

**ГИБРИДНЫЙ АЛГОРИТМ ОБУЧЕНИЯ КОНВОЛЮЦИОННОЙ НЕЙРОННОЙ СЕТИ**

И. А. Иванов\*, Е. А. Сопов

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева  
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31  
\*E-mail: ilyaiv92@gmail.com

*Задачи распознавания шаблонов в изображениях на сегодняшний день решаются во многих областях – от анализа земной поверхности спутниками до анализа лица человека в системах человеко-машинного взаимодействия. Одним из самых успешных алгоритмов по анализу и распознаванию изображений является конволюционная нейронная сеть. Алгоритм обратного распространения ошибки, используемый при обучении таких сетей, является градиентным алгоритмом, а следовательно, зачастую сходится к локальному оптимуму. Предлагается гибридный алгоритм обучения конволюционной сети, нацеленный на поиск глобального оптимума и состоящий из двух этапов. На первом этапе ведется поиск подобласти глобального оптимума с помощью генетического алгоритма, а на втором этапе используется алгоритм обратного распространения ошибки для нахождения глобального оптимума. Генетический алгоритм обладает внутренней процедурой поддержания разнообразия популяции решений, что позволяет более полно исследовать пространство поиска и находить разные удачные конфигурации конволюционной сети. Найденное генетическим алгоритмом решение используется на втором этапе в качестве начального приближения весов сети. Далее сеть обучается с помощью алгоритма обратного распространения ошибки. Разработанный гибридный алгоритм был протестирован на задаче распознавания эмоций, было проведено его сравнение с традиционным алгоритмом обратного распространения ошибки. Сравнение проводилось по точности классификации, а также по F-мере для задачи распознавания эмоций в двух постановках: классификация, зависящая от спикера, и классификация, независимая от спикера. Гибридный алгоритм показал лучшую эффективность по обоим критериям в обеих постановках задачи по сравнению со стандартным алгоритмом обучения конволюционной нейронной сети.*

*Ключевые слова:* конволюционная нейронная сеть, генетический алгоритм, алгоритм обратного распространения ошибки, человеко-машинное взаимодействие.

Sibirskii Gosudarstvennyi Aerokosmicheskii Universitet  
imeni Akademika M. F. Reshetneva. Vestnik  
Vol. 17, No. 4, P. 871–877**HYBRID ALGORITHM FOR CONVOLUTIONAL NEURAL NETWORK LEARNING**

I. A. Ivanov\*, E. A. Sopov

Reshetnev Siberian State Aerospace University  
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation  
\*E-mail: ilyaiv92@gmail.com

*The problems of image pattern recognition are nowadays solved in many areas – beginning with satellite analysis of the Earth surface all the way to human face analysis in the human-machine interaction systems. One of the most successful algorithms of image analysis and recognition is the convolutional neural network. The back propagation algorithm used for training of these networks belongs to the class of gradient algorithms; therefore it often converges to local minima. In this paper we propose a hybrid algorithm for convolutional network training, which is aimed at searching for global optimum, and consisting of two stages. At the first stage a genetic algorithm is used to search for the subdomain of the search space that includes the global optimum, whereas at the second stage a back propagation algorithm is used to find the actual global optimum. Genetic algorithm includes an internal procedure that maintains diversity of the solutions population, which allows examining the search space more thoroughly and finding various successful convolutional network configurations. The solution found by the genetic algorithm is used at the second stage as an initial approximation of network weights. After that the network is trained using a back propagation algorithm. The developed hybrid algorithm has been tested on the emotion recognition problem; it has been compared to a traditional back propagation algorithm. A comparison was made on the classification accuracy, as well as the F-measure for the emotion recognition problem in two formulations – speaker-dependent classification, and speaker-independent classification. Hybrid algorithm has yielded a better performance according to both criteria, and in both problem formulations compared to the standard algorithm used for convolutional neural network training.*

*Keywords:* convolutional neural network, genetic algorithm, back propagation algorithm, human-machine interaction.

**Введение.** Задачи распознавания изображений в настоящее время имеют огромное практическое применение в различных областях науки и техники. Алгоритмы глубокого обучения, в частности конволюционные (сверточные) нейронные сети, предложенные Яном Лекуном, являются одним из наиболее эффективных методов решения задач распознавания изображений. Конволюционная сеть имитирует механизм анализа изображений, используемый зрительным отделом коры головного мозга человека.

Важная часть конволюционной сети – алгоритм обучения. Традиционно для этой цели используется алгоритм обратного распространения ошибки (ОРО), в основе которого лежит аппроксимация градиента некоторой функции потерь, зависящей от регулируемых весов сети. Данный алгоритм хорошо зарекомендовал себя при решении задач различного рода, классический пример – распознавание рукописных цифр [1].

