

ВИМІРЮВАННЯ, КОНТРОЛЬ І ТЕХНІЧНА ДІАГНОСТИКА

УДК 004.942:622.286

ФОРМАЛЬНІ ОСНОВИ КОМП'ЮТЕРНОЇ ІНТЕРПРЕТАЦІЇ ДАНИХ ГЕОФІЗИЧНИХ ДОСЛІДЖЕНЬ СВЕРДЛОВИН

М.Я. Бестильний, О.Ф. Козак, В.М. Юрчишин

ІФНТУНГ, 76019, м. Івано-Франківськ, вул. Карпатська, 15, тел. (0342) 727132,
e-mail: pz@nuing.edu.ua

В запропонованому дослідженні виконано аналіз формально-логічних методів створення інтелектуальних технологій видобування знань на основі даних геофізичних досліджень. Процес аналітичної обробки та інтерпретації геолого-геофізичних даних розглядається з точки зору підвищення ефективності процесу комп'ютерної обробки геолого-геофізичних даних та виявлення нових знань в даних шляхом розробки інтелектуальних формально-логічних методів аналізу даних на основі концепції Data Mining. Проаналізовано існуючі методи видобування даних та пошуку знань в геолого-геофізичних даних. Розробка новітніх методів видобування знань на основі геолого-геофізичних даних з використанням штучних нейронних мереж дозволить застосовувати спеціальні імітаційні моделі для генерування геолого-геофізичних даних, що забезпечить можливість проведення на їх основі експериментальних досліджень технологій аналізу даних в цілому. Розробка інтелектуальної системи визначення колекторських властивостей гірських порід на основі геолого-геофізичних даних з використанням нейротехнології дозволить виконати інтеграцію визначених формально-логічних методів інтелектуальної технології видобування знань та інтелектуальних систем визначення колекторських властивостей гірських порід.

Ключові слова: нафтогазовий об'єкт, штучні нейронні мережі, геолого-геофізичні дані, інтерпретація даних та знань, видобування даних та знань, інтелектуальна система.

В предлагаемом исследовании выполнен анализ формально-логических методов создания интеллектуальных технологий извлечения знаний на основе данных геофизических исследований. Процесс аналитической обработки и интерпретации геолого-геофизических данных рассматривается с точки зрения повышения эффективности процесса компьютерной обработки геолого-геофизических данных и выявления новых знаний в данных путем разработки интеллектуальных формально-логических методов анализа данных на основе концепции Data Mining. Проанализированы существующие методы добычи данных и поиска знаний в геолого-геофизических данных. Разработка новых методов извлечения знаний на основе геолого-геофизических данных с использованием искусственных нейронных сетей позволит применять специальные имитационные модели для генерирования геолого-геофизических данных, что обеспечит возможность проведения на их основе экспериментальных исследований технологий анализа данных в целом. Разработка интеллектуальной системы определения коллекторских свойств горных пород на основе геолого-геофизических данных с использованием нейротехнологии позволит выполнить интеграцию определенных формально-логических методов интеллектуальной технологии извлечения знаний и интеллектуальных систем определения коллекторских свойств горных пород.

Ключевые слова: нефтегазовый объект, искусственные нейронные сети, геолого-геофизические данные, интерпретация данных и знаний, добыча данных и знаний, интеллектуальная система.

In this research the analysis of formal-logical methods of creating intelligent technologies for the extraction of knowledge from logging data is done. The analytical processing and interpretation of geological and geophysical data is considered in terms of improving process efficiency of computer processing of geological and geophysical data and revealing of new knowledge in the data by developing intelligent formal logical methods of data analysis based on the concept of Data Mining. Existing methods for data mining and knowledge search in geological and geophysical data were explored. The development of new methods of knowledge extraction based on geological and geophysical data using artificial neural networks will use special simulation models to generate geological and geophysical data, providing the possibility of their experimental studies on data mining in general. Developing smart technologies for identifying reservoir properties of rocks which are based on geological and geophysical data

using neurotechnology will integrate certain formal logical methods of intellectual knowledge extraction technologies and intelligent systems for determining reservoir properties of rocks.

Key words: oil and gas facility, artificial neural networks, geological and geophysical data, interpretation of data and knowledge, data mining and knowledge mining.

