

Качество классификации данных в условиях противоречивости экспертных оценок

Я. И. Кучин*, Р. И. Мухамедиев**, ***, К. О. Якунин**

*Институт информационных и вычислительных технологий
050000, Казахстан, Алматы, ул. Пушкина, 125

**Казахский национальный технический университет имени К. И. Сатпаева,
050000, Казахстан Алматы улица Сатпаева 22а

***ISMA Высшая школа менеджмента информационных систем,
LV-1019, Latgales priekšpilsēta, Rīga, Lomonosova iela 1, k-6

e-mail: ykuchin@mail.ru, ravil.muhamedyev@gmail.com, yakunin.k@mail.ru

Аннотация. В процессе добычи урана методом подземного скважинного выщелачивания важное значение имеет правильная классификация литологических слоев на базе данных электрического каротажа. Для решения этой задачи в настоящее время часто применяется ручной труд. Сопутствующие ошибки, ведущие к экономическим потерям, делают актуальной автоматизацию данного процесса. Для автоматизации процесса интерпретации данных каротажа используются методы машинного обучения с учителем. Рассматривается проблема противоречивости экспертной интерпретации и связанные с этим ограничения в качестве работы нейронных сетей. Для оценки методов машинного обучения, в «идеальном случае», введено понятие «синтезированной» скважины, — набора специально сгенерированных данных, имитирующих показатели электрического каротажа и экспертное оценивание. На основе этих данных нейронные сети с прямым распространением сигнала показывают высокое качество классификации (около 96% точности). При использовании реальных данных показатель точности существенно ниже — около 55%. Приведены результаты оценки трех скважин тремя различными экспертами и оценено их согласие между собой и с данными кернового апробирования. Приведены качественные показатели интерпретации данных указанных скважин с помощью искусственной нейронной сети. Предложено использовать идентификатор эксперта в качестве дополнительного параметра при обучении нейронной сети.

Ключевые слова: машинное обучение, обучение с учителем, экспертное оценивание, «синтезированная» скважина.

1. Введение

Казахстан является мировым лидером в производстве урана, с 2009 г. обеспечивая более трети общего объема данного ресурса. При этом наблюдается постоянный

рост объемов производства — за последние семь лет производство урана в Казахстане возросло почти в 3,5 раза.

Добыча урана на месторождениях Казахстана ведется методом подземного скважинного выщелачивания, который относится к числу малозатратных, экологически безвредных способов добычи. При этом экономические показатели процесса добычи зависят от скорости и точности интерпретации геофизических данных. Ошибочный или неточный анализ геофизических данных приводит к потерям скважин, неоправданным трудозатратам и в конечном счете снижает экономические показатели добычи.

Использование информационных технологий в процессе интерпретации данных геофизического исследования скважин (ГФИС) урана позволяет отойти от традиционных методов сбора и обработки информации, решая проблему обработки больших объемов данных и уменьшая субъективность интерпретации отдельными экспертами. Однако классические программные средства, позволяющие автоматизировать процессы сбора, обработки и хранения данных ГФИС, не обладают достаточными диагностирующими способностями.

Возможность интерпретации результатов ГФИС может предоставить современные разработки в области так называемого слабого искусственного интеллекта (ИИ). Таким инструментом на текущий момент являются методы машинного обучения (МО), точнее, методы обучения с учителем (*supervised learning*), к числу которых относится несколько групп алгоритмов, включая и искусственные нейронные сети (ИНС). Несмотря на относительную простоту их построения и функционирования, системы машинного обучения позволяют накапливать уже известные закономерности ГФИС, обобщать факты и давать вполне корректные оценки в ситуациях, когда на входе представлены зашумленные данные.

На практике при применении МО для анализа данных внимание исследователей должно акцентироваться не только на самом процессе интерпретации, но и на этапах подготовки исходных данных. Качество, состав и формат входных данных напрямую влияют на процесс обучения системы и впоследствии — на качество классификации.

Наравне с предобработкой данных значительно улучшить результат позволяет постобработка данных интерпретации, учитывающая специфику данных ГФИС и физические особенности области исследования, результаты работы других алгоритмов распознавания, в том числе имитирующих методику распознавания экспертами. Правильная подготовка (предобработка) исходных данных для интерпретации и постобработка результатов интерпретации позволят повысить надежность литологической классификации, что, в свою очередь, улучшит экономические показатели всего процесса добычи урана. Применение методов МО в задаче класси-

фикации каротажных данных базируется на использовании экспертного оценивания. Другими словами, основанием для обучения автоматических классификаторов служат мнения экспертов. Однако сравнение мнений экспертов между собой и с данными кернового апробирования показывает существенное расхождение. Это расхождение тем выше, чем менее критична определяемая порода. В настоящей работе сделана попытка оценки данного расхождения и определения границ автоматической классификации при применении искусственных нейронных сетей (ИНС). Работа состоит из четырех разделов. В первом разделе приводится краткий литературный обзор и формулируются проблемы, возникающие при применении ИНС в задаче классификации каротажных данных на урановых месторождениях.. Во втором разделе описывается метод синтезирования каротажных данных. В третьем разделе показаны результаты применения некоторых методов машинного обучения к синтезированным данным. В четвертом разделе выполнена сравнительная оценка экспертного оценивания. В заключении обсуждаются полученные результаты и сформулированы рекомендации для проведения следующей серии вычислительных экспериментов.

2. Обзор

Начиная с 1970-х гг. искусственные нейронные сети стали применяться в задачах петрографии как средство анализа каротажных данных, в литологии, оценке минерально-сырьевой базы, сейсмическом зондировании [1] и т. п. [2–11]. Применению нейронных сетей в решении практических задач интерпретации каротажных данных в области нефтедобычи посвящена работа [12]. В работах [13–15] описаны некоторые результаты применения нейронных сетей прямого распространения для интерпретации данных геофизического исследования скважин при добыче урана.

