

Классификация результатов измерений технологического процесса учета расхода газа при помощи алгоритмов машинного обучения

Classification of the gas flow metering results using machine learning algorithms



С. Е. Абрамкин,
к. т. н., доцент, кафедра автоматизации
и процессов управления
✉ seabramkin@etu.ru

S. E. Abramkin,
PhD in engineering sciences,
associate professor, department
of automation and control processes



А. К. Петрова,
старший преподаватель,
кафедра инновационного менеджмента
✉ ak72p@yandex.ru

A. K. Petrova,
dsenior teacher, innovation management
department

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
St. Petersburg state electrotechnical university «LETI»

В статье рассматривается задача классификации результатов технологического процесса учета расхода газа на предмет наличия в них отклонений. Классификация осуществлена при помощи LSTM-нейросети, обученной на размеченных данных, осуществлено сравнение с другими алгоритмами машинного обучения, а также комбинирование с регрессионной LSTM-нейросетью. Применение данного алгоритма, с целью предсказательного обслуживания, заблаговременного выявления и предотвращения причин некоторых видов отклонений, направлено на повышение эффективности управления газотранспортной системой.

The article discusses the problem of classifying the results of the technological process of gas flow metering for the presence of deviations in them. The classification was carried out using an LSTM neural network trained on labeled data, compared with other machine learning algorithms, and combined with a regression LSTM neural network. The use of this algorithm, for the purpose of predictive maintenance, early identification and prevention of the causes of certain types of deviations, is aimed at increasing the efficiency of gas transportation system management.

Ключевые слова: газотранспортная сеть, баланс газа, классификация отклонений, алгоритмы машинного обучения, LSTM нейронные сети.

Keywords: gas transportation network, gas balance, classification, machine learning algorithms, LSTM neural networks.

Достижение равенства входящих и исходящих потоков газа в газотранспортной сети (ГТС) является актуальной задачей (рис. 1) [1]. При наличии такого равенства в системе достигается баланс, при неравенстве — присутствует небаланс [2, 3]. Для решения задачи управления балансом в ПАО «Газпром» используются современные методы, в том числе алгоритмы машинного обучения [4, 5], наиболее эффективные для прогнозирования и классификации многомерных временных зависимостей нелинейных величин [6]. Например, такие алгоритмы могут быть применены для классификации данных по расходу газа по признаку наличия или отсутствия отклонений, поскольку такие отклонения могут привести к нарушению баланса в ГТС.

Рассмотрим применение алгоритмов машинного обучения для классификации результатов технологи-

ческого процесса учета расхода газа в ГТС на предмет наличия отклонений. В качестве такого алгоритма была выбрана LSTM-нейросеть, поскольку данные входного массива являются многомерными и нелинейными. Обучение осуществлялось на статистических данных Калининградского участка Север-Западной ГТС за 2015 г.

Применение LSTM-нейросети для классификации данных

Для применения алгоритмов необходимо разметить данные результатов измерения. На рис. 2 приведены размеченные экспертами данные технологического процесса учета расхода газа на одной из ГРС — Краснознаменской, использовались сравнительные шкалы, справа находится шкала для давления, единица

