

*Петрин Дмитрий Александрович, студент-магистр, Калужский филиал
ФГБОУ ВО «Московский государственный технический университет имени
Н.Э. Баумана (национальный исследовательский университет)»*

*Белов Юрий Сергеевич, к.ф.-м.н., доцент, Калужский филиал ФГБОУ ВО
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»*

ТЕХНОЛОГИИ РАБОТЫ С РЕДУЦИРОВАННЫМИ ОБУЧАЮЩИМИ ДАНЫМИ В ЗАДАЧАХ КЛАССИФИКАЦИИ

Аннотация: Данная статья посвящена обзору современных подходов, используемых для решения задач классификации изображений в условиях ограниченного набора обучающих данных. Приводятся ограничения традиционного подхода к классификации – свёрточной нейронной сети (CNN). Рассматриваются архитектурные особенности, а также преимущества и недостатки каждого из алгоритмов. Делается вывод о перспективах использования данных алгоритмов.

Ключевые слова: нейронная сеть, КААРМА, машина опорных векторов, k-ближайших соседей, классификация.

Annotation: This article is devoted to a review of modern approaches used to solve the problems of image classification in a limited set of training data. The limitations of the traditional classification approach, convolutional neural network (CNN), are given. The architectural features, as well as the advantages and disadvantages of each of the algorithms are considered. The conclusion is drawn about the prospects of using these algorithms.

Keywords: neural network, KAARMA, support vector machine, k-nearest neighbors, classification.

В настоящее время задача обработки и анализа изображений является одной из самых сложных и востребованных. Данная задача решается различными методами машинного обучения при помощи нейронных сетей самых разных архитектур. Нейронные сети обучаются на тренировочном наборе данных, а затем происходит оценка точности модели на тестовой выборке.

Однако многие алгоритмы, такие как CNN (свёрточные нейронные сети) имеют ряд ограничений. Точность CNN в значительной степени определяется размером обучающего набора данных [4, с. 100-106]. Чтобы добиться высокой точности, нужны объемы данных большого размера [1; 3]. К сожалению, не всегда возможно создать такой набор данных (например, из-за редкости наблюдаемых явлений или событий). К тому же, необходимость разделения на обучающую и тестовую выборки также снижает размер возможного датасета для обучения.

Именно поэтому существуют методы классификации изображений, которые обладают достаточно хорошей точностью при обучении на ограниченном наборе данных. В данной статье будут рассмотрены особенности, достоинства и недостатки современных подходов, таких как: алгоритм К-ближайших соседей (K-NN), машина опорных векторов (SVM), а также различные подходы, основанные на рекуррентных нейронных сетях.

Алгоритм К-ближайших соседей (K-NN) классифицирует выборки на основе k ближайших выборок [5] в обучающем наборе данных (рис. 1).

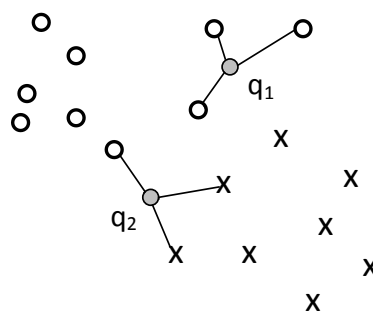


Рис.1 Пример классификации на основе 3-х соседей

Данный алгоритм имеет две стадии: первая — это определение ближайших соседей, а вторая - определение класса на основе использования этих соседей. К ближайших соседей выбираются на основе так называемой метрики расстояния. Эта метрика показывает на сколько удалены между собой различные классы и элементы, относящиеся к этим классам. Метрика может задаваться самыми различными способами. Главное, чтобы она удовлетворяла 4 критериям метрики. Наиболее популярными метриками являются:

- евклидова метрика $D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- косинусное расстояние

$$D(X, Y) = \cos(\hat{X}, \hat{Y}) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}}$$

- манхэттенское расстояние $D(X, Y) = \sum_{i=1}^n |x_i - y_i|$,
- расстояние Минковского $D(X, Y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p}$ [5].

