

DOI: 10.51981/2588-0039.2023.46.021

КЛАСТЕРИЗАЦИЯ УНЧ-ВОЛН МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ ПО ДАННЫМ СПУТНИКА THEMIS-A

Д.А. Шубин, А.В. Рубцов, Д.Ю. Климущкин

Институт Солнечно-Земной Физики СО РАН, Иркутск, Россия; e-mail: d.shubin@iszf.irk.ru

Аннотация

В настоящей работе проведена кластеризация ультранизкочастотных волн по данным спутника THEMIS-A за один полный обзор магнитосферы Земли. Выделение волн производилось автоматически, после чего проводилась кластеризация методом Gaussian Mixture. Сравнительный анализ с использованием метода t-SNE позволил соотнести получившиеся кластеры с ожидаемым разделением волн по поляризации на полоидальные, тороидальные и компрессионные. Показано, что результаты кластеризации лишь частично совпадают с таким разделением.

Введение

Исследование ультранизкочастотных (УНЧ) волн в магнитосфере Земли представляет большое значение для понимания процессов переноса энергии заряженных частиц, летящих от Солнца на Землю (McPherron, 2005). Изучение волн на основе визуального исследования не всегда позволяют обнаружить неочевидные и интересные закономерности среди таких волн (Климущкин и др., 2021; Menk, 2011). Однако современные методы работы с большими данными позволяют это сделать и по-новому взглянуть на взаимодействие УНЧ-волн с частицами. В последнее время подобные методы, использующие машинное обучение, находят всё больше применения в анализе данных (например, Бернгардт, 2022).

Детектирование волн

В настоящем исследовании использовались данные зонда А многоспутниковой миссии Time History of Events and Macroscale Interactions during Substorms (THEMIS; Angelopoulos, 2008) с января 2017 года по февраль 2018 года. За этот период спутник совершил один полный обзор магнитосферы, т. е. охватил наблюдениями все MLT. Спутник THEMIS-A может достигать апогея в $15 R_E$ (радиус Земли) и покидать магнитосферу Земли. Чтобы анализировать данные только внутри магнитосферы, мы ограничили рассматриваемую область пространства $8 R_E$, что позволило избежать пересечения магнитопаузы в рассматриваемый интервал времени.

Спутниковые данные были взяты с открытого сервиса CDAWeb (<https://cdaweb.gsfc.nasa.gov/>). Для последующего анализа был разработан алгоритм, который позволяет детектировать УНЧ-волны в диапазоне Pc4 (период 45–150 с) и Pc5 (период 150–600 с) автоматически. Для этого в спектре колебаний магнитного поля за 15-минутный интервал определялось наличие пиков, чье отношение ширины пика на полувысоте к частоте пика не превышало 0,4. Пики на близких частотах у последовательно идущих 15-минутных интервалов считались одной волной (Якимчук и др., 2023).

Таким образом был сформирован набор данных, содержащий параметры наблюдения волны, усреднённые за время её наблюдения: пространственные характеристики (MLT, MLAT, L-оболочка спутника) и параметры магнитного поля (значения амплитуд по всем трём компонентам).

Кластеризация

Для выделения схожих групп данных применяют кластеризацию – метод машинного обучения, позволяющий независимо от пользователя разделить данные на группы (кластеры). Наибольшую известность получили методы Kmean, DBSCAN и Gaussian Mixture, позволяющие получить наиболее качественные результаты на разных наборах данных (Singh et al., 2015). Тем не менее, не существует универсального метода кластеризации, так как разные методы будут давать отличающийся результат на одних и тех же данных. Поэтому при использовании кластерного анализа нужно определить наиболее подходящий метод и ожидаемое количество кластеров.

Вопрос о количестве кластеров решается в ходе выполнения поставленной задачи, в то время как определить нужный метод для кластеризации более сложная задача. С этим позволяет справиться метод снижения размерности t-SNE (van der Maaten, 2014). В этом случае под размерностью мы понимаем количество параметров, характеризующих данный объект. Метод позволяет спроецировать объект в многомерном пространстве параметров на двумерную плоскость. Важная особенность метода заключается в том, что объекты, которые находились рядом в исходном многомерном пространстве, окажутся рядом и на полученной

проекции. Следует отметить, что оси координат получившейся проекции не имеют физической интерпретации и служат лишь для визуального отображения исходных объектов на плоскости.

