

УДК 004.89

А. М. Голубков, Д. М. Клионский
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Применение метода каскадной редукции к решению задачи распознавания лиц

Представлен метод последовательной (каскадной) редукции пространства признаков с целью преобразования данного пространства в линейное пространство. Предложенное преобразование позволяет отделять точку в пространстве от других точек с помощью линейного дискриминанта. Показывается, что с использованием свойств разделимости, возникающих в многомерных пространствах, привести к линейному пространству можно множество точек с практически любым распределением. Показан пример применения данного результата в решении задачи распознавания лиц, в частности, в задаче идентификации человека, где каждое изображение представляется в виде вектора признаков, позволяя таким образом с помощью линейного дискриминанта определить, имеется ли искомым человек в базе данных лиц без использования перебора по кластерам, т. е. операция производится за константное время. Сравнивается качество распознавания при использовании только выходов с нейронной сети и с использованием каскадной редукции. Описывается процедура проведения экспериментов, а также детали программной реализации, в частности, выбор нейронной сети для генерирования признакового описания и способ подключения блока каскадной редукции к выходу данной сети.

Каскадная редукция, метод главных компонент, линейный дискриминант, распознавание лиц, классификация изображений

Задача классификации является одной из наиболее популярных и востребованных задач, решаемых методами машинного обучения. Задача состоит в сопоставлении входному набору данных правильных меток (лейблов) классов. Рассматривая бинарную классификацию, можно сказать, что данная задача решается путем построения кривой, разделяющей пространство данных на два класса, при этом кривая может быть сколько угодно сложной формы (в том числе и с пересечениями). Задачей алгоритмов классификации в таком случае является построение уравнения такой кривой, в частности, подбор таких параметров, чтобы кривая максимально полно описывала разницу между двумя классами. Решение данной задачи легко представить, если признаков, по которым происходит классификация, всего два. Тем не менее с ростом количества признаков, описывающих классифицируемые объекты, решение представляется все более сложным, так как растет размерность пространства, в котором необходимо строить разделяющую кривую, т. е. растет как количество самих параметров кривой, так и ее сложность.

В реальных задачах очень часто встречаются объекты, которые описываются десятками, сотнями, а иногда и тысячами параметров. Напри-

мер, в задаче классификации изображений изображения часто представляются 128-мерными или 256-мерными признаковыми векторами [1]. В задачах построения рекомендаций пользователи описываются тысячами и десятками тысяч параметров [2]. Таким образом, усложняется задача построения разделяющей кривой и, как следствие, уменьшается качество классификации и увеличивается ее время (так как необходимо строить сложную кривую). Для решения данной проблемы был предложен способ преобразования пространства признаков таким образом, чтобы любую точку в данном пространстве можно было бы отделить с достаточно высокой вероятностью от остального множества. Строить разделяющую гиперплоскость быстрее по времени, чем разделяющую многомерную кривую, при этом качество классификации не меняется или становится лучше, так как точки (объекты, описываемые признаками) становятся линейно разделяемыми.

В статье описывается способ преобразования изначально нелинейного пространства признаков в линейное с помощью каскадной редукции, а также дальнейшая классификация точек в этом пространстве с помощью линейного дискрими-

нанта. Вариант использования данного способа приводится на примере решения задачи идентификации человека по изображению лица.

Отделение точки от множества в многомерном пространстве. Теорема о разделимости утверждает [3], что при рассмотрении n -мерного шара единичного радиуса и множества точек, которые равномерно распределены в объеме данного шара, вероятность того, что одну точку этого множества можно отсечь с помощью разделяющей n -мерной гиперплоскости от остального множества растет с увеличением размерности пространства n . Данное утверждение выражается следующим неравенством:

$$P(M, n) \geq \max_{\varepsilon \in (0,1)} (1 - (1 - \varepsilon)^n) \left(1 - \frac{\rho(\varepsilon)^n}{2} \right)^m, \quad (1)$$

где $\rho(\varepsilon) = (1 - (1 - \varepsilon)^2)^{1/2}$; M – множество одинаково распределенных точек; m – мощность данного множества; n – размерность пространства.