Однако при решении более сложных задач, например задач анализа изображений большого размера, распознавания объектов сложной структуры или неясных объектов, алгоритм обратного распространения имеет тенденцию к стагнации, т. е. обученная с помощью него сеть теряет обобщающую способность (например, все время распознает один и тот же объект, даже когда на изображении его нет).

В данной работе предлагается новый, гибридный алгоритм обучения конволюционной сети, созданный на базе традиционного алгоритма обратного распространения ошибки и генетического алгоритма оптимизации (ГА). Преимущество ГА состоит в том, что он, благодаря своей внутренней процедуре поиска, не только ищет оптимальное решение, но и поддерживает разнообразие найденной популяции решений. Благодаря этому удастся избежать стагнации при решении различных задач оптимизации [2; 3].

Так как количество параметров (весов) конволюционной сети очень велико, ГА не может должным образом покрыть все пространство поиска. Его задача – поддерживать разнообразие найденных решений и найти область в пространстве параметров сети, соответствующую предполагаемому глобальному оптимуму. После ГА в работу вступает алгоритм ОРО, который с помощью градиентного спуска находит оптимальные веса сети в найденной генетическим алгоритмом области.

В качестве практической задачи классификации была выбрана задача распознавания эмоций. Эта задача является частью более общей задачи человеко-машинного взаимодействия. Сложность данной задачи обусловлена тем, что приходится работать с большими изображениями, и в особенности с тем, что эмоции являются неявными признаками, распознать которые можно только с помощью системного анализа всего изображения лица человека.

**Обзор используемых алгоритмов.** Генетический алгоритм оптимизации [4–6] основан на биологической идее естественного отбора: индивиды с большей пригодностью для выживания имеют больше возможностей для воспроизводства. В качестве индивидов в ГА выступают решения, представляющие собой вектор входных переменных задачи оптимизации.

Значения входных переменных кодируются в двоичную систему, таким образом каждое решение (индивид) представляет собой бинарную строку, в которой закодированы значения входных переменных и значение пригодности, которому это решение соответствует [7]. Функция пригодности может совпадать с целевой оптимизируемой функцией, если решается задача максимизации, либо быть обратной ей в случае задачи минимизации (например, минимизация ошибки классификации).

ГА – это итеративная процедура оптимизации, на каждой итерации которой последовательно производятся следующие операции: селекция, скрещивание и мутация [8; 9]. Оператор селекции выполняет отбор наиболее перспективных решений по значению их функции пригодности. Чем выше пригодность решения, тем выше вероятность его отбора. Основные виды селекции – ранговая, пропорциональная, турнирная. Оператор скрещивания производит рекомбинацию бинарных строк, которыми представлены решения, выбранные на этапе селекции, генерируя новое поколение решений. Скрещивание бывает трех основных типов: одноточечное, двухточечное, равномерное. Операторы селекции и скрещивания вместе обеспечивают эволюционную динамику решений, давая более пригодным индивидам больше возможностей для воспроизводства. Наконец, оператор мутации с некоторой малой вероятностью меняет случайный бит (несколько битов) в бинарной строке, представляющей решение. Этот оператор направлен на поддержание разнообразия популяции решений, предотвращение стагнации алгоритма, т. е. ситуации, когда вся популяция состоит из одинаковых решений и дальнейшая оптимизация отсутствует. Данная проблема преждевременной стагнации ГА и застревания в области локального оптимума известна и ранее исследовалась [10].

Конволюционная нейронная сеть, используемая в данной работе, была предложена в 80-е годы прошлого века. В ней заимствовались идеи обработки входящего изображения зрительным отделом коры головного мозга человека. Данный алгоритм получил большую популярность вследствие высокой эффективности его применения для распознавания рукописных цифр (он до сих пор применяется в системах автоматизированного анализа ~10 % банковских чеков в США).

Конволюционная сеть состоит из последовательно расположенных слоев двух типов – конволюционный слой (слой свертки) и слой субдискретизации. На каждом конволюционном слое происходит разбиение входного изображения на фрагменты, после чего производится свертка каждого фрагмента с набором карт признаков (feature maps). На слое субдискретизации происходит уменьшение входного изображения с некоторым масштабом путем усреднения близлежащих пикселей либо иным способом. Варьируемыми параметрами конволюционной сети являются количество слоев, число и размер карт признаков на конволюционных слоях, масштаб уменьшения входного изображения на слоях субдискретизации, величина шага обучения, наличие или отсутствие полносвязного перцептрона на выходе конволюционной сети.

