DOI: 10.51790/2712-9942-2020-1-3-1

АЛГОРИТМ КЛАССИФИКАЦИИ ЛИЦ НА ОСНОВАНИИ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ И МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

A. О. Алексанян 1 , С. О. Старков 2 , К. В. Моисеев 3

- ¹ Обнинский институт атомной энергетики филиал Национального исследовательского ядерного университета «МИФИ», г. Обнинск, Российская Федерация, andron@alexanyan.tech
- ² Обнинский институт атомной энергетики филиал Национального исследовательского ядерного университета «МИФИ», г. Обнинск, Российская Федерация, sergeystarkov56@mail.ru

 ³ ООО «Группа Аурум», г. Обнинск, Российская Федерация, K.V.Mois@gmail.com

Аннотация: данная статья затрагивает проблему распознавания лиц при решении задачи идентификации, где в качестве входных данных для последующей классификации используются векторапризнаки, полученные в результате работы сети глубокого обучения. Немногие существующие алгоритмы способны проводить классификацию на открытых наборах (open-set classification) с достаточно высокой степенью надежности.

Общепринятым подходом к проведению классификации является применение классификатора на основании порогового значения. Такой подход обладает рядом существенных недостатков, что и является причиной низкого качества классификации на открытых наборах. Из основных недостатков можно выделить следующие. Во-первых, отсутствие фиксированного порога — невозможно подобрать универсальный порог для каждого лица. Во-вторых, увеличение порога ведет к снижению качества классификации. И, в-третьих, при пороговой классификации одному лицу может соответствовать сразу большое количество классов. В связи с этим мы предлагаем использование метода главных компонент в качестве дополнительного способа понижения размерности, вдобавок к выделению ключевых признаков лица сетью глубокого обучения, для дальнейшей классификации векторов-признаков.

Геометрически применение метода главных компонент к векторам-признакам и проведение дальнейшей классификации равносильно поиску пространства меньшей размерности, в котором проекции исходных векторов будут хорошо разделимы. Идея понижения размерности логически вытекает из предположения, что не все компоненты N-мерных векторов-признаков несут значимый вклад в описание человеческого лица и что лишь некоторые компоненты образуют большую часть дисперсии. Таким образом, выделение только значимых компонентов из векторов-признаков позволяет производить разделение классов на основании самых вариативных признаков, без изучения при этом менее информативных данных и без сравнения вектора в пространстве большой размерности.

Ключевые слова: компьютерное зрение, распознавание лиц, классификация лиц, метод главных компонент, сверточные нейронные сети.

Для цитирования: Алексанян А. О., Старков С. О., Моисеев К. В. Алгоритм классификации лиц на основании сверточной нейронной сети и метода главных компонент. *Успехи кибернетики*. 2020;1(3):6-14. DOI: 10.51790/2712-9942-2020-1-3-1.

A FACE CATEGORIZATION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORKS AND PRINCIPAL COMPONENT ANALYSIS

Andron O. Alexanyan 1, Sergey O. Starkov 2, Konstantin V. Moiseev 3

- ¹ Obninsk Institute for Nuclear Power Engineering, National Research Nuclear University, Obninsk, Russian Federation, andron@alexanyan.tech
- ² Obninsk Institute for Nuclear Power Engineering, National Research Nuclear University, Obninsk, Russian Federation, sergeystarkov56@mail.ru
 - ³ Aurum Group, OOO, Obninsk, Russian Federation, K.V.Mois@gmail.com

Abstract: the study objective is face recognition for identification purposes. The input data to be classified are attribute vectors generated by a deep learning neural network. The few existing algorithms can perform sufficiently reliable open-set classification.

The common approach to classification is using a classification threshold. It has several disadvantages leading to the low quality of open-set classifications. The key disadvantages are as follows. First, there is no set threshold: it is impossible to find a common threshold suitable for every face. Second, the higher the threshold, the lower the quality of classification. Third, with the threshold classification more than one class can match a face.

