

Том I, с. 435–438

УДК: 551.510.413.5 + 550.2(063)

МОДЕЛИРОВАНИЕ КОЛИЧЕСТВА ГЕОМАГНИТНЫХ БУРЬ С ИСПОЛЬЗОВАНИЕМ НЕЛИНЕЙНОЙ ПУАССОНОВСКОЙ РЕГРЕССИИ

А. Э. Пекина, Ю. С. Масленникова, В. В. Бочкарев

Казанский федеральный университет, 420008, г. Казань, ул. Кремлёвская, 18

E-mail: pekina.96@mail.com

Аннотация. В данной работе предложена модель прогнозирования динамики временного ряда количества геомагнитных бурь с использованием нелинейной пуассоновской регрессии. Исследуемый ряд количества геомагнитных бурь был сформирован на основе данных индекса Dst за период с 1964 г. по 2018 г. Для построения авторегрессионной модели была использована искусственная нейронная сеть, обучение которой было адаптировано с учетом статистических особенностей исследуемого ряда. В работе показано, что флуктуации ряда количества геомагнитных бурь близки к закону Пуассона, поэтому для оценки качества прогностической модели была предложена новая мера на основе анализа отношения среднеквадратичных отклонений оцененных параметров модели. В результате работы, построен прогноз числа геомагнитных бурь на 1 неделю вперед, при этом показано, точность прогноза выше на 20% по сравнению со стандартными подходами к обучению нейросетевых прогностических систем. Аналогичный подход может успешно применяться для прогнозирования динамики редких событий в задачах солнечно-земной физики.

Ключевые слова: геомагнитная активность; нелинейная пуассоновская регрессия; искусственные нейронные сети

MODELLING OF MAGNETIC STORMS COUNT DATA USING NONLINEAR POISSON REGRESSION

A. E. Pekina, Y. S. Maslennikova, V. V. Bochkarev

Abstract. The paper proposes a prediction model of magnetic storms counts data using nonlinear Poisson regression. The investigated time series were converted from Dst index data for the time interval from 1964 to 2018. An artificial neural network was used as a basis of the nonlinear autoregressive model, training procedures were adapted using statistical properties of the investigated time series. It is shown that the fluctuations of geomagnetic storms count data are close to the Poisson's distribution, a special measure of the quality was proposed for estimating the prediction efficiency. It bases on the analysis of the standard deviation ratio for the estimated model parameters. The prediction model was tested on the geomagnetic storms counts dynamics. It is shown that the prediction accuracy for the 1-week counts is higher on 20% in comparison with the standard approaches for the training of artificial neural network systems. A similar approach can be successfully applied for the forecasting of a count data dynamics in other applications of the atmospheric and solar-terrestrial physics.

Keywords: geomagnetic activity; nonlinear poisson regression; artificial neural networks

Введение

Геомагнитная активность является важной частью солнечно-земной физики и космической погоды. Основными проявлениями геомагнитной активности являются сильные возмущения — магнитные суббури и магнитные бури, а также слабые возмущения — разнообразные типы магнитных пульсаций [1]. Индекс геомагнитной активности Dst (Disturbance storm-time) является ключевым параметром, который характеризует возмущение геомагнитного поля в магнитных бурях. Он отражает среднечасовое возмущение горизонтальной составляющей напряженности магнитного поля Земли, величина возмущения определяется по четырем низкоширотным обсерваториям, равномерно распределенным по долготе, поэтому на индекс Dst слабо влияют суббури [2]. Таким образом, по индексу Dst можно однозначно сказать о наличии бури, поэтому он был выбран для прогнозирования динамики количества геомагнитных бурь.

