

Построение классификаторов с использованием искусственных нейронных сетей и принципа ADABOOST

А. В. Стадник, А. В. Кравчук, К. И. Гулина

*Кафедра прикладной математики и информатики
Международный университет природы, общества и человека «Дубна»
ул. Университетская, д. 19, Дубна, Московская область, Россия, 141980*

Проблема построения различного рода детекторов объектов на изображениях до сих пор остаётся актуальной задачей, несмотря на набор достаточно сильных методов, описанных в литературе. Одним из методов, ставших стандартом для построения эффективных и быстрых классификаторов, является каскад Виолы–Джонса, который до сих пор является основополагающим для поиска объектов на изображении в реальном времени и его реализация была включена в открытую библиотеку компьютерного зрения OpenCV. Для экспериментов в данной работе использовалась база данных изображений CMU Face Database.

При прикладном использовании алгоритмов в компьютерном зрении существенным фактором становится вычислительная сложность. Предпочтительно использовать в качестве классификаторов пороговые решающие функции или Хаар-признаки, вычислительная сложность которых мала. Однако, на практике ADABOOST, как жадный алгоритм, не всегда даёт эффективную комбинацию классификаторов. В данной работе рассмотрен подход к построению классификаторов сравнимой эффективности, на примере задачи детектирования лица.

Для построения детектора были исследован подход, предполагающий разбиение процесса детекции на два отдельных этапа: этап построения дескриптора изображения и этап классификации.

Для этапа, отвечающего за классификацию, были рассмотрены две возможности: двухслойная нейронная сеть, т.е. использование многослойного перцептрона в качестве «сильного» классификатора, и каскад из нескольких таких сетей увеличивающего размера.

Для этапа формирования дескриптора также в работе исследовались две возможности. В качестве первой был построен фиксированный базис Хаара, дающий вектор признаков в качестве дескриптора входного изображения. Данный базис был построен с использованием принципа ADABOOST. Второй возможностью, исследованной в работе, было построение базиса признаков Хаара из меньшего количества необходимых признаков, более точно отражающего характерные особенности объектов, который был получен с использованием преобразования Карунена–Лоэва. Для получения признаков Хаара собственные вектора были подвергнуты квантованию.

В результате построен классификатор, сравнимый по эффективности с каскадом Хаара.

Ключевые слова: нейронные сети, ADABOOST, признаки Хаара, детекция объектов, обработка изображений.

1. Введение

Проблема построения различного рода детекторов объектов на изображениях до сих пор остаётся актуальной задачей, несмотря на набор достаточно сильных методов, описанных в литературе. Одним из методов, ставших стандартом для построения эффективных и быстрых классификаторов, является каскад Виолы–Джонса. Хотя метод был разработан и представлен в 2001 г. Полом Виолой и Майклом Джонсом [1, 2], он до сих пор является основополагающим для поиска объектов на изображении в реальном времени [2]. Этот метод показал себя настолько эффективным в задаче детектирования человеческого лица на изображениях (ставшей эталонным примером для разработки и тестирования детекторов

объектов), что его реализация была включена в открытую библиотеку компьютерного зрения OpenCV [3]. Для экспериментов в данной работе использовалась база данных изображений CMU Face Database [4].

При прикладном использовании алгоритмов в компьютерном зрении существенным фактором становится вычислительная сложность. Предпочтительно использовать в качестве классификаторов пороговые решающие функции или Хаар-признаки, вычислительная сложность которых мала. Однако на практике ADABOOST, как жадный алгоритм, не всегда даёт эффективную комбинацию классификаторов.

В данной работе рассмотрен подход к построению классификаторов сравнимой эффективности на примере задачи детектирования лица.

2. Иллюстрация подхода на модельной задаче

Для демонстрации подхода была рассмотрена следующая модельная задача классификации. На рис. 1 слева представлены 2 исходных множества, а также схематично показаны пороговые решающие правила к которым сходится ADABOOST. Алгоритм сошёлся на 10 решающих правилах, из которых 6 независимых. Справа на рисунке представлен результат классификации с соответ-

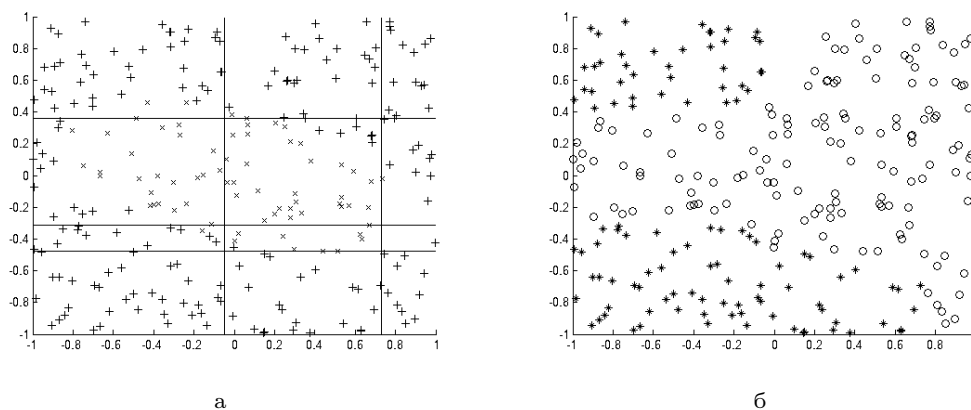


Рис. 1. Модельная задача (а) и классификация ADABOOST (б)