For this reason, we proposed to apply the principal component analysis as an extra dimensionality reduction tool besides identifying the key face attributes by a deep learning neural network for subsequent classification of the attribute vectors. In geometric terms, the principal component analysis application to attribute vectors with subsequent classification is similar to a search for a low-dimension space where the projections of the source vectors can be easily separated. The dimensionality reduction concept is based on the assumption that not all the components on N-dimensional attribute vectors are relevant for the human face representation, and only some of them produce the larger part of the dispersion. Therefore, by selecting only the relevant components of the attribute vectors we can separate the classes using the most variable attributes while skipping the less informative data and not comparing the vectors in a high-dimensional space.

Keywords: computer vision, face recognition, face classification, principal component analysis, convolutional neural networks.

Cite this article: Alexanyan A. O., Starkov S. O., Moiseev K. V. A Face Categorization Algorithm Based on Convolutional Neural Networks and Principal Component Analysis. *Russian Journal of Cybernetics*. 2020;1(3):6–14. DOI: 10.51790/2712-9942-2020-1-3-1.

Введение

Область компьютерного зрения, одного из направлений современного искусственного интеллекта, последние годы находится в стадии быстрого развития [1–5]. Это обусловлено многими факторами, один из которых — прикладной характер данной области. Большинство наработок являются полностью применимыми для реального бизнеса и способствуют его улучшению и росту. Как следствие, крупнейшие компании инвестируют огромные средства в развитие отрасли компьютерного зрения, что неминуемо приводит к постоянному совершенствованию ее подходов и алгоритмов. Однако вопрос качества результатов стоит остро даже несмотря на постоянные доработки и исследования. В частности, это актуально для одного из наиболее активно развиваемых направлений компьютерного зрения — распознавания лиц. Принятые в индустрии подходы к классификации лиц показывают хорошее качество классификации на тестовых наборах данных, однако при решении реальных задач ситуация другая: качество классификации резко падает, что не позволяет полномасштабно внедрять и использовать созданные программные продукты.

Ниже рассматривается алгоритм классификации лиц, повышающий качество классификации при решении задач распознавания лиц. Достижение поставленной цели предполагает последовательное выполнение следующих задач:

- 1. Ознакомиться с существующими подходами и алгоритмами классификации лиц, проанализировать их слабые места.
 - 2. Разработать алгоритм классификации, повышающий качество обработки.
 - 3. Реализовать разработанный алгоритм с помощью языка программирования Python.

В качестве объектов классификации выступают результаты работы сетей глубокого обучения [2]. На выходе таких сетей получается числовой вектор (так называемый вектор-признак), абстрактно характеризующий изображенный на фотографии объект (в нашем случае — человеческое лицо) [6]. Соответственно, задача классификации лиц сводится к задаче классификации векторов-признаков.

Задачи из области распознавания лиц можно разбить на два класса: класс задач по идентификации лиц и класс задач по верификации лиц [5]. При решении задач верификации на вход программы подается два лица и необходимо найти ответ на вопрос: «Являются ли предъявленные лица одним и тем же человеком?» При проведении тестирований алгоритмов распознавания лиц решается именно задача верификации: из тестового набора фотографий последовательно выбираются пары фотографий, производится верификация и происходит расчет итоговой точности алгоритма. При решении задач идентификации производится гораздо более сложная обработка: отдельно взятое лицо сравнивается со всей базой известных лиц и необходимо найти ответ на вопрос: «Соотносится ли исследуемое лицо с одним из известных лиц и если да, то с каким?» В реальных условиях происходит решение задачи идентификации: новые лица сверяются с уже существующей базой лиц. Задача идентификации является более сложной по сравнению с задачей верификации — именно по этой причине качество распознавания лиц падает по сравнению с результатами тестов при решении реальных задач.

