УДК 539.16:691

ИСПОЛЬЗОВАНИЕ КАРТ КОХОНЕНА ДЛЯ ОЦЕНКИ РАДИОАКТИВНОСТИ СТРОИТЕЛЬНОГО СЫРЬЯ

П.Э. Соколов

USAGE OF KOHONEN MAPS FOR THE RATE OF THE RADIOACTIVITY OF BUILDING RAW MATERIAL

P.E. Sokolov

Аннотация. В работе предпринята попытка структурировать сырьевые строительные материалы по радиоактивности с помощью самоорганизующихся карт Кохонена. Анализ проводился на данных об удельных активностях естественных радионуклидов в сырьевых строительных материалах Волгоградской области, а также по расчетным величинам удельных эффективных активностей естественных радионуклидов. В отличие от традиционных методов анализа использованная методика оценки радиоактивности основана на количественных характеристиках. В результате обучения сети на данных об удельных активностях сырьевых материалов получена карта кластеров с сегментацией по эффективной удельной активности естественных радионуклидов. На основании полученных результатов, сделаны выводы о возможности и целесообразности использования примененного алгоритма для классификации и анализа данных о радиоактивности строительного сырья.

Ключевые слова: *радиоактивность; естественные радионуклиды; удельная активность; самоорганизующаяся карта; карта Кохонена.*

Abstract. In the research, there has been made an attempt to structure raw construction materials due to their radioactivity by means of Kohonen self-organizing maps. The analysis was based on the data of specific activities of natural radionuclides in construction raw materials of the Volgograd region as well as on rated magnitudes of specific effective activity of natural radionuclides. Unlike traditional methods of the analysis the applied technique of radioactivity estimation is grounded upon quantitative performances. As a result of tutoring a web on the data of raw materials specific activities there has been obtained the map of clusters with segmentation on effective specific activity of natural radionuclides. On the basis of the obtained results there has been drawn a conclusion on a possibility and expediency of the applied algorithm usage for grading and analysis of construction raw material radioactivity data.

Keywords: *radioactivity; natural radionuclides; specific activity; self-organizing map; Kohonen map.*

Введение

Естественные радионуклиды, содержащиеся в строительных материалах для возведения зданий и сооружений различного назначения создают значительную дозу облучения населения [1]. Их основными источниками являются природные изотопы ^{40}K, ^{226}Ra и ^{232}Th [2]. Эти радионуклиды находятся во всех объектах окружающей среды, в том числе и в горных породах, которые служат сырьем для производства строительных материалов.

Обработка и анализ данных о радиоактивности, как строительных материалов, так и сырья для их производства обычно осуществлялась в рамках прикладной статистики, при этом в основном решались задачи по обработке небольших объемов данных. Наряду с этим существуют методы, которые позволяют производить количественный и качественный

анализ больших объемов данных. К одному из таких методов можно отнести карты Кохонена [3].

Карты Кохонена – это одна из разновидностей нейронных сетей, которые используют неконтролируемое обучение (обучение без учителя). При таком обучении обучающее множество состоит лишь из значений входных переменных, а в процессе обучения не происходит сравнения выходных нейронов с эталонными значениями. Такая сеть «учится» понимать структуру данных.

Самоорганизующиеся карты Кохонена – являются особым видом нейронных сетей, которые используются для решения задач кластеризации данных. Этот метод был предложен финским ученым Тойво Кохоненом в 1984 г. [3]. Основной принцип работы сетей – введение в правило обучения нейрона информации относительно его расположения, т.е. топологических свойств.

Самоорганизующиеся карты признаков (англ.: self organizing map – SOM) представляет собой не только эффективный алгоритм кластеризации, но и позволяют представлять ее результаты в виде двумерных карт, где расстояния между объектами соответствуют расстоянию между векторами в многомерном пространстве, а сами значения признаков отображаются различными оттенками. Таким образом, двумерная карта позволяет за счет различных цветовых оттенков показывать три измерения.

Самоорганизующиеся карты используют для решения задач моделирования, прогнозирования, поиска закономерностей в больших массивах данных, выявлении наборов независимых признаков, сжатия информации и т.д. Наиболее распространенное применение сетей Кохонена – решение задач классификации «без учителя», т.е. кластеризации. При такой постановке задачи дан набор объектов, каждому из которых сопоставлена строка таблицы (вектор значений признаков). Исходное множество требуется разбить на классы, т.е. для каждого объекта найти класс, к которому он принадлежит. В результате получения новой информации о классах возможна коррекция существующих правил классификации объектов.

Часто карты Кохонена применяются для разведочного анализа данных и обнаружения новых явлений. При обнаружении новых явлений сеть Кохонена распознает кластеры в обучающих данных и относит эти данные к тем или иным кластерам. Если после этого сеть встретится с набором данных непохожим ни на один из известных образцов, то она не сможет классифицировать такой набор, тем самым выявить его новизну.

Задачи исследования:

- сбор и предварительная обработка данных, отражающих удельные активности естественных радионуклидов в горных породах - сырье для производства строительных материалов;
- кластеризация сырья для производства строительных материалов по радиоактивности;
- определение кластеров, к которым относятся сырьевые материалы по радиоактивности и по виду сырья;
- анализ полученных результатов.

Постановка задачи

Использовать показатели удельных активностей естественных радионуклидов сырьевых строительных материалов для оценки их радиоактивности на основе методов кластеризации данных.

Методы исследований

Для проведения анализа использованы данные радиационного обследования месторождений по добыче сырья предназначенного для производства строительных материалов. На каждом месторождении было отобрано от 3 до 9 проб сырья. В дальнейшем каждая из отобранных проб была подвергнута гамма-спектрометрическому анализу. Для

этой цели использовался гамма-спектрометрический комплекс «ГАММА ПЛЮС-Р». В его состав входят сцинтилляционный детектор на основе кристалла NaI(Tl) размерами ∅63×63 мм, персональный компьютер, оборудованный амплитудно-цифровой платой и программное обеспечение «PROGRESS 3.00» [4]. Данный комплекс позволяет определять удельные активности естественных радионуклидов, которые нормируются соответствующими нормативными документами [5-7]. К ним относятся: ^{40}K, ^{226}Ra, ^{232}Th. Программное обеспечение позволяет определять удельные активности ЕРН и на основании полученных результатов производить расчет их эффективной удельной активности, используя для этого формулу:

$$A_{\text{эфф}} = 0{,}09 A_K + A_{Ra} + A_{Th}, \quad (1)$$

где A_K – удельная активность ^{40}К, Бк/кг; A_{Ra} – удельная активность ^{226}Ra, Бк/кг; A_{Th} – удельная активность ^{232}Th, Бк/кг.

В качестве исследуемых материалов использовались горные породы различных месторождений Волгоградской области, которые применяются в качестве сырья для производства строительных материалов.

Инструментальное средство анализа данных о результатах радиоактивности строительного сырья – аналитическая платформа Deductor – программный продукт, разработанный российской компанией [8]. Для обучения и применения сети Кохонена использовалась аналитическая платформа Deductor Studio 5.3 версии Academic. Она имеет все необходимое для применения традиционного алгоритма самоорганизующихся карт, разработанного Кохоненом: инициализация карты (задания начальных векторов), обучения (модификации векторов), вычисления ошибки квантования (ошибки отнесения к классу) и визуализации (классификации) [9].

Параметры модели (самоорганизующейся карты Кохонена): подобраны эвристическим путем с учетом рекомендаций по выбору оптимальной архитектуры нейронной сети и необходимости высокого качества визуализации и интерпретации результатов. Размерность карты - 12×16 шестиугольных ячеек, узлы которой представлены искусственными нейронами. Выбранное автором в рамках проведенного исследования количество кластеров, позволяет произвести анализ и интерпретацию результатов. Качество обучения нейросетевой модели лежащей в основе карты Кохонена – 100 % распознанных примеров на обучающем множестве.

Проведение исследований

Результаты исследований были сведены в одну таблицу, сохраненную в текстовом формате с разделителями в виде табуляции. Фрагмент исходных данных приведен в таблице 1. Подготовленный таким образом файл данных, возможно, использовать для импорта в аналитическую программу Deductor Studio версии Academic. После импорта текстового файла, данные были подвергнуты анализу на предмет пригодности для дальнейшего использования. Для этого применялись следующие отработчики: «Качество данных», «Заполнение пропусков», «Редактирование выбросов» и «Дубликаты и противоречия».

Имеющиеся исходные данные были разбиты на два набора, первый из которых использовался для обучения (обучающий), а второй для тестирования (тестовый). Для инициализации карты (задания начальных значений узлам сети) использовались значения из собственных векторов, равномерно распределенные в диапазоне значений входных данных.

Обучение проводилось несколько раз с различными значениями параметров обучения – варьировались радиусы соседства для обоих этапов обучения и для каждого набора значений параметров обучения использовалось несколько вариантов случайной начальной инициализации.

Таблица 1 – Фрагмент исходных данных о радиоактивности строительного сырья

№ п/п	Район	Месторождение	Наименование сырья	Число проб	Удельные активности радионуклидов, Бк/кг		
					40K	226Ra	232Th
1	Урюпинский	Южно-Добринское	глина	3	972,4	35,2	36,9
2	Жирновский	Дорошевское	песчаник	3	541,8	16,7	27,4
3	Михайловский	Себряковское	мел	5	62,6	9,8	4,4
4	Камышинский	Камышинское-4	песок	5	79,0	10,7	8,7
5	Фроловский	Арчединское-1	известняк	8	68,0	9,1	7,9

Для обученной карты проводилась визуализация, то есть процедура постановки в соответствие всем векторам, подаваемым на вход соответствующих им узлов нейронной сети (узлов победителей). Таким образом, мы можем в случае полных данных классифицировать их.

Нейронные сети работают только с числовыми данными, поэтому важным этапом при подготовке данных является преобразование и кодирование данных. С целью полного и всестороннего анализа имеющихся данных информация о виде сырья столбец Наименование сырья был закодирован целыми числовыми идентификаторами с присвоением ему названия Сырье. Для глины установлено значение – 1, для песчаника – 2, для мела – 3, для известняка – 4, для песка – 5 и для доломита – 6. Данная операция была проведена с использованием отработчика «Калькулятор». Общее количество сырьевых материалов составило – 95.Таким образом, был подготовлен набор данных пригодных для дальнейшего анализа с применением самоорганизующихся карт Кохонена.

Дальнейший анализ проводился с использованием отработчика «Карта Кохонена». Для этого ряду столбцов, таким как Район и Месторождение было присвоено назначения «Информационное», в дальнейшем при анализе они не учитывались. Столбцам Сырье и Число проб – целое, а столбцам удельных активностей ^{40}K, ^{226}Ra, ^{232}Th, а также эффективной удельной активности $A_{эфф}$ – вещественное. Столбцы Сырье, Число проб, ^{40}K, ^{226}Ra и ^{232}Th использовались как входные, а $A_{эфф}$ как выходное. Радиус обучения составлял в начале обучения – 4, а в конце – 0,1. При этом способ начальной инициализации карты – из собственных векторов, из обучающего множества и случайными значениями. Кроме этого изменялась функция соседства. Она принималась ступенчатой либо гауссовой. Наилучший результат показал вариант, при котором начальная инициализация карты проводилась из собственных векторов, а функция соседства принималась гауссовой.

Общее количество записей было разделено на два множества: обучающее и тестовое. В обучающем множестве содержалось 85 записей (89,5 %), а в тестовом 10 записей (10,5 %). По результатам анализа данных программа распознала 60 % тестовых записей. При этом все полученные данные были разделены на пять кластеров. Результаты кластеризации приведены в таблице 2.

Результаты моделирования

На рисунке 1 приведены результаты кластеризации для данных о радиоактивности строительного сырья Волгоградской области, представленные в виде карт Кохонена. Изучаемым материалам, показатели, радиоактивности которых представляют собой многомерные векторы, сопоставляются определенные участки на полученных картах. При этом вектора, схожие в исходном многомерном пространстве, оказываются рядом и на картах. Для того чтобы отразить относительную величину того или иного компонента вектора, соответствующие участки карты окрашиваются в определенный цвет в зависимости от выбранной цветовой схемы.

Таблица 2 – Результаты кластеризации сырьевых строительных материалов по радиоактивности

Номер кластера	Размер	Сырье	Удельная активность естественных радионуклидов, Бк/кг			$A_{эфф}$, Бк/кг
			^{40}K	^{226}Ra	^{232}Th	
1	7	известняк, песок	43,8÷70,8	6,5÷14,5	4,1÷11,0	18,9÷33,2
2	30	глина, песчаник	320,3÷745,9	14,0÷37,1	16,4÷39,7	72,1÷145,3
3	17	мел, известняк, песок, доломит	27,7÷104,7	3,9÷49,7	3,1÷14,3	10,4÷70,6
4	22	глина, песчаник	499,9÷972,4	13,7÷35,2	30,1÷61,1	122,8÷196,8
5	19	мел, известняк, песок	40,0÷117,2	5,5÷34,1	3,6÷17,5	18,6÷49,4

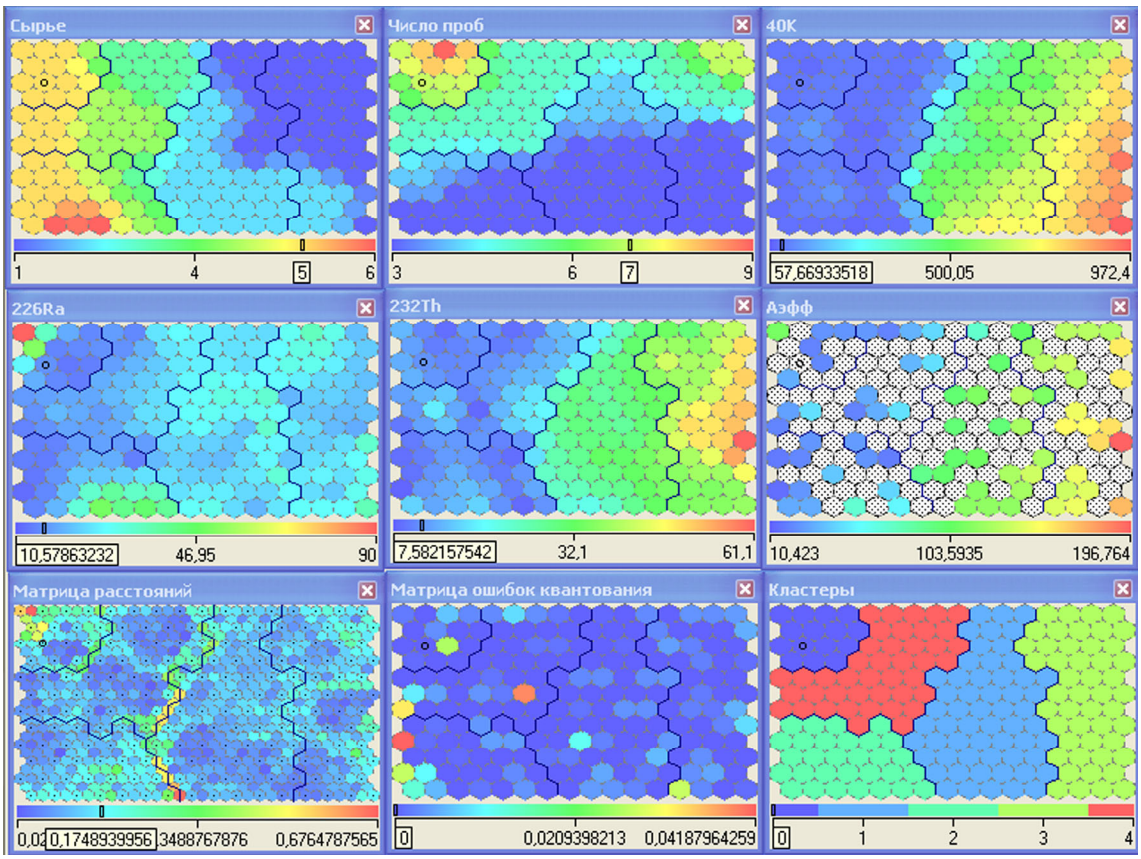

Рисунок 1 – Кластеризация сырьевых строительных материалов по радиоактивности

В результате работы алгоритма получены следующие карты (см. рисунок 1):
- карта входов нейронов: Сырье, Число проб, ^{40}K, ^{226}Ra и ^{232}Th;
- карта выходов нейронов: $A_{эфф}$;
- специальные карты: Матрица расстояний, Матрица ошибок квантования и Кластеры.

На карте входов нейронов веса нейронов подстраиваются под значения входных переменных и отображают их внутреннюю структуру. Для каждого входа рисуется своя карта, раскрашенная в соответствии со значениями веса нейрона.

При анализе данных используют несколько карт входов. На одной из карт выделяют область определенного цвета – это означает, что соответствующие входные параметры имеют приблизительно одинаковое значение соответствующего входа. Цветовое распределение нейронов из этой области анализируется на других картах для определения схожих или отличительных характеристик.

На карту выходов нейронов проецируется взаимное расположение исследуемых входных данных. Нейроны с одинаковыми значениями выходов образуют кластеры - замкнутые области на карте, которые включают нейроны с одинаковыми значениями выходов.

Специальные карты – это карты кластеров, матрица расстояний, матрица плотности попадания и другие карты, которые характеризуют кластеры, полученные в результате обучения сети Кохонена.

Между всеми рассмотренными картами существует взаимосвязь – все они являются разными раскрасками одних и тех же нейронов. Каждый параметр из обучающей выборки имеет одно и тоже расположение на всех картах.

Обсуждение результатов

Анализируя данные, приведенные в таблице 2 и на рисунке 1 можно отметить следующее.

В первый кластер входят 7 видов сырья. Из которых на долю песка приходится 6, а известняк представлен одним видом. Для материалов входящих в этот кластер характерны низкие значения удельных активностей и эффективных удельных активностей ЕРН. При этом интервалы изменения значений составляют 1,5-2 раза.

Второй кластер является самым большим по количеству входящих в него материалов. Для этого кластера характерны высокие значения удельных активностей и эффективных удельных активностей ЕРН. Из 30 видов сырья входящих в этот кластер на глину приходится 8 и 22 на песчаник. Таким образом, преобладающим материалом данного кластера является песчаник, а глины, попавшие в него имеют очень близкие к песчанику значения удельных активностей ЕРН.

Третий кластер представлен 17 видами сырья, из которых на мел приходится 3 вида сырья, на известняк – 3, на песок – 9 и на доломит – 2. По величинам удельных активностей и эффективных удельных активностей этому кластеру характерен широкий диапазон изменения значений.

В четвертый кластер входят 22 вида сырья, из которых глин – 16, а песчаников – 6. В этом кластере преобладающими являются глины. Значения удельных активностей и эффективных удельных активностей сырья входящего в этот кластер являются самыми высокими из исследуемых.

Пятый кластер представлен 19 видами сырья, из которых 8 приходятся на мел, 6 на известняк и 5 на песок. В отличие от первого кластера радиационные характеристики этих видов сырья несколько выше, а значения удельных активностей ^{226}Ra и ^{232}Th изменяются в более широких пределах.

Необходимо отметить, что все исследованные материалы соответствуют I классу материалов по радиоактивности и могут применяться для производства строительных материалов и для всех видов строительства без ограничения [5-7].

Заключение

Самоорганизующиеся карты Кохонена представляют возможность выявить и наглядно продемонстрировать скрытые взаимосвязи или различия между радиационными

характеристиками сырьевых строительных материалов. Полученные описание и обученную карту Кохонена можно использовать для получения характеристики любого другого сырьевого материала, для которого имеется тот же набор входных данных. Достоинство метода в том, что он позволяет работать с большим числом переменных, и позволяет визуально отображать результаты анализа в двумерном виде. Подтверждением данного утверждения являются приведенные результаты исследований.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Соколов П.Э. Проблемы радиационной безопасности в производстве и использовании строительных материалов. Волгоград: Изд-во ВолгГАСУ, 2003. 85 с.

2. Крисюк Э.М. Радиационный фон помещений. М.: Энергоатомиздат, 1989. 120 с.

3. Кохонен Т. Самоорганизующиеся карты. М.: БИНОМ. Лаборатория знаний, 2014. 655 с.

4. Измерение активности гамма-излучающих радионуклидов на сцинтилляционном спектрометре с использованием пакетов программ SM и EXPRESS: методические рекомендации. М.: ВНИИФТРИ, 1993. 31 с.

5. ГОСТ 30108-94. Материалы и изделия строительные. Определение удельной эффективной активности естественных радионуклидов. М.: Стандартинформ, 2007. 11 с.

