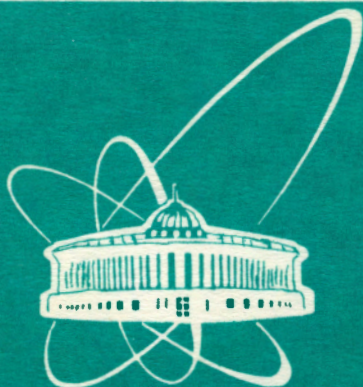


93-404



сообщения
объединенного
института
ядерных
исследований
дубна

P10-93-404

Е.Г.Зазнобина¹, Г.А.Ососков

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ
В АНАЛИЗЕ ДАННЫХ ВЕРТИКАЛЬНОГО
ЗОНДИРОВАНИЯ ИОНОСФЕРЫ

¹Ивановский госуниверситет, Россия

1993

1 Введение

Метод искусственных нейронных сетей (НС), ставший столь популярным в последние годы, позволяет успешно и быстро решать широкий спектр задач большой комбинаторной сложности. Это объясняется способностью нейронной сети эволюционировать к некоторым равновесным состояниям, соответствующим либо некоторым образам (проблемы распознавания, ассоциативная память), либо оптимальным значениям переменных (проблемы оптимизации).

В терминах формальных математических описаний равновесным состояниям НС соответствуют минимумы так называемой энергетической функции сети, причем для НС Хопфилда (ХНС) [1], рассматриваемых в данной работе, решению исходной задачи соответствует именно глобальный минимум, а не один из локальных, избегнуть попадания в которые и является необходимым. Поэтому при выборе наиболее подходящей модификации ХНС важно так учесть специфику конкретной задачи, чтобы создать начальную конфигурацию сети, энергетическая функция которой уже близка к своему глобальному минимуму, что существенно сокращает поиск решения (см., напр. [2, 3]).

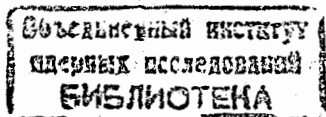
НС в самом общем виде — это структура вычислительных элементов, называемых нейронами по аналогии с мозгом благодаря их столь же тесной взаимосвязанности и взаимозависимости.

Каждый нейрон в любой момент времени характеризуется (см., напр. [4]):

- уровнем активации,
- топологией связей с другими нейронами (для ХНС все нейроны связаны друг с другом),
- мерой взаимодействия между ними, называемой синаптической силой связи (или весом. Матрица весов ХНС симметрична с нулевой диагональю),
- выходным уровнем, который связан с уровнем активации посредством некоторой функции обычно сигмоидального типа.

Все названные величины могут изменяться с течением времени по определенным правилам динамики НС, что в совокупности означает эволюцию системы как целого и приводит к решению задачи, определившей эту сеть. Так, например, новое значение уровня активации обычно определяется через некоторую функцию, аргументами которой являются величина прежнего уровня активации и комбинация (чаще всего сумма) входных воздействий с их весовыми коэффициентами. Веса в свою очередь также перевычисляются по новым значениям нейронов. Таким образом, НС является очень гибкой вычислительной системой, и вся проблема сводится к поиску переходных функций и удачной начальной конфигурации системы, обеспечивающей быструю сходимость к минимуму энергетической функции.

Известны результаты успешного применения ХНС в физике высоких энергий для автоматической идентификации треков заряженных частиц и связанных с ними "событий". В этой связи использовались разные модели НС — роторная [5], модифицированная роторная [2, 3], модель деформируемого шаблона [6]. В данной ра-



боте мы рассмотрим применение ХНС к разрешению сходной проблемы — распознаванию треков на ионограммах вертикального зондирования ионосферы. Значительная сложность экспериментального материала потребовала его предварительной обработки с помощью модификации клеточного автомата [7] и применения локального подхода, аналогичного [3].