В работах по распознаванию лиц большинство исследователей использует классификацию с помощью фиксированного порогового значения — как для идентификации, так и для верификации. Такой подход и обуславливает резкое падение качества при переходе к реальным задачам. В случае верификации происходит простое сравнение расстояния между векторами-признаками, характеризующими каждое лицо, и неким заданным порогом (который обычно выбирается на основании заранее проведенных экспериментов). В случае идентификации же ситуация осложняется сразу двумя аспектами:

- 1. Каждое исследуемое лицо сравнивается со всей известной базой лиц, что подразумевает много-кратное сравнение расстояния между векторами с неким порогом. Очевидно, что одновременно может попасться несколько людей из банка лиц, проходящих пороговое значение. Однако одно лицо не может сразу соотноситься с несколькими известными людьми. Соответственно, возникает неоднозначность выбора и повышается вероятность ошибки.
- 2. В практических задачах редко представляется возможным провести предварительный эксперимент с целью выявления оптимального порога классификации: это требует дополнительных денежных, человеческих и временных ресурсов. В связи с этим порог классификации выбирается исходя из эмпирических соображений исследователя, что не позволяет минимизировать количество ошибок. Стоит отметить, что для каждой задачи подбирать пороговое значение рекомендуется индивидуально.

В связи с вышеизложенным актуальность разработки алгоритма, отличного от классификации фиксированным порогом, становится явной: это позволит повысить качество распознавания лиц как в уже существующих системах, так и в будущих разработках.

Описание алгоритма

Разработанный алгоритм основан на методе главных компонент (англ. principal component analysis, PCA) [1] — одном из основных способов понижения размерности. Идея о снижении размерности векторов-признаков вытекает из сущности самого объекта исследования: на выходе сверточной нейронной сети мы получаем вектора-признаки большой размерности, абстрактным образом описывающие человеческое лицо. Закономерно возникает мысль о том, что не все компоненты векторапризнака несут в себе существенный вклад в описание лица: некоторые элементы могут быть малоинформативными в терминах поставленной задачи.

Другой ключевой идеей предлагаемого подхода является его итеративность: главные компоненты каждого известного лица можно многократно пересчитывать при нахождении нового экземпляра класса (т. е. если найдено еще одно изображение данного человека, то главные компоненты пересчитываются с учетом новой информации).



Рис. 1. Пример фотографий из сгенерированного датасета

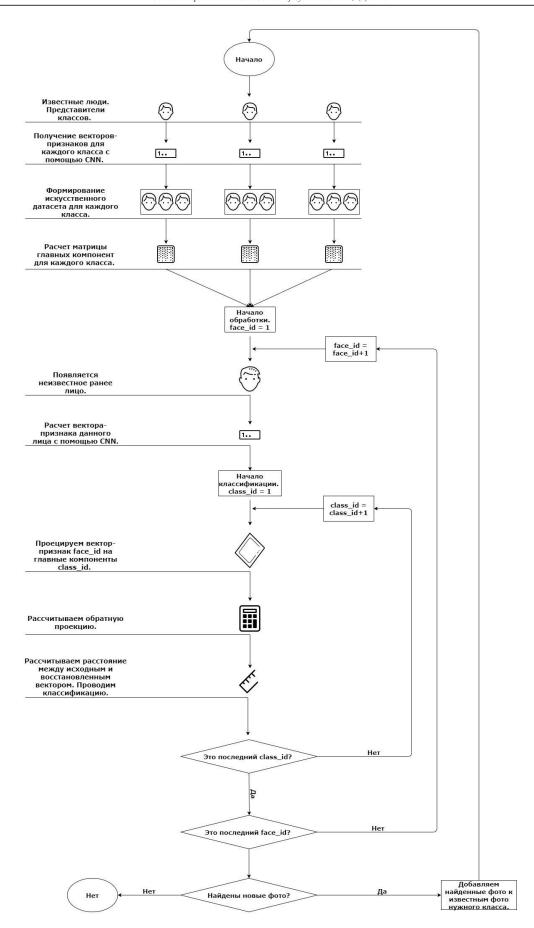


Рис. 2. Схема алгоритма

При реализации данного алгоритма возникает потребность в решении еще одной задачи: для эффективного применения метода главных компонент необходимо как можно больше разных фотографий одного и того же человека — это позволяет точнее вычислить компоненты, по которым наблюдается максимальный разброс дисперсий. Однако в большинстве случаев в базе лиц хранится лишь одна (максимум несколько) фотографий каждого пользователя. Для решения данной проблемы была разработана процедура генерации искусственного датасета. Основной принцип заключается в том, что для каждой фотографии пользователя (как хранящейся в банке лиц, так и найденной на очередной итерации разработанного алгоритма классификации) создаются множественные копии с различными преобразованиями лица (размытие, обрезка, поворот, добавление дополнительных элементов и так далее). Процедура генерации искусственного датасета позволяет получить достаточно первичного материала для использования метода главных компонент на первой итерации, при этом не привлекая дополнительные денежные средства, а также человекоресурсы.

