

УДК 621.855

В.О. Бонар

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ВИЯВЛЕННЯ АНОМАЛІЙ OBD2 ДАНИХ ТРАНСПОРТНИХ ЗАСОБІВ МЕТОДОМ НАЙБЛИЖЧИХ СУСІДІВ

V. O. Bonar

OBD2 VEHICLE DATA ANOMALY DETECTION USING NEAREST NEIGHBOORS METHOD

Підтримка справного технічного стану транспортних засобів є актуальною проблемою сьогодення. Виявлення аномалій у сигналах транспортних засобів допомагає передбачити можливі несправності у його окремих компонентах. Збір даних для аналізу проводиться за допомогою системи OBD2, яка є доступною в більшості транспортних засобів. Далі зібрані дані потрібно систематизувати та проаналізувати для виявлення аномальних екземплярів. Метод найближчих сусідів визначає, що нормальні дані знаходяться у щільному сусідстві, аномалії ж найчастіше є ізольованими від решти (1). Дистанція може бути обчислена багатьма методами. Евклідова відстань є найбільш популярною та вимірює довжину прямої лінії між точками (2)

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

, де n - кількість вимірів. Манхеттенська відстань є ще одним популярним методом, що вимірює відстань між точками (2)

$$d(x, y) = \sum_{i=1}^n |y_i - x_i|$$

, де n - кількість вимірів.

Іншим важливим аспектом для методу найближчих сусідів є вибір кількості найближчих сусідів (коефіцієнт k), які будуть використовуватись для подальшої класифікації поточного екземпляру. Цей параметр обирається експериментально і залежить від якості і природи вхідних даних.

Для проведення експерименту виявлення аномалій транспортного засобу було виделено 4 параметри з набору даних (3) для аналізу - швидкість, оберти двигуна, витрата палива та абсолютне положення дросельної заслінки. Для визначення відстані використано Евклідову відстань, кількість найближчих сусідів (коефіцієнт k) для перевірки вибрано 5.

Важливим кроком при застосуванні методу найближчих сусідів є нормалізація даних для уникнення упереджених результатів через метрики з більшим значенням. Наприклад, швидкість транспортного засобу може досягати 200 км/год, тоді як оберти двигуна можуть досягати набагато більших значень, що спричинить нерівномірний вплив на кінцевий результат обчислювальної відстані. Для нормалізації було використано мінімально-максимальний метод

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

, де x_{min} та x_{max} - мінімальні та максимальні значення певної метрики в даному наборі даних.

Висновок. Було проведено експеримент виявлення аномалій при аналізі швидкості, обертів двигуна, витрати палива та абсолютного положення дросельної заслінки транспортного засобу. Дані були проаналізовані використовуючи мову програмування Python та реалізації методу найближчих сусідів з пакету sklearn - NearestNeighbors (4). Отримані результати у вигляді діаграм показують залежність аномалій від кожного аналізованого параметру (рис. 1) та розподіл нормальних даних (рис. 2).

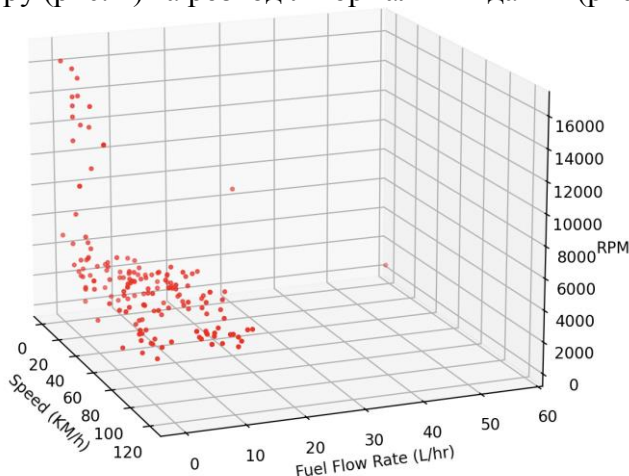


Рисунок 1. Розподіл аномальних даних

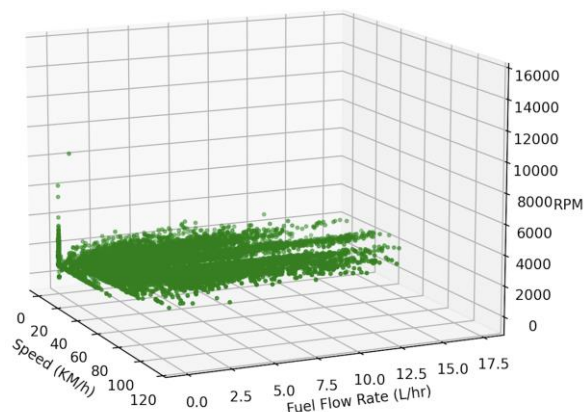


Рисунок 2. Розподіл нормальних даних

Результати проведеного експерименту виявлення аномалій методом найближчих сусідів в сигналах транспортного засобу будуть використані в подальших дослідженнях віддаленої діагностики транспортних засобів.

Література

1. Zhao, Ming & Chen, Jingchao & Li, Yang. (2018). A Review of Anomaly Detection Techniques Based on Nearest Neighbor. 10.2991/cmsa-18.2018.65.
2. Suwanda, R & Syahputra, Zulfahmi & Zamzami, Elviawaty. (2020). Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K. Journal of Physics: Conference Series. 1566. 012058. 10.1088/1742-6596/1566/1/012058.
3. OBD-2 Vehicular Dataset Trace? Kaggle: веб-сайт. URL: <https://www.kaggle.com/datasets/brunops2/obd-2-vehicular-dataset-trace> (дата звернення 9.11.2024)
4. NearestNeighbors. Scikit Learn: веб-сайт. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html> (дата звернення 9.11.2024)