

# Применение механизма внимания для прогнозирования спроса на газ

А. К. Петрова, С. Е. Абрамкин

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

E-mail ak72p@yandex.ru

**Аннотация.** В статье рассматривается задача прогнозирования спроса на газ с применением гибридной нейросетевой архитектуры. Отмечены ограничения классических рекуррентных сетей (RNN/LSTM) при обработке временных рядов, в частности, последовательный характер вычислений и деградация долгосрочных зависимостей. В качестве решения предложен подход, сочетающий LSTM-слой с механизмом само-внимания (self-attention), позволяющий модели автоматически выделять наиболее релевантные временные интервалы исторических данных при формировании прогноза.

**Ключевые слова:** прогнозирование расхода газа, механизм внимания, LSTM, временные ряды, нейронные сети, трансформеры, машинное обучение

## I. ВВЕДЕНИЕ

Точное и своевременное прогнозирование спроса на газ является важной задачей для обеспечения надёжности и экономической эффективности функционирования газораспределительных систем.

В условиях растущей волатильности погодных условий, изменения режимов потребления и интеграции возобновляемых источников энергии традиционные методы прогнозирования (статистические модели, авторегрессии) зачастую не обеспечивают достаточной точности из-за нелинейности и многомерности зависимостей в реальных данных [1–5].

Современные подходы на основе машинного обучения, в частности рекуррентные нейронные сети с механизмами внимания, позволяют автоматически выявлять сложные временные паттерны и учитывать влияние разнородных факторов (температура, давление, сезонность) без ручного формирования признаков.

### A. Ограничения RNN и LSTM

Несмотря на успехи LSTM в обработке временных рядов, архитектура имеет ограничения [6–10]:

- Последовательная обработка. LSTM обрабатывает временной ряд строго слева направо, что затрудняет параллелизм и замедляет обучение на больших наборах данных.
- Проблема с длинными зависимостями. Несмотря на то, что LSTM лучше, чем базовая RNN, долгосрочные зависимости всё ещё могут быть забыты.
- Сжатие информации. Весь контекст временного ряда сжимается в единственный вектор перед тем, как сделать предсказание, что может привести к потере информации.

### B. Механизм внимания (Attention Mechanism)

Вычисление оценок внимания (attention scores). Для каждого элемента вычисляется оценка, показывающая, насколько релевантна эта часть входа для текущего элемента.

Применение softmax. Оценки преобразуются в вероятности через функцию softmax, так что они суммируются к 1.

Взвешенная сумма. Входные элементы взвешиваются по вероятностям внимания и суммируются для получения контекстного вектора.

Математически, механизм внимания можно описать следующим образом (1):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q_i K_j}{\sqrt{d_k}}\right) \cdot V \quad (1)$$

где  $Q$  (*Query*) – запросы, текущее слово/позиция, для которой вычисляется внимание, то есть векторы, которые «спрашивают»: чему сейчас нужно уделить внимание;  $K$  (*Key*) – ключи, все позиции в последовательности, то есть векторы, с которыми сравнивается запрос: «на что можно обратить внимание»;  $V$  (*Value*) – векторы содержимого, которые содержат фактическую информацию, которую необходимо извлечь;  $d_k$  размерность вектора ключа (и обычно запроса).

Результат – контекстуально обогащенное представление запроса, где  $Q, K, V$  – это три различных представления входных данных.

## II. АЛГОРИТМ ВНИМАНИЯ НА ПРИМЕРЕ ПРОГНОЗА РАСХОДА ГАЗА

В исследовании использован многомерный временной ряд, содержащий данные о функционировании газораспределительной системы. Набор данных включает четыре признака, регистрируемых с часовой дискретностью: давление газа, температура газа, расход газа и температура окружающего воздуха.

Перед обучением данные подвергнуты стандартной предобработке, выполнена нормализация признаков методом StandardScaler для приведения к нулевому среднему и единичной дисперсии. Исходная выборка разделена на обучающую, валидационную и тестовую части в пропорции 70:15:15 с сохранением временного порядка (без перемешивания), что обеспечивает корректную оценку обобщающей способности модели на новых данных. Общий объём выборки составил около 8000 наблюдений за 1 год, что обеспечивает

достаточную репрезентативность для обучения нейросетевой архитектуры.

