

ИСПОЛЬЗОВАНИЕ УПРАВЛЯЕМОГО ДАННЫМИ ПОДХОДА ДЛЯ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ СИГНАЛОВ, ПРИНИМАЕМЫХ СРЕДНЕШИРОТНЫМИ РАДАРАМИ КОГЕРЕНТНОГО РАССЕЯНИЯ

О.И. Бернгардт

ИСЗФ СО РАН, 664033, г. Иркутск, ул. Лермонтова 126А

E-mail: bering@iszf.irk.ru

Аннотация. В докладе представлено развитие самосогласованного, управляемого данными подхода к классификации данных, получаемых на среднеширотных радарах когерентного рассеяния ИСЗФ СО РАН. Алгоритм самостоятельно размечает данные, определяет оптимальное количество классов сигналов, наблюдаемых радарами, и обучает двухслойную классифицирующую нейронную сеть. При траекторных расчетах используется метод геометрической оптики и международные ссылочные модели ионосферы и магнитного поля Земли. Модель обучена на сигналах, приходящих с главного лепестка диаграммы направленности радаров. При обучении, для адаптации части данных, полученных с повышенным спектральным разрешением, производится их искусственное загрубление до стандартного разрешения. Показано, что количество различных классов в данных составляет до 35. Показано, что наиболее важными для классификации параметрами являются расчетные высота рассеяния и наклон траектории в точке рассеяния, а наименее важными - спектральная ширина принятого сигнала и расчетное количество отражений от нижележащей поверхности, что подтверждает важность траекторных расчетов для точного решения задачи классификации.

Ключевые слова: декаметровый радар; СЕКИРА; ионосфера; автоматическая классификация

USING A DATA-DRIVEN APPROACH FOR AUTOMATIC CLASSIFICATION OF SIGNALS RECEIVED BY MID-LATITUDE COHERENT SCATTER RADARS

О.И. Berngardt

Abstract. The paper presents the a self-consistent, data-driven approach to classifying data obtained at the ISTP SB RAS mid-latitude coherent scatter radars. The algorithm labels the data by itself, determines the optimal number of signal classes observed by the radars, and trains a two-layer classifying neural network. The trajectory calculations use the geometric optics method and international reference models of the ionosphere and the Earth's magnetic field. The model is trained on signals coming from the main lobe of the radar antenna pattern. During training, to adapt part of the data obtained with improved spectral resolution, they are artificially decreased to a standard resolution. It is shown that the number of different classes in the data is up to 35. It is shown that the most important parameters for classification are the calculated scattering height and the trajectory inclination at the scattering point. The least important are the spectral width of the received signal and the calculated number of reflections from the underlying surface, which confirms the importance of trajectory calculations for an accurate solution of the classification problem.

Keywords: decameter radar; SECIRA; ionosphere; automatic classification

Введение

Российская сеть когерентных радаров СЕКИРА [1] состоит из радаров, близких радарам международной сети SuperDARN [2] по программному и аппаратному обеспечению и включает в себя радары EKB и MAGW ИСЗФ СО РАН. Основными типами рассеянных сигналов являются: ионосферное рассеяние от магниторентированых неоднородностей, рассеяние от подстилающей поверхности (земной и морской), рассеяние на метеорных следах, и другие [2]. Предлагаемая работа является улучшением метода автоматической классификации данных [3], уменьшающим количество начальных приближений.

Модель распространения сигнала и входные параметры модели

Предложенный в работе [3] подход заключается в использовании неразмеченных данных для создания их классификатора. Получаемая при этом нейронная сеть представляет собой схему похожую на автоэнкодер [4], но с множеством голов-декодеров, где каждая голова (декодер) тренируется отдельно разметкой, созданной неким кластеризатором по отдельному

эксперименту. Считается [4], что автоэнкодер позволяет выделить скрытые переменные в данных, необходимые для описания этих данных.

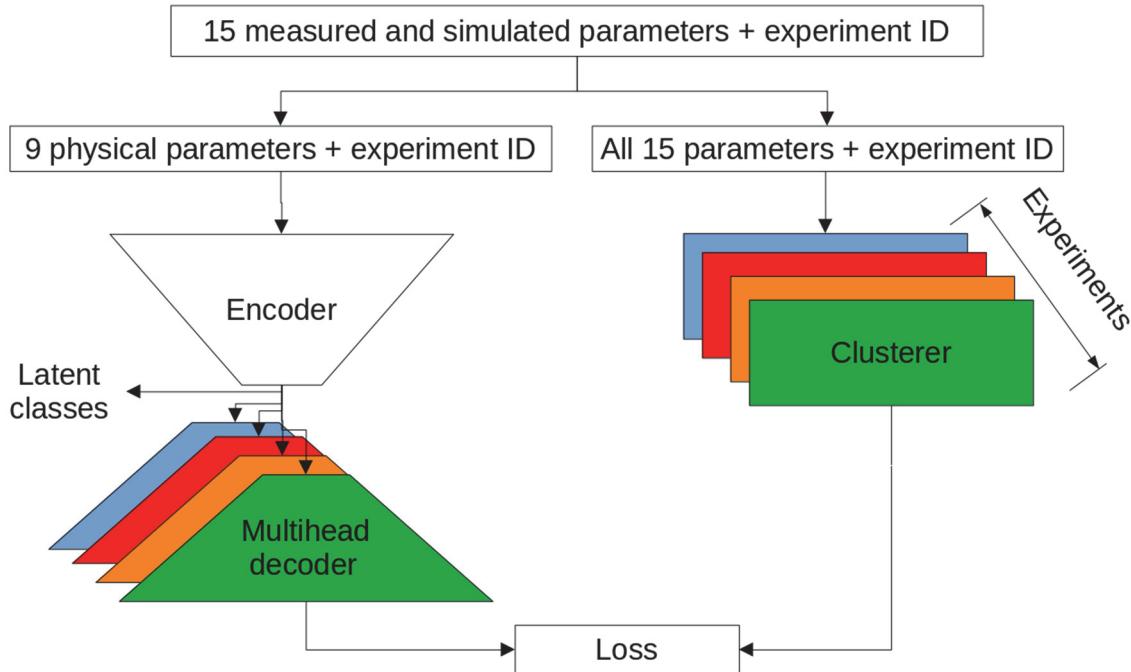


Рис. 1. Схема обучения классификатора (Encoder). Каждый кластеризатор и декодер ('голова') соответствуют одному эксперименту (фиксированный луч, частотный канал и день). Количество голов декодера и кластеризаторов равно числу экспериментов и составляет порядка 15 тыс. Разные цвета соответствуют различным экспериментам

Подход состоит из двух этапов. Первый этап (кластеризатор) – это разделение данных на слабо пересекающиеся классы. На этом этапе используется метод смеси гауссовых функций (GM), при котором предполагается, что данные в каждом кластере подчиняются многомерному гауссовому распределению с неизвестными параметрами, а их число определяется из минимума байесовского информационного критерия (BIC).

Второй этап (классификатор) алгоритма [3] – обучение классификатора на данных, размеченных на первом этапе, и оптимизация нейронной сети по ширине (числу нейронов) методом [5].

При построении классификатора использовались следующие измеряемые и модельные данные. Измеряемые радаром параметры: время, дальность до рассеивателя; доплеровская скорость V , измеренная спектральная ширина W – определяются по сигналу алгоритмом FITACF [6]; угол места – определяется после угломестной калибровки радара по метеорам. Параметры, полученные в результате моделирования распространения радиоволны методом геометрической оптики в модельной ионосфере, описываемой международными моделями IRI и IGRF: высота рассеяния; эффективная высота рассеяния; угол наклона (синус угла места) траектории по отношению к горизонту в 4x точках по дальности (1/4,2/4,3/4,4/4 измеренной дальности); угол между направлением распространения радиоволны и магнитным полем Земли (косинус угла); мода распространения – количество отражений от нижележащего слоя (ионосферного или земной поверхности) во время распространения до рассеивателя; высота рассеяния.

В качестве входных данных классификатора использовались только 9 из этих параметров, связанных с высотой рассеяния, траекторией радиоволны и спектральными характеристиками неоднородностей. Для кластеризации использовались все перечисленные параметры.

Для построения единой модели, не зависящей от характеристик радара, при подготовке данных необходимо компенсировать различия их характеристик и режимов. Учет формы зондирующего сигнала в работе компенсируется искажением (аугментацией) данных о спектральной ширине в зависимости от типа используемого сигнала. Параметры искажения

были подобраны для обеспечения удовлетворительного совпадения спектральных ширин экспериментально измеряемых 7-импульсной последовательностью и аугментированных данных, получаемых при измерениях 16-импульсной последовательностью. Учет расстояния между антенными решетками радара компенсируется отсевом 'плохих' сигналов, приходящих не с главного лепестка диаграммы направленности [7].

Выбор архитектуры новой сети-классификатора основывался на трех принципах: для аппроксимации непрерывных функций достаточно широкой двухслойной сети [8]; в качестве функций активации нейронной сети желательно использовать вычисление абсолютных значений для сохранения преемственности с широкоизвестными алгоритмами [9]; в качестве преобразования выходов сети к вероятностям возможно использовать нормировку неотрицательных величин на их сумму.

Все это позволило решить задачу меньшим числом входных параметров, и обеспечить намного лучшее качество прогноза кластеров по сравнению с предыдущими моделями [3].