Основная проблема прогнозирования индекса Dst состоит в малом количестве событий, соответствующих буре, и большом количестве магнитоспокойных дней. Пути решения данной проблемы предложены в работе [3], где прогнозирование уровня возмущений магнитного поля осуществляется с использованием дополнительных факторов таких как, например, солнечный ветер. Однако подобный подход не учитывает характер флуктуаций количества геомагнитных бурь от времени, который близок к распределению ряда редких событий. Поэтому для прогнозирования таких рядов эффективнее использовать подходы на основе пуассоновской авторегрессии. В настоящее время появился ряд моделей нелинейной пуассоновской авторегрессии на основе искусственных нейронных сетей, которые применялись для прогнозирования динамики редких событий [4, 5]. Целью настоящей работы являлась разработка нейросетевой модели прогнозирования динамики временного ряда количества геомагнитных бурь с использованием нелинейной пуассоновской регрессии. Для достижения поставленной цели была решена задача улучшения методов обучения нейронной сети с учетом статистических особенностей ряда, предложена оптимальная мера оценки качества прогноза, а также проведено сравнение результатов прогнозирования со стандартными подходами.

Предварительный анализ исследуемого ряда

Почасовой индекс Dst рассчитывается в Мировом центре данных по геомагнетизму и управляется при Центре анализа данных по геомагнетизму и космическому магнетизму при Киотском университете (Япония). Данные индекса были получены с интернет-ресурса [6] за период с 1964 по 2018 год через каждый час. Данные индекса представляют собой значения в диапазоне от +100 нТл до -600 нТл. Показано [2], что в магнитоспокойные дни значения индекса не превышают диапазона ± 20 нТл, а во время магнитных бурь они могут иметь большие отрицательные значения. Исходный ряд значений за выбранный временной период содержит 173400 отсчетов.

Предварительная обработка данных происходит в несколько этапов. Для выявления значимых геомагнитных бурь было выбрано пороговое значение, равное -60 нТл. Затем из дискретного ряда значений индекса был построен бинарный ряд с соответствующим порогом: значение заменялось на 0, если индекс Dst превышал порог, и на 1 – в противоположном случае. Все последовательности подряд идущих единичных значений принимались за бурю. Далее в интервале времени размером 7 дней вычислялось количество найденных бурь. Результирующий ряд представлен на рис. 1, он использовался для построения модели прогноза.

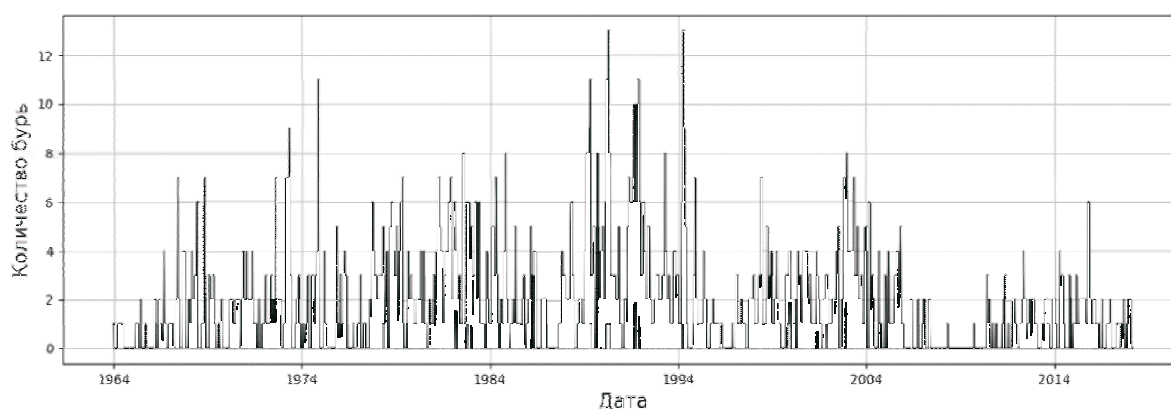


Рис. 1. Ряд количества бурь за неделю, сформированный на основе индекса Dst за период с 1964 г. по 2018 г.

Моделирование и результаты прогноза

Анализируемый ряд количества бурь, рассчитанный на основе индекса Dst, имеет размах значений в пределах 10, следовательно является хорошим примером временного ряда числа

редких событий. Поэтому для прогнозирования его динамики была предложена модель пуассоновской авторегрессии. Для получения наиболее точного значения параметра распределения λ предлагается использовать метод максимального правдоподобия. Более подробно данный метод с применением нейронных сетей описан в статье [5].