ствующими весами. Как видно, сами решающие правила подобраны достаточно эффективно, однако весовые коэффициенты, составляющие финальный сильный классификатор, не являются оптимальными. На рис. 2 ниже представлены результаты классификации этих модельных данных в случаях, когда пороговые решающие правила, определённые ADABOOST, скомбинированы с использованием однослойной (слева) и двухслойной (справа) нейронных сетей.

Таким образом, альтернативный подход к получению эффективных классификаторов состоит из двух этапов:

- 1) ADABOOST;
- 2) одно- или двухслойная искусственная нейронная сеть. Нейронная сеть обучается на результатах классификаторов ADABOOST, позволяя скомбинировать их более эффективным способом.

3. Принцип работы каскада Виолы-Джонса

Разработанный подход к детектированию объектов на изображении комбинирует четыре ключевые концепции:

- 1) простые прямоугольные функции, называемые функциями Хаара;

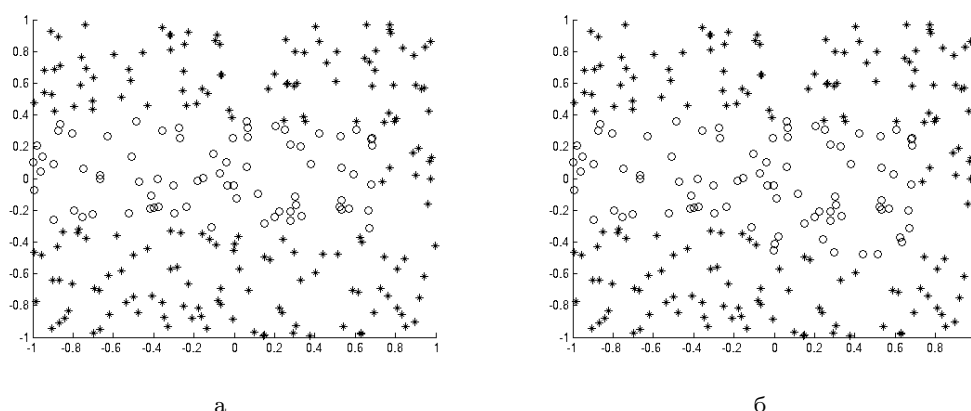


Рис. 2. Классификация с использованием однослойной (а) и двухслойной (б) нейронных сетей

- 2) интегральное изображение для быстрого вычисления функций Хаара;
- 3) метод машинного обучения ADABOOST;
- 4) каскадный классификатор для эффективного совмещения множественных функций.

Особенности, которые использовали Виола и Джонс, базируются на признаках Хаара (Haar features). Для выбора конкретных используемых функций Хаара и установления пороговых уровней используется метод машинного обучения ADABOOST, который комбинирует много «слабых» классификаторов с целью создания одного «сильного» классификатора. ADABOOST выбирает набор слабых классификаторов для объединения и присваивает каждому из них свой вес [5, 6]. Эта взвешенная комбинация и является сильным классификатором. В результирующем каскаде объединяются серии классификаторов ADABOOST как последовательность фильтров, что особенно эффективно для классификации областей изображения. Каждый фильтр является отдельным классификатором ADABOOST с небольшим количеством слабых классификаторов [7].

Принятый порог на каждом уровне устанавливается достаточно низким, чтобы пройти все (или почти все) лицевые образцы в тренировочном наборе. Фильтры на каждом уровне обучены классифицировать тренировочные изображения, которые прошли все предыдущие этапы. Во время работы, если какой-либо из этих фильтров не пропускает область изображения, то область сразу же классифицируется как «не лицо». Когда фильтр пропускает область изображения, она переходит к следующему фильтру в последовательности. Область изображения, прошедшая через все фильтры, классифицируется как «лицо».

4. Детекторы на основе нейронной сети

Классификаторы, используемые как части каскада, достаточно просты. Функционально они являются линейными отображениями с решающим порогом, или однослойный линейный перцептрон, т.е. нейронная сеть самой простой конфигурации. Это приводит к тому, что количество этих классификаторов весьма значительно. В каскадах из библиотеки OpenCV [2] из более порядка 3000.

Для построения детектора были выбраны два следующих принципа:

- 1) фиксированный базис Хаара, дающий фиксированный вектор признаков в качестве дескриптора входного изображения;
- 2) использование многослойного перцептрона в качестве «сильного» классификатора.

При выборе фиксированного базиса первый был построен на основании визуального восприятия изображения; он не обязательно обладает полнотой и не содержит достаточно мелких деталей, несущих высокочастотную информацию (последняя часто не существенна для определения типа объекта). Возможно быстрое вычисление разложения по указанному базису через интегральное изображение. Удовлетворяющий этим требованиям базис из 112 признаков Хаара, использовавшийся в работе, представлен в виде единого изображения на рис. 3.



Рис. 3. Базис из 112 признаков Хаара, использовавшийся в работе, представленный в виде единого изображения

Выбранный базис построения дескриптора совместно с искусственной нейронной сетью конфигурации 112-16-2 (112 входов, 12 нейронов скрытого слоя, 2 нейрона выходного слоя) позволили получить детектор лиц с характеристиками $detection\ rate/false\ positive\ rate \sim 1.4 \cdot 10^{-5}$. В данном контексте для детектора на первый план выходит низкий уровень ложного срабатывания или максимум отношения $detection/false\ positive\ rate$ как критерия оптимизации детектора.

На рис. 4 можно увидеть результаты работы детектора (слева).

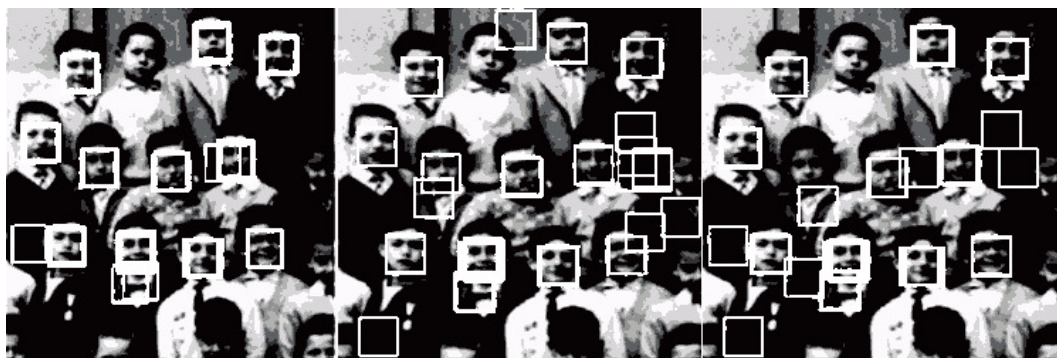


Рис. 4. Слева — результат дескриптора с базисом на рис. 3 + ИНС 112-16-2; далее — результат детекции при использовании первых 110 прямоугольных признаков из каскада Виолы–Джонса, подаваемых на вход линейного слоя нейронов; справа — аналогичный результат при использовании двухслойной нейронной сети прямого распространения в качестве классификатора

5. Собственные признаки Хаара и детектор на их основе

Идея уменьшения размера дескриптора изображения привела к построению базиса признаков Хаара из меньшего количества необходимых признаков, более точно отражающего характерные особенности объектов. Такой базис был получен с использованием преобразования Карунена–Лоэва и сформирован из набора 48 собственных векторов. Для получения признаков Хаара собственные вектора были подвергнуты квантованию, которое заключалась в обнулении пикселей в интервале, не превышающем среднее значение на половину среднеквадратичного отклонения. Пиксели, превышающие этот порог в ту или иную сторону, квантовались до единицы с соответствующим знаком. Полученные «собственные признаки

Хаара» сформировали базис для получения дескриптора изображения. Базис состоит из признаков более сложной формы, но более точно отражающих особенности человеческого лица. Можно также использовать аппроксимацию данного базиса, оставив только существенные области для каждого из признаков, уменьшив тем самым количество прямоугольных областей, необходимых для вычисления.

Результаты демонстрируют отношение $false\ alarm/detection\ rate \sim 3.7 \cdot 10^{-6}$ (рис. 5).

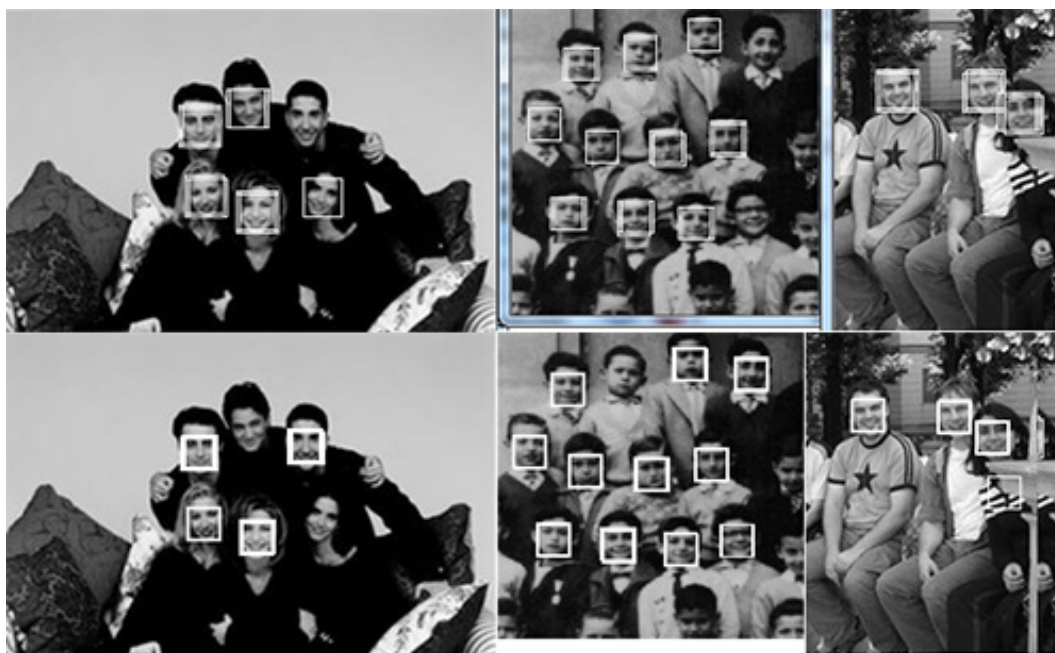


Рис. 5. Собственный Хаар-базис, результаты (внизу): $false\ alarm/detection\ rate \sim 3.7 \cdot 10^{-6}$; ИНС {48-16-1}, сравнение с каскадом Виолы-Джонса

Литература

1. Viola P., Jones M. J. Rapid Object Detection using a Boosted Cascade of Simple Features // Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. — Vol. 1. — 2001. — Pp. 511–518.
2. OpenCV. — 2013. — <http://opencv.org>.
3. Rowley H. A., Baluja S., Kanade T. Neural Network-Based Face Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1998. — Vol. 20. — Pp. 23–38.
4. CMU Image Data Base: Face. — <http://vasc.rh.cmu.edu/idb/html/face/>.
5. Freund Y., Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting // Journal of Computer and System Sciences. — 1997. — Vol. 55, No 1. — Pp. 119–139.
6. Freund Y., Schapire R. A Short Introduction to Boosting // Journal of Japanese Society for Artificial Intelligence. — 1999. — Vol. 14, No 5. — Pp. 771–780.
7. Viola P., Jones M. J. Robust real-time face detection // International Journal of Computer Vision. — 2004. — Vol. 57, No 2. — Pp. 137–154.

UDC 681.518:004.9

Creating Classifiers using Artificial Neural Networks and the ADABOOST Principle

A. V. Stadnik, A. V. Kravchuk, K. I. Gulina

*Department of Applied Mathematics and Informatics
Dubna International University for Nature, Society and Man
19, Universitetskaya str., Dubna, Moscow region, Russia, 141980*

The problem of constructing various types of object detectors in images is still an urgent task, despite the relatively strong set of methods described in the literature. One of the methods that have become standard for the construction of efficient and fast classification, is a Viola-Jones cascade, which is still fundamental to search for objects in the image in real time and which implementation has been included in the open-source computer vision library OpenCV. For the experiments in this study we used the database of images CMU Face Database.

In practice, when we use of the algorithms in computer vision the computational complexity becomes a significant factor. Preferably, one should use threshold decision rules or Haar-features as classifiers, which gives small the computational complexity. In this paper, the approach to the construction of classifiers of comparable performance for the problem of detecting faces.

For the construction of the detector were studied approach involves separating detection process into two stages: construction the descriptor of image, and classification stage.

For the phase, which responsible for the classification, were considered two possibilities: a two-layer neural network, i.e. using multilayer perceptron as a “strong” classifier, and a cascade of several such networks of different size.

For the phase, which responsible for forming the descriptor, we also have investigated two possibilities. First one – fixed Haar-basis, which gives us a feature-vector of the descriptor of input image. This basis was constructed using the ADABOOST principle. The second possibility, investigated in this paper, was the construction of the basis of fewer required Haar-features, every of which more accurately reflects the object characteristics, which was obtained by using Karhunen-Loeve transform. In order to get Haar-features from eigenvectors, they have been quantized.

As a result, the classifier built with efficiency which comparable to the Haar cascade.

Key words and phrases: neural networks, ADABOOST, Haar-features, object detection, image processing.