



ПРАКТИКА АТРИБУТНОГО ПРОГНОЗИРОВАНИЯ ТРЕБУЕТ СОВЕРШЕНСТВОВАНИЯ

В.В. Масюков, М.А. Попов, С.П. Тюнегин

*ООО “Славнефть – Научно-производственный центр”,
170002, Тверь, просп. Чайковского, 21А, Россия, e-mail: tvergw@slavneft.ru*

В статье дан обзор современного состояния технологий атрибутного прогнозирования с теоретической и практической точек зрения. Предложены новые, удобные оценки точности, надежности и информативности прогнозирования. Показана важность использования “эффективного коэффициента корреляции”, учитывающего неравномерность расположения точек на кросс-плоте. Обсуждается и сравнивается прогнозирование по нейронным сетям и регрессионным моделям.

Сейсмическая интерпретация, атрибутное прогнозирование

INDUSTRIAL APPLICATIONS OF ATTRIBUTE PREDICTION NEED IMPROVEMENTS

V.V. Masjukov, M.A. Popov, S.P. Tunegin

Slavneft R& D Department, 170002, Tver, Chaykovskovo 21A, Russia, tvergw@slavneft.ru

The paper presents a brief review of state-of-the-art attribute prediction technologies including their theoretical and practical aspects. New, more convenient estimations of reliability, accuracy and informativity of geological attribute predictions have been suggested. Utility of “effective correlation coefficient” for scattered data has been shown. In the article one also discusses and compares neural networks and regression predictions.

Seismic interpretation, attribute prediction

ВВЕДЕНИЕ

Методические рекомендации [Левянт и др., 2006] по использованию сейсморазведки 3D для подсчета запасов нефти и газа содержат ряд важных положений по использованию атрибутного прогнозирования (в том числе основанные на публикациях [Шленкин и др., 2004; Масюков А.В., Масюков В.В., 2005; Шленкин и др., 2005]). Однако технологии атрибутного прогнозирования постоянно совершенствуются, и за прошедшие семь лет получены новые важные результаты. В частности, в статье [Масюков и др., 2007] предложен “эффективный коэффициент корреляции”, учитывающий неравномерное расположение точек на кросс-плоте, а в статьях [Масюков и др., 2006, 2008] доказана важность использования наиболее адекватных геолого-геофизическому прогнозированию скейлинговых итерационных методов интерполяции. Тем не менее в научных статьях и производственных отчетах уважаемых отечественных и зарубежных компаний, работающих на российском нефтегазовом рынке, обнаруживаются грубые ошибки и откровенное непонимание (или незнание) их авторами теоретических основ атрибутного прогнозирования и последних достижений в этой области. Основная цель настоящей статьи – последовательно изложить теоретические основы, описать современное состояние и практические аспекты технологий атрибутного прогнозирования в простой и понятной для специалистов-практиков форме. Обзор наиболее важных результатов основывается на статье [Шленкин и др., 2006]. Кроме того, в статье вводятся новые, удобные оценки надежности, информативности, точности атрибутного прогнозирования, обсуждаются оценки его качества и подтверждаемости, а также сравнивается

прогнозирование на основе регрессионных моделей и нейронных сетей. Содержание данной статьи может представлять интерес для специалистов широкого круга деятельности, использующих в работе корреляционный и регрессионный анализ.

ВЕРОЯТНОСТЬ ОШИБКИ ВТОРОГО РОДА – ПРИНЯТИЕ СЛУЧАЙНОСТИ ЗА ЗАКОНОМЕРНОСТЬ

Атрибутное прогнозирование является “простым и понятным методом прогнозирования” [Ампилов, 2004]. Оно основывается на количественной оценке статистической связи между измеренными (оцененными) в скважинах параметрами геологической среды и атрибутами сейсмического поля и последующем использовании этой связи для построения прогнозных карт. Какова вероятность того, что на самом деле связи нет, и мы приняли случайность за закономерность? Этот принципиальный вопрос достаточно подробно рассмотрен в работе [Шленкин и др., 2006].

Известно [Тейлор, 1983], что при конкретном значении коэффициента корреляции Пирсона между двумя выборками вероятность ошибки второго рода (т. е. вероятность принятия случайности за закономерность) можно оценить с помощью таблиц, например табл. 1. Эти результаты справедливы при анализе статистической связи между двумя конкретными (фиксированными, определенными) признаками. Однако атрибутное прогнозирование предполагает поиск статистических связей, основанный зачастую на переборе значительного числа различных атрибутов. При таком подходе уже нельзя непосредственно использовать табл. 1.

Вероятности (в %) случайного получения выборочного коэффициента корреляции больше R для двух некоррелированных выборок*

$n \backslash R$	0.600	0.625	0.650	0.675	0.700	0.725	0.750	0.775	0.800	0.825	0.850	0.875	0.900	0.925	0.950
5	28.5	26.1	23.7	21.4	19.1	16.9	14.8	12.8	10.9	9.1	7.3	5.7	4.2	2.7	1.5
6	20.6	18.2	15.9	13.8	11.7	9.8	8.1	6.6	5.2	3.9	2.8	2.0	1.2	0.7	0.3
7	15.7	13.6	11.7	9.8	8.2	6.7	5.3	4.1	3.1	2.2	1.5	0.9	0.5	0.2	0.1
8	11.7	9.9	8.3	6.8	5.5	4.4	3.4	2.5	1.8	1.3	0.8	0.5	0.2	0.1	
9	8.8	7.2	5.9	4.6	3.6	2.7	2.0	1.4	0.9	0.6	0.3	0.2	0.1		
10	6.6	5.2	4.1	3.1	2.3	1.7	1.2	0.8	0.5	0.3	0.1	0.1			
11	5.3	4.2	3.2	2.4	1.8	1.3	0.9	0.6	0.4	0.2	0.1	0.1			
12	4.0	3.1	2.3	1.7	1.2	0.8	0.6	0.4	0.2	0.1	0.1				
13	3.1	2.4	1.7	1.2	0.8	0.6	0.4	0.2	0.1	0.1					
14	2.4	1.8	1.3	0.9	0.6	0.4	0.2	0.1	0.1						
15	1.9	1.4	1.0	0.7	0.4	0.3	0.2	0.1							
16	1.5	1.0	0.7	0.5	0.3	0.2	0.1								
17	1.2	0.8	0.5	0.3	0.2	0.1	0.1								
18	0.9	0.6	0.4	0.2	0.1	0.1									
19	0.7	0.5	0.3	0.2	0.1	0.1									
20	0.6	0.4	0.2	0.1	0.1										
21	0.5	0.3	0.2	0.1	0.1										
22	0.3	0.2	0.1	0.1											
23	0.3	0.2	0.1												
24	0.2	0.1	0.1												
25	0.2	0.1	0.1												
26	0.1	0.1													

* Красным цветом выделен недопустимый диапазон при уровне значимости в 5 %; синим – допустимый, с вероятностью ошибки второго рода менее 5 %.

Предположим, что у нас есть N измерений одной величины, и мы перебираем m случайных признаков (выборки такого же объема) и вычисляем коэффициент корреляции. Это – реализация схемы Бернулли, и вероятность получить хотя бы один раз в m испытаниях коэффициент корреляции больше, чем r_0 , вычисляется по известной формуле $P_m(r > r_0) = 1 - (P(r \leq r_0))^m$, где $P(r \leq r_0) = 1 - P(r > r_0)$ находится с помощью табл. 1. В работе [Масюков и др., 2007] показано, что при принятом уровне значимости в 5 %, т. е. требование $P(n, R) < 5$ при переборе m атрибутов должно быть заменено более жестким условием

$$P(n, R) < 100(1 - \sqrt[m]{0.95}). \quad (1)$$

Графики зависимости вероятности $P_m(r > 0.7)$ от числа испытаний m (количества перебираемых атрибутов) при объеме выборки $N = 5, 10, 15$ приведены на рис. 1. Из рисунка видно, что при объеме выборки (числе скважин) $N = 10$ вероятность ошибки второго рода (при полученном коэффициенте корреляции большем 0.7) превышает 50 % при переборе более 30 случайных атрибутов. При пяти скважинах совсем легко получить высокий коэффициент корреляции, но доверять полученной зависимости не стоит. Естественно, прогнозирование при $P > 0.5$ (с надежностью менее 50 %) нецелесообразно в принципе. На самом деле уровень значимости (пороговое значение вероятности ошибки второго рода) не должен превышать 5–10 %.

Нельзя использовать табл. 1 для ответа на вопрос: “От чего зависит данный признак в наибольшей степени?” при переборе некоторого (достаточно большого) числа признаков. Очевидно, что при неограниченном переборе произвольных признаков мы рано или поздно получим значение выборочного коэффициента корреляции не меньшее, чем наперед заданное (например, 0.7). Читатель может в этом наглядно убедиться, исследуя корреляцию между двумя столб-

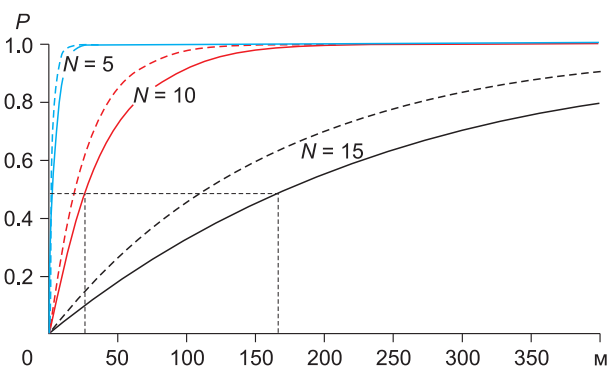


Рис. 1. Графики зависимости вероятности ошибки второго рода (получение коэффициента корреляции $r > 0.7$ при отсутствии связи) от количества перебираемых атрибутов m при числе скважин $N = 5$ (синий), $N = 10$ (красный) и $N = 15$ (черный). Соответствующими пунктирными линиями показаны зависимости при отбрасывании одной (наиболее выпадающей из регрессии) скважины.

цами в MS Excel, в которых заданы случайные числа (функция СЛЧИС). После нескольких пересчетов всегда удается получить высокий коэффициент корреляции (если объем выборки невелик, например, меньше 10, что часто соответствует количеству скважин на этапе разведки месторождения). *Следовательно, при поиске статистической связи (путем перебора достаточно большого числа признаков) для оценки вероятности ошибки второго рода (надежности прогнозирования) необходимо пользоваться формулой (1) и графиками рис. 1, а не основываться только на полученном (достигнутом за счет перебора) значении коэффициента корреляции.*

В этой связи уместно обсудить технологии, основанные на поиске и использовании так называемых комплексных атрибутов. Действительно, если ввести в качестве атрибута некоторую нелинейную комбинацию даже небольшого числа атрибутов (создать “комплексный атрибут”) с непрерывно меняющимися параметрами, то фактически эту комбинацию в точках скважин можно рассматривать как некоторый датчик случайных чисел (при различном выборе параметров). В этой связи также представляются очень опасными какие-либо дополнительные операции с атрибутами (например, сглаживания с разной базой и т. п.), поскольку фактически это также сводится к перебору большого количества вариантов. Поэтому мы считаем, что использование такого подхода недопустимо при атрибутом прогнозировании из-за очень высокой вероятности ошибки второго рода. Заметим, что использование нелинейных регрессий сводится к линейной регрессии, если в качестве регрессора рассматривать соответствующую нелинейную функцию атрибута. Поэтому с точки зрения технологии достаточно ограничиться линейной регрессией. Следует также иметь в виду, что перебор различных вариантов при поиске самих вариантов нелинейных связей также увеличивает ошибку второго рода.

Необходимо добавить, что часто на практике при анализе корреляционной связи и построении прогноза отбрасывается одна или несколько наиболее выпадающих скважин. Это допустимо лишь при очень серьезных основаниях и дополнительных исследованиях (например, скважина попадает в ярко выраженную крайнюю зону, в которой нельзя полностью доверять сейсмическим данным). *При выбрасывании скважин, естественно, вероятность ошибки второго рода (принятие случайности за закономерность) возрастает (это показывают пунктирные зависимости на рис. 1).*

Какие же атрибуты использовать? Ведь некоторые производители программных продуктов заявляют о наличии у них нескольких сотен сейсмических атрибутов. Наиболее предпочтительным в настоящее время [Ампил, 2004] считается известный подход [Milligan, 1978], основанный на методе главных компонент и использовании в качестве атрибутов факторов сейсмического поля. На расчет факторов подается набор конформных или пропорциональных срезов в целевом интервале. В результате сейсмическое поле $A(x, y, t)$ в некотором окне, привязанном к целевому интервалу, представляется в специальном виде

$$A(x, y, t) = \sum_{k=1}^R F_k(x, y) \varphi_k(t), \text{ где } F_k(x, y) - \text{ факторы}$$

сейсмического поля. Факторы являются независимыми (в отличие от большинства стандартных сейсмических атрибутов, которые часто сильно коррелируют

между собой), и можно сказать, что они содержат в себе всю информацию о сейсмическом поле (зная базисные функции φ_k и факторы $F_k(x, y)$, можно восстановить исходное поле). При возрастании номера фактора повышаются его пространственные частоты, но уменьшается его вес (вклад в энергию) в разложении сейсмического поля. Так, при учете в разложении по факторам 99,9 % общей энергии сейсмического поля и расчете факторов по 100 слайсам обычно достаточно ограничиться 7–10 членами разложения и, следовательно, при поиске корреляционных связей рассматривать только 7–10 карт факторов. Этот подход существенно снижает произвол при выборе атрибутов, уменьшая вероятность ошибки второго рода. Естественно, можно использовать и небольшое количество других физически осмысленных атрибутов – мгновенные (или доминирующие) частоты, RMS амплитуды, структурные атрибуты (карты временных мощностей целевых интервалов) и т. п. Выбор атрибутов зависит от специфики прогнозируемого параметра. Например, при попытке прогнозирования трещиноватости естественно использовать различные оценки когерентности сейсмического поля [Левянт и др., 2010; Шленкин и др., 2012].

НАДЕЖНОСТЬ, ИНФОРМАТИВНОСТЬ И ТОЧНОСТЬ ПРОГНОЗИРОВАНИЯ

В течение десяти лет в НПЦ “Славнефть” используется внутренняя инструкция по выполнению атрибутивного прогнозирования и представлению его результатов. Опыт показал, что включенные в инструкцию дополнительные количественные оценки атрибутивного анализа (информативность, надежность и условная точность) являются понятными, удобными и эффективными.

Информативность прогнозирования оценивается параметром

$$I = 1 - \sigma/\sigma_0, \quad (2)$$

где σ – среднеквадратичная ошибка прогноза, σ_0 – среднеквадратичный разброс скважинных данных (или среднеквадратичная ошибка прогноза по среднему). Поскольку среднеквадратичную ошибку можно считать мерой неопределенности, то введенная характеристика прогноза дает некоторую оценку уменьшения неопределенности. Информативность также показывает, насколько полученный прогноз лучше прогноза по среднему (который мы имели бы при использовании только скважинных данных). Если $\sigma \geq \sigma_0$, то $I = 0$, если $\sigma = 0$, то $I = 1$ (или 100 %).

Надежность найденной связи оценивается параметром

$$N = 1 - P/P_0. \quad (3)$$

Здесь P_0 – уровень значимости (обычно принимается за 5–10 %), а P – вероятность (в %) случайного получения коэффициента корреляции R при отсутствии связи, вычисляемая по статистической табл. 1 в зависимости от R , числа скважин и числа атрибутов. Если $P \geq P_0$, то $N = 0$, если $P = 0$, то $N = 1$ (или 100 %). Использование этой характеристики идет в русле основных тенденций в нефтегазовой отрасли по внедрению количественных оценок рисков (например, повсеместное внедрение оценок рисков запасов углеводородов при их расчете объемным способом). Однако следует иметь в виду, что при оценке надежности статистической связи уровень значимости должен

Диапазоны значений параметров информативности $I = 1 - \sigma/\sigma_0$, надежности $N = 1 - P/P_0$ и условной точности прогноза $T = 1 - \sigma/\Delta$

Параметры	Диапазоны значений			
	“хорошая”	“приемлемая”	“слабая”	“недопустимая”
I (информативность связи)	$0.75 \leq I \leq 1$	$0.5 \leq I \leq 0.75$	$0.25 \leq I \leq 0.5$	$I < 0.25$
N (надежность)	$0.7 < N \leq 1$	$0.4 \leq N < 0.7$	$0 \leq N < 0.4$	$N < 0$
T (условная точность)	$0.5 < T \leq 1$	$0.25 \leq T \leq 0.5$	$0 \leq T \leq 0.25$	$T < 0$

составлять 5–10 %, не более, в то время как оценки запасов приводятся с уровнем значимости и 10 и 50 %.

Условная точность прогноза оценивается параметром

$$T = 1 - \sigma/\Delta, \quad (4)$$

где Δ – это желательная предельно допустимая погрешность, определяемая заказчиком в соответствии с геологическими особенностями и рекомендациями Министерства природных ресурсов (например, при средней мощности пласта 20 м значение Δ можно принять 3–5 м, что недопустимо при средней мощности пласта в 5 м). Мы рекомендуем выбирать значение Δ равным 40 % от среднего (или максимального) значения предсказываемой характеристики. Этот параметр показывает, как абсолютная точность прогноза, оцениваемая значением σ , соотносится с желательной предельно допустимой погрешностью. Если $\sigma \geq \Delta$, то $T = 0$, а если $\sigma = 0$, то $T = 1$ (или 100 %).

Расчет приведенных выше параметров реализован в программном комплексе CLUST+ [Масюков и др., 2004] и в виде макросов в MS Excel. На основании опыта использования и рекомендаций министерства предложены диапазоны “хороших”, “приемлемых”, “слабых” и “недопустимых” значений введенных выше характеристик прогноза, которые показаны в табл. 2.

ИСПОЛЬЗОВАНИЕ “ЭФФЕКТИВНОГО КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ”

Существуют различные количественные оценки тесноты статистической связи. Это и коэффициент линейной связи Пирсона (наиболее известный и популярный выборочный коэффициент корреляции), и коэффициент ранговой корреляции Спирмена, и мн. др. В MS Excel используется коэффициент детерминированности r^2 (иногда он называется критерием тесноты связи), который численно равен $1 - D/D_0$, где

D_0 – дисперсия исходных данных, а D – дисперсия разности исходных данных и их прогнозных значений (остаточная дисперсия). Очевидно (и это легко проверить в том же MS Excel), что коэффициент детерминированности совпадает с квадратом коэффициента корреляции Пирсона, если в качестве двух выборов подать исходные данные и их прогнозные оценки. Что лучше использовать при атрибутом прогнозировании? Оказывается, ни одна из известных традиционных оценок статистической связи не учитывает неравномерность расположения точек на кросс-плоте, и их вычисление недостаточно для принятия решения о возможности использования найденной корреляционной связи для прогноза.

Действительно, на рис. 2 (см.: [Шленкин и др., 2006]) представлены два кросс-плота – (А) и (Б), полученные на базе одинакового количества скважин и содержащие одинаковый коэффициент корреляции Пирсона. Но если на кросс-плоте (А) отключить одну отдельно стоящую точку, то коэффициент корреляции резко падает, более того, зависимость превращается в убывающую (В). Легко показать, что если кросс-плот представляет собой достаточно компактное облако точек и отдельно стоящую точку, то при удалении точки от облака коэффициент корреляции Пирсона стремится к 1. Опытный геолог-интерпретатор, безусловно, не будет использовать для прогноза “плохой” вариант (А), так как он, как говорят, “висит на одной точке”. Если используются два регрессора, то недопустимо, чтобы прогноз “висел” на двух точках, и т. д.

В статье [Масюков и др., 2007] введен “эффективный коэффициент корреляции”. На самом деле это просто удобная числовая характеристика для оценки значимости корреляционных связей. В соответствии с общим подходом к проверке значимости рассматривается гипотеза об отсутствии связи между прогнозируемым параметром и атрибутами. Пусть почти все скважины имеют близкие значения пара-

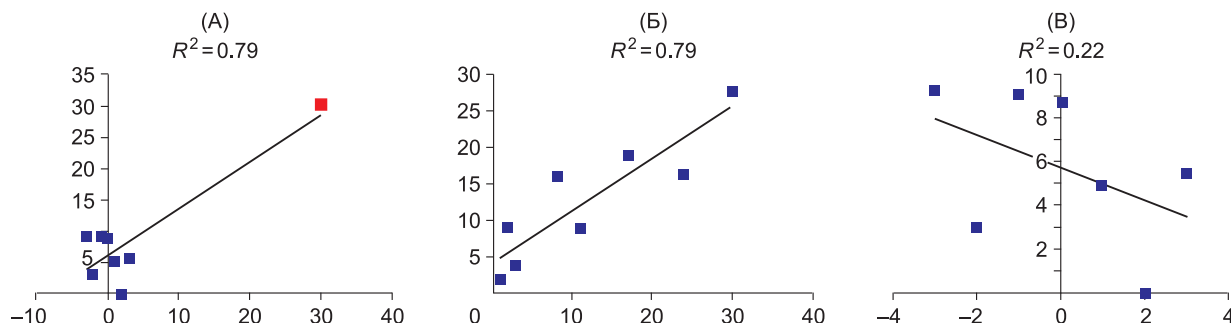


Рис. 2. Коэффициенты корреляции Пирсона для кросс-плотов (А) и (Б) одинаковы, но при отбрасывании красной точки на кросс-плоте (А) регрессионная зависимость кардинально меняется (В).

метра и каждого из атрибутов, а несколько скважин – сильно отличающиеся. Тогда за счет свободного члена регрессии близкие по значениям скважины образуют на кросс-плоте “облако” с малым коэффициентом корреляции. Остальные же коэффициенты регрессии позволят получить высокий коэффициент корреляции с вероятностью 1, если число различающихся по значениям скважин не больше числа атрибутов (за счет минимизации регрессионной невязки, а не за счет наличия связи между параметром и атрибутами). Определим скважины, число которых равно числу атрибутов и отбрасывание которых максимизирует отношение остаточной дисперсии D_R (без пересчета модели) к дисперсии параметра D . Под дисперсией D_R здесь, естественно, понимается средний квадрат отклонений точек кросс-плота “факт–модель” от его тренда. Введем эффективный коэффициент корреляции

$$R_e = \sqrt{1 - D'_R/D'}, \quad (5)$$

вычисляемый после отбрасывания этих скважин. В рамках рассмотренной гипотезы полученный коэффициент является мерой статистической значимости, т. е. он должен быть подставлен в табл. 1 для проверки надежности корреляционной связи вместо обычного коэффициента корреляции в случае неравномерного расположения точек скважин на кросс-плоте.

Уточним алгоритм вычисления эффективного коэффициента корреляции в соответствии с [Масюков и др., 2007]. Пусть p_i – отклонение значения параметра в i -й скважине от среднего по всем скважинам значения, а Δ_i – невязка прогноза (модели) в i -й скважине. Тогда обычный коэффициент корреляции (Пирсона) равен

$$R = \sqrt{1 - D_R/D} = \sqrt{1 - \sum_i \Delta_i^2 / \sum_i p_i^2}, \quad (6)$$

а эффективный коэффициент корреляции при одном регрессоре (одном атрибуте) равен

$$R_e = \sqrt{1 - D'_R/D'} = \sqrt{1 - \max_j (D_R - \Delta_j^2) / (D - p_j^2)}. \quad (7)$$

В случае нескольких регрессоров максимизацию надо повторять столько раз, сколько атрибутов используется в регрессионной модели. Если под корнем в формуле (7) отрицательное значение, то $R_e = 0$.

Заметим, что идея использования эффективного коэффициента корреляции близка идее подхода Cross-Validation для оценки надежности прогнозирования, основанного на последовательном выключении части исходных данных (скважин), выполнении прогноза и последующем анализе невязок “факт–прогноз” в отброшенных точках (скважинах). Фактически предложенная нами оценка показывает, насколько хуже могла бы быть связь (коэффициент корреляции), если выкинуть из рассмотрения число точек, равное числу атрибутов.

Введенный эффективный коэффициент корреляции обладает следующими очевидными свойствами:

- если коэффициент Пирсона равен по модулю 1, то эффективный коэффициент корреляции тоже равен 1;

- если при одном регрессоре (атрибуте) на кросс-плоте существует точка, при отбрасывании которой средний квадрат отклонения от линии регрессии (построенной с учетом этой точки) становится боль-

ше дисперсии выборочных значений параметра, то эффективный коэффициент корреляции равен 0;

- при подстановке эффективного коэффициента корреляции в табл. 1 надо взять строку со значением n , равным числу скважин, так как количество используемых регрессоров уже учтено при вычислении R_e .

