

Обнаружение лиц на изображении с помощью адаптивных методов

Ф.М. Жданов, М. Ю. Карасёва,

Московский Государственный Университет им. М.В. Ломоносова, Москва, Россия

zhdanov_fm@mail.ru, loginus@yandex.ru

Аннотация

Задача обнаружения лица на изображении является ключевой при решении задач распознавания, автоматического слежения за передвигающимся в поле зрения камеры субъектом, распознавания человеческих эмоций по изображению лица. Целью данной работы является сравнение методов разделения изображений на лица и "не-лица". Проводится сравнение двух эффективных подходов к построению классификатора: искусственная нейронная сеть и метод опорных векторов (SVM). На основе полученных результатов делаются выводы о возможностях обоих подходов.

Ключевые слова: Обнаружение лица, SVM, нейросеть, задача классификации, адаптивные методы.

1. ВВЕДЕНИЕ

Задача обнаружения лица обычно решается в два этапа. Сначала на изображении производится поиск фрагментов-претендентов для классификации. Затем производится разделение найденных фрагментов на классы лиц и "не-лиц". В настоящее время в этих целях используются такие методы, как Метод Главных Компонент (PCA), линейный дискриминантный анализ (LDA), искусственные нейронные сети, метод опорных векторов (SVM)[1],[2]. Все эти методы являются адаптивными, т.е. для построения классификатора требуется некоторое количество обучающих изображений лиц и "не-лиц", которые вручную выбираются из базы изображений. В рамках данной статьи проводится сравнение двух эффективных методов классификации: нейросеть и метод опорных векторов.

Алгоритм преобразования входных изображений в векторы, используемые для обучения, описан в части 2. В части 3 данной статьи приведена общая постановка задачи классификации и изложено функционирование методов. Далее в части 4 говорится об архитектуре и параметрах используемых методов, а также проведено сравнение методов. Часть 5 посвящена сравнению двух классификаторов.

2. ПОДГОТОВКА ВХОДНЫХ ДАННЫХ

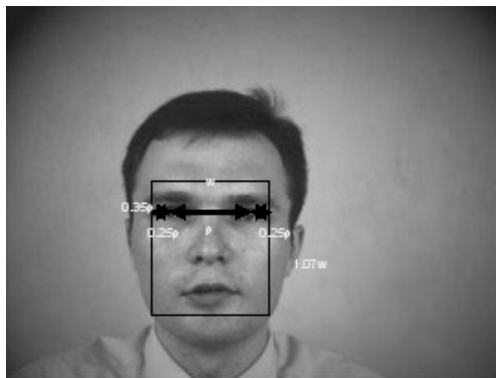


Рис. 1. Выделение изображения лица

Для обучения и тестирования использовались черно-белые (gray-scale) изображения лиц размером 22x24 пикселя. Из имеющихся изображений 320x240 пикселей выделялись изображения лица, а затем полученные фрагменты

приводились к размеру 22x24. Для этого использовался следующий алгоритм:

а) вручную на каждом изображении размечаются центры глаз;

б) обозначая расстояние между центрами глаз за ρ , отступим от них влево и вправо $0,25\rho$ - получим ширину фрагмента $w=1,5\rho$ (см. рис. 1)

с) высота фрагмента h считается равной $1,07w$, причем расстояние от центров глаз до верхнего края фрагмента равно $0,35\rho$.

д) фрагмент масштабируется к размеру 22x24.

Для анализа выбирались равномерно освещенные лица. Допускался поворот лица во всех ракурсах на 10-15 градусов относительно фронтального положения. Кроме того, допускалась также некоторая мимика лица: во время съемки человек произносил слова. Примеры исходных изображений показаны на рис. 2.



Рис. 2. Примеры исходных изображений

Для использования методов классификации имеющиеся изображения были преобразованы в векторы. Эти векторы подавались на вход классификатора. В [3] предложен алгоритм преобразования изображения в вектор. Этот алгоритм удобен тем, что переводит вектора из пространства изображений в пространства бинарных векторов очень высокой размерности: 135168. Число единиц в каждом векторе равно 528. В таком пространстве множества векторов-признаков могут оказаться линейно разделимыми. Для сохранения векторов изображения в файлах используется их краткое взаимно-однозначное представление – векторы размерности 528, каждая координата которых соответствует номеру единичной координаты исходного бинарного вектора.

3. ОПИСАНИЕ МЕТОДОВ

В общем виде задача классификации формулируется следующим образом[4]:

Имеется: **пространство векторов** X , m -мерное евклидово пространство R^m векторов-признаков изображения.

пространство ответов $Y=\{1,-1\}$, где $y_i = 1$ означает,

что вектор x_i соответствует изображению лица, а

$y_j = -1$, что x_j соответствует изображению "не-лица".

пространство F функций $f: X \rightarrow Y$, или пространство функций-классификаторов. Требуется по некоторому обучающему набору $(x_i, y_i), i = 1, 2, \dots, N$ найти функцию f , так чтобы достигался минимум среднеквадратической ошибки

$$\sum_{i=1}^N (y_i - f(x_i))^2 \rightarrow \min_{f \in F} \quad (1)$$

Рассмотрим два подхода к построению классификатора.