K-NN опирается только на локальную информацию относительно ранее обработанных образцов. Однако K-NN не проверяет признаки, используемые для классификации, а только лишь применяет метрику расстояния. Поэтому главным недостатком данного алгоритма является то, что K-NN очень чувствителен к преобладающим признакам, даже если они не имеют значения, что существенно сказывается на точности классификации.

Машина опорных векторов (SVM) – широко известный алгоритм, который позволяет решить задачу классификации (в том числе изображений).

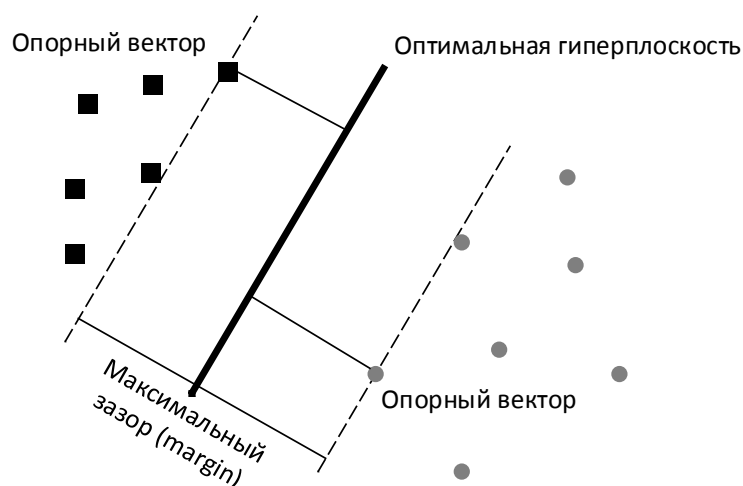


Рис.2 Пример работы алгоритма SVM

Основная идея метода SVM состоит в том, чтобы перевести исходные входные векторы в пространство большей размерности и найти разделяющую гиперплоскость (рис. 2) с максимальным зазором (англ. margin) в этом пространстве (так называемая «оптимальная» гиперплоскость) [6, с. 103-113]. Параллельно «оптимальной» гиперплоскости строятся две гиперплоскости по обеим сторонам. Разделяющей будет та гиперплоскость, которая максимизирует зазор до двух параллельных гиперплоскостей. SVM работает в предположении, что чем больше зазор между этими гиперплоскостями, тем меньше будет средняя ошибка классификатора. Точки, являющиеся ближайшими к параллельным гиперплоскостям, называются опорными векторами.

Простейший случай для SVM - линейный, когда гиперплоскость принадлежит пространству входных данных x . В этом случае пространство гипотез является подмножеством всех гиперплоскостей вида: $f(x) = w \cdot x + b$ [5]. В самом общем случае SVM находит гиперплоскость в пространстве, отличном от пространства входных данных. Эта гиперплоскость принадлежит пространству признаков, индуцированному ядром K (ядро определяет скалярное произведение в этом пространстве). Через ядро K пространство гипотез определяется как набор «гиперплоскостей» в пространстве признаков, индуцированных K . Его также можно рассматривать как набор функций в

Рис.3 Сканирование изображения пиксельной RNN

Она сканирует изображение по одной строке по одному пикселю за раз (рис. 3) в каждой строке. Впоследствии она предсказывает условные распределения по возможным значениям пикселей. Распределение пикселей изображения вычисляется как произведение условных распределений, и эти значения распределяются по всем пикселям изображения. Цель состоит в том, чтобы назначить вероятность $p(x)$ каждому пикселю ($n \times n$) изображения. Это можно сделать, записав вероятность пикселя x_i как $p(x) = \prod_{i=1}^{n^2} p(x_i | x_1, \dots, x_{i-1})$. Это вероятность i -го пикселя с учетом вероятности всех ранее рассмотренных пикселей. Кроме того, цвет пикселя x_i определяется совместно всеми тремя цветовыми каналами: красным, зеленым и синим (RGB). Таким образом, каждый цвет зависит от других цветов, а также от ранее обработанных пикселей.