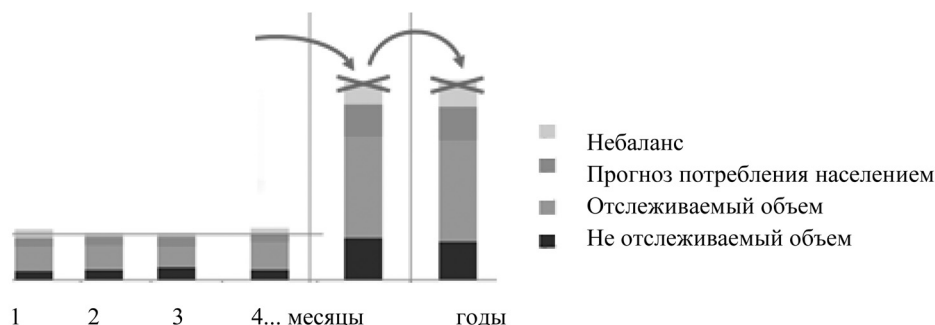


Рис. 1. Фрагмент доклада руководителя службы Межрегионгаз

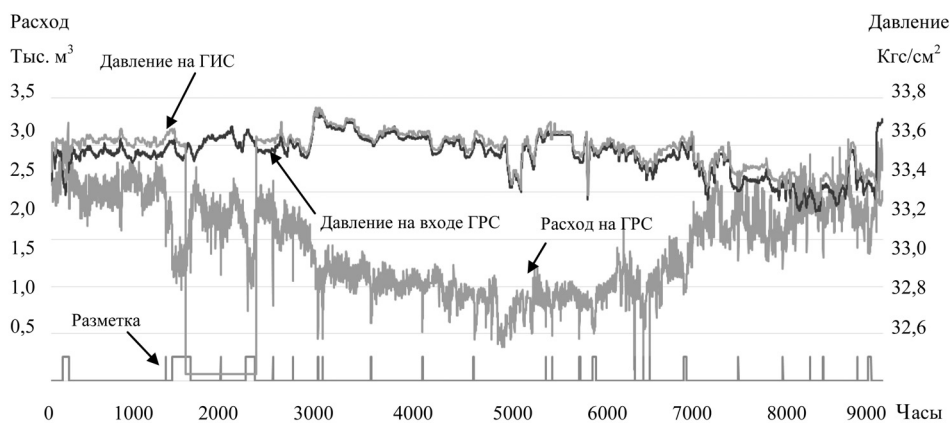


Рис. 2. Разметка данных для ГРС Краснознаменская

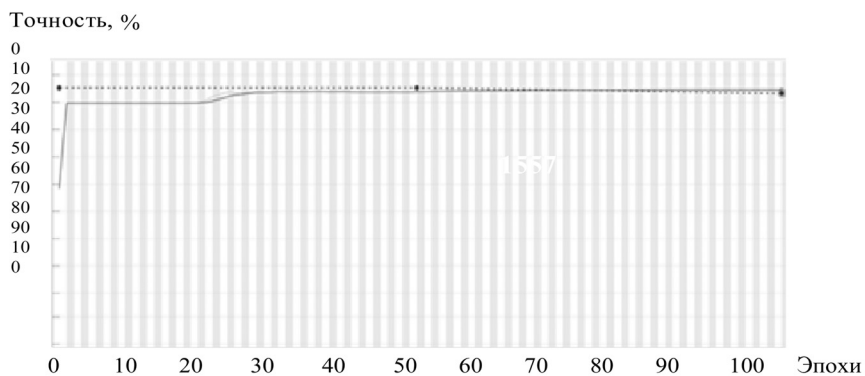


Рис. 3. График обучения LSTM-нейросети

измерения — кгс/см², слева — для расхода, единица измерения — млн м³, ГИС — газоизмерительная станция на входе участка.

Следующим этапом на размеченных данных была обучена классификационная LSTM-нейросеть, данные предварительно были нормированы по математическому ожиданию и среднеквадратическому отклонению.

В качестве вектора входных факторов LSTM-нейросети использовались следующие данные:

- 1) расход с шагом — 1, тыс. м³;
- 2) фактор времени (дата, час);
- 3) разметка отклонений.

Параметры LSTM-нейросети:

- 1) два рабочих слоя: один рекуррентный с модулями LSTM, один полносвязный слой прямого распространения, на выходе — результат классификации по наличию/отсутствию отклонений в данных;
- 2) число нейронов в LSTM — 128;

- 3) алгоритм обучения — ADAM (adaptive moment estimation);
- 4) время обучения — 100 эпох;
- 5) начальная скорость обучения (initial learning rate) равна 0,01;
- 6) количество наблюдений в каждом пакете — MiniBatchSize=1.

График обучения LSTM-нейросети представлен на рис. 3, ошибка обучения — на рис. 4.

Итогом работы нейросети являются значения вероятности наступления события, в данном случае, отклонения, для каждого анализируемого значения измерения. Если рассчитанная нейросетью вероятность превышает некоторое заданное пороговое значение (здесь 0,9), то такой результат измерений классифицируется как отклонение.

Следующим этапом обученная LSTM-нейросеть была применена к тестовой выборке для классификации данных. Результаты классификации представлены

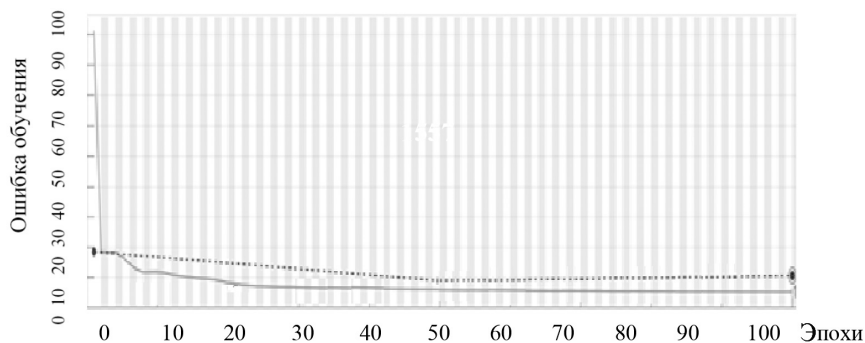


Рис. 4. Ошибка обучения LSTM-нейросети

Истинные классы	0	1557	14
	1	52	129
		0	1
		Предсказанные классы	

Рис. 5. Результаты классификации LSTM-нейросетью, матрица спутанности

в матрице спутанности (confusion matrix), рис. 5. На рис. 5 видно, что нейросеть классифицирует наличие (129) или отсутствие (1557) отклонений. Количество нераспознанных отклонений (52) или принятых за отклонения данных, таковыми не являющихся (14), может быть улучшено качеством разметки или гиперпараметров нейросети.