В [Масюков и др., 2007] показаны преимущества эффективного коэффициента корреляции перед коэффициентами корреляции Спирмена и Пирсона. Возвращаясь к рис. 2, отметим, что эффективный коэффициент корреляции для кросс-плота (А), в предположении, что использован один регрессор, равен 0 (что соответствует интуитивным представлениям о недопустимости использования такой зависимости для прогноза).

На большом количестве синтетических и реальных данных мы пришли к выводу, что эффективный коэффициент корреляции оказывается адекватной числовой характеристикой для количественного описания надежности кросс-плота. Его использование определяет допустимость или недопустимость построения прогноза на основе найденной корреляционной связи и формализует предпочтения, отдаваемые той или иной зависимости при одинаковых коэффициентах корреляции Пирсона. Вычисление эффективного коэффициента корреляции легко программируется и может быть реализовано в виде макроса в MS Excel.

РЕГРЕССИОННАЯ МОДЕЛЬ ИЛИ НЕЙРОННЫЕ СЕТИ?

Прогнозирование на основе нейронных сетей известно достаточно давно и включено во многие геофизические программные продукты. Мы долгое время с некоторым недоверием относились к нейронным сетям (как “вещи в себе”), поскольку оценить надежность прогнозирования и явно представить работу нейронных сетей достаточно сложно. Однако в 2012 г. в наш программный продукт для атрибутного прогнозирования был добавлен модуль прогноза по нейронным сетям. Упрощенная схема нейронной сети показана на рис. 3.

Нами был выбран наиболее проработанный в теоретическом плане алгоритм обучения нейронной сети, так называемый алгоритм обратного распространения [Gurpey, 1997]. Существует значительное количество алгоритмов обучения, которые делятся на два больших класса: детерминистские и стохастические. Алгоритм обратного распространения является детерминистическим и, с математической точки зрения, сводится к поиску минимума функционала, который обычно ищется методом градиентного спуска.

Смысл обучения сводится к последовательному изменению значений синапсов сети (весов) с целью уменьшения невязки “факт–прогноз” (разницей между известными значениями предсказываемого параметра в скважинах и выходными значениями нейронной сети). Следует отметить, что ограниченное число скважинных данных (данных для обучения) приводит к необходимости уменьшения числа нейронов сети (уменьшение числа степеней свободы). Очевидно, что при числе нейронов, превышающем число скважин, мы всегда имеем коэффициент корреляции “факт–прогноз”, равный 1. То же самое мы получим, если на вход регрессионной модели подать количество атрибутов (причем любых, отличных в точках скважин), равное количеству скважин. При этом полученный

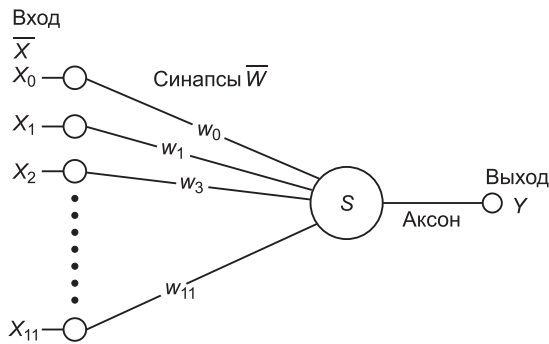


Рис. 3. Схема простейшей нейронной сети. На вход подаются значения с карт атрибутов, они суммируются с весами (синапсами) и после нелинейного преобразования аксоном на выходе получается прогнозируемое значение.

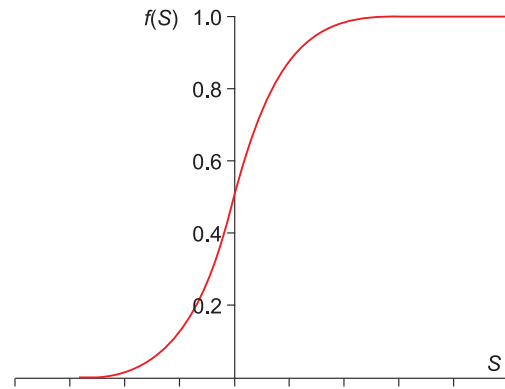


Рис. 4. Вид активационной функции (аксона) типа сигмоид, часто используемой в нейронных сетях.

единичный коэффициент корреляции, конечно, не может свидетельствовать о наличии статистической связи и надежности прогнозирования.

С целью придания большей универсальности модулю прогнозирования на основе нейронных сетей было решено использовать различные типы аксонов (пользовательский параметр). Это, в частности, наиболее популярный в различных приложениях аксон сигмоид, а также линейный аксон, который в частном случае соответствует простой линейной регрессионной модели, что позволяет тестировать процесс обучения нейронной сети и верифицировать результаты прогнозирования.

Текущее состояние нейрона определяется как взвешенная сумма его входов

$$s = \sum_{i=0}^{10} x_i \cdot w_i \text{ или } s = \sum_{i=0}^{10} x_i^2 \cdot w_i. \quad (8)$$

Выход нейрона-аксона есть функция его состояния

$$Y = f(s). \quad (9)$$

Функция f называется активационной и может иметь различный вид. Одной из наиболее распространенных активационных функций является нелинейная функция с насыщением (сигмовидная функция, или сигмоид), показанная на рис. 4. Аналитический вид этой функции

$$f(s) = \frac{1}{1 + e^{-\alpha s}}. \quad (10)$$

При уменьшении α сигмоид становится более пологим, в пределе при $\alpha = 0$ вырождаясь в горизонтальную линию на уровне 0.5, при увеличении α сигмоид приближается по внешнему виду к функции единичного скачка. Из выражения для сигмоида очевидно, что выходное значение нейрона лежит в диапазоне $[0, 1]$. На выходе нейронной сети (см. рис. 3) находится один нейрон (аксон), значение функции активации которого и есть прогнозируемое значение. Заметим, что конкретное значение константы α в алгоритме несущественно, так как его выбор соответствует просто перенормировке весов (синапсов).

Верификация разработанного нами модуля прогнозирования по нейронным сетям путем сравнения с уже сделанными прогнозами на основе классичес-

кого регрессионного атрибутивного анализа позволила сделать следующие выводы.

Обычно (примерно в 80–90 % случаев) нейронные сети дают более высокий коэффициент корреляции Пирсона и меньшее среднеквадратичное отклонение (RMS). Более того, при использовании сигмоида прогнозная карта получается в заданных пределах, в отличие от регрессионной прогнозной карты, значения на которой часто могут быть нереалистичными (например, значительные отрицательные значения при прогнозе эффективных толщин). При этом использование одних и тех же атрибутов приводит к визуально достаточно близким прогнозам по нейронным сетям и по регрессионным моделям. Казалось бы, это свидетельствует о явном преимуществе нейронных сетей.

Однако в ряде случаев прогнозная карта, полученная при использовании нейронной сети, оказывается очень контрастной. На гистограмме распределения значений на карте наблюдаются два ярко выраженных максимума, соответствующих выбранным пользователем минимуму и максимуму (что обеспечивает высокий коэффициент корреляции, так как скважины разбиваются на кросс-плоте на две плотные группы – с самыми маленькими и самыми большими значениями). В некотором смысле этот эффект близок к так называемому эффекту паралича сети, при котором рабочая точка на сигмоиде смещается в область насыщения, производная стремится к 0, и процесс поиска минимума целевой функции методом градиентного спуска практически останавливается. Есть и другая теоретическая и хорошо известная опасность при работе с нейронными сетями и обучении обратным распространением, связанная с возможностью попадания в локальные экстремумы целевой функции. Существуют методы борьбы с этим явлением, один из которых – добавление в метод градиентного спуска “инерции” – смещение происходит не по градиенту на текущем шаге, а по его сумме с градиентом на предыдущем шаге (естественно, с незначительным весом – у нас пользовательский параметр “инерция”).

В силу отмеченных особенностей нейронных сетей следует очень тщательно и критически подходить к оценке полученных прогнозов, проводить ее не только по формальным критериям, но и с использо-

ванием всей имеющейся априорной геолого-геофизической информации. При оценке надежности прогнозирования (вероятности ошибки второго рода) следует учитывать параметры нейронной сети (которые часто в интерпретационных пакетах закрыты для пользователя). Прежде всего это касается числа синапсов, так как оно увеличивает число степеней свободы, которое необходимо знать (например, при использовании табл. 1). При оценке надежности прогнозирования на основе эффективного коэффициента корреляции (5) следует иметь в виду, что он учитывает число используемых атрибутов, а не степеней свободы, которых у нейронной сети может быть существенно больше. *Поэтому мы считаем (и реализовали это в нашей программе), что число синапсов должно равняться числу используемых атрибутов. В этом случае эффективный коэффициент корреляции работает корректно.* В противном случае требуется введение дополнительных критериев надежности прогнозирования по нейронным сетям или, как минимум, корректное использование табл. 1 с учетом известного количества степеней свободы.

