предположения о том, что независимые компоненты направляющего вектора $a = (\alpha_1, \dots, \alpha_n)$ распределены, как и прежде, по нормальному закону с нулевыми математическими ожиданиями, но с разными неизвестными случайными дисперсиями $r_i \geq 0, i = 1, \dots, n$ характеризующимися некоторым априорным распределением, и подлежащими оцениванию в процессе обучения по байесовскому принципу вместе с параметрами разделяющей гиперплоскости (a, b) .

В данном методе оптимизационная задача обучения эквивалентна критерию [3, 15]:

$$\begin{cases} J_{RFM}(a, b, r, \delta | C, \mu) = \sum_{i=1}^n [(\frac{1}{r_i})(\alpha_i^2 + (\frac{1}{\mu})) + ((\frac{1}{\mu}) + 1 + \mu) \ln r_i] + \\ \quad + C \sum_{j=1}^N (\delta_j) - \rightarrow \min(a, b, r, \delta) \\ y_i(\sum_{i=1}^n (a_i x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N, C = 2c \end{cases}$$

Итерационная процедура обучения обычно сходится за 10-15 шагов, а сам алгоритм демонстрирует способность подавлять неинформативные признаки за счет выбора очень маленьких весов в решающем правиле.

5.2.2 Метод опорных признаков с регулируемой селективностью

Данный метод базируется на другом предположении об априорном распределении $\Psi(a)$ независимых компонент направляющего вектора а $a = (\alpha_1, \dots, \alpha_n)$. Вид априорной плотности $\Psi(a)$ представляет собой комбинацию распределения Лапласа при значениях нормы компонент и нормального распределения при больших значениях $|\alpha_i| \geq \mu$

В данном методе мы имеем следующий вид оптимизационной задачи обучения [3, 17]:

$$\begin{cases} J_{SFM}(a, b, \delta | C, \mu) = 2\mu \sum_{|\alpha_i| \leq \mu} |\alpha_i| + \sum_{|\alpha_i| \geq \mu} (\alpha_i^2 + \mu^2) + C \sum_{j=1}^N (\delta_j) - > \min(a, b, \delta) \\ y_i (\sum_{j=1}^N (a_j x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N, C = 2c \end{cases}$$

Задача является задачей выпуклого программирования в пространстве $R^n \times R^{N+1}$, поскольку каждое слагаемое целевой функций выпукло в R^n , а система ограничений для заданной обучающей выборки образует выпуклый многогранник.

Оптимальные ненулевые значения компонент направляющего вектора, как и в классическом методе опорных векторов, выражаются в виде линейной комбинации опорных объектов обучающей выборки, но, в отличие от SVM, решение задачи явным образом указывает множество строго нулевых компонент, а, следовательно, множество неактивных или неопорных признаков, т.е. признаков, не участвующих в принятии решения о классе нового объекта. По этой причине, по аналогии с методом опорных векторов, предложенный метод назван методом опорных признаков с регулируемой селективностью или **Selective Support Feature Support Vector Machine**

6 Сравнительный анализ

6.1 Метод опорных векторов (SVM)

Преимущества SVM:

- SVM имеет свойство *разреженности*, то есть можно исключить из рассмотрения нулевые λ_i и построить компактный классификатор (решающую функцию) [1, 60]
- Метод имеет модификацию преобразования множественной классификации в последовательность бинарных классификаций
- SVM позволяет работать с линейно неразделимыми обучающими выборками
- Данный метод показывает один из наилучших на данный момент результатов по точности классификации в сочетании с формированием вектора признаков изображения на основе метода LBP [1, 78]

Все найденные недостатки метода SVM были указаны в пункте 5.