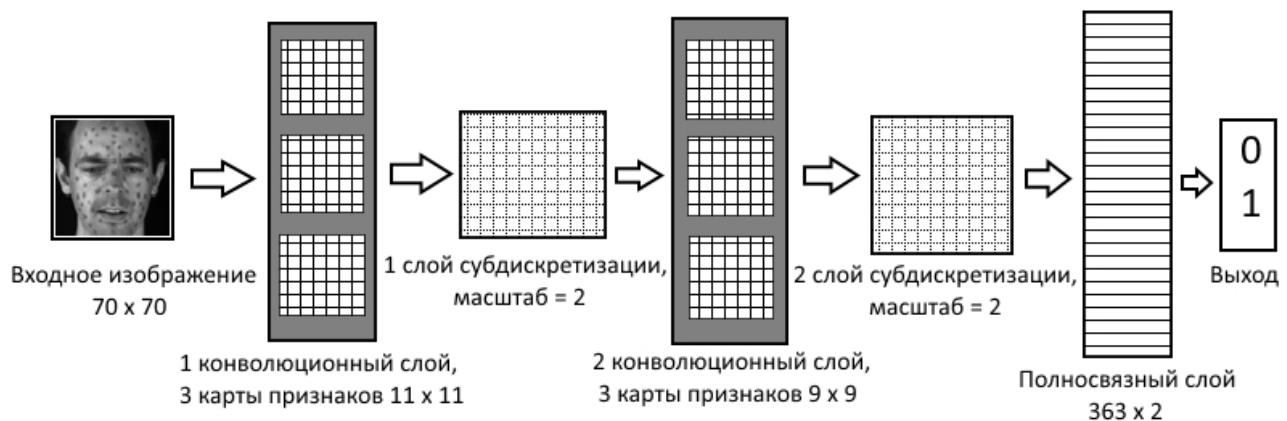
Все вышеперечисленные параметры определяют структуру сети. В процессе обучения происходит настройка весов сети, число которых в зависимости от структуры сети может быть очень велико [11]. Веса конволюционной сети включают в себя веса карт признаков на конволюционных слоях, а также веса полносвязного перцептрона на выходе сети, если он присутствует. Несмотря на большое количество весов конволюционной сети, их число оказывается существенно меньшим по сравнению с полносвязной нейронной сетью. Это обусловлено тем, что конволюционная сеть, обучаясь, не запоминает положение каждого пикселя на изображении, как это делает полносвязная нейронная сеть, а ищет некие шаблоны во фрагментах изображения с помощью карт признаков, что по сути имитирует процесс распознавания образов человеческим мозгом. Кроме того, такая процедура требует существенно меньше вычислительных ресурсов.

**Методология.** В данной работе предлагается новый, инновационный гибридный алгоритм обучения конволюционной нейронной сети, основанный на алгоритме обратного распространения ошибки и коэволюционном генетическом алгоритме оптимизации.

Коэволюционный ГА отличается от стандартного тем, что в нем вместо одной популяции используется несколько популяций, эволюция в которых идет отдельно друг от друга (отсюда и название коэволюция). Каждой популяции соответствует свой стандартный ГА с некоторыми заданными параметрами,

обеспечивающий процедуру оптимизации. Через фиксированное количество итераций происходит обмен решениями между популяциями (миграция). Коэволюционный ГА не требует тонкой настройки параметров, и в этом плане он предпочтительнее стандартного ГА.

В данной работе использовалась конволюционная сеть фиксированной структуры, представленной на рисунке. Макропараметры сети, такие как количество и размерность ядер свертки каждого конволюционного слоя, и масштаб уменьшения входного изображения на каждом слое субдискретизации выбирались вручную в зависимости от размера входных изображений. Размерность полносвязного слоя рассчитывалась автоматически в зависимости от размера изображения на предыдущем слое. Ручной выбор параметров сети связан с тем, что в процессе прохождения через сеть с входным изображением производится ряд операций по его разбиению на фрагменты и масштабированию. В процессе выполнения этих операций может возникнуть ситуация, когда изображение будет иметь нецелый пиксельный размер, например  $23,5 \times 23,5$ , что приведет к сбою работы сети. Поэтому необходимо заранее выбирать такие параметры сети, которые не приводили бы к подобной ситуации. Значения параметров устанавливались одинаковыми для сетей с исследуемыми алгоритмами обучения для проведения объективного сравнения эффективности (табл. 1).



Структура конволюционной нейронной сети для входных изображений размера  $70 \times 70$  пикселей, 2 класса на выходе

Таблица 1

Значения параметров конволюционной нейронной сети для различных размеров входных изображений

| Параметр                              | Размер входных изображений |       |       |
|---------------------------------------|----------------------------|-------|-------|
|                                       | 40×40                      | 50×50 | 70×70 |
| Число карт признаков на конв. слоях   | 3                          | 3     | 3     |
| Размерность ядра свертки 1 конв. слоя | 7×7                        | 9×9   | 11×11 |
| Размерность ядра свертки 2 конв. слоя | 6×6                        | 6×6   | 9×9   |
| Масштаб слоев субдискретизации        | 2                          | 2     | 2     |
| Число нейронов полносвязного слоя     | 108                        | 192   | 363   |