Из данного неравенства видно, что вероятность успешного отделения (классификации) точки растет пропорционально росту размерности пространства. Данную зависимость иллюстрируют следующие графики зависимости вероятности отделения от размерности пространства (рис. 1).

Из данных графиков на рис. 1, а–г видно, что благодаря теореме о разделимости возможно рас-

считать такую размерность пространства и такое количество данных, при которых любую точку (или любой кластер точек) можно будет отделить от остального множества гиперплоскостью [4].

Таким образом, задача сводится к двум шагам: преобразование многомерного пространства признаков к равномерно распределенному внутри единичного шара множеству точек с учетом значений, полученных с помощью неравенства (1), а затем применение линейного дискриминанта для отделения требуемой точки от остального множества.

Построение каскадной редукции. Для преобразования многомерного пространства признаков в пространство признаков, равномерно распределенных на единичной сфере, применяется метод анализа главных компонент (Principal Component Analysis – PCA) [5]. При этом количество компонент для преобразования пространства вычисляется из неравенства (1), исходя из имеющего набора данных. Например, исходя из рис. 1, при размере набора данных равного 10 000 вероятность отделения точки от множества стремится к 100 % при размерности пространства равного 28 и более. Для успешного применения этого метода данные нужно предварительно центрировать:

$$X_c = X - \bar{X},$$

где $\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{128})^T$, X – исходные неотцентрированные данные и описывает среднее

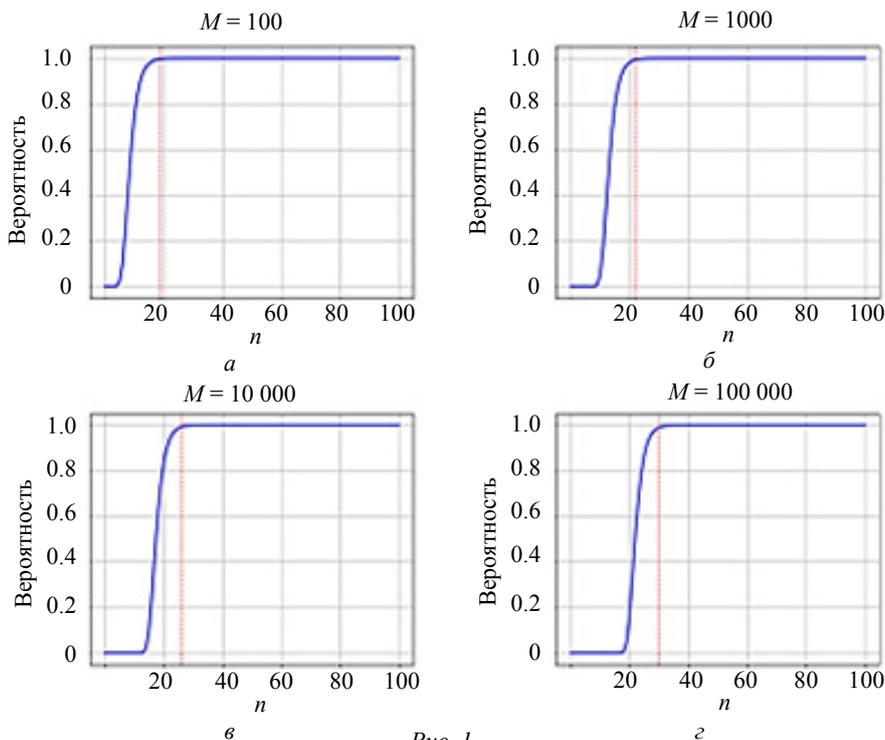


Рис. 1

значение каждого признака на всем наборе данных, после чего нормализовать:

$$X_n = \frac{X_c}{X_{std}}$$

где $X_{std} = (X_{std_1}, X_{std_2}, \dots, X_{std_N})$ и описывает стандартное отклонение каждого признака на всем наборе данных. Это необходимо для нивелирования масштаба значений признаков в признаковом векторе, т. е. для того чтобы метод PCA основывался именно на дисперсии данных, а не на их абсолютных значениях.

Далее необходимо сократить размерность центрированных и нормализованных векторов. Для решения данной задачи существуют два наиболее распространенных метода: PCA и t-SNE [6]. Последний предназначен для удобной визуализации многомерных данных с сохранением их относительной структуры, в то время как алгоритм PCA работает значительно быстрее, ориентируясь на сохранение максимальной дисперсии в данных. В алгоритме PCA сокращение размерности признаков векторов происходит за счет расчета собственных векторов и собственных чисел ковариационной матрицы данных (с размерами (количество признаков \times количество признаков)), а затем сортировки собственных векторов в соответствии с убыванием по значениям соответствующих им собственных чисел.

После центрирования, нормализации и PCA необходимо провести обеление (whitening), т. е. расположить точки нового пространства в единичной сфере. Для проведения данной операции необходимо разделить каждый полученный ранее признаковый вектор на соответствующее ему сингулярное значение (которое рассчитывается как квадратный корень из соответствующего вектору собственного значения):

$$X_{whiten} = \frac{X_{PCA}}{\sqrt{\lambda}}$$

где $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ – сингулярные значения. Таким образом получаются точки в пространстве размерностью n , которые с вероятностью, близкой к 100 %, отделены друг от друга гиперплоскостью. Общая схема преобразования данных представлена на рис. 2.

Для отделения точки от множества других точек используется линейный дискриминант Фишера [7], так как он обучается единожды на имеющемся наборе данных, работает за линейное время и не требует большого количества оперативной памяти для хранения.

Базовая модель системы линейного разделения выбранной точки $y \in M$ от любой другой точки $x \in M, x \neq y$ следующая:

$$l_y(x) = \left\langle \frac{y}{\|y\|}, x \right\rangle - \|y\| < 0, x \in M \setminus \{y\}. \quad (2)$$

Данная процедура разделения является одноитерационной и не требует какого-либо обучения. Тем не менее данная конструкция подразумевает распределение, близкое к распределению на шаре. В более общем случае, когда мы имеем дело с распределением на эллипсоиде, можно использовать линейный дискриминант Фишера:

$$l_y(x) = \left\langle \frac{w(y, M)}{\|w(y, M)\|}, x \right\rangle - c < 0; \quad (3)$$

$$w(y, M) = \Sigma^{-1} \left(y - \frac{1}{M-1} \sum_{x \neq y} x \right),$$

где y – отделяемый элемент; x – любой элемент из множества M ; Σ – ковариационная матрица множества M ; c – параметр, задающий границу «разделимости», т. е. то, при каком минимальном c элементы можно считать разделимыми. Стоит отметить, что если M центрировано на 0, то $\frac{1}{M-1} \sum_{x \neq y} x = 0$ и (3) превращается в (2) при сле-

дующем преобразовании: $x \rightarrow \Sigma^{1/2} x$.