Отметим, что добыча урана на месторождениях Казахстана ведется методом подземного скважинного выщелачивания, который относится к числу малозатратных, экологически безопасных способов добычи [16].

При этом экономические показатели процесса добычи зависят от скорости и точности интерпретации геофизических данных. В основном применяются электрические методы: каротаж методом кажущихся сопротивлений (КС), методом потенциалов самопроизвольной поляризации (ПС) и индукционный каротаж (ИК). Широкое применение кернового опробования при анализе пород невозможно в силу медленности процесса получения данных. Результаты каротажа представляются в виде каротажных диаграмм, на основании которых эксперт делает заключение о глубине залегания и качестве пород.

Ошибочный или неточный анализ геофизических данных приводит к потерям скважин, неоправданным трудозатратам и в конечном счете снижает экономиче-

ские показатели добычи. Поскольку интерпретация данных каротажа носит во многом эмпирический характер, т. е. точные закономерности отсутствуют, становится очевидным возможность применения обучаемых систем, в частности нейронных сетей. Однако, несмотря на достоинства ИНС, главное из которых — это способность ИНС решать слабоформализованные задачи [17], в процессе их использования имеются следующие проблемы:

1. Неоднозначность мнений экспертов.
2. Необходимость большого и равного количества примеров из разных классов.
3. Невозможность нейронной сети объяснить полученный результат.
4. Необходимость тщательной предварительной подготовки данных (очистка от аномальных значений, нормирование, сглаживание).

Несмотря на указанные проблемы, задача автоматической интерпретации является крайне актуальной с учетом большого числа пробуриваемых скважин и требований по оперативной обработке данных. По приблизительным оценкам, возможный экономический результат при использовании системы автоматической интерпретации данных каротажа на месторождениях Казахстана может составить до 2 млн долл. США в год (в 2014 г.). Данная оценка получена на основе приблизительной стоимости одной скважины (около 40 тыс. долл.) и примерного количества скважин, пробуриваемых ежегодно (от 1 до 2 тыс.). Предполагая возможность неправильной интерпретации данных и, соответственно, потери скважин в пределах 1%, можно оценить экономические потери величиной около 2 млн долл. При точной интерпретации потери скважин могут существенно снизиться, что и приведет к указанному выше экономическому эффекту.

В ряде работ авторов начиная с 2011 г. рассматриваются задачи и проблемы автоматической интерпретации каротажных данных на урановых месторождениях. В работах [19, 20] приведены результаты вычислительных экспериментов с применением нейронных сетей в качестве алгоритма классификации каротажных данных. В работах [21, 22] рассмотрены несколько методов машинного обучения и приведены сравнительные результаты. В частности, показано, что нейронные сети прямого распространения показывают более высокое качество классификации по сравнению с алгоритмами kNN и SVM. В работах [23, 24] предложен подход, основанный на ансамбле алгоритмов машинного обучения для решения указанной задачи. Перечисленные выше проблемы 2 и 4 нашли практическое решение в процессе вычислительных экспериментов. В настоящей работе уделено внимание проблеме 1 — оценке влияния неоднозначности экспертного оценивания результатов электрического каротажа на обучение автоматических классификаторов, более конкретно, — нейронных сетей.

3. Синтезированная (искусственная) скважина

В силу особенностей классификации каротажных данных точная оценка качества классификации на основании объективных данных невозможна, поскольку получение точных данных о реальном распределении пород вдоль оси скважины затруднено. Керновое опробование делается не для всех скважин и не по всей глубине интерпретации. По этой причине система машинного обучения (классификации) опирается на данные экспертного оценивания. То есть система классификации обучается в основном, используя в качестве правильных ответов мнения экспертов. При этом, однако, возникает проблема оценки качества самой экспертной интерпретации, ее непротиворечивости.

Проведенные ранее эксперименты показали в среднем невысокое качество классификации при использовании нейронных сетей прямого распространения (точность порядка 65%). Возникает вопрос: связано ли это с особенностями самих нейронных сетей в задаче литологической классификации или это является следствием противоречивости в экспертном оценивании?

Для анализа возможностей алгоритмов машинного обучения можно синтезировать некоторые данные, соответствующие по тем или иным параметрам реальным данным каротажа. Используя эти синтезированные данные, удовлетворяющие некоторым базовым физическим принципам работы каротажных приборов и свойств вмещающих пород, можно проверить качество системы машинного обучения. Разумеется, такая оценка не претендует на исчерпывающую полноту, но в то же время может дать представление о предельных возможностях алгоритмов и некоторое, косвенное, представление об экспертных оценках. С другой стороны, еще один путь, позволяющий оценить границы противоречивости экспертного оценивания, заключается в сравнении экспертных оценок нескольких экспертов.

Единственным способом прямого и достоверного определения литологического состава пород вдоль ствола скважины является отбор керна. Однако и он не дает стопроцентно достоверной информации, поскольку процент извлечения, как правило, не превышает 80% для осадочных пород, характерных для урановых месторождений Казахстана. Кроме того, погрешности возникают при привязке керна по глубине, а также при самом описании данных керна. Эти погрешности трудно определить и учесть. Ввиду отсутствия достоверных сведений о литологическом строении разреза ствола скважины, когда и данные электрического каротажа, и отбор керна дают лишь приблизительные значения, представляется перспективным моделирование записанного сигнала каротажа при заданном распределении пород вдоль ствола скважины с известными физическими свойствами. Это позволит получить набор пусть искусственных, но полностью определенных значений каротажа, четко «привязанных» к определенным породам.

Физические свойства каждой породы, в частности кажущееся сопротивление, лежат в определенном диапазоне. Распределение внутри диапазона может быть изучено на этапе разведки и в лабораторных исследованиях.

На этапе разведки отбирается керн, проводятся лабораторные анализы, изучаются физические свойства горных пород, в том числе характерные значения кажущегося сопротивления (КС) и для каждой породы [25, 26]. Типичное распределение показано на рис. 1. (График построен специалистами ТОО «Геотехно-сервис»).

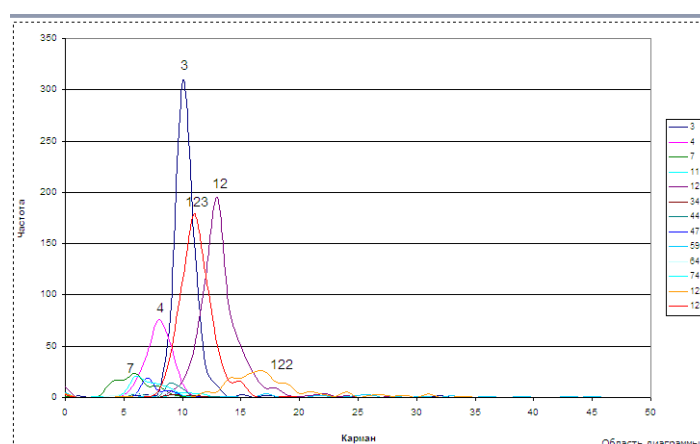


Рисунок 1. Распределение литотипов по значению КС

Можно отметить, что интервалы значений кажущегося сопротивления для каждого литотипа значительно перекрываются. Также видно, что для построения модели в первом приближении можно взять нормальное распределение внутри интервала. Со значениями ПС все несколько сложнее, поскольку они сильно зависят от свойств бурового раствора и минерализации грунтовых вод.

Задав распределение мощности пластов, для каждого 10-сантиметрового пропластка внутри пласта выбирается кажущееся сопротивление из заданного диапазона с соответствующим распределением, что позволяет получить кривую распределения кажущегося сопротивления вдоль ствола скважины. Для моделирования зарегистрированной кривой КС необходимо также учитывать параметры скважинного прибора (тип зонда, расстояние между электродами), диаметр скважины, свойства бурового раствора и др. В итоге можно получить смоделированную запись каротажа сопротивлений, которая соответствует заданному распределению пород, при этом достоверность информации о распределении пород будет известна со стопроцентной точностью. Аналогичным образом моделируется и кривая ПС. Это дает возможность проверить работу различных алгоритмов машинного обуче-

ния и улучшить их. Максимальное приближение этой модели к реальным условиям позволит повысить качество распознавания на реальных данных и определить верхний предел точности распознавания.

На первом этапе построения модели генерируется набор пластов исходя из типичных мощностей пластов этих пород на моделируемом горизонте данного месторождения, например для песков мощности пластов 1–5 м, для глин 0.2–2 м, для песчаников, 0.2–0.5 м. На втором этапе, для каждого 10-и см пропластка внутри пласта выбирается значение КС из заранее определенного интервала, с вероятностью, заданной нормальным распределением. Мы получаем «реальное» значение КС и ПС для каждого 10-и см пропластка вдоль ствола скважины в моделируемом горизонте. Зарегистрированное при каротаже значение КС и ПС зависит от многих параметров, характеристик бурового раствора и грунтовых вод, диаметра скважины, параметров скважинного прибора. В данной модели мы ограничились только учетом влияния скважинного прибора на зарегистрированные значения.

Для реализации описанной модели разработана программа генерации каротажных данных, позволяющая сгенерировать данные произвольного количества синтетических (искусственных) скважин. Полученные данные, в свою очередь, обрабатываются моделью скважинного прибора длиной 1 м.

В графическом виде результаты генерации данных КС и данных, полученных моделью скважинного прибора, приведены на рис. 2.

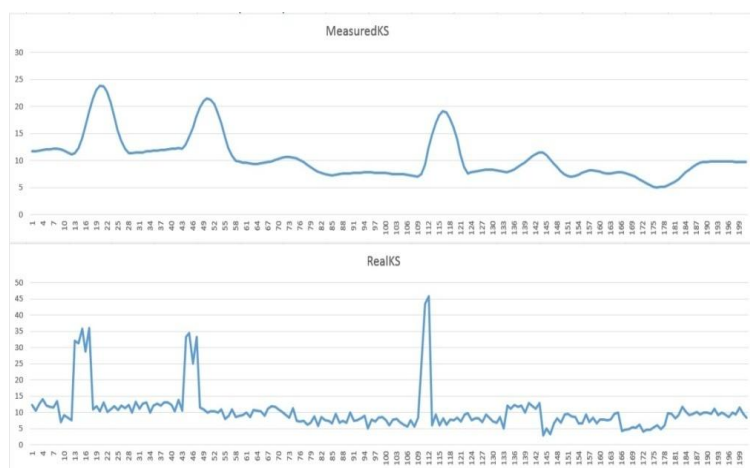


Рисунок 2. Генерируемые данные КС (внизу) и результат их «измерения» (сверху)

Видно, что в результате использования прибора пласт низкой мощности с высоким показателем КС при измерении имеет значительно меньший показатель из-за усреднения. Для того чтобы оценить адекватность модели, по некоторым скважи-

нам из обучающей выборки были смоделированы «измеренные» значения КС и ПС для заданного распределения литотипов вдоль ствола скважины. Результат для одной из скважин показан на рис. 3. Видно хорошее совпадение с реальными данными. Кривая КС получается несколько более сглаженная по сравнению с реальными данными, но это вероятно из-за неучета некоторых факторов, например изменения диаметра скважины, помехи и пр.

Кривая ПС более выразительная, чем реальная кривая, но качественную кривую ПС практически невозможно встретить в реальных данных потому, что это требует особого приготовления раствора, определенных условий измерения, что трудновыполнимо на практике.

4. Результаты применения алгоритмов МО к данным синтезированной скважины

Синтезированные по указанным выше правилам данные были обработаны и проинтерпретированы с помощью искусственной нейронной сети (табл. 1) и алгоритма k-NN. Кривая обучения показана на рис. 4. Алгоритм k-NN демонстрирует несколько худшие результаты обучаемости (рис. 5, 6).

Результаты свидетельствуют о том, что искусственные нейронные сети способны показать высокий результат классификации на данных синтезированной скважины. Соответственно, синтезированная скважина упрощает выявление различий в алгоритмах классификации.

Таблица 1. Результаты работы алгоритма искусственной нейронной сети на данных синтезированных скважин

Условный номер скважины	Accuracy	Kappa	WM recall	WM precision	T1 Score
1	0.9122	0.889	0.9174	0.8605	0.8880
2	0.9295	0.911	0.9196	0.9053	0.9124
3	0.9391	0.923	0.9411	0.8993	0.9197
4	0.9545	0.942	0.9621	0.9281	0.9448
5	0.9591	0.949	0.9625	0.9355	0.9488
6	0.9678	0.96	0.9713	0.9458	0.9584
7	0.9740	0.967	0.9733	0.9621	0.9676
8	0.9748	0.968	0.9761	0.9623	0.9692
9	0.9789	0.973	0.9788	0.9733	0.9760
10	0.9792	0.974	0.9783	0.9744	0.9763
11	0.9801	0.975	0.9775	0.9762	0.9767
12	0.9809	0.976	0.9812	0.9781	0.9796
Average	0.9608	0.9506	0.9616	0.9417	0.9515

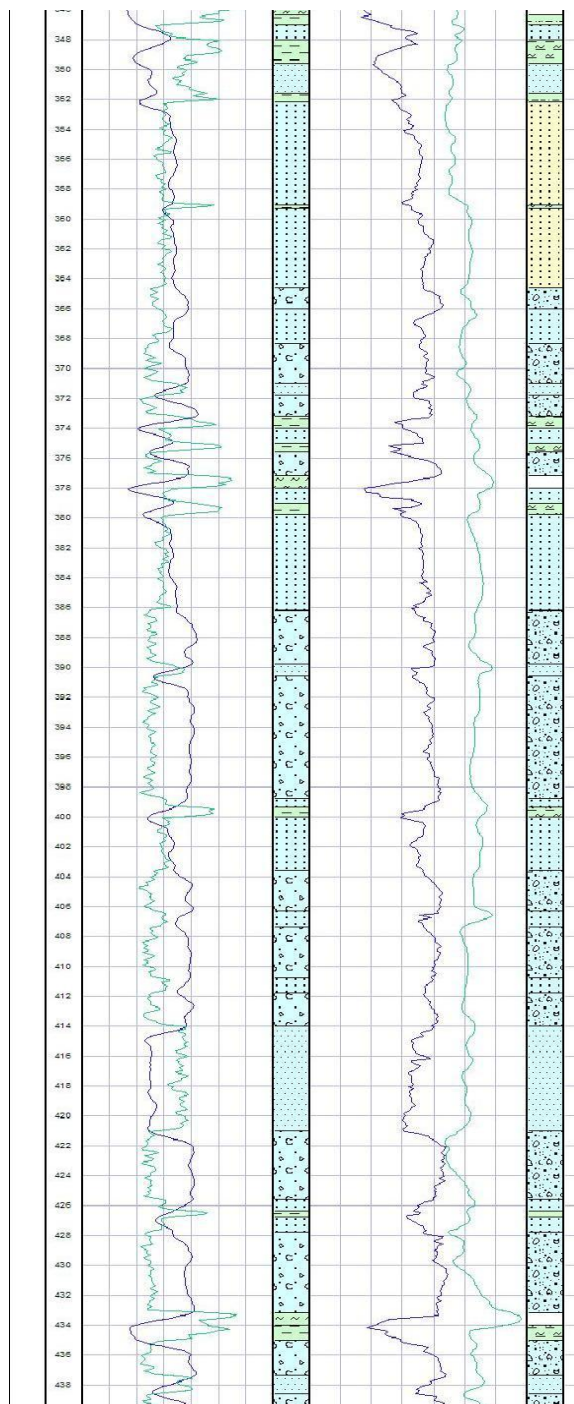
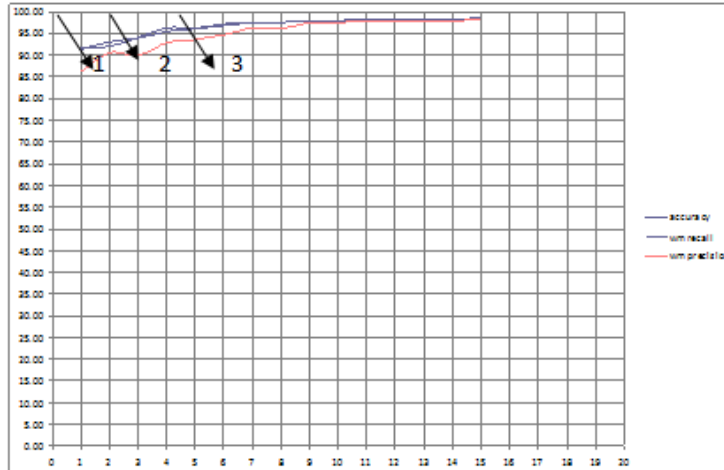
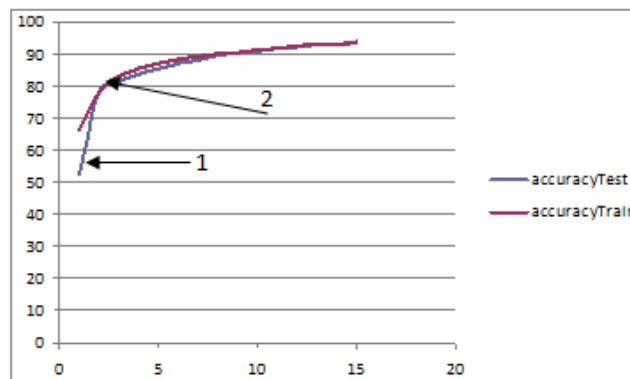


