

ТЕЛЕКОММУНИКАЦИИ И РАДИОТЕХНИКА TELECOMMUNICATION AND RADIO ENGINEERING

Научная статья
УДК 621.396; 004.89
<https://doi.org/10.25686/2306-2819.2022.3.6>

Методика и алгоритм определения периодов оперативного прогнозирования динамики максимально применимых частот КВ-связи на основе алгоритма машинного обучения XGBoost

Н. А. Конкин

Поволжский государственный технологический университет,
Российская Федерация, 424000, Йошкар-Ола, пл. Ленина, 3
konkinna@volgatech.net

Аннотация. Дальняя КВ-связь получает дальнейшее развитие, благодаря расширенному внедрению IT-технологий и методов активной сенсорной диагностики её линий и радиоканалов. Рассмотрена задача применения алгоритма машинного обучения XGBoost для анализа излученности верхней границы полосы прозрачности линии связи. Анализируются вопросы длительности периодов прогнозирования максимально применимой частоты для заданных ошибок в случае применения настройки гиперпараметров и без нее. Для реализации алгоритма и его расширенного применения создан класс программ на языке Python. Верификация разработки выполнена на основе экспериментальных данных, полученных в Поволжском государственном технологическом университете на среднеширотной радиолинии, протяжённостью 2 600 км. Установлена возможность существенного (7 раз) увеличения периода прогноза.

Ключевые слова: КВ-связь; зондирование; ионосфера; максимально применимая частота; машинное обучение; временной ряд; XGBoost; гиперпараметр; гиперопт

Финансирование: работа выполнена при поддержке гранта Российского научного фонда, проект № 22-19-00073

Благодарности: автор выражает искреннюю благодарность доктору физ.-мат. наук, профессору В. А. Иванову, доктору физ.-мат. наук, профессору Н. В. Рябовой за обсуждение работы и ценные замечания.

Введение. КВ-радиосвязь получает дальнейшее развитие, благодаря применению IT-технологий в части машинного обучения и нейронных сетей, а также развитию сенсорной диагностики радиоканалов. Диагностика применяется с целью определения текущего состояния упоря-

доченного по частоте множества каналов связи [1]. Активные сенсоры подразделяются на вертикальные и наклонные. Вертикальные позволяют проводить исследование характеристик радиоканалов для систем связи из области радиусом до 500 км от точки зондирования. На наклонных

© Конкин Н. А., 2022.

Для цитирования: Конкин Н. А. Методика и алгоритм определения периодов оперативного прогнозирования динамики максимально применимых частот КВ-связи на основе алгоритма машинного обучения XGBoost // Вестник Поволжского государственного технологического университета. Сер.: Радиотехнические и инфокоммуникационные системы. 2022. № 3 (55). С. 6-16. DOI: <https://doi.org/10.25686/2306-2819.2022.3.6>

трассах связи параметры каналов определяются состоянием среды в контрольной (средней) точке зондирования. Активные сенсоры представляют собой приёмо-передатчики, расположенные на концах линии связи. Диагностика осуществляется сложными сигналами с их оптимальной обработкой при приёме. Серия таких сенсоров разработана в Поволжском государственном технологическом университете (ПГТУ) [2, 3]. Обработанный сигнал отображается в виде частотных зависимостей задержек для принятых скачковых мод. По данным характеристикам определяются параметры частотного множества каналов и полоса доступных частот связи (полоса прозрачности). Её верхняя граница называется максимально применимой частотой (МПЧ). Известно, что наиболее надёжная связь обеспечивается каналами с рабочими частотами 0,8...0,9 МПЧ [4]. Из-за изменчивости в геофизическом времени среды (ионосферы) полоса прозрачности и её верхняя граница являются случайными динамическими величинами, поэтому в настоящее время одной из актуальных задач является корректный учёт на основе машинного обучения изменчивости полосы прозрачности линии КВ-связи (особенно её верхней границы), определяющей область её допустимых рабочих частот.

На сегодняшний день существует множество моделей и методов прогнозирования временных рядов. Модели можно разделить на две группы: статистические и структурные. Первые являются особым классом математических моделей, в котором исследуются недетерминированные процессы. К данным моделям относятся, например, регрессионные модели, авторегрессионные модели, модели экспоненциального сглаживания и т. д. Вторая группа моделей строится на базе некоторой структуры, принципы функционирования которой формируются при обучении модели. К данным моделям относятся нейронные сети, модели на классификационно-регрессионных деревьях, модель опорных векторов, нечёткой логики и др.

Оба класса моделей могут применяться для решения поставленной задачи использования технологии машинного обучения [5]. Однако с применением технологии машинного обучения XGBoost (eXtreme Gradient Boosting) возникает задача выбора оптимального периода оперативного прогнозирования МПЧ [6]. В составе оптимальных периодов определены те, что имеют минимальную и максимальную продолжительность.

Цель работы – создание алгоритма оценки оптимальных периодов прогнозирования динамики изменения МПЧ ионосферных линий КВ-связи на основе технологии машинного обучения XGBoost и её экспериментальная верификация на среднеширотной трассе протяжённостью 2 600 км.