Таблица 2

Параметры генетического алгоритма

| Параметр                               | Значение  |
|--|-----------|
| Размер популяции                       | 50        |
| Число итераций                         | 50        |
| Интервал поиска (по каждой переменной) | $[-3; 2]$ |
| Точность разбиения интервала поиска    | 0,0001    |
| Процент элитарных решений              | 4 %       |

Предлагаемый гибридный алгоритм обучения конволюционной сети состоит из двух этапов. На первом этапе происходит оптимизация весов сети с помощью генетического алгоритма оптимизации. Используемые в экспериментах параметры ГА представлены в табл. 2. В качестве индивида в ГА выступает конволюционная нейронная сеть с определенным набором весов. Количество входных переменных ГА (весов сети) для изображений размером  $40 \times 40$ ,  $50 \times 50$ ,  $70 \times 70$  пикселей равно 483, 579 и 1104 соответственно. Применение ГА на 1 этапе обучения сети позволяет более полно охватить пространство поиска и избежать стагнации. Задача ГА – в общем пространстве поиска найти область потенциального оптимума, чтобы на втором этапе с помощью градиентного спуска найти сам оптимум.

Важный момент – выбор оптимизируемого критерия для генетического алгоритма. Так как каждый индивид ГА – это конволюционная сеть, а в качестве целевой функции выступает задача классификации, то естественно было бы выбрать точность классификации максимизируемым критерием (или ошибку классификации минимизируемым критерием соответственно). Изначально в ходе выполнения данной работы так и было сделано, однако точность классификации оказалась неудовлетворительным критерием, так как ГА в большинстве экспериментов впадал в стагнацию, т. е. все конволюционные сети – индивиды популяции, спустя некоторое число итераций оказывались одинаковыми, что прекращало процесс оптимизации. В связи с этим автором данной работы был предложен иной критерий оптимизации – усредненная  $F$ -мера, которая для случая двух классов вычисляется по формуле (1):

$$F = 2 \cdot \frac{\text{точность} \cdot \text{полнота}}{\text{точность} + \text{полнота}}. \quad (1)$$

Точность (*precision*) и полнота (*recall*) вычисляются по формулам (2) и (3), понятия TP (*true-positive*, истинно положительное решение), TN (*true-negative*, истинно отрицательное решение), FP (*false-positive*, ложноположительное решение) и FN (*false-negative*, ложноотрицательное решение) следуют из табл. 3:

$$\text{Точность} = \frac{TP}{TP + FP}; \quad (2)$$

$$\text{Полнота} = \frac{TP}{TP + FN}. \quad (3)$$

$F$ -мера более адекватно отражает способность системы классификации к обобщению, т. е. классификации объектов, не использовавшихся для обучения системы. Формула (1) используется для вычисления

$F$ -меры каждого класса в задаче с двумя классами. В случае большего числа классов задачу всегда можно свести к двум классам, используя схему «один против всех» (*one-vs-all*), в которой все классы, кроме 1-го, заменяются на один класс – непервый, затем все классы, кроме 2-го, заменяются на невторой класс и т. д.

Для ускорения работы генетического алгоритма при вычислении  $F$ -меры каждой сети в популяции на ее вход подавалась не вся обучающая выборка, а лишь часть.

Таблица 3

Таблица сопряженности, иллюстрирующая понятия точности и полноты

| Класс $i$ |   | Истина |    |
|-----------|---|--------|----|
|           |   | 1      | 0  |
| Оценка    | 1 | TP     | FP |
|           | 0 | FN     | TN |

В конце 1 этапа из конечной популяции генетического алгоритма выбиралась конволюционная сеть,  $F$ -мера которой была максимальной. Веса выбранной сети, оптимизированные генетическим алгоритмом, брались в качестве начального приближения для 2 этапа обучения. На 2 этапе обучение выбранной сети продолжалось с помощью традиционного алгоритма ОРО. Так как данный алгоритм является градиентным, идея его использования после начального приближения с помощью ГА состоит в том, что он должен в суженной ГА области поиска найти глобальный оптимум, донстроить сеть. Так как алгоритм ОРО в качестве начального приближения использует случайно сгенерированные значения весов конволюционной сети, он склонен застревать в локальных оптимумах. Получение начального приближения весов с помощью генетического алгоритма призвано решить эту проблему.

**Описание задачи классификации эмоций.** Задача распознавания эмоций, решаемая в данной работе, относится к более общему классу задач человеко-машинного взаимодействия. В качестве исходных данных использовалась база данных эмоций SAVEE (Surrey Audio-Visual Expressed Emotion) [12]. Из видеопоследовательностей отбирались отдельные кадры через фиксированный промежуток времени, размер кадров уменьшался до  $40 \times 40$ ,  $50 \times 50$  и  $70 \times 70$  пикселей соответственно. Таким образом, были получены 3 выборки с различным размером изображений. В базе SAVEE присутствуют 7 классов эмоций: злость, отвращение, страх, счастье, нейтральная, грусть, удивление. Задача была переформулирована для классификации отдельных типов эмоций по схеме *one-vs-all*.