2 Физические основы

Как известно, наличие ионосферного слоя, отражающего радиоволны, есть необходимое условие радиосвязи на определенных частотах. Для прогнозирования условий радиосвязи следует знать состояние ионосферы. С этой целью применяются различные способы радиозондирования ионосферы: на участок ионосферного слоя посылается излучение с определенными характеристиками и фиксируются параметры отраженного сигнала (амплитуда, поляризация, доплеровский сдвиг частоты и т.д.). Одним из самых распространенных источников информации о состоянии ионосферы является ионограмма вертикального зондирования (ВЗ), представляющая собой последовательность отсчетов частоты сигнала, его амплитуды и времени прихода после отражения от ионосферного слоя (время задержки сигнала соответствует высоте слоя и обычно так и называется). На рис. 1 представлена типичная ионограмма ВЗ. По оси абсцисс изменяется частота сигнала, по оси ординат — высота слоя.

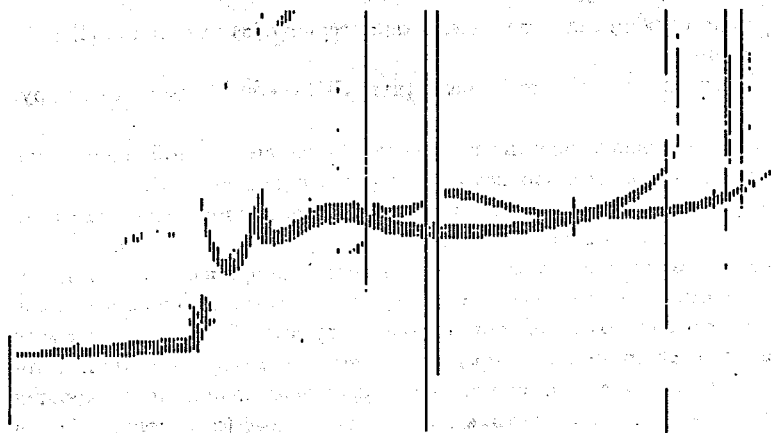


Рис. 1. Ионограмма вертикального зондирования

Видно, что точки-отсчеты формируют некие расплывчатые кривые достаточно разнообразной формы, которые и представляют основной интерес и будут да-

лее по устоявшейся терминологии называться треками. Один и тот же трек на разных частотах обычно имеет неодинаковую протяженность по высоте, может неоднократно пересекаться с другими треками. Некоторые точки в результате погрешностей аппаратуры могут отсутствовать, зато заметно явно излишнее присутствие помех как точечных, так и вертикальных, поражающих отдельные частоты. Однако на рис. 1 отражена не вся полезная информация — ведь каждый отсчет характеризуется еще амплитудой. Если представить ее изменяющейся по оси, перпендикулярной плоскости рисунка, то картинка приобретает трехмерность и выглядит примерно, как на рис. 2, немного напоминая горный пейзаж.