В методології та технології обробки та аналізу геолого-геофізичних даних (ГГД) умовно можна виділити три етапи розвитку: створення основних методів обробки та інтерпретації даних за відсутності засобів обчислювальної техніки, що суттєво стримувало теоретичні та методичні дослідження; перші спроби перетворення аналогових вхідних даних в цифрову форму та цифрової реєстрації результатів спостережень, розробка ряду алгоритмічних та програмних засобів обробки та інтерпретації даних для різних класів ЕОМ; сучасний етап, для якого характерне широке впровадження у виробничу практику комп'ютеризованих технологій – від реєстрації даних до комплексної інтерпретації результатів польових досліджень, підрахунку запасів нафтогазових родовищ, контролю за їх розробкою тощо [1–3].

На сьогоднішній день для аналізу ГГД використовується широкий спектр програмного забезпечення. Його можна поділити на дві групи: 1) програмні пакети загального призначення – математичні, графічні та текстові процесори; складність використання таких пакетів полягає в тому, що кожний з них має свій формат вхідних та вихідних даних, а, отже, реалізація певної послідовності процедур вимагає від користувача додаткових дій для узгодження даних; 2) спеціалізовані програмні засоби різного рівня інтегрованості – автономні спеціалізовані програми, програмні комплекси та системи обробки за окремим методом, системи комплексної обробки даних, геоінформаційні системи (ГІС) [4–5].

Динаміка розвитку комп'ютерної техніки, ускладнення задач, які постають перед геологічною галуззю, призводить до того, що виникає необхідність у розробці нових, більш досконалих та інтелектуальних систем і, відповідно, проведення теоретичних досліджень, пов'язаних із створенням методологічного забезпечення та теоретичного обґрунтування їхнього функціонування.

Таким чином, для нафтогазової предметної області невирішеною є задача комплексної інтелектуальної обробки даних та знань про нафтогазові об'єкти [6] на основі використання існуючих моделей даних та знань, що використовуються при геофізичних дослідженнях свердловин.

Метою даної статті є розробка формально обґрунтованого методу інтерпретації даних та знань про нафтогазові об'єкти на етапі інтерпретації даних геофізичних досліджень свердловин.

Успіхи сучасного розвитку комп'ютерних технологій обробки ГГД значною мірою зобов'язані розвитку сукупності дисциплін, які вивчають властивості інформації, а також організацію процесів її отримання, зберігання, оброб-

ки та передачі з використанням ЕОМ – комп'ютерних наук, основними напрямками яких є:

- розробка та специфікація моделей процесів та явищ реального світу для отримання нової інформації про закономірності їх виникнення та розвитку;

- алгоритмізація та програмування моделей для їх інтерпретації у середовищі ЕОМ;

- організація обчислювального та імітаційного експерименту з моделлю;

- організація інтелектуального предметно-орієнтованого інтерфейсу користувача з інтерпретуючим середовищем ЕОМ;

- організація мережевих структур передачі інформації з множинним доступом на основі концепції відкритих систем;

- створення баз даних предметної області (ПО) та організація з їх допомогою процесів зберігання та пошуку інформації;

- створення нових інформаційних технологій на основі концепції штучного інтелекту.

Широке використання обчислювальної техніки у геологічному виробництві привело до створення окремого розділу комп'ютерних наук, який отримав назву "геоінформатика". Насамперед це пов'язано із специфікою геології як сфери виробничої діяльності, яка принципово відрізняється від інших виробничих галузей – її єдиною продукцією є інформація, причому унікальна за своїми обсягами, різноманітністю та способами одержання. Ці способи не мають аналогів у інших виробничих галузях. Геологія сьогодні – це індустрія інформації, одержаної в процесі зондування Землі з чотирьох рівнів – космічного, повітряного, наземного та свердловинного. На основі одержаної первинної інформації виробляється основна інформаційна продукція – геологічні карти, моделі геологічних об'єктів, прогностичні оцінки, рекомендації геокористування та прогнозування його наслідків. Геологічні організації є фактично підприємствами з одержання та переробки інформації про Землю.

Геоінформатика базується на інтеграції геології, геохімії, геофізики, прикладної математики та обчислювальної техніки. Вона вивчає методи і закономірності реєстрації, зберігання, передачі, обробки та інтерпретації геологічної інформації. Метою досліджень у межах геоінформатики є створення комп'ютерних технологій для вивчення будови земної кори та її еволюції, а також прогнозу та пошуку корисних копалин, охорони довкілля.

