УДК 004.853

В.А. Сергеев, М.Ю. Леонтьев, В.Ю. Ислентьева, С.В. Сухов

АПРОБАЦИЯ ГЕНЕРАТИВНЫХ ПОДХОДОВ К ОБМЕНУ ЗНАНИЯМИ В НЕЙРОННЫХ СЕТЯХ¹

Сергеев Вячеслав Андреевич, доктор технических наук, профессор, окончил физический факультет Горьковского государственного университета им. Н.И. Лобачевского. Директор Ульяновского филиала Института радиотехники и электроники им. В.А. Котельникова Российской академии наук, заведующий кафедрой «Радиотехника, опто- и наноэлектроника» Ульяновского государственного технического университета (УлГТУ). Имеет монографии, статьи и изобретения в области исследования характеристик полупроводниковых приборов и интегральных схем, измерения их тепловых параметров. [e-mail: sva@ulstu.ru].

Леонтьев Михаил Юрьевич, окончил аспирантуру факультета математики, информационных и авиационных технологий Ульяновского государственного университета (УлГУ). Младший научный сотрудник Научно-исследовательского технологического института им. С.П. Капицы УлГУ, ведущий инженер УФИРЭ им. В.А. Котельникова РАН. Имеет 8 статей, 1 свидетельство о регистрации программы для ЭВМ. Область научных интересов – машинное обучение и робототехника. [e-mail: ulstaer@gmail.com].

Ислентьева Виктория Юрьевна, студент факультета информационных систем и технологий УлГТУ. Имеет 2 статьи. Область научных интересов – машинное обучение. [e-mail: viksaskin@yandex.ru].

Сухов Сергей Владимирович, кандидат физико-математических наук, окончил филиал Московского государственного университета им. М.В. Ломоносова в г. Ульяновске. Старший научный сотрудник УФИРЭ им. В.А. Котельникова РАН, доцент кафедры «Прикладная математика и информатика» УлГТУ. Имеет монографию, 90 статей, 2 патента на изобретения. Область научных интересов: оптика, вычислительная нейробиология, машинное обучение. [e-mail: ssukhov@ulireran.ru].

Аннотация

Способность к взаимодействию и обмену знаниями между искусственными нейронными сетями (ИНС) является критичной для дальнейшего развития сферы искусственного интеллекта. Один из методов обмена знаниями между нейронными сетями может быть основан на генерации искусственных тренировочных данных с помощью дополнительных генеративных моделей, получивших в последнее время широкое распространение. Для повышения эффективности подобных методов необходимо дальнейшее улучшение качества генерируемых искусственных тренировочных данных, создаваемых генеративными моделями. В данной работе предлагаются и тестируются несколько методов улучшения характеристик одного из классов генеративных моделей – вариационных автокодировщиков

¹ Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научного проекта № 18-47-732006.

(ВАК). Улучшения касались как самой процедуры тренировки ВАК, так и изменения архитектуры модели. Методы улучшения качества генерируемых тренировочных данных ВАК проанализированы с точки зрения их эффективности для обмена знаниями между ИНС. Эксперименты по обмену знаниями между ИНС были проведены с использованием общедоступных наборов тренировочных данных.

Ключевые слова: генеративные сети, вариационные автокодировщики, нейронные сети, машинное обучение, классификация изображений, обмен знаниями.

doi: 10.35752/1991-2927-2020-2-60-73-80

EVALUATION OF GENERATIVE APPROACHES TO KNOWLEDGE SHARING IN NEURAL NETWORKS

Viacheslav Andreevich Sergeev, Doctor of Sciences in Engineering, Professor; graduated from the Faculty of Physics of Gorky State University n.a. Lobahevsky; Director of the Ulyanovsk Branch of the Kotel'nikov Institute of Radio-Engineering and Electronics of the Russian Academy of Sciences; Head of the Department of Radioengineering, Opto- and Nanoelectronics of Ulyanovsk State Technical University (UISTU); an author of monographs, papers, and inventions in the field of modeling and researching semiconductor devices and integrated circuits characteristics and measuring its thermal parameters. e-mail: sva@ulstu.ru.