Метод t-SNE позволяет определить необходимый класс методов, которые следует использовать с текущими данными. Если на проекции есть группы отдельно стоящих точек, "облаков" (где каждая точка — это некоторый многомерный объект), то это говорит о значительном пространственном разделении между группами. В таком случае следует использовать методы кластеризации, в которых фактором определения кластеров является их пространственное расположение, такие как Kmean и Spectral. В противном случае, когда такие "облака" отсутствуют, применяются иные статистические методы, наиболее известным из которых является Gaussian Mixture.

Результаты кластеризации

Среди УНЧ-волн обычно выделяют три основные группы: полоидальные (колебания происходят в радиальном направлении), тороидальные (колебания происходят в азимутальном направлении) и компрессионные (колебания вдоль силовых линий, значительно меняется модуль магнитного поля) (Климушкин и др., 2021). В настоящее время широко распространено предположение, что эти волны генерируются разными источниками, что проявляется в виде разного распределения в пространстве (Anderson et al., 1990). В настоящем исследовании мы проверили это предположение с помощью методов кластеризации. Для этого использовались данные о средней амплитуде колебаний магнитного поля в каждом из трёх направлений и их отношения между собой: b_a (азимутальное направление), b_r (радиальное), $b_{||}$ (направление вдоль силовой линии) и $b_a/b_{||}$, $b_r/b_{||}$.

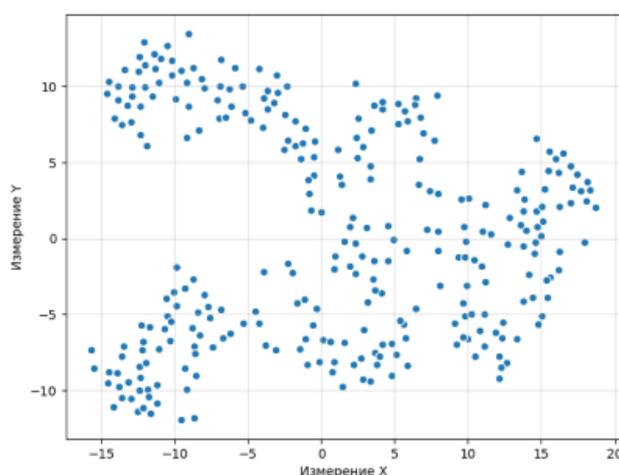


Рисунок 1. Двумерная проекция набора данных УНЧ-волн по измерениям магнитного поля спутником THEMIS-A с января 2017 г. по февраль 2018 г., полученная с помощью метода t-SNE.

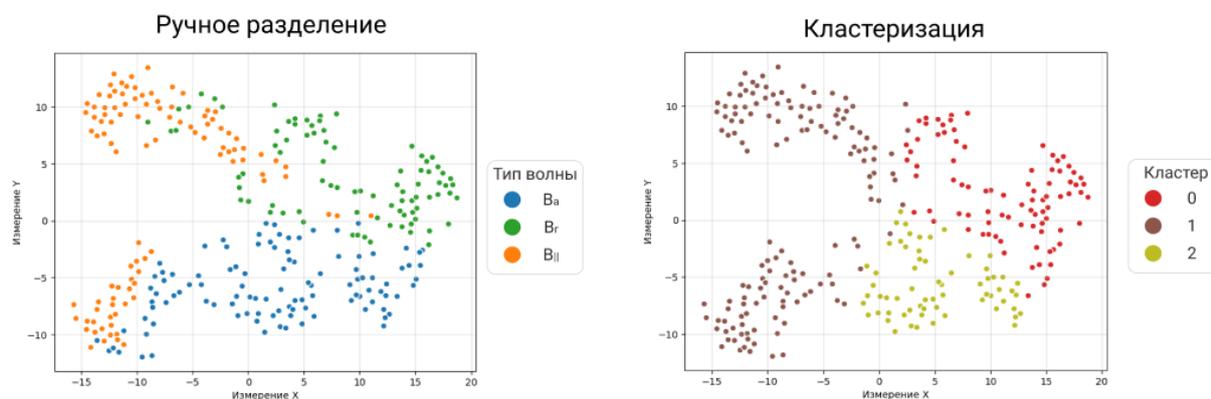


Рисунок 2. Результаты разделения волн на 3 кластера. Слева — ручное разделение по величине амплитуд колебаний магнитного поля по каждой из компонент. Справа — разделение на кластеры с помощью метода Gaussian Mixture.