Компьютерные эксперименты по обучению и предсказанию осуществлялись в среде Matlab [7].

В ходе компьютерных экспериментов было выявлено, что если рассмотреть результаты предсказания на графике данных с разметкой, то можно заметить интересную особенность: нейросеть начинает повышать вероятность наступления некоторых отклонений заранее, за несколько часов до появления отклонения (рис. 6).

Данная возможность делает этот инструмент полезным для заблаговременного предсказания некоторых видов отклонений, повышая тем самым эффективность управления балансом и всей ГТС в целом.

Важно сравнить эффективность классификации алгоритмов машинного обучения и LSTM-нейросети.

С целью сравнения, к статистическим данным по расходу были применены различные алгоритмы машинного обучения.

Результаты сравнения показаны в табл. 1.

Сравнение показало, что LSTM-нейросеть классифицирует более точно, возможность для улучшения существует в более точной разметке и настройке гиперпараметров.

Следующим этапом предложено соединить классификационную LSTM-нейросеть с регрессионной, полученной в [8]. Структура новой нейросети представлена на рис. 7, левый выход — классификационный, правый — регрессионный.

На вход подается тот же набор факторов, что и для регрессионной нейросети [8], в сочетании с разметкой для классификационного выхода. Суммарная ошибка работы обоих выходов представлена на рис. 8.

Применение такой комбинированной нейросети позволит одновременно выявлять отклонения, определять их значения и динамику, а также классифицировать их.

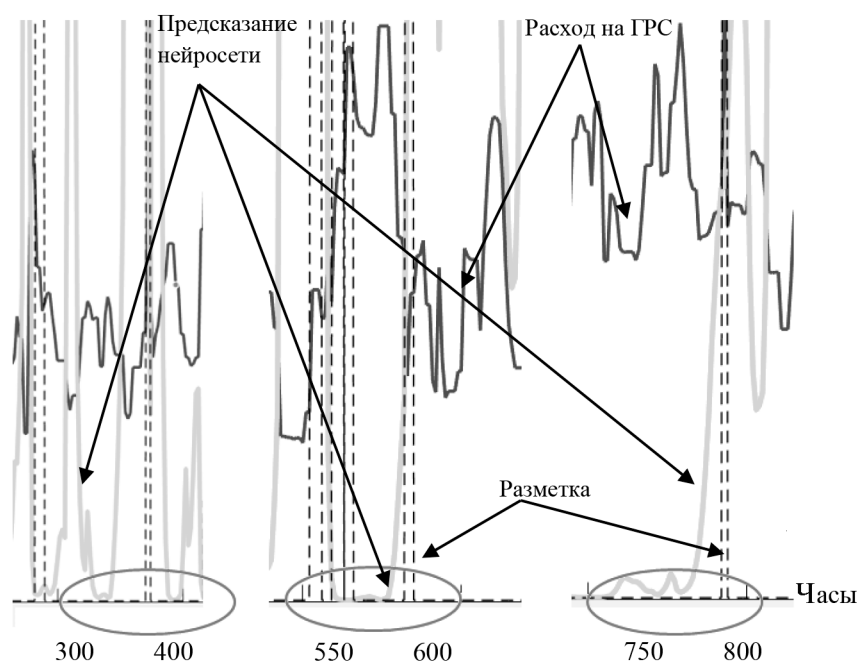


Рис. 6. Заблаговременное повышение вероятности наступления отклонения

Таблица 1

Результаты сравнения классификации алгоритмами машинного обучения

Алгоритм машинного обучения	Точность (test), %
LSTM нейросеть	76
Нейросеть прямого распространения	67
Деревья решений	53
SVM	45
Наивный Байес	38

Заключение и дальнейшие перспективы исследования

В статье рассматривается задача классификации результатов измерения данных технологического процесса учета расхода газа на предмет наличия в них отклонений. Классификация осуществлена при помощи LSTM-нейросети, обученной на размеченных данных, в сравнении с другими алгоритмами машинного обучения. Компьютерные эксперименты показали перспективность использования LSTM-нейросети. Возможным представляется применения комбинированной нейросети с двумя выходами: регрессионным и классификационным.