База данных SAVEE включает в себя видеозаписи 4 спикеров. Задача распознавания эмоций может быть поставлена двумя способами. Данные, соответствующие одному спикеру, могут присутствовать как в обучающей, так и в тестовой выборке. Такая постановка задачи называется зависимой от спикера (*speaker-dependent*). Либо обучающая и тестовая выборка могут содержать данные разных спикеров и не пересекаться между собой. Такая постановка называется независимой от спикера (*speaker-independent*) [13–15]. Задача классификации эмоций во второй постановке более сложная, так как система должна научиться определять эмоцию у человека, которого она ранее не видела. В данной работе были проведены эксперименты по решению задачи распознавания эмоций в обеих постановках.

**Результаты экспериментов.** Сравнение эффективности проводилось для конволюционных сетей, обученных с помощью алгоритма обратного распро-

странения ошибки, генетического алгоритма и гибридного алгоритма (ГА + ОРО).

Задача распознавания эмоций была сформулирована как задача классификации «один против всех» для каждого типа эмоций, например, классы «счастье» и «несчастье», в который входят объекты выборки, принадлежащие ко всем остальным классам, кроме «счастья». Сравнимые критерии – точность классификации и *F*-мера. Критерий *F*-меры был выбран потому, что он более адекватно оценивает качество классификации при классах, неоднородно представленных в выборке (число объектов одного класса существенно больше числа объектов другого класса).

Результаты сравнения средней точности распознавания эмоций по итерациям обучения в постановке, зависимой от спикера, представлены в табл. 4, средней *F*-меры – в табл. 5. Аналогичные результаты для постановки, независимой от спикера, представлены в табл. 6 и 7.

Таблица 4

**Сравнение средней точности (%) распознавания эмоций по итерациям обучения в конволюционной нейронной сети (КНС): обратное распространение ошибки (ОРО); генетический алгоритм (ГА); ГА + ОРО; классификация, зависимая от спикера; цветовая схема: темно-серый – лучшая точность, светло-серый – средняя точность, белый – худшая точность**

| Алгоритм обучения КНС                       |             | ОРО   |       |       | ГА    |       |       | ГА + ОРО |       |       |
|---|-------------|-------|-------|-------|-------|-------|-------|----------|-------|-------|
| Размер изображений ( $n \times n$ пикселей) |             | 40    | 50    | 70    | 40    | 50    | 70    | 40       | 50    | 70    |
| Эмоция                                      | Злость      | 92,27 | 93,39 | 96,45 | 85,48 | 79,62 | 88,89 | 95,31    | 95,87 | 97,31 |
|   | Отвращение  | 87,91 | 89,64 | 96,73 | 79,77 | 80,94 | 82,64 | 91,45    | 94,36 | 89,67 |
|   | Страх       | 87,28 | 88,24 | 92,04 | 82,59 | 78,1  | 84,64 | 90,15    | 91,61 | 87,09 |
|   | Счастье     | 88,3  | 90,4  | 93,3  | 87,53 | 84,22 | 74,72 | 92,98    | 94,06 | 89,02 |
|   | Нейтральная | 87,43 | 88,31 | 93,59 | 87,53 | 71,32 | 61,65 | 92,98    | 97,29 | 86,82 |
|   | Грусть      | 87,31 | 88,67 | 91,4  | 88,43 | 85,66 | 85,94 | 90,79    | 92,09 | 89,52 |
|   | Удивление   | 88,97 | 90,51 | 93,96 | 83,44 | 85,59 | 73,77 | 91,82    | 94,29 | 94,56 |
| Средний показатель                          |             | 90,77 |       |       | 81,55 |       |       | 92,34    |       |       |

Таблица 5

**Сравнение средней *F*-меры по итерациям обучения: ОРО; ГА; ГА + ОРО; классификация, зависимая от спикера; цветовая схема: темно-серый – лучшая точность, светло-серый – средняя точность, белый – худшая точность**