Стоит отметить, что виды преобразований исходного лица, а также количество копий реализованы в виде параметров алгоритма. Т. е. исследователь может менять данные параметры в зависимости от задачи, от вычислительных мощностей и особенностей аппаратного обеспечения.

На рис. 2 схематично изображен пример системы, основанной на разработанном алгоритме. Вначале происходит обработка известных лиц с помощью нейронной сети, генерация искусственного датасета и расчет главных компонент для каждого класса (т. е. для каждого известного человека). Далее, если системе предъявляется неизвестное лицо, сначала происходит обработка этого лица нейронной сетью и получения вектора-признака. После этого производится проецирование вектора-признака неизвестного лица на матрицу главных компонент каждого класса, расчет обратной проекции и принятие решения об отнесении к тому или иному классу на основании расстояния между исходным и восстановленным вектором. Таким образом обрабатываются все неизвестные лица. Если неизвестных лиц больше нет и для известных классов были найдены новые фото, происходит пополнение датасета и перерасчет главных компонент. Если на текущей итерации новых фото не было найдено, то перерасчет не происходит и система либо ждет новые неизвестные лица, либо заканчивает работу (в зависимости от решаемой задачи).



Рис. 3. Ошибка 1 рода. Нужно было найти 3 фотографии, а найдено лишь 2.

Оценка преимуществ и области применения

Применение описанного алгоритма обладает рядом преимуществ по сравнению с классификацией пороговым значением, а также многими другими алгоритмами классификации, которые потенциально можно использовать для решения поставленной задачи. Алгоритм позволяет избавиться от лишнего «информационного шума» в векторах-признаках, выделить самые главные характеристики (в терминах рассеяния дисперсии) и в результате повысить качество классификации. За счет своей итеративности алгоритм позволяет многократно корректировать результаты классификации, оставляя только те лица, принадлежность к конкретной группе которых подтверждается на каждой новой итерации алгоритма. Технология генерации искусственного датасета, являющаяся одной из ступеней разработанного алгоритма, позволяет без дополнительных усилий получать размеченные наборы фотографий, что несет большую ценность для индустрии компьютерного зрения. Разработанный алгоритм инвариантен к тому, какая сверточная нейронная сеть служит для получения векторов-признаков. Единственное условие — размерность выходного вектора-признака должна быть N * 1, где N - любое натуральное число. Это предположительно позволяет применять алгоритм практически для любых задач, связанных с распознаванием лиц (как для верификации, так и для идентификации).

Оценивая точность любого алгоритма классификации, необходимо учитывать два вида возможных ошибок:

- 1. Лицо классифицировано неправильно, т. к. относится к данному классу, а алгоритм отнес его к другому классу. Такую ситуацию будем называть ошибкой первого рода -FalsePositive (рис. 3).
- 2. Лицо классифицировано неправильно, т. к. относится к другому классу, а алгоритм отнес его к данному классу. Такую ситуацию будем называть ошибкой второго рода -FalseNegative (рис. 4).



Рис. 4. Ошибка 2 рода. Нужно было найти 2 фотографии, а найдена лишь 1: вторая найдена неверно.

