

Применение импульсной нейронной сети на этапе предварительной обработки данных в условиях анализа поведения по видеоданным

С. Е. Ильин

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

s.ilin2086@gmail.com

Аннотация. Обсуждаются особенности использования импульсной нейронной сети в контексте решения задачи по предварительной обработке данных с целью анализа поведения. Конкретизируется специфика решения задачи в условиях работы с видеоданными. Предлагается вариант предварительной обработки указанных данных с применением импульсной нейронной сети. Производится его апробация и оценка полученных результатов.

Ключевые слова: искусственные нейронные сети; импульсные нейронные сети; предварительная обработка данных; анализ видеоданных; анализ поведения

I. ВВЕДЕНИЕ

В условиях современности принципиальное значение обретают вопросы эффективного применения и совершенствования методов обработки данных. Развитие технологий больших данных, высокопроизводительных вычислительных систем, искусственного интеллекта, цифровых двойников и др. создает условия для новых направлений прикладного применения обработки данных и масштабирования существующих задач. Активное развитие агентов на базе искусственного интеллекта и рост востребованности мультимодального машинного обучения позволяют делегировать обработку данных интеллектуальным системам, в то же время расширяя их возможности. Открываются перспективы освобождения человека от выполнения ряда рутинных операций в сфере повседневной работы с данными, а также повышения качества прогнозирования будущего.

Важную роль в обозначенных процессах играют видеоданные. В настоящее время они не только выступают одним из ключевых форматов передачи и хранения данных, но и актуализируют ряд комплексных исследовательских задач, многие из которых еще не получили окончательного решения. К числу таких задач относятся идентификация представленных в кадре объектов и субъектов; отслеживание объектов; распознавание и прогнозирование действий; классификация действий; анализ протяженных по времени событий и процессов; и мн. др. [9, р. 1–2]. Вариативность качества, средств и ракурсов съемки, возможность принятия во внимание освещения и глубины резкости изображаемого пространства, наличие или отсутствие звукового сопровождения, а также иные особенности видеоданных обуславливают рост разнообразия методов их анализа.

Среди проблем, рассматриваемых с опорой на видеоданные, специального внимания заслуживают проблемы в области исследования поведения. В

обобщенном виде поведение можно охарактеризовать при помощи следующей модели. Субъект поведения s в момент времени t реализует в отношении объекта поведения o действие a , которое приводит к результату f . Далее эту модель можно расширить путем включения в нее целей субъекта поведения; контекста и условий действий; данных о порядке следования действий, их внутренней организации; и др. [2, р. 5–6]. Особый интерес представляет вопрос о внутренней организации поведения. Его прояснение способствует более глубокому пониманию содержания поведения, помогает выявлять характерные для него паттерны, а также, в пределах, корректировать и прогнозировать их.

Задача настоящей работы связана с обозначенным вопросом. Она предполагает уточнение структуры анализируемого поведения путем последовательного использования метода разложения данных на эмпирические моды (англ. Empirical Mode Decomposition, EMD) и импульсной нейронной сети (англ. Spiking Neural Network, SNN) в условиях предварительной обработки видеоданных о поведении.

II. ОБЗОР ЛИТЕРАТУРЫ

В контексте поставленной задачи целесообразно выделить два блока исследований, заслуживающих специального обсуждения. С одной стороны, это исследования, посвященные методу EMD. С другой стороны, – исследования, посвященные SNN. Традиции обсуждения обеих тем насчитывают десятилетия продуктивной научной работы и оформляются, по меньшей мере, к рубежу XX и XXI в. Метод EMD принадлежит к семейству методов разложения сигналов на компоненты с целью получения дополнительной информации об анализируемых объектах. В отличие от некоторых других методов обозначенного семейства, он не предполагает использования заранее выбранного базиса, благодаря чему дает результаты, свободные от привносимых извне теоретических установок. Указание на эмпирический характер мод, вынесенное в название метода, призвано подчеркнуть данный момент. Эмпирические моды, или внутренние модовые функции исходного сигнала, выступают результатом разложения сигнала на компоненты с применением EMD [3].

В исследовательской литературе можно найти указания на преимущества и недостатки указанного метода. К его преимуществам относится тот факт, что он хорошо приспособлен к анализу нестационарных процессов, позволяет учитывать локальные особенности сигнала и восстанавливать его с определенной

точностью. Кроме того, он гарантирует, что при конечной длительности сигнала количество мод будет конечным [1, с. 58]. К потенциальным недостаткам метода EMD в его классической реализации относится риск усложнения математической интерпретации получаемых результатов, а также риск т. н. смешивания мод (англ. *mode mixing*), при котором в состав одной моды попадают компоненты, разнящиеся по своим характеристикам, либо когда один компонент ошибочно распределяется по нескольким модам. Для преодоления обозначенных недостатков было предложено несколько модифицированных вариантов метода, таких как множественное разложение на эмпирические моды (англ. *Ensemble Empirical Mode Decomposition, EEMD*) и комплементарное множественное разложение с адаптивным шумом (англ. *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, CEEMDAN*). Однако у них также были обнаружены некоторые потенциальные ограничения, – например, в случае EEMD, – существенное увеличение времени анализа сигнала в сравнении с исходным вариантом метода [5].