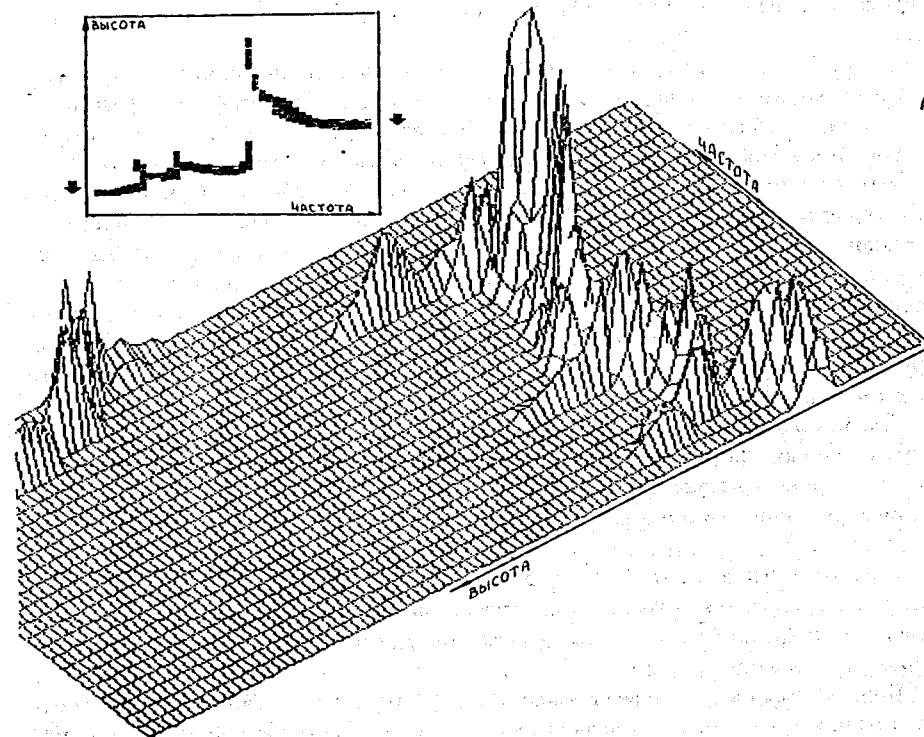


Рис. 2. Трехмерный вид фрагмента трека с учетом амплитуды сигнала

Распознающие способности человека позволяют проанализировать такую ионограмму, т. е. идентифицировать все треки, классифицировать их на основе формы, числа и ряда других параметров, чтобы сделать прогноз состояния ионосферы.

Однако массовость подобных исследований требует обработки столь больших массивов информации, что идея автоматизации этого процесса давно уже стала настоящей проблемой. Конечно, создать на ЭВМ полный аналог восприятия и распознавания человеком визуальной информации пока невозможно; поэтому речь идет о создании алгоритмов, моделирующих этот процесс лишь функционально с достаточной степенью надежности.

3 Постановка задачи

Мы сосредоточимся в этой работе на проблеме распознавания и идентификации треков, оставляя оценку параметров ионосферных процессов на последующие этапы.

Наша задача состоит в том, чтобы с минимальными временными затратами, без использования априорной информации отследить все имеющиеся треки и аппроксимировать их по точкам наибольшей амплитуды. Если вернуться к аналогии с горным пейзажем, то это означает зафиксировать траверсы всех "хребтов".

Это типичная задача распознавания без наведения, решаемая обычно методами кластер-анализа (см., напр. [8]). В общем смысле она аналогична проблеме идентификации треков, снятых с разного рода детекторов в физике высоких энергий [8, 9], хотя и имеет некоторые усложняющие отличия, не позволяющие непосредственно применить уже известные методы. Так, наличие помех и пропусков информации, с одной стороны, и возможность пересечения треков – с другой, являются непреодолимыми препятствиями для обычной кластеризации по критерию наименьшего или критического расстояния.

Очень мощный метод динамической кластеризации [8] требует для столь объемных данных чрезмерных временных затрат и мыслим лишь после предварительной процедуры значительного сжатия этих данных при условии сохранения их информативности и исключения явных ошибок типа удаления целых кластеров или их частей.

Предложенный в работе [10] метод так называемых масок дает хорошие результаты, но становится ненадежным в случае большой протяженности треков по высоте (диффузности) и содержит риск пропуска полезной информации в силу априорного задания масок.

Применяемый в настоящее время в СибИЗМИР [11] трехэтапный алгоритм отслеживания треков может успешно работать лишь при отсутствии сильной диффузности и пересечений последних.

Метод, использующий ХНС, позволяет снять эти ограничения. Однако его применение также требует существенного предварительного сжатия данных, т.к. сеть с количеством нейронов, равным числу точек в исходном массиве данных, потребовала бы чрезвычайно больших вычислительных мощностей.

Существенной проблемой является также поиск начальной конфигурации сети.

Таким образом, специфика задачи делает необходимым разбиение процесса распознавания на три стадии:

- Предварительное сжатие данных, существенно сокращающее их объем за счет отсеивания помех и точек, распределенных вокруг предполагаемого "траверса" трека, т.е. выделение гладких трек-элементов.
- Определение начальной конфигурации системы, близкой к минимуму энергетической функции сети.
- Идентификация треков методом НС.

В двух последующих разделах рассматриваются проблемы алгоритмической реализации этих стадий.

4 Предварительное сжатие данных

Сжатие также осуществляется в три этапа, каждому соответствует полный просмотр данных. Ввиду их значительного объема в работе непосредственно участвует только малая часть всех данных, а именно точки на трех соседних частотах. Таким образом, данные как бы пропускаются через фильтр, роль которого играет модификация клеточного автомата [7] с логикой, обеспечивающей "выживание" точек, расположенных по траверсам треков, и "вымирание" всех остальных. В перспективе такой способ допускает выполнение всех трех этапов за один просмотр, т.е. возможность анализа и сжатия данных синхронно с их поступлением (в режиме on-line).

Первый этап. Данные разбиваются на сегменты (первичные кластеры) по критерию близости точек на каждой частоте независимо от других частот; одновременно исключаются вертикальные и точечные помехи: для каждой точки-отсчета производится проверка на существование непосредственных соседей на предыдущей и последующей частотах. Если таковых не оказывается, точка отсеивается ("умирает") (см. рис. 3а).

Второй этап. Происходит заполнение случайных пропусков данных: если у отсутствующей точки-отсчета есть две соседние на предыдущей и последующей частотах (или высотах), то она "рождается" и включается в список данных. Таким образом, на каждой частоте остаются непрерывные сегменты "чистых" данных, разделенные пустыми промежутками (см. рис. 3б).

Третий этап. В каждом сегменте на каждой частоте "выживает" одна точка с максимальной амплитудой. Совокупность этих точек и есть искомые траверсы треков (см. рис. 4).

Как видим, алгоритм с применением клеточного автомата позволил успешно осуществить сжатие данных: их объем сократился в 5 – 10 раз, устранены помехи и оставлены точки максимальной амплитуды, расположенные по траверсам треков. В таком виде данные уже пригодны для анализа с помощью нейронной сети.

5 Алгоритм идентификации треков

По причинам, изложенным во введении, для решения проблемы была выбрана роторная модель ХНС [5] с модификациями, предложенными в [2, 3]. Каждому нейрону сопоставлен в соответствие вектор \vec{v}_i , модуль которого характеризует интенсивность воздействия данного нейрона на остальные, а направление в идеале должно совпадать с касательной к треку. Поскольку трек имеет непостоянную кривизну и даже изломы, взаимодействовать могут лишь нейроны некоторого локального участка, имеющего более-менее постоянную кривизну. Это достигается введением робастного множителя в весовую функцию w_{ij} , играющую роль жесткого фильтра. Переходя к полевой терминологии, введем векторную величину \vec{h}_i как поле, созданное в точке i нейроном j :

$$\vec{h}_{ij} = w_{ij} \vec{v}_j .$$

Тогда общее поле \vec{H}_i в произвольной точке i определится как векторная сумма полей от всех нейронов:

$$\vec{H}_i = \sum_j \vec{h}_{ij} = \sum_j w_{ij} \vec{v}_j .$$

Учитывая замечание о локальности, естественно предположить, что трек образуют нейроны, собственный вектор которых близок к полемому вектору в данной точке по направлению. За меру близости удобно взять скалярное произведение векторов \vec{v}_i и \vec{H}_i . Тогда энергетическая функция системы запишется в виде

$$E = -\frac{1}{2} \sum_i \vec{H}_i \vec{v}_i = -\frac{1}{2} \sum_{i,j} \vec{v}_i w_{ij} \vec{v}_j .$$

Ее минимум соответствует оптимальному распределению точек по трекам при надлежащем выборе весовой функции.

Поскольку в нашей задаче невозможно применить подход [2, 3] с глобальной аппроксимацией трека окружностью, мы вводим скользящее окно просмотра, размер которого выбирается так, чтобы в его пределах локальная аппроксимация трека дугой окружности была бы достаточно точной. Теперь можно определить весовую функцию из следующих геометрических соображений: если две точки принадлежат одному треку, то угол между ними мал, а расстояние не должно превышать диагонали выбранного окна. Как и в [2], вектор \vec{v}'_j получен зеркальным отображением вектора \vec{v}_j относительно хорды L_{ij} , стягивающей дугу локальной окружности. Направление и модуль вектора \vec{v}'_j определяются следующим образом [2, 3]:

$$\vec{v}'_j = \begin{pmatrix} \cos 2\varphi_{ij} & \sin 2\varphi_{ij} \\ \sin 2\varphi_{ij} & -\cos 2\varphi_{ij} \end{pmatrix} \cdot \vec{v}_j ,$$

где φ_{ij} есть угол между хордой L_{ij} и осью абсцисс Ox .

Таким образом, весовая функция может быть записана в виде

$$w_{ij} = |\vec{v}_i| |\vec{v}_j| \cos \Theta_{ij} , \quad \Theta_{ij} = \begin{cases} 1, & |L_{ij}| < R_c \\ 0, & |L_{ij}| > R_c \end{cases} .$$

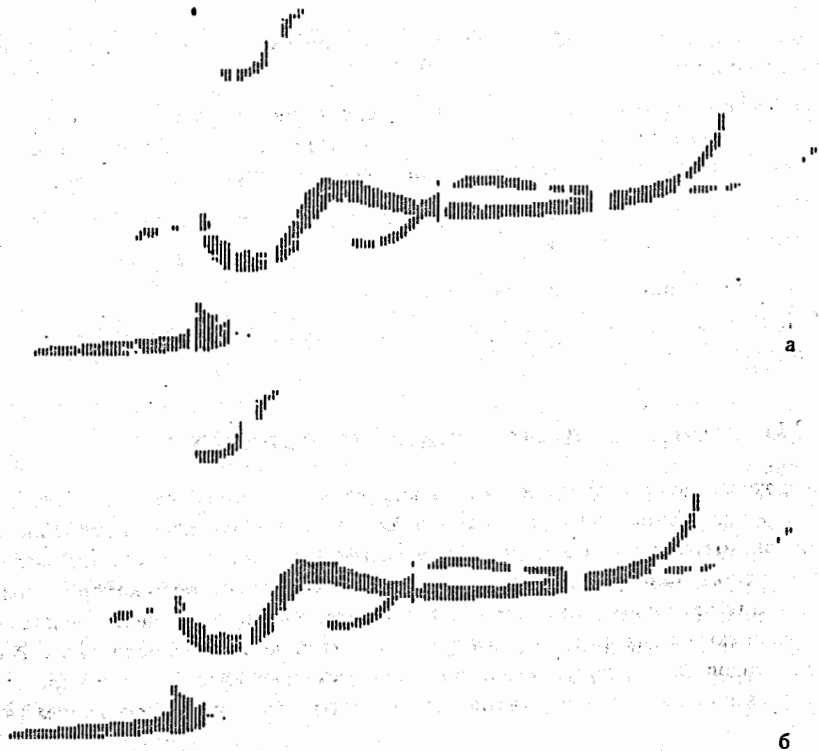


Рис. 3. Прослеживание треков по частотам (пояснения в тексте)



Рис. 4. Исходные данные после сжатия

Здесь Θ — угол между вектором \vec{v}_i и преобразованным вектором \vec{v}_j , $|L_{ij}|$ — расстояние между точками i и j , R_c — пороговое значение этого расстояния равное диагонали окна просмотра.

Вернемся к вопросу о начальной конфигурации. Здесь чрезвычайно полезной оказалась процедура предварительного гистограммирования, аналогичная той, что была использована в [3]. Метод, позволивший определить начальные значения векторов нейронов по сжатым данным измерений, заключается в последовательном просмотре исходных данных буквально с точки зрения каждого i -го нейрона. Все пространство вокруг последнего (в пределах окна просмотра) делится на угловые интервалы в 1° . Затем подсчитывается число точек (n_i), попадающих в каждый интервал.

За начальное направление вектора \vec{v}_i принимается значение угла центра тяжести трех соседних интервалов, средний из которых содержит максимальное число точек:

$$\alpha = \frac{\alpha_1 n_1 + \alpha_2 n_2 + \alpha_3 n_3}{n_1 + n_2 + n_3}$$

Модуль вектора считается равным среднему арифметическому от суммы точек в указанных трех интервалах.

Судя по результатам работы [3], следовало ожидать, что полученная начальная конфигурация должна быть достаточно близкой к минимуму энергетической функции.

