

Особенности подготовки и обработки данных для прогнозирования событий в распределительных электрических сетях

Синяков С.А., руководитель отдела разработки ПО МНПП «АНТРАКС»

В 2006 году, выступая на конференции по маркетингу, математик Клайв Хамби высказал тезис, который сейчас определяет развитие не только отдельно взятых IT-фирм, но и целых отраслей: «Данные — это новая нефть. Как и нефть, они ценные, но не сами по себе, а благодаря продуктам на их основе. Данные должны быть обработаны, проанализированы для того, чтобы извлечь из них ценность, повысить рентабельность бизнеса». Что мы видим спустя полторы декады — большие объемы данных, или Big Data, становятся частью бизнес-модели в различных направлениях деятельности. Аналитики компании IDC прогнозируют, что к 2025 году во всем мире объем информации, который накоплен компаниями и потребителями, достигнет отметки в 163 зеттабайта. По сравнению с 2016 годом этот показатель вырастет в 10 раз.

По наблюдениям Ассоциации больших данных, в Российской Федерации данные от таких отраслей, как телекоммуникации или IT используются для обучения нейросетей гораздо чаще, чем данные таких отраслей, как Металлургия, Рetail или Нефтегазовая отрасль. Данные же от электроэнергетики пока никак не фигурируют в отчетах. С чем это связано и как можно изменить ситуацию — пойдет речь в статье.

Для начала отметим основные особенности, связанные с подготовкой и обработкой данных в распределительных электрических сетях:

- события в электроэнергетике происходят достаточно редко; отсутствуют описания происходящих событий в открытом доступе (только в журнале диспетчера), откуда следует проблема классификация данных — сопоставить реальные события и полученные данные; фактически отсутствуют специализированные для нейросетей устройства регистрации нормального/аномального поведения распределительной электросети;
- нет понимания необходимых типов данных для сбора и для использования в алгоритмах нейросетевого анализа;
- в текущих временных рядах, получаемых из штатных регистраторов преимущественно нет достаточной для построения моделей нейросети частоты дискретизации;
- недостаточно наложен процесс сотрудничества разработчиков искусственного интеллекта с пользователем;
- в отрасли отсутствует понимание необходимости сбора данных нормального режима работы РЭС;
- в применяемых системах мониторинга отсутствует подходящая организация процесса сбора и хранения данных.

Постараемся рассмотреть все по порядку. Для чего мы хотим собирать данные? В первую очередь, актуальные данные помогают нам строить статистические модели и производить аналитические выводы, что в свою очередь дает повод для внесения изменений

в систему с целью достижения лучших показателей эффективности самой системы. Более качественный анализ, подразумевающий разделение классифицированных данных по соответствующим кластерам и использование методов искусственного интеллекта, может дать нам больше — предиктивную аналитику.

Цифровизация электроэнергетики идет полных ходом, продолжаясь уже не первый год, и на первый взгляд может показаться, что с введением новых цифровых РЭС количество данных, пригодных для проведения статистического анализа должно расти прямо пропорционально их количеству. На деле все гораздо сложнее, ведь отрасль электроэнергетики характеризуется нестабильной и, на первый взгляд, не поддающейся прогнозированию частотой возникновения аварийных ситуаций. Добавляя к этому отсутствие специализированных устройств регистрации данных и недостаточное понимание типов данных, необходимых для качественного анализа, мы сталкиваемся с проблемой получения, сбора и обработки данных, а впоследствии с проблемой построения архитектуры нейросети для дальнейшего прогнозирования. Вдобавок ко всему, во многих даже цифровых РЭС не уделяется достаточно внимания организации хранения данных, а также электронному документообороту.