Mikhail Iurevich Leontev, completed postgraduate studies at the Faculty of Mathematics, Information, and Aviation Technologies at Ulyanovsk State University; Junior Research Associate of the Scientific and Research Institute of Technology n.a. S.P. Kapitsa; Leading Engineer of the Ulyanovsk Branch of the Kotel'nikov Institute of Radio-Engineering and Electronics of RAS; an author of eight articles, a certificate of registration for computer program. The area of his research interests relates to learning and robotics. e-mail: ulstaer@gmail.com.

Victoriia Iurevna Islenteva, a student of the Faculty of Information Systems and Technologies of Ulyanovsk State Technical University; an author of two articles. The area of her research interests relates to machine learning. e-mail: viksaskin@yandex.ru.

Sergei Vladimirovich Sukhov, Candidate of Sciences in Physics and Mathematics; graduated from the Ulyanovsk Branch of Lomonosov Moscow State University; Senior Staff Scientist at the Ulyanovsk Branch of the Kotel'nikov Institute of Radio-Engineering and Electronics of RAS; Associate Professor at the Department of Applied Mathematics and Informatics of UISTU; an author of a monograph, 90 articles, two patents for inventions in the field of optics, computational neuroscience, and machine learning. e-mail: ssukhov@ulireran.ru.

Abstract

The ability to interact and share knowledge among artificial neural networks (ANNs) is critical for further development of the artificial intelligence. One of the methods of knowledge sharing among ANNs may rely on the generation of training data using auxiliary generative models, which have recently become widespread. To make such methods more efficient, it is necessary to further improve the quality of artificial training data developed by generative models. The article proposes and tests several methods of enhancing characteristics of one of the classes of generative models, viz. variational autoencoders (VAE). It was improved both the VAE training procedure and changes in the model architecture. The improving methods of the quality of generated training data have been analyzed in terms of the efficiency of knowledge sharing among ANNs. To test the process of knowledge sharing among ANNs, the public training data sets have been used.

Key words: generative networks, variational autoencoders, neural networks, machine learning, image classification, knowledge sharing.

ВВЕДЕНИЕ

Задачи взаимодействия агентов, обладающих искусственным интеллектом (ИИ), обсуждаются уже продолжительное время [1]. В последние годы подходы, требующие взаимодействия нескольких агентов с ИИ, начали развиваться особенно интенсивно [2]. В задачах обучения с подкреплением несколько агентов могут действовать сообща, параллельно разрабатывая прото-

кол обмена информацией, чтобы максимизировать совместное вознаграждение [2]. Разработка алгоритмов взаимодействия подобных агентов и обмена информацией между ними необходима в целом ряде приложений, таких как совместная доставка грузов дронами, координация между автономными транспортными средствами, задачи управления городским движением и т. д.

Другой пример взаимодействующих искусственных нейронных сетей (ИНС) можно найти в задачах обучения с учителем: знания нескольких нейронных сетей, обученных на разных наборах данных, необходимо объединить в одной сети. При этом предполагается, что использование ансамбля сетей является слишком ресурсоемким. Действительно, эксплуатация ансамбля, состоящего из N сетей, будет требовать в N раз больше вычислительных ресурсов, чем одна сеть сходной архитектуры, содержащая объединенное (консолидированное) знание. Решение задачи объединения знаний нескольких ИНС ранее уже решалась неитеративными методами [3–6]. Точность классификации нейронной сети с объединенным знанием при этом оказывалась недостаточно высокой. В работе [5] нами была продемонстрирована возможность обмена знаниями между ИНС с помощью дополнительных генеративных моделей, что позволило значительно повысить точность классификации нейронной сети с объединенным знанием. В данной работе мы предлагаем развитие генеративного метода обмена знаниями путем улучшения качества генеративных моделей. Ключевым аспектом данного подхода является генерация искусственных тренировочных данных, как можно более схожих с реальными тренировочными данными, что в свою очередь открывает широкие возможности для взаимодействия различных ИНС через внешнюю память и сети передачи данных на любых расстояниях по существующей инфраструктуре.

