УДК 004.855.5

Восстановление сигнала с помощью импульсного мультисенсорного предиктивного кодирования

А. Т. Хакимов1, Б.Б. Батуев2

1Московский физико-технический институт (национальный исследовательский университет)

2Институт радиотехники и электроники имени В.А. Котельникова Российской академии наук

В данной работе предложен метод обучения импульсной нейронной сети с использованием синаптической пластичности и механизма распада весов для решения задачи одновременного восстановления аудио- и визуальных сигналов на основе предиктивного кодирования. Основу предложенного подхода составляет концепция предиктивного кодирования [4], которая предполагает, что нейронная сеть предсказывает входные данные на основе текущего состояния и корректирует свои предсказания с учетом ошибки. Метод восстановления сигналов основан на биологически вдохновленных подходах, таких как кодирование через выборки отсчетов [2], и использует архитектуру, схожую с генеративными автоэнкодерами [1].

Для нейронов была выбрана модель leaky-integrate-and-fire (LIF) [3], которая позволяет моделировать динамику нейронов с утечкой мембранного потенциала, обеспечивая более реалистичное поведение нейронов, учитывая их биологические свойства. Входные нейроны обрабатывают сигналы из двух сенсорных подсистем: визуальной и аудиальной. Их мембранный потенциал 𝑣 изменяется во времени по следующему дифференциальному уравнению:

, (1)

где – мембранный потенциал нейрона, – постоянная времени, а возвращаемая ошибка предсказания определяется как:

, (2)

где – входной временной сигнал изображения или аудио для нейрона с индексом , a – обратная связь от скрытого слоя, корректирующее текущее представление входа. В данной модели было взято = 10 мс для расширения частотного диапазона спайков и для более быстрого обновления весов.

Скрытые нейроны в свою очередь получают суммарный вход от обеих сенсорных подсистем, их динамика описывается уравнением:

, (3)

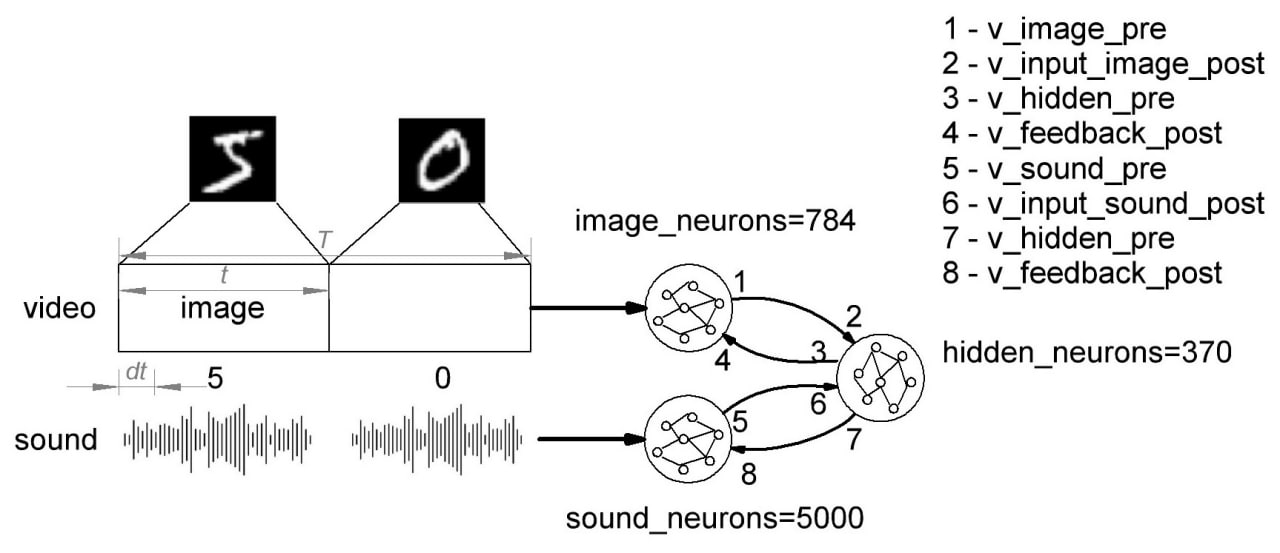
где и представляют вклад (текущий мембранный потенциал на нейроне) входных сигналов от визуальной и аудиоподсистем соответственно. Эти величины вычисляются через синаптические связи, в которых вклад от предшествующего нейрона определяется произведением синаптического веса на его активность. Аналогичным образом задаются обратные синапсы, что обеспечивает механизм предиктивного кодирования. В данной модели размер скрытого слоя был равен 370 нейронам, количество входных нейронов для аудиосигнала было равно 5000, количество входных нейронов для изображения было равно 784.

Рис. 1. Схема мультисенсорного предиктивного кодирования. Мультисенсорность реализована за счет интегрирования аудио и видео сигналов в скрытый слой.

(4)

Синаптические веса в модели обновляются по модифицированную правилу Хебба, которое включает механизм распада весов. Это правило записывается следующим образом:

, (5)

где – скорость обучения, и – активности пред- и постсинаптических нейронов соответственно, – коэффициент распада, предотвращающий неограниченный рост синаптических весов при положительной корреляции между активностями.

В качестве демонстрационного примера восстановления изображения и аудиоряда были использованы датасеты MNIST [5] (изображения 28x28) и Free Spoken Digit Dataset [8] (аудиоряды произношения цифр) (рис.2).