Таким образом, мы пришли к основному выводу, что нейронные сети являются эффективным, но довольно тонким инструментом прогнозирования и более сложным для объективной априорной оценки качества прогнозирования по сравнению с регрессионным анализом.

ДОПОЛНИТЕЛЬНЫЕ ТЕХНИЧЕСКИЕ ТОНКОСТИ АТРИБУТНОГО ПРОГНОЗИРОВАНИЯ

1. Прежде всего необходимо тщательно подойти к анализу “кондиционности” используемых атрибутов. При работе со срезами сейсмического куба следует убрать (вырезать) краевую зону, если на ней имеют место ярко выраженные краевые эффекты, иначе эта зона может повлиять на корректные расчеты факторов и кластеров сейсмического поля. Следует проанализировать атрибуты на предмет случайных значительных отскоков (выбросов) и, при необходимости, убрать их медианным фильтром. Подчеркнем, что во многих интерпретационных пакетах визуализация карт не совсем честная (не точка в точку) и используется некоторое подглаживание, поэтому такие дефекты часто просто не видны. Заметим также, что бороться с этими выбросами сглаживанием атрибутов не следует, так как ошибочный “выскоч” при этом

просто “размазывается” на большую площадь. В ряде случаев следует пропустить некоторые дифференциальные процедуры (например, рассчитать оператор Лапласа) при анализе качества используемых атрибутов. Это позволит выявить локальные резкие изменения, например, связанные с ошибками корреляции горизонтов, и т. п. Существенным моментом, уже много лет включенным в основной граф сейсмической обработки, является подавление футпринтов системы наблюдения (видимых на седиментационных срезах как регулярная прямоугольная сетка). Подавление помехи осуществляется в спектральной области, в которой она локализуется в виде отдельных регулярных максимумов [Масюков и др., 2010; Shlionkin et al., 2008].

Только после такой тщательной проверки (контроля качества) можно переходить к анализу статистической связи атрибутов и предсказываемого петрофизического параметра. Безусловно, надо понимать, что возможности интерпретации зависят от качества исходных данных (рис. 5), поэтому желательно максимально повысить соотношение “сигнал–шум” и с помощью препроцессинга атрибутного анализа.

2. Характер связи между параметрами коллектора и сейсмическими атрибутами может меняться по площади. Это проявляется в тех случаях, когда на кросс-плоте (“параметр коллектора – атрибут”) можно выделить несколько регрессионных зависимостей [Шленкин и др., 2005а], т. е. точки ложатся не на одну прямую, а на несколько прямых. Более того, для разных групп скважин корреляция может быть высокой с разными наборами атрибутов. Поиск нескольких зависимостей следует проводить на основе объективных критериев для выделения нескольких групп скважин, например, на основе кластерного анализа, как это показано на рис. 6. Категорически нельзя выделять несколько групп скважин по принципу “выше–ниже” линии регрессии (если для этого нет серьезных геологических обоснований).

После получения прогнозных карт по двум разным регрессиям необходима сшивка прогнозных карт на основе карт весов по формуле

$$RESULT = MAP1 * W1 + MAP2 * W2, \quad (11)$$

где $MAP1$, $MAP2$ – прогнозныe карты, полученные по первой и второй группам скважин соответственно, а $W1$ и $W2$ – карты весов, которые в сумме дают еди-

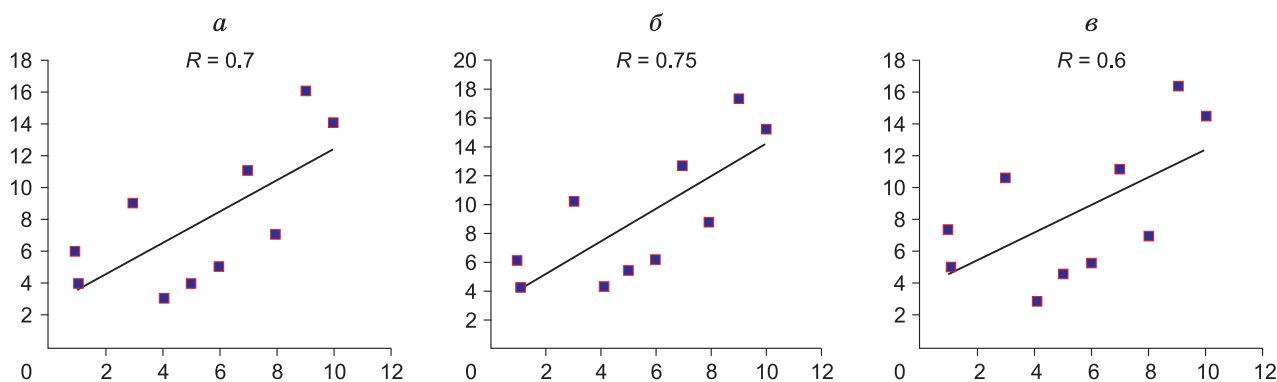


Рис. 5. Иллюстрация важности препроцессинга при атрибутном анализе. Зависимость петрофизического параметра по десяти скважинам от некоторого атрибута (а); в атрибут внесен случайный аддитивный шум (12 %, что соответствует соотношению сигнал–помеха для футпринтов системы наблюдений) (б, в). То есть при наличии шума мы можем как принять данную зависимость, так и отвергнуть ее в соответствии с табл. 1 при 5%-м уровне значимости.

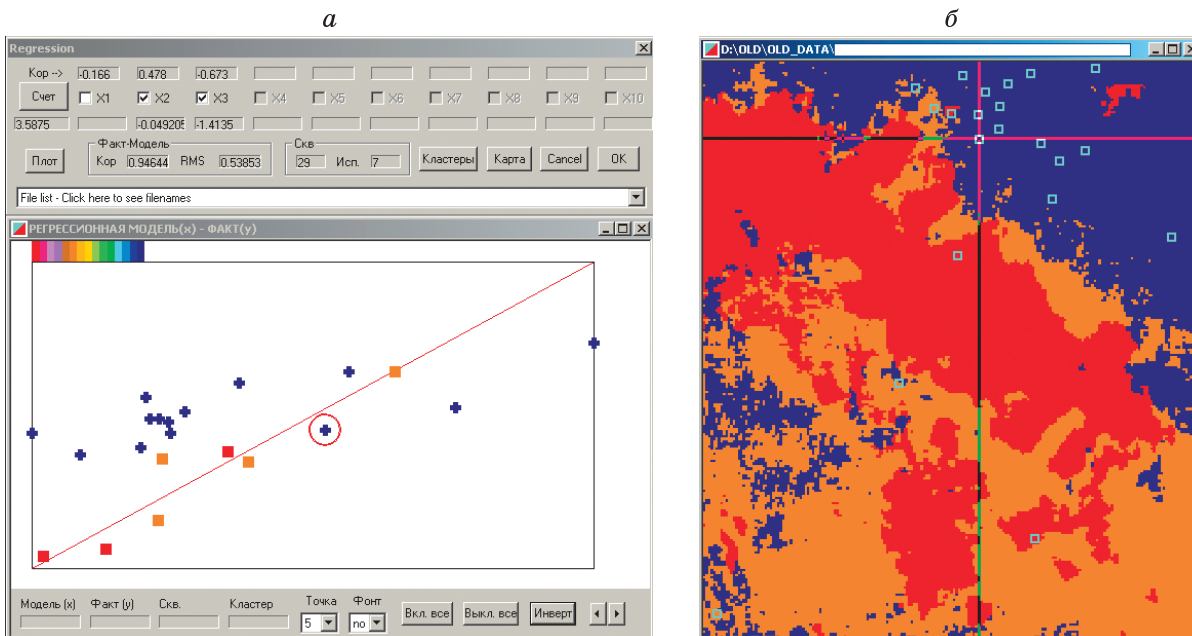


Рис. 6. Характерный кросс-плот (а) эффективной толщины коллектора и двух факторов сейсмического поля, на котором целесообразно выделение двух разных регрессионных зависимостей на основе кластерного анализа сейсмического поля (б). Скважины, попавшие в синий кластер, исключены при расчете кросс-плота (а), так как они имеют свой линейный тренд.