Рисунок 3. Смоделированные (слева) и реальные(справа) данные каротажа



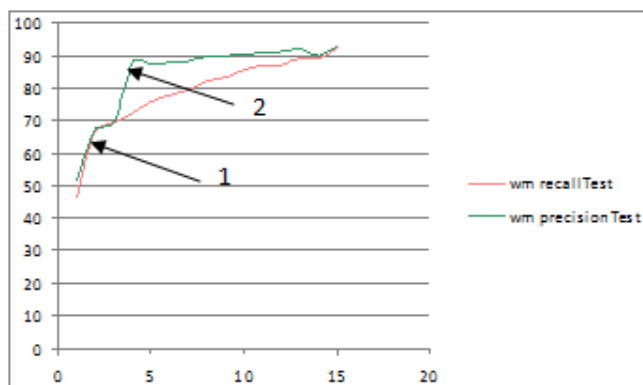
1 — accuracy, 2 — wm recall, 3 — wm precision

Рисунок 4. Кривая обучения нейросетевого классификатора для случая синтезированной скважины. По горизонтальной оси показано количество «скважин», участвовавших в обучении



1 — accuracyTest, 2 — accuracyTrain

Рисунок 5. Изменение показателя accuracy при увеличении числа обучающих примеров синтезированной скважины для алгоритма k-NN



1 — *wm recallTest*, 2 — *wm precisionTest*

Рисунок 6. Изменение показателей *recall* и *precision* при увеличении числа обучающих примеров синтезированной скважины для алгоритма *k-NN*

Хотя нейронные сети прямого распространения показывают высокие результаты классификации на искусственных данных, к сожалению, на реальных данных результат значительно хуже. Причина этого заключается в существенной противоречивости экспертного оценивания. Оценка противоречивости мнений экспертов приведена в следующем разделе.

5. Сравнение экспертных оценок

Противоречивость экспертных оценок вносит дополнительные трудности в процесс обучения системы МО. Несмотря на то, что именно экспертные оценки используются для обучения, эксперты проводят классификацию по-разному. Для сравнительного анализа качества экспертных оценок были выбраны три скважины с номерами 2100, 2104 и 4939. Данные каротажа указанных скважин были переданы трем экспертам, условно обозначенным буквами *D*, *L* и *T*. Кроме этого, для скважин 2100 и 4939 известны данные кернового опробования (*kern*). Основываясь на полученных от экспертов данных литологического расчленения и данных кернового опробования, был проведен расчет основных показателей качества (*accuracy*, *recall*, *precision*, *Kappa*) при попарном сравнении, когда данные одного из экспертов принимались за эталон, а данные второго эксперта с ними сравнивались (табл. 2).

Видно, что среднее значение точности для экспертов составляет $accuracy = 0.67$, а усредняющий показатель разброса $T1\ Score = 0.6$. Сравнивая оценки экспертов с данными кернового опробования, получаем цифры $accuracy = 0.5$, а $T1\ Score = 0.27$.

В то же время, учитывая субъективность экспертов, можно отметить, что по скважине 2100 и 2104 эксперты *D* и *L* больше всего согласны между собой (*accuracy* > 0.8). Со значениями усредненного ядра больше всего совпадает мнение эксперта *D* (*accuracy* = 0.69). По скважине 4939 эксперты *D* и *T* больше всего согласны между собой (*accuracy* = 0.77). С результатами усредненного ядра больше всего совпадают оценки эксперта *L* (*accuracy* = 0.60).

Таблица 2. Парное сравнение

Пары экспертов	<i>Accuracy</i>	<i>Kappa</i>	<i>Recall</i>	<i>Precision</i>	<i>T1 Score</i>
2100					
<i>D</i> vs <i>L</i>	0.81	0.70	0.63	0.61	0.618018
<i>D</i> vs <i>T</i>	0.71	0.54	0.42	0.49	0.455325
<i>L</i> vs <i>T</i>	0.80	0.67	0.46	0.52	0.493058
4939					
<i>D</i> vs <i>L</i>	0.3317	0.16	0.5979	0.4975	0.543099
<i>D</i> vs <i>T</i>	0.7706	0.66	0.7941	0.6879	0.737195
<i>L</i> vs <i>T</i>	0.3762	0.19	0.6086	0.5495	0.577542
2104					
<i>D</i> vs <i>L</i>	0.8409	0.76	0.8445	0.8205	0.832327
<i>D</i> vs <i>T</i>	0.6551	0.49	0.5493	0.5424	0.545828
<i>L</i> vs <i>T</i>	0.7213	0.57	0.5845	0.6011	0.592684
Average experts	0.67	0.53	0.61	0.59	0.60
2100					
<i>Kern</i> vs <i>D</i>	0.693	0.39	0.3624	0.3487	0.355418
<i>Kern</i> vs <i>L</i>	0.6444	0.23	0.2775	0.2423	0.258708
<i>Kern</i> vs <i>T</i>	0.6505	0.21	0.274	0.2441	0.258187
4939					
<i>Kern</i> vs <i>D</i>	0.1749	0.04	0.2011	0.2295	0.214363
<i>Kern</i> vs <i>L</i>	0.6089	0.37	0.3066	0.3986	0.346599
<i>Kern</i> vs <i>T</i>	0.2096	0.04	0.2004	0.2231	0.211142
Average kern	0.50	0.21	0.27	0.28	0.27

При сравнениях данных экспертов с данными ядрового опробования показатели качества существенно ниже, чем при сравнении данных экспертов между собой (табл. 3).

Анализ показал, что для некоторых критичных пород (глина) мнения экспертов совпадают в 70–95% случаев, однако значительно хуже согласуются с данными ядрового опробования. При выделении глин совпадение экспертов с ядром составляло 30–50%.

Таблица 3. Средние значения качественных показателей

	<i>Accuracy</i>	<i>Kappa</i>	<i>Recall</i>	<i>Precision</i>	<i>T1 Score</i>
<i>Experts</i>	0.67	0.53	0.61	0.59	0.60
<i>Kern</i> vs <i>experts</i>	0.50	0.21	0.27	0.28	0.27

Это во многом связано с тем, что, исходя из требования процесса скважинного выщелачивания, критично определить непроницаемые породы. Для урановых месторождений Казахстана это в основном глины или алевролиты. Кроме того, для глин характерно минимальное значение кажущегося сопротивления (КС) и максимальное значение потенциала (ПС), что облегчает их определение. Выделение же различных пропластков в проницаемых породах, например выделение в пласте среднезернистого песка пропластков разнозернистых песков, некритично с точки зрения технологии и до некоторой степени определяется предпочтениями конкретного эксперта.