Передумовою появи геоінформатики стали об'єктивні фактори. Ресурс родовищ, які могли б бути відкриті класичними методами, майже вичерпані. Успіх геологічної розвідки пов'язаний, насамперед, з інтегрованим системним аналізом різномірної та різнорівневої інформації. Проблеми такого інтегрованого аналізу (комплексування) входять до задач геологічної

інформатики. Їх не можна штучно вкласти ні у геофізику, ні у геохімію, ні у будь-яку іншу класичну дисципліну, оскільки геоінформатика їх охоплює, а джерелом нової інформації в ній є синтез даних, одержаних цими науками. Методологічною основою геоінформатики є математичні методи та інформаційні технології.

Одним з найбільш високотехнологічних продуктів на ринку програмного забезпечення нафтогазової галузі є технології ГІС – системи, в яких зберігаються дані, зареєстровані на площині досліджень. Вони використовуються для об'ємного моделювання, побудови геологічних карт, розрізів, аналізу багат шарових електронних карт.

Основними завданнями та проблеми геоінформатики є:

1. Оптимізація геологічної інформації, визначення оптимальних планів проведення геологорозвідувальних робіт, встановлення методів та обсягів, які б забезпечували результативність при мінімальних затратах.

2. Створення автоматизованих робочих місць (АРМ) геолога.

3. Розробка методів математичного моделювання геологічних об'єктів та процесів для вирішення прогнозно-пошукових задач та геоекологічних прогнозів.

4. Розробка методів та технологій інтегрованого (комплексного) аналізу різнорівневої та різнопараметричної інформації для тих самих задач.

5. Розробка комп'ютерних технологій моделювання будови та еволюції земної кори.

6. Розробка автоматизованих інформаційно-аналітичних систем для геологічних даних.

7. Розробка технологій цифрових геологічних карт та методів аналізу картографічної інформації.

8. Розробка експертних систем у геології для прогнозу корисних копалин та наслідків геористування.

9. Розробка програмного забезпечення польових комплексів, систем збору та передачі інформації.

Враховуючи сучасний розвиток науки та техніки, доцільною є розробка систем нового класу в геоінформатиці на основі теорії штучного інтелекту, призначених для комплексного вирішення задач аналізу наявної геолого-геофізичної інформації шляхом моделювання інтелектуальної діяльності фахівця ПО [7].

В геофізиці предметом вимірювання є фізичні властивості гірських порід та характеристики природних або штучно створених фізичних полів у цих породах. Засобами їх вимірювання є вимірювальні прилади – як аналогові, так і цифрові.

Результати вимірювань – це числові величини, виражені у відповідних одиницях, що і є елементом первинних *геолого-геофізичних даних*. Первинна *геолого-геофізична інформація* – це відомості про фізичні властивості, явища геологічного середовища, або фізичні поля деякого геологічного об'єкта та залежності між ними [8].

Основним результатом таких досліджень є відтворення просторового положення геологічних границь, характеристика літолого-стратиграфічного складу геологічних шарів, їх фізичних властивостей, зображених у вигляді відповідних геологічних розрізів, геологічних схем і карт, карт розподілу характеристик порід.

Перетворення цієї інформації про об'єкт у дані про його будову здійснюються шляхом багатокрокового процесу обробки та аналізу (на кожному проміжному кроці) результатів вимірювань. Машинна реалізація такого процесу виконується із застосуванням процедур перетворення первинних даних у більш загальну інформацію, яка доступна для безпосередньої оцінки людиною та подальшого використання. Конкретний склад та послідовність відповідних процедур залежать від конкретного методу та об'єму вхідних даних. Визначена сукупність таких процедур утворює методику обробки даних. В залежності від галузі розрізняють методики обробки геологічних та геофізичних даних. У межах окремої галузі використовуються: у геофізиці – методики обробки сейсмічних, гравімагнітних, електро- та радіометричних даних; у геології – методики обробки даних геохімічного аналізу, гранулометричних даних та ін.

У різноманітності форм прикладного програмного забезпечення, призначеного для вирішення задач обробки та аналізу ГГД, можна виділити ряд типів, кожний з яких може бути зіставлений з певним класом задач, які розв'язуються, та зв'язаний із рівнем розвитку обчислювальної техніки та технології програмування:

1. Незалежні програми, призначені для вирішення однієї фіксованої задачі.

2. Бібліотеки програм — набір програм, кожна з яких виконує певну функцію і має визначений інтерфейс програмування (API).

3. Багатофункціональні програми, які є розвитком продуктів першого типу. Це незалежні програми, які виконують певний набір функцій залежно від вхідних керуючих параметрів та даних, