

ЭЛЕКТРОННАЯ КОМПОНЕНТНАЯ БАЗА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ УПРАВЛЕНИЯ

УДК 004.27

КЛАССИФИКАЦИЯ ЗВУКА ИМПУЛЬСНОЙ НЕЙРОННОЙ СЕТЬЮ, УДОВЛЕТВОРЯЮЩЕЙ ОГРАНИЧЕНИЯМ НЕЙРОПРОЦЕССОРА АЛТАЙ И ПРИ ЭТОМ ИСПОЛЬЗУЮЩЕЙ ПОЛНЫЙ ДИНАМИЧЕСКИЙ ДИАПАЗОН ВЕСОВ

© 2022 г. А. Г. Сбоев^{1,2,*}, Р. Б. Рыбка¹, А. В. Серенко¹,
А. Г. Селиванов¹, А. О. Софонов³

¹ Национальный исследовательский центр “Курчатовский институт”, Москва, Россия

² Национальный исследовательский ядерный университет “МИФИ”, Москва, Россия

³ ООО “Мотив НТ”, Новосибирск, Россия

*E-mail: Sboev_AG@nrcki.ru

Поступила в редакцию 15.03.2022 г.

После доработки 20.03.2022 г.

Принята к публикации 20.03.2022 г.

Нейроморфные вычислительные устройства являются аппаратной базой для реализации импульсных (импульсных) нейронных сетей с чрезвычайно низким энергопотреблением. Это делает актуальной разработку методов получения импульсных нейронных сетей для решения задач машинного обучения. Представлен метод получения импульсной нейронной сети, удовлетворяющей ограничениям для размещения на нейропроцессоре “Алтай”: синаптические веса являются 8-битными целыми числами, при этом веса входящих синапсов каждого нейрона могут по модулю принимать не более заданного количества уникальных значений. Предложенный метод основан на преобразовании нейронной сети, обученной с учетом ограничений, в импульсную. Учет ограничений обеспечивается с помощью квантования и кластеризации синаптических весов в отличие от существующих подходов, в которых рассматриваются сети с тернарными весами, не использующими весь доступный динамический диапазон весов. При преобразовании в импульсную сеть пороговые потенциалы импульсных нейронов устанавливаются так, чтобы нормировать количество испускаемых импульсов. На примере задачи классификации городских шумов импульсная сеть, полученная предложенным методом, достигает точности классификации от 42 до 67% (по 10 разбиениям кросс-валидации, в среднем 54%), что на 5% ниже, чем исходная сеть до учета ограничений нейропроцессора.

DOI: 10.56304/S2782375X22030093

ВВЕДЕНИЕ

Малоэнергозатратные нейроморфные вычислительные устройства, основанные подобно биологическим нейронным сетям на передаче информации посредством дискретных импульсов-спайков [1], актуальны для целого ряда практических приложений, включая управление робототехническими и другими автономными устройствами, критически зависимыми от энергопотребления. Одним из направлений настройки нейрочипа под конкретную задачу является размещение на нем нейронной сети, полученной на стороннем вычислительном устройстве с учетом ограничений, налагаемых конструкцией используемого нейрочипа.

Каноническим примером нейроморфной архитектуры является IBM TrueNorth [2], позволяющий перенести на аппаратную реализацию про-

цесс вычисления выходов (*inference*) обученной импульсной нейронной сети. Сходные с TrueNorth принципы работы и ограничения на реализуемые нейронные сети используются в последующих его аналогах [3], в том числе в нейропроцессоре “Алтай”, недавно разработанном компанией “Мотив Нейроморфные Технологии”.

В нейропроцессорах рассматриваемого вида синапс характеризуется типом аксона, и каждый нейрон ставит в соответствие каждому типу аксона целое число, характеризующее силу синаптической связи. У TrueNorth четыре типа аксона, а у “Алтая” – четыре для возбуждающих связей и четыре для тормозящих. Таким образом, ограничения, налагаемые нейропроцессором на веса нейронной сети, состоят в том, что, во-первых, веса целочисленные и имеют ограниченную разрядность, и, во-вторых, веса входящих синапсов

каждого нейрона могут принимать ограниченное количество уникальных значений (но набор значений, которые могут принимать веса, можно задавать независимо для каждого нейрона). Например, в нейропроцессоре “Алтай” каждому нейрону можно задать свой набор из четырех 8-битных положительных значений веса (одно из которых имеет смысл задать равным нулю), тогда вес какого-либо синапса этого нейрона может принимать любое из заданных значений со знаком “плюс” или со знаком “минус” (всего семь уникальных значений, если среди заданного набора значений есть ноль). Набор возможных значений весов нейрона можно увеличить с помощью поддерживаемой нейропроцессором “Алтай” синхронизации мембранных потенциалов нескольких нейронов, которая эффективно позволяет использовать для размещения одного нейрона исходной сети наборы весов, принадлежащие нескольким нейронам на нейропроцессоре.