#### А. Механизм внимания

Анализ сходства запросов и ключей:  $Q_i, K_j$ . Скалярное произведение  $Q_i \cdot K_j$  измеряет, насколько «похож» запрос  $Q_i$  на ключ  $K_j$ . Чем больше значение, тем сильнее этот ключ (а значит и соответствующее ему значение  $V_j$ ) релевантен текущему запросу.

Масштабирование: деление на  $\sqrt{d_k}$ . При большой размерности скалярные произведения становятся по модулю большими, softmax даёт очень резкие распределения и плохие градиенты. Деление на  $\sqrt{d_k}$  нормализует значения, чтобы softmax вёл себя стабильнее.

Применение softmax для определения весов внимания. К каждому запросу применяется softmax по всем ключам:

$$a_{ij} = \text{softmax}\left(\frac{Q_i \cdot K_j}{\sqrt{d_k}}\right)$$

Это даёт неотрицательные веса, которые суммируются в 1 по  $j$ .

$a_{ij}$  можно интерпретировать как «вероятность» или долю внимания, которую запрос  $i$  уделяет элементу  $j$ .

Взвешенная сумма значений: умножение на  $V$ . После softmax получаем матрицу весов внимания  $A$  и умножаем её на  $V$ :

$$\text{Attention}(Q, K, V) = A \cdot V$$

где  $A = \text{softmax}(QK^T/\sqrt{d_k})$ .

$$\text{Для одного запроса: } \text{output}_i = \sum_j a_{ij} V_j$$

То есть выход – это контекстный вектор: смесь значений  $V_j$ , усреднённых с весами внимания.

Работу механизма внимания иллюстрирует рис. 1. На графике по оси  $X$  отложены дни (от день 1 к день 30), по оси  $Y$  – значения спроса на газ. Синяя линия показывает исторический временной ряд. Для конкретного дня (например, день 30), показаны веса внимания как вертикальные полосы, где высота полосы показывает, насколько модель обращает внимание на соответствующий день при прогнозировании спроса на день 30. Обычно высокие веса приходятся на дни с похожими температурными условиями и меньшие веса на дни с отличающимися условиями [11–15].

График показывает, что для прогноза спроса на день 30 модель обращает наибольшее внимание на дни 17–18, и 27–30, которые имели схожие условия (низкие температуры). Дни в начале периода имеют низкие веса внимания, так как их условия менее релевантны.

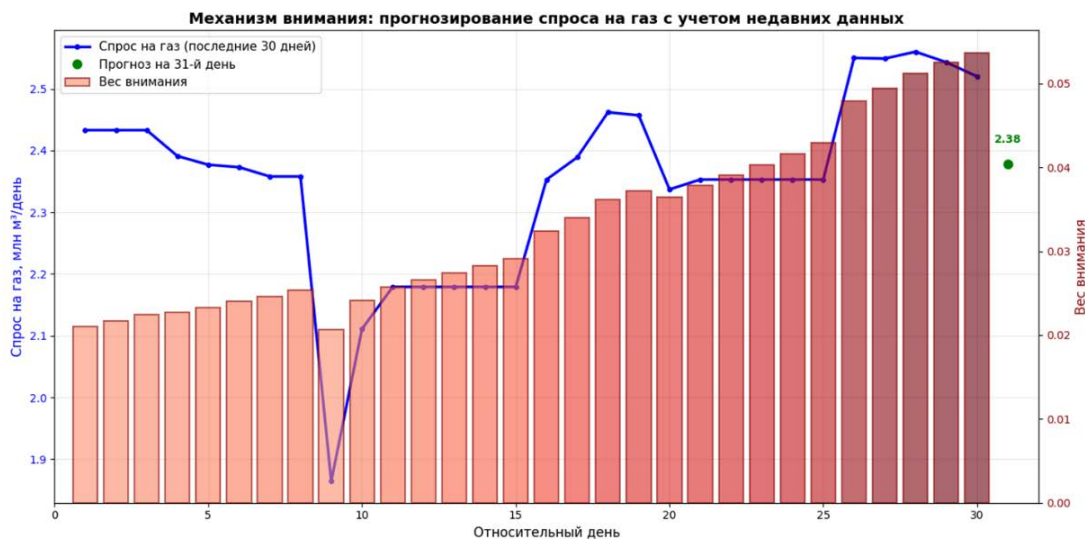


Рис. 1. Механизм внимания

#### В. Модель нейронной сети с механизмом внимания

Определим простую модель нейронной сети с механизмом внимания и обучим ее на подготовленных данных. Для демонстрации мы можем использовать комбинацию LSTM-слоя (Long Short-Term Memory) и слоя внимания, что является распространенным подходом для временных рядов. Краткое описание структуры модели:

- LSTM слой: для извлечения временных зависимостей из входных последовательностей.
- Attention слой: чтобы модель могла фокусироваться на наиболее релевантных частях выходных данных LSTM при предсказании.

- Dense слой: для выдачи конечного предсказания.

Среда реализации: Язык программирования: Python 3.x (стандартная версия Colab), фреймворк для глубокого обучения: TensorFlow 2.x (с Keras API), ключевые библиотеки: pandas (для работы с данными), numpy (для численных операций), scikit-learn (для нормализации данных, PCA и метрик), matplotlib, seaborn (для визуализации)? tensorflow.keras (для построения и обучения нейронных сетей)

Гиперпараметры: оптимизатор Adam, Learning Rate 0.001, количество эпох: 50, размер батча: 32, функция потерь: Mean Squared Error (MSE).

Как работает модель:

Входной слой (Input Layer): подаются наши подготовленные данные. Это последовательность данных за look\_back (24) предыдущих часа, где для каждого часа есть features (4) признака: давление, температура газа, расход газа, температура воздуха. Форма данных: (количество\_примеров, look\_back, features).

Слой LSTM (LSTM Layer): входная последовательность проходит через слой LSTM. LSTM - это тип рекуррентной нейронной сети, которая хорошо справляется с обработкой последовательностей, запоминая важную информацию из прошлого и забывая менее важную. LSTM возвращает не только последнее скрытое состояние, но и скрытое состояние для каждого временного шага входной последовательности. LSTM извлекает временные зависимости и паттерны из входных данных. Выход LSTM на этом этапе - это набор скрытых состояний, каждое из которых представляет собой контекст для соответствующего временного шага, учитывающий все предыдущие.

Слой Внимания (Attention Layer): это ключевой компонент, который позволяет модели «взвешивать» важность различных моментов времени во входной

последовательности. В данном случае используется self-attention (само-внимание), где каждый элемент выходной последовательности LSTM взаимодействует со всеми остальными элементами этой же последовательности.

Плотный слой (Dense Layer, 32 нейрона, relu активация): к каждому временному шагу выхода слоя внимания применяется обычный полносвязный слой. Он преобразует взвешенные контекстные векторы в более абстрактное представление. Это добавляет нелинейность и позволяет модели извлекать более сложные комбинации признаков.

Слой глобального усредняющего пулинга (GlobalAveragePooling1D) берет выход предыдущего Dense слоя и усредняет его по временному измерению. То есть, из последовательности векторов он делает один вектор, который является средним значением всех векторов последовательности.

Выходной слой (Output Layer, 1 нейрон): принимает агрегированный вектор от GlobalAveragePooling1D и выдает окончательное предсказание расхода газа.

Рис. 2 иллюстрирует сравнение фактического и предсказанного расхода на тестовой выборке.