Допустим, что мы собрали достаточно данных хорошего качества в виде аварийных событий (более 500 тысяч штук), провели

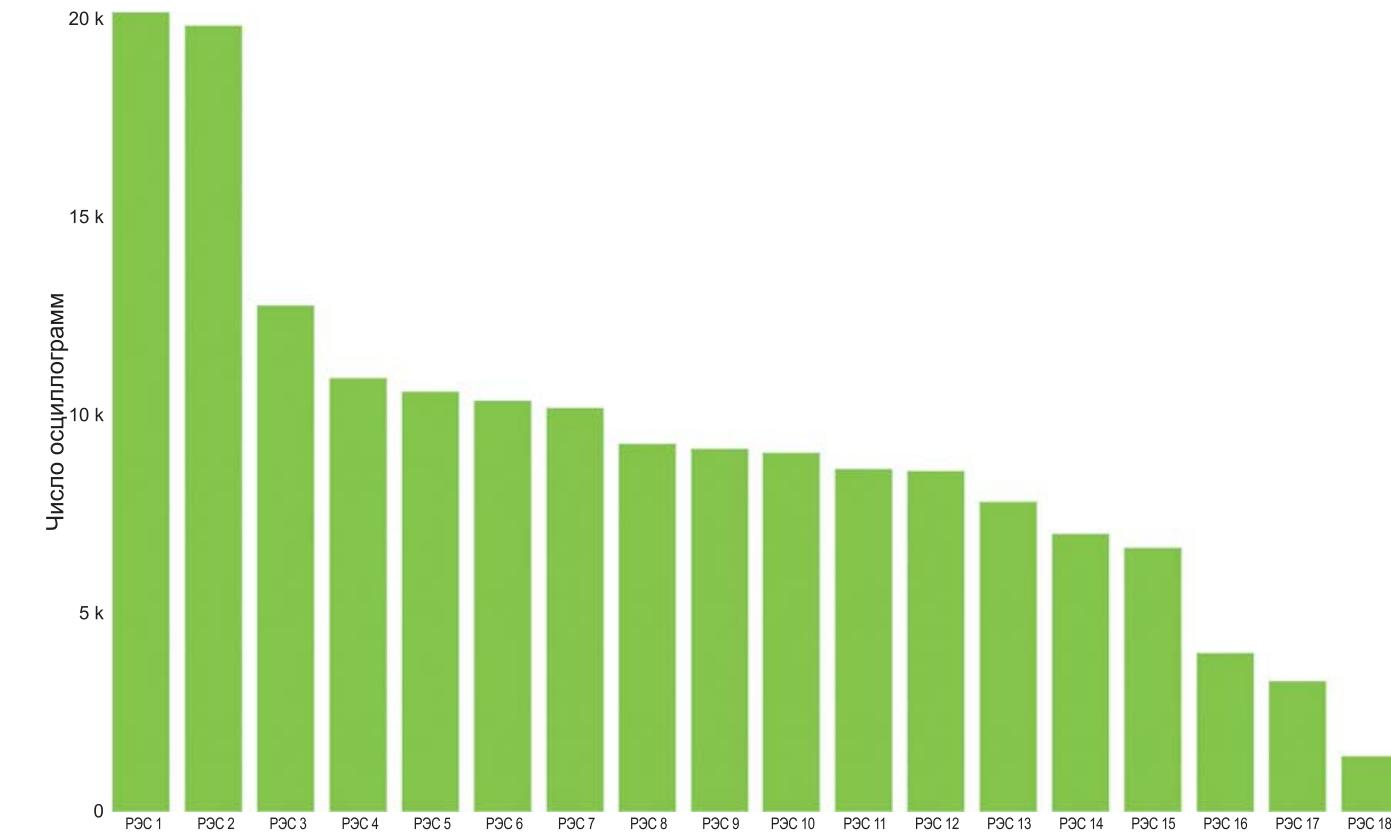


Рис. 1. Количество данных с некоторых РЭС, в которых установлены устройства АНТРАКС

анализ и разработали модель нейросети, которая с учетом топологии и особенностей цифрового РЭС способна решить задачу прогнозирования, но далее, чтобы наложить данные на реальные события, необходимо изучить эти 500 тысяч событий из записей журналов диспетчера и сопоставить их с набором данных.

На данный момент такого полноценного обучения нейросети не существует, так как невозможно специально создавать сотни и тысячи аварийных ситуаций в реальной работающей распределительной сети.

