

С Е К Ц И Я 16

СОВРЕМЕННЫЕ ТЕХНИКА И ТЕХНОЛОГИИ ТРАНСПОРТИРОВКИ И ХРАНЕНИЯ НЕФТИ И ГАЗА

ЧИСЛЕННОЕ СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ПОЛИЭТИЛЕНОВЫХ ГАЗОПРОВОДОВ

Афанасьев Д.Ю.

Научный руководитель – доцент А.Г. Зарубин

Национальный исследовательский Томский политехнический университет, г. Томск, Россия

Аннотация: В статье проведено численное моделирование технического состояния полиэтиленовых газопроводов, приведена классификация с помощью Gaussian process classification (GPC).

Ключевые слова: газопровод, классификация, нормальное распределение, машинное обучение, классификация данных.

Газораспределительная организация, в соответствии с действующим законодательством, в первую очередь отвечает за безопасность эксплуатации различных объектов газоснабжения, а также обеспечение поставок газа потребителям. Требования к обследованию, эксплуатации, техническому обслуживанию, диагностированию газопроводов, установлены нормативными документами и законодательными актами Ростехнадзора РФ, в том числе и расчетный срок эксплуатации газопроводов. Следовательно, эксплуатирующей организацией до достижения окончания срока эксплуатации должно быть организовано техническое диагностирование объекта, и уже по его результатам будет принято решение — продлить срок эксплуатации на определенный период или вывести этот объект из эксплуатации, если он не соответствует требованиям. Все работы выполняются аккредитованными организациями и по действующим принятым методикам, утвержденным Ростехнадзором.

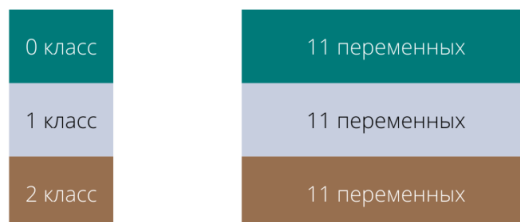
В настоящее время, используя методы машинного обучения и численного моделирования, можно достаточно точно предсказать техническое состояние газопровода, имеющего необходимый набор входных параметров. Задачи подобного плана решаются с использованием проекционных методов, которые основаны на статистическом подходе к обработке информации – методом главных компонент (PCA), PLS-дискриминации (PLS-DA) и других.

Выборка входных данных формирует некоторое распределение, и на сегодняшний день наиболее известное распределение — распределение Гаусса, которое чаще называют нормальным распределением.

Распределение предоставляет математическую функцию с определенными параметрами, которую можно использовать для расчета вероятности любого отдельного наблюдения из выбранных данных. Оно описывает группировку или плотность наблюдений, называемую функцией плотности вероятности. Таким образом, используя нормальное распределение, можно рассчитать вероятность того, что наблюдение будет иметь значение, равное или меньшее заданного значения [3].

Построить модель машинного обучения можно на основе распределения Гаусса, которая предскажет, к какому классу (техническому состоянию) будет относиться трубопровод с новым набором переменных [2]. Однако необходимо убедиться в том, что эта модель на самом деле работает, и ее прогнозы будут достаточно точны прежде, чем применить модель к новому набору. И для этого в первую очередь необходимо разделить выборку на обучающую и проверочную в соотношении, указанном на рисунке 1.

Обучающий набор 70 %



Проверочный набор 30 %



Рис.1 Структура данных

Проверочный набор не использовался для построения модели, но у каждого трубопровода известен класс, к которому его относят. Таким образом, можно сделать прогноз для каждого трубопровода в проверочном наборе и сравнить его с фактической меткой (уже известным классом). Оценить качество модели можно, вычислив точность (ассигасу) – процент трубопроводов, для которых модель правильно спрогнозировала класс.

Вторым шагом стало применение метода главных компонент (РСА). Используя этот метод, появляется возможность учесть главные взаимодействия и получить несколько более полную картину. Алгоритм находит первые две главные компоненты и визуализирует датасет в этом новом двумерном пространстве с помощью диаграммы рассеяния, показанной на рисунке 2, где по оси у отложена вторая главная компонента (pc2), а по оси x – первая главная компонента (pc1), затем алгоритм пользуется информацией о классах, чтобы выделить точки разным цветом.

Заключительным этапом стало использование алгоритма GaussianProcessClassifier. На выходе получается следующая классификация [1]:

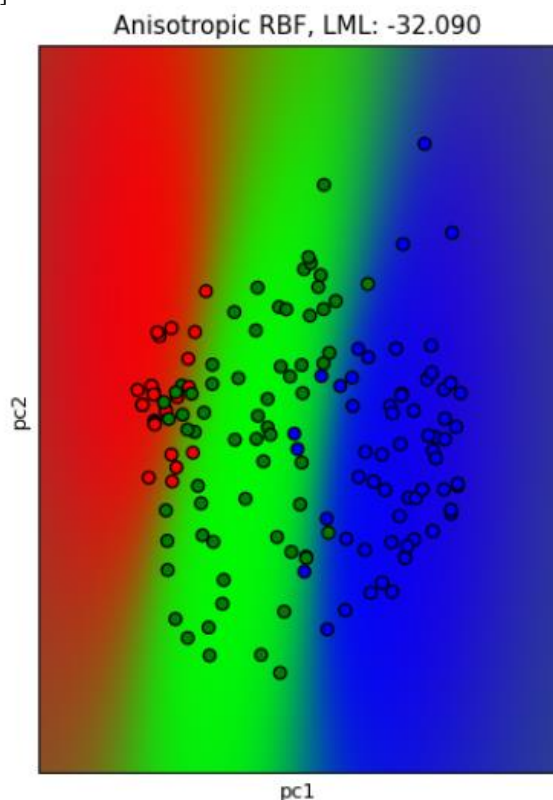


Рис.2 Gaussian process classification (GPC). Анизотропное ядро RBF в двумерной версии

Достаточно четкое разделение границ классификации свидетельствует о том, что правильно выбраны необходимые параметры газопровода. Проведенное статическое моделирование иллюстрирует прогнозируемую вероятность GP-классификации для анизотропного ядра RBF (радиальной базисной функции). Высокая точность достигается за счет применения метода главных компонент, для сокращения размерности данных [3].

В ходе настоящей работы был применен метод GPC к обучающему и проверочному наборам данных. В результате чего было выявлено, что, модель может демонстрировать точность при определении технического состояния полиэтиленовых газопроводов более 0.8.

Литература

1. Gaussian process classification (GPC) on iris dataset [Электронный ресурс] – Режим доступа: https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpc_iris.html#sphx-glr-auto-examples-gaussian-process-plot-gpc-iris-py.
2. Аралов О. В. Методология управления качеством сложных технических систем на объектах магистрального трубопроводного транспорта нефти и нефтепродуктов //Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. – 2019. – Т. 9. – №. 6. – С. 608-625.
3. Müller A. C., Guido S. Introduction to machine learning with Python: a guide for data scientists. – "O'Reilly Media, Inc.". – 2016.