Цель данной работы – исследовать эффективность импульсной нейронной сети, удовлетворяющей ограничениям для реализации на нейропроцессоре “Алтай”.

В качестве тестовой задачи рассмотрена классификация звуков Urban Sound. Анализ звуковых сигналов является востребованной областью применения автономных устройств, требующих энергоэффективной реализации нейронных сетей. Набор данных UrbanSound8k выбран по причине того, что классификация данных из этого набора является достаточно сложной задачей, при решении которой можно оценить эффективность нейронной сети: максимальная точность, полученная в литературе без применения аугментации данных и предобучения на других наборах данных, составляет 73% и достигнута сверточной нейронной сетью [4].

Получение импульсной нейронной сети, удовлетворяющей ограничениям нейропроцессора, проведено с помощью подхода “train-and-constrain” [5], позволяющего воспользоваться мощью и универсальностью традиционных нейронных сетей, обучаемых обратным распространением ошибки:

- обучение традиционной нейронной сети;
- наложение на настраиваемые параметры (синаптические веса) нейронной сети ограничений нейропроцессора;
- преобразование нейронной сети в импульсную нейронную сеть, где отображение последовательности входных импульсов в последовательности выходных импульсов должно воспроизводить отображение входных векторов в выходные векторы традиционной нейронной сети.

Однако существующая реализация получения модели нейронной сети по принципу train-and-constrain, предложенная разработчиками True-

North, – *Energy-efficient deep neuromorphic network (Eedn)* [5] – не вполне использует даже эти ограниченные возможности нейрочипа: веса, которые настраивает Eedn, тернарные, они принимают значения только -1 , 0 и 1 . Обойти это ограничение и расширить диапазон значений весов нейронной сети можно дублированием синаптических связей: направляя импульс одного нейрона одновременно на несколько входных синапсов другого, можно добиться того, что эффективный вес связи между этими нейронами будет произвольной комбинацией указанных чисел. Однако такой подход требует большего количества синаптических связей, а следовательно, для размещения сети на чипе потребуется больше ядер.

В данной работе показано, что модель импульсной нейронной сети, удовлетворяющую ограничениям по точности и количеству весов, можно получить и без дублирования входов, если вместо тернарных весов задействовать весь набор из четырех значений, которые могут принимать веса, и задавать эти четыре значения независимо для каждого нейрона.

При этом звукозаписи из набора данных UrbanSound кодируются спектральными признаками. Импульсная нейронная сеть получена на основе традиционной нейронной сети. Ограничение на разрядность значений весов налагается на полученную нейронную сеть с помощью квантования весов. Ограничение на количество уникальных значений весов нейрона налагается с помощью кластеризации весов. При преобразовании традиционной нейронной сети в импульсную, чтобы не изменять веса, уже удовлетворяющие ограничениям нейропроцессора, нормировка выходов импульсной сети [6] осуществляется не с помощью нормировки весов, как описано в литературе, а путем задания порогов нейронов.

МЕТОДЫ

В качестве тестовой классификационной задачи для исследования применимости предложенного подхода рассмотрен набор данных UrbanSound8k [7]. Он состоит из 8732 образцов звука длительностью до 4 с из 10 классов, которые нужно классифицировать по 10 типам городских звуков: *air conditioner*, *car horn*, *children playing*, *dog bark*, *drilling*, *engine idling*, *gun shot*, *jackhammer*, *siren* и *street music* (рис. 1). Примеры форм звуковых сигналов для каждого класса приведены на рис. 2, логарифмических спектрограмм звуков каждого класса – на рис. 3.