Постановка задачи

Предположим, что несколько ИНС были натренированы на распознавание наборов изображений $\{x\}_1, \{x\}_2, \dots, \{x\}_N$ непересекающихся множеств классов $C_1 = \{c\}_1, C_2 = \{c\}_2, \dots, C_N = \{c\}_N$. Также предполагается, что после обучения ИНС доступ к изображениям x был утрачен. Такая ситуация может возникнуть при онлайн-обучении на потоке данных без их сохранения (например, для автономных транспортных средств) или в случае, когда каждый из наборов обучающих данных $\{x\}_i$ является конфиденциальным (например, для медицинских экспертных систем). Каждая сеть в итоге может распознавать только изображения определенного множества ранее выученных ею классов C_i (например, одна сеть умеет хорошо распознавать дома, другая – деревья, третья – животных). Необходимо, чтобы одна из сетей получила знания от остальных сетей и смогла распознавать изображения объединенного множества набора классов $C_{total} = C_1 \cup C_2 \cup \dots \cup C_N$ (например, деревья и животные, не забыв при этом как выглядят дома).

Один из путей решения поставленной задачи мы видим в следующем. На первом шаге необходимо выгрузить знания из всех имеющихся ИНС в виде объектов (изображений) $\tilde{x}$, схожих с теми, на которых они изначально обучались. Этим изображениям при-

сваиваются соответствующие метки классов c . Другими словами, предлагается создать реконструированные наборы обучающих данных (РНОД) $\{\tilde{x}, c\}_i$. Затем, объединив все реконструированные наборы, получить один результирующий набор данных $\{x, c\}_{total} = \{x, c\}_1 \cup \{x, c\}_2 \cup \dots \cup \{x, c\}_N$ и на нем дообучить целевую ИНС. Проблема данного подхода – это получение РНОД.

Если известна категория использованных ранее обучающих данных, то в качестве РНОД можно использовать образцы, схожие с образцами из исходного набора. Например, если ИНС обучались классификации изображений, то в качестве РНОД можно использовать базу ImageNet [7], содержащую 1000 классов изображений. Такой подход был использован в экспериментах по «воровству функционала» ИНС [8–10]. Изображения из огромной консолидирующей базы подавались на вход сети-«жертвы», подобно атаке подбора пароля по словарю. Полученные на выходе сети предсказания затем использовались как обучающие метки; пары «изображение-метка» использовались для обучения сети-«подделки».

В то же время, не всегда доступный набор изображений может адекватно аппроксимировать исходный набор обучающих данных. В некоторых случаях ИНС могут быть обучены классифицировать не изображения, а произвольный набор данных (звуки, электрические сигналы и т. д.). В этом случае для создания РНОД могут быть использованы генеративные методы.

Один из генеративных методов может быть основан на байесовском подходе [11]. Генеративный байесовский подход был ранее использован нами в задачах последовательного обучения ИНС [12]. Искусственные тренировочные данные создавались путем наблюдения активности нейронов на выходе ИНС. Недостаток данного подхода заключается в необходимости сохранения части изначальных тренировочных данных.

В последние годы были разработаны модели ИНС, которые могут воспроизводить данные, сходные как по типу, так и по вероятностному распределению с исходными тренировочными данными. Одним из видов подобных сетей являются генеративно-состязательные сети (ГСС) [13, 14]. ГСС успешно используются при генерации изображений, прогнозировании и в других приложениях. ГСС имеет в основе своей архитектуры генератор и дискриминатор. Последний пытается различить настоящие и сгенерированные данные. Первый, наоборот, пытается подобрать образцы той или иной категории с целью заставить дискриминатор принять их как настоящие входные данные. Генератор обученной ГСС в дальнейшем может быть использован для синтеза искусственных обучающих данных, что открывает дорогу к его применению для обмена знаниями между ИНС.

Другим типом генеративных моделей ИНС может служить вариационный автокодировщик (ВАК) [11, 14]. ВАК старается воспроизвести на выходе ту же самую ве-

роятность распределения данных $p(x)$, что и на входе. Отличие от обычных автокодировщиков состоит в том, что входной сигнал сначала переводится в случайную переменную, которая выбирается из определенного (обычно нормального) распределения.

И ГСС, и ВАК уже включают в свою архитектуру отдельные части, отвечающие за генерацию данных. Таким образом, ГСС и ВАК могут быть использованы для воссоздания РНОД и последующего обмена знаниями. Процедуру переноса знания с помощью генеративных моделей можно описать с помощью следующего алгоритма:

1. Каждый классификатор, знаниями которого предполагается делиться, дополняется генеративной сетью.
2. Классификаторы и генеративные сети обучаются на исходных наборах обучающих данных (ИНОД) $\{x, c\}_i$.
3. Для сетей, которые предполагают объединить знания, генеративные модели создают синтетические данные $\{\tilde{x}\}_i$.
4. Предварительно обученные классификаторы используются для восстановления меток для сгенерированных данных, формируя РНОД $\{\tilde{x}, c\}_i$. Набор данных с соответствующими метками $\{\tilde{x}, c\}_i$ представляет собой выгруженное из ИНС знание.

5. Реконструированные наборы данных каждой ИНС объединяются для создания полного набора $\{\tilde{x}, c\}_{total} = \{\tilde{x}, c\}_1 \cup \{\tilde{x}, c\}_2 \cup \dots \cup \{\tilde{x}, c\}_N$.

6. Сгенерированные данные с метками $\{\tilde{x}, c\}_{total}$ используются для дообучения целевой нейронной сети.

Нужно отдельно подчеркнуть, что создавать РНОД нужно не только для сетей, передающих знания, но и для сетей, получающих знания. Это связано с необходимостью устранения эффекта катастрофической интерференции: при обучении сети на новых данных она может быстро забыть все, что было приобретено ранее. Для предотвращения катастрофической интерференции не-

обходимо «напоминать» сети о предыдущих знаниях с помощью заранее созданного РНОД [12].

Качество обмена знаниями напрямую зависит от качества РНОД. Под качеством подразумевается схожесть вероятности распределения восстановленных данных $p(\tilde{x})$ с вероятностным распределением изначальных данных $p(x)$. В следующем разделе мы рассмотрим различные пути улучшения качества сгенерированных данных.

УЛУЧШЕНИЕ КАЧЕСТВА ГЕНЕРАТИВНЫХ МОДЕЛЕЙ

И ГСС, и ВАК обладают определенными недостатками. Например, изображения, полученные от ГСС, стремятся к одномодальному распределению, что характеризуется отсутствием их разнообразия. Типичным недостатком изображений, получаемых от ВАК, является их размытость. В данном исследовании мы остановимся подробнее на ВАК и рассмотрим ряд подходов по улучшению качества генерируемых изображений. Применяемые подходы условно можно разделить на две категории – функциональные и архитектурные.

1 Функциональные подходы

Данные подходы основаны на изменении функциональности процессов обучения и генерации и не затрагивают архитектуру самой сети.

1.1 Функции потерь *perceptual loss* и *reconstruction loss*

Стандартной функцией потерь при обучении ВАК является ошибка реконструкции (*reconstruction loss*), измеряющая отличие объекта на входе сети и восстановленного объекта на выходе (рис. 1а). Ошибка реконструкции может измеряться как среднеквадратичная ошибка или как бинарная перекрестная энтропия.

Вместо ошибки реконструкции в некоторых работах предлагается использовать «ошибку восприятия» (*perceptual loss*) [15]. Вкратце её основное отличие от

Рис. 1. Способы подсчета функции потерь: а) на основе ошибки реконструкции; б) на основе ошибки восприятия

Рис. 2. Сгенерированные реконструкции исходных изображений: а) при использовании функции reconstruction loss; б) при использовании функции perceptual loss

стандартной функции reconstruction loss заключается в том, что область подсчета ошибки вместо входного и выходного слоев ВАК включает скрытые слои дополнительного классификатора (рис. 16). Результатом применения данной функции ошибки являются менее размытые изображения рукописных цифр на выходе генератора (рис. 2). На рисунке 2 в качестве ИНОД использовался набор изображений рукописных цифр MNIST [16].

1.2 Отбор лучших генераций

Суть метода состоит в том, что во время процесса генерации в РНОД отправляются не все сгенерированные изображения, а только те, которым классификатор назначил высокую оценку (например, уверенность в выбранном классе выше 90%). Данный подход будет отбрасывать «неудавшиеся» реконструкции изображений по заданному порогу, прежде чем отправить их в РНОД.

2 Архитектурные подходы

Применяя различные архитектурные конфигурации модели классификатора и ВАК, можно влиять на результат обмена знаниями.

2.1 Классификатор отдельно от ВАК

Классификатор может находиться отдельно от ВАК и использоваться как для восстановления меток сгенерированных изображений, так и для отбора лучших генераций. Схематично архитектура данного способа размещения изложена на рисунке 3. При таком подходе появляется возможность обучать классификатор и ВАК независимо, в том числе и параллельно.

2.2 Классификатор в составе ВАК

Слои классификатора можно совместить для компактности вместе с ВАК, как схематично представлено на рисунке 4.

Для повышения эффективности данного подхода можно использовать поочередное обучение слоев, которое заключается в том, что вначале обуча-

ются только сверточные слои в составе классификатора. Затем сверточные слои классификатора замораживаются и обучается уже вся сеть сразу (кроме замороженных слоев).

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В наших экспериментах мы апробировали несколько моделей ВАК применительно к задаче переноса знаний между нейронными сетями. Две нейронные сети, ИНС_1 и ИНС_2, представлялись в виде совокупности двух подсетей – классификатора и ВАК. Это могла быть как единая архитектура, когда классификатор входит в состав слоев ВАК (рис. 4), так и раздельная архитектура, когда классификатор не имеет общих слоев с ВАК (рис. 3). Во время процедуры обмена знаниями генеративные подсети сетей ИНС_1 и ИНС_2 создавали РНОД_1 и РНОД_2, на которых дообучалась подсеть-классификатор ИНС_2. После обучения ИНС_2 должна была обладать объединенным знанием. Для оценки качества переноса знаний в ИНС_2 производилось тестирование обученного классификатора на объединенном тестовом наборе данных.

Рис. 3. Раздельная архитектура ИНС: ВАК (слева), классификатор (справа)

Рис. 4. Совмещенная архитектура ИНС

Рис. 5. Схема обмена знания между ИНС

Цель экспериментов – определить наиболее оптимальные функциональные и архитектурные подходы улучшения качества ВАК для целей переноса знаний между ИНС. Описанная выше схема графически представлена на рисунке 5.

Эксперименты проводились на наборе рукописных цифр MNIST [16]. Набор MNIST состоит из 60 тысяч тренировочных изображений десяти классов и 10 тысяч проверочных изображений. Исходный набор MNIST (ИНОД_0) был поделен случайным образом на 2 части по 5 классов в каждой. Полученные 2 части соответственно образовали ИНОД_1 и ИНОД_2.

