

УДК 004.032.26

МЕТОДИКА ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ОБУЧЕНИЯ НЕЙРОННЫХ КАРТ КОХОНЕНА ДЛЯ РАСПОЗНАВАНИЯ ВОЗМУЩЕНИЙ ГЕОАКУСТИЧЕСКОЙ ЭМИССИИ

А.В. Шадрин

Институт космофизических исследований и распространения радиоволн ДВО РАН,
684034, Камчатский край, п. Паратунка, ул. Мирная, 7

E-mail: SugerMas@yandex.ru

Данная работа посвящена методике обучения карт Кохонена на примере геоакустического сигнала в поддиапазоне 1500–6000 Гц. Описаны параметры обучения карты кохонена, которые позволили классифицировать аномалии в геоакустической эмиссии на различные типы.

Ключевые слова: геоакустическая эмиссия, геоакустический сигнал, возмущение, нейронная карта Кохонена, обучение

© Шадрин А.В., 2014

MSC 68T10

THE TECHNIQUE OF INCREASE THE EFFICIENCY OF LEARNING NEURAL KOHONEN MAPS FOR RECOGNITION OF PERTURBATIONS GEOACOUSTIC EMISSION

A.V. Shadrin

Institute of Cosmophysical Researches and Radio Wave Propagation Far-Eastern Branch,
Russian Academy of Sciences, 684034, Kamchatskiy Kray, Paratunka, Mirnaya st., 7,
Russia

E-mail: SugerMas@yandex.ru

This work is dedicated to technique of training Kohonen maps on the example of geoacoustical signal in the subrange 1500-6000 Hz. Describes the parameters of learning the Kohonen maps to classify anomalies in geoacoustical signal on different types.

Key words: geoacoustical emission, geoacoustic signal, disturbance, neural Kohonen maps, learning

© Shadrin A.V., 2014

Введение

Долгие годы ученые многих стран мира изучают природные катастрофические явления. Одной из главных целей исследований – прогноз этих событий, который позволит снизить человеческий и экономический ущерб. Наука прогнозирует ураганы, наводнение и другие стихийные бедствия, и лишь землетрясения наносят удар совершенно непредсказуемо, убивая людей там, где они чувствуют себя наиболее защищенными, — в собственных домах [1]. И хотя землетрясение возникает внезапно, в настоящее время научно доказанным является факт, что необходимо некоторое конечное время для накопления энергии разрывов пород в его очаге [2],[3].

Одним из перспективных направлений исследований с целью выявления аномалий, предшествующих землетрясениям, является регистрация и анализ возмущений геоакустической эмиссии. Данное явление обусловлено деформационными процессами из очагов будущих землетрясений. Исследования геоакустической эмиссии производятся гидрофонами, ориентированными по сторонам света и установленными в небольших водоемах на Камчатке [4], [5]. Диапазон исследуемых частот от 0.1 Гц до 10 кГц (используются частотные поддиапазоны 0.1–10 Гц, 10–50 Гц, 50–200 Гц, 200–700 Гц, 700–1500 Гц, 1500–6000 Гц, 6000–10000 Гц.). Далее сигналы накапливаются за 4 секунды и записываются на жесткий диск ЭВМ. Анализ данных показал, что в период 2001 – 2004 г.г. из 74 землетрясений с магнитудой $M > 4$, произошедших на эпицентральной расстоянии до 250 км от пунктов наблюдений, 34 предварялись в суточном временном интервале сильным увеличением уровня геоакустической эмиссии в килогерцовом диапазоне [4], [5]. Установлено, что амплитуда возмущений эмиссии зависит от магнитуды землетрясения и местоположения его эпицентра. Кроме возмущений геоакустической природы, системы регистрируют сигналы, обусловленные плохой погодой, в первую очередь действием осадков и сильного ветра. Частотный диапазон таких воздействий также составляет сотни герц - единицы килогерц и близок к диапазону возмущений перед землетрясениями. В связи с этим, для более детального исследования поведения геоакустической эмиссии перед сейсмическими событиями, а также ее выявления на фоне погодных аномалий, возникла необходимость классификации сигнала на основные типы: дождь, ветер, аномалии деформационной природы. Из-за особенности регистрации сигнала, использование стандартных частотных методов обработки не является эффективным. По этой причине автор воспользовался одной из разновидностей нейронных сетей - картами Кохонена, в основе которых лежит классификация и кластеризация данных.

В работе рассмотрена методика обучения карт Кохонена на примере геоакустического сигнала, а так же описаны методы решения проблем, которые возникли во время обучения. Для отработки методики обучения был взят поддиапазон 1500–6000 Гц, т.к. в нем наиболее четко проявляются все основные события, которые необходимо классифицировать. Обучение сети проводилось на сигнале за 2007 год, критериями обуславливающий выбор послужили: запуск цифровой системы регистрации данных в этом году, и высокая активность сигнала.

Выбор начальных параметров обучения сети

На первом этапе обучающая выборка формировалась из элементарных воздействий аномалий различной природы длительностью 80 сек (20 накопленных за 4 с

отсчетов). Затем сеть обучалась с использованием программы MATLAB. Рекомендаций по обучению карт Кохонена не приводятся, в связи с этим использовались параметры обучения, предложенные программой по умолчанию. Размерность сети была выбрана двумерная [20 20].

После обучения сети было исследовано распределение событий по нейронам, которое показало, что сеть равномерно распределила события, но не смогла классифицировать аномалии разных типов отдельными нейронами. Видно, что на одни и те же нейроны приходится одинаковое количество срабатываний для сигналов разных типов (рис. 1).

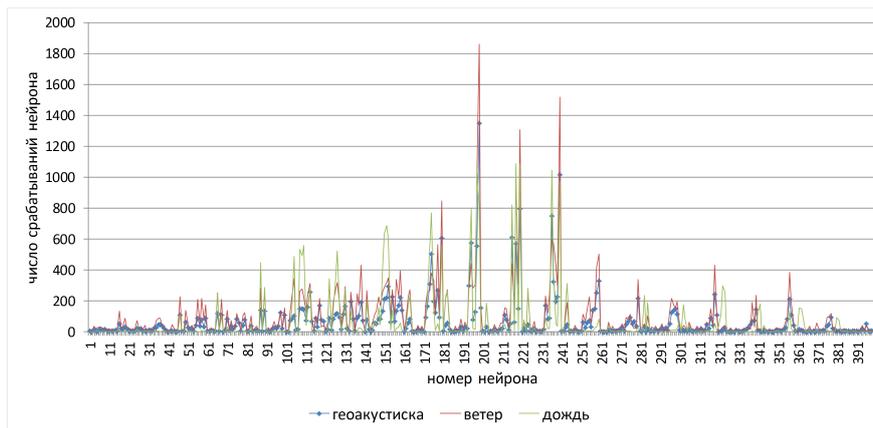


Рис. 1. Распределение событий по нейронам с использованием рекомендованных параметров обучения сети

Методика повышения эффективности обучения карт Кохонена

Проанализировав работу нейронной сети, была разработана методика, позволяющая обучить сеть и разделить сигнал на различные классы.

Из-за огромной размерности обучающей выборки использовался следующий алгоритм для отбора данных. Подсчитывалось математическое ожидание (МО) сигнала и его среднеквадратическое отклонение (СКО), далее вычислялось их отношение друг другу. В обучающую выборку попадали лишь те сигналы, отношение МО и СКО у которых было уникально.

В связи с тем, что нельзя выявить определенной зависимости нейрона сети от типа входного сигнала, каждый нейрон сети представляется в виде отношения распределения приходящихся на него аномалий. Рассмотрим это на примере вычисления распределения возмущений. На нейрон номер 1 приходится 5 срабатываний аномалий деформационной природы, 1 срабатывание от ветра и 0 - дождя. Обрабатываем результат, в данном случае оценка будет представлена в следующем виде: 1 аномалии деформационной природы, 0 - ветра и 0 - дождя. Если распределение событий на нейрон будет одинаковое (5 срабатываний аномалий деформационной природы, 5 срабатывание от ветра и 5 - дождя), то оценка будет $\{0.3 \ 0.3 \ 0.3\}$. Если два события имеют равное количество срабатываний, то $\{0.5 \ 0.5 \ 0\}$ или $\{0 \ 0.5 \ 0.5\}$ или $\{0.5 \ 0 \ 0.5\}$. Необходимо подобрать параметры сети таким образом, что бы после обучения сети число нейронов с результатом $\{0.3 \ 0.3 \ 0.3\}$ было минимально, а с $\{0 \ 0 \ 1\}$, $\{0 \ 1 \ 0\}$ и $\{1 \ 0 \ 0\}$ – максимально.

При обучении сети очень важно выбрать правильное количество эпох обучения, которое позволит избежать недообучения и переобучения карты. В системе MATLAB есть набор инструментов позволяющий наблюдать за процессом обучения сети. Один из них - это графическое отображения расположения весовых коэффициентов сети в пространстве векторов обучающей выборки во время обучения сети. К сожалению, инструментарий позволяет увидеть, как охватывает обученная сеть в пространстве входные данные только по двум первым координатам. Часто сеть скачкообразно накрывает все облако данных, это очень наглядно и заметно. Эпоха обучения, при которой весовые коэффициенты «охватили» входные данные, принимается за отсчетную точку. На рис. 2 показано скачкообразное изменение положения весовых коэффициентов карты.

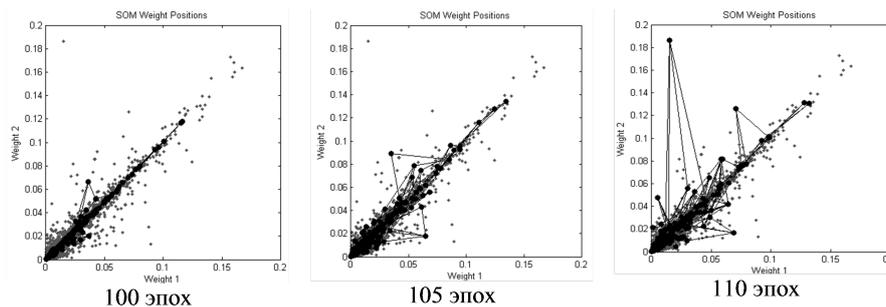


Рис. 2. Скачкообразное изменение положения весовых коэффициентов карты на разных этапах обучения. Маленькими точками отображена обучающая выборка, жирными точками, соединёнными линиями, изображены весовые коэффициенты

После обучения сети подсчитывается суммарное число срабатываний нейронов по аномалиям деформационной природы. Сеть обучается с пошаговым увеличением числа эпох относительно отсчетной точки, до тех пор, пока максимальное число срабатываний нейронов не достигнет максимального уровня и перестанет расти. Если продолжить увеличивать число эпох обучения дальше, то суммарное число срабатываний начнет снижаться, как следствие сеть становится переобученной.

Может возникнуть ситуация, в которой можно получить сеть с одинаковым числом срабатываний, но с разным числом эпох обучения. Для оценки качества сети в этом случае можно воспользоваться методиками оценки нейронных сетей, предложенными в [6]-[8]. Это качественная (средняя ошибка квантования) и количественная (топографическая ошибка) оценки. Качественная показывает способность нейронной сети раскрывать скрытую структуру и кластеризовать данные. Данная оценка может быть применена как мера разрешающей способности карты. Рассчитывается средняя ошибка квантования по следующей формуле:

$$eq = \frac{1}{N} \sum_{i=1}^N d(X, W_w)$$

где N - общее число входных векторов, которые участвуют в оценке карты; X - текущий вектор из входной выборки; W_w - вектор весов нейрона победителя, на текущее входное воздействие.

Количественная оценка позволяет определить непрерывность отображения входных векторов на пространстве карты [6]-[8]. Она измеряет пропорцию всех векторов данных, для которых первый и второй нейроны победители - несмежные. Чем меньше эта ошибка, тем лучше карта сохраняет топологию. Рассчитывается топографическая ошибка следующим образом:

$$et = \frac{1}{N} \sum_{i=1}^N u(X)$$

где X – текущий вектор из входной выборки; $u(X)$ – функция, принимающая значение 0 если первый и второй нейрон победители сети - смежные, и 1 в противном случае.

Результат обучения

После обучения сети с наилучшими характеристиками было повторно исследовано распределение событий по нейронам, которое показало, что на этот раз сеть смогла классифицировать большую часть аномалий разных типов отдельными нейронами (рис. 3).

