**Паттерны геологических объектов на основе нейро–технологий**

О. М. Гафуров1, Д. О. Гафуров2

1Международная лаборатория систем технического зрения Томского государственного университета, Томск, Россия;

2Красноярский научно-исследовательский институт геологии и минерального сырья, Красноярск, Россия.

**Аннотация.** Рассматривается область компьютерных наук, образовавшихся на пересечении искусственного интеллекта, статистики и теории баз данных и обозначаемых как «Data Mining» (обнаружение знаний в данных). Данный термин «Data Mining» обозначает процесс получения из "сырых" данных новой, потенциально полезной информации о предметной области, в нем наряду с классическими методами математического анализа и численного моделирования применяется теория нейронных сетей.Изложена методика создания паттернов геологических объектов на основе геолого-геофизической и геохимической информации, описаны возможности прогнозирования на основе созданных паттернов.

*Ключевые слова: нейронная сеть, паттерн, эталон, Data Mining*

*Keywords: neural network, pattern, standard, Data Mining*

**ВВЕДЕНИЕ.** Продуктивные отложения, входящие в целостную природную систему как их составляющие, в свою очередь, состоят из множества индивидуальных элементов, находящихся в историко-генетической связи друг с другом. Эти элементы испытывают качественные и количественные изменения на фоне непрерывно - прерывистого движения и фиксируются на страницах геологической истории в виде литолого-геохимических комплексов, обусловленных определёнными генетическими пространственно – временными структурными связями с окружающей геологической средой.

Для создания геолого-генетической модели важно правильное дифференцирование рассматриваемой толщи на составляющие компоненты. Их выделение должно находиться в диалектической взаимосвязи как с природной системой в целом, так и с продуктивной толщей в частности.

Используя известный ряд - химический элемент, минерал, горная порода, геологическое тело - необходимо добавить в него геологический комплекс, характеризующийся только ему присущими особенностями, однако сходный с другими, главным образом в отношении геохимических образований.

Подход, ориентируемый на объединение принципов обучаемости и интерпретируемости, соответствует целям и задачам Data Mining. B основу современной технологии Data Mining (discovery-driven data mining) положена концепция шаблонов (паттернов), отражающих фрагменты многоаспектных взаимоотношений в данных.

Эти шаблоны представляют собой закономерности, свойственные подвыборкам данных, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

В геоинформационной система «НейроИнформГео» реализована методика интеллектуального анализа и интерпретации геофизических и геохимических параметров на основе нейросетевых методов с ранжированием входных признаков по уровню значимости, обеспечивающим оценку информативного вклада каждого параметра для точности прогноза геологических объектов, а также методы комплексного анализа на основе обученных нейросетей.

Разработанная методика предполагает осуществление настройки и обучения нейронных сетей на эталонном месторождении, которое введено в разработку с установленным геологическим комплексом и представительным массивом геолого-геофизических данных описываемых данный эталонный комплекс. Обученные на эталонах нейронные сети формируют паттерн, который может быть использованы на этапах разведки и доразведки залежей полезных ископаемых или прогнозирования рудных тел, а также, с учетом статистических данных, на поисковом этапе на неизученных, малоизученных разработкой территориях.

**Определение паттерна.** Накопление опыта прогноза и интерпретации можно обеспечить в виде формирования, сохранения и корректного применения паттернов, представляющих собой библиотеку обученных нейронных сетей на эталонных геологических объектах, месторождениях или разработанных площадях [4].

Достоверность прогноза на каждом шаге обеспечивается правильно определенными параметрами новых пробуренных скважин или горных выработок. Полученная информация по новым скважинам или выроботкам принимается как опорная, имеющая для прогноза по геохимическим и геофизическим полям высшую абсолютную степень достоверности. Субъективными факторами в виде мнения геологов и геофизиков–интерпретаторов пренебрегают (т.е. прогнозные зоны не дорисовываются). Пошаговое дополнение знаний и прогноз эффективности бурения новых скважин должен быть добавлен их моделированием и пересчетом новых зон на основе предыдущего накопленного опыта и оценкой экономической эффективности точек их заложения.