Второй блок исследований, требующих анализа в связи с решением поставленной задачи, составляют работы в области применения SNN. В их основе лежит идея об искусственных нейронных сетях, характеристики которых ближе к биологическим нейронным сетям, чем предполагают модели искусственных нейронных сетей на базе глубокого обучения. Предлагается дополнительно принимать во внимание особенности кодирования сигналов биологическими нейронами, специфику функционирования их мембран, роль нейромедиаторов в передаче сигналов и др. Созданные благодаря этому модели искусственных нейронов могут существенно отличаться по степени биофизической достоверности и сложности применения. Базовой моделью традиционно является т. н. интегративно-пороговая модель. В ней функция активации нейрона применяется лишь в тех случаях, когда превышено требуемое пороговое значение входного сигнала. Примером модели с высокой степенью биофизической достоверности может выступать модель Ходжкина–Хаксли [13, р. 6].

Изменения, привнесенные в SNN по сравнению с нейронными сетями на базе глубокого обучения, делают их перспективным инструментом для анализа временных рядов и предоставляют в распоряжение исследователей дополнительные варианты кодирования сигнала. В то же время, остается открытым вопрос выбора предпочтительных архитектур SNN в зависимости от специфики решаемой задачи, – а также некоторые другие методологические вопросы [13, р. 15–19].

Приведенные исследования позволяют предположить, что метод EMD может быть использован на этапе выделения признаков поведения в видеоданных, а SNN – на этапе фильтрации данных о поведении. Подобную комбинацию не удастся обнаружить в исследовательской литературе, поэтому задача оказывается перспективно ограничить подготовкой и тестированием базового варианта применения SNN в условиях предварительной обработки видеоданных о поведении средствами EMD.

III. МЕТОДОЛОГИЯ И РЕЗУЛЬТАТЫ

Работа с данными при решении поставленной задачи осуществлялась по следующему алгоритму.

1. Выбор видеоданных для анализа.
2. Импорт видеоданных.
3. Преобразование видеоданных в формат числового массива с информацией об изменении яркости белого цвета пикселей на протяжении видео.
4. Проверка содержимого массива на соответствие требованиям метода EMD. В случае соответствия – разложение данных на эмпирические моды. Иначе – возврат к п. 1.
5. Анализ результатов разложения данных на эмпирические моды. Выбор сценариев фильтрации сигнала эмпирических мод с помощью SNN.
6. Фильтрация данных с помощью SNN.
7. Анализ результатов фильтрации данных. Если задача не решена – возврат к п. 1 или п. 5.

В качестве видеоданных для анализа был привлечен фрагмент короткометражного фильма с участием Ч. Чаплина “Зарабатывая на жизнь” (англ. “*Making a Living*”) 1914 г. выпуска [6]. Пример кадра в составе указанного фрагмента представлен на рис. 1:



Рис. 1. Пример кадра в составе анализируемых данных

Выполнение действий, направленных на преобразование и обработку видеоданных, производилось с использованием языка Python, а также библиотек Matplotlib [4], NumPy [7], opencv-python [8] и др. Реализация метода EMD была взята из библиотеки PyEMD [10], [11]. Для работы с SNN использовался нейронный симулятор Brian 2 [12].

При обращении к EMD дополнительное внимание было решено уделить скорости анализа данных. В связи с этим в качестве метода EMD был выбран его первоначальный вариант в реализации PyEMD. Пример работы метода представлен на рис. 2:

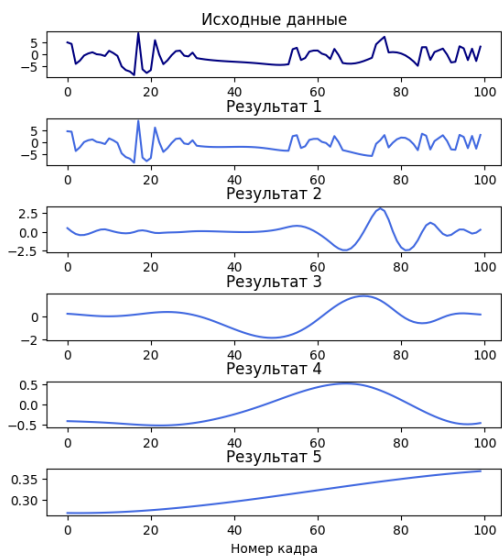


Рис. 2. Разложение данных об изменении яркости белого цвета выбранного пикселя на эмпирические моды

