

# Применение Методов Машинного Обучения к Задаче Автоматического Распознавания Пола и Возраста Людей по Изображению Лица

Лев Шмаглит, Владимир Хряшев  
Физический факультет

Ярославский Государственный Университет имени П.Г.Демидова, Ярославль, Россия

[Lev\\_shmaglit@yahoo.com](mailto:Lev_shmaglit@yahoo.com), [vhr@yandex.ru](mailto:vhr@yandex.ru)

## Аннотация

В данной работе представлен алгоритм автоматического распознавания пола и возраста людей по изображению лица. В его основе лежат современные методы машинного обучения, такие как линейный дискриминантный анализ и метод опорных векторов. Разработанный алгоритм позволяет решать задачу разбиения анализируемых объектов на три класса («мужчины», «женщины» и «дети») с точностью 89%.

**Ключевые слова:** Выделение Лиц, Распознавание Пола, Машинное Обучение.

## 1. ВВЕДЕНИЕ

Алгоритмы, осуществляющие автоматический анализ и распознавание лица человека, находят применение в системах технического зрения, робототехнике, системах видеонаблюдения и контроля доступа, в интерфейсах взаимодействия человек-компьютер [6]. Одной из наиболее актуальных задач, решаемых данным классом алгоритмов, является классификация анализируемых объектов по полу (gender recognition) [4]. Распознавание пола востребовано в технологии Digital Signage – представление информации с электронных носителей (дисплеев, проекционных систем), установленных в общественных местах. Кроме того, распознавание пола может использоваться, к примеру, для сбора и оценки демографических показателей, а также как важный этап предобработки при решении задачи идентификации личности, поскольку оно позволяет вдвое (в случае одинакового числа мужчин и женщин в базе данных) уменьшить число кандидатов при анализе, и, таким образом, вдвое ускорить процесс идентификации. Следует отметить, что методы, применяемые в задаче распознавания пола, являются универсальными, и поэтому могут быть успешно использованы для решения других задач в области распознавания образов [6].

Для организации полностью автоматической системы, распознавание пола используется совместно с алгоритмом выделения лиц, который отбирает кандидатов для анализа [3]. От качества работы алгоритма выделения лиц во многом зависит итоговый результат всей системы, поскольку неточности при определении местоположения лица на изображении могут привести к принятию ошибочного решения на этапе распознавания.

Выделенные фрагменты подвергаются предобработке для приведения их к единому разрешению и яркостным характеристикам. Нами использовалось масштабирование до разрешения 40×40 пикселей. Эта величина была подобрана экспериментально, как компромисс между скоростью и качеством распознавания. Для устранения различий в степени

освещенности и контрасте к выделенным фрагментам применялась процедура выравнивания гистограммы яркости.

На последнем этапе выделенные и обработанные фрагменты поступают на вход классификатора, который принимает решение об их принадлежности к тому или иному классу. В нашей работе, в отличие от классической постановки задачи распознавания пола, рассматриваются три класса: «мужчины», «женщины» и «дети». Для решения такой задачи предлагается организовать систему из трех двухклассовых классификаторов («мужчины» - «женщины»; «мужчины» - «дети»; «женщины» - «дети»), обученных на наборе тренировочных изображений, и принимающих совместное решение путем простого голосования. Подобный подход позволяет упростить классификатор и ускорить процесс обучения без потери качества распознавания.

Дальнейшее повествование построено следующим образом: вначале приведено описание предлагаемого алгоритма (рассмотрены используемые методы обучения классификатора и этап автоматического выделения лиц), затем представлена методика и результаты его тестирования.

## 2. АВТОМАТИЧЕСКОЕ ВЫДЕЛЕНИЕ ЛИЦ

Анализируя произвольное изображение, необходимо сначала определить, имеются ли на этом изображении лица, где находится каждое лицо и каков его размер. Эту задачу решают алгоритмы выделения лиц. Их работа заключается в сканировании входного изображения окном, имеющим определенную форму и различный масштаб, и в определении к какому классу относится изображение внутри этого окна («лицо», либо «не лицо»). За последние несколько лет было предложено множество алгоритмов выделения лиц, использующих различные подходы [3].

