

УДК 621.315.1

Е.М. Воронкова, А.Г. Темников
Национальный исследовательский университет «МЭИ»
Москва, Россия

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ОТКЛЮЧЕНИЙ В РАСПРЕДЕЛИТЕЛЬНЫХ СЕТЯХ

Аннотация. В статье рассматривается возможность использования алгоритмов машинного обучения для детализации причин технологических нарушений климатического характера для воздушных линий электропередачи в распределительных сетях.

E.M. Voronkova, A.G. Temnikov
National Research University «Moscow Power Engineering Institute»
Moscow, Russia

APPLICATION OF MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION OF OUTAGES IN DISTRIBUTION NETWORKS

Abstract. The article considers the possibility of using machine learning algorithms to detail the climatic nature causes of technological disturbances for overhead transmission lines in distribution networks.

Исследования в области надёжности электрических сетей, за последние годы всё больше акцентируют внимание на влиянии климатических факторов на процессы передачи и распределения электрической энергии. Тем не менее, спрогнозировать изменение конкретного климатического фактора – достаточно сложная задача, в частности это относится к характеристикам скоростей ветра, грозовой деятельности региона, так как данные показатели являются вариативными и зависящими от большого количества факторов. Однако в настоящее время с учетом больших статистических баз данных можно выявить конкретные тенденции и зависимости, позволяющие заблаговременно выявлять опасные факторы для конкретного региона для повышения устойчивости объектов энергетики, а также объектов в других секторах экономики.

В случае для распределительных и системообразующих электрических сетей тенденции влияния изменения климата можно наблюдать по данным аварийности. На сегодняшний день на территории Российской Федерации согласно Постановлению

Правительства № 24 [1] субъектами электроэнергетического рынка предоставляется сводная информация об аварийных отключениях в сетях территориальных зон деятельности организации. Данная информация представляет собой сведения о техническом состоянии электрических сетей, предоставляемые непосредственно для потребителей по итогам актов расследований аварий.

Детализация климатических причин отключений по статистическим данными устанавливает все возможные причины воздействия климата на территории конкретного региона, позволяет оценить потенциальный вклад меняющихся климатических условий, устанавливает вклад каждого фактора в общий уровень аварийности, а также тенденцию их влияния и возможные последствия. Поэтому на примере данных аварийности электросетевых компаний Центрального и Приволжского Федеральных округов с 2011 по 2020 гг. [2] были выявлены технологические нарушения для воздушных линий электропередачи (ВЛ), причинами которых являлись именно климатические факторы.

Стоит понимать, что предоставляемая субъектами информация не всегда сразу пригодна для анализа, что в первую очередь связано с различными вариантами типов и форматов файлов, несоответствием столбцов, а также изменениями о необходимой предоставляемой информации на основании Постановлений. Поэтому целью данной работы является применение информационных технологий, как одного из способа технологического развития, для получения готовых баз данных по аварийным отключениям.

Самой главной задачей работы является выполнение классификации аварий по типам причин климатического характера, а также по типам поврежденного оборудования, так как ручная обработка занимает достаточно длительное время, не позволяющее оценить всю ситуацию для региона. Для достижения данной цели, первоначально необходимо было выполнить предварительную подготовку данных, которая включает в себя:

1. преобразование различных форматов данных;
2. ручная обработка данных для представления информации в удобном формате, а также исключения данных, однозначно не относящихся к критерию «климатические данные»;
3. ручная классификация части данных (размеченные данные);

В результате такой предварительной подготовки по данным [2] было выявлено порядка 300 тысяч значений аварийных отключений, из которых было выделено 25% размеченных данных (данных, на основе

которых можно выполнить машинное обучение по классификации) с конкретной причиной аварии, типом поврежденного электрооборудования и другими признаками. Очевидно, что дальнейшая ручная обработка может занять достаточно длительное время. Поэтому для ускорения, а также упрощения получения необходимого набора данных предлагается использовать различные алгоритмы машинного обучения для естественных языков.

Детализация всех возможных причин повреждения по статистическим данным достаточно обширная и в рамках классификации текста машинного обучения будет затруднительна, в виду небольшой выборки для отдельных причин (например, паводок встречается всего несколько раз). Поэтому для найденных в ходе ручной обработки данных причин отключения необходимо выполнить их группировку признаков по близости происхождения и последствиям.

В результате было решено выделить 9 основных кластеров причин повреждения оборудования, для которых будет выполняться анализ: ветер; гроза; ветер, гроза; ветер, сопутствующие факторы; ветер, гроза, сопутствующие факторы; гроза, сопутствующие факторы; осадки/осадки и другое; ГИО, сопутствующие факторы; прочее.

Таким образом, для полученных 25% размеченных данных с девятью возможными типами причин отключений необходимо выполнить машинное обучение для классификации оставшейся части неразмеченных данных. При классификации текстовой информации нужно выделить несколько этапов:

1. Предварительная обработка текста от знаков препинания, чисел, знаков, стемминг, лемматизация и др.;
2. Выбор подходящего алгоритма классического машинного обучения для размеченных данных в зависимости от необходимой точности и скорости обучения;
3. Проверка точности прогнозирования предсказанных данных с реальными;
4. Применение выбранного алгоритма обучения для неразмеченных данных.

Опыт использования алгоритмов машинного обучения находит применение во многих сферах деятельности [3], однако его выбор определяется типом данных. В нашем случае для текстовой информации, содержащей сведения о причине отключения, а также типе поврежденного оборудования, было применено четыре модели классификаторов: логистическая регрессия, наивный Байесовский процесс, метод опорных векторов, нейронные сети [4].

Для выбранных алгоритмов было выполнено разделение данных на обучающую и тестовую выборки в соотношении 70 на 30%, отобрано одинаковое количество аварий каждой категории для обучения и выполнено само обучение. Полученные результаты показали, что наиболее эффективным методом классификации причин отключений для нашей задачи является алгоритм логистической регрессии с точностью 90%, а также одномерная нейронная сверхточная сеть с точностью 90-94%. Для ускорения процесса обучения и дальнейшего прогнозирования было решено остановиться на алгоритме логистической регрессии.

Аналогичное обучение при классификации типов поврежденного оборудования составило 98% (методом логистической регрессии).

Для оценки точности предсказанных значений в работе была рассчитаны матрицы ошибок модели на тренировочной и тестирующей части выборки, результаты которых показали, что практически все предсказанные значения соответствуют реальным: особенно при классификации оборудования.

Далее непосредственно было выполнено прогнозирование причин аварий и типов поврежденного оборудования на основе выбранного алгоритма машинного обучения и выполнено суммирование количества отключений соответствующей категории по годам. Полученные результаты продемонстрированы для Владимирской области (рис. 1). Аналогичный расчёты были выполнены и для оставшихся регионов.

Также для всех отключений, связанных с фактором «гроза», было представлено региональное распределение отключений за 10 лет (рис. 2), для уточнения региона с большим воздействием грозовой активности – Кировская область.

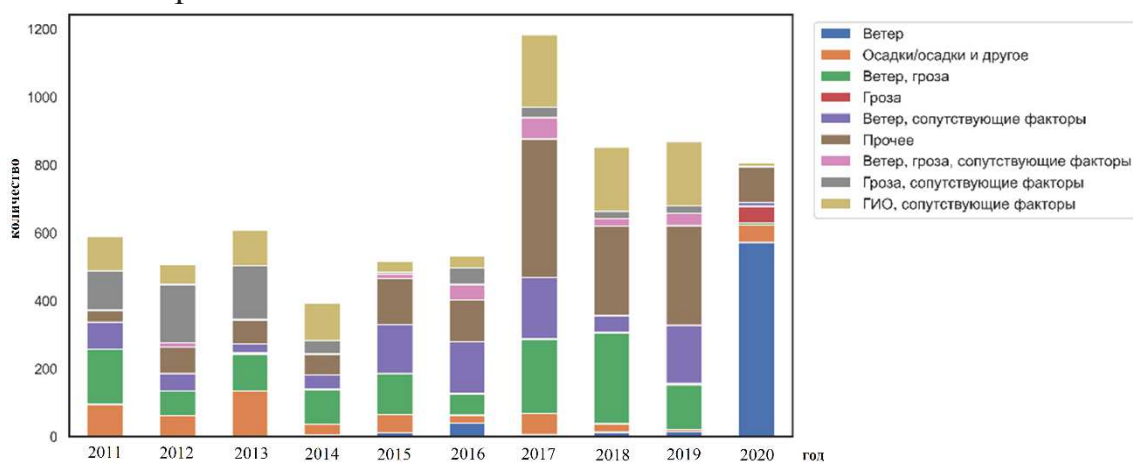


Рис. 1 – Распределение аварий по причинам климатического характера для Владимирской области с 2011 по 2020 гг.

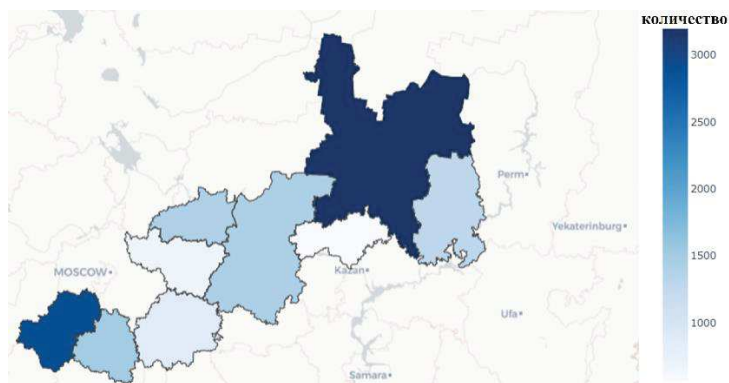


Рис. 2 – Карта распределения аварийности по грозе за 10 лет

Очевидно, что отключения, не зависящие от действий человека, сложно предупредить. Именно поэтому необходимо своевременно оценивать воздействия различных параметров, в том числе факторов окружающей среды, для разработки мероприятий по повышению надежности. Однако, стоит понимать, что анализ аварийности по климатическим факторам является первоначальной оценкой уязвимости ВЛ к действию природных факторов. Для более детального учета необходимо также обладать информацией о характеристике местности расположении опор по трассе ВЛ (равнинная местность, вершина холма, в долине, на склоне и т.д.). В журналах аварийности отсутствует информация о территориально-географическом месте поврежденного участка, что затрудняет детализацию районов по наиболее опасным климатическим факторам. Тем не менее такой анализ дает общее представление о влиянии климатических факторов на электрические сети.

Список использованных источников

1. Постановление Правительства РФ от 21 января 2004 г. № 24 «Об утверждении стандартов раскрытия информации субъектами оптового и розничных рынков электрической энергии».
2. Сводные данные об аварийных отключениях, 2011-2020. URL: https://mrsk-cp.ru/for_consumers/service_area/network_condition/svodnye-dannye-ob-avariynykh-otklyucheniy_akh/.
3. Ишиев Н.Л., Ракунов С.В. Practical Application of Machine Learning and GIS. URL: <https://arcreview.esri-cis.ru/2020/09/29/practical-experience-using-machine-learning-algorithms/>.
4. A Complete Guide to Machine Learning Algorithms. URL: <https://medium.com/swlh/a-complete-guide-to-machine-learning-acd426e2a373>.