Подготовка базы экспериментальных данных. Временные ряды МПЧ в общем случае можно рассматривать как нестационарные стохастические процессы [7]. В отличие от детерминированных, стохастические процессы (ряды) имеют случайную составляющую, которая влияет на корреляцию между отсчётами ряда. Организованные в базе данных временные ряды МПЧ для целей машинного обучения представляют формат датасета, который является структурированной версией исходного набора МПЧ и включает в себя временные метки и набор признаков. Вычисление максимальной применимой частоты коротковолновой линии связи выполняется при обработке очищенной от помех частотной зависимости групповой задержки принимаемых мод сигнала (ионограммы) [8]. Этот процесс реализован на сетке частот бинарного файла, получаемых в результате выполнения этапов зондирования и очистки ионограммы. В зависимости от заданного шага частоты при приёме КВ-сигнала определяется количество подряд идущих парциальных каналов справа налево, при этом значение частоты последнего отсчёта приравнивается к МПЧ. Полученные таким образом значения аккумулируются в базе данных.

Машинное обучение (МО) представляет собой инструмент для создания математических моделей, которые используются для описания набора экспериментальных данных. В машинном обучении выделяют задачи классификации и регрессии. Классификация позволяет определять принадлежность вновь поступающей информации к какому-либо набору значений (классу). Прогнозирование МПЧ [9, 10] радиолинии имеет прямую зависимость от времени, поэтому относится к задаче регрессии, в которой зависимой переменной выступает максимально применимая частота, а независимой – время. Обязательной составляющей МО является подготовка набора данных, а именно тренировочной и тестовой выборки, размеры которых могут определяться в процентном соотношении, при котором обучающей выборке отдаётся часть, составляющая более 60 %. В данной работе подготовка выборки произведена по дням, ограничением тестовой составляющей является превышение порога средней абсолютной ошибки прогноза. Для написания программ, реализующих операции машинного обучения и дальнейшую обработку, в работе используется язык программирования Python. Сформированная база данных значений максимальной применимой частоты используется в дальнейшем в качестве дата сета для выполнения обучения и тестирования моделей алгоритма XGBoost.

XGBoost и параметры оптимизации. XGBoost является сложным составным алгоритмом, который использует ансамбли деревьев принятия решений при минимизации функции потерь.

Функция потерь определяет ошибку (отклонение) между теоретически создаваемой моделью и практически полученным набором данных результатов зондирования. Это является важной составляющей машинного обучения и определяет конечную точность модели. Для этих целей существует множество метрик, таких как средняя абсолютная ошибка (mean

absolute error – MAE), среднеквадратическая ошибка (mean square error – MSE), корень из среднеквадратической ошибки (root mean square error – RMSE)), средняя процентная ошибка (mean absolute percentage error – MAPE). Наиболее используемой является метрика MAE, так как трактуемая ею погрешность является более понятной относительно исследуемых величин и менее чувствительной к сильным отклонениям. MSE более чувствительна к сильным отклонениям и не имеет размерность исследуемого ряда, в то время как RMSE позволяет сохранить высокую чувствительность к выбросам и находится в размерности исследуемого ряда. Метрика MAPE является относительной и позволяет сравнивать наборы данных разных единиц измерения. Первым этапом проведенного исследования выполнена предварительная обработка временных рядов с целью интерполяции пропущенных значений и выравнивания аномальных величин, поэтому для оценки потерь метода машинного обучения XGBoost выбрана метрика средней абсолютной ошибки.

Градиентный бустинг (Gradient boosting) последовательно добавляет модели, целью которых является уменьшение величины ошибки предыдущих моделей (деревьев), в общем виде это представлено в выражении (1).

$$obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \omega(f_k), \quad (1)$$

где $l(y_i, \hat{y}_i)$ – метрика оценки отклонений (потерь), y_i – тестовые данные, \hat{y}_i – прогнозные данные, $\omega(f_k)$ – функция регуляризации, f_k – функция, содержащая структуру дерева и её точность.

Так как при бустинге деревья суммируют прогнозы предыдущих деревьев, то прогнозные данные y_i имеют добавочный элемент (оценка точности листа t -го дерева) $f_t(x_i)$, оптимизирующий дерево, таким образом:

$$\begin{aligned}\hat{y}_i^{(0)} &= 0, \\ \hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i), \\ \hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i), \\ &\dots\end{aligned}\quad (2)$$

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i).$$

В итоге целевая функция (1) будет иметь вид:

$$\begin{aligned}obj^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^t \omega(f_i) = \\ &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \omega(f_t) + const.\end{aligned}\quad (3)$$

Далее добавим уравнение функции потерь, рассмотрим вариант с базовой функцией алгоритма XGBoost — среднеквадратичная ошибка (MSE — mean squared error) формула (4), заменим её элемент $l(y_i, \hat{y}_i)$ и запишем в виде полинома второй степени:

$$l(y_i, \hat{y}_i^{(t)}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (4)$$

$$\begin{aligned}obj^{(t)} &= \\ &= \sum_{i=1}^n \left[y_i - (\hat{y}_i^{(t-1)} + f_t(x_i)) \right]^2 + \sum_{i=1}^t \omega(f_i) = \\ &= \sum_{i=1}^n \left[2(\hat{y}_i^{(t-1)} - y_i) f_t(x_i) + f_t(x_i)^2 \right] + \\ &\quad + \omega(f_t) + const.\end{aligned}\quad (5)$$

Далее XGBoost аппроксимирует полученный полином при помощи ряда Тейлора второй степени (градиентный бустинг):

$$\begin{aligned}obj^{(t)} &= \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \right. \\ &\quad \left. + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) + const,\end{aligned}\quad (5)$$

где $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$, $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$.