В [7] этот набор данных сопровождали разделением на 10 подмножеств (фолдов), рекомендованных для использования при кросс-валидации классификаторов. Это деление устроено так, чтобы образцы, которые являются фрагментами

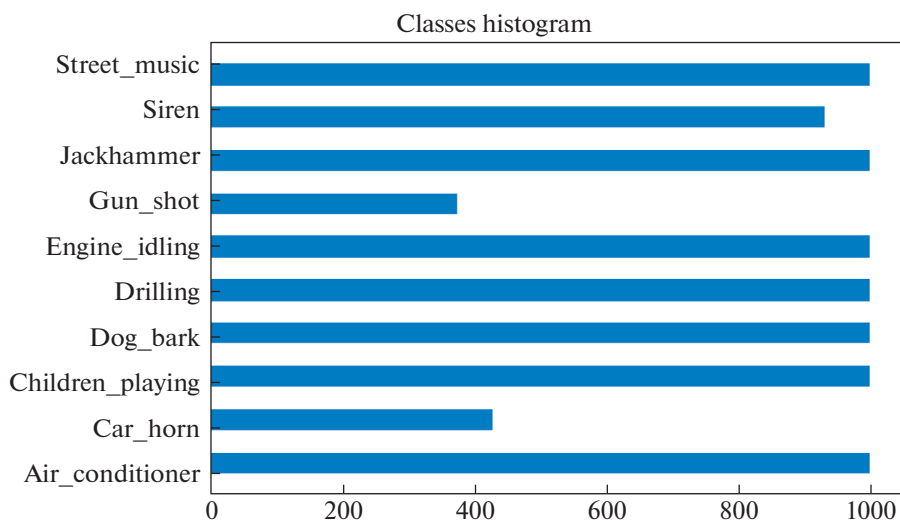


Рис. 1. Количество звукозаписей каждого из классов в наборе данных UrbanSound8k.

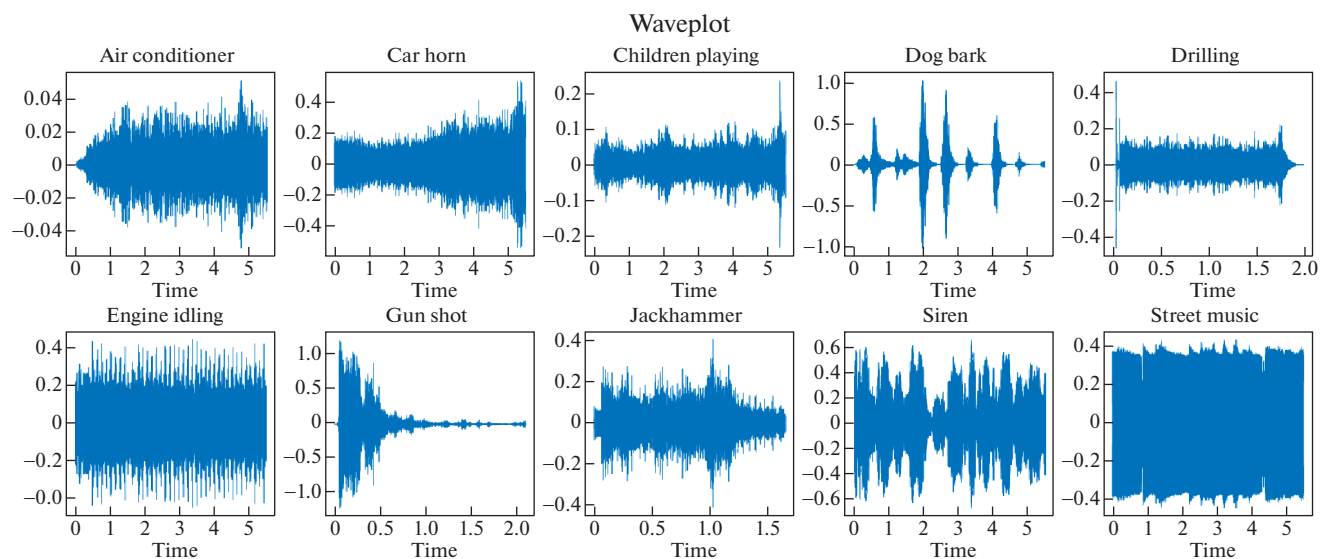


Рис. 2. Примеры форм звуковых сигналов каждого из классов.

одних и тех же звукозаписей, не попали одновременно в обучающую и тестовую выборки одного разбиения кросс-валидации и не привели таким образом к переоцененной точности классификатора. Проведенные эксперименты следуют этому рекомендованному разделению.

