

УДК 621.311.61

Исследование влияния точности входных данных на работу алгоритма рекурсивного уменьшения признаков в задаче прогнозирования состояния здоровья аккумуляторной батареи

Арбузина А.А. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Поляков Н.А. (ИТМО)

Введение. Показателем работоспособности аккумуляторной батареи, характеризующим степень ее деградации, является состояние здоровья SOH (State of Health). Ввиду того, что величину данного параметра невозможно получить с помощью датчика, а также того, что старение аккумуляторной батареи нелинейно зависит от тока и температуры, SOH может только оцениваться. Большое распространение получили методы оценки SOH с помощью моделей машинного обучения в силу возможности решения нелинейных задач с высокой точностью [1]. Для оценки SOH моделями машинного обучения производится инженерия новых показателей здоровья (признаков) аккумуляторной батареи на основе «сырых» данных (тока, напряжения, температуры) [2], и на основе вклада каждого показателя здоровья в прогнозирование SOH осуществляется отбор релевантных из них. Он необходим для дальнейшего упрощения модели, уменьшения вероятности переобучения и ускорения работы алгоритма оценки [3]. Автором данной работы было замечено, что при изменении точности входных данных отбор релевантных признаков с помощью алгоритма рекурсивного уменьшения признаков (Recursive Feature Elimination (RFE)) также меняется. Кроме того, поскольку RFE, который является оберточным методом, склонен к переобучению [4], необходимо осуществлять проверку его работы другими методами отбора признаков. Таким образом, в данной работе исследуется зависимость выбора показателей здоровья аккумуляторной батареи от точности входных данных с помощью алгоритма RFE и производится проверка релевантных показателей другими методами отбора.

Основная часть. Целью работы является исследование изменения важности показателей здоровья для наборов входных данных. Вариативность входных данных обусловлена переходом от лабораторных условий к реальным, где их точность определяется разрешающей способностью установленных датчиков и разрядностью АЦП микроконтроллера. В связи с поставленной целью в работе выделены следующие задачи:

- 1) Произвести отбор показателей здоровья с помощью алгоритма RFE, где в качестве базового алгоритма используется алгоритм случайного леса (Random Forest (RF)), для наборов входных данных;
- 2) Установить зависимость выбора показателей здоровья от точности входных данных;
- 3) Осуществить сравнение отбора показателей здоровья другими базовыми алгоритмами, используемыми в RFE;
- 4) Произвести проверку отобранных показателей здоровья с помощью расчета коэффициента корреляции между каждым признаком и SOH;
- 5) Произвести проверку отобранных показателей здоровья с помощью альтернативного алгоритма.

Выводы. Проведено исследование влияния точности входных данных на выбор показателей здоровья алгоритмом рекурсивного уменьшения признаков RFE. Произведена проверка отобранных алгоритмом RFE показателей здоровья.

Список использованных источников:

1. Antonopoulos, Ioannis & Robu, Valentin & Couraud, Benoit & Kirli, Desen & Norbu, Sonam & Kiprakis, Aristides & Flynn, D. & Elizondo-Gonzalez, Sergio & Wattam, Steve. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic

- review. *Renewable and Sustainable Energy Reviews* (2020). 130. 109899. 10.1016/j.rser.2020.109899.
2. X. Hu, Y. Che, X. Lin, S. Onori, Battery health prediction using fusion-based feature selection and machine learning, *IEEE Transactions on Transportation Electrification* 7 (2) (2020) 382–398.
 3. M. Kuhn, K. Johnson. *Applied Predictive Modeling*, 2013.
 4. Barnali Sahu, Satchidananda Dehuri, Alok Jagadev. A Study on the Relevance of Feature Selection Methods in Microarray Data, *The Open Bioinformatics Journal* (2018) 117-139. DOI <http://dx.doi.org/10.2174/1875036201811010117>