В нашей работе был использован современный алгоритм выделения лиц, базирующийся на обучающей сети SNoW (Sparse Network of Winnows) [5]. Этот алгоритм отличается высокой точностью в определении местоположения лица и низкой ошибкой классификации. Рассмотрим подробнее его структуру и принцип работы.

Алгоритм на базе SNoW можно условно разбить на три этапа:

### 2.1 Переход от значений пикселей к локальным SMQT признакам

SMQT [5] (Successive Mean Quantization Transform) – это преобразование, которое позволяет извлечь из локальной области изображения составляющую, не зависящую от освещенности. Оно заключается в квантовании области изображения с порогом квантования, равным среднему значению пикселей, входящих в эту область. При этом

локальная область задается как блок разрешением  $3 \times 3$  пикселя.

Для нахождения лиц на изображении используется окно детектора разрешением  $32 \times 32$  пикселя. Сканирование осуществляется с шагом в один пиксель, как по горизонтали, так и по вертикали. Для того чтобы выделять лица разного размера, изображение многократно масштабируется с коэффициентом масштабирования 1,2. Чтобы избежать краевых эффектов, к окну детектора применяется овальная маска, содержащая 648 пикселей. Каждому из этих 648 пикселей соответствует свой SMQT признак.

Переход к SMQT признакам позволяет алгоритму адаптироваться к изменению освещенности объекта, так как SMQT признаки извлекают из изображения не зависящие от освещенности компоненты.

## 2.2 Классификатор на базе обучающей сети SNoW

Обучающая архитектура SNoW (Sparse Network of Winnos) представляет собой разреженную сеть линейных элементов в пространстве признаков [5]. Большим достоинством сети SNoW является возможность создания весовых таблиц для классификации. Пусть  $x$  – значение одного пикселя входного изображения,  $W$  – набор SMQT признаков  $M(x)$ , тогда классификатор может быть получен посредством использования весовой таблицы для не лиц  $h_x^{nonface}$ , весовой таблицы для лиц  $h_x^{face}$ , и определения порога  $\theta$ . Решающее правило такого классификатора может быть представлено в виде:

$$\theta = \sum_{x \in W} h_x^{nonface}(M(x)) - \sum_{x \in W} h_x^{face}(M(x)).$$

Поскольку обе таблицы работают в одном домене, их можно объединить в одну весовую таблицу:  $h_x = h_x^{nonface} - h_x^{face}$ .

Для обучения классификатора использовалась база тренировочных изображений лиц и не лиц. Изображения лиц были получены с помощью web-камеры. Далее на них вручную были отмечены три точки: правый глаз, левый глаз и центральная точка внешней границы верхней губы. После этого лицо деформировалось в блок  $32 \times 32$  пикселя с различным расположением опорных точек. Деформация необходима для того, чтобы сымитировать различное положение лица по отношению к камере. В результате описанной процедуры был получен обучающий набор лиц емкостью порядка 1 миллиона образцов. База обучающих изображений не лиц первоначально состояла из случайным образом сгенерированных изображений. После того, как она была дополнена ошибками классификации, ее емкость также составила порядка 1 миллиона образцов.

Каждому обучающему изображению соответствует свой набор SMQT признаков. Пусть тренировочная база состоит из  $i = 1, 2, \dots, K$  наборов SMQT признаков  $M_i(x)$  и соответствующих каждому набору значений классов  $c_i$  («лицо» или «не лицо»). Тогда весовые таблицы лиц и не лиц могут быть обучены с помощью закона обновления весов, получившего название Winnow Update Rule [5]. Изначально обе таблицы содержат нули. При первом обращении к элементу таблицы в процессе обучения ему присваивается

значение 1. Три параметра системы задаются вручную: порог  $\gamma$ , повышающий коэффициент  $\alpha > 1$  и понижающий коэффициент  $\beta < 1$ . Их значения были заданы следующим образом:  $\alpha = 1,005$ ,  $\beta = 0,995$ ,  $\gamma = 200$ . Если

$$\sum_{x \in W} h_x^{face}(M_i(x)) \leq \gamma \text{ и } c_i - \text{лицо, то веса обновляются}$$

согласно формуле:

$$h_x^{face}(M_i(x)) = \alpha h_x^{face}(M_i(x)), \quad \forall x \in W.$$

$$\text{Если } \sum_{x \in W} h_x^{face}(M_i(x)) > \gamma \text{ и } c_i - \text{лицо, то значение весов}$$

уменьшается:

$$h_x^{face}(M_i(x)) = \beta h_x^{face}(M_i(x)), \quad \forall x \in W.$$

Эта процедура повторяется до достижения неизменного значения весов. Обучение весовой таблицы для не лиц происходит точно таким же образом. И, наконец, общая весовая таблица находится по формуле  $h_x = h_x^{nonface} - h_x^{face}$ .