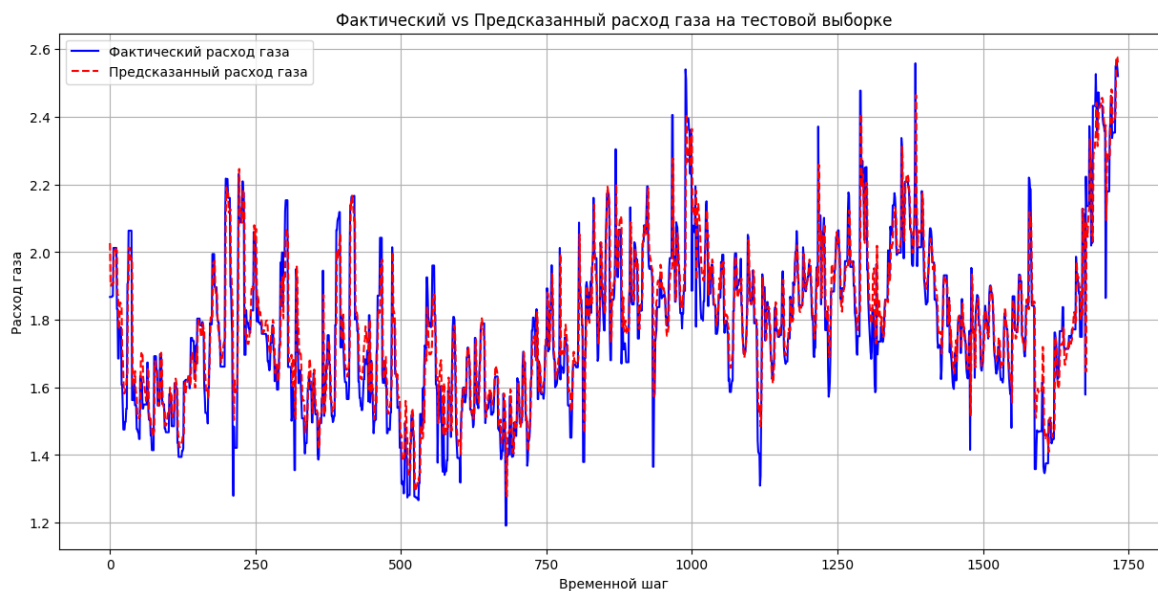


Рис. 2. Сравнение фактического и предсказанного расхода на тестовой выборке

Для сравнения в табл. 1 приведены метрики качества других моделей: ARIMA (AutoRegressive Integrated Moving Average) – традиционная статистическая модель для временных рядов, и базовая рекуррентная нейронная сеть LSTM, способная улавливать долгосрочные зависимости во временных рядах.

ТАБЛИЦА I. СРАВНЕНИЕ МЕТРИК КАЧЕСТВА МОДЕЛЕЙ

Model	MAE	RMSE	R <sup>2</sup>
LSTM	0.1656	0.2289	0.9850
LSTM with Attention	0.1508	0.2078	0.9877
ARIMA	0.4908	0.6121	0.8934

Для оценки каждой модели использовались следующие метрики:

MAE (Mean Absolute Error) – средняя абсолютная ошибка, RMSE (Root Mean Squared Error), корень из среднеквадратической ошибки, и R<sup>2</sup> (Coefficient of

Determination), коэффициент детерминации, показывающий долю дисперсии зависимой переменной, объясняемую моделью. Видно, что гибридная модель лучше других предсказывает расход газа, что подтверждают метрики качества.

### III. ДОПОЛНИТЕЛЬНЫЕ АСПЕКТЫ И СОВРЕМЕННЫЕ ПОДХОДЫ

Для прогнозирования временных рядов с механизмом внимания могут быть применены нейронные сети с архитектурой трансформера, полностью основанной на механизме внимания, без рекуррентных соединений. Она впервые была предложена в статье "Attention is All You Need" (Vaswani et al., 2017) [16-18] и стала основой для больших языковых моделей. По аналогии с рекуррентными нейронными сетями (PHC / RNN) трансформеры предназначены для обработки последовательностей, таких как текст на естественном

языке, и решения таких задач как машинный перевод и автоматическое реферирование. В отличие от РНС, трансформеры не требуют обработки последовательностей по порядку. Например, если входные данные – это текст, то трансформеру не требуется обрабатывать конец текста после обработки его начала. Благодаря этому трансформеры распараллеливаются легче чем РНС и могут быть быстрее обучены.

Преимущество трансформеров заключается в отсутствии рекуррентных блоков, поэтому они требуют меньше времени на обучение по сравнению даже с улучшенными архитектурами РНС / RNN, такими как «долгая краткосрочная память» (LSTM).

#### IV. ЗАКЛЮЧЕНИЕ

Разработана и обучена гибридная модель краткосрочного прогнозирования спроса на газ, объединяющая рекуррентный слой LSTM и механизм само-внимания. Attention-компонент обеспечивает интерпретируемость решений, автоматически выделяя исторические интервалы с условиями, наиболее релевантными для текущего прогноза.

Вместе с тем предложенный подход имеет ряд ограничений. Модель чувствительна к аномальным выбросам (внезапные технологические отключения, экстремальные погодные явления), требует периодического дообучения на свежих данных для предотвращения деградации качества прогнозов («дрейфа» модели), а её точность напрямую зависит от полноты, частоты и достоверности входных метеорологических и телеметрических измерений.

В перспективе планируется тестирование а архитектур, полностью основанных на механизме внимания (Transformer, Informer, TFT), расширение пула входных признаков (календарные факторы, праздничные дни, макроэкономические индикаторы) и промышленная апробация алгоритма на реальных телеметрических потоках газотранспортных предприятий.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Воронцов К. В. Машинное обучение: курс лекций. М.: МГУ, 2024. 412 с.
- [2] Семен Козлов. Transformer - новая архитектура нейросетей для работы с последовательностями. Хабр (30 октября 2017). <https://habr.com/ru/articles/341240/> Дата обращения 3 ноября 2020.
- [3] Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2018. 480 с.
- [4] LeCun Y., Bengio Y., Hinton G. "Deep learning" (Nature, 2015)
- [5] Schmidhuber J. "Deep learning in neural networks: An overview" *Neural Networks*, Volume 61, January 2015, Pages 85-117 <https://doi.org/10.1016/j.neunet.2014.09.003>
- [6] Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. 1997. Vol. 9. Iss. 8. Pp. 1735-1780. DOI:10.1162/neco.1997.9.8.1735. PMID 9377276.
- [7] Gers F.A. Learning to Forget: Continual Prediction with LSTM // *Neural Computation*. 2000. Vol. 12. Iss. 10. Pp. 2451-2471. DOI: 10.1162/089976600300015015.
- [8] Hochreiter S. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies / S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber // *A Field Guide to Dynamical Recurrent Neural Networks*. 2001. Pp. 464–467.
- [9] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions // *International Journal of Forecasting*, 37(1), 388-427.
- [10] Gagliano L. Bispectrum and Recurrent Neural Networks: Improved Classification of Intercital and Preictal States / L. Gagliano, E.B. Assi, D.K. Nguyen, M. Sawan // *Scientific Reports*. 2019. Vol. 9: 15649. Режим доступа: <https://doi.org/10.1038/s41598-019-52152-2>.
- [11] Hewamalage H., Bergmeir C., & Bandara K. (2021). Recurrent Neural Networks for Time Series Forecasting: Status and Future Directions // *International Journal of Forecasting*, 37(1), 388-427.
- [12] Cover of Machine Learning for Brain Disorders, Machine Learning for Brain Disorders [Internet]. Colliot O, editor. New York, NY: Humana; 2023. <https://www.ncbi.nlm.nih.gov/books/NBK597507/figure/ch1.Fig3/?report=objectonly>
- [13] Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. 736 с.
- [14] Хайкин С. Нейронные сети. Полный курс /Пер. с англ. М.: Вильямс, 2006. 1104 с.
- [15] Vapnik V.N. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer Verlag, 2000. P. 334.
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention is All you Need (англ.) // *Advances in Neural Information Processing Systems* 30 / I. Guyon, U. v. Luxburg, S. Bengio, H. Wallach, R. Fergus, S.V.N. Vishwanathan, R. Garnett. 2017. 15 p. - arXiv:1706.03762
- [17] Назаров А.В., Лоскутов А.А. Нейросетевые алгоритмы прогнозирования и оптимизации систем. СПб.: Наука и Техника, 2003. 384 с.
- [18] Holland J.H. Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Cambridge: 1st MIT Press ed., 1992. 232 p.