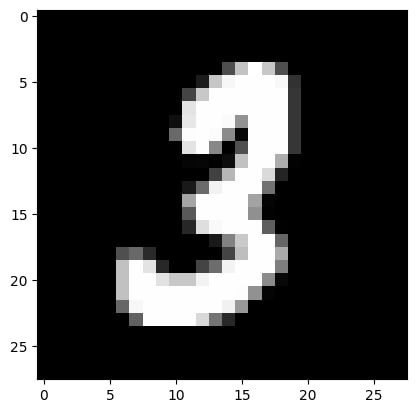


Рис.2. Оригинал изображения и реконструированное (восстановленное) изображение для цифры 3

Для оценки качества восстановления использовалась среднеквадратичная ошибка (MSE), которая для изображений вычислялась между оригинальным и восстановленным изображениями, а для аудиоряда между соответствующими временными отсчетами оригинального и восстановленного аудиосигналов. Для изображений полученные значения среднеквадратичной ошибки (MSE) в диапазоне от 0.02 до 0.04 и высокая корреляция в диапазоне от 0.96 до 0.98 указывают на то, что модель успешно восстанавливает структуру цифр, сохраняя их геометрические особенности. Низкий уровень MSE подтверждает, что пиксельные паттерны набора данных MNIST сохраняются через скрытый слой, несмотря на спайковое кодирование. Для аудио более высокий MSE в диапазоне от 0.07 до 0.10 объясняется сложностью временных зависимостей в аудиосигналах. Тем не менее, корреляция в диапазоне от 0.89 до 0.94 показывает, что модель улавливает основные частотные и амплитудные паттерны (рис.3), хотя сильно шумит в случае отсутствия сигнала.

Для оценки точности представления сигнала на скрытом слое использовалась модель логистической регрессии. Для этого из средней активности нейронов скрытого слоя в пределах каждого из объектов формировалась обучающая выборка, на которой обучалась логистическая регрессия, после чего сравнивалась с метрикой accuracy логистической регрессии обученная на исходном сигнале. В результате логистическая регрессия по скрытому слою на тестовой выборке из 200 изображений показала значение от 0.74 до 0.79 по метрике accuracy, тогда как классическая регрессия показала значение 0.91 по метрике accuracy. Реализация вышеописанного эксперимента доступна на платформе GitHub [9].

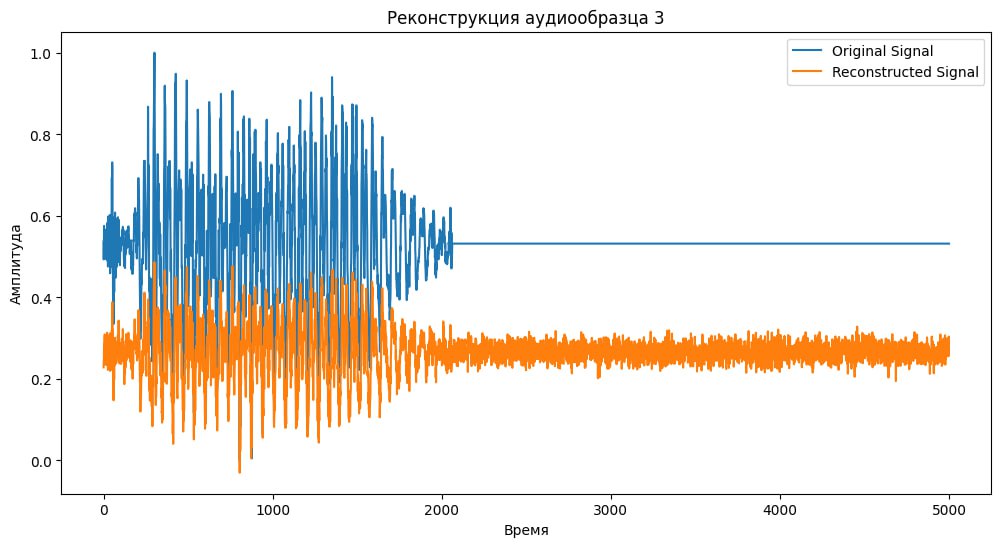


Рис.3. Исходный аудиоряд и его реконструкция (восстановление) для цифры 3

Литература

1. *Antsiperov V. E.* Generative Model of Autoencoders Self-Learning on Images Represented by Count Samples // Automation and Remote Control. 2022. Vol. 83. No. 12. Pp. 1959-1983. DOI: 10.1134/s00051179220120098.
2. *Antsiperov V. E., Kershner V. A*. Image Coding by Count Sample, Motivated by the Mechanisms of Light Perception in the Visual System // Communications in Computer and Information Science. 2022. Vol. 1534. Pp. 715-729. DOI: 10.1007/978-3-030-96040-7\_54.
3. *Abbott LF*. Lapicque's introduction of the integrate-and-fire model neuron (1907) // Brain Res Bull. 1999. V.50. №5-6. C 303-304. DOI: 10.1016/s0361-9230(99)00161-6.
4. *Friston K.* The free-energy principle: A unified brain theory? // Nature Reviews Neuroscience. 2010. Vol. 11, № 2.
5. *LeCun Y., Cortes C., Burgess C.J.C.* MNIST handwritten digit database // AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist. 2010. Vol. 7.
6. *Trappenberg T.P.* Fundamentals of Computational Neuroscience // Fundamentals of Computational Neuroscience. 2022.
7. *Abbott LF*. Lapicque's introduction of the integrate-and-fire model neuron (1907) // Brain Res Bull. 1999. V.50. №5-6. C 303-304. DOI: 10.1016/s0361-9230(99)00161-6.
8. *Zohar J.* Free Spoken Digit Dataset (FSDD) [Electronic resource]. 2020. URL: <https://github.com/Jakobovski/free-spoken-digit-dataset>.
9. Predictive Coding Multisensor v3 [Electronic resource]. URL: https://github.com/buligar/Predictive-coding/blob/main/Predictive%20coding%20Multisensor%20v3.ipynb.