

Шицелов Анатолий Вячеславович

магистр, инженер

Центр прогнозирования и развития

ФГБОУ ВО «Югорский государственный университет»

г. Ханты-Мансийск, ХМАО – Югра

АЛГОРИТМЫ АДАПТИВНОГО БУСТИНГА КАК ИНСТРУМЕНТ ПРОГНОЗИРОВАНИЯ АВАРИЙ НА ПРИМЕРЕ РАЗЛИВА НЕФТИ В ХМАО – ЮГРЕ

Аннотация: в статье представлен общий подход к прогнозированию участков, на которых произойдёт разлив нефти в регионе с использованием модели черного ящика и алгоритмов адаптивного бустинга. Предложены подходы к формированию обучающей выборки и особенности работы с алгоритмом AdaBoost.

Ключевые слова: прогноз, машинное обучение, алгоритм.

Введение

Прогнозирование аварий, в частности, участков, на которых может произойти разлив нефти в ближайшее время, помогает эффективнее бороться с авариями на нефтепроводах. Задача прогнозирования аварий нефтепроводов, относится к классу задач, которые сложно формализовать, в связи с этим для решения предлагается использовать методы «Больших данных».

«Большие данные» это совокупность методов и инструментариев для обработки как структурированных, так и неструктурированных данных больших объёмов. Среди основных методов, которые можно использовать для прогнозирования аварий в регионе, стоит выделить: машинное обучение, в частности, алгоритмы адаптивного бустинга и прогнозная аналитика.

С точки зрения системного анализа алгоритмы адаптивного бустинга это каскад простых алгоритмов (например, линейной регрессии), но скомбинированных оптимальным образом для решения сложных задач, которые можно деконструировать на множество более простых [1; 2].

Прогнозирование аварий на примере разлива нефти в ХМАО-Югре

Общий алгоритм использования алгоритмов адаптивного бустинга, в частности, AdaBoost для прогнозирования аварий в регионе, можно представить в следующем виде.

1. Определяются первичные данные, которые формируют обучающую выборку и эталонные значения [3]. Эталонные значения соответствуют параметрам, которые прогнозируются:

$$I_{ij} \quad i = 1 \dots n \quad j = 1 \dots m, \quad (1)$$

где n – количество эталонных (прогнозируемых) значений в одном образе, j – количество образов. Обучающая выборка формируется из параметров, которые влияют на прогнозируемое событие:

$$O_{lj} \quad l = 1 \dots k \quad j = 1 \dots m, \quad (2)$$

где k – количество параметров, которые влияют на прогнозируемое событие. Значение обучающей выборки и эталонных значений нормируют в соответствии с планируемой использованием функции активации.

2. Определить типы параметров: число или класс. Для чисел провести нормализация, т. е. привести все числовые данные к диапазону от 0 до 1 так как все параметры в исходных данных лежат в диапазоне больше или равному 0. Для классов необходимо определить способ их представления: тут несколько подходов:

- закодировать каждый класс как позицию в разражённом векторе;
- сопоставить каждому классу некоторый числовой вектор.

Так как выборка для обучения мала (около 2000 примеров), то этого не достаточно для сопоставления каждому классу числового вектора, так как это потребует построение еще одной рекуррентной нейронной сети. Первый же вариант позволяет преобразовать все классы из тренировочных данных в вектор, сопоставив каждому классу его позицию в векторе. Будем кодировать класс как номер поэзии в векторе, установив в позицию класса 1, а все остальные компоненты вектора опустив в 0.

3. Настройка параметров AdaBoost схожа с настройкой нейронной сети, так же необходимо определить шаг обучения и установить максимальное число слоев, только в адаптивном бустинге вместо функции активации необходимо выбрать нижележащий алгоритм машинного обучения каскад из которых будет строиться, поэтому правила для нейронных сетей применимы и для AdaBoost. Поэтому большинство правил предложенный в [4] будут применимы и здесь.

4. Инициализируется процесс обучения выгладит следующим образом:

– обучить нижележащий алгоритм. Если его точности достаточно, то закончить обучение иначе добавить еще один слой;

– обучить нижележащий алгоритм машинного обучения в новом слое и проверить точность, если достигнута необходимая точность, то закончить иначе добавить еще слой и повторить данный шаг, пока не будет достигнут лимит слоев или заданная точность.

Список литературы

1. Freund Y. A Short Introduction to Boosting / Y. Freund, R.E. Schapire // Journal of Japanese Society for Artificial. – 1999. – №14 (5). – 771–780 p.

2. Polikar R. Ensemble Based Systems in Decision Making // IEEE Circuits and Systems Magazine. – 2006. – Vol. 6. – №3. – P. 21–45.

3. Татьянкин В.М. Обучающая выборка в задаче распознавания образов при использовании нейронных сетей / В.М. Татьянкин, И.С. Дюбко // Вестник Югорского государственного университета. – 2015. – №2 (37). – С. 94–98.

4. Татьянкин В.М. Подход к формированию архитектуры нейронной сети для распознавания образов // Вестник Югорского государственного университета. – 2016. – №2 (41). – С. 61–64.