Изучение результатов разложения данных на эмпирические моды позволило получить следующие ключевые наблюдения. Во-первых, было обнаружено, что в случае анализа поведения моды с низкой частотой встречаемости событий аккумулируют информацию об общих поведенческих трендах, тогда как моды с высокой частотой – детали происходящего. Во-вторых, моды с низкой частотой встречаемости событий содержат больше информации об устойчивом контексте (фоне) поведения. В-третьих, вне обращения к модам с высокой частотой встречаемости событий данные, фиксируемые первой группой мод, оказываются в значительной степени абстрактными (лица могут быть размыты; фиксируются перемещения в пространстве, но не движения, которые их вызывают; и др.). Наконец, в-четвертых, данные со средней и высокой частотой встречаемости событий могут содержать не только детали происходящего, но и шум, который не имеет отношения к поведению. Наличие такого шума, а также риск смешивания мод в условиях применения исходного варианта метода EMD создают предпосылки для фильтрации полученных данных.

В общем случае фильтрация данных может осуществляться с применением различных методов и распространяться на различные аспекты содержания данных. В рамках настоящей работы роль шума была отведена неинформативным изменениям яркости белого цвета пикселей, зафиксированным в одной из эмпирических мод с высокой частотой встречаемости событий. Решение признать указанные изменения неинформативными было принято по той причине, что они в большей степени затрагивали фон, а не содержание поведения. Решение применить к ним фильтр обуславливалось тем, что при их наличии значимые аспекты поведения было сложнее фиксировать. Пример обсуждаемых данных представлен на рис. 3:

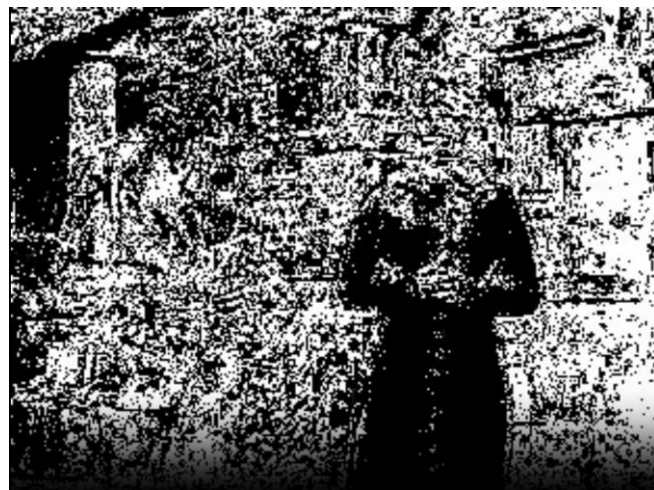


Рис. 3. Пример данных в составе эмпирических мод, допускающий фильтрацию

Роль фильтра была отведена варианту SNN, объединяющему в себе 3 группы интегративно-пороговых нейронов. Группы 0 и 1 получали на вход идентичный сигнал с характеристиками пикселей фильтруемой эмпирической моды и реагировали на него в соответствии с заложенными в них особенностями (мембранный потенциал, пороговое значение, рефрактерный период и др.). Результатом их реакции выступали импульсы (спайки), которые передавались в нейронную группу 2. Группа 2 объединяла сигналы от групп 0 и 1. В данном случае реакция на стимулы уточнялась за счет использования адаптивного порога, который выступал дополнительным средством фильтрации данных эмпирической моды. Содержание сигнала после фильтрации определялось наличием реакции нейронов группы 2 на входные стимулы. В отсутствие реакции, изменения значений яркости белого цвета в эмпирической моде игнорировались. Детали процесса фильтрации данных отражены на рис. 4, который показывает, как сигнал изменяется при переходе между нейронами из разных групп. Изменения, которые проходят фильтрацию, отмечены на нижнем графике вертикальными пунктирными линиями. Из 100 рассматриваемых изменений пикселя, которые фиксируются нейронами группы 0 и 1, фильтрацию проходит только треть, а именно, 34 изменения.

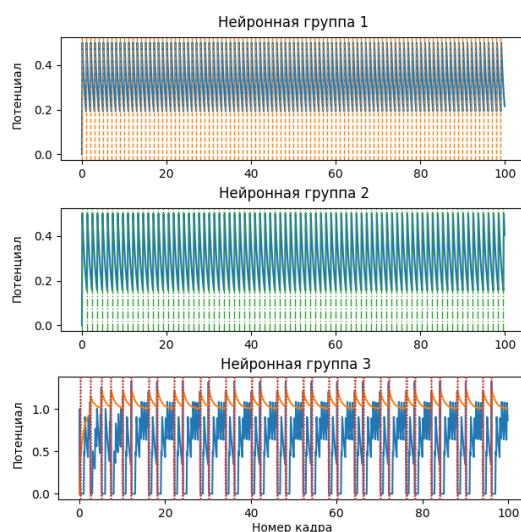


Рис. 4. Пример фильтрации данных в составе эмпирической моды

Отталкиваясь от полученного результата, можно сделать вывод о применимости SNN в качестве фильтра для отсеивания неинформативных изменений эмпирических мод в составе видеоданных о поведении. Полученный вывод актуализирует дополнительные вопросы касательно работы с соответствующим способом фильтрации и располагает к дальнейшему углублению темы.