3.1. Нейросеть

Архитектура сети – трёхслойный перцептрон следующего вида (рис. 3):

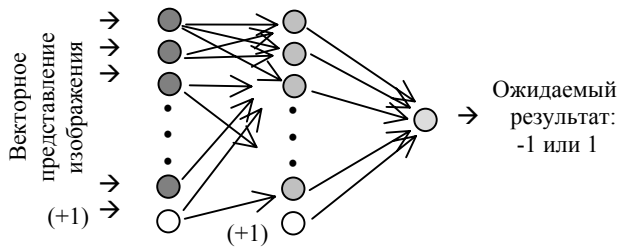


Рис. 3. Вид нейросети

Процесс настройки весов сети должен привести к минимизации среднеквадратической ошибки (1).

Алгоритм нахождения весов методом градиентного спуска:

- 1) Обнуляем ΔW и ΔV .
- 2) Считываем входные векторы X , вычисляем для них Y .
- 3) Найдём сигнал ошибки $(Y - T)$. Если он достаточно мал на всех парах обучающего множества, то считаем обучение сети успешно законченным, в противном случае:
- 4) Умножаем сигнал ошибки на производные по весам и прибавляем это к будущей модификации веса ΔW и ΔV .
- 5) Прибавляем к весам накопленные ранее ΔW и ΔV , умноженные на η - коэффициент скорости обучения. Замечание: можно ещё прибавлять сюда ΔW и ΔV , полученные на предыдущей итерации, умножив их на γ - коэффициент инерции. Тогда ошибка будет уменьшаться более плавно, не кидаясь из стороны в сторону.
- 6) Перейти к шагу (1).

Предусмотрен также выход из этого цикла по числу итераций (если ошибка не убывает до нужной величины).

3.2. SVM

Метод опорных векторов или SVM [5] основан на том, что ищется линейное разделение классов. В этом случае функция решения $f(x) = \text{sgn}(\bar{w} \cdot \bar{x} + b)$, и производится поиск параметров \bar{w} и b . Видно, что $\bar{w} \cdot \bar{x} + b = 0$ - уравнение разделяющей гиперплоскости. Проводя параллельные гиперплоскости $\bar{w} \cdot \bar{x} + b = \pm 1$ (см. рис. 4), получим, что для проведения оптимальной разделяющей гиперплоскости, надо максимизировать расстояние между этими двумя плоскостями, $\frac{2}{\|\bar{w}\|}$, с условиями, что между ними нет точек

данных, т.е. $y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1$. Методом множителей Лагранжа эта задача сводится к поиску коэффициентов $\alpha_i \geq 0$:

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\bar{x}_i \cdot \bar{x}_j) \rightarrow \max \quad c$$

ограничениями $\sum_{i=1}^N \alpha_i y_i = 0$. Эта задача решается с

помощью алгоритма последовательной оптимизации SMO[6] (Sequential Minimal Optimization). Этот метод сводит решаемую задачу максимизации функции N переменных к задаче с минимально возможным количеством α - с двумя - и является одним из наиболее эффективных методов обучения.

4. ПАРАМЕТРЫ МЕТОДОВ

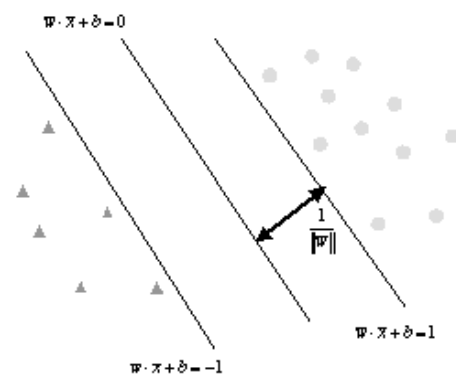


Рис. 4. Иллюстрация к разделению на классы

4.1. Нейросеть

Количество нейронов в слоях трёхслойного перцептрона равно соответственно 529, 101, 1, с учётом добавления дополнительного нейрона во всех слоях, кроме последнего. Дополнительный единичный вход необходим для реализации свободного коэффициента уравнения плоскости (рис. 4). Обучение нейросети производилось с помощью метода градиентного спуска, а для нахождения хорошего начального приближения использовался метод "отжига" [7],[8].

4.2. SVM

Было установлено, что возможно провести линейное разделение векторов. Параметр штрафа C [6] для SVM выбирался эмпирически, наилучшие результаты метод показал при $C=1$.

5. СРАВНЕНИЕ МЕТОДОВ

Доля изображений лиц составляла 24,6 % как в обучающих, так и в тестовых выборках. Для обучения SVM и нейросети использовались одни и те же векторы размерности 528. Среднюю ошибку распознавания мы считаем равной = общее количество ошибочно классифицированных изображений / общее кол-во изображений. Тестирование адекватности распознавания проводилось на векторах, не участвовавших в обучении.

В результате сравнения двух описанных выше методов можно составить следующие таблицы:

Таблица 1.

Для обучения были использованы 100 изображений, из них 25 лиц. Тестовое множество: 1000 изображений, из них 246 лиц.