**ПАТТЕРН ГЕОЛОГО-ГЕОФИЗИЧЕСКОГО КОМПЕКСА:**

* Изучение всей объемной части геосреды производится геохимическими и геофизическими методами (геохимия, магниторазведка, гравиразведка и т.д.). В процессе изучения естественных геохимических и геофизических полей (магнитометрия, гравиметрия и т.д.) или на основе принудительно возбужденных полей (сейсморазведка, электроразведка) образуется трехмерная матрица, где реальные геологические тела и объекты заменены цифровыми образами их физических и геохимических полей.
* Введем понятие физической модели среды в виде функции Qn(х, у, z), опи­сывающей распределение в области В физического параметра, создающего геофизическое поле fn (х, у, z) в окружающем пространстве с параметрами геосреды pi, . р k, а также соотношение:

fn (х, у, z ) = f(Q n, Pi,.., Рk), (1)

 связывающее Qn и fn , т.е. позволяющее рассчитывать поле данного геофизического параметра (прямая задача геофизики).

* Для расчета поля геофизического параметра f n (х, у, z) необходимы значения Оn (х, у, z) и параметров геосреды pi, ., рk , а также положение граничных (контактных) поверхностей, для сейсморазведки это значение плотности горных пород и границы их разностей, в геологии - пластов глин, песчаников и т.д., т.е. геологических тел, объектов.
* Задача построения геологической модели Qn (x, у, z) по измеренному геофизическому параметру fn(х, у, z) называется обратной задачей. Необходимо найти распределение параметров геосреды р1… рk и положение контактных поверхностей.
* Представим, что формирование ответов для каждого шага есть выска­зывание Р(1), Р(2), Р (3), ..., каждое из которых может быть либо истинным, либо ложным, т.е. вести нас к решению задачи либо уводить от нее. Принцип финит­ной (конечной) индукции утверждает, что для доказательства истинности выска­зываний Р (n) для всех n0 Р достаточно установить, во-первых, истинность Р (1) и. во-вторых, истинность бесконечной последовательности импликаций:

Р (1) => Р (2) => Р (3) => ... Р (п) => Р (п+1) => ... (2)

Машина должна на каждом шаге правильно рассчитывать параметры и начальная постановка задачи должна быть поставлена адекватно реальной геологической среде. Расчет параметров проводиться на каждом цикле нейрокомпьютерной про­граммой с опорой на обучающую выборку.

**МАШИННОЕ ОБУЧЕНИЕ ДЛЯ СОЗДАНИЯ ПАТТЕРНА.**

На первом этапе формирования обученных нейронных сетей (паттернов) определяется эталонный геологический объект или месторождение, которое введено в разработку и следовательно имеет представительный массив геолого-геофизической информации, который позволит корректно обучить и протестировать нейронные сети, создать «паттерн» для дальнейшего прогноза промысловых и геологических параметров на близлежащих ЛУ слабоизученных разведочным бурением и горными выработками.

Реализация методики создания паттерна в геоинформационной системе “НейроИнформГео” выглядит следующим образом [5]:

1. На вход из базы данных вызываются в едином масштабе с общей начальной точкой сетки различных геофизических и (или) геохимических параметров по площади или в плоскости разреза, при интерпретации 3D импортируется кубы сейсмических атрибутов;
2. Все входные параметры (карты (сетки) сейсмических атрибутов, геофизичесих, геохимических и др. параметров) нормируются на диапазон [-1...1].
3. В целях формирования эталона используются скважины или горные выработки в конкретной точке объекта , в которых известно содержание полезного ископаемого и значение наблюдаемого параметра геохимического и (или) геофизического поля.
4. Скважины или точки горной выработки разбиваются на классы прогнозируемого геологического признака.
5. Для каждой *i*-ой точки, где *i=1,…,* *,* на основе геофизических (геохимических) атрибутов рассчитывается множество точек, лежащих в круге радиуса , с центром в *i*-ой точке наблюдения. Выбор радиуса доверительного интервала зависит, от геологических условий и параметров пространства около точки наблюдения (скважины или точки горной выработки). При этом считается, что все точки, попавшие в доверительное пространство возле одной точки наблюдения, несут одинаковую информационную нагрузку.
6. Формируется эталон в виде вектора с усредненными значениями геофизических и (или) геофизических параметров и весами, обеспечивающими оценку информативного вклада каждого параметра. Точки наблюдения с доверительным радиусом, используются для оформления образа или классов.
7. Все точки наблюдения, находящиеся в полосе доверительной вероятности в первой итерации, разбиваются на две части. Часть скважин используется в качестве эталонной выборки, другая часть случайным образом формирует тестовую выборку для проверки качества прогноза.
8. По методу обучения нейросетей алгоритмом обратного распространения ошибки формируется многослойная нейронная сеть с  входами и одним выходным нейроном, принимающим значение OUT  значению выделенного класса k. На множестве  нейронная сеть обучается. Архитектура сети подбирается эмпирически, обеспечивается минимум ошибки обучения.
9. Для каждого из  входных параметров определяются показатели значимости - , характеризующие информативный вклад данного параметра (атрибута) в результаты. Нейронная сеть вычисляет градиент функции оценки по входным сигналам и обучаемым значениям сети. Таким образом, показатель значимости  параметра при решении *q*-о примера определяется по следующей формуле:

 (3)

где, при решении *q*-о примера показатель значимости показывает насколько изменится значение функции оценки решения сетью *q*-о примера если текущее значение параметра  заменить на ближайшее выделенное значение   для  параметра. Конечный показатель значимости  параметра рассчитывается как общее среднее:

, (4)

 где *n* число примеров. Таким образом, вычисленное значение показателя значимости для  параметра по существу представляет, в линейном приближении абсолютную величину изменения функции оценки при удалении из сети входного параметра (сигнала). Неинформативные параметры удаляются из сети, нейронная сеть переобучается.

1. Проверка обучения осуществляется сопоставлением множества спрогнозированных значений классов в тестовых скважинах с множеством присвоенных значений классов тестовым скважинам множества . На множестве  по несовпадениям ответов сети с известными ответами рассчитывается ошибка обобщения сети. В случае неудовлетворительных результатов обучение повторяют.
2. В случае удовлетворительного решения задачи осуществляется прогнозирование на малоизученных ЛУ. Прогнозирование осуществляется по соответствующему эталону набору параметров. Программа разбивает всё многомерное признаковое пространство на классы принадлежности или подобия тому или другому эталону и формирует на выходе карту эталонов для площадной интерпретации или формирует разрез с выделением комплексов (классов) на временную или глубинную модель.

**ЗАКЛЮЧЕНИЕ**. При решении задач прогнозирования геологических объектов технология применения эталонов и паттернов опробовано на ряде золоторудных месторождениях по комплексу данных литохимической съемки по первичным и вторичным ореолам рассеяния на территории Восточной Сибири. А также на ряде нефтянных и газовых объектах Иркутской области (Непско-Ботуобинская НГО), Красноярского края (Байкитское НГО, Пур-Тазовское НГО) и ряде меторождений Республики Саха (Якутия), Томской области, Тюменской области, Ханты-Мансийского автономного округа, в различных сейсмогеологических условиях. Получен Патент РФ, созданы програмные продукты и базы данных [1-3].

**Список Литературы**

1. Патент РФ на изобретение № 2477499 «Способ определения мест заложения эксплуатационных скважин при разработке месторождений углеводородов» заявка №2011125437. Гафуров О.М., Гафуров Д.О., Гафуров А.О., Панков М.В., Битнер А.К., Красильникова Н.Б. – 2012 г.
2. Свидетельство о государственной регистрации программы для ЭВМ в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам: "Программа обработки сейсмоакустических и геофизических данных в формате сеток .dat на основе обучаемых нейронных сетей для формирования эталонов и расчета паттернов" №2011613919 , Гафуров О.М., Гафуров Д.О., Гафуров А.О., 2011 г.
3. Свидетельство о государственной регистрации базы данных в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам: "База данных блока автоматизированного формирования эталонов в интеллектуальной геоинформационной системе "НейроИнформГео"" №201162031, Гафуров О.М., Гафуров Д.О., Гафуров А.О., 2011 г.
4. Denis Gafurov and Oleg Gafurov. Development and Practical Application of Neural Information Technologies, Patterns // MATEC Web of Conferences. — 11 October 2016. — Vol. 79. VII Scientific Conference with International Participation «Information-Measuring Equipment and Technologies» (IME&T 2016), no. 01086. — P. 12.
5. Denis Gafurov and Oleg Gafurov. Method of Creation of “Core-Gisseismic Attributes” Dependences With Use of Trainable Neural Networks // MATEC Web of Conferences. — 11 October 2016. — Vol. 79. VII Scientific Conference with International Participation «Information-Measuring Equipment and Technologies» (IME&T 2016), no. 01055. — P. 10.