Звукозаписи подавали на нейронную сеть следующим образом: каждый образец из набора данных делили на фрагменты длиной 500 мс со сдвигом на 250 мс. Фрагменты кодировали и подавали на нейронную сеть как отдельные входные векторы. На этапе обучения на нейронную сеть подавали фрагменты, принадлежащие звукозаписям из обучающей выборки. Эталонным классом фраг-

мента является класс той звукозаписи, к которой он принадлежит. При классификации звукозаписей обученной нейронной сетью класс определяется как медианное значение меток классов, полученных для всех фрагментов этой звукозаписи (рис. 4).

Для представления звука использовали набор спектральных характеристик, взятый аналогично подходу [4]:

- мел-спектр;
- мел-частотные кепстральные коэффициенты (количество используемых коэффициентов – 128);

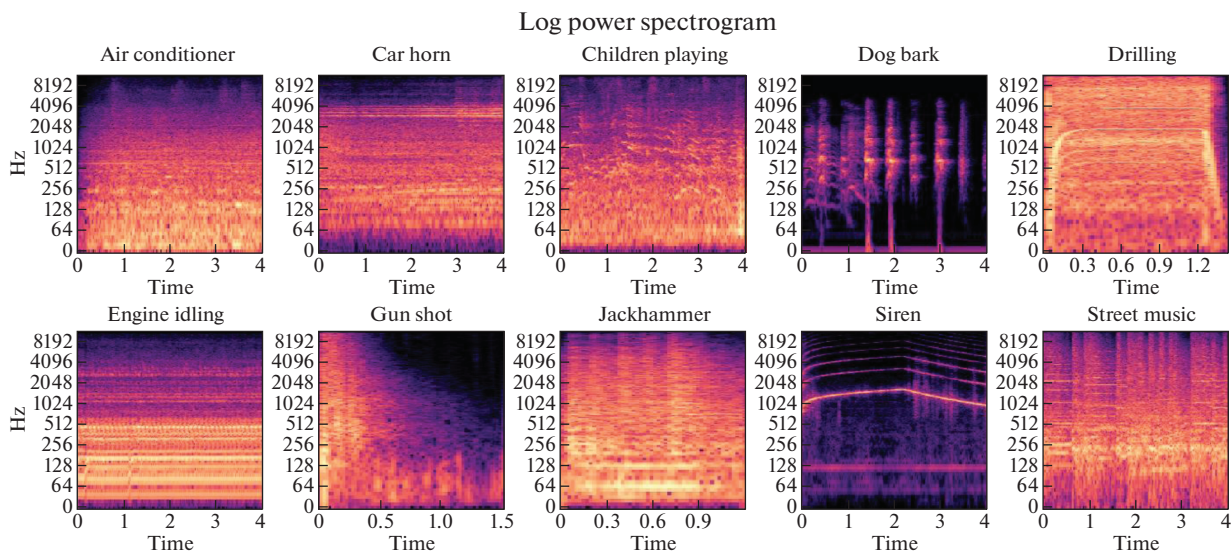


Рис. 3. Примеры логарифмических спектрограмм звуковых сигналов каждого из классов.