В итоге после удаления констант и упрощения, получаем следующее уравнение:

$$\begin{aligned}obj^{(t)} &= \\ &= \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t).\end{aligned}\quad (7)$$

Регуляризирующий член уравнения (7) в модели XGBoost описывается выражением:

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \quad (8)$$

где γ и λ — коэффициенты регуляризации, T — количество лепестков дерева.

Таким образом, к основным параметрам настройки модели XGBoost относятся `n_estimators` — число деревьев, `eta` — размер шага обучения, `gamma` — минимальное значение функции потерь для разделения листа на поддеревья, `max_depth` — максимальная глубина дерева, `lambda/alpha` — L1 L2 регуляризация (ограничение веса листа дерева). Кроме того, возможно настраивать выбор столбцов (признаков) для каждого дерева (`colsample_bytree`), выбор столбцов для каждого уровня глубины дерева (`colsample_bylevel`), максимальную глубину дерева (`max_depth`), ветвление дерева (Гессиян функции) (`min_child_weight`), деление обучающей выборки с целью предотвращения переобучения (`subsample`).

Алгоритм определения оптимального прогнозного периода временного хода МПЧ на основе данных наклонного зондирования ионосферы. В работе были созданы методика и алгоритм определения оптимального прогнозного периода временного хода МПЧ на основе данных наклонного зондирования ионосферы. На рис. 1 представлена блок-схема разработанного алгоритма. Исходный датасет представляет собой вычисленные по ионограммам значения максимальной применимой частоты. Далее данные подвергаются первичной обработке, в которую входят операции фильтрации и интерполяции временных ходов МПЧ. На следующем этапе для экономии временных ресурсов использовалась библиотека `hyperopt` [11], позволяющая автоматизировать процесс подбора гиперпараметров модели XGBoost. Данная библиотека позволяет подробно описывать пространство поиска, используя алгоритм случайного поиска, TPE (Tree of Parzen Estimators — Дерево оценок Парцена) и адаптивный TPE. Настройке подвергались такие параметры модели, как: `n_estimators`, `eta`, `gamma`, `max_depth`, `lambda`, `alpha`, `colsample_bytree`, `colsample_bylevel`, `max_depth`, `min_child_weight`, `subsample`. Перед аккумулялированием моделей в базе данных выполняется проверка точности

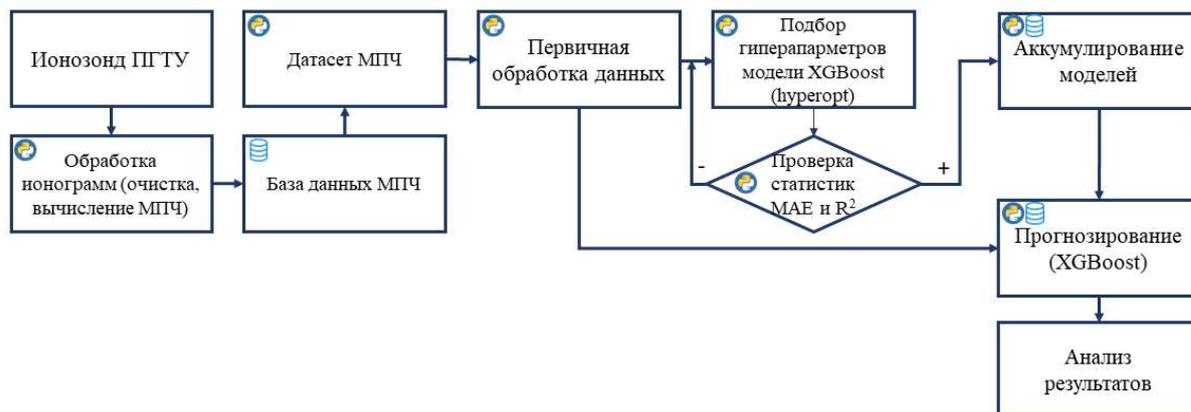


Рис. 1. Алгоритм определения оптимального прогнозного периода временного хода МПЧ на основе данных наклонного зондирования ионосферы

Fig. 1. Algorithm for determining the optimal predictive period of the muf time variation based on oblique sounding of the ionosphere

по значениям средней абсолютной ошибки (MAE) и коэффициента детерминации (R^2). Полученная таким образом выборка моделей представляет собой набор файлов формата .sav, содержащих коэффициенты (параметры) модели машинного обучения. На этапе прогнозирования используются модели XGBoost, аккумулярованные в базе данных. Стоит отметить, что модели ранжируются в зависимости от периода прогноза, таким образом, итерации прогнозного блока зависят от количества прогнозных периодов и моделей для каждого прогнозного периода. В итоге конечной считается модель, результаты которой соответствуют пороговым значениям средней абсолютной ошибки и коэффициента детерминации на этапе прогнозирования. Финальный процесс анализа результатов осуществляет проверку стабильности полученных моделей, т. е. выполняет заданное количество итераций цикла, за который оценивается среднее значение средней абсолютной ошибки и коэффициента детерминации.

Экспериментальная верификация алгоритма. В работе была проведена экспериментальная верификация разработанного алгоритма и определены оптимальные прогнозные периоды временного хода МПЧ на основе данных, полученных в ПГТУ для среднеширотной радиолонии

протяжённостью 2 600 км. Временной ряд состоял из значений МПЧ, полученных за 30 дней каждые 5 минут (всего 8 640 значений временного ряда). Для прогнозирования применялись развитые в работе методика и алгоритм, использующие технологию машинного обучения XGBoost. Учитывалось, что при машинном обучении с учителем обязательной является подготовка тренировочного и тестового набора данных. В проведённом анализе размер тестовой части изменялся от одного дня до дня, когда значение средней абсолютной ошибки превышало пороговое значение (в эксперименте оно составляло 1 МГц, что обусловлено определённой в предыдущих исследованиях мощностью случайной компоненты временного ряда МПЧ), соответственно количество дней в тренировочной выборке варьируется от 29 дней в сторону уменьшения, при условии 30-дневного датасета. Кроме того, проводилось сравнение эффективности алгоритма XGBoost с оптимизацией параметров и без неё.