Сопоставление экспертных оценок наталкивает на мысль о разделении скважин, проинтерпретированных разными экспертами в рамках одной выборки для уменьшения ее противоречивости. Другой возможностью является использования ID эксперта как одного из параметров при обучении.

Эти же три скважины были проинтерпретированы с помощью FeedForward нейронной сети, состоящей из 26 входных нейронов (глубина, координаты, плавающие окна по кривым КС и ПС) и двух скрытых слоев: 52 и 78 нейронов, реализованной с помощью библиотеки Keras.

Каротажные кривые после предварительной нормировки, подаются в виде плавающего окна, поскольку для определения литотипа важным критерием является не только собственно значение в точке, но и форма кривой (экстремумы, перегибы). Кроме того, поскольку литологические разрезы соседних скважин достаточно хорошо коррелируют (этот факт также используют эксперты для выделения пород), было принято решение также использовать координаты скважин как один из параметров обучения. Использование такого подхода позволило получить точность распознавания, сопоставимую с экспертными оценками по «кernовым» скважинам (табл. 4).

Таблица 4. Результаты работы алгоритма искусственной нейронной сети на данных kernовых скважин

Номер скважины	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
2100	0.4273	0.4623	0.2701	0.3409
2104	0.7053	0.7961	0.6053	0.6877
4939	0.3277	0.2573	0.1181	0.1619
Average	0.5092	0.5069	0.3409	0.4076

Однако стоит отметить, что если рассматривать отдельно непроницаемые породы (глины, алевролиты), то точность их распознавания сетью по данным kernовых скважин уступает экспертам: *Precision* = 0.30, и особенно «хромает» полнота

$Recall = 0.22$. То есть большинство непроницаемых пластов, критичных с точки зрения технологии добычи, остаются невыделенными.

Таблица 5. Точность выделения непроницаемых пород экспертами и нейронной сетью для керновых скважин

Номер скважины	Эксперт D		Эксперт L		Эксперт T		ANN	
	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>	<i>Precision</i>	<i>Recall</i>
2100	0.3182	0.7000	0.2727	0.5455	0.2727	0.5455	0.4516	0.4375
2104	Сравнение невозможно, так как данные относятся к разным горизонтам							
4939	0.4146	0.3366	0.4146	0.3301	0.3049	0.2907	0.1613	0.0060
Average	0.3664	0.5183	0.3437	0.4378	0.2888	0.4181	0.3064	0.2275

В целом же невысокая точность, как экспертов, так и нейронной сети по этим скважинам, обусловлена тем, что интерпретация по ним изначально была проведена не на основе формы каротажных кривых, как это обычно происходит, а по результатам кернового опробования (описания, лабораторные пробы), т. е. информации, которая была фактически недоступна ни нейронной сети, ни экспертам.

6. Заключение

Интерпретация данных электрического каротажа скважин на урановых месторождениях решает важную производственную задачу — выбор места для установки фильтра при добыче урана методом подземного скважинного выщелачивания. В процессе интерпретации эксперт выделяет залегающие породы и, по существу, выполняет литологическую классификацию, формируя строение скважины по глубине. Накопленные классификации послужили основой для обучения автоматических систем на базе методов машинного обучения, в первую очередь с использованием нейронных сетей. Однако проведенные ранее эксперименты по классификации данных электрического каротажа скважин на урановых месторождениях показали, что средние показатели качества автоматического классификатора на базе нейронных сетей прямого распространения составляют порядка 65% (ассигасу).

Вместе с тем в настоящей работе с применением синтезированных данных показано, что качество нейронной классификации должно быть значительно выше (до 90%). Одной из причин низкого качества автоматического классификатора может быть расхождение мнений экспертов при обработке одних и тех же данных каротажа. На практике оценка данных осуществляется экспертами независимо, каждый эксперт оценивает данные однократно и результаты оценивания не перекрываются. Для выявления разнородности мнений был проведен небольшой эксперимент, в котором участвовали трое экспертов. Каждый из них оценил данные каротажа одних и тех же трех скважин. Эксперимент показал существенное расхождение мнений

экспертов. Примерно в 30% случаев интерпретации экспертов отличаются друг от друга. В силу ограниченности эксперимента нельзя с высокой достоверностью утверждать, что такое расхождение имеет место всегда. Однако учитывая то обстоятельство, что качество классификации в «идеальных» условиях также примерно на 30% лучше, чем на реальных данных, можно предположить, что указанная закономерность не случайна. Для уменьшения ее влияния и повышения качества нейросетевой классификации предложено в будущих экспериментах учитывать идентификатор эксперта как один из параметров в процессе обучения и применения сети.

Благодарности

Работа финансируется грантом BR05236447 «Интеллектуальные системы управления и принятия решений для разработки месторождений урана и нефти».