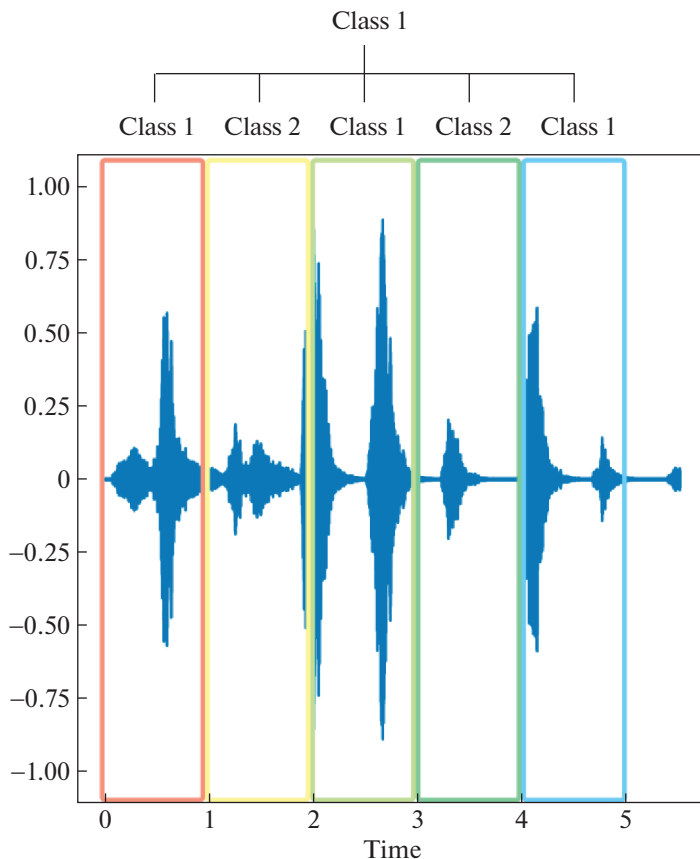


Рис. 4. Пример определения класса у поделенного на фрагменты примера.

- центроиды сетки тональностей (*tonal centroid features, tonnetz*) [8];
- хромограммы [9];
- спектральный контраст [10].

Фрагменты звукозаписей длительностью 0.5 с, подаваемые на вход нейронной сети, кодировали перечисленными выше спектральными характеристиками с помощью библиотеки *librosa*; при этом их разбивали на окна по 0.128 с, для которых

определяли перечисленные выше характеристики (всего 281 число) и каждую из характеристик усредняли между окнами. Затем каждую характеристику нормировали так, чтобы в обучающей выборке она принимала значения от 0 до 1 (нормировка *min-max-scale*).

Для традиционной нейронной сети, на основе которой далее будет получена импульсная сеть, выбрана следующая полносвязная схема: входы (конкатенация перечисленных в предыдущем разделе признаков всего 281) – слой из 1000 – из 500 – из 100 – из 10 нейронов. Связи между 1000- и 500-нейронными, а также между 500- и 100-нейронными слоями при обучении подвергали *dropout* с вероятностью 0.5. Нейроны всех слоев, кроме выходного, имеют функцию активации *ReLU*, выходного – *softmax*. Параметров *bias* нейроны не имеют.

Обучение нейронной сети проводили методом обратного распространения ошибки с использованием оптимизатора Adam, функции потерь *categorical cross-entropy*, и раннего останова при условии неухудшения функции потерь в течение четырех эпох.

После обучения веса зафиксировали в значениях из той эпохи, в которой достигалась наименьшая функция потерь, и для полученной сети проводили квантование весов.

Для того чтобы привести веса нейронной сети к значениям, позволяющим размещение в памяти нейропроцессора, веса подвергали дискретизации и приведению к целым числам.

Дискретизацию весов проводили с помощью процедуры дообучения с ограничениями *quantization-aware training* [11], реализованной в программном пакете *Tensorflow model optimization*, чтобы снизить влияние дискретизации на точность классификации. В ходе данной процедуры дообучения выходы нейронной сети вычисляли на основе дискретизованных весов. После каждого акта вычисления функции ошибки и изменения весов соответственно градиенту функции ошибки веса снова дискретизировали. Дискретизация приводит веса каждого слоя к набору значений, расположенных через равные интервалы в диапазоне от минимального до максимального значения веса нейронов данного слоя в исходной сети. Интервал дискретизации весов выбирали так, чтобы количество возможных значений дискретизованного веса равнялось 256. Таким образом, веса, полученные в результате дообучения с дискретизацией, остаются вещественными, и диапазон их значений не меняется, но количество дискретных значений, которые может принимать вес, соответствует тому, которое может содержаться в 8-битной ячейке памяти нейропроцессора. Дообучение проводили с теми же параметрами, что и обучение, в течение трех эпох.

Чтобы привести веса к целочисленным значениям, полученные после дообучения веса были умножены на коэффициент шкалирования, выбранный для каждого слоя так, чтобы минимальное для данного слоя значение дискретизованного веса, умноженное на коэффициент шкалирования, равнялось -128 , а максимальное значение веса слоя после шкалирования стало равным 127. Таким образом, полученные в результате квантования веса каждого слоя могут принимать 256 значений, расположенных через равные интервалы в диапазоне от -128 до 127, т.е. являются 8-битными целыми числами.

Выполнение ограничения на количество уникальных абсолютных значений весов нейрона обеспечивается с помощью дообучения с кластеризацией, основанного на процедуре *clustering-aware training* [12], реализованной в пакете *Tensorflow model optimization* и модифицированной так, чтобы применять кластеризацию отдельно для каждого нейрона, применять ее к абсолютным значениям весов и после кластеризации округлять веса до целых.

Перед дообучением для весов, полученных в результате квантования, выполняли следующий алгоритм:

- для весов каждого нейрона получали их абсолютные значения;
- кластеризовали абсолютные значения весов методом *k-means*;
- заменяли каждый вес на центроид кластера, к которому было отнесено абсолютное значение этого веса, сохраняя при этом исходный знак веса;
- округляли полученное значение веса до целого.

В ходе дообучения выходы нейронной сети вычисляли на основе весов, кластеризованных по приведенному алгоритму. Изменение весов, вычисленное на основе градиента функции потерь, применяли вместо весов к центроидам кластеров, к которым веса принадлежат. После каждого изменения центроиды кластеров вновь округляли до целых, и весам присваиваются значения ближайших центроидов (с сохранением знаков весов).

Дообучение с кластеризацией проводили с теми же параметрами, что и исходное обучение, в течение одной эпохи. Полученные в результате веса остаются 8-битными целыми числами, но вдобавок к этому абсолютные значения весов каждого нейрона принимают не более заданного количества значений, которое равно числу кластеров при кластеризации.

После кластеризации нейронную сеть преобразовали в импульсную с помощью программного пакета *SNN Conversion Toolbox* [13]. Импульсная сеть имеет точно такую же топологию, что и

исходная традиционная сеть, и те же синаптические веса, только формальные нейроны заменяются импульсными нейронами *Integrate-and-Fire*, без утечки и без рефрактерности.

Один входной вектор подавали на сеть в течение 200 мс, при этом моделирование импульсной сети проводили с временным шагом $\Delta t = 1$ мс (что соответствует временному шагу TrueNorth и нейропроцессора «Алтай»). Такая длительность подачи выбрана для того, чтобы обеспечить потенциальную возможность классификации звука в режиме реального времени: поскольку классификация проводится по фрагментам длительностью 500 мс, пересекающимся на половину своей длительности, то следующий фрагмент поступит на сеть через 250 мс после предыдущего. При длительности подачи одного входного вектора в 200 мс сеть заведомо успеет обработать его до подачи следующего.

Входные данные кодировали частотным способом: на i -й входной синапс нейрона подается случайная последовательность импульсов, имеющая на каждом временном шаге вероятность импульса $x_i \cdot v \cdot \Delta t$, где x_i — величина i -го компонента входного вектора, а $v = 1000$ Гц. Таким образом, минимально возможная величина x_i (равная 0 после нормировки входных данных, описанной выше) соответствует отсутствию импульсов, а максимальная величина компонент входного вектора (равная 1) соответствует наличию импульса на каждом временном шаге.

Параметры импульсной сети выбирали таким образом, чтобы минимизировать потери точности, вызванные переходом от исходной нейронной сети с непрерывными выходными значениями к импульсной нейронной сети, выходов нейронов которой является количество импульсов, испущенное в течение подачи входного вектора, — дискретное число от 0 до 200. Потери точности в данном случае вызваны двумя возможными причинами, по которым количество спайков импульсного нейрона может не быть пропорционально выходу соответствующего нейрона исходной сети: невозможностью импульсного нейрона испустить более одного импульса за временной шаг [14] (в результате чего может оказаться, что нейроны, выдававшие различные величины активации в исходной сети, после преобразования в импульсную сеть будут все испускать по 200 импульсов), а также дискретностью импульсного представления выхода сети (в результате чего незначительно различающиеся величины активации нейронов исходной сети могут соответствовать одному и тому же количеству спайков после преобразования в импульсную сеть).

Для того чтобы минимизировать обе причины потерь, пороговый потенциал нейрона θ выбирали независимо для каждого слоя следующим об-

разом. У исходной сети вычисляли заданный процентиль P ($\{w \cdot x\}$) произведения входного вектора x на вектор весов w (до применения функции активации) среди всех нейронов данного слоя и по всем векторам тренировочной выборки. θ устанавливали равным данному процентилю, если рассматриваемый слой является первым в сети, а если нет, то дополнительно деленному на вычисленное аналогичным образом значение θ для предыдущего слоя.

В результате максимальная активация нейрона исходной сети (за вычетом только заданной доли самых высоких активаций) соответствует максимально возможному спайкованию нейрона после преобразования в импульсную сеть. Отбрасывание заданной доли максимальных выходов приводит к тому, что соответствующие нейроны импульсной сети во время подачи соответствующих входных векторов будут все испускать по 200 спайков, что может вызвать ошибки в классификации этих входных векторов, но зато чем большая доля выходов отбрасывается при вычислении порога, тем меньше ошибки, вызванные дискретностью импульсного представления, во время подачи остальных входных векторов.