Для максимального охвата различных комбинаций параметров модели применялись десять подходов с постепенным увеличением количества итераций. Так как при использовании модели XGBoost каждый раз генерируются новые деревья принятия решений, то конечная точность од-

ной и той же модели будет меняться в некоторых пределах, которые зависят от набора параметров и датасета. Вследствие этой особенности на каждой итерации производился контроль точности модели, выполнялось сохранение моделей с наименьшим параметром MAE. Сформированный набор представляет собой базу данных, которая использовалась далее при проведении прогноза. Алгоритм прогноза выполняет перебор всех моделей базы данных и выбирает модель с лучшим значением метрик (коэффициента детерминации модели и средней абсолютной ошибки) [1]. Итоговая зависимость представлена на рис. 2, который описывает изменение статистических метрик при прогнозировании в диапазонах от одного до десяти дней. Количество дней ограничено десятью, так как уже при восьми днях прогноза появляется резкий скачок за пороговые пределы по метрикам оценки точности. К статистическим метрикам относится среднее значение оценок точности (средняя абсолютная ошибка

и среднеквадратичная ошибка) коэффициента детерминации моделей, который определяет степень схожести исходного временного хода с модельным путём оценки дисперсии случайной величины и дисперсии ошибки модели согласно выражению:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \quad (9)$$

где $SS_{res} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ – сумма квадратов остатков регрессии, $y_i - \hat{y}_i$ – значения временного хода МПЧ и спрогнозированной модели. $SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$ – общая сумма квадратов (реального временного хода МПЧ), $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ – среднее значение реального временного ряда.

Данные метрики приведены для случаев с использованием оптимизированных моделей машинного обучения и с использованием моделей без настройки параметров (на рис. 2 приставка original).

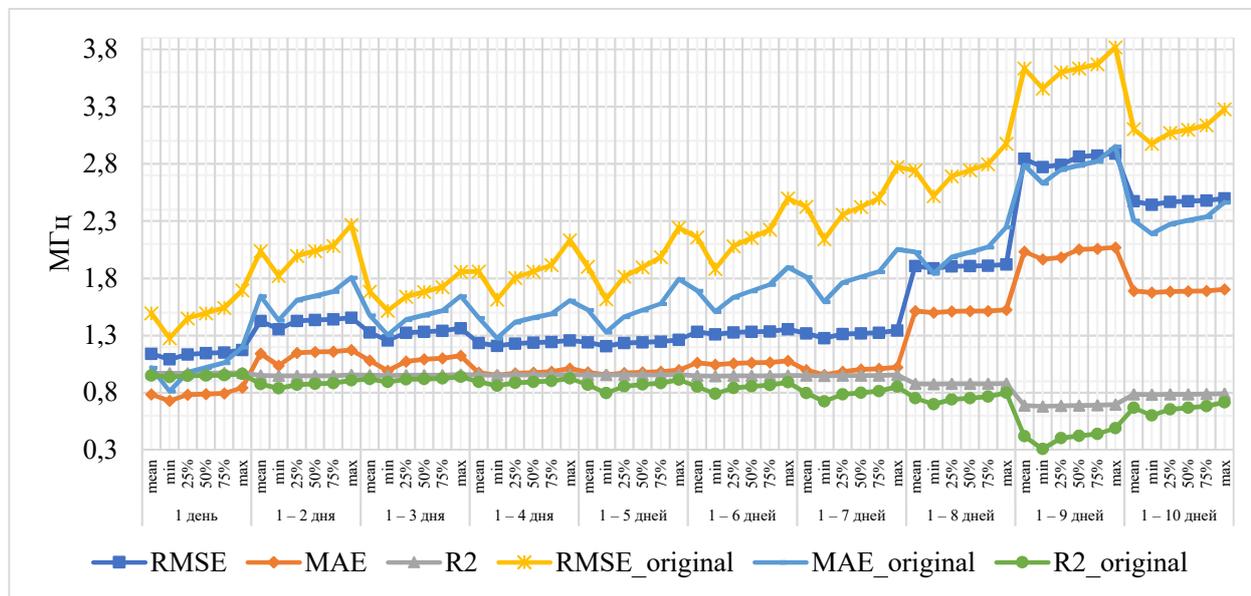


Рис. 2. Результаты прогноза по метрикам средней абсолютной ошибки (MAE), среднеквадратичной ошибки (RMSE) и коэффициента детерминации (R^2) для десяти дней. По оси абсцисс – статистические метрики среднего значения (mean), перцентилей (25, 50, 75 %), максимальных (max) и минимальных (min) значений для каждого прогнозного диапазона (от одного до десяти дней)

Fig. 2. The results of the forecast based on the metrics of mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) for ten days are presented. x-axis plotted statistical metrics of the mean (mean), percentiles (25, 50, 75 %), maximum (max) and minimum (min) values for each forecast range (from 1 to 10 days).