Чаще всего для построения такой сети используются слои RowLSTM и Diagonal BiLSTM. Благодаря архитектуре LSTM (долгой краткосрочной памяти), пиксельная RNN может моделировать локальную информацию, полученную от изображения, такую как текстура или цвет, при этом понимая глобальный контекст рассматриваемого пикселя, например, принадлежит ли пиксель тому или иному объекту.

Схожим образом работает и RNN-CNN, которая является рекуррентной свёрточной нейронной сетью, оптимизированной для классификации изображений. Архитектура сети представлена на рис. 4. Она включает в себя две крупные части: часть CNN извлекает семантические представления из изображений; часть RNN моделирует отношения между изображениями и метками, а также зависимость между метками.

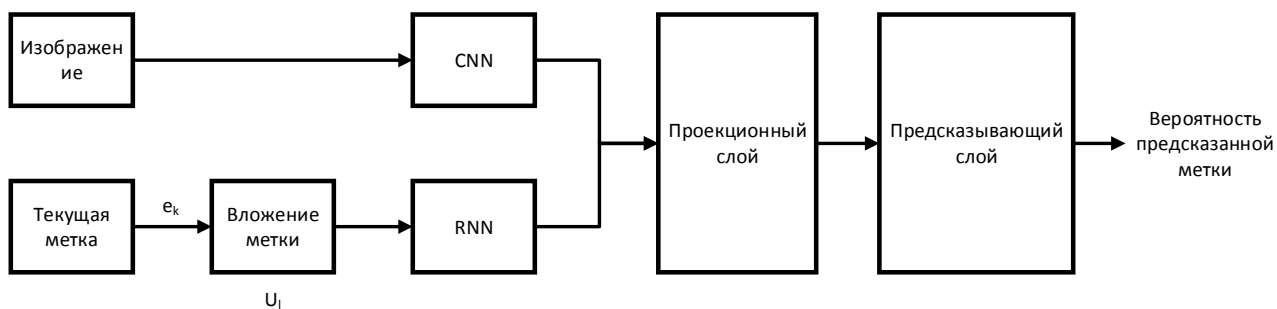


Рис. 4 Архитектура сети RNN-CNN

Для представления изображения (в удобной для обработки форме) используется свёрточная нейронная сеть, а рекуррентный слой извлекает информацию о ранее предсказанных метках. Вероятность выходной метки вычисляется в соответствии с представлением изображения и выходным сигналом рекуррентного слоя. RNN используется в качестве компактного, но мощного способа представления зависимости совместного появления меток (когда на одном изображении присутствуют объекты классификации нескольких распознаваемых типов) в этом пространстве. Для этого данный слой использует скрытые состояния, принимая на входе вложенные метки. Метка k представляется как вектор $e_k = [0, \dots, 0, 1, 0, \dots, 0]$, k -й элемент которого равен 1, а остальные равны 0. Вложенная метка получается путем умножения вектора e_k на специальную матрицу вложенных меток U_l , где k -я строка U_l — это вложение метки k .

В отличие от обычной CNN, RNN-CNN способна адаптивно фокусировать свое внимание на разных частях изображения в зависимости от других объектов, которые были идентифицированы [3]. В результате, такая сеть лучше понимает контекст, используемый для точной классификации нескольких объектов.

Несмотря на то, что рекуррентные системы являются достаточно успешными в обработке и классификации изображений (в частности, при ограниченном размере обучающей выборки), они обладают очень большой вычислительной сложностью. Тому есть несколько причин. Во-первых, градиент ошибки должен быть взят во времени, что требует многократных

вычислений ошибок и обновлений весов. Во-вторых, большинство рекуррентных систем требуют многократного просмотра последовательности изображений для достижения полностью оптимального решения. Этот недостаток в сочетании с обратным распространением градиента ошибки во времени экспоненциально увеличивает временную сложность обучения. Однако, если последовательности изображений можно вводить в рекурсивную систему ограниченное число раз и приблизиться к обучению за один проход, то возможно использовать логический вывод рекуррентных систем и решить проблему дорогостоящих вычислений.