Литература

- [1] *Van der Baan M. and Jutten C.* Neural networks in geophysical applications // *Geophysics*. 2000. Vol. 65. No. 4. P. 1032–1047.
- [2] *Baldwin J. L., Bateman R. M. and Wheatley C. L.* Application of a neural network to the problem of mineral identification from well logs // *The Log Analyst*. 1990. No. 3. P. 279–293.
- [3] *Benaouda B., Wadge G., Whitmark R. B., Rothwell R. G., MacLeod C.* Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs — an example from the Ocean Drilling Program // *Geophysical Journal International*. 1999.
- [4] *Saggaf M. M., Nebrija Ed. L.* Estimation of missing logs by regularized neural networks // *AAPG Bulletin*. 2003. No. 8. P. 1377–1389.
- [5] *Тененев В. А., Якимович Б. А., Сенилов М. А., Паклин Н. Б.* Интеллектуальные системы интерпретации геофизических исследований скважин // *Штучний інтелект*. 2002. No. 3. С. 338.
- [6] *Klaus Yelbig and Sven Treitel.* Computational Neural Networks For Geophysical Data Processing. Ed. Mary M. Poulton. 2001.
- [7] *Borsaru M., Zhou B., Aizawa T., Karashima H., Hashimoto T.* Automated lithology prediction from PGNA and other geophysical logs // *Applied Radiation and Isotopes*. 2006. No. 64. P. 272–282.
- [8] *Rogers S. J., Chen H. C., Kopaska-Merkel D. C. T., Fang J. H.* Predicting permeability from porosity using artificial neural networks // *AAPG Bulletin*. 1995. No. 79. P. 1786–1797.
- [9] *Kapur L., Lake L., Sepehrnoori K., Herrick D., Kalkomey C.* Facies prediction from core and log data using artificial neural network technology // *Transactions of the 39th Society of Professional Well Log Analysts Annual Logging Symposium*. 1998. P. 1.

- [10] Алешин С. П., Ляхов А. Л. Нейросетевая оценка минерально-сырьевой базы региона по данным геофизического мониторинга // *Нові технології*. 2011. № 1 (31). С. 39–43.
- [11] Rogers S. J., Fang J. H., Karr C. L., and Stanley D. A. Determination of lithology from well logs using a neural network // *AAPG Bulletin*. 1992. No. 76(5). P. 731–739.
- [12] Костиков Д. В. Инструментальные средства интерпретации геофизических исследований скважин на основе преобразованных каротажных диаграмм с помощью многослойной нейронной сети: дисс. ... к.т.н. — М. : РГБ, 2007. 189 с.
- [13] Muhamediyev R., Amirgaliev E., Iskakov S., Kuchin Y., Muhamedyeva E. Integration of Results of Recognition Algorithms at the Uranium Deposits // *Journal of ACIII*. 2014. Vol. 18. No. 3. P. 347–352.
- [14] Амиргалиев Е. Н., Искаков С. Х., Кучин Я. В., Мухамедиев Р. И. Интеграция алгоритмов распознавания литологических типов // *Проблемы информатики. Сибирское отделение РАН*. 2013. № 4 (21). С. 11–20.
- [15] Амиргалиев Е. Н., Искаков С. Х., Кучин Я. В., Мухамедиев Р. И. Методы машинного обучения в задачах распознавания пород на урановых месторождениях // *Известия НАН РК*. 2013. № 3. С. 82–88.
- [16] Яшин С. А. Подземное скважинное выщелачивание урана на месторождениях Казахстана // *Горный журнал*. 2008. № 3. С. 45–49.
- [17] Нейрокомпьютеры: учеб. пособие для вузов. — М. : МГТУ им. Н. Э. Баумана, 2004.
- [18] Development of methods of data boreholes interpretation by using artificial neural network (On request of “Geotechnoserviss” ltd) (in russian – «Разработка методики для интерпретации данных ГИС с помощью нейронных сетей») (2011).
- [19] Kuchin Y. I., Muhamedyev R. I., Muhamedyeva E. L. et al. The analysis of the data of geophysical research of boreholes by means of artificial neural networks // *Computer Modelling and New Technologies*. 2011. Vol. 15. No. 4. P. 35–40.
- [20] Muhamedyev R. I., Kuchin Y. I., and Muhamedyeva E. L. Geophysical research of boreholes: Artificial neural networks data analysis // *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*. P. 825–829. IEEE, 2012.
- [21] Amirgaliev E., Isabaev Z., Iskakov S., Kuchin Y., Muhamediyev R., Muhamedyeva E., Yakunin K. Recognition of rocks at uranium deposits by using a few methods of machine learning // *Soft Computing in Machine Learning*. P. 33–40. Springer, Cham, 2014.
- [22] Muhamedyev R. I. et al. Comparative analysis of classification algorithms // 2015 9th International Conference on Application of Information and Communication Technologies (AICT). — IEEE, 2015. P. 96–101.
- [23] Muhamediyev R., Amirgaliev E., Iskakov S., Kuchin Y., Muhamedyeva E. Integration of Results of Recognition Algorithms at the Uranium Deposits // *JACIII*. 2014. Vol. 18. No. 3. P. 347–352

- [24] *Muhamedyev R., Iskakov S., Gricenko P., Yakunin K., Kuchin Y.* Integration of results from Recognition Algorithms and its realization at the uranium production process // 8th IEEE International Conference AICT. — Astana, 2014. P. 188–191.
- [25] Методические рекомендации по комплексу геофизических методов исследования скважин при подземном выщелачивании урана. — Алматы : ЗАО НАК «Казатомпром». ТОО ИВТ, 2003. 36 с.
- [26] Техническая инструкция по проведению геофизических исследований в скважинах на пластово инфильтрационных месторождениях урана. — Алматы : ТОО ГРК, 2010.

Авторы:

Ян Игоревич Кучин — master of computer science engineering, Институт информационных и вычислительных технологий МОН РК, Казахстан, Алматы

Равиль Ильгизович Мухамедиев — доктор инженерных наук, профессор, Казахский национальный технический университет имени К. И. Сатпаева, Казахстан, Алматы; ISMA University, Latvia, Riga

Кирилл Олегович Якунин — докторант PhD, Казахский национальный технический университет имени К. И. Сатпаева, Казахстан, Алматы

The quality of data classification in the context of inconsistency of expert assessments

Y. I. Kuchin, K. O. Yakunin, R. I. Muhamedyev

*Institute of Information and Computational Technologies.
Kazakhstan, Almaty, Pushkin str. 125.*

e-mail: ykuchin@mail.ru, ravil.muhamedyev@gmail.com, yakunin.k@mail.ru

Abstract. In the process of uranium mining using the in-situ leaching method, the correct classification of lithological layers based on electrical logging data is important. To solve this problem, manual labor is often used today. The attendant errors leading to economic losses make the automation of this process relevant. To automate the process of interpreting logging data, machine learning methods are used, more precisely, "supervised learning". The paper deals with the problem of inconsistency of expert interpretation and the related limitations in the quality of work of neural networks. To assess the methods of machine learning in the "ideal case", the concept of a "synthesized" well was introduced, a set of specially generated data that simulates electrical logging data and expert assessments. Based on these data, neural networks with direct signal propagation show high quality of classification (about 96% accuracy). When using real data, the accuracy rate is considerably lower, about 55%. The paper shows that one of the important reasons for the low quality of automatic classification is the inconsistency of expert assessments. The results of the evaluation of three wells by three different experts are given and their agreement among themselves and with core testing data is evaluated. The qualitative indicators of data interpretation of these wells using an artificial neural network are given. Due to the revealed subjectivity of expert estimation, it was sug-

gested in further experiments to use an expert identifier as an additional parameter when training a neural network.

Keywords: machine learning, supervised learning, expert estimation, "synthesized" well, accuracy, precision, recall, artificial neural networks.

References

- [1] Van der Baan M. and Jutten C. (2000) *Geophysics*, **65**(4):1032–1047.
- [2] Baldwin J. L., Bateman R. M. and Wheatley C. L. (1990) *The Log Analyst*, 3:279–293.
- [3] Benaouda B., Wadge G., Whitmark R. B., Rothwell R. G., MacLeod C. (1999) Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs — an example from the Ocean Drilling Program. *Geophysical Journal International*.
- [4] Saggaf M. M., Nebrija Ed. L. (2003) *AAPG Bulletin*, 8:1377–1389.
- [5] Tenenev V. A., Yakimovich B. A., Senilov M. A., Paklin N. B. (2002) *Shtuchnyy intellekt*, 3:338.
- [6] Klaus Yelbig and Sven Treitel. (2001) *Computational Neural Networks For Geophysical Data Processing*. Ed. Mary M. Poulton.
- [7] Borsaru M., Zhou B., Aizawa T. et al. (2006) *Applied Radiation and Isotopes*, 64:272–282.
- [8] Rogers S. J., Chen H. C., Kopaska-Merkel D. C. t, Fang J. H. (1995) *AAPG Bulletin*. 79:1786–1797.
- [9] Kapur L., Lake L., Sepehrnoori K., Herrick D., Kalkomey C. (1998) Facies prediction from core and log data using artificial neural network technology. *Transactions of the 39th Society of Professional Well Log Analysts Annual Logging Symposium*. P. 1.
- [10] Aleshyn S. P., Lyakhov A. L. (2011) *Novi tekhnolohiyi*, **1**(31):39–43.
- [11] Rogers S. J., Fang J. H., Karr C. L., and Stanley D. A. (1992) *AAPG Bulletin*. **76**(5):731–739.
- [12] Kostikov D. V. Instrumental'nyye sredstva interpretatsii geofizicheskikh issledovaniy skvazhin na osnove preobrazovannykh karotazhnykh diagramm s pomoshch'yu mnogo-sloynoy neyronnoy seti. Tesis. M., 2007.
- [13] Muhamediyev R., Amirgaliev E., ..., Muhamedyeva E. (2014) *Journal of ACIII*, **18**(3):347–352.
- [14] Amirgaliyev Ye. N., Iskakov S. Kh., Kuchin Ya. V., Mukhamediyev R. I. (2013) *Problemy informatiki. Sibirskoye otdeleniye RAN*, **4**(21):11–20.
- [15] Amirgaliyev Ye. N., Iskakov S. Kh., Kuchin Ya. V., Mukhamediyev R. I. (2013) *Izves. NAN RK*, 3:82–88.
- [16] Yashin S. A. (2008) *Gornyy zhurnal*. 3:45–49.
- [17] *Neyrokomp'yutery: ucheb. posobiye dlya vuzov*. (2004) Moscow, Izd-vo MGTU im. N. E. Bauman
- [18] Development of methods of data boreholes interpretation by using artificial neural network (On request of "Geotehnoserviss" ltd) (2011).
- [19] Kuchin Y. I., Muhamedyev R. I. et al (2011) *Comp. Model. and New Technol.*, **15**(4):35–40.
- [20] Muhamedyev R. I., Kuchin Y. I., and Muhamedyeva E. L. (2012) Geophysical research of boreholes: Artificial neural networks data analysis. In 2012 Joint 6th Intern. Conf. on Soft Computing and Intelligent Systems (SCIS) and 13th Intern. Symp. on Advanced Intelligent Systems (ISIS), P. 825–829.
- [21] Amirgaliev E., Isabaev Z., Iskakov S., ..., and Yakunin K. (2014) Recognition of rocks at uranium deposits by using a few methods of machine learning. *Soft Computing in Machine Learning*. P. 33–40.
- [22] Muhamedyev R. I. et al. (2015) Comparative analysis of classification algorithms. In 2015 9th International Conference on Application of Information and Communication Technologies (AICT), P. 96–101.
- [23] Muhamediyev R., Amirgaliev E., Iskakov S., Kuchin Y., Muhamedyeva E. (2014) *JACIII*, **18**(3):347–352
- [24] Muhamedyev R., Iskakov S., Gricenko P., Yakunin K., Kuchin Y. (2014) In 8th IEEE International Conference AICT Integration of results from Recognition Algorithms and its realization at the uranium production process.. Astana. P. 188–191.
- [25] (2003) Metodicheskiye rekomendatsii po kompleksu geofizicheskikh metodov issledovaniya skvazhin pri podzemnom vyshchelachivanii urana. Almaty, ZAO NAK «Kazatomprom». TOO IVT.
- [26] (2010) Tekhnicheskaya instruktsiya po provedeniyu geofizicheskikh issledovaniy v skvazhinakh na plastovo infil'tratsionnykh mestorozhdeniyakh urana. Almaty, TOO GRK.