ИНС_1 и ИНС_2 имели одинаковую архитектуру. Классификаторы и ВАК в ИНС_1 и ИНС_2 представляли собой сверточные сети. Классификатор, кодировщик и декодер вариационного автокодировщика содержали по четыре сверточных слоя. Кодировщик перево-

дил изображения MNIST в 32-мерный скрытый вектор признаков. ИНС_1 и ИНС_2 были натренированы в течение 20 эпох на наборах ИНОД_1 и ИНОД_2 соответственно с использованием стандартного оптимизатора Adam, входящего в состав пакета библиотеки Keras. Затем ВАК, входящие в ИНС_1 и ИНС_2, генерировали изображения (пять тысяч изображений на класс) с тем же вероятностным распределением, что и изначальные изображения MNIST. Пропуская сгенерированные изображения через соответствующие классификаторы, мы получали «мягкие» метки (вероятности соответствия сгенерированных изображений классам). Сгенерированные изображения совместно с соответствующими метками образовывали

ИНОД_1 и ИНОД_2. Далее ИНОД_1 и ИНОД_2 случайным образом смешивались между собой, образуя ИНОД_0 размерностью в 50 тысяч реконструированных изображений для дальнейшей тренировки ИНС_2. Обученная таким образом ИНС_2 тестировалась на исходном наборе проверочных данных (ИНОД) базы MNIST.

По данной схеме далее был проведен ряд экспериментов с применением различных вышеописанных модификаций ВАК. В случае использования ошибки восприятия (perceptual loss) мы использовали первый и второй сверточные слои из состава классификатора для вычисления ошибки. Каждый эксперимент повторялся 10 раз при случайной инициализации параметров для получения должной статистики. Результаты каждого подхода занесены в таблицу. Для сравнения классификация базы MNIST была выполнена и ансамблем ИНС_1 и ИНС_2 без их объединения (см. табл.).

Таблица

Точность классификации ИНОД для различных модификаций ВАК

Название подхода	средняя точность ± стандартное отклонение
Ансамбль классификаторов	0,936 ± 0,019
Классификатор в составе ВАК и reconstruction loss	0,953 ± 0,012
Классификатор в составе ВАК и perceptual loss	0,955 ± 0,016
Классификатор в составе ВАК и reconstruction+perceptual loss	0,939 ± 0,023
Классификатор отдельно от ВАК и perceptual loss	0,965 ± 0,026
Классификатор отдельно от ВАК, отбор лучших образцов и perceptual loss	0,977 ± 0,008

Ансамбль нейросетей показал худшую точность классификации (табл.) по сравнению с любым из рассмотренных генеративных методов объединения знаний. Большой процент ошибочной классификации возникает из-за отсутствия контрпримеров из других наборов данных при обучении отдельных сетей ИНС_1 и ИНС_2. Кроме того, эксперимент показал, что класси-

фикация набора MNIST ансамблем двух сетей занимает в ≈ 2 раза больше времени, чем классификация MNIST сетью с объединенным знанием. При объединении знаний большего количества ИНС выигрыш в быстродействии становится еще более значительным. Справедливости ради следует сказать, что сама процедура объединения знаний довольно времязатратна, но она

выполняется один раз и не используется при дальнейшей эксплуатации ИНС. Веса классификатора вместе с весами ВАК будут занимать в памяти в 2–3 раза больший объем, чем отдельный классификатор. Поэтому выигрыш в памяти по сравнению с ансамблем сетей возникнет при объединении знаний трех и более ИНС.

Из таблицы можно заключить, что к ухудшению результатов классификации как по среднему арифметическому, так и по стандартному отклонению приводит одновременное применение сразу двух функций подсчета ошибки (reconstruction loss + perceptual loss). Между собой же они не имеют сильных отличий в отношении средней точности классификации. Наличие отдельных классификатора и генеративной сети приводит к лучшей точности классификации по сравнению с ИНС, имеющими совмещенные слои. Метод с дополнительным отбором изображений в РНОД приводит к наибольшей точности. Для сравнения, классификатор с той же архитектурой, что ИНС_1 или ИНС_2, обученный сразу на всем наборе MNIST, показывает еще большую точность классификации $0,991 \pm 0,001$. Однако такой тип обучения требует наличия всего набора данных (ИНОД_1 и ИНОД_2), что по условиям нашей задачи является невозможным.

ЗАКЛЮЧЕНИЕ

В данной работе мы апробировали генеративные подходы по обмену знаниями между ИНС в задачах обучения с учителем. Генеративный подход обмена знаниями необходим в случае, когда тренировочные данные, использованные для обучения отдельных сетей, более недоступны. Методы были протестированы на задачах классификации изображений с использованием общедоступных баз данных. Генеративные методы объединения знаний показали большую точность, чем использование ансамбля сетей. Предполагается, что в перспективе генеративные методы смогут достигнуть точности обмена знаниями, сопоставимой с точностью нейронной сети, изначально натренированной на всем объеме данных. Для достижения подобной точности необходимо как можно аккуратнее реконструировать исходный набор данных. Мы показали, что, изменяя процедуру обучения генеративных ИНС или их архитектуру, удастся улучшить качество РНОД и, соответственно, качество обмена знаниями между нейронными сетями. Методы улучшения качества сгенерированных изображений были опробованы на ВАК.

Помимо преодоления ограничений в сфере машинного обучения, развитие генеративных методов обмена знаниями между ИНС в перспективе будет полезно в том числе и в смежных областях, таких как, например, медицина (диагностика заболеваний по рентгеновским снимкам, доступ к которым является медицинской тайной), транспорт (управление беспилотниками и обмен опытом между ними), информационная безопасность и защита конфиденциальных данных (создание обезличенных баз тренировочных данных).

СПИСОК ЛИТЕРАТУРЫ

1. Tan M. Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents // *Proceedings of the tenth international conference on machine learning*. 1993. pp. 330–337.
2. Learning to communicate with deep multi-agent reinforcement learning / J.N. Foerster, Y.M. Assael, N. De Freitas, S. Whiteson // *Advances in neural information processing systems*. 2016. pp. 2145–2153.
3. Smith J., Gashler M. An Investigation of How Neural Networks Learn from the Experiences of Peers Through Periodic Weight Averaging // *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2017. pp. 731–736.
4. Akhlaghi M.I., Sukhov S.V. Knowledge Fusion in Feedforward Artificial Neural Networks // *Neural Processing Letters*. 2018. № 1 (48). pp. 257–272.
5. The problem of neural networks communication / M.I. Leontev, V.I. Islenteva, A.V. Mikheev, K.V. Sviatov, S.V. Sukhov // *Journal of Physics: Conference Series*. 2019. (1368). pp. 52–33.
6. Leontev M.I., Islenteva V.I., Sukhov S.V. Non-iterative Knowledge Fusion in Deep Convolutional Neural Networks // *Neural Processing Letters*. 2020. № 1 (51), pp. 1–22.
7. ImageNet Large Scale Visual Recognition Challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei // *International Journal of Computer Vision*. 2015. № 3 (115). pp. 211–252.
8. Orekondy T., Schiele B., Fritz M. Knockoff Nets: Stealing Functionality of Black-Box Models // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. pp. 4949–4958.
9. Copycat CNN: Stealing Knowledge by Persuading Confession with Random Non-Labeled Data / J.R. Correia-Silva, R.F. Berriel, C. Badue, A.F. de Souza, T. Oliveira-Santos // *International Joint Conference on Neural Networks (IJCNN)*. 2018. pp. 1–8.
10. Shi Y., Sagduyu Y., Grushin A. How to steal a machine learning classifier with deep learning // *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*. 2017. pp. 1–5.
11. Kingma D.P., Welling M. Auto-Encoding Variational Bayes // *NeuroImage*. 2013. (147). pp. 302–313.
12. Overcoming Catastrophic Interference with Bayesian Learning and Stochastic Langevin Dynamics / M. Leontev, A. Mikheev, K. Sviatov, S. Sukhov // *International Symposium on Neural Networks*. 2019. pp. 370–378.
13. Progressive growing of GANs for improved quality, stability, and variation / T. Karras, T. Aila, S. Laine, J. Lehtinen // *arXiv preprint arXiv:1710.10196*. 2017.
14. Razavi A., Oord A. Van Den, Vinyals O. Generating diverse high resolution images with VQ-VAE-2 // *Advances in Neural Information Processing Systems*. 2019. pp. 14837–14847.

15. Hou X. Deep feature consistent variational autoencoder // 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). 2017. pp. 1133–1141.

16. LeCun Y., Corinna C., Chris B. MNIST handwritten digit database // URL: <http://yann.lecun.com/exdb/mnist/index.html> (дата обращения: 17.08.2013).

REFERENCES

1. Tan M. Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents. *Proceedings of the 10th International Conference on Machine Learning*, 1993. pp. 330–337.
2. Foerster J.N., Y.M. Assael, N. De Freitas, S. Whiteson. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. *Advances in Neural Information Processing Systems*, 2016, pp. 2145–2153.
3. Smith J., Gashler M. An Investigation of How Neural Networks Learn from the Experiences of Peers Through Periodic Weight Averaging. *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2017, pp. 731–736.
4. Akhlaghi M.I., Sukhov S.V. Knowledge Fusion in Feedforward Artificial Neural Networks. *Neural Processing Letters*, 2018, no. 1 (48), pp. 257–272.
5. Leontev M.I., Islenteva V.I., Mikheev A.V., Sviatov K.V., Sukhov S.V. The Problem of Neural Networks Communication. *Journal of Physics: Conference Series*, 2019, (1368), pp. 52–33.
6. Leontev M.I., Islenteva V.I., Sukhov S.V. Non-Iterative Knowledge Fusion in Deep Convolutional Neural Networks. *Neural Processing Letters*, 2020, no. 1(51), pp. 1–22.
7. Russakovsky O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 2015, no. 3 (115), pp. 211–252.
8. Orekondy T., Schiele B., Fritz M. Knockoff Nets: Stealing Functionality of Black-Box Models. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4949–4958.
9. Correia-Silva, J.R., R.F. Berriel, C. Badue, A.F. de Souza, T. Oliveira-Santos. Copycat CNN: Stealing Knowledge by Persuading Confession with Random Non-Labeled Data. *International Joint Conference on Neural Networks (IJCNN)*. 2018, pp. 1–8.
10. Shi Y., Sagduyu Y., Grushin A. How to Steal a Machine Learning Classifier with Deep Learning. *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*. 2017, pp. 1–5.
11. Kingma D.P., Welling M. Auto-Encoding Variational Bayes. *NeuroImage*, 2013, (147), pp. 302–313.
12. Leontev M., Mikheev A., Sviatov K., Sukhov S. Overcoming Catastrophic Interference with Bayesian Learning and Stochastic Langevin Dynamics. *Proc. of International Symposium on Neural Networks*. 2019, pp. 370–378.
13. Karras T., T. Aila, S. Laine, J. Lehtinen. *Progressive growing of GANs for Improved Quality, Stability, and Variation*. arXiv preprint arXiv:1710.10196. 2017.
14. Razavi A., Oord A. Van Den, Vinyals O. Generating Diverse High Resolution Images with VQ-VAE-2. *Advances in Neural Information Processing Systems*, 2019, pp. 14837–14847.
15. Hou X. Deep Feature Consistent Variational Autoencoder. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017, pp. 1133–1141.
16. LeCun Y., Corinna C., Chris B. *MNIST Handwritten Digit Database*. Available at: <http://yann.lecun.com/exdb/mnist/index.html> (accessed 17.08.2013).