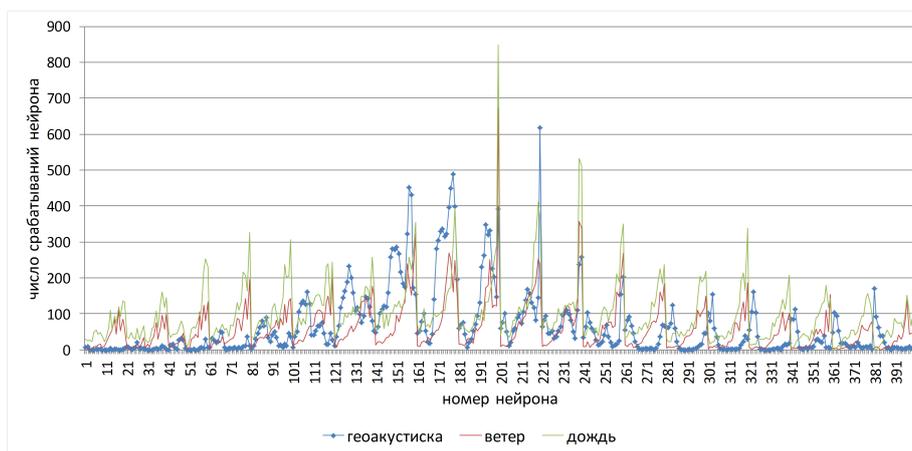


Рис. 3. Распределение событий по нейронам с использованием наилучших параметров обучения сети

Выбрав сеть с наилучшими характеристиками, проанализировали сигнал за 2007. Полученная оценка сети сравнивалась с каталогом землетрясений [9]. После сравнения оценивалась точность классификации сети (табл. 1).

Таблица

Процент правильно классифицированных сигналов за 2007 год

	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Аномалии деформационной природы	35	10	4	7	1	4	3	3	4	12	17
Метео аномалии	92	94	96	97	99	98	98	98	98	96	91

Как видно из таблицы сеть хорошо справилась с метео аномалиями и неудовлетворительно с аномалиями деформационной природы. Это связано с тем, что основная масса аномалий схожа по свойствам с сигналом обусловленными метео аномалиями. По этой причине для улучшения результата распознавания необходимо провести дополнительную работу по улучшению обучения сети. Кроме того, многократные эксперименты показали, что для достижения более точного результата, необходимо комплексно анализировать все поддиапазоны сигнала.

Заключение

Таким образом, созданная методика позволяет использовать карты Кохонена для классификации метео аномалий в геоакустическом сигнале. Для повышения точности выделения аномалий деформационной природы, необходимо провести дополнительные исследования.

Библиографический список

1. М. Родкин. Прогноз непредсказуемых катастроф // Вокруг света. 2008. № 6. С. 88-100.
2. А. Г. Соболев, А. В. Пономарев. Физика землетрясений и предвестники. М.: Наука. 2003. 270 с.
3. Добровольский И.П., Зубков С.И., Мячкин В.И. Об оценке размеров зоны проявления предвестников землетрясений. Моделирование предвестников землетрясений. М.: Наука. 1980. С. 7–44.
4. А. В. Купцов, Ю. В. Марапулец, Б. М. Шевцов. Анализ изменений геоакустической эмиссии в процессе подготовки сильных землетрясений на Камчатке // Исследовано в России. 2004. 262. С. 2809-2818. URL: <http://zhurnal.ape.relarn.ru/articles/2004/262.pdf>.
5. Купцов А.В., Ларионов И.А., Шевцов Б.М. Особенности геоакустической эмиссии при подготовке камчатских землетрясений // Вулканология и сейсмология. 2005. № 5. С.45-59.
6. J. Schatzmann. Using Self-Organizing Maps to Visualise Clusters and Trends in Multidimensional Datasets BEng thesis, Imperial College. June 19. 2003. URL: <http://mi.eng.cam.ac.uk/~js532/papers/schatzmann03soms.pdf>
7. Vesanto J. Data Exploration Process Based on the Self-Organizing Map, ActaPolytechnicaScandinavica // Mathematics and Computing Series. 2002. №. 115. pp. 96.
8. Arsuaga Uriarte, F. Diaz Martin. Topology Preservation in SOM //PWASET. 2006. V. 15. 187-191.
9. Мищенко М.А. Статистический анализ возмущений геоакустической эмиссии, предшествующих сильным землетрясениям на Камчатке // Вестник КРАУНЦ. Физико-математические науки. 2011. № 1 (2). С. 56–64.

Поступила в редакцию / Original article submitted: 15.11.2014