

УДК 004.85

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ ЭКОЛОГИЧЕСКОЙ ОПАСНОСТИ

Турсуков Н.О. (Национальный исследовательский университет ИТМО)

Научный руководитель – к.т.н., преподаватель ФБИТ Виксин И.И.
(Национальный исследовательский университет ИТМО)

В работе были проанализированы методы машинного обучения и их применение для предотвращения экологических катастроф. Были определены условия, сформирован набор данных и проведен эксперимент по тестированию методов мониторинга окружающей среды, позволяющих прогнозировать экологические угрозы.

Введение. Системы мониторинга позволяют оценивать состояние почвы, атмосферы и гидросферы. При достижении определенного порога концентрации веществ в окружающей среде, может возникнуть экологическая угроза. Для прогнозирования угроз и минимизации рисков, создаются системы раннего предупреждения на основе машинного обучения. При этом в критически важных объектах очень важно безошибочно определить концентрацию потенциально опасных веществ и не спутать его с другими веществами, находящимися в границах определенной площади. Следовательно, нужно четко разграничить опасные вещества от безвредных. Методы машинного обучения позволяют предсказать увеличение концентрации вредных веществ, как при поломке датчиков, так и при неполноте собираемых датчиками данных.

Основная часть. Задачи машинного обучения, применимые к мониторингу критически важных объектов:

- кластеризация;
- классификация.

В ходе исследования существующих решений применений машинного обучения было выявлено, что предпочтение отдается изучению случайного леса и опорных векторов в силу возможности эффективно обрабатывать данные с большим числом признаков. Многослойный перцептрон не был рассмотрен из-за неспособности классифицировать данные находящиеся вне обучающей выборки, что в случае неопределенности и недостоверности данных может негативно сказаться на результатах мониторинга.

Для тестирования методов машинного обучения были определены факторы, влияющие на критический параметр. В качестве веществ, концентрацию которых необходимо было спрогнозировать, были взяты ультрадисперсные частицы классов PM2.5 и PM10, находящиеся в воздухе. Наборы данных собирались с действующих станций мониторинга окружающей среды.

Для формирования набора исходных данных были определены параметры окружающей среды, которая была представлена как область трехмерного пространства ограниченного периметра. Параметры влияют на область как извне (метеорологические параметры), так и изнутри, во время функционирования критически важного объекта. Данные, собираемые станцией, включают в себя метеорологические: направление ветра и его скорость, температура, давление и влажность, а также концентрацию частиц (PM10 и PM2.5). Используемые в тестировании данные собирались каждые пять минут. Для выборок были взяты данные за год и за месяц (~9000 показателей).

По существующим наборам данных было построено нормальное распределение концентрации частиц по выборке. В качестве статистического метода исследования данных был проведен корреляционно-регрессионный анализ, с целью определения зависимости между изучаемыми параметрами. В результате были определены коэффициенты корреляции и детерминации. Одной из протестированных моделей была регрессионная модель, для

оценки которой использовались метрики: абсолютное отклонение, средняя квадратическая ошибка и квадратный корень из средней квадратической ошибки.

Выводы. В ходе работы были исследованы существующие решения по прогнозированию техногенных катастроф, собраны и нормализованы наборы данных, пригодные для анализа методов машинного обучения. Также было проведено их тестирование с помощью полученных наборов, с дальнейшей целью разработки методов, решающих задачу мониторинга области в условиях неопределенности и недостоверности данных.

Турсуков Н.О. (автор)

Подпись

Викснин И.И. (научный руководитель)

Подпись