Эксперимент показал, что максимальным является семидневный прогнозный период, в течение которого значение средней абсолютной ошибки находится в заданных пределах (1 МГц) с отклонением в 28 %, коэффициент детерминации составляет 0,95 (95 %) для модели с подгонкой параметров. При использовании алгоритма XGBoost без настройки параметров максимальным является прогноз на один день, MAE составляет 1 МГц, коэффициент детерминации 95 %. Дальнейшее увеличение прогнозного диапазона (8, 9, 10 дней) вызвало резкий скачок по всем метрикам в сторону ухудшения модели, по MAE в среднем на 0,6 МГц, при снижении точности модели на 12 % ниже пороговой отметки в 80 %. Наилучшим для прогнозирования является один день, т. к. минимальное значение MAE может достигать 0,72 МГц, при коэффициенте детерминации 97 % для модели с подгонкой параметров. Кроме того, в среднем значение MAE на 0,17 МГц ниже результатов, соответствую-

ющих максимальному периоду прогноза. В состав оптимальных периодов прогнозирования вошли интервалы от одного до семи дней, за исключением один–два дня, в который значение ошибки превысило 1 МГц. Также стоит отметить, что использование оптимизированной модели XGBoost позволило производить прогноз на интервал времени от одного до семи дней с отклонением по MAE до 28 % от порогового значения 1 МГц, в то время как неоптимизированная модель по метрике MAE в среднем имеет отклонение 43 %.

На рис. 3 представлен результат прогнозирования оптимального периода (1 день) для данного дата сета и набора параметров модели XGBoost, проведено сравнение результатов настроенной модели (XGBoost tuned), модели со стандартными настройками (untuned) и реального временного хода максимальной применимой частоты. Выигрыш оптимизированной модели по метрикам составил 0,25 МГц (MAE) и 2 % (R^2).

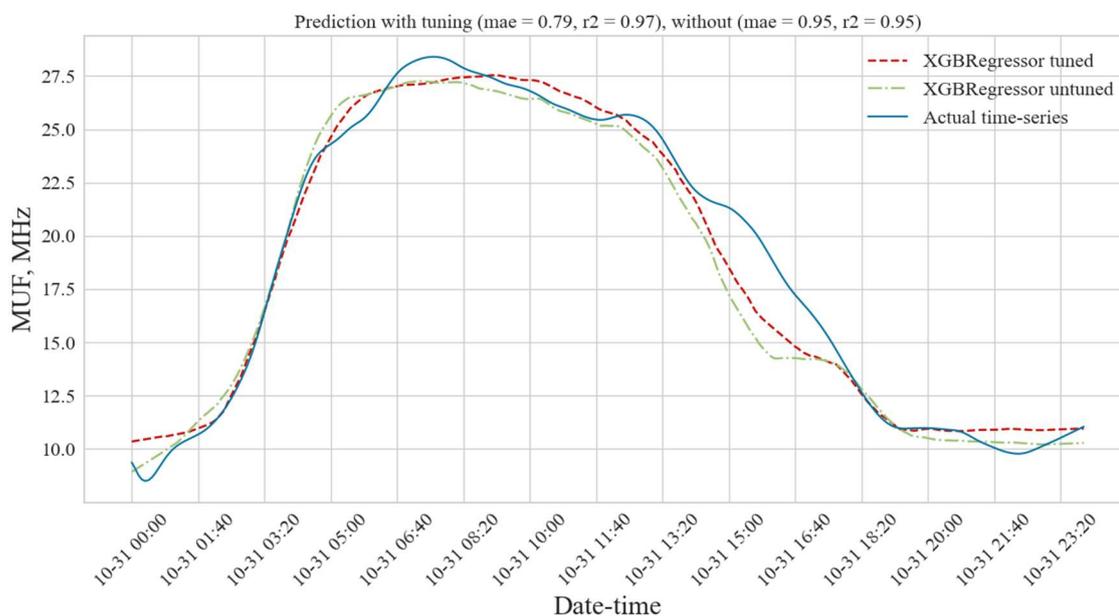


Рис. 3. График зависимости спрогнозированного временного и тестового ходов МПЧ, где XGBRegressor tuned – прогноз с настроенными параметрами, XGBRegressor untuned – прогноз со стандартными параметрами, Actual time-series – реальный ход максимальной применимой частоты со сглаженной случайной компонентой.

Fig. 3. Plot of the predicted MUF time and test cycles, where XGBRegressor tuned is the forecast with tuned parameters, XGBRegressor untuned is the forecast with standard parameters, Actual time-series is the actual behavior of the maximum usable frequency

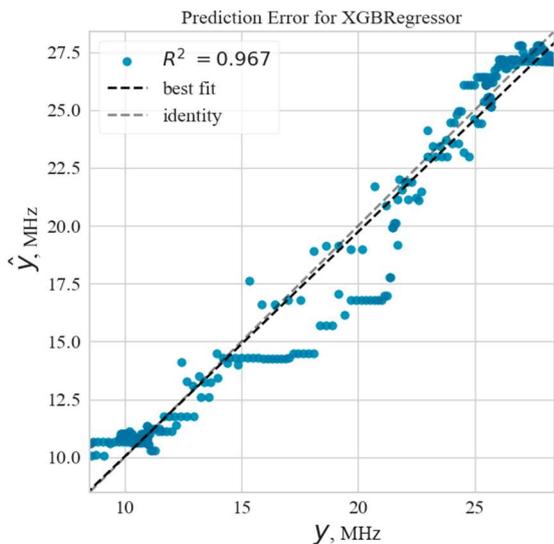


Рис. 4. График зависимости прогнозных значений МПЧ (\hat{y}) от тестовых (график ошибок прогноза)
 Fig. 4. Graph of the dependence of the predicted MUF (\hat{y}) values on the test ones (plot of prediction errors)