Таблица 1

Сравнительная статистика алгоритмов классификации

Алгоритм	Найдено верно, шт.	FP , шт. FN, шт.		Accuracy, %	Precision, %	Recall, %	
Порог (0.5)	22	116	0	15,9	15,9	100	
Порог (0.6)	29	99	7	26,8	28,2	84,7	
Порог (0.65)	49	89	33	28,6	35,5	59,7	
Алгоритм (1 шаг)	86	52	1	61,9	62,3	98,8	
Алгоритм (2 шаг)	95	43	1	68,3	68,9	98,9	
Алгоритм (3 шаг)	98	40	1	70,5	71	99	

Ниже представлена статистика тестирования алгоритма пороговой классификации с разными пороговыми значениями (таблица 1), а также разработанного алгоритма на первых двух итерациях. Тестирование проводилось на датасете из 951 фотографии, на которых изображено 4275 лиц. Всего было 8 классов (т. е. 8 известных пользователей) и найти необходимо было 138 лиц. Для генерации векторов-признаков использовалась сверточная нейронная сеть SphereFace [3]. Для генерации искусственного датасета использовались операции поворота изображения, добавления очков, искажения и размытия с помощью библиотеки OpenCV для языка программирования Python [7]. На каждый вид операции создавалось по 3 копии с различной степенью преобразований. Для комплексной оценки алгоритма использовалось 3 метрики: качество (accuracy), точность (precision) и полнота (recall) —

$$Accuracy = rac{TP}{TP + FP + FN},$$

$$Precision = rac{TP}{TP + FP},$$

$$Recall = rac{TP}{TP + FN},$$

где TP (True Positive) — количество верно найденных фото; FP (False Positive) — количество ошибок 1 рода; FN (False Negative) — количество ошибок 2 рода.

Из приведенной статистики хорошо видно сразу несколько слабых мест классификации пороговым значением.

Во-первых, при слишком маленьком пороге (0.5) алгоритм находит недостаточно нужных фотографий, а при слишком высоком (0.65) сильно возрастает количество ошибок 2 рода. По мере увеличения порога ситуация будет только ухудшаться: качество классификации будет уменьшаться за счет роста количества ошибок второго рода.

Во-вторых, из первого пункта вытекает, что повышать качество классификации за счет увеличения порога практически невозможно: увеличение порога приведет к снижению качества классификации.

В-третьих, на больших наборах данных практически невозможно подобрать фиксированный порог, который бы позволял произвести оптимальную классификацию. Иллюстрацию этого явления можно увидеть в таблице 2. В приведенной таблице по отдельности рассмотрена статистика классификации по 3 пользователям (т. е. по 3 классам) в разрезе классификации с двумя порогами -0.5 и 0.65.

Из таблицы 2 видно, что порог 0.5 слишком низкий для Пользователей 1 и 3, однако достаточно много фотографий Пользователя 2 нашлись верно даже с этим порогом (притом почти без ошибок второго рода). Однако, увеличив порог до 0.65, мы наблюдаем рост качества классификации у Пользователей 1 и 3, но количество ошибок 2 рода у Пользователя 2 резко выросло.

Сравнительная статистика алгоритмов классификации

Алгоритм	Пользователь 1		Пользователь 2			Пользователь 3			
	Верно	FP	FN	Верно	FP	FN	Верно	FP	FN
Порог 0.5	0	5	0	5	16	0	1	4	0
Порог 0.65	4	1	0	15	6	11	4	1	0

Проводя статистический анализ на основе разработанного алгоритма, можно наблюдать прямо противоположную ситуацию. Уже на первой итерации алгоритм показывает в несколько раз более высокое качество классификации. Помимо этого, есть еще несколько моментов, указывающих на эффективность алгоритма.

Во-первых, в отличие от классификации порогом, разработанный алгоритм уже на первой итерации нашел хотя бы по одной фотографии для каждого класса (при классификации порогом даже с порогом 0.65 для одного класса не было найдено ни одной фотографии). Это позволяет произвести перерасчет главных компонент для каждого класса и запустить вторую итерацию алгоритма, точность на которой выше максимальной точности порогового классификатора на 39,7 процентных пункта (в соответствии со статистикой из таблицы 1). Стоит отметить, что для задач компьютерного зрения улучшением считается даже процентное изменение точности распознавания.

Во-вторых, также стоит обратить внимание на стабильно низкий показатель ошибки 2 рода в разработанном алгоритме. В течение всех трех итераций разработанного алгоритма количество ошибок 2 рода оставалось неизменным на уровне 1 фотографии. Таким образом, все найденные лица (а на второй итерации их было уже 98) классифицированы правильно. Для сравнения, при увеличении порога с 0.6 до 0.65 в алгоритме пороговой классификации произошел рост количества ошибок 2 рода с 7 до 33.

