

Метод комплексного анализа природных данных с применением искусственного интеллекта

В.В. Геппенер
Санкт-Петербургский
государственный

электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
geppener@mail.ru

Б.С. Мандрикова
Институт космофизических
исследований и распространения
радиоволн ДВО РАН
Паратунка, Камчатский край, Россия
555bs5@mail.ru

Н.Г. Пятайчук
Санкт-Петербургский
государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Санкт-Петербург, Россия
pet.ai.4.uk@yandex.ru

Аннотация—Предложен метод комплексного анализа сложно структурированных природных данных, основанный на комбинации нейронных сетей с конструкциями вейвлет-преобразования. На примере данных вторичных космических лучей, подтверждена эффективность метода для детектирования и оценки параметров аномальных вариаций. Эмпирически доказано, что совмещение нейронных сетей Автокодировщик и пакетных вейвлет-разложений позволяет извлекать сложные зависимости в данных и подавлять шум.

Ключевые слова— данные сложной структуры, нейронные сети, вейвлет-преобразование, космические лучи.

1. ВВЕДЕНИЕ

Со стремительным развитием техносферы, основанном на прогрессе науки и технологий, все сильнее ощущается неудовлетворительность существующих методов обработки и анализа сложно структурированных данных. Особая сложность возникает при исследовании природных данных ввиду отсутствия знаний об информативной составляющей сигнала и, как следствие, невозможности эффективного вычищения шума различной природы. Примером таких исследований являются факторы в данных геофизического мониторинга при выполнении прогноза космической погоды. Аномальные события в околоземном космическом пространстве способны создать радиационную опасность для пилотов высотной авиации и космонавтов, сбой радиосвязи, а также могут привести к потере ориентации и разрушению спутников [1]. Космические лучи являются одним из значимых факторов космической погоды. В настоящее время исследованием космических лучей занимается большое число ученых и научных групп, однако вопрос оперативного обнаружения аномальных вариаций с требуемой точностью остается открытым [1].

В докладе представлен метод комплексного анализа данных вторичных космических лучей и детектирования аномальных вариаций на основе применения нейронной сети (НС) Автокодировщик, пакетных вейвлет-разложений и непрерывного вейвлет-преобразования. Сеть Автокодировщик эффективна для подавления шума и извлечения зависимостей в сложных данных [2]. Вейвлет-преобразование позволяет детально изучить частотно-временную структуру сигнала, выделить информативные компоненты и подавить шум [3]. Данная работа является продолжением исследования [4]. Для снижения риска наступления ложной тревоги предложено правило, дающее результат по совокупности решений.

2. ОПИСАНИЕ МЕТОДА

Предлагаемый метод комплексного анализа сложно структурированных данных включает следующие операции:

1. Преобразование данных КЛ на основе Автокодировщика:

$$\hat{f} = h^{(2)}(\omega^{(2)}(h^{(1)}(\omega^{(1)}F + b^{(1)})) + b^{(2)}), \quad (1)$$

где $h^{(1)}$ передаточная функция энкодера, $\omega^{(1)}$, $\omega^{(2)}$ матрицы весов, $h^{(2)}$ – передаточная функция декодера, $b^{(1)}$, $b^{(2)}$ – векторы смещения.

2. Применение непрерывного вейвлет-преобразования: $W\hat{f}(u, s) = \int_{-\infty}^{+\infty} \hat{f}(t) \frac{1}{\sqrt{s}} \Psi^*\left(\frac{t-u}{s}\right) dt$, где Ψ – вейвлет, s – масштаб, u – сдвиг по времени, $s, u \in R$ (R – действительные числа) $s \neq 0$.

3. Применение пороговой функции:

$$P_{T_s^l, NN} [W\hat{f}(u, s)] = \begin{cases} W\hat{f}(u, s), & \text{если } |W\hat{f}(u, s)| \geq T_s^l, \\ 0, & \text{если } |W\hat{f}(u, s)| < T_s^l, \end{cases} \quad (2)$$

где $T_s^l = q \times \sigma_s^l$, σ_s^l – среднеквадратическое отклонение (СКО) коэффициентов, рассчитанное в скользящем окне длины l , q – пороговый коэффициент.

4. Оценка интенсивности аномалий в момент времени $t=u$:

$$E(u)_{NN} = \sum_s P_{T_s^l, NN} [W\hat{f}(u, s)]. \quad (3)$$

5. Выполняем разложение функции f в вейвлет-пакеты: $W_j^0 = \bigoplus_{i=0}^l W_{j_i}^{P_i}$, $\{\Psi_{j_i}^{P_i}(2^{j_i}t - m)\}_{m \in N}$ есть базис пространства $W_{j_i}^{P_i}$;

6. Определяем ветви дерева, соответствующие компонентам сигнала, базис $O_{j_i}^p$ пространства $W_{j_i}^{P_i}$ есть базис:

$$O_{j_i}^p = \begin{cases} \{\Psi_{j_i}^{P_i}(2^{j_i}t - m)\}_{m \in Z}, \\ \sum_{m \in I_M^p} |\langle X, \Psi_{j_i, m}^p \rangle|^2 \geq \sum_{m \in I_M^{2p}} |\langle X, \Psi_{j_i+1, m}^{2p} \rangle|^2 + \sum_{m \in I_M^{2p+1}} |\langle X, \Psi_{j_i+1, m}^{2p+1} \rangle|^2, \\ \{\Psi_{j_i}^{2p}\}_{m \in Z} \cup \{\Psi_{j_i+1}^{2p+1}\}_{m \in Z}, \\ \sum_{m \in I_M^p} |\langle X, \Psi_{j_i, m}^p \rangle|^2 < \sum_{m \in I_M^{2p}} |\langle X, \Psi_{j_i+1, m}^{2p} \rangle|^2 + \sum_{m \in I_M^{2p+1}} |\langle X, \Psi_{j_i+1, m}^{2p+1} \rangle|^2, \end{cases}$$

где множество индексов I_M^l , $l = P, 2P, 2P+1$ определяются так: индекс $m \in I_M^l$, если $|\langle X, \Psi_{j_i, m}^l \rangle| \geq T_{j_i}$ порог $T_{j_i} = K * \sigma_{j_i}^l$,

$$\sigma_{j_i}^l = \sqrt{\frac{1}{l} \sum_{m=1}^l \left(\langle X, \Psi_{j_i, m}^l \rangle - \langle X, \Psi_{j_i, m}^l \rangle \right)^2}.$$

Получаем представление сигнала в виде:

$$\tilde{f}(t) = \sum_{(j_i, m) \in Q_M^l} \langle X, \Psi_{j_i, m}^l \rangle \Psi_{j_i, m}^l(t) = \sum_{(j_i, m) \in Q_M^l} v_{j_i, m} \Psi_{j_i, m}^l(t), \quad (4)$$

где $(j_i, m) \in Q_M^l : j_i \in O_M^p, m \in I_M^l$.

7. Применение непрерывного вейвлет-преобразования:

$$W\hat{f}(u, s) = \int_{-\infty}^{+\infty} \hat{f}(t) \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t-u}{s} \right) dt, \quad (4)$$

где Ψ – вейвлет, s – масштаб, u – сдвиг по времени, $s, u \in R$ (R – действительные числа) $s \neq 0$.

8. Применение пороговой функции:

$$P_{T_s^l}^{pac} [W\hat{f}(u, s)] = \begin{cases} W\hat{f}(u, s), & \text{если } |W\hat{f}(u, s)| \geq T_s^l, \\ 0, & \text{если } |W\hat{f}(u, s)| < T_s^l, \end{cases} \quad (5)$$

где $T_s^l = q \times \sigma_s^l$, σ_s^l – средноквадратическое отклонение (СКО) коэффициентов, рассчитанное в скользящем окне длины l , q – пороговый коэффициент.

9. Оценка интенсивности аномалий в момент времени $t=u$:

$$E(u)_{pac} = \sum_s P_{T_s^l}^{pac} [W\hat{f}(u, s)]. \quad (6)$$

10. Аномалия в данных есть, если выполняются условия: $E(u)_{NN} > T_{аном}$ и $E(u)_{pac} > T_{аном}$ при $T_{аном} = 1.5$.

