

*Дуля И.С.,
студент магистратуры
2 курс, Институт прикладной математики
и компьютерных наук
Томский государственный университет
Россия, г. Томск*

РАЗРАБОТКА МОДЕЛИ РАСПОЗНАВАНИЯ АКТИВНОСТИ ЧЕЛОВЕКА НА ОСНОВЕ МЕХАНИЗМОВ ВНИМАНИЯ

***Аннотация:** Статья посвящена разработке модели распознавания человеческой активности на предприятиях на основе сигналов носимых устройств. Было проведено сравнительное исследование разработанной модели и наиболее эффективных моделей глубокого обучения, применяемых в HAR. Было показано, что предложенная модель имеет более высокую точность по сравнению с современными гибридными решениями.*

***Ключевые слова:** распознавание активности человека, сигналы носимых устройств, глубокое обучение, нейронные сети, механизмы внимания.*

***Annotation:** The article is devoted to developing a model for human activity recognition in enterprises based on signals from wearable devices. A comparative study of the developed model and the most effective deep learning models used in HAR was shown that the proposed model has higher accuracy compared to modern hybrid solutions.*

***Key words:** human activity recognition, wearable sensors, deep learning, neural networks, attention mechanisms.*

1 Модель на основе механизмов внимания

Преимущество моделей, основанных на механизмах внимания, состоит в том, что они не используют рекуррентные связи. Соответственно, процессы обучения и прогнозирования не имеют итерационный характер, что открывает возможности повышения производительности за счёт параллельных вычислений. Предложенная модель на основе механизма внимания использует внимание трёх типов: сенсорное внимание, self-attention и глобальное временное внимание. Эти механизмы внимания позволяют получить признаковое представление, по которому осуществляется окончательная классификация с помощью полносвязного блока (рисунок 1).

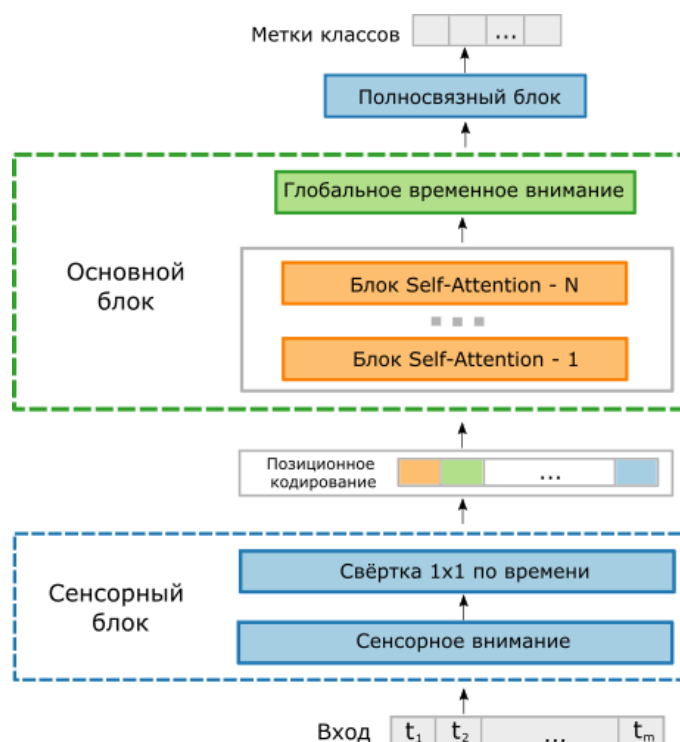


Рисунок 1. Предложенная модель на основе механизмов внимания

Входными данными для модели являются сегменты временного ряда, полученные с помощью метода скользящего окна. На первом этапе к входным данным применяется слой сенсорного внимания, оценивающий коэффициенты внимания для разных сенсорных модальностей и возвращающий взвешенное представление значений датчиков в соответствии с их оценкой внимания. Таким образом, оценка внимания определяет вклад каждой из сенсорной модальности в представление признаков, используемое

последующими слоями. После этого происходит преобразование взвешенных значений датчиков в векторы размера d с помощью применения свертки 1×1 . Далее кодируется позиционная информация в последовательности с помощью процедуры positional encoding. Закодированная информация передается n блокам self-attention. Представление, сгенерированное блоками self-attention, передается слою глобального временного внимания. Этот уровень определяет оценки внимания во временном измерении, чтобы создать окончательное признаковое представление, которое используется полностью связными блоками для итоговой классификации.

2 Сенсорное внимание

Для оценки вклада различных сенсорных модальностей используется слой сенсорного внимания. Необходимость данного слоя обусловлена тем, что для классификации отдельных видов активностей, нужны данные только определенных сенсорных модальностей. Например, для распознавания действия, выполняемого только руками, не требуется информация с датчиков, расположенных на ноге субъекта. Слои сенсорного внимания изучают такие взаимосвязи, используя двумерную свертку по времени и значениям разных датчиков.

На первом этапе процедуры вход преобразуется в одноканальное изображение. Затем к входу применяются k свёрточных фильтров с отступами, которые выводят изображение с k каналами. Затем оно преобразуется обратно в один канал путем применения временной свертки 1×1 . Оценка внимания для отдельных сенсоров определяется из следующего уравнения:

$$S_k^{(t_i)} = \frac{\exp(q_k^{(t_i)})}{\sum_k \exp(q_k^{(t_i)})}, \quad (1)$$

где k – индекс сенсора, q_k – значения, полученные после применения свертки 1×1 , t_i – индекс временного шага. На рисунке 2 представлена схема работы слоя сенсорного внимания. Из схемы видно, что после слоя сенсорного внимания применяется слой свёртки 1×1 по времени. Свёртка 1×1 по времени

применяется для того, чтобы смоделировать межсенсорное взаимодействие и увеличить размерность элемента последовательности. Это позволяет повысить качество обработки последующими слоями self-attention.