Критерий сравнения	Нейросеть	SVM
Ошибка классификации лиц	21,14 %	3,66 %
Ошибка классификации "не-лиц"	0,4 %	0,13 %
Общая (средняя) ошибка классификации	5,5 %	1 %
Время обучения	~20 минут	1 сек

Таблица 2.

Для обучения были использованы 300 изображений, из них 74 лица. Тестовое множество: 3000 изображений, из них 738 лиц.

Критерий сравнения	Нейросеть	SVM
Ошибка классификации лиц	16,53 %	0,68 %
Ошибка классификации "не-лиц"	3,45 %	12,86 %
Общая (средняя) ошибка классификации	6,7 %	9,87 %
Время обучения	~ 32 часа	4 сек

6. ЗАКЛЮЧЕНИЕ

Подводя итог, отметим, что оба метода могут быть применены к данной задаче в зависимости от различных условий.

Для обоих рассмотренных алгоритмов можно отметить высокую скорость работы в режиме эксплуатации (~ 1-2 мс).

Преимуществом нейросети является то, что при увеличении обучающей выборки она способна улучшать качество обобщения. Эксперимент показал, что чем больше обучающих пар, тем выше качество распознавания при том же соотношении размеров обучающей выборки и тестовой. Кроме того, ошибка второго рода у нейросети меньше, чем у SVM.

Одним из основных преимуществ SVM можно считать высокую скорость обучения алгоритма, а, следовательно, и возможность использовать для обучения достаточно большое количество векторов. Кроме того, ошибка первого рода у SVM меньше, чем у нейросети.

Результаты данной статьи могут быть полезны при создании автоматической системы обнаружения лиц на изображениях. Для этого необходимо так выбрать 1) алгоритмы выделения фрагментов на изображении и 2) алгоритмы классификации фрагментов, чтобы их комбинация дала наилучшие результаты – хорошее качество распознавания и высокую скорость работы.

Авторы выражают благодарность научному руководителю Лукьянице А.А. за помощь в проведении исследования. Авторы также благодарны Буряку Д.Ю. и Мурынину А.Б. за предоставленные материалы.

7. БИБЛИОГРАФИЯ

1. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey", IEEE transactions on

pattern analysis and machine intelligence, vol. 24, No.1, 2002. pages 34-58.

2. Вежнев В., Дегтярева А., "Обнаружение и локализация лица на изображении", CGM Journal, 2003.

3. D. Roth, M.-H. Yang, N. Ahuja, "A SNoW-Based Face Detector", Advances in Neural Information Processing Systems 12, pp. 855-861, MIT Press, 2000

4. А.Б. Мерков, "О статистическом обучении", 2006 <http://www.recognition.mccme.ru/pub/RecognitionLab.html/slt.html>

5. Vladimir N. Vapnik "An Overview of Statistical Learning Theory", IEEE transactions on neural networks, vol.10, no.5, pp. 988-999, September 1999;

6. John C. Platt "Fast Training of Support Vector Machines using Sequential Minimal Optimization", In Advances in Kernel Methods - Support Vector Learning, Edited by Scholkopf, B., Burges, C., and Smola, A., MIT Press, 1998., pp.41-65

7. Уоссермен Ф. "Нейрокомпьютерная техника: теория и практика". <http://www.codenet.ru/progr/alg/nks>

8. Р.Каллан. "Основные концепции нейронных сетей". 2002 г., издательство Вильямс. Серия: Основы вычислительных систем. ISBN 5-8459-0210-X

Об авторах

Жданов Федор Михайлович – студент 4-го курса факультета вычислительной математики и кибернетики Московского Государственного Университета им. М.В. Ломоносова.

Адрес: Москва, 119899, Воробьевы горы, МГУ, 2-й учебный корпус, факультет ВМиК, кафедра автоматизации научных исследований.

E-mail: zhdanov_fm@mail.ru

Карасёва Мария Юрьевна – выпускница факультета вычислительной математики и кибернетики Московского Государственного Университета им. М.В. Ломоносова

Адрес: Москва, 119899, Воробьевы горы, МГУ, 2-й учебный корпус, факультет ВМиК, кафедра автоматизации научных исследований.

E-mail: loginus@yandex.ru

Face Detection in Images using Learning Methods

Abstract

The paper is devoted to a face detection problem. It appears in applications based on face recognition, subject tracking, and express recognition. To build automated face recognition system, 1) effective algorithms for fragment extracting from image and 2) methods to classify these fragments on faces and not-faces images are needed. The purpose of this paper is to compare two effective approaches for classification: artificial neural network and Support Vector Machine (SVM). After comparing we conclude about possibilities of two approaches.

Keywords: Face detection, SVM, artificial neural network, classification problem, learning methods.

About the authors

Zhdanov Fedor Mikhailovich is a student of Computational Mathematics and Cybernetics faculty, Moscow State University. His contact e-mail is zhdanov_fm@mail.ru.

Karaseva Maria Yurievna is graduated from Computational Mathematics and Cybernetics faculty, Moscow State University. Her contact e-mail is loginus@yandex.ru