Для получения прогноза ряда количества бурь обучение нейронных сетей было адаптировано, помимо стандартной функции ошибки на основе минимизации средней квадратичной ошибки (mean squared error, далее MSE) было также реализовано обучение по методу максимального правдоподобия (maximum likelihood estimation, далее MLE). Модель прогнозирования была выполнена с использованием нейронной сети с архитектурой 14-10-5-1-1 (цифры соответствуют количеству нейронов в каждом слое сети). Обучение происходило с использованием оптимизатора Adadelta [7]. Результат прогнозирования ряда количества бурь на одну неделю вперед (на 1 шаг) представлен на рис. 2.

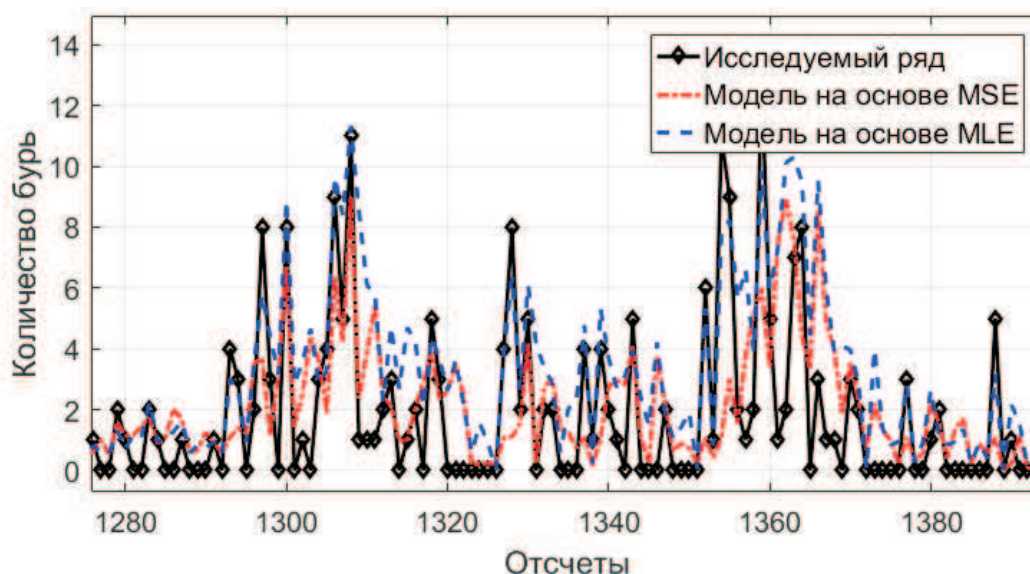


Рис. 2. Результат прогнозирования ряда количества бурь на одну неделю вперед (чёрный – исходный ряд, прерывистые цветные линии - ряды спрогнозированных значений двумя моделями)

Оценить точность полученных параметров с использованием стандартных подходов (например, по значениям среднеквадратичного отклонения) не представляется возможным. Поэтому для оценки эффективности работы модели был разработан метод сравнения двух временных рядов с учетом того, что ряды предположительно имеют распределение Пуассона. Основная идея алгоритма состоит в оценке условного разброса значений ряда, иными словами, в оценке дисперсии ряда при условии спрогнозированного значения параметра распределения: $D[x|\hat{\lambda}]$. Теоретическое СКО исходного ряда должно быть $\sim \sqrt{\lambda}$. Тогда оцениваемое условное СКО $\sim \sqrt{\hat{\lambda}}$.

Результаты оценки предложенной метрики изображены на рис. 3 (слева – результат прогнозирования модели с обучением по минимизации MSE, справа – результат модели с обучением по максимизации MLE). Черная прерывистая кривая – это целевая функция $\sim \sqrt{\lambda}$. Сплошная цветная кривая – это $A\sqrt{\hat{\lambda}}$, где A – коэффициент пропорциональности, равный $A = \text{mean}(\sqrt{\lambda}/\sqrt{\hat{\lambda}})$. При распределении ряда, соответствующего закону Пуассона, корректный прогноз будет иметь коэффициент пропорциональности, близкий к 1. Маркерами нанесены значения оценок $\sqrt{\hat{\lambda}_i}$. Значения должны находиться в допустимом диапазоне $A\sqrt{\hat{\lambda}_i} \pm (A\sqrt{\hat{\lambda}_i}/\sqrt{n_i})$, где n_i – количество значений в исследуемом ряду, близких к λ_i .