Несмотря на все эти трудности и нетривиальность задачи, мы в компании АНТРАКС сможем ее решить, и вот что мы делаем. Мы имеем достаточное количество устройств, которые адаптированы и сертифицированы для цифровизации РЭС, и которые уже установлены на объектах несколько лет. Эти приборы способны не только регистрировать аварийные события (ОЗЗ и МФЗ), но и записывать данные в виде осцилограмм аварийных процессов, протекающих в распределительной сети. Также нами разработана

система мониторинга воздушных линий — КОМОРСАН, архитектура БД которой позволяет обращаться к архивным данным (рисунок 1).

Мы взяли некоторое количество временных рядов, записанных нашими устройствами ИКЗ-В34Л, установленными в разных местах, и попробовали обучить нейросеть на имеющихся данных.

Результаты были положительны, и мы поняли, что характеристики приборов позволяют нам использовать их как регистраторы данных.

Собраны более 170 тысяч осцилограмм с более чем 8 тысяч устройств ИКЗ-В34Л. В текущий момент эти данные делятся алгоритмами искусственного интеллекта на условные кластеры, внутри которых определены характеристики данных для определенного типа события, такого как пробой изолятора, например, или коммутационные перенапряжения. Некоторые кластеры содержат внутри себя шум с точки зрения физических параметров,

рий диспетчера, искусственный интеллект в отличие от человека сможет не просто отличать их от нормального режима работы, но и оценивать относительную вероятность причастности шума к конкретному событию в распределительной сети (рисунок 2).

Тут мы подобрались к основной проблеме подготовки данных для распределительных сетей — это их разметка.

Для дальнейшего развития продвинутого подхода в электроэнергетике нам требуется убедить персонал в необходимости сотрудничать с искусственным интеллектом — от персонала требуется помочь размечать происходящие события.

Алгоритмы должны явно понимать, чем данные об обрыве провода отличаются от данных, связанных, скажем, с ускоренным старением изоляции и повреждением электрооборудования. В нашем проекте с нами сотрудничает несколько РЭС в разных частях страны, которые заинтересованы в развитии цифровизации своих объектов. Из этих pilotных РЭС мы получаем информацию о событиях, запи-