| Алгоритм обучения КНС                       |             | ОРО   |       |       | ГА    |       |       | ГА + ОРО |       |       |
|---|-------------|-------|-------|-------|-------|-------|-------|----------|-------|-------|
| Размер изображений ( $n \times n$ пикселей) |             | 40    | 50    | 70    | 40    | 50    | 70    | 40       | 50    | 70    |
| Эмоция                                      | Злость      | 0,759 | 0,796 | 0,898 | 0,773 | 0,683 | 0,804 | 0,885    | 0,894 | 0,939 |
|   | Отвращение  | 0,628 | 0,705 | 0,794 | 0,663 | 0,679 | 0,664 | 0,76     | 0,861 | 0,677 |
|   | Страх       | 0,514 | 0,568 | 0,744 | 0,702 | 0,622 | 0,684 | 0,674    | 0,75  | 0,51  |
|   | Счастье     | 0,517 | 0,632 | 0,763 | 0,759 | 0,683 | 0,582 | 0,782    | 0,824 | 0,579 |
|   | Нейтральная | 0,698 | 0,721 | 0,849 | 0,759 | 0,642 | 0,533 | 0,782    | 0,946 | 0,713 |
|   | Грусть      | 0,479 | 0,573 | 0,691 | 0,737 | 0,673 | 0,7   | 0,689    | 0,738 | 0,632 |
|   | Удивление   | 0,504 | 0,602 | 0,772 | 0,685 | 0,67  | 0,571 | 0,689    | 0,797 | 0,79  |
| Средний показатель                          |             | 0,677 |       |       | 0,679 |       |       | 0,758    |       |       |

Таблица 6

**Сравнение средней точности (%) по итерациям обучения: ОРО; ГА; ГА + ОРО; классификация, независимая от спикера; цветовая схема: темно-серый – лучшая точность, светло-серый – средняя точность, белый – худшая точность**

| Алгоритм обучения КНС                       |             | ОРО   |       |       | ГА    |       |       | ГА + ОРО |       |       |
|---|-------------|-------|-------|-------|-------|-------|-------|----------|-------|-------|
| Размер изображений ( $n \times n$ пикселей) |             | 40    | 50    | 70    | 40    | 50    | 70    | 40       | 50    | 70    |
| Эмоция                                      | Злость      | 93,31 | 94,75 | 96,71 | 87,22 | 81,41 | 85,96 | 95,26    | 94,26 | 89,94 |
|   | Отвращение  | 92,8  | 94,37 | 96,47 | 84,52 | 77,88 | 81,39 | 94,95    | 94,22 | 90,91 |
|   | Страх       | 90,15 | 90,83 | 94,11 | 81,96 | 87,88 | 80,08 | 90,61    | 93,13 | 95,1  |
|   | Счастье     | 90,27 | 92,71 | 94,74 | 86,64 | 42,42 | 75,03 | 95,81    | 89,12 | 90,29 |
|   | Нейтральная | 92,93 | 94,15 | 93,05 | 85,66 | 77,12 | 70,22 | 95,81    | 96,35 | 94,73 |
|   | Грусть      | 90,58 | 91,78 | 93,71 | 90,5  | 88,27 | 85,63 | 94,18    | 94,31 | 91,72 |
|   | Удивление   | 90,66 | 92,51 | 95,71 | 91,26 | 70,36 | 79,76 | 94,04    | 89,05 | 96,34 |
| Средний показатель                          |             | 93,16 |       |       | 80,53 |       |       | 93,34    |       |       |

Таблица 7

Сравнение средней *F*-меры по итерациям обучения: ОРО; ГА; ГА + ОРО; классификация, независимая от спикера; цветовая схема: темно-серый – лучшая точность, светло-серый – средняя точность, белый – худшая точность

| Алгоритм обучения КНС                          |             | ОРО   |       |       | ГА    |       |       | ГА + ОРО |       |       |
|--|-------------|-------|-------|-------|-------|-------|-------|----------|-------|-------|
| Размер изображений<br>( $n \times n$ пикселей) |             | 40    | 50    | 70    | 40    | 50    | 70    | 40       | 50    | 70    |
| Эмоция   | Злость      | 0,796 | 0,846 | 0,911 | 0,761 | 0,693 | 0,771 | 0,874    | 0,853 | 0,739 |
|  | Отвращение  | 0,798 | 0,845 | 0,905 | 0,711 | 0,661 | 0,674 | 0,877    | 0,843 | 0,763 |
|  | Страх       | 0,662 | 0,687 | 0,817 | 0,69  | 0,685 | 0,663 | 0,725    | 0,822 | 0,841 |
|  | Счастье     | 0,639 | 0,74  | 0,837 | 0,73  | 0,389 | 0,585 | 0,867    | 0,547 | 0,597 |
|  | Нейтральная | 0,835 | 0,862 | 0,82  | 0,74  | 0,649 | 0,598 | 0,918    | 0,924 | 0,884 |
|  | Грусть      | 0,719 | 0,762 | 0,83  | 0,76  | 0,756 | 0,717 | 0,851    | 0,855 | 0,795 |
| Удивление                                      |             | 0,595 | 0,69  | 0,837 | 0,795 | 0,522 | 0,616 | 0,775    | 0,475 | 0,864 |
| Средний показатель                             |             | 0,783 |       |       | 0,675 |       |       | 0,795    |       |       |

Из представленных результатов видно, что генетический алгоритм в представленных постановках задачи классификации проигрывает по критериям эффективности алгоритму обратного распространения ошибки. Гибридный же алгоритм в среднем превышает по эффективности алгоритм ОРО.