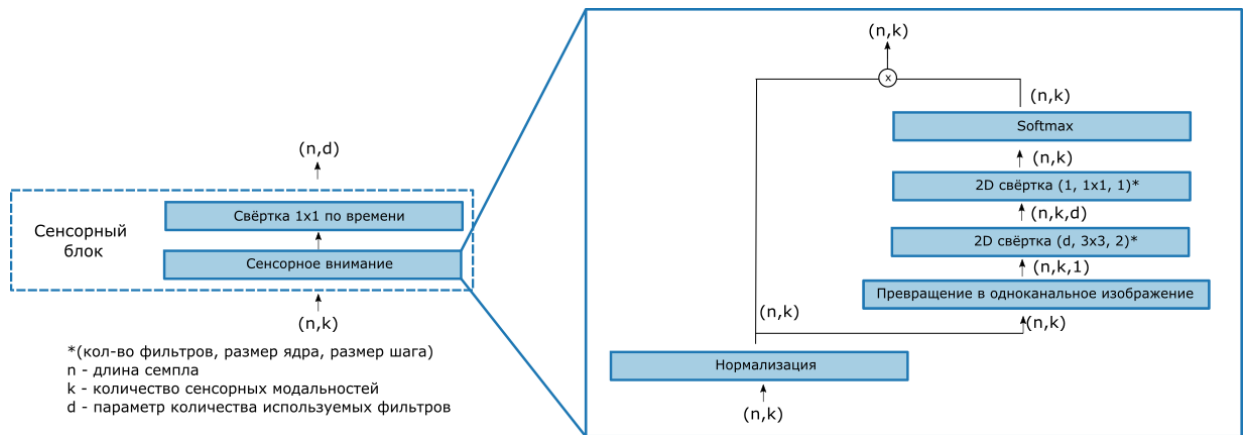


Рисунок 2. Схема работы слоя сенсорного внимания

В дополнение к получению оценок внимания, позволяющих определить вклад каждого из датчиков, этот механизм позволяет получить карты признаков, делая модель более интерпретируемой.

3 Глобальное временное внимание

Глобальное временное внимание использует признаковое представление, созданное блоками self-attention, и оценивает коэффициенты внимания для каждого временного шага. Оценки коэффициентов внимания, полученные в (3), используются для создания средневзвешенного представления всех временных шагов в окне. Далее на основе полученного вектора признаков осуществляется окончательная классификация с помощью полносвязного блока.

$$g_t = \tanh(W_g \cdot s_t + b_g), \quad (2)$$

$$\alpha_t = \frac{\exp(g_t^T \cdot g_s)}{\sum_k \exp(g_k^T \cdot g_s)}, \quad (3)$$

$$c_i = \sum_t \alpha_t \cdot s_{ti}. \quad (4)$$

$W_g \in R^{d \times d}$, $b_g \in R^{d \times 1}$, $g_s \in R^{d \times 1}$ – параметры, настраиваемые в ходе обучения, $s_t \in R^{d \times 1}$ – выходной вектор предыдущего слоя ($t = 1, \dots, n$), α_t – оценки внимания, c_i – выходные значения слоя глобального временного

внимания ($i = 1, \dots, d$). Параметр g_s в (3) фиксирует временной контекст при обучении параметров внимания. Выходные значения c_i определяются как взвешенные суммы в соответствии с оцененными коэффициентами внимания (4).

4 Результаты

В таблице 1 продемонстрировано сравнение точности современных гибридных моделей и предложенной модели на основе механизма внимания.

Таблица 1.

Макро F1-score для тестовых субъектов

Набор данных	ConvAE	CNN-LSTM	DeepConvLSTM	Предложенная модель
PAMAP2	0,870	0,896	0,923	0,968
Opportunity	0,628	0,660	0,644	0,673
USC-HAD	0,468	0,439	0,475	0,551
Skoda	0,871	0,864	0,927	0,970

Предложенная модель обеспечивает значительное улучшение показателей по сравнению с рассмотренными моделями на эталонных наборах данных. Предлагаемая модель может фиксировать пространственно-временные характеристики данных датчиков более эффективно, чем другие гибридные модели, за счёт чего и объясняется более высокое качество классификации.

Использованные источники:

1. Дуля И.С. Распознавание активности человека на предприятиях на основе данных носимых устройств [Электронный ресурс] // GitHub Inc. – Электрон. дан. – 2022. – URL: <https://github.com/dulyaivan/HAR> (дата обращения: 15.05.2022).

2. Attention Is All You Need [Electronic resource] / A. Vaswani [et al.] // ArXiv. – 2017. – Electronic data. – URL: <https://arxiv.org/abs/1706.03762> (access date: 10.05.2022).
3. DanHAR: dual attention network for multimodal human activity recognition using wearable sensors / W. Gao [et al.] // Applied soft computing. – 2021. – № 111. – P. 107728–107753.
4. Multi-modality sensor data classification with selective attention / X. Zhang [et al.] // In proceedings of the 27th International joint conference on artificial intelligence (13 – 19 July 2018). – Stockholm, Sweden, 2018. – P. 3111–3117.
5. Rubwurm M. Self-attention for raw optical satellite time series classification [Electronic resource] / M. Rubwurm, M.Korner // ArXiv. – 2020. – Electronic data. – URL: <https://arxiv.org/pdf/1910.10536.pdf> (access date: 08.05.2022).