Сравнение всех трех метрик классификации отражено на рисунке 5 в разрезе по алгоритмам классификации и метрикам.

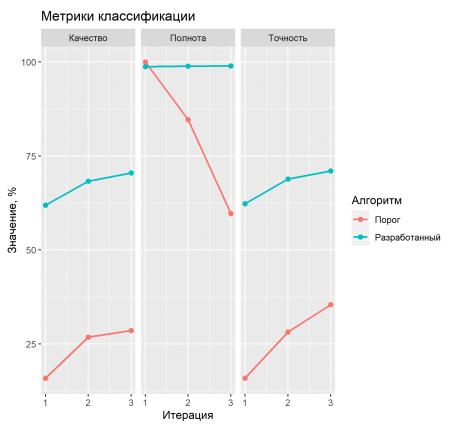


Рис. 5. Метрики классификации

Заключение

На основе метода главных компонент реализован алгоритм, повышающий качество классификации лиц при решении задач распознавания.

С помощью вспомогательной процедуры генерации искусственного датасета удалось улучшить качество поиска главных компонент и, как следствие, повысить качество задачи распознавания.

Алгоритм реализован на языке программирования Python с использованием библиотеки OpenCV [7], фреймворка PyTorch [4], СУБД MySQL [8].

ЛИТЕРАТУРА

- 1. Айвазян С. А. Теория вероятностей и прикладная статистика. М., 2001.
- 2. Ле Мань Xa. Сверточная нейронная сеть для решения задачи классификации. *Труды МФТИ*. 2016;8(3):91-97.
- 3. Liu W., Wen Y., Yu Z., Li M., Raj B., Song L. Sphereface: Deep Hypersphere Embedding for Face Recognition. *CoRR*. 2017. Vol. abs/1704.08063. Режим доступа: http://arxiv.org/abs/1704.08063.
- 4. PyTorch Documentation. Режим доступа: https://pytorch.org/docs/stable/index.html.
- 5. Günther M., Hu P., Herrmann C., Chan C. H., Jiang M. et al. Unconstrained Face Detection and Open-Set Face Recognition Challenge. *IEEE International Joint Conference on Biometrics (IJCB)*. 2017. Vol. abs/1708.02337. Режим доступа: http://arxiv.org/abs/1708.02337.
- Sumit Saha. A Comprehensive Guide to Convolutional Neural Networks the ELI5 way. Режим доступа: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5way-3bd2b1164a53.
- 7. OpenCV-Python Tutorials. Режим доступа: https://docs.opencv.org/master/d6/d00/tutorial py root.html.
- 8. Документация по MySQL. Режим доступа: http://www.mysql.ru/docs/.

REFERENCES

- 1. Aivazyan S. A. Teoriya veroyatnostei i prikladnaya statistika. M., 2001.
- 2. Le Manh Ha. Convolutional Neural Networks for Classification. Proceedings of MIPT. 2016;8(3):91–97.
- 3. Liu W., Wen Y., Yu Z., Li M., Raj B., Song L. Sphereface: Deep Hypersphere Embedding for Face Recognition. *CoRR*. 2017. Vol. abs/1704.08063. Available at: http://arxiv.org/abs/1704.08063.
- 4. PyTorch Documentation. Available at: https://pytorch.org/docs/stable/index.html.
- 5. Günther M., Hu P., Herrmann C., Chan C. H., Jiang M. et al. Unconstrained Face Detection and Open-Set Face Recognition Challenge. *IEEE International Joint Conference on Biometrics (IJCB)*. 2017. Vol. abs/1708.02337. Available at: http://arxiv.org/abs/1708.02337.
- 6. Sumit Saha. A Comprehensive Guide to Convolutional Neural Networks the ELI5 way. Available at: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.
- 7. OpenCV-Python Tutorials. Available at: https://docs.opencv.org/master/d6/d00/tutorial_py_root.html.
- 8. MySQL Documentation. Available at: http://www.mysql.ru/docs/.