## 2.3 Организация каскадной структуры

С целью увеличения быстродействия алгоритма полный классификатор SNoW разбивается на более слабые классификаторы, которые организуются в каскадную структуру (алгоритм Split-up SNoW). При этом не требуется дополнительного обучения слабых классификаторов.

Пусть все возможные значения одного признака ограничиваются множеством  $P_i, i = 1, 2, \dots, 2^N$ , тогда вклад каждого признака в разделение классов можно оценить по формуле:

$$v_x = \sum_{i=1}^{2^N} |h_x(P_i)|, \quad \forall x \in W. \quad \text{Отсортировав}$$

признаки по соответствующим им значениям  $v_x$ , можно получить рейтинг лист признаков (importance list). Пусть  $W' \in W$  – это подпространство, содержащее определенное количество признаков с наибольшим рейтингом. Тогда:  $\theta' = \sum_{x \in W'} h_x(M(x))$  можно определить как слабый

классификатор, который выделяет все лица из тренировочной базы, но имеет больший по сравнению с полным классификатором уровень неверного выделения. Увеличить количество слабых классификаторов в каскаде можно за счет тех же самых операций, выбирая подпространства следующим образом:  $W' \in W'' \in \dots \in W$ .

Каскад состоит из 5 слоев. Разбиения происходят при 20, 50, 100, 200 и 648 признаках. Признаки в каждый слой каскада отбираются согласно их рейтингу (importance list). Каждый слой отбирает кандидатов для следующего слоя и передает следующему слою значения уже посчитанных для этих кандидатов SMQT признаков.

## 3. ПРОЦЕДУРА РАСПОЗНАВАНИЯ ПОЛА

На этапе распознавания пола и возраста нами применялись два различных метода обучения классификатора: метод опорных векторов (MOB) [1] и KDDA (Kernel Direct Discriminant Analysis) [2]. Эти методы являются одними из наиболее эффективных в задаче разделения изображений на два класса [6].

### 3.1 Линейный дискриминантный анализ

Рассмотрим алгоритм построения классификатора KDDA, основанный на линейном дискриминантном анализе. Основная идея заключается в том, чтобы представить входные обучающие изображения в виде  $n$ -мерных векторов, а затем найти такое подпространство, в котором проекции векторов, принадлежащих разным классам, были бы расположены как можно дальше друг от друга, а проекции векторов, принадлежащих одному и тому же классу, наоборот, – ближе друг к другу.

Однако в случае анализа лиц классы могут оказаться линейно неразделимыми. Для решения проблемы линейной неразделимости классов, алгоритм KDDA использует неявное проецирование векторов-признаков в пространство потенциально намного более высокой размерности (еще выше, чем пространство изображений), в котором классы могут оказаться линейно разделимыми (рис. 1).

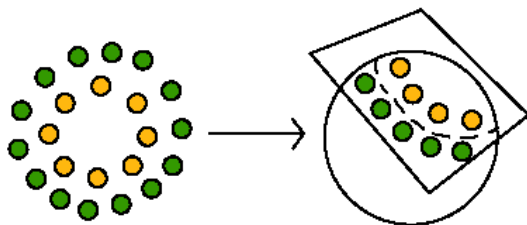


Рис 1: Пример разделения линейно неразделимых классов с помощью перехода к пространству более высокой размерности.

Проецирование осуществляется с помощью аппарата ядерных функций. Неявное проецирование с помощью ядерных функций не приводит к усложнению вычислений, что позволяет успешно использовать линейный классификатор для линейно неразделимых классов.

В качестве ядра в нашей работе использовалась радиальная базисная функция Гаусса:

$$k(z_1, z_2) = \exp\left(-\frac{\|z_1 - z_2\|^2}{\sigma^2}\right).$$

В полученном пространстве признаков считается, что объекты, принадлежащие различным классам линейно разделимы, и применяется D-LDA [2]. Такой выбор связан с тем, что применение классического LDA в задачах распознавания невозможно. Например, в нашей работе обучающая выборка состоит из изображений размером  $40 \times 40 = 1600$  пикселей, значит размерность матриц межклассовой и внутриклассовой дисперсии ( $S_{BTW}$  и  $S_{WTN}$ ) составляет  $1600 \times 1600$ . Кроме того, что работа непосредственно с матрицей  $S_{WTN}$  затруднительна из-за её размерности, обучающая выборка должна состоять из 1600 изображений для каждого класса, иначе матрица  $S_{WTN}$  будет вырожденной. Для решения данной проблемы в LDA использовано предварительное уменьшение размерности с помощью метода главных компонент. Таким образом, размер матрицы  $S_{WTN}$  уменьшается за счет удаления близких к нулю собственных значений.

В алгоритме D-LDA применяется другой способ для решения данной проблемы. Основная идея алгоритма заключается в том, что в «нулевом» пространстве  $S_{WTN}$  может содержаться значимая информация для дискриминантного анализа, если проекция  $S_{BTW}$  не нулевая в данном направлении, и поэтому полезная информация будет утеряна. В отличие от классического LDA, при поиске матрицы для отображения пространства изображений на пространство признаков вначале диагоназируется матрица  $S_{BTW}$ , при этом уменьшается размер матрицы за счет удаления нулевых и близких к нулю собственных значений, которые не содержат полезной информации. Затем диагоназируется матрица  $S_{WTN}$ . Последовательность диагонализации матриц имеет значение, только если матрица  $S_{WTN}$  вырожденная, что в нашем случае имеет место быть. Матрица  $A$  для проецирования пространства изображения на пространство признаков в алгоритме D-LDA выбирается из следующего условия:

$$A = \arg \max_A \frac{|A^T S_{BTW} A|}{|A^T S_{BTW} A + A^T S_{WTN} A|}.$$

Пространство признаков строится на основе собственных векторов матрицы  $A$  на этапе обучения. На этапе тестирования для принятия решения, к какому классу относится вектор в пространстве признаков, используется метод ближайших соседей.

### 3.2 Метод опорных векторов

Применение метода опорных векторов (МОВ) [1] к задаче разделения анализируемых изображений на два класса заключается в поиске оптимальной гиперплоскости в признаковом пространстве, отделяющей один класс от другого. В качестве признаков служат значения пикселей изображения, представленные в виде  $n$ -мерного вектора. Оптимальной считается гиперплоскость, которая максимизирует ширину полосы между классами (рис. 2).

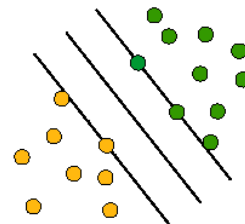


Рис 2: Пример разделяющей полосы классификатора на базе МОВ.

Разделяющая гиперплоскость определяется как линейная комбинация небольшого набора тренировочных векторов, называемых опорными векторами. Обозначим набор собственных векторов как  $\{X_1, \dots, X_m\}$ , а соответствующие им коэффициенты линейной комбинации –  $\{\alpha_1, \dots, \alpha_m\}$ . МОВ – это линейный классификатор, поэтому для разделения линейно неразделимых классов, применяется проецирование векторов-признаков в пространство потенциально намного более высокой размерности. Формально решающее правило

классификатора на базе МОВ для входного изображения  $X$  может быть записано следующим образом:

$$f(X) = \operatorname{sgn}\left(\sum_{i=1}^m y_i \alpha_i k(X_i, X) + b\right),$$

где  $k(\cdot, \cdot)$  – это ядерная функция, а  $b$  – смещение.

В качестве ядерной функции, как и для метода KDDA, использовалась радиальная базисная функция Гаусса.

#### 4. МЕТОДИКА И РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ

Для обучения и проведения тестирования была составлена база, состоящая из 2770 изображений, содержащих фронтально расположенные лица «мужчин», «женщин» и «детей». На каждом изображении из этой базы были выделены лица алгоритмом SNoW. При этом уровень выделения составил 98,2%. Ложные выделения были отсеяны вручную, после чего к оставшимся изображениям была применена предобработка. Далее было отобрано по 400 лиц каждого класса. Эти лица были разделены на обучающий набор (по 300 лиц каждого класса) и тестовый набор (по 100 лиц). Были обучены два классификатора: один – с помощью алгоритма KDDA, и второй – на базе метода опорных векторов.