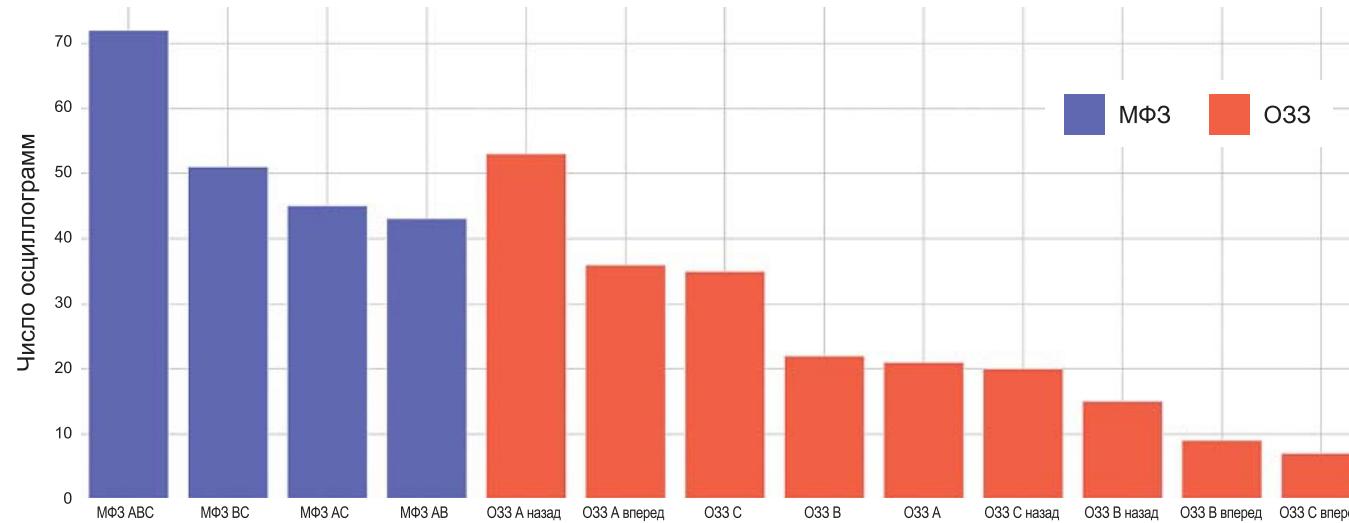


Рис. 2. Пример распределения аварий по типам для отдельно взятого РЭС

санных в оперативных журналах диспетчеров, затем наши data-инженеры размещают собранные на их пилотных фидерах данные и наши разработчики обучают нейросеть на полученные тем самым события, одновременно проверяя их принадлежность к кластерам данных, упомянутых выше. И если события попадают в какой-либо неопределенный кластер, уже содержащий сотни или тысячи временных рядов, то нейросеть переобучается и точность определения события резко возрастает.

В основу формирования data-сета для обучения нейросети должны ложиться все аномалии относительного нормального режима работы, зафиксированные в распределительной сети (рисунок 3). Для этого необходим «отпе-

чаток пальца» РЭС. В пилотных регионах мы обязательно сни- маю нормальный режим работы сети. Зачастую энергетики счита- ют, что регистрация нормального режима работы — это излишние затраты. Однако непрерывные временные ряды с достаточной частотой дискретизации могут дать понимание о характере ра- боты распределительной сети и режим, кажущийся нормаль- ным, при обработке искусствен- ным интеллектом может показать нам индивидуальный рисунок РЭС, связанный с особенностя- ми его инфраструктуры.

Именно отклонения от нормы, дают нам возможность осуществ- лять предиктивную аналитику поведения распределительной сети. А с привязкой к топологии мы получаем прогностику по от-

дельным участкам сети или по отдельным элементам инфра- структуры. Например, возможно выявлять часто повреждаемые участки сети с селекцией по типу аварий и причине возникнове- ния или, скажем, предотвращать ускоренное старение изоляции и повреждение электрооборудо- вания.

Несомненно, потенциал при- менения предиктивной аналити- ки в распределительных сетях огромен, как, к слову, в работе трансформаторов и электродвига- телей. Для того чтобы его реа- лизовать, мы с вами должны со- действовать подготовке данных для разработчиков интеллек- туальных систем, а также вно- сить новые критерии к данным и устройствам, входящим в сис- темы цифровизации. Р