6. СанПиН 2.6.1.2523 Нормы радиационной безопасности (НРБ-99/2009): [утв. гл. гос. санитар. врачом Рос. Федерации 07.07.2009: введ. 01.09.2009]. М.: Минздрав России, 2009. 93 с.

7. Основные санитарные правила обеспечения радиационной безопасности (ОСПОРБ 99/2010): Санитарные правила и нормативы СП 2.6.1.2612-10. М.: Федеральный центр гигиены и эпидемиологии Роспотребнадзора, 2010. 83 с.

8. BaseGroup Labs. Технологии анализа данных. [Электронный ресурс]. Режим доступа свободный. URL: http://basegroup.ru/ (дата обращения: 15.03.2017).

9. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. С.-Пб.: Питер, 2010. 704 с.

REFERENCES

1. Sokolov P.E. *Problemy radiatsionnoy bezopasnosti v proizvodstve i ispol'zovanii stroitel'nykh materialov* [Problems of radiation safety in the production and use of construction materials.]. Volgograd: Izd-vo VolgGASU, 2003. 85 p.

2. Krisyuk E.M. *Radiatsionnyy fon pomeshcheniy* [Background radiation areas]. Moscow: Energoatomizdat, 1989. 120 p.

3. Kokhonen T. *Samoorganizuyushchiesya karty* [Self-organizing maps]. Moscow: BINOM. Laboratoriya znaniy, 2014. 655 p.

4. *Izmerenie aktivnosti gamma-izluchayushchikh radionuklidov na stsintillyatsionnom spektrometre s ispol'zovaniem paketov programm SM i EXPRESS: metodicheskie rekomendatsii* [Measurement of activity of gamma-emitting radionuclides in the scintillation spectrometer using software packages such as SM and EXPRESS: guidelines.]. Moscow: VNIIFTRI Publ., 1993. 31 p.

5. GOST 30108-94. *Materialy i izdeliya stroitel'nye. Opredelenie udel'noy effektivnoy aktivnosti estestvennykh radionuklidov* [Construction materials and products. Determination of specific effective activity of natural radionuclides.]. Moscow: Standartinform Publ., 2007. 11 p.

6. SanPiN 2.6.1.2523 *Normy radiatsionnoy bezopasnosti (NRB-99/2009): utv. gl. gos. sanitar. vrachom Ros. Federatsii 07.07.2009: vved. 01.09.2009* [Radiation safety standards (NRB-99/2009): ratified. sec. of state nurse. doctor Russian Federation 07.07.2009: introduced 01.09.2009]. Moscow: Minzdrav Rossii, 2009. 93 p.

7. *Osnovnye sanitarnye pravila obespecheniya radiatsionnoy bezopasnosti (OSPORB 99/2010): sanitarnye pravila i normativy SP 2.6.1.2612-10* [Basic sanitary rules for ensuring radiation safety (OSPORB 99/2010): sanitary regulations SP 2.6.1.2612-10]. Moscow: Federal'nyy tsentr gigieny i epidemiologii Rospotrebnadzora, 2010. 83 p

8. BaseGroup Labs. *Tekhnologii analiza dannykh* [BaseGroup Labs. Data analysis technologies]. Available at: URL: http://basegroup.ru/

9. Paklin N.B., Oreshkov V.I. *Biznes-analitika: ot dannykh k znaniyam* [Business intelligence: from data to knowledge]. St.-Peterburg: Piter Publ., 2010. 704 s.

ИНФОРМАЦИЯ ОБ АВТОРЕ

Соколов Петр Эдуардович

Волгоградский государственный технический университет, Институт архитектуры и строительства, г. Волгоград, Россия, кандидат технических наук, доцент,

E-mail: pr7391@yandex.ru

Sokolov Petr Eduardovich

Volgograd State Technical University, Institute of Architecture and Construction, Volgograd, Russia, PhD in Engineering, associate Professor

E-mail: pr7391@yandex.ru

Корреспондентский почтовый адрес и телефон для контактов с автором статьи:
400074, г. Волгоград, ул. Академическая, 1, ИАиС, ВолгГТУ, каф. СМиСТ, ауд. В-2.
Соколов П.Э. 8(8442) 96-99-57