В табл. 1 и 2 представлены результаты распознавания для каждого класса при использовании алгоритмов классификации на базе KDDA и МОВ соответственно.

**Таблица 1:** Распознавание на базе классификатора KDDA.

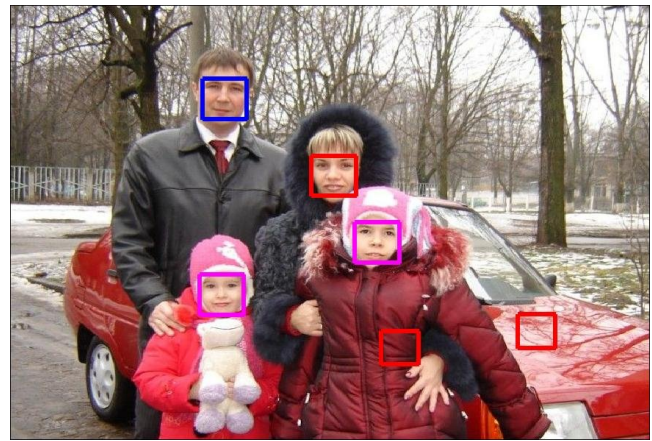
Решение \ Входное изображение	Мужчины	Женщины	Дети
Мужчины	88%	12%	0%
Женщины	7%	83%	10%
Дети	7%	23%	74%

**Таблица 2:** Распознавание на базе классификатора МОВ.

Решение \ Входное изображение	Мужчины	Женщины	Дети
Мужчины	94%	5%	1%
Женщины	8%	81%	11%
Дети	1%	7%	92%

Классификатором KDDA было верно распознано 245 фрагментов из 300, таким образом, его средний уровень распознавания составил 81,6%. Классификатор МОВ показал уровень распознавания 89% (267 верных фрагментов из 300). При этом оба алгоритма допускают больше ошибок при различении «детей» от «женщин».

Пример работы системы распознавания пола и возраста представлен на рис. 3.



**Рис 3:** Распознавания пола и возраста людей на тестовом изображении.

Здесь фрагменты, распознанные как класс «мужчины», условно выделены синим цветом; «женщины» – красным; «дети» – фиолетовым. Из рис. 3 видно, что лица были распознаны верно, однако, было допущено два ложных выделения, которые были распознаны как класс «женщины».

#### 5. ЗАКЛЮЧЕНИЕ

Итак, был представлен алгоритм распознавания пола и возраста по изображению лица. Для обучения классификатора использовались два различных метода: линейный дискриминантный анализ и метод опорных векторов. Наибольшую эффективность продемонстрировал классификатор на базе МОВ. Уровень распознавания в задаче разбиения изображений на три класса составил порядка 90%.

Дальнейшая работа будет направлена на увеличение уровня распознавания путем применения интеллектуальной предобработки, обучения алгоритма на собственных ошибках, увеличения обучающего набора изображений, оптимизации параметров классификаторов.

#### 6. ССЫЛКИ

- [1] Burges C. A Tutorial on Support Vector Machines for Pattern Recognition // Data Mining and Knowledge Discovery, V. 2, P. 121-167, 1998.
- [2] Gao H., Davis J. Why direct LDA is not equivalent to LDA // Pattern Recognition Letters 39, № 5, P. 1002-1006, 2006.
- [3] Kriegman D., Yang M.H., Ahuja N. Detecting faces in images: A survey // IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 24, № 1, P. 34-58, 2002.
- [4] Makinen E., Raisamo R. An experimental comparison of gender classification methods // Pattern Recognition Letters 29, № 10, P. 1544-1556, 2008.
- [5] Nilsson M., Nordberg J., Claesson I. Face Detection Using Local SMQT Features and Split Up SNoW Classifier // Proceedings of IEEE Int. Conf. ICASSP, V. 2, P. 589-592, 2007.
- [6] Потапов А.С. Распознавание образов и машинное восприятие: общий подход на основе принципа минимальной длины описания. - СПб.: Политехника. — 2007.