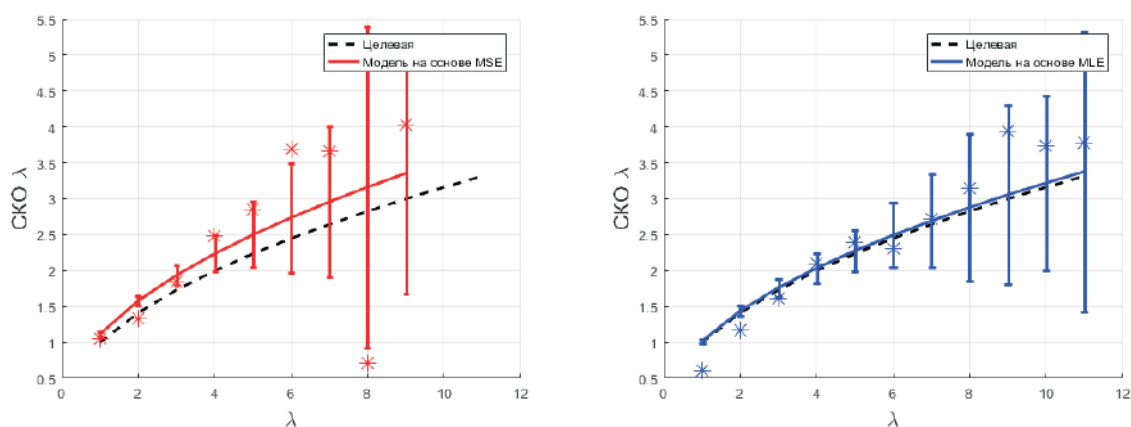


Рис. 3. Мера оценки качества для двух рядов, для модели с обучением по минимизации MSE (слева) и модели с обучением по максимизации MLE (справа)

Коэффициент пропорциональности при сравнении модельного ряда, построенного с функцией ошибки ММП для Пуассоновского распределения, с исходным рядом – 1.01, а для модельного ряда, построенного с МНК - 1.25.

Заключение

Предложена модель прогнозирования ряда количества геомагнитных бурь на основе нелинейной пуассоновской регрессии, предложена мера оценки качества прогноза для ряда с законом распределения флуктуаций, близким к Пуассону. Предложенный подход показал точность прогноза на 20% выше, по сравнению со стандартными. Стоит также отметить, что подобный подход может быть успешно использован для прогнозирования динамики редких событий и в других задачах солнечно-земной физики.

Список литературы

1. Авакян С. В., Воронин Н. А., Дубаренко К. А. Влияние магнитных бурь на аварийность систем электроэнергетики, автоматики и связи // Научно-технические ведомости СПбГПУ. 2012. № 3-2 (154).
2. Sugiura M. Hourly values of equatorial Dst for the IGY, Ann. Int. Geophys. Year, 35, 9-45, Pergamon Press, Oxford, 1964.
3. O.V. Khabarova. Current Problems of Magnetic Storm Prediction and Possible Ways of Their Solving. Sun and Geosphere, 2007; 2(1): 32-37.
4. Fallah N., Nourijelyani K., Eshraghian M. R., Gu H., Seyyedsalehi S. A. Nonlinear Poisson regression using neural networks: a simulation study // Neural Comput & Applic (2009) 18:939–943
5. Pekina A, Maslennikova Y, Bochkarev V. Probability analysis of the vocabulary size dynamics using google books ngram corpus// CEUR Workshop Proceedings. - 2018. - Vol.2268, Is. - P.202-207
6. OMNI2 database. National Space Science Data Center / Space Physics Data Facility. – Access mode: <https://omniweb.gsfc.nasa.gov/form/dx1.html>.
7. Николенко С.И., Кадури А. А., Архангельская Е. О. Глубокое обучение. Погружение в мир нейронных сетей. – Издательство "Питер", 2018. – 480 с.