Задача идентификации треков сводится к организации такой эволюции состояний нейронной сети, которая обеспечивает ее сходимость к конфигурации, минимизирующей энергетическую функцию системы.

Итеративная процедура поиска такой конфигурации НС строится, как последовательное вычисление поля в точках, соответствующих нейронам, определение каждым нейроном своего состояния и нового значения вектора для следующего шага итерации.

Новое значение модуля вектора i -го нейрона определяется по известной формуле [4]

$$|\vec{v}_i| = \tanh(|\vec{H}_i|/T), \text{ где } T = 1,5$$

Новое направление вектора принимается совпадающим с направлением поля в этой точке. Вычисленный таким образом вектор участвует в определении векторов других нейронов уже на данном шаге эволюции НС. На следующем шаге вся процедура повторяется. После нескольких итераций (достаточность их числа определяется по малости изменений энергетической функции с новой итерацией) векторы нейронов с некоторой погрешностью выстраиваются по направлению касательных к треку в точках локализации нейронов.

Окончательное восстановление треков проводится с помощью довольно простой процедуры интерпретации, сводящейся к поиску соседних по треку точек. Две точки считаются связанными (принадлежащими одному треку), если их взаимодействие вносит минимальный вклад в энергетическую функцию. В сложных случаях пересечений треков, когда энергетическая функция начинает осциллировать не убывая, применяется дополнительный критерий гладкости $dR/dl \rightarrow \min$, где R — локальная кривизна, l — длина "спорного" участка, позволяющий выбрать оптимальную линию продолжения трека.

6 Результаты применения и заключение

Построенная таким образом нейронная сеть применялась в обработке реальных данных, полученных в СибИЗМИР при вертикальном зондировании ионосферы зондом с непрерывной линейной частотной модуляцией [11]

В качестве примера на рис. 4-6 представлены данные одной из ионограмм:

- после сжатия с применением клеточного автомата (4),
- начальная конфигурация НС после гистограммирования (5),

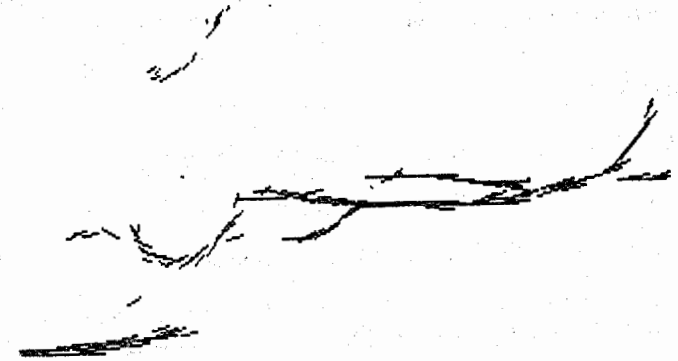


Рис. 5. Данные после гистограммирования

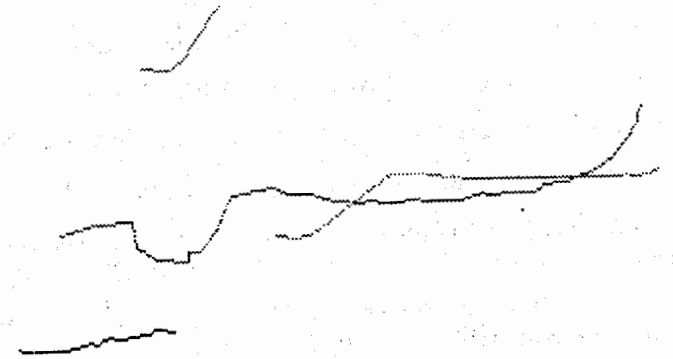


Рис. 6. Треки после прослеживания

- результат работы НС, достигнутый на четвертой итерации (6).

Сходимость алгоритма оказалась достаточно быстрой: в среднем 3 - 5 итераций, что подтвердило предположение о близости начальной конфигурации к состоянию минимума энергетической функции сети. Таким образом, затраты на этап определения начальной конфигурации с помощью гистограммирования, занимающий около 90% времени работы алгоритма, полностью окупаются скоростью сходи-

мости алгоритма. В среднем неоптимизированный алгоритм требует на обработку одной ионограммы 20–30 с для ПЭВМ класса IBM PC/AT для сетей с количеством нейронов 150–250 в зависимости от объема ионограммы. Ошибки идентификации составляют менее 5%.

На рис. 6 приведен пример завершающей работы алгоритма — отдельные треки в виде гладких кривых (применен метод медианного сглаживания [12]), пригодные для последующего структурно-лингвистического анализа. Хорошие результаты получены также на данных, содержащих диффузные треки и пересечения.

В заключение авторы считают своим долгом выразить глубокую признательность сотрудникам СиБИЗМИР проф. В.Е.Носову и И.А.Галкину за постановку задачи и предоставление экспериментальных материалов, а также С.Багиняну (ЛВТА ОИЯИ) за полезные консультации.

Литература

- [1] Hopfield J.J. Proc. Nat. Acad. Sci. USA 79(1982), 2554.
- [2] Glasov A. et al. JINR Comm., E10-92-352, Dubna, 1992.
- [3] Baginyan S. et al. JINR Comm. E10-93-86, Dubna, 1993.
- [4] Кисель И.В., Нескоромный В.Н. и Ососков Г.А. *Применение нейронных сетей в экспериментальной физике*. ЭЧАЯ, т. 24 (1993), вып. 6, стр. 1551.
- [5] Peterson C. *Neural networks and high energy physics*. Lund preprint LU TP 90-6, 1990.
- [6] Ohlsson M. et al. Comp. Phys. Commun., 71 (1992) 77.
- [7] Glazov A. et al. Nucl. Instr. and Meth., A329 (1993) 262-268.
- [8] Иоселиани Ц.И., Ососков Г.А. и др. Сообщ. ОИЯИ, P10-86-666, Дубна, 1986.
- [9] Chernov N. et al. Comp. Phys. Comm., 74 (1993) 217-227.
- [10] Галкин И.А. Программное обеспечение системы автоматической обработки ионограммы вертикального зондирования. СиБИЗМИР, P20-87, Иркутск, 1987.
- [11] Брынько И.Г., Галкин И.А., Грозов В.П. и др. ЛЧМ-зонд и его потенциальные возможности. СиБИЗМИР, П2-90, Иркутск, 1990.
- [12] Баранчук М.В. и др. Сообщ.ОИЯИ P11-90-9, Дубна, 1990.

Рукопись поступила в издательский отдел
5 ноября 1993 года.

Применение нейронных сетей в анализе данных вертикального зондирования ионосферы

Рассматривается модификация хопфилдовой нейронной сети (ХНС) для решения задачи обработки данных ионосферных измерений, образующих треки сложной структуры на фоне значительных помех. Последнее потребовало предварительной фильтрации и сжатия данных, что было выполнено методами, основанными на применении клеточного автомата и медианного сглаживания. Использовалась роторная модель ХНС, для построения начальной конфигурации которой применялось угловое гистограммирование в области с размерами, определяемыми средней локальной кривизной трека. Весовые функции связей нейронов строились с учетом близости направлений роторов к треку в той же локальной области. Удачный выбор начальной конфигурации обеспечивал быструю сходимость сети к состоянию, соответствующему глобальному минимуму ее энергетической функции. Ошибки итоговой идентификации треков составляют в среднем 5%.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 1993

Перевод авторов

Neural Network Application for Data Analysis of Vertical Ionosphere Locating

Hopfield neural network (HNN) is considered as a mean for processing data obtained by ionosphere measurements. Data look like tracks of a complicated structure with a considerable background. That required preliminary data filtering and compression. Both are accomplished by applying a cellular automaton and median smoothing. The specially modified rotor HNN model is used. For constructing its initial configuration an angle histogramming is applied in a domain which size is determined by the mean local track curvature. Neuron weight functions are constructed as a measure of the direction closeness of rotors and "their" tracks. The proper choice of the initial configuration guarantees the network fast convergence to the stable state, corresponding to the global minimum of its energy function. Mean errors of the track resulting identification are about 5%.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 1993