В дальнейшем, на основе соответствующим образом размеченных данных, отклонения могут быть классифицированы по причинам их возникновения с целью предсказательного обслуживания, заблаговременного выявления и предотвращения причин отклонений для повышения эффективности управления ГТС.

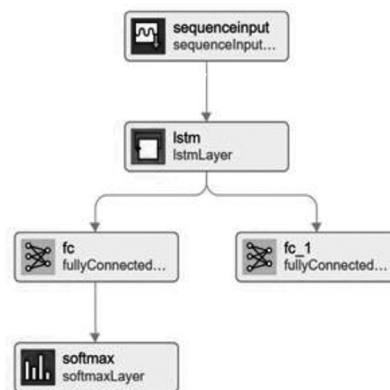


Рис. 7. Структура LSTM-нейросети с двумя выходами

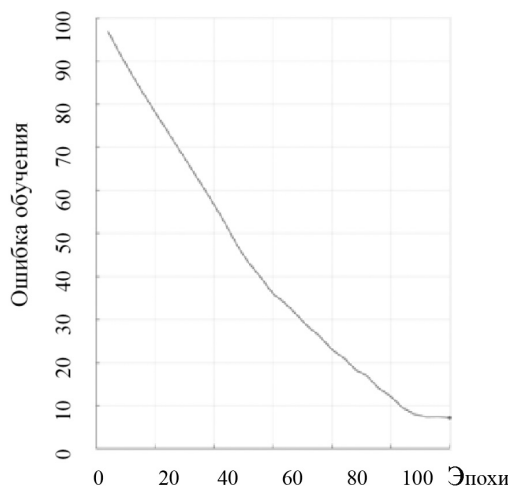


Рис. 8. Суммарная ошибка LSTM-нейросети с двумя выходами

Список использованных источников

1. Отчет «Газораспределение и реализация газа на внутреннем рынке Российской Федерации». <https://groznyrg.ru/pdf/dokladGustov.pdf>.
2. А. Р. Саликов. Разбаланс в сетях газораспределения // Газ России. 2015. № 4. С. 36-41.
3. Г. А. Хворов, С. И. Козлов, Г. С. Аكوпова, А. А. Евстифеев. Сокращение потерь природного газа при транспортировке по магистральным газопроводам ОАО «Газпром» // Газовая промышленность. 2013. № 12. С. 66-69.
4. Сайт ПАО «Газпром». <https://gasandmoney.ru/novosti/gazprom-budet-vychislyat-nebalans-gaza-pri-pomoshhi-big-data>.
5. Цифровая экономика Российской Федерации. <https://digital.gov.ru/ru/activity/directions/858>.
6. С. Хайкин. Нейронные сети. Полный курс. М: Вильямс, 2018, 1104 с.
7. В. П. Дьяконов, В. В. Круглов. MATLAB 6.5 SP1/7/7 SP1/7 SP2 + Simulink 5/6. Инструменты искусственного интеллекта и биоинформации. М.: Солон-пресс, 2006.
8. А. К. Петрова. Детектирование аномалий в технологических процессах газотранспортной системы с использованием сетей с долгой краткосрочной памятью (LSTM) // Сборник докладов Международной конференции «Проектирование и обеспечение качества информационных процессов и систем». СПбГЭТУ «ЛЭТИ», 2022. С. 95-98.

References

1. Report «Gas distribution and gas sales on the domestic market of the Russian Federation». <https://groznyrg.ru/pdf/dokladGustov.pdf>.
2. A. R. Salikov. Imbalance in gas distribution networks // Gas of Russia. 2015. № 4. P. 36-41.
3. G. A. Khorov, S. I. Kozlov, G. S. Akopova, A. A. Evstifeev. Reducing losses of natural gas during transportation through main gas pipelines of OJSC Gazprom // Gas industry. 2013. № 12. P. 66-69.
4. Website of PJSC Gazprom. <https://gasandmoney.ru/novosti/gazprom-budet-vychislyat-nebalans-gaza-pri-pomoshhi-big-data>.
5. Digital economy of the Russian Federation. <https://digital.gov.ru/ru/activity/directions/858>.
6. S. Haykin. Neural networks. Complete course. M.: Williams, 2018. 1104 p.
7. V. P. Dyakonov, V. V. Kruglov. MATLAB 6.5 SP1/7/7 SP1/7 SP2 + Simulink 5/6. Artificial intelligence and bioinformation tools. M.: Solon-press, 2006.
8. A. K. Petrova. Detection of anomalies in the technological processes of the gas transmission system using networks with a long short-term memory (LSTM) // Conference paper «Information processes and systems' design and quality assurance». St. Petersburg state electrotechnical university «LETI». 2022. P. 95-98.