6.2 Модификации метода опорных векторов

6.2.1 1-norm SVM (LASSO SVM)

Преимущества LASSO SVM:

- Отбор признаков с параметром селективности μ : чем больше μ , тем меньше признаков останется

Недостатки LASSO SVM:

- LASSO начинает отбрасывать значимые признаки, когда ещё не все шумовые отброшены
- Нет эффекта группировки (grouping effect): значимые зависимые признаки должны отбираться вместе и иметь примерно равные веса w_j

6.2.2 Doubly Regularized SVM (ElasticNet SVM)

Преимущества Doubly Regularized SVM:

- Отбор признаков с параметром селективности μ : чем больше μ , тем меньше признаков останется
- Есть эффект группировки

Недостатки Doubly Regularized SVM:

- Шумовые признаки также группируются вместе, и группы значимых признаков могут отбрасываться, когда ещё не все шумовые отброшены

6.2.3 Метод релевантных признаков с регулируемой селективностью

Преимущества RFM:

- Отбор признаков с параметром селективности μ : чем больше μ , тем меньше признаков останется
- Есть эффект группировки
- Лучше отбирает набор значимых признаков, когда они только совместно обеспечивают хорошее решение

Недостатки RFM:

- Достаточно трудоемкий метод, в сравнение с базовым SVM

6.2.4 Метод опорных признаков с регулируемой селективностью

Преимущества SFM:

- Отбор признаков с параметром селективности μ
- Есть эффект группировки
- Значимые зависимые признаки ($|w_j| > \mu$) группируются и входят в решение совместно (как в Elastic Net),
- Шумовые признаки ($|w_j| < \mu$) подавляются независимо (как в LASSO)

Недостатки SFM:

- Достаточно трудоемкий метод, в сравнение с базовым SVM

6.2.5 Практическое сравнение модификаций SVM

В диссертации [3, 19] приводятся результаты экспериментального исследования рассмотренных методов обучения с регулируемой селективностью, а именно, метода релевантных признаков и метода опорных признаков. Основной целью экспериментального исследования является анализ предложенных методов в сравнении с существующими методами SVM, Lasso SVM и Elastic Net SVM по их способности сокращать признаковое описание объектов распознавания и, в конечном итоге, повышать обобщающую способность обучения при относительно малой обучающей выборке и большом числе признаков.

Экспериментальное исследование имеет общепринятую структуру, включающую в себя серию модельных экспериментов и пример решения прикладной задачи. Модельные эксперименты диссертации основаны, по своей структуре, на исследовании, проведенном авторами метода Elastic Net SYM для иллюстрации преимущества модульноквадратичной функции штрафа Elastic Net по сравнению с традиционным модульным Lasso. Прикладной задачей является задача распознавания рака легких. Структура четырех модельных задач настоящей диссертации соответствует четырем простым требованиям к селективному обучению - (5).

Результаты модельных экспериментов, изложенные в диссертации, наглядно иллюстрируют весьма важный факт, что ни один из существующих методов селективного обучения, включая предлагаемые, не удовлетворяет сразу всем этим весьма неизошренным требованиям.

Предложенный метод опорных признаков хорошо справился с подавлением шумовых признаков при линейно независимых информативных признаках, существенно улучшив и без того неплохой результат существующего метода Elastic Net SVM. Однако метод релевантных признаков в этих условиях оказался далеко не так эффективен. Требование выделять группу информативных признаков, которые только вместе обеспечивают достаточную точность распознавания, оказалось весьма проблематичным, как для существующего Elastic Net SVM, так и для предложенного метода опорных признаков, селективность которых обеспечивается за счет использования в целевых функциях критериев обучения штрафа модуля. Вместе с тем требование

не доставило никаких сложностей для второго предложенного метода релевантных признаков, селективность которого имеет отличную от привычного модуля природу.

Полученные в рамках диссертационной работы результаты экспериментального исследования иллюстрируют полезность рассмотренных методов.

6.3 Методы формирования вектора признаков изображения

6.3.1 Метод локальных бинарных шаблонов (LBP)

Преимущества LBP:

- Метод наиболее информативен с точки зрения формирования вектора признаков по сравнению с другими методами признакового описания
- Локальные бинарные шаблоны инвариантны к небольшим изменениям в условиях освещенности и небольшим поворотам классифицируемого изображения, что обуславливает их широкое распространение для решения задач определения таких атрибутов личности, как "пол" "раса" и "возрастная группа" [1, 37]
- Важным достоинством метода LBP является простота реализации LBP, что позволяет использовать его в задачах обработки изображений в реальном времени
- LBP позволяет сформировать пространство признаков большой размерности (порядка нескольких тысяч), обеспечивая высокую концентрацию информации об исходном изображении, и создает предпосылки для более точной классификации по возрасту
- Метод дает возможность использовать только те шаблоны, которые хранят больше информации о локальных особенностях изображения [1, 11]