Для анализа отклонений полученной модели от реального временного хода МПЧ на рис. 4 представлен график ошибок прогнозирования, который позволяет оценить дисперсию прогноза и выявить наиболее сложные участки прогноза. В ходе анализа определено, что наибольшая зашумлённость наблюдается в диапазоне от 14 до 18 МГц на временном участке с 11:00 до 17:00. Высокие вариации данного диапазона соответствуют периоду постзенитного движения солнца вплоть до заката, что согласуется с нашими

предыдущими исследованиями с помощью других методов прогнозирования и данными исследований, полученных другими учёными.

Заключение. Разработан алгоритм оценки оптимальных периодов прогнозирования динамики МПЧ ионосферных линий КВ-связи на основе технологии машинного обучения XGBoost. Установлено, что настроенные по гиперпараметрам модели для диапазона от одного до семи дней дают лучший результат по метрике средней абсолютной ошибки на 25 % (среднее MAE составило 1,43 МГц), по коэффициенту детерминации на 8 % (среднее значение R^2 составило 0,94). Экспериментально определено, что минимальным прогнозным периодом является диапазон одного дня, при этом значение средней абсолютной ошибки составило 0,78 МГц (в пике до 0,84 МГц), а коэффициент детерминации 0,97 (97 % сходимость прогнозной модели и практического временного хода МПЧ). При этом максимальный прогнозный период составляет семь дней и для него среднее значение MAE составляет 0,99 МГц, а коэффициент детерминации равен 0,95. Для практической реализации разработанного алгоритма создан класс программ на языке программирования Python, предполагающий их расширенное применение для частотного обеспечения систем КВ-связи.

Список источников

1. Кислицын А. А., Рябова Н. В., Конкин Н. А. Метод интеллектуального анализа данных для прогнозирования значений полосы когерентности изменяющегося трансionoсферного радиоканала // Международная Байкальская молодежная научная школа по фундаментальной физике: Труды XVII Конференции молодых ученых, Иркутск, 05–10 сентября 2022 года. Иркутск: ИСЗФ СО РАН, 2022. С. 361-363.
2. Зондирование ионосферных каналов высокочастотной связи с поверхности Земли / В. А. Иванов, Д. В. Иванов, Н. В. Рябова и др. // Вестник Марийского государственного технического университета. Серия: Радиотехнические и инфокоммуникационные системы. 2008. № 1(2). С. 3-20.
3. Новые задачи ионосферной КВ-связи. Развитие методик, аппаратурных решений и экспериментальные результаты их исследования / Д. В. Иванов, В. А. Иванов, Н. В. Рябова и др. // Распространение радиоволн: Труды XXVII Всероссийской открытой научной конференции, Калининград, 28 июня – 03 июля 2021 года. Калининград: Балтийский федеральный университет имени Иммануила Канта, 2021. С. 68-85.
4. Алгоритмы обнаружения полезного сигнала на фоне помех и их верификация для универсального цифрового ионозонда, созданного по SDR-технологии / Д. В. Иванов, В. А. Иванов, Н. В. Рябова и др. // Международная Байкальская молодежная научная школа по фундаментальной физике: Тезисы докладов XV Конференции моло-

дых ученых, Иркутск, 11–16 сентября 2017 года. – Иркутск: Институт солнечно-земной физики Сибирского отделения Российской академии наук, 2017. С. 67.

5. Чучуева И. А. Модель прогнозирования временных рядов по выборке максимального подобию: специальность 05.13.18 "Математическое моделирование, численные методы и комплексы программ": диссертация на соискание ученой степени кандидата технических наук / Чучуева Ирина Александровна. Москва, 2012. 154 с.

6. *Chen T., & Guestrin C.* XGBoost: A Scalable Tree Boosting System // In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM. 2016. Pp. 785–794. <https://doi.org/10.1145/2939672.2939785>

7. *Иванов В. А., Рябова Н. В.* Линейное прогнозирование структурных функций ионосферных радиоканалов с использованием данных зондирования многоэлементным ЛЧМ-сигналом // Вестник Поволжского государственного технологического университета. Серия: Радиотехнические и инфокоммуникационные системы. 2017. № 1(33). С. 26-37. DOI 10.15350/2306-2819.2017.1.26.

8. Технология программно-определяемого радио в задаче последовательного зондирования ионосферных каналов высокочастотной связи / Д. В. Иванов, В. А. Иванов, Н. В. Рябова и др. //

Радиотехника и электроника. 2016. Т. 61. № 7. С. 629–638. DOI 10.7868/S0033849416070068.

9. Моделирование и исследование ионограмм наклонного зондирования ВЧ-радиоканалов для трасс различной протяженности с использованием цифрового ионозонда на платформе USRP / В. В. Овчинников, А. Р. Лашевский, Д. В. Иванов и др. // Системы синхронизации, формирования и обработки сигналов. 2017. Т. 8. № 1. С. 104-108.