Полученные результаты указывают на то, что гибридный алгоритм обучения конволюционной нейросети, сочетающий в себе эволюционную эвристику и градиентный спуск, эффективнее алгоритма ОРО. Однако данный результат справедлив только для решаемой в данной работе задачи распознавания эмоций, и требуются дальнейшие исследования разработанного алгоритма на других практических задачах классификации для доказательства или опровержения гипотезы об общей эффективности разработанного алгоритма. В первую очередь, аналогичное сравнение алгоритмов обучения необходимо провести на стандартных, общепринятых задачах, например, на задаче классификации рукописных цифр из базы MNIST.

Генетический алгоритм не показал хорошей схожимости к глобальному оптимуму при оптимизации параметров конволюционной нейросети. Вероятная причина этого – слишком большая размерность пространства поиска (~ 500–1000 переменных). Задача ГА в разработанном гибридном алгоритме состоит в поиске подобласти в пространстве входных переменных, в которой скорее всего находится глобальный оптимум, и со своей задачей ГА справляется. Однако очевидно, что для поиска точного расположения глобального оптимума генетический алгоритм в данной ситуации непригоден, так как для этого потребуются неоправданно высокие вычислительные затраты. Поэтому необходимо использовать процедуру градиентного спуска либо иную процедуру для поиска глобального оптимума в выделенной подобласти.

**Заключение.** Конволюционные нейронные сети с успехом применяются для решения задач распознавания шаблонов в изображениях на протяжении многих лет. Стандартным алгоритмом обучения таких сетей является алгоритм обратного распространения ошибки, являющийся по сути алгоритмом градиентного спуска. Алгоритмы градиентного спуска сходятся к глобальному оптимуму, только если начальная точка поиска находится вблизи данного оптимума. В противном случае алгоритм сходится к одному из локальных оптимумов.

В данной работе предложен гибридный алгоритм обучения конволюционной нейронной сети, призванный решить вышеописанную проблему. Разработанный алгоритм состоит из двух этапов: поиск подобласти глобального оптимума генетическим алгоритмом и градиентный спуск к глобальному оптимуму в этой подобласти с помощью алгоритма обратного распространения ошибки.

Разработанный гибридный алгоритм превзошел по эффективности стандартный алгоритм обратного распространения ошибки на задаче распознавания эмоций, однако для вынесения вердикта о его эффективности необходимо провести дальнейшие сравнения на других тестовых задачах классификации.

**Благодарности.** Исследование выполнено при финансовой поддержке РФФИ и Правительства Красноярского края в рамках научного проекта № 16-41-243036. Работа также поддержана грантом Президента РФ (МК-3285.2015.9), Российским фондом фундаментальных исследований и Красноярским краевым фондом поддержки научной и научно-технической деятельности (проект 16-41-240822).

**Acknowledgments.** The reported study was funded by RFBR and Government of Krasnoyarsk Region according to the research project № 16-41-243036. This research was also supported by the President of the Russian Federation grant (МК-3285.2015.9) and by Russian Foundation for Basic Research and Krasnoyarsk Regional Fund for Support of Scientific and Technical Activities (project 16-41-240822).

### Библиографические ссылки

1. Comparison of learning algorithms for handwritten digit recognition / Y. LeCun [et al.] // International Conference on Artificial Neural Networks. 1995. P. 53–60.
2. Семенкина М. Е., Семенкин Е. С., Попов Е. А. Исследование эффективности самоконфигурируемого генетического алгоритма выбора эффективного варианта системы управления космическими аппаратами // Программные продукты и системы. 2013. № 4. С. 208–213.
3. Иванов И. А., Сопов Е. А. Исследование эффективности самоконфигурируемого коэволюционного алгоритма решения сложных задач многокритериальной оптимизации // Системы управления и информационные технологии. 2013. Т. 51, № 1.1. С. 141–145.