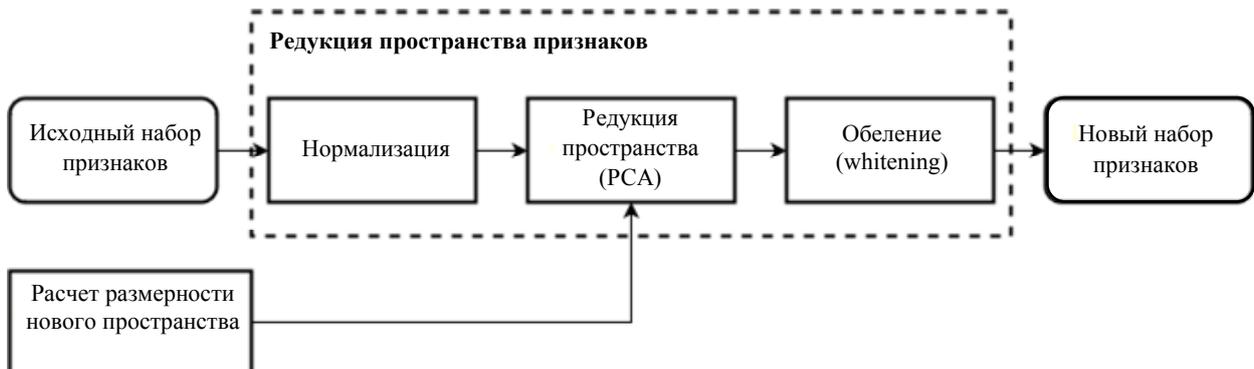


Рис. 2

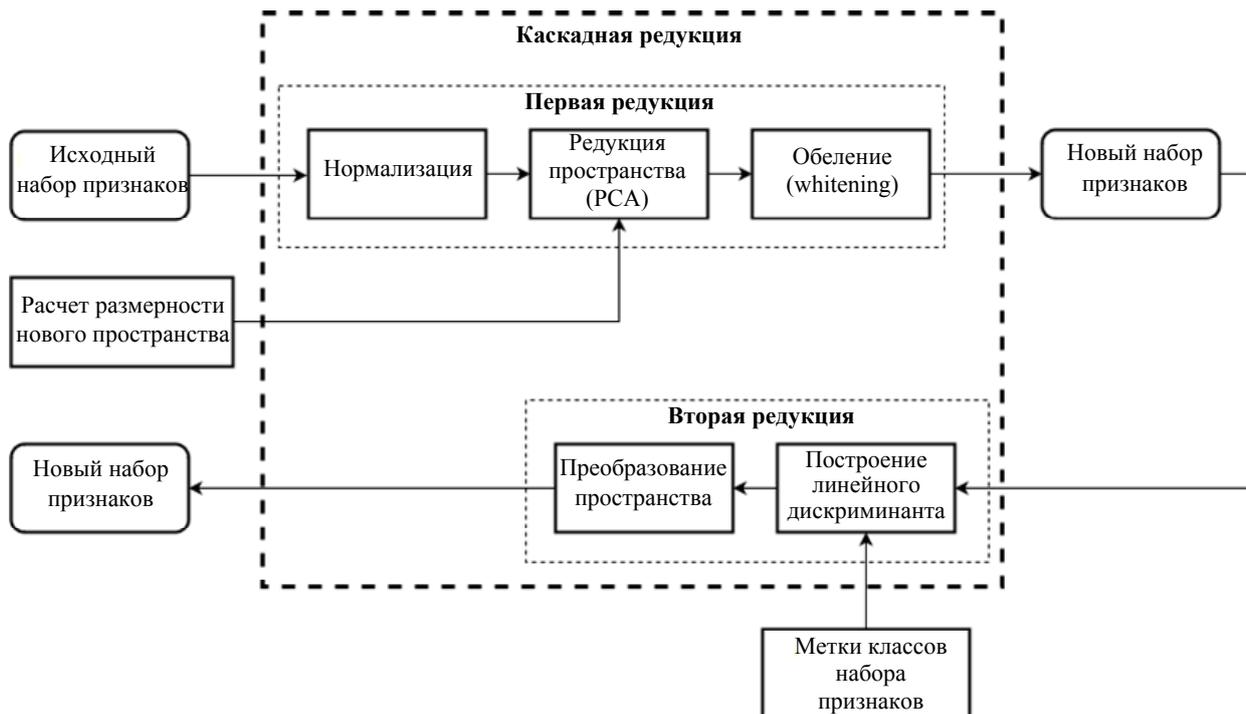


Рис. 3

Таким образом, с помощью линейного дискриминанта Фишера происходит вторая редукция пространства, с помощью которой и выполняется отделение выбранной точки от остального множества, а общая схема преобразования пространства признаков методом каскадной редукции выглядит следующим образом (рис. 3).

Применение каскадной редукции к решению задачи распознавания образов. Далее будет рассмотрен пример применения каскадной редукции в решении задачи идентификации человека по лицу. Задача состоит в следующем: имеется база данных изображений лиц. Системе идентификации на вход подается новое изображение лица (ранее отсутствовавшее в базе), после чего система определяет, какому лицу из базы данных (а точнее, какому именно человеку) соответствует данное изображение. Как видно из описания задачи, идентификация – частный вариант задачи классификации, где в качестве объектов выступают изображения лиц, а в качестве меток классов – уникальные идентификаторы людей (например, номера паспортов). Данную задачу можно решить с помощью сверточной нейронной сети: использовать ее для получения векторов признаков изображений, а в качестве последнего слоя использовать слой с функцией активации Softmax [8] для определения вероятности принадлежности изображения каждому человеку в базе. Несмотря на простоту реализации у данного

подхода есть существенный минус: при добавлении нового лица в базу данных необходимо не только доучивать сверточную нейронную сеть распознавать новое лицо, но и увеличивать размер выходного слоя на один нейрон, что потребует изменения исходного кода, описывающего архитектуру, а это может быть затратно по времени. Также, учитывая, что количество выходных нейронов нейронной сети равно количеству людей в базе данных, выходной слой может разрастись до значительных размеров, вплоть до миллионов нейронов. Это, во-первых, замедлит работу нейронной сети, во-вторых, потребует большое количество оперативной памяти для хранения, а в третьих, «размозет» вероятность принадлежности входящего изображения по всем выходным нейронам в случае «неуверенности» сети.

Таким образом, было принято решение использовать сверточную нейронную сеть только как способ получения признаков описания изображения, а далее применять классификатор на основе каскадной редукции для идентификации входящего изображения лица. Схема работы такой системы показана на рис. 4.

Как видно из схемы, перед тем, как попасть на вход каскадной редукции, изображение проходит три этапа: загрузку, выравнивание и преобразование в вектор признаков. На этапе выравнивания обнаруживается и выделяется лицо, отмечая остальную часть изображения. В данном экспе-

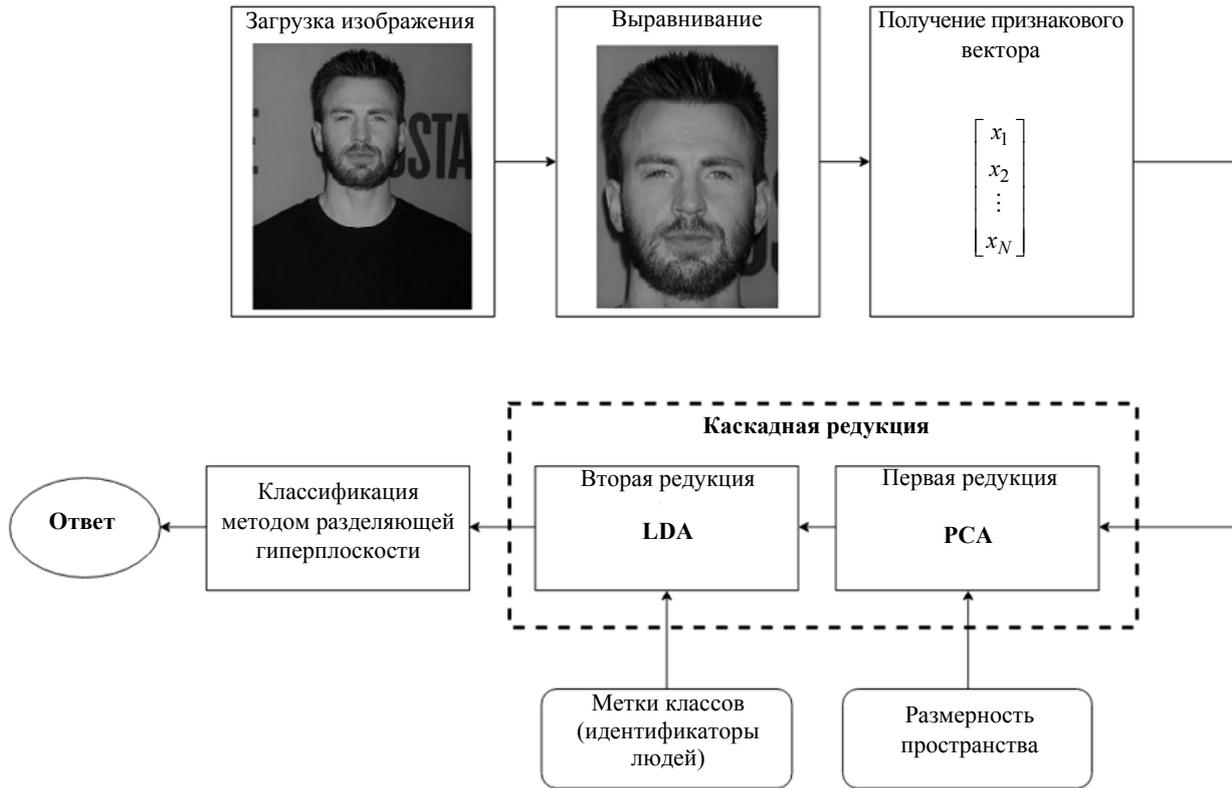


Рис. 4

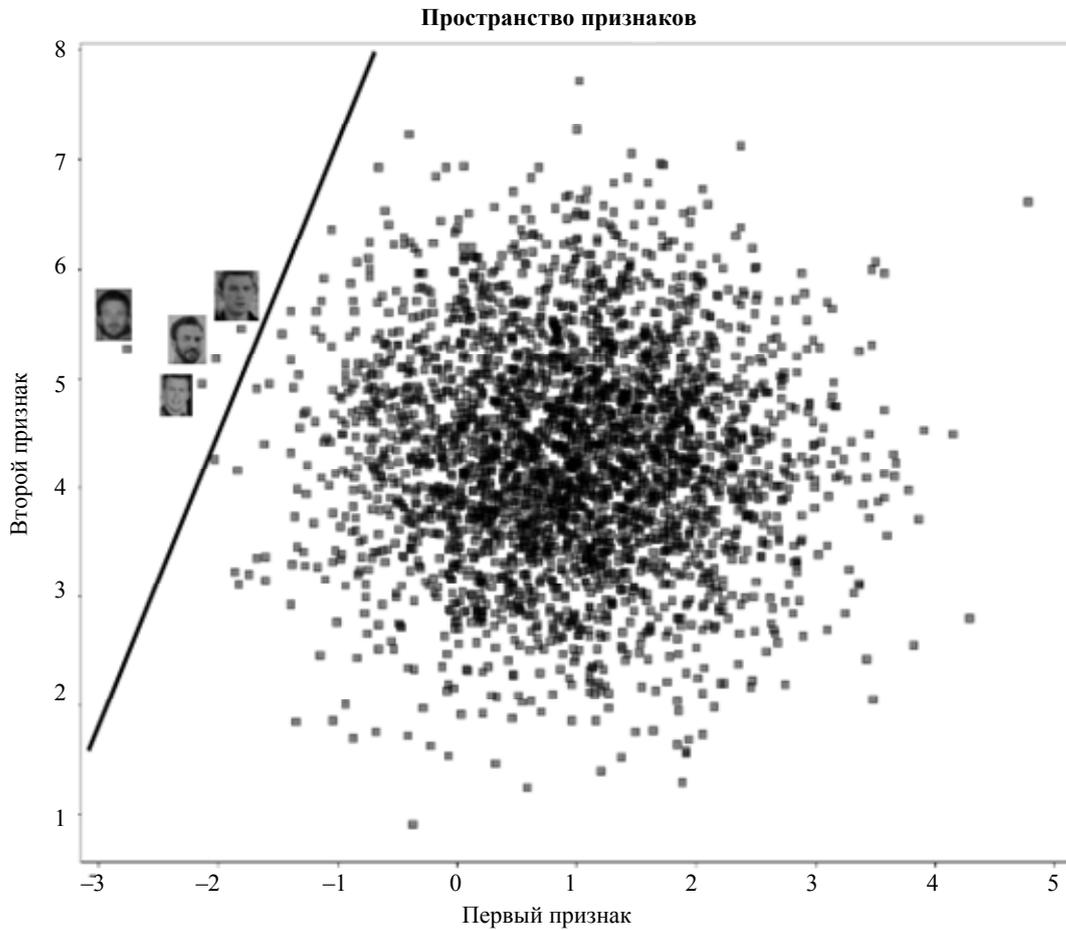


Рис. 5

Метод	Вероятность правильного ответа	Время работы, мс	Количество изображений для тренировки	Количество изображений для тестирования	Набор данных
Каскадная редукция	0.9935	2.2	8000	4000	LFW
Сверточная нейронная сеть Inception ResNet v2	0.9116	2.0	8000	4000	
Eigenfaces	0.9992	14.4	8000	4000	
Эластичные графы	0.9992	68.4	8000	4000	
Метод Виолы–Джонса	0.9970	86.6	8000	4000	

рименте для выполнения этой задачи использовалась хорошо зарекомендовавшая себя сверточная нейронная сеть МТСNN [9]. Таким образом, на вход сети подается только изображение самого лица. Далее другая сверточная нейронная сеть на основе полученного изображения генерирует вектор признаков. Для выполнения этого этапа была выбрана сеть Inception ResNet V2 [10] как одна из самых лучших на текущий момент сетей для генерации признаков и распознавания изображений. Данная сеть, как и сеть МТСNN, была загружена изначально предобученными на наборе данных ImageNet [11], считающимся стандартным набором изображений для предварительного обучения сверточных нейронных сетей. После получения признакового описания данные векторы попадают на вход каскадной редукции, которая, как было описано выше, преобразует их в линейное пространство признаков (здесь PCA – Principal Component Analysis, метод главных компонент и LDA – Linear Discriminant Analysis, линейный дискриминантный анализ). Последним этапом работы системы является непосредственно классификация с помощью разделяющей гиперплоскости, как показано на рис. 5.

На рис. 5 представлен пример работы разделяющей гиперплоскости. По оси абсцисс и оси ординат указаны значения первого и второго признаков соответственно. В проведенном эксперименте в базе данных имелось три изображения представленного человека. После загрузки изображения и проведения каскадной редукции пространство признаков выстроилось таким образом, что при построении разделяющей гиперплоскости (в данном случае для наглядности показаны только два признака и, следовательно, в качестве гиперплоскости выступает прямая) по ее левую сторону оказались все изображения данного класса, а по другую сторону – все остальные изображения. Таким образом удалось определить класс (человека), которому принадлежит лицо на поступившей на вход фотографии.

По представленной схеме было проведено множество экспериментов в режиме кроссвалидации на

четырёх фолдах с использованием базы данных LFW (Labeled Faces in the Wild) [12], широко применяющейся для тестирования алгоритмов распознавания лиц. Результаты экспериментов и сравнение с другими используемыми для решения данной задачи методами представлены в таблице.

Как видно из таблицы, метод каскадной редукции выигрывает по качеству классификации сеть Inception ResNet v2 и при этом лишь незначительно уступает в качестве остальным методам. По времени работы данный метод значительно превосходит все остальные, кроме сверточной нейронной сети, так как метод каскадной редукции использует данную сеть в качестве своей составной части.

В данной работе был представлен способ классификации признаковых описаний изображений с использованием метода каскадной редукции: последовательно проводя редукцию пространства признаков, данные в этом пространстве становятся линейно разделимыми с достаточно большой вероятностью, при этом размерность нового пространства рассчитывается исходя из количества данных для обучения и необходимой вероятности отделения, используя представленную формулу (1). Эксперименты, основанные на задаче идентификации человека по фотографии его лица, проведенные на наборе данных LFW, показали, что данный метод практически не уступает другим методам идентификации по качеству классификации, при этом значительно (на один или два порядка) обходя их по скорости работы. Метод каскадной редукции легко реализуем технически и пригоден к использованию с любыми признаковыми описаниями, полученными от любой сверточной нейронной сети при условии, что размерность вектора признаков не меньше рассчитанной по формуле (1). Дальнейшие исследования подразумевают проведение экспериментов на основе решения других типов задач классификации (верификация, поиск похожих лиц), а также на основе других типов входных данных и нейронных сетей в качестве генератора признаковых описаний.

СПИСОК ЛИТЕРАТУРЫ

1. Going Deeper with Convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // <https://arxiv.org/pdf/1409.4842.pdf> (дата обращения: 03.03.2019).
2. The Deconfounded Recommender: A Causal Inference Approach to Recommendation / Y. Wang, D. Liang, L. Charling, D. M. Blei. URL: <https://arxiv.org/pdf/1808.06581.pdf> (дата обращения: 06.03.2019).
3. One-Trial Correction of Legacy AI Systems and Stochastic Separation Theorems / A. N. Gorban, R. Burton, I. Romanenko, I. Y. Tyukin. URL: <https://arxiv.org/pdf/1610.00494.pdf> (дата обращения: 06.03.2019).
4. Correction of AI systems by linear discriminants: Probabilistic foundations / A. N. Gorban, A. Golubkov, B. Grechuk, E. M. Mirkes, I. Y. Tyukin. URL: <https://arxiv.org/pdf/1811.05321.pdf> (дата обращения: 28.11.2018).
5. Jolliffe I. T., Cadima J. Principal Component Analysis: A Review and Recent Developments // Philosophical Transactions of the Royal Society A. 2016. Vol. 374 (2065). P. 201–215.
6. Maaten L., Hinton G. Visualizing Data using t-SNE // J. of Machine Learning Research. 2008. Vol. 9. P. 2579–2605.
7. Welling M. Fisher Linear Discriminant Analysis. URL: <https://www.ics.uci.edu/~welling/teaching/273ASpring09/Fisher-LDA.pdf> (дата обращения: 14.02.2019).
8. Activation Functions: Comparison of Trends in Practice and Research for Deep Learning / C. E. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall. URL: <https://arxiv.org/pdf/1811.03378.pdf> (дата обращения: 03.03.2019).
9. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks / K. Zhang, Z. Zhang, Z. Li, Y. Qiao // IEEE Signal Processing Letters. 2016. Vol. 23, № 10. P. 1499–1503.
10. Szegedy C., Ioffe S., Vanhoucke V. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. URL: <https://arxiv.org/pdf/1602.07261.pdf> (дата обращения: 06.03.2019).
11. ImageNet: A Large-Scale Hierarchical Image Database / J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei. URL: http://www.image-net.org/papers/imagenet_cvpr09.pdf (дата обращения: 06.03.2019).
12. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments / G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller. URL: http://w.tamaraberg.com/papers/Huang_eccv2008-lfw.pdf (дата обращения: 03.03.2019).

A. M. Golubkov, D. M. Klionskiy
Saint Petersburg Electrotechnical University «LETI»

CASCADE REDUCTION METHOD APPLIED TO FACE RECOGNITION PROBLEM

Sequential (cascade) feature space reduction method to convert current space to linear space method is shown. Method is able to separate one point from other points using linear discriminant. Using the properties of separation in large spaces, every distribution of points could be convert to linear space statement is shown. It is shown that using these properties and the cascade reduction method it is possible to solve the face recognition problem in constant time, without cluster searching. Classification quality comparison of cascade reduction method with different face recognition methods is made. Experiment procedure and details of implementation is made, in particular types of convolutional neural networks that were used. The results shown by experiments says that is could be useful to use cascade reduction method to solve tasks with strong inference time requirements.

Cascade reduction, Principal Component Analysis, Linear Discriminant Analysis, face recognition, images classification