10. Экспериментальная оценка ключевых параметров множества парциальных каналов КВ-связи с использованием USRP платформы / Д. В. Иванов, В. А. Иванов, Н. В. Рябова и др. // Международная Байкальская молодежная научная школа по фундаментальной физике. Взаимодействие полей и излучения с веществом: Труды Международной Байкальской молодежной научной школы и XVI Конференции молодых ученых, Иркутск, 16–21 сентября 2019 года. Иркутск: Федеральное государственное бюджетное учреждение науки Ордена Трудового Красного Знамени Институт солнечно-земной физики Сибирского отделения Российской академии наук, 2019. С. 241-244.

11. *Bergstra J., Yamins D., Cox D. D.* (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures // To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013). Atlanta, Georgia, USA, 2013. JMLR: W&CP vol. 28. Pp. 1–9.

Статья поступила в редакцию 15.08.2022; одобрена после рецензирования 05.09.2022; принята к публикации 15.09.2022

Информация об авторе

КОНКИН Никита Александрович – аспирант кафедры радиотехники и связи, Поволжский государственный технологический университет. Область научных интересов – анализ данных дистанционного зондирования ионосферы. Автор десяти 45 научных публикаций. ORCID: 0000-0001-7052-748X

Автор заявляет об отсутствии конфликта интересов.

Автор прочитал и одобрил окончательный вариант рукописи.

Scientific article

UDC 621.396; 004.89

<https://doi.org/10.25686/2306-2819.2022.3.6>

Methodology and Algorithm for Determining the Periods of Operational Forecasting of the Maximum Usable Short Wave Communication Frequencies Based on the XGBoost Machine Learning Algorithm

N. A. Konkin

Volga State University of Technology,
3, Lenin sqr., Yoshkar-Ola, 424000, Russian Federation
konkinna@volgatech.net

Keywords: HF-communication; sounding; ionosphere; maximum usable frequency; machine learning; time series; XGBoost; hyperparameter; hyperopt

ABSTRACT

Introduction. HF radio communication is further developed due to the use of IT technologies in terms of machine learning and neural networks, as well as the development of sensor diagnostics of radio channels. The paper investigates the application of the XGBoost machine learning algorithm to analyze the variability of the upper limit of the transparency band of a HF communication line (MUF). **The goal** is to develop an algorithm for determining the optimal periods of MUF prediction in the case of hyperparameter tuning and without it. Based on the prepared MUF time series, the optimal (minimum and maximum possible) periods of MUF forecasting were determined based on training samples compiled from experimental data of radio sounding of short-wave communication channels obtained using equipment on a mid-latitude radio link with a length of 2600 km. A database of machine learning models for the XGBoost algorithm has been formed, which will avoid the procedure for pre-tuning the prediction algorithm. To assess the accuracy of forecasting, the metrics of the mean absolute error (MAE) and the coefficient of determination were used. It has been established that hyperparameter-tuned models for the range from 1 to 7 days give the best result in terms of the mean absolute error metric by 25% (average MAE was 1.43 MHz), in terms of determination coefficient by 8% (average value was 0.94). It was experimentally determined that the minimum forecast period is the range of one day, while the value of the average absolute error was 0.78 MHz (at the peak up to 0.84 MHz), and the coefficient of determination was 0.97 (97% convergence of the forecast model and the practical time course of the MUF). At the same time, the maximum forecast period is 7 days and for it the average MAE value is 0.99 MHz, and the coefficient of determination is 0.95. For the practical implementation of the developed algorithm, a class of programs was created in the Python programming language, which assumes their extended use for frequency support of HF communication systems.

Funding: this work was supported by the grant № 22-19-00073 from the Russian Science Foundation.

Acknowledgements: the author wishes to sincerely thank doctor of physical and mathematical sciences, professor V. A. Ivanov, doctor of physical and mathematical sciences, professor N. V. Ryabova for their discussion and fruitful comments.

REFERENCES

1. Kislitsin A.A., Ryabova N.V., Konkin N.A. Metod intellektual'nogo analiza dannyh dlya prognozirovaniya znachenij polosy kogerentnosti izmenyayushchegosya transionosfernogo radiokanala [Method of Intelligent Data Analysis for Predicting the Coherence Bandwidth of a Variable Transionospheric Radio Channel]. *Mezhdunarodnaya Bajkal'skaya molodezhnaya nauchnaya shkola po fundamental'noj fizike: Tru-dy XVII Konferencii molodyh uchennyh* [International Baikal Youth Scientific School on Fundamental Physics: Proceedings of the XVII Conference of Young Scientists, Irkutsk, September 05–10, 2022. Irkutsk: Institute of solar-terrestrial physics of siberian branch of russian academy of sciences (ISTP SB RAS), 2022. Pp. 361-363. (In Russ.).
2. Ivanov V.A., Ivanov D.V., Ryabova N.V. et al. Zondirovanie ionosfernnyh kanalov vysokochastotnoj svyazi s poverhnosti Zemli [Ionospheric High-Frequency Communication Channels Sounding from the Earth Surface]. *Vestnik Marijskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Radiotekhnicheskie i infokommunikacionnye sistemy* [Vestnik of Mari State Technical. Ser. «Radio Engineering and Infocommunication Systems». 2008. № 1(2). Pp. 3-20. (In Russ.).