4. Eiben A. E., Raue P.-E., Ruttkay Z. Genetic algorithms with multi-parent recombination // *Proceedings of the International Conference on Evolutionary Computation, The Third Conference on Parallel Problem Solving from Nature*. 1994. P. 78–87.
5. Ting C.-K. On the Mean Convergence Time of Multi-parent Genetic Algorithms Without Selection // *Advances in Artificial Life*. 2005. P. 403–412.
6. Akbari Z. A multilevel evolutionary algorithm for optimizing numerical functions // *IJEC*. 2010. 2. P. 419–430.
7. Панфилов И. А., Базанова Е. П., Сопов Е. А. Исследование эффективности работы генетического алгоритма оптимизации с альтернативным представлением решений // *Вестник СибГАУ*. 2013. 4(50). С. 68–71.
8. Сопов Е. А., Семенкин Е. С. Исследование эффективности модифицированного вероятностного генетического алгоритма на десептивных функциях «ловушках» // *Системы управления и информационные технологии*. 2011. № 3 (45). С. 90–95.
9. Ахмедова Ш. А., Семенкин Е. С. Кооперативный бионический алгоритм безусловной оптимизации // *Программные продукты и системы*. 2013. № 4 (104). С. 133–137.
10. Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs. 3rd Ed. Springer-Verlag, 1996.
11. Flexible, high performance convolutional neural networks for image classification / D. C. Ciresan [et al.] // *Proceedings of the International Joint Conference on Artificial Intelligence*. 2011. Vol. 22 (1). P. 1237–1242.
12. Haq S., Jackson P. J. B. Speaker-dependent audio-visual emotion recognition // *Proceedings Int. Conf. on Auditory-Visual Speech Processing*. 2009. P. 53–58.
13. Sidorov M., Ultes S., Schmitt A. Emotions are a personal thing: Toward speaker-adaptive emotion recognition // *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2014. P. 4803–4807.
14. Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles / B. Schuller [et al.] // *Proc. INTERSPEECH*. 2005. P. 805–808.
15. Brester C., Sidorov M., Semenkina E. Speech-based emotion recognition: Application of collective decision making concepts // *Proceedings of the 2nd International Conference on Computer Science and Artificial Intelligence*. 2014. P. 216–220.
3. Ivanov I. A., Sopov E. A. [Effectiveness study of the self-configuring of co-evolutionary algorithm for solving complex multi-objective optimization problems]. *Sistemy upravleniya i informatsionnye tekhnologii*. 2013, Vol. 51, No. 1.1, P. 141–145 (In Russ.).
4. Eiben A. E., Raue P.-E., Ruttkay Z. Genetic algorithms with multi-parent recombination. *Proceedings of the International Conference on Evolutionary Computation, The Third Conference on Parallel Problem Solving from Nature*. 1994, P. 78–87.
5. Ting C.-K. On the Mean Convergence Time of Multi-parent Genetic Algorithms Without Selection. *Advances in Artificial Life*. 2005, P. 403–412.
6. Akbari Z. A multilevel evolutionary algorithm for optimizing numerical functions. *International Journal of Industrial Engineering Computations* 2. 2010, P. 419–430.
7. Panfilov I. A., Bazanova E. P., Sopov E. A. [Effectiveness study of the genetic optimization algorithm with alternative solutions representation]. *Vestnik SibGAU*. 2013, No. 4(50), P. 68–71 (In Russ.).
8. Sopov E. A., Semenkina E. S. [Effectiveness study of the modified probabilistic genetic algorithm on deceptive “trap” functions]. *Sistemy upravleniya i informatsionnye tekhnologii*. 2011, No. 3 (45), P. 90–95 (In Russ.).
9. Akhmedova Sh. A., Semenkina E. S. [Cooperative bionic algorithm of unconditional optimization]. *Programmnye produkty i sistemy*. 2013, No. 4 (104), P. 133–137 (In Russ.).
10. Michalewicz Z. Genetic Algorithms + Data Structures = Evolution Programs, 3rd Edition. *Springer-Verlag*. 1996.
11. Ciresan D. C., Meier U., Masci J., Gambardella L. M., Schmidhuber J. Flexible, high performance convolutional neural networks for image classification. *Proceedings of the International Joint Conference on Artificial Intelligence*. 2011, Vol. 22 (1), P. 1237–1242.
12. Haq S., Jackson P. J. B. Speaker-dependent audio-visual emotion recognition. *Proceedings Int. Conf. on Auditory-Visual Speech Processing*, 2009, P. 53–58.
13. Sidorov M., Ultes S., Schmitt A. Emotions are a personal thing: Toward speaker-adaptive emotion recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2014, P. 4803–4807.
14. Schuller B., Muller R., Lang M., Rigoll G. Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. *Proc. INTERSPEECH*. 2005, P. 805–808.
15. Brester C., Sidorov M., Semenkina E. Speech-based emotion recognition: Application of collective decision making concepts. *Proceedings of the 2nd International Conference on Computer Science and Artificial Intelligence*. 2014, P. 216–220.

## References

1. LeCun Y., Jackel L. D., Bottou L., Brunot A., Cortes C., Denker J. S., Drucker H., Guyon I., Muller U. A., Sackinger E., Simard P., Vapnik V. Comparison of learning algorithms for handwritten digit recognition. *International Conference on Artificial Neural Networks*. 1995, P. 53–60.
2. Semenkina M. E., Semenkina E. S., Popov E. A. [Effectiveness study of the self-configuring genetic algorithm for selecting an effective variant of spacecraft control system]. *Programmnye produkty i sistemy*. 2013, No. 4, P. 208–213 (In Russ.).