Недостатки LBP:

- При небольшом количестве разбиений изображения или его отсутствии теряется информация об расположении локальных особенностей изображения
- При формировании вектора признаков изображения получается пространство большой размерности (более 2000), поэтому необходимо вводить модификации для снижения данной размерности и учитывать симметричность лица и различную информативность отдельных участков изображения лица [1, 46]
- Особое значение в рассмотрении метода равномерных шаблонов выявлено чисто эмпирически, поэтому существует необходимость описать более формально данный подход и сформировать унифицированное функциональное представление изображений

6.3.2 Метод построения гистограммы направленных градиентов (HOG)

Данные методы похожи на метод локальных бинарных шаблонов, поэтому они наследуют часть преимуществ и недостатков последнего.

Недостатки:

- Размерность пространства признаков при использовании метода HOG плохо поддается сокращению, а процедура вычисления градиента интенсивности заметно сложнее процедуры формирования LBP [1, 39]
- Методы более пригодны для решения задач выделения заданного объекта на изображении, а не для определения атрибутов личности

6.3.3 Метод построения активной модели формы (ASM) и активной модели внешности (AAM)

Преимущества:

- Методы позволяют получать хорошие результаты при определении таких атрибутов личности, как "пол" и "раса" так как именно взаимное расположение важных антропометрических точек совместно с текстурой изображения позволяют определить принадлежность личности к определенному полу и расе [1, 44]

Недостатки [1, 45]:

- Данный подход формирования вектора по признаку "возраст" может использоваться только для определения возрастной группы, а не для вычисления прогнозируемого возраста, так как расположение важных антропометрических точек и расстояние между ними практически не меняются при изменениях в возрасте на несколько лет
- Проблемой данного метода является сложность автоматического выделения на изображении характерных антропометрических точек, так как для их выделения изображение анализируется на уровне текстуры (интенсивности пикселей) ввиду отсутствия другой информации об изображении

7 Краткий вывод

Использование метода опорных векторов является существенно менее трудоемким и дает большую точность классификации при определении атрибута "пол" на неподготовленных изображениях. Наиболее перспективным методом формирования вектора признаков является метод локальных бинарных шаблонов (LBP) как для задачи определения атрибутов "пол" и "раса" так и для атрибута "возраст". Их сочетание дает точность для атрибута "пол" порядка 87%, для атрибута "раса" около 81.3%, а при прогнозировании возраста личности можно достичь точности в 53% [1, 124-128].

Существующие на сегодня методы классификации изображений лиц по атрибуту "пол" в большинстве случаев позволяют правильно классифицировать примерно две трети реальных изображений, что дает предпосылку к созданию модификаций с целью повышения качества классификации.

8 Источники

1. Рыбинцев А.В. Исследование, модификация и разработка методов компьютерного зрения для задач определения атрибутов личности по изображению лица: диссертация на соискание ученой степени кандидата технических наук. - М.: НИУ МЭИ, 2018.
2. Лепский А.Е., Броневиц А.Г. Математические методы распознавания образов: курс лекций. - Таганрог: Технологический институт Южного федерального университета, 2009.
3. Татарчук А.И. Байесовские методы опорных векторов для обучения распознаванию образов с управляемой селективностью отбора признаков: автореферат диссертации на соискание ученой степени кандидата физико-математических наук. - М.: ФГБУН «Вычислительный центр им. А.А. Дородницына Российской академии наук», 2014.
4. Математические методы распознавания образов: Доклады 13-й Всероссийской конференции, посвящённой 15-летию РФФИ. - М., 2007.
5. Воронцов К.В. Машинное обучение: курс лекций [Электронный ресурс]: Линейные методы классификации и регрессии: метод опорных векторов.
- URL: <http://machinelearning.ru/wiki/images/archive/a/a0/20160310092432!Voron-ML-Lin-SVM.pdf> (дата обращения: 02.05.2021).