На Рис. 1 представлен результат применения предлагаемого метода к минутным данным нейтронного монитора ст. Оулу. Оранжевыми вертикальными линиями отмечены моменты регистрации геомагнитных бурь. На Рис. 1б,в представлен результат применения операций (2) и (3), соответственно, к исходным данным нейтронного монитора. На Рис. 1д,е представлен результат применения операций (2) и (3) к данным нейтронного монитора после применения нейронной сети. На Рис. 1з,и представлен результат операций (5) и (6) к данным нейтронного монитора после вейвлет-фильтрации по формуле (4). Результаты показывают, что предобработка данных на основе нейронной сети (операция (1)) (Рис. 1з,д) и выделение когерентных структур с использованием вейвлет-пакетов (операция (4)) (Рис. 1з,и) позволяют существенно снизить уровень шума и выделить аномальные вариации – аномальные понижения и повышения интенсивности космических лучей.

Результат показывает эффективность предлагаемого метода для анализа вторичных космических лучей и выделения аномальных событий в космических лучах в период магнитных бурь.

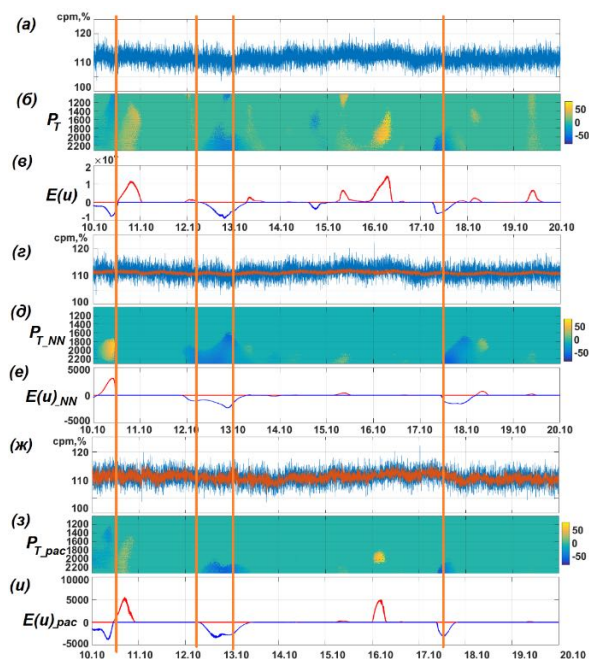


Рис. 1. Результат применения метода к данным НМ ст. Оулу за 10-20 октября 2021 г.: а) данные НМ; б) результат оп. (2) к данным НМ; в) результат оп. (3) к данным НМ; г) синим цветом изображены данные НМ, оранжевым – результат оп. (1); д) результат оп. (3) к данным НМ после аппроксимации НС; е) результат оп. (3) к данным НМ после аппроксимации НС; ж) синим цветом изображены данные НМ, оранжевым – результат оп. (4); з) результат оп. (5) к данным НМ после вейвлет-фильтрации по формуле (4); и) результат оп. (6) к данным НМ после вейвлет-фильтрации

3. ЗАКЛЮЧЕНИЕ

Предложенный метод комплексного анализа сложно структурированных природных данных показал свою эффективность для выделения аномальных особенностей различной интенсивности и продолжительности. Апробация метода с использованием данных вторичных космических лучей подтвердила его результативность для детектирования аномальных событий в сигнале в периоды магнитных бурь. Результаты показали, что совмещение нейронной сети Автокодировщик с пакетными вейвлет-разложениями позволяет извлекать сложные зависимости в данных и подавлять шум.

БЛАГОДАРНОСТИ

Работа выполнена в рамках Государственного задания по теме “Физические процессы в системе ближнего космоса и геосфер при солнечных и литосферных воздействиях” (2021–2023 гг.), регистрационный номер АААА-А21-121011290003-0.

ЛИТЕРАТУРА

- [1] Кузнецов, В.Д. Космическая погода и риски космической деятельности / В.Д. Кузнецов, // Космическая техника и технологии. – 2014. – № 3(6). – С. 3-13.
- [2] Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. – «ДМК Пресс», 2018. – 652 с.
- [3] Малла, С. Вейвлеты в обработке сигналов / С. Малла. – «Мир», 2005. – 671 с.
- [4] Geppener, V. Combination of wavelet transform and Autoencoder for complex data analysis and anomaly detection / V. Geppener, B. Mandrikova // International Conference on Information Technology and Nanotechnology (ITNT). – 2021. – P. 1-4. DOI: 10.1109/ITNT52450.2021.9649436.