IV. ЗАКЛЮЧЕНИЕ

На фоне стремительного развития средств обработки информации и увеличения объемов создаваемых человечеством видеоданных важную роль в современном мире обретает анализ этих данных. К числу явлений, которые находят отражение в видеоданных, следует отнести поведение. Будучи иерарически организованным, поведение фиксируется в совокупностях событий и процессов, которые могут существенно различаться по продолжительности и частоте встречаемости. Для учета соответствующих фактов в условиях предварительной обработки данных о поведении может быть использовано разложение указанных данных на эмпирические моды. Полученные в результате компоненты данных о поведении в дальнейшем могут фильтроваться с помощью импульсных нейронных сетей. Будучи тонким инструментом работы с данными, такие сети предоставляют в распоряжение исследователей множество вариантов совершенствования получаемых с их помощью наблюдений. Базовые варианты обращения к импульсным нейронным сетям в условиях предварительной обработки данных о поведении могут быть дополнены и расширены путем усложнения используемых моделей нейронов, привлечения эффективных стратегий организации нейронных связей, а также последовательного привлечения механизмов памяти и обучения в импульсных нейронных сетях.

СПИСОК ЛИТЕРАТУРЫ

- [1] Алимуратов А. К. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов // Труды МФТИ. 2015. Т. 7. №3. С. 56-68.
- [2] Cao L. In-Depth Behavior Understanding and Use: The Behavior Informatics Approach [Electronic resource] // ArXiv. 2020. P. 1-29. DOI: <https://doi.org/10.48550/arXiv.2007.15516> (accessed: 14.04.2025).
- [3] Huang N. E., Shen Zh., Long S. R., Wu M. C., Shih H. H., Zheng Q., Yen N.-Ch., Tung Ch. Ch., Liu H. H. The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis // Proceedings of the Royal Society of London. Series A: Mathematical, physical and engineering sciences. 1998. P. 903-995.
- [4] Hunter J. D. Matplotlib: A 2D Graphics Environment // Computing in Science & Engineering. 2007. Vol.9, No.3. P. 90-95.
- [5] Liu T., Luo Zh., Huang J., Yan Sh. A Comparative Study of Four Kinds of Adaptive Decomposition Algorithms and Their Applications // Sensors. 2018. Vol.18(7). P. 1-51. DOI: <https://doi.org/10.3390/s18072120> (accessed: 14.04.2025).
- [6] Making a Living [Electronic resource]. L.A.: Keystone Film Company, 1914. URL: https://ia802201.us.archive.org/13/items/CC_1914_02_02_MakingALiving/CC_1914_02_02_MakingALiving_512kb.mp4 (accessed: 10.04.2025).
- [7] NumPy: The Fundamental Package for Scientific Computing with Python [Electronic resource]. URL: <https://numpy.org/> (accessed: 10.04.2025).
- [8] opencv-python [Electronic resource]. URL: <https://pypi.org/project/opencv-python/> (accessed: 10.04.2025).
- [9] Pattichis M. S., Jatla V., Cerna A. E. U. A Review of Machine Learning Methods Applied to Video Analysis Systems [Electronic resource] // ArXiv. 2023. Pp. 1-5. DOI: <https://doi.org/10.48550/arXiv.2312.05352> (accessed: 16.04.2025).
- [10] Pele O., Werman M. A Linear Time Histogram Metric for Improved SIFT Matching. Computer Vision – 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008. Marseille: Springer, 2008. P. 495-508.
- [11] Pele O., Werman M. Fast and Robust Earth Mover's Distances // IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009. Kyoto: IEEE, 2009. P. 460-467.
- [12] Stimberg M., Brette R., Goodman D. F. M. Brian 2, an Intuitive and Efficient Neural Simulator // eLife 8:e47314. 2019. P. 1-41. DOI: <https://doi.org/10.7554/eLife.47314> (accessed: 15.04.2025).
- [13] Yamazaki K., Vo-Ho V.-Kh., Bulsara D., Le N. Spiking Neural Networks and Their Applications: A Review // Brain Sciences. 2022, No.12(863). P. 1-30. DOI: [10.3390/brainsci12070863](https://doi.org/10.3390/brainsci12070863) (accessed: 15.04.2025).