3. Ivanov D. V., Ivanov V. A., Ryabova N. V. et al. Novye zadachi ionosfernoj KV-svyazi. Razvitie metodik, apparatnyh reshenij i eksperimental'nye rezultaty ih issledovaniya [Advances in Ionospheric HF Communication. Development of Methods, Hardware Solutions and Research Findings]. *Rasprostranenie radiovoln: Trudy XXVII Vserossijskoj otkrytoj nauchnoj konferencii* [Propagation of radio waves: Proceedings of the XXVII All-Russian Open Scientific Conference. Kaliningrad, June 28 - 03, 2021]. Kaliningrad: Immanuel Kant Baltic Federal University, 2021. Pp. 68-85. (In Russ.).
4. Ivanov D.V., Ivanov V.A., Ryabova N.V. et al. Algoritmy obnaruzheniya poleznogo signala na fone pomekh i ih verifikaciya dlya universal'nogo cifrovogo ionozonda, sozdannogo po SDR-tekhnologii [The useful signal detection algorithms on the interference background and their verification for sdr based digital ionosonde]. *Mezhdunarodnaya Bajkal'skaya molodezhnaya nauchnaya shkola po fundamental'noj fizike: Tezisy dokladov XV Konferencii molodyh uchenyh, Irkutsk, 11–16 sentyabrya 2017 goda* [Baikal Young Scientists' International School on Fundamental Physics: Abstracts of the XV Conference of Young Scientists, Irkutsk, September 11–16, 2017]. Irkutsk: Institute of Solar-Terrestrial Physics of Siberian Branch of Russian Academy of Sciences (ISTP SB RAS), 2017. Pp. 67. (In Russ.).
5. Chuchueva I. A. Model' prognozirovaniya vremennyh ryadov po vyborke maksimal'nogo podobiya Diss. kand. tech. nauk [Time series forecasting model based on the maximum similarity sample. dissertation for the degree of candidate of technical sciences. Moscow, 2012. 154 p. (In Russ.).
6. Chen T., & Guestrin C. XGBoost: A Scalable Tree Boosting System // In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM. 2016. Pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
7. Ivanov V.A., Ryabova N.V. Linejnoe prognozirovanie strukturnykh funkcij ionosfernykh radiokanalov s ispol'zovaniem dannyh zondi-rovaniya mnogoelementnym LCHM-signalom [Linear Prediction of Structure Functions of Ionospheric Radio Channels Using the Data of the Probing by a Multielement LFM – Signal]. *Vestnik Povolzhskogo gosudarstvennogo tekhnologicheskogo universiteta. Seriya: Radiotekhnicheskie i infokommunikacionnye sistemy* [Vestnik of Volga State University of Technology. Ser. Radio Engineering and Infocommunication Systems. 2017. № 1(33). Pp. 26-37. DOI 10.15350/2306-2819.2017.1.26. (In Russ.).
8. Ivanov D. V., Ivanov V. A., Ryabova N. V. et al. Tekhnologiya programmno-opredelyaemogo radio v zadache posledovatel'nogo zondirovaniya ionosfernykh kanalov vysokochastotnoj svyazi [Software-Defined Radio Technology in the Problem Concerning with the Successive Sounding of HF Ionospheric Communication Channels]. *Radiotekhnika i elektronika* [Journal of Communications Technology and Electronics]. 2016. Vol. 68. № 7. Pp. 629–638. DOI 10.7868/S0033849416070068. (In Russ.).
9. Lashchevskii A. R., Ivanov D. V., Ivanov V. A. et al. Modelirovanie i issledovanie ionogramm naklonnogo zondirovaniya VCH-radiokanalov dlya trass razlichnoj protyazhennosti s ispol'zovaniem cifrovogo ionozonda na platforme USRP [Modeling and study of ionograms of oblique sounding of HF radio channels for paths of various lengths using a digital ionosonde on the USRP platform]. *Sistemy sinhronizacii, formirovaniya i obrabotki signalov* [Systems of Synchronization, Formation and Processing of Signals]. 2017. № 1. Pp. 104-108. (In Russ.).
10. Ivanov D. V., Ivanov V. A., Ryabova N. V. et al. Eksperimental'naya ocenka klyuchevykh parametrov mnozhestva parcial'nykh kanalov KV-svyazi s ispol'zovaniem USRP platformy [Experimental Evaluation of Key Parameters of a Set Of Partial Communication Channels, Using USRP Platform]. *Mezhdunarodnaya Bajkal'skaya molodezhnaya nauchnaya shkola po fundamental'noj fizike: Trudy XVII Konferencii molodyh uchenyh* [Baikal young scientists' international school on fundamental physics. Interaction of fields and radiation with matter: XVI Young Scientists' Conference, Irkutsk, September 16–21], 2019. Irkutsk: Institute of solar-terrestrial physics of siberian branch of russian academy of sciences (istp sb ras), 2019. Pp. 241-244. (In Russ.).
11. Bergstra J., Yamins D., Cox D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures // To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013). Atlanta, Georgia, USA, 2013. JMLR: W&CP vol. 28. Pp. 1–9.

The article was submitted 15.08.2022; approved after reviewing 05.09.2022; accepted for publication 15.09.2022

For citation: Konkin N. A. Methodology and Algorithm for Determining the Periods of Operational Forecasting of the Maximum Usable Short Wave Communication Frequencies Based on the XGboost Machine Learning Algorithm. *Vestnik of Volga State University of Technology. Ser.: Radio Engineering and Infocommunication Systems*. 2022. No 3 (55). Pp. 6–16. DOI: <https://doi.org/10.25686/2306-2819.2022.3.6>

Information about the author

Nikita A. Konkin – PhD student at the Department of Radio Engineering and Communications of Volga State University of Technology. Research interests – the analysis of ionospheric sounding data. The author of 45 scientific publications. ORCID 0000-0001-7052-748X

Author declare that they have no conflict of interest.
Authors read and approved the final manuscript.