ническую карту, т. е. $W1 + W2 = I$. Карту $W1$ можно построить простой интерполяцией (“посадкой”) нулевой карты на единичные значения для первой группы скважин и нулевые значения для второй группы скважин. Заметим, что тогда вторую карту весов можно получить по формуле $W2 = I - W1$.

Иногда интерпретатор получает несколько альтернативных прогнозов, выполненных при использовании различных атрибутов на основе регрессионных моделей или нейронных сетей. При этом объективные оценки надежности, информативности и точности прогнозов достаточно близкие. В таком случае мы советуем тщательно проанализировать полученные результаты на предмет соответствия априорной геолого-геофизической информации и устроить “консилиум”, на котором принять экспертное решение, а именно: сложить полученные прогнозные карты с некоторыми весами (в сумме равными 1 для сохранения “посадки” на скважинные данные).

3. Одним из ключевых моментов любого прогноза является фитинг – “посадка” прогнозной карты на фактические значения в точках скважин (интерполяция невязок). Выбор алгоритма интерполяции при фитинге чрезвычайно важен [Масюков и др., 2006]: цена вопроса на площади в 100 км² – сотни тысяч тонн нефти. В работах [Масюков А.В., Масюков В.В., 2005; Шленкин и др., 2005б; Masjukov A.V., Masjukov V.V., 2005] показаны преимущества новых скейлинговых итерационных методов интерполяции, направленных на автоматическое обнаружение и использование разномасштабных пространственных трендов исходных данных. Заметим, правда, что некоторые до сих пор используемые методы скейлинговых интерполяций (типа Convergence) не выдерживают никакой критики из-за принципиальных ошибок алгоритма (несимметричности результата при про-

странственной симметрии исходных данных). В [Масюков и др., 2008] также показано, что любые методы интерполяции, содержащие масштабные параметры (включая крайгинги), неизбежно порождают артефакты при неравномерном расположении скважин по площади (рис. 7).

В качестве примера использовалась экспоненциальная модель вариограммы $C(h) = (1 + h/\sigma) \exp(-h/\sigma)$. На отрезке (1, 3) крайгинг при любой длине автоковариации $\sigma > 0.15$ дает хорошую аппроксимацию модельной функции, так как на этом отрезке плотность данных удовлетворяет теореме отсчетов. На отрезке (7, 10) приемлемая аппроксимация получается при $\sigma > 0.5$. Однако на отрезке (4, 6) крайгинг-интерполяция очень плохая – при любом σ . Только мультимасштабная интерполяция дает везде приемлемую аппроксимацию.

4. Нужно ли точно “сажать” прогнозные карты на скважины? Этот вопрос подробно рассмотрен в статье [Шленкин и др., 2005а]. Показано, что при точной “посадке” на все скважины (например, структурного плана) при увеличении разбуренности структурный план не уточняется, а постоянно меняется. При высокой плотности эксплуатационных скважин в структурном плане присутствуют высокочастотные составляющие, которые не могут найти геологического объяснения (часто их просто не должно быть). Причина очевидна: данные инклинометрии – координаты пластопересечений и глубинные отметки – имеют некоторые ошибки. Поэтому вместо точной “посадки” целесообразно использовать аппроксимационный фитинг (приближенную “посадку” на скважинные данные). Пользователь задает базу аппроксимационного фитинга, при этом “посадка” на одиноко стоящие скважины (удаленные от других на расстояние, большее базы фитинга) будет точной, на близко сто-

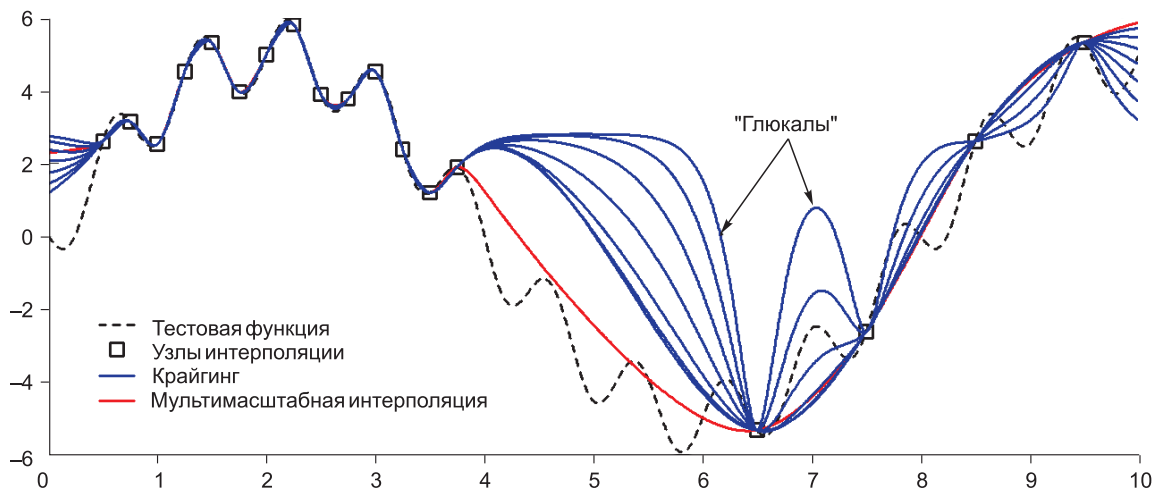


Рис. 7. Иллюстрация в одномерном тесте неизбежного возникновения артефактов при интерполяции с масштабным параметром. Только мультимасштабная интерполяция дает везде приемлемую аппроксимацию.

ящие скважины с сильно отличающимися значениями – неточной, аппроксимационной, т. е. поверхность пройдет по тренду, который определяется этими скважинами. Таким образом, на карте не будут присутствовать высокочастотные эффекты, связанные с резкими перепадами значений в близкорасположенных скважинах. По существу база фитинга определяется тем, на каком расстоянии от скважины мы можем принимать во внимание (учитывать) полученные в ней значения. Конкретный выбор значения базы аппроксимационного фитинга зависит от конкретных геологических условий (для морских, выдержанных по мощности, отложений он может быть около 500 м, для изменчивых, континентальных отложений – существенно меньше).

Заметим, что для устойчивого определения тренда и значений прогнозной карты в краевой области целесообразно привлекать скважинные данные с соседних площадей. Для этого перед “посадкой” (фитингом) необходимо расширить площадь карты путем экстраполяции, выполнить фитинг на полный сет скважин (с учетом не попадающих в область данной съемки), а затем просто обрезать финальную карту.

5. В свете рекомендаций [Левянт и др., 2006] нам представляется целесообразным вынесение на финальные прогнозные карты областей (контуров) “неудовлетворительного”, “удовлетворительного” и “хорошего” прогноза (относительная ошибка ε больше 50 %, от 50 до 20 %, меньше 20 % соответственно). Технически это выполняется следующим образом. Сначала для получения карты относительных ошибок нужно разделить константную карту со значением, равным RMS прогноза во всех точках карты, на саму полученную прогнозируемую карту. Затем сделать локальный фитинг полученной карты на нулевые значения в точках скважин для учета эффекта “посадки” прогнозной карты на скважинные значения. Значения в точках скважин на карте относительных ошибок должны оказаться, таким образом, нулевыми (при точном фитинге) и близкими к 0 (при аппроксимационном фитинге). База локального фитинга выбирается исходя из геологических особенностей предсказываемого параметра (например, его латеральной из-

менчивости – для морских отложений база “посадки” (влияния скважины) может быть большой, для континентальных отложений – малой). Фактически база показывает, на каком расстоянии от скважины мы можем принимать во внимание измеренные в ней значения. С полученной таким образом карты относительных ошибок прогноза (с учетом “посадки” на скважины) снимаются линии уровня (в 20 и 50 %) и наносятся на финальную прогнозируемую карту.

КАЧЕСТВО ПРОГНОЗА И ЕГО ПОДТВЕРЖДАЕМОСТЬ

Понятия “качество прогноза” и “подтверждаемость прогноза” часто используются очень вольно [Медведев, 2004]. Пусть для определенности строится прогноз эффективной толщины и в качестве результата прогнозирования рассматривается карта. При малом числе скважин, большой области прогнозирования и небольшом среднем значении эффективной толщины по площади (менее 5 м) качество прогноза объективно не может быть очень “хорошим” (даже при самом тщательном и профессиональном подходе). Если слово “хорошо” относится к оценке самого прогноза, а не оценивать корректность и правильность используемых технологий. По существу качество прогноза определяется картой относительных ошибок, построение которой описано выше. Для количественного описания качества прогноза мы предлагаем использовать отношение общей площади областей “хорошего” ($\varepsilon < 20\%$) и “удовлетворительного” ($50\% > \varepsilon > 20\%$) прогноза к общей площади области прогнозирования (при положительных прогнозных значениях эффективной толщины). В общую площадь, по-видимому, не следует включать области, в которых прогноз нулевой (например, зоны глинизации). Предложенная оценка качества прогноза является объективной. Уменьшение среднеквадратичной ошибки прогноза приводит к увеличению областей “хорошего” и “удовлетворительного” прогноза и к улучшению качества прогноза. Увеличение разбуренности площади, как правило, также приводит к повышению качества прогноза, так как должны увеличиваться области “хорошего” и “удовлетворительного”

прогноза (хотя бы из-за посадки на вновь пробуренные скважины). Из предложенного определения видно, что качество прогноза меняется от 0 до 1 (от 0 до 100 %). В пределе (при увеличении плотности скважин или уменьшении RMS) оно стремится к 100 %. Нам представляется, что эта характеристика дополняет введенные выше характеристики “информативности”, “надежности” и “условной точности” прогноза. Заметим, что ее нельзя использовать независимо, без учета, например, “надежности” найденной корреляционной связи. Действительно, если взять число регрессоров, равное числу скважин, то коэффициент корреляции будет равен 1, $RMS = 0$, относительная ошибка равна 0, и вся прогнозная карта будет являться областью “хорошего” прогноза, а следовательно, качество прогноза будет равно 100 %. Корректная оценка “надежности” в рассмотренном случае делает невозможным использование такого прогноза в принципе.

Очевидно, что подтверждаемость прогнозирования является статистической оценкой (нельзя по одной скважине говорить о “хорошей” или “плохой” подтверждаемости). Для оценки подтверждаемости при мониторинге бурения обычно используются оценки невязки “факт–прогноз” – среднеквадратичные и средние относительные ошибки. Но значение среднеквадратичной ошибки следует соотносить с мощностью пласта, а относительная ошибка при вскрытой нулевой толщине и ненулевыми прогнозными значениями равна бесконечности. Поэтому для исключения данного эффекта все полученные в результате бурения невязки суммируются (по модулю), суммируются также все вскрытые толщины, а затем вычисляется общая относительная ошибка. Эти оценки, безусловно, являются объективными и полезными, но они оценивают результат бурения, не учитывая объективную неопределенность самого прогноза, которая количественно выражается в виде карт относительных ошибок прогноза. Мы предлагаем для оценки подтверждаемости использовать новую величину

$$G = \frac{N_{good}}{N} 100 \% , \quad (12)$$

показывающую процентное отношение числа скважин N_{good} с “хорошим” соответствием “факт–прогноз” (в которых фактическая относительная ошибка прогноза оказалась меньше значения с карты относительных ошибок прогноза) к общему числу пробуренных новых скважин N . Эта величина (в отличие от традиционных) учитывает, как наблюденная невязка “факт–прогноз” соотносится с неопределенностью самого прогноза в этой точке. Например, если скважина пробурена в точку, в которой значение с карты относительных ошибок прогноза равно 80 %, а реально полученная относительная ошибка составила 50 %, то это должно рассматриваться как “хорошее” соответствие “факт–прогноз”, и согласно (12) увеличивать подтверждаемость. Нам представляется это логичным, так как в рассмотренном примере скважина бурилась в разведочных целях в область высокой неопределенности прогноза. Заметим, что использование традиционных подходов приводит к ухудшению подтверждаемости в этом примере.

Не секрет, что иногда прогнозные значения эффективной толщины умышленно занижаются для того, чтобы результат бурения оказывался лучше прогноза. Совместное использование предложенных ха-

рактеристик препятствует реализации такого подхода, так как при занижении прогнозных значений уменьшаются области “хорошего” и “удовлетворительного” прогноза, и, следовательно, понижается его качество.

Важно понимать, что сама по себе карта относительных ошибок прогноза имеет качественный характер, так как абсолютная ошибка оценивается через RMS. Можно было бы через RMS рассчитывать доверительный интервал с заданным уровнем доверительной вероятности, но это все было бы приближенно, качественно, так как опирается на некоторые статистические гипотезы, далекие от реальности. Поэтому оценки качества и подтверждаемости также достаточно условны. На самом деле важны не конкретные значения с карты относительных ошибок и не конкретные числовые значения введенных характеристик. Важно, чтобы они работали “логично”, с точки зрения специалиста-практика, соответствовали предельным случаям и давали возможность единообразного сравнения различных прогнозов.

Безусловно, целесообразность широкого практического использования предложенных числовых характеристик атрибутивного прогнозирования должна быть проверена независимым анализом большого количества данных, а также дискуссией специалистов-практиков. Нам представляется важным введение количественных характеристик качества прогноза и его подтверждаемости и при 3D моделировании, но эта проблема требует отдельного рассмотрения и выходит за рамки настоящей работы.

ДИСКУССИЯ И ВЫВОДЫ

“...подрядчик не может гарантировать и не гарантирует точности, правильности и полноты каких-либо результатов интерпретации, рекомендаций и характеристик залежей...”

*Из отчета иностранной компании,
работающей на российском
нефтегазовом рынке*

Мы уверены, что необходимо повсеместное использование “эффективного коэффициента корреляции”, так как он учитывает неравномерность расположения скважин на кросс-плоте (в отличие от других оценок статистической связи, например коэффициента Пирсона или Спирмена). Именно данный коэффициент должен быть использован для оценки надежности прогнозирования (особенно при выполнении прогноза на основе нейронных сетей). Авторы статьи считают, что для повышения доверия к прогнозированию на основе нейронных сетей пользователь должен иметь доступ к их подробному описанию и знать конкретную архитектуру сети при выполнении своего прогноза.

Предложенные в данной статье удобные оценки “условной точности прогноза”, “информативности” и “надежности статистической связи”, как нам представляется, должны сопровождать (помимо кросс-плотов и традиционных оценок корреляционной связи и соотношений “факт–прогноз”) все практические результаты атрибутивного прогнозирования для возможности более полной их объективной оценки и сравнения.

Чрезмерное увлечение некоторыми специалистами “трехмерным” прогнозированием значительных по размеру объектов при малом количестве скважинных данных вызывает у нас серьезную обеспокоенность. Во-первых, многие используемые для построения

трехмерных моделей алгоритмы являются псевдотрехмерными (например, основаны на двумерной интерполяции по пропласткам в предположении о конформном или пропорциональном залегании пластов). Во-вторых, 3D-моделирование часто основано на инверсии сейсмических данных, которая сама по себе является некорректной задачей (неединственность решения). При этом результаты существенно зависят как от выбора алгоритма инверсии, так и от методики ее реализации. Вместе с тем очевидно, что двумерный прогноз (карт) эффективных характеристик объектов (эффективной толщины, эффективной пористости и т. п.) более надежен, чем построение кубов литологии, пористости и т. п. Если с трудом (на пределе пороговых значений уровня значимости) удастся получить двумерные прогнозные карты интегральных характеристик пласта для 5–10 скважин, то надежность трехмерных картинок, несомненно, является крайне неудовлетворительной. Возникает ощущение, что, понимая тщетность попыток трехмерного прогнозирования с высоким пространственным разрешением, некоторые разработчики просто пытаются прикрыть это “геостохастикой”.

Опыт работы с различными заказчиками показывает, что техническое задание для сейсмической обработки и интерпретации формируется часто в соответствии с их вкусовыми пристрастиями. В технических заданиях нередко фигурируют модные, но трудоемкие наукообразные процедуры, ненадежные и неоднозначные, не дающие серьезного профицита для геолого-геофизического прогнозирования. Например, часто требуется не только проведение инверсии, но и пересчет куба акустических импедансов в куб плотности, пористости, литологии, и все это – для юрских отложений, для целевого пласта, который практически не разрешается в сейсмической записи и интегральные характеристики которого (эффективная толщина, эффективная пористость) предсказываются с большим трудом. Зачастую отсутствуют четко сформулированные требования к представлению результатов прогнозирования. Поэтому исполнитель просто формально делает инверсию в соответствии с техническим заданием, при этом результат практически не важен. Затем исполнитель, построив некоторую нелинейную зависимость по 4–6 точкам, приводит кросс-плот, “висящий” на одной точке, и пересчитывает по найденной (крайне ненадежной) зависимости один куб из другого, получая красивые трехмерные картинки. Полученные из кубов петрофизических характеристик карты, характеризующие целевой интервал, не будут соответствовать прогнозам интегральных характеристик, а часто будут просто противоречить элементарным геологическим закономерностям. Наиболее красноречиво о таком положении дел говорит эпиграф к этому разделу. Можно ли что-то изменить?