Именно на данной идее и основывается алгоритм КААРМА (ядерный алгоритм адаптивного авторегрессионного скользящего среднего). КААРМА представляет собой рекуррентную сеть (рис. 5). Она опирается на предыдущие выборки, чтобы влиять на обработку последующих. По мере анализа новых выборок, алгоритм адаптируется для изучения динамики последовательности фрагментов изображения, следуя по траектории этой последовательности до тех пор, пока не достигнет решения. КААРМА эффективно извлекает пространственную информацию в изображении за счет использования ядерных функций (фильтров). Благодаря этому можно получать лучшие признаки, чтобы в дальнейшем находить более сложные решения [2].

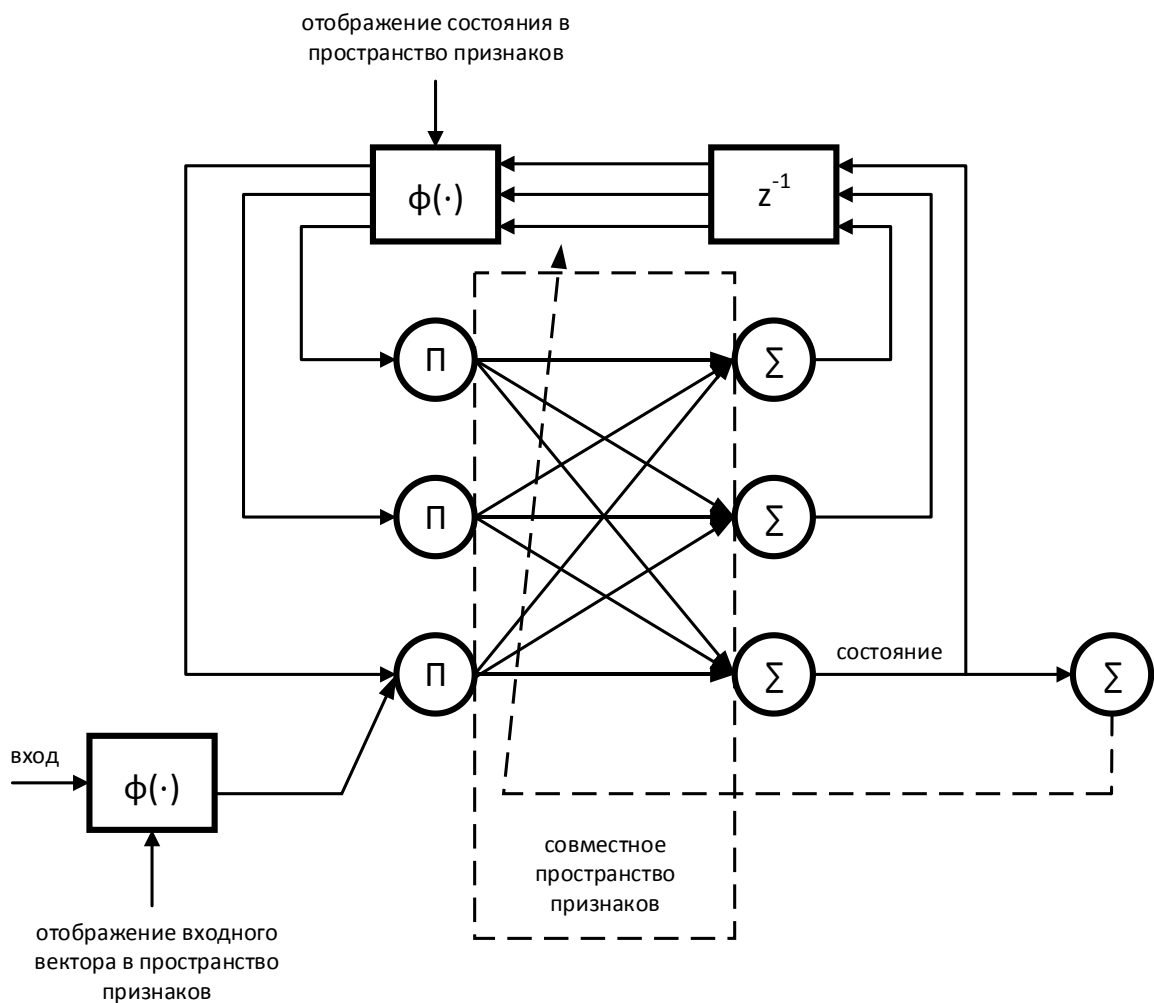


Рис. 5 Сеть КААРМА

Кроме того, алгоритм хорошо оптимизирован при обучении на небольшом количестве выборок. Данный алгоритм показал хорошую точность (более 90%) на наборе данных JAFFE (выражения лиц японских женщин) даже при небольших размерах обучающих данных [2]. Тем не менее, главным недостатком алгоритма КААРМА является то, что он сравнительно новый. А значит потребуются провести дополнительные исследования, чтобы изучить возможности применения алгоритма для разных типов изображений. Также необходимо понять, как КААРМА решает другие задачи, связанные с классификацией изображений.

Таким образом, подводя итог, можно сделать вывод, что каждый из рассмотренных алгоритмов имеет свои достоинства и недостатки. Каждый из них можно применять в областях, где создание датасетов затруднено редкостью тех или иных данных, например в медицине. Но не стоит забывать, что

рассмотренные алгоритмы не являются универсальными. Они хорошо себя показывают только на ограниченных по размеру данных. При наличии достаточного объема обучающих выборок точность некоторых из них может уступать все той же CNN. С этой точки зрения из всех рассмотренных алгоритмов наиболее перспективными являются рекуррентные нейронные сети, основанные на алгоритме КАARMA. Поэтому, при должном изучении и дальнейших оптимизациях КАARMA имеет хорошие перспективы для широкого распространения и внедрения в различные системы.

Библиографический список:

1. Cudic M., Department of Electrical and Computer Engineering University of Florida Gainesville, FL U.S.A Using a Recurrent Kernel Learning Machine for Small-Sample Image Classification. – URL: <https://ufdc.ufl.edu/AA00063111/00001> (дата обращения 12.11.2019).

2. Kan Li, Jose C. The Kernel Adaptive Autoregressive-Moving-Average Algorithm in IEEE transactions on neural networks and learning systems, vol. 1, no. 1, January 2015. – URL: https://www.researchgate.net/profile/Jose_Principe/publication/275717876_The_Kernel_Adaptive_Autoregressive-Moving-Average_Algorithm/links/554fd9e508ae739bdb908843/The-Kernel-Adaptive-Autoregressive-Moving-Average-Algorithm.pdf (дата обращения 12.11.2019).

3. Белов Ю.С., Демин И.С., Особенности обработки спутниковых снимков для обучения сверточной нейронной сети //Электронный журнал: наука, техника и образование. 2019. № 1 (23). С. 62-69.

4. Гришанов К.М., Белов Ю.С. Модель свёрточной нейронной сети в задачах машинного зрения // Электронный журнал: наука, техника и образование. 2017. № СВ1 (11). С. 100-106.

5. Гудфеллоу Я., Бенджио И., Курвилль А.; Глубокое обучение /перевод с английского А. А. Слинкина.— 2-е изд. — Москва : ДМК Пресс,

2018. — 652 с. — URL: <https://e.lanbook.com/book/107901> (дата обращения: 12.11.2019).

6. Демидова Л.А., Соколова Ю.С.; Использование SVM-алгоритма для уточнения решения задачи классификации объектов с применением алгоритмов кластеризации // Вестник Рязанского государственного радиотехнического университета. — 2015. — № 1(51). — С. 103-113. — URL: <https://e.lanbook.com/journal/issue/298285> (дата обращения: 12.11.2019).

7. Тай З.Л., Романов А.К., Ткаченко А.В., Рыбкин С.В. Обнаружение дорог по спутниковым снимкам с использованием нейронных сетей // Электронный журнал: наука, техника и образование. 2019. № 2 (24). С. 93-98.