На Рис. 1 показаны результаты применения метода t-SNE. Так как на графике нет чёткого разделения на облака, то было принято решение в дальнейшем использовать метод кластеризации Gaussian Mixture. Количество кластеров было выбрано равным трём из предположения о делении волн в магнитосфере на

группы с преобладающими колебаниями в разных направлениях: полоидальные, тороидальные и компрессионные. Мы сравнили результаты кластеризации с ручным разделением по направлению колебаний с наибольшей амплитудой (Рис. 2).

Полученное разделение на кластеры в общих чертах согласуется с ожидаемым разделением. Можно заметить, что группы полоидальных и тороидальных (зелёные и синие точки на Рис. 2, слева) выделились в соответствующие отдельные кластеры (красные и золотые точки на Рис. 2, справа), а компрессионные волны оказались объединены с частью поперечных волн в третий кластер. Этот результат можно объяснить сменой поляризации волн и ложным определением типа волны из-за небольшого отклонения от магнитного экватора (Леонович и Мазур, 2016).

Заключение

Мы показали, что кластеризация УНЧ-волн с помощью современных методов машинного обучения даёт результат, сходный с общепринятым разделением волн по поляризации. Однако имеются и заметные различия, которые могут быть интерпретированы, как с точки зрения физической природы, так и из-за особенностей используемых методов. Это открывает перспективы в продолжении исследования и более детального анализа найденных различий между группами, поиска закономерностей и факторов, которые на это влияют.

Работа выполнена при финансовой поддержке Минобрнауки России.

Список литературы

- Бернгардт О.И. Первый сравнительный анализ метеорного эха и спорадического рассеяния, идентифицированных самообучившейся нейронной сетью по данным радаров ЕКВ и MAGW ИСЗФ СО РАН. *Солнечно-земная физика*, 8, 66–76 (2022). <https://doi.org/10.12737/szf-84202206>
- Климушкин Д.Ю., Магер П.Н., Челпанов М.А., Костарев Д.В. Взаимодействие долгопериодических УНЧ-волн и заряженных частиц в магнитосфере: теория и наблюдения (обзор). *Солнечно-земная физика*, 7, 35–69 (2021). <https://doi.org/10.12737/szf-74202105>
- Леонович А.С., Мазур В.А. *Линейная теория МГД-колебаний магнитосферы*. — М.: ФИЗМАТЛИТ, 2016. — 480 с. — ISBN978-5-9221-1687-9
- Якимчук А.И., Рубцов А.В., Климушкин Д.Ю. Распределение поляризации поперечных УНЧ-волн по данным Van Allen Probe A: существуют ли раздельно тороидальные и полоидальные волны в магнитосфере? *Солнечно-земная физика* (2023), принята к печати.
- Anderson B.J., Engebretson M.J., Rounds S.P., Zanetti L.J., Potemra T.A. A statistical study of Pc 3–5 pulsations observed by the AMPTE/CCE Magnetic Fields Experiment, 1. Occurrence distributions. *Journal of Geophysical Research: Space Physics*, 95, 10495–10523 (1990). <https://doi.org/10.1029/JA095iA07p10495>
- Angelopoulos V. The THEMIS Mission. *Space Science Reviews*, 141, 5–34 (2008). <https://doi.org/10.1007/s11214-008-9336-1>
- McPherron R.L. Magnetic Pulsations: Their Sources and Relation to Solar Wind and Geomagnetic Activity. *Surveys in Geophysics*, 26, 545–592 (2005). <https://doi.org/10.1007/s10712-005-1758-7>
- Menk F.W. (2011). Magnetospheric ULF Waves: A Review. In: Liu W., Fujimoto M. (eds) *The Dynamic Magnetosphere*. IAGA Special Sopron Book Series, vol. 3. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0501-2_13
- Garima, H. Gulati, P.K. Singh, Clustering techniques in data mining: A comparison, 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2015, pp. 410–415.
- van der Maaten L.J.P. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research*, 15, 3221–3245 (2014)