Мы уверены, что только государственное вмешательство может изменить ситуацию в этой важной сфере народного хозяйства, связанной с поиском и добычей углеводородного сырья. Если Министерством природных ресурсов будут утверждены не просто рекомендации, но и обязательные требования (аналог ГОСТов качества) к представлению результатов прогнозирования, и если за соблюдением этих требований будет следить Государственный комитет по запасам, то ситуация может измениться к лучшему. Естественно, к разработке такого рода требований должно быть привлечено широкое научное сообщество

специалистов-практиков. Соблюдение этих условий приведет, с одной стороны, к упрощению, ясности, прозрачности используемых методик и подходов (так как позволит объективно их сравнивать), с другой стороны, к исключению из практики технологий “мыльных пузырей” – красивых картинок, полученных специфическими и крайне неустойчивыми процедурами. Заметим, что после разработки и утверждения критериев контроля качества реализовывать их расчет в виде макроса MS Excel или автономной программы для свободного распространения технически очень просто. Повышение эффективности и качества прогнозирования при этом очевидно, затрат же практически нет.

В заключение хотелось бы обсудить мнение, звучащее в последнее время на многих геофизических конференциях, – о целесообразности развития и других некоммерческих отечественных программных продуктов, включающих лучшие достижения отечественной математики и геофизики в области обработки и интерпретации сейсмических данных. Это является, с нашей точки зрения, крайне актуальной государственной задачей, при этом не требует значительных капиталовложений. Ее решение позволит обеспечить всех специалистов, работающих на российском нефтегазовом рынке, самым современным, бесплатным и единым базовым программным арсеналом.

В течение многих лет мы имеем счастливую возможность сравнивать технические процедуры обработки и интерпретации сейсмических данных, которые реализованы в различных (самых современных) программных продуктах. В результате такого сравнения мы пришли к выводу, что в подавляющем большинстве случаев отечественные программные разработки по техническим параметрам не уступают западным программным продуктам. Объективным минусом отечественных разработок пока является их слабая интегрированность (зарубежные программные продукты позиционируются как “умеющие все”: от обработки до интерпретации, построения гидродинамической модели, оценки рисков, сопровождения бурения и т. п.). Использование отечественных автономных программ существенно повышает производительность обработки и интерпретации сейсмических данных, несмотря на некоторые затраты времени для загрузки-выгрузки данных. Поэтому нам представляется чрезвычайно важным объединить все современные передовые отечественные программные разработки с целью их свободного распространения. Для реализации этой идеи необходима последовательная и повсеместная борьба с лоббированием интересов иностранных компаний, с общепринятой практикой “откатов” при закупке программного обеспечения, а также выполнение положений Этического кодекса геофизиков России (протокол № 9 ЕАГО от 26 августа 2004 г.), в частности призыва: “Помни о своей принадлежности к отечественной геофизике, уважай и преумножай ее славные традиции. Будь объективен и критичен в своих оценках зарубежных конкурентов”.

Литература

- Ампилов Ю.П. Сейсмическая интерпретация: опыт и проблемы. М.: Геоинформмарк, 2004. 277 с.
- Левянт В.Б., Ампилов Ю.П., Глоговский В.М. и др. Методические рекомендации по использованию данных сейсморазведки 2D, 3D для подсчета запасов нефти и газа. М.: СГЭ, 2006.

- Левянт В.Б., Хромова И.Ю., Козлов Е.А. и др.** Методические рекомендации по использованию данных сейсморазведки для подсчета запасов углеводородов в условиях карбонатных пород с пористостью трещинно-кавернозного типа / Федеральное агентство по недропользованию. М., 2010.
- Масюков А.В., Масюков В.В., Шленкин В.И.** Атрибутное прогнозирование: алгоритм интерполяции // Пути реализации нефтегазового потенциала ХМАО. 2004. Т. 2. С. 289–294.
- Масюков А.В., Масюков В.В.** Итерационный метод интерполяции, основанный на масштабируемом сглаживании // Математическое моделирование. 2005. Т. 17, № 2. С. 46–56.
- Масюков А.В., Масюков В.В., Шленкин В.И.** Повышение точности геолого-геофизического моделирования на основе скейлинговой интерполяции // Каротажник. 2006. № 2–4. С. 143–145.
- Масюков А.В., Масюков В.В., Шленкин В.И.** Проверка значимости корреляционных связей в геолого-геофизическом прогнозировании // Технологии сейсморазведки. 2007. № 1. С. 80–86.
- Масюков А.В., Масюков В.В., Шленкин В.И.** Крайгинг как частный случай радиальных базисных функций // Докл. X юбилейной Междунар. науч.-практ. конф. “Геомодель-2008” (21–26 сент. 2008 г., Геленджик). Геленджик, 2008. С. 116–119.
- Масюков В.В., Масюков А.В., Шленкин В.И.** Совершенствование технологии подавления регулярных фут-принтов, связанных с системой сейсмических наблюдений // Докл. XII Междунар. науч.-практ. конф. “Геомодель-2010” (13–17 сент. 2010 г., Геленджик). Геленджик, 2010. С. 24–27.
- Медведев А.Л., Федоров В.В., Масюков В.В.** Оценка точности и подтверждаемости атрибутного прогнозирования параметров пластов-коллекторов // Технологии сейсморазведки. 2004. № 2. С. 105–109.
- Тейлор Дж.** Введение в теорию ошибок. М.: Мир, 1983.
- Шленкин В.И., Федоров В.В., Масюков А.В., Масюков В.В.** Об оценке точности корреляционного прогнозирования параметров геологического разреза // Геофизический вестник. 2004. № 4. С. 19–22.
- Шленкин В.И., Тюнегин С.П., Масюков А.В., Масюков В.В.** Методика уточнения структурных построений для эксплуатационных площадей Широкого Приобья на основе представлений о пространственном спектре ошибки сейсмического горизонта // Технологии сейсморазведки. 2005а. № 1. С. 57–66.
- Шленкин В.И., Федоров В.В., Масюков А.В., Масюков В.В.** Методика объективного сравнения методов интерполяции // Геофизический вестник. 2005б. № 1. С. 17–21.
- Шленкин С.И., Федоров В.В., Масюков В.В.** Исследование и совершенствование технологий атрибутного прогнозирования свойств терригенных коллекторов на основе данных ГИС и 3D-сейсморазведки // Каротажник. 2006. № 2–4. С. 454–468.
- Шленкин С.И., Масюков А.В., Масюков В.В., Козлова А.Н.** Вычисление кубов когерентности и сингулярностей // Технологии сейсморазведки. 2012. № 2. С. 5–11.
- Gurney K.** An introduction to neural networks. London: UCL Press, 1997. 234 p.
- Masjukov A.V., Masjukov V.V.** A new fast iterative method for interpolation of multivariate scattered data // Computational Methods in Applied Mathematics. 2005. V. 5, N 3. P. 1–18.
- Milligan S.D.** Statistical Grouping of Acoustic Reflection Profiles // J. Acoust. Soc. Am. 1978. V. 64. P. 795–807.
- Shlionkin S.I., Masjukov A.V., Masjukov V.V., Tararin G.V.** Seismic Automatic Suppression of Acquisition Footprints in Seismic 3D // International Conference SEG-EAGE-EAIG. Saint Peterburg, 2008. A024.

*Поступила в редакцию 10 июля 2013 г.,
в окончательном варианте – 6 августа 2013 г.*

КОРОТКО ОБ АВТОРАХ

МАСЮКОВ Владимир Владимович – кандидат физико-математических наук, доцент, ведущий математик-программист ООО “Славнефть – НПЦ”.
E-mail: masvol@mail.ru

ПОПОВ Михаил Альбертович – начальник отдела интерпретации ООО “Славнефть – НПЦ”.

ТЮНЕГИН Сергей Петрович – кандидат геолого-минералогических наук, главный геолог ООО “Славнефть – НПЦ”.