Модель была обучена на данных радаров ЕКВ и MAGW за 2021 год. Итоговая нейронная сеть имеет 49 нейронов в первом слое и 35 нейронов во втором, что означает наличие в данных 35 различных классов, из которых с радиофизической точки зрения удалось уверенно проинтерпретировать 19. Модель достигает качества повторения кластеризации 0.92 по метрике AUC-PR, что существенно превышает качество 0.68 предыдущих сетей [3].

Итоговая модель (Encoder на рис. 1) для определения по параметрам сигнала его класса имеет аналитический вид:

$$y_k = \left| \sum_{j=1}^{49} C_{kj} \left| \sum_{i=1}^9 A_{ij} x_i + B_j \right| + D_k \right|, \quad (1)$$

$$k_{detected} = \operatorname{argmax}(y_k) \quad , \quad (2)$$

где A_{ij} , B_j , C_{kj} , D_k - коэффициенты, которые ищутся в результате обучения сети, x_i - входные параметры, а $k_{detected}$ - номер скрытого класса, к которому будет принадлежать измеренный сигнал с параметрами x_i . Количество параметров модели равно 2240, что значительно лучше предыдущей модели [3].

Одним из актуальных вопросов при идентификации типов рассеянных сигналов является выбор параметров, которые необходимо для этого использовать. В рамках применения перестановочного метода, популярного в машинном обучении, к построенной модели классификатора было показано, что чаще всего важными для классификации являются: высота, на которой рассеивается сигнал, наклон траектории распространения радиоволны в точке рассеяния, а также примерно в равной мере угол с магнитным полем Земли в точке рассеяния и угол места в середине траектории распространения сигнала. Наименее важными параметрами являются мода распространения сигнала и спектральная ширина принятого сигнала. Таким образом, для классификации рассеянных сигналов наиболее важным является корректное и точное моделирование процесса распространения радиоволны.

Заключение

В работе сделана попытка решить задачу автоматической классификации данных радаров когерентного рассеяния. В рамках самосогласованного управляемого данными подхода был разработан метод автоматического построения такого классификатора. Количество свободных параметров модели 2240, количество обнаруженных классов сигналов 35.

Предложенный метод позволил автоматически классифицировать примерно 85 % всех данных, принимаемых в главном лепестке диаграммы направленности на радарах ЕКВ и MAGW в спокойном 2021 году и примерно 75 % данных радара MAGW за первую половину возмущенного 2023 года. Параметры трех вариантов обученной модели доступны по адресу <https://github.com/berng/WrappedClassifier/tree/master/v.3.0>. Результаты обработки сигналов предложенным алгоритмом в реальном масштабе времени доступны на сайте <http://sdrus.iszf.irk.ru/node/107>. Детальное обсуждение метода и результатов приведено в работе [10].

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 24-22-00436, <https://rscf.ru/project/24-22-00436/>.

Список литературы

1. ISTP SB RAS decameter radars / O.Berngardt, V.Kurkin, D.Kushnarev et al. // Solar-Terrestrial Physics. – 2020. – Vol. 6. – No. 2. – P. 63–73.
2. Review of the accomplishments of mid-latitude Super Dual Auroral Radar Network (Super-DARN) HF radars / N.Nishitani, J.M.Ruohoniemi, M.Lester et al. // Progress in Earth and Planetary Science. – 2019. – Vol. 6. – No. 1.
3. Self-trained artificial neural network for physical classification of ionospheric radar data / O.I. Berngardt, O.A. Kusonsky, A.I. Poddelsky, A.V. Oinats // Advances in Space Research. – 2022. – Vol. 70. – No. 10. – P. 2905–2919.
4. Hinton G. E., Salakhutdinov R. R. Reducing the Dimensionality of Data with Neural Networks // Science. – 2006. – Vol. 313. – No. 5786. – P. 504–507.
5. Berngardt O.I. Minimum number of neurons in fully connected layers of a given neural network (the first approximation) // arXiv e-prints. – 2024. – P. arXiv:2405.14147.–2405.14147.
6. A comparison of SuperDARN ACF fitting methods / A.J. Ribeiro, J.M. Ruohoniemi, P.V. Ponomarenko et al. // Radio Science. – 2013. – Vol. 48. – No. 3. – P. 274–282.
7. Interferometric evidence for the observation of ground backscatter originating behind the CUTLASS coherent HF radars / S.E. Milan, T.B. Jones, T.R. Robinson et al. // Annales Geophysicae. – 1997. – Vol. 15. – No. 1. – P. 29–39.
8. Kolmogorov A.N. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition // Dokl. Akad. Nauk SSSR. – 1957. – P. 953–956.
9. Ponomarenko P.V., Waters C.L., Menk F.W. Factors determining spectral width of HF echoes from high latitudes // Annales Geophysicae. – 2007. – Vol. 25. – No. 3. – P. 675–687.
10. Berngardt O.I. Data-driven approach to mid-latitude coherent scatter radar results classification // Solar-Terrestrial Physics(in press). – 2025.