Точность решения классификационной задачи оценивали с применением кросс-валидации: независимо обучали 10 нейронных сетей, и для каждой сети свое подмножество из 10 фолдов набора данных использовали в качестве тестовой выборки, а остальные 9 фолдов — в качестве обучающей. При этом каждую кластеризованную и импульсную сеть тестировали на той же тестовой выборке, что и исходную сеть, из которой они были получены. Для дообучения сети при квантовании и кластеризации весов и для вычисления нормировочных коэффициентов при преобразовании в импульсную сеть использовали ту же обучающую выборку, что и для обучения исходной сети.

Перед квантованием, кластеризацией и преобразованием в импульсную для исходной сети было проведено исследование влияния различных комбинаций входных признаков на точность сети.

Чтобы оценить потери точности на каждом из этапов получения импульсной сети, точность классификации оценивали для исходной сети, после квантования, кластеризации и преобразования в импульсную сеть.

Кластеризацию проводили с различным числом кластеров, при этом у одного из кластеров центроид был фиксирован равным нулю. Рассматривали количество кластеров, равное четырём, что соответствует количеству уникальных значений, которые вес может принимать по модулю в нейропроцессоре «Алтай», а также семи, что соответствует использованию весов двух нейронов нейропроцессора для размещения каждого

Таблица 1. Средняя (по фолдам кросс-валидации) точность сети на разных наборах признаков

Признаки					Accuracy, %	F1-macro, %
MFCC	Mel	Contrast	Chroma	Tonnetz		
✓	✓				53	53
✓	✓	✓			56	57
✓	✓	✓	✓		56	57
✓	✓			✓	54	54
✓	✓	✓	✓		58	60
✓	✓	✓	✓	✓	56	57
✓	✓	✓		✓	56	57
✓		✓	✓		55	55
✓		✓	✓	✓	56	57
✓	✓	✓	✓	✓	60	60

Таблица 2. Средняя, минимальная и максимальная (по фолдам кроссвалидации) метрики точности для исходной нейронной сети после квантования весов, кластеризации весов и после преобразования в импульсную сеть

Нейронная сеть	Accuracy, %			F1-macro, %			F1-micro, %		
	max	mean	min	max	mean	min	max	mean	min
Исходная	67.95	57.94	49.62	69.07	58.84	47.04	67.95	57.94	49.62
Квантованная	70.47	63.59	51.65	68.43	60.15	30.68	70.47	63.59	51.65
Кластеризованная (4 кластера)	55.57	44.22	35.94	56.78	40.02	27.73	55.57	44.22	35.94
Кластеризованная (7 кластеров)	61.84	52.82	43.46	64.48	52.84	41.12	61.84	52.82	43.46
Кластеризованная (10 кластеров)	65.7	55.71	44.15	66.5	56.89	44.0	65.7	55.71	44.15
Импульсная	67.1	53.5	41.62	67.45	54.5	37.92	67.1	53.5	41.62

нейрона исходной сети (при этом у обоих нейронов одно значение из набора весов фиксированно равно нулю, а остальные три выбираются независимо), и десяти, что соответствует размещению одного нейрона исходной сети на трех нейронах “Алтая”.

Преобразование в импульсную сеть проводили с различными величинами процента: 99.9% (что соответствует отбрасыванию 0.1% максимальных активаций при вычислении порога нейрона), 99 и 98%. Кроме того, исследовали влияние применения *softmax* к количеству спайков, испущенных выходным слоем импульсной сети.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Результаты тестирования исходной нейронной сети, обученной на различных комбинациях входных признаков (табл. 1), показывают, что наилучшая точность достигается при использовании всех рассматриваемых признаков, что соответствует результатам [4]. Поэтому дальнейшие результаты получены с использованием всех входных признаков. Точность исходной нейронной сети не отличается значительно от литератур-

ных результатов, полученных кросс-валидацией с рекомендованным разбиением на фолды. Максимальная точность, достигнутая нейронными сетями без применения аугментации данных и предобучения на других наборах данных, составляет 73% [4].

При преобразовании нейронной сети в импульсную оптимальная доля активаций исходной сети, используемых при определении порогов импульсных нейронов, равна 99%, и далее приведены результаты, полученные с ней.

Применение *softmax* к количествам спайков выходного слоя приводит к результатам классификации, полностью идентичным результатам без применения *softmax*, ввиду ее монотонности. Поэтому приведенные далее результаты получены без применения *softmax*, с классификацией фрагмента звукозаписи по выходному нейрону, испустившему больше всего импульсов.

Средняя, минимальная и максимальная (по фолдам кросс-валидации) точности для каждого из этапов получения импульсной сети по метрикам *accuracy*, *F1-macro* и *F1-micro* приведены в табл. 2.

Квантование весов сети не приводит к существенному изменению ее точности. Кластеризация весов вызывает потери точности, которые зависят от количества используемых кластеров.

ЗАКЛЮЧЕНИЕ

Предложен алгоритм получения импульсной нейронной сети, удовлетворяющей ограничениям нейропроцессора “Алтай” на синаптические веса. Точность нейронной сети, полученной с помощью предложенной процедуры, зависит от количества уникальных значений, которые могут принимать по модулю веса. При 10 значениях точность классификации звуков Urban Sound составила 54%, что на 5% ниже исходной формальной нейронной сети аналогичной топологии без ограничений нейропроцессора.

Предложенная технология в перспективе может применяться в составе аппаратного *end-to-end* решения для обработки и анализа звуковых сигналов. Дальнейшие исследования будут направлены на разработку методов прямого импульсного кодирования звукозаписей без предварительного преобразования в высокоуровневые признаки.

Вычислительные эксперименты проведены с использованием оборудования центра коллективного пользования “Комплекс моделирования и обработки данных исследовательских установок мега-класса” НИЦ “Курчатовский институт”, <http://ckp.nrcki.ru/>.

СПИСОК ЛИТЕРАТУРЫ

1. *Paugam-Moisy H., Bohte S.M.* // Handbook of Natural Computing / Eds. Rozenberg G. et al. Berlin; Heidelberg: Springer, 2012. P. 335. https://doi.org/10.1007/978-3-540-92910-9_10
2. *Merolla P.A., Arthur J.V., Alvarez-Icaza R. et al.* // Science. 2014. V. 345. № 6197. P. 668. <https://doi.org/10.1126/science.1254642>
3. *Zou C.* // China Semiconductor Technology International Conference (CSTIC). 2021. P. 1. <https://doi.org/10.1109/CSTIC52283.2021.9461550>
4. *Salamon J., Bello J.P.* // IEEE Signal Proces. Lett. 2017. V. 24. № 3. P. 279.
5. *Steven K. Esser, Paul A. Merolla, John V. Arthur et al.* // Proceedings of the National Academy of Sciences. 2016. P. 201604850.
6. *Peter U. Diehl, Guido Zarrella, Andrew Cassidy et al.* // IEEE International Conference on Rebooting Computing (ICRC) / IEEE. 2016. P. 1–8.
7. *Salamon Justin, Jacoby Christopher, Bello Juan Pablo* // Proceedings of the 22nd ACM International Conference on Multimedia, New York, USA: Association for Computing Machinery, 2014. P. 1041. <https://doi.org/10.1145/2647868.2655045>
8. *Harte Christopher, Sandler Mark, Gasser Martin.* // Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, New York, USA: Association for Computing Machinery, 2006. P. 21. <https://doi.org/10.1145/1178723.1178727>
9. *Ellis Daniel P.W.* Chroma feature analysis and synthesis. 2007. <http://labrosa.ee.columbia.edu/matlab/chroma-syn/>
10. *Jiang Dan-Ning, Lu Lie, Tao Jianhua et al.* // Music Type Classification by Spectral Contrast Feature. 2002. V. 1. P. 113. <https://doi.org/10.1109/ICME.2002.1035731>
11. *Jacob Benoit, Kligys Skirmantas, Chen Bo et al.* // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 2704.
12. *Han Song, Mao Huizi, Dally William J.* // arXiv preprint. arXiv:1510.00149. 2015.
13. *Rueckauer Bodo, Lungu Iulia-Alexandra, Hu Yuhuang et al.* // Frontiers in Neuroscience. 2017. V. 11. P. 682. <https://doi.org/10.3389/fnins.2017.00682>
14. *Diehl Peter, Neil Dan, Binas Jonathan et al.* // IEEE International Joint Conference on Neural Networks (IJCNN). 2015. <https://doi.org/10.1109/IJCNN.2015.7280696>