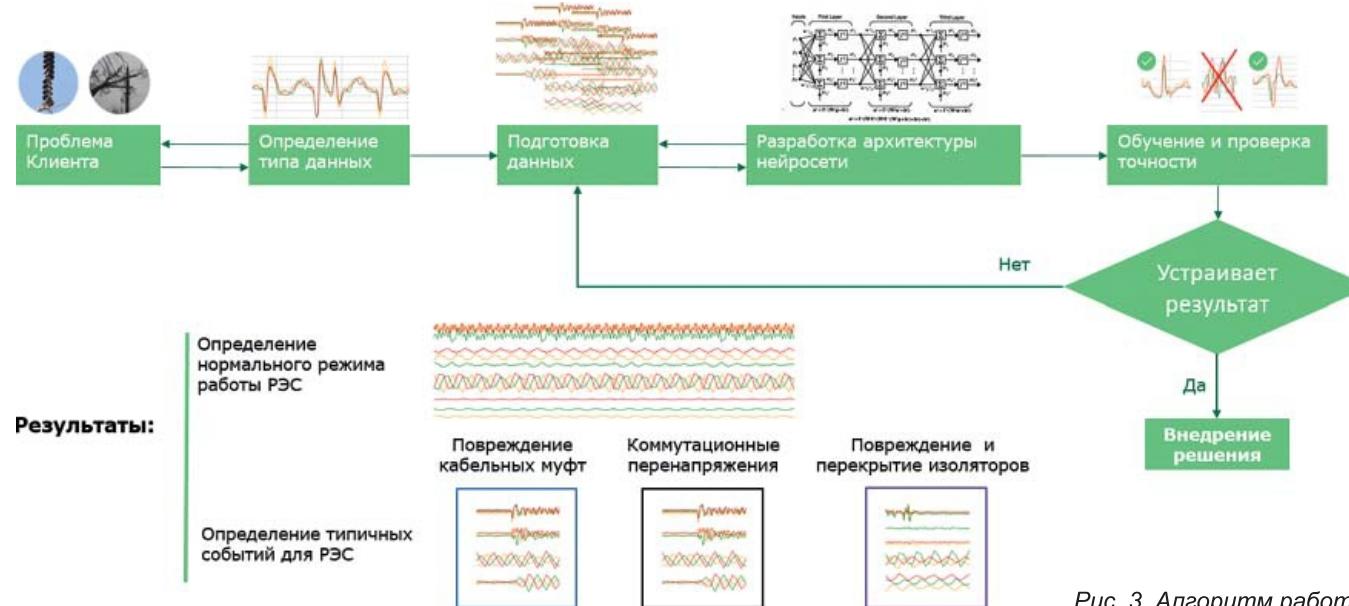
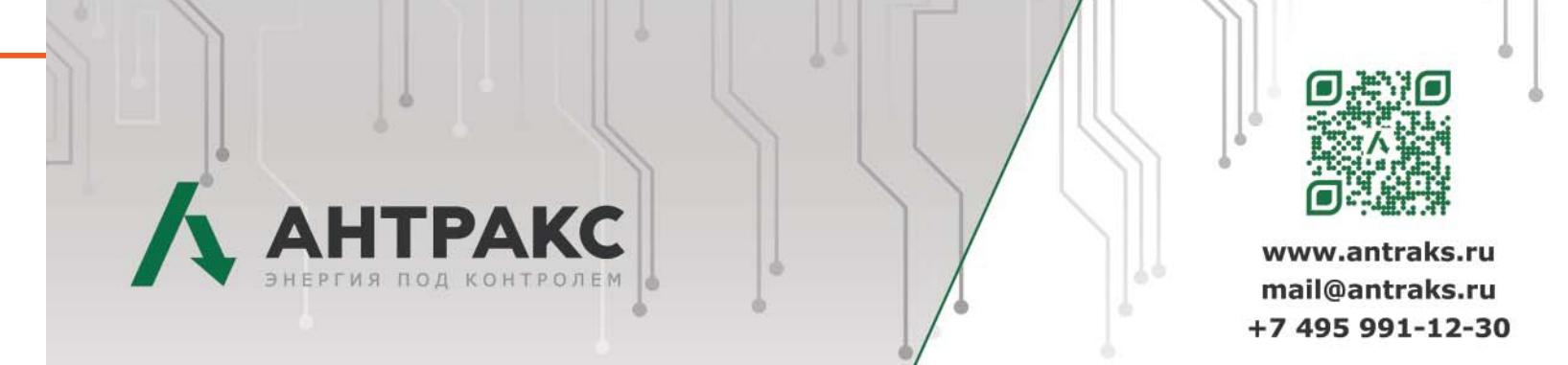


Рис. 3. Алгоритм работы



МНПП «АНТРАКС» – один из ведущих в России разработчиков и производителей оборудования, интеллектуальных систем для контроля и автоматизации электросетевой инфраструктуры.

Комплексные решения компании АНТРАКС позволяют:

- 1 Улучшить надежность и безопасность ваших энергосистем
- 2 Максимально снизить показатели SAIDI и SAIFI на каждый вложенный рубль
- 3 Снизить уровень недоотпуска электроэнергии при аварии
- 4 Сократить продолжительность и объем отключений электроэнергии:
 - селективное определение любых аварийных процессов и локализация места аварии;
 - дистанционное отключение и включение фидеров;
 - отключение поврежденного участка в цикле АПВ
- 5 Распознать и предупредить аварийные отключения на ранней стадии развития
- 6 Продлить срок службы оборудования в сети
- 7 Уменьшить влияние человеческого фактора и сэкономить на выездах ОВБ



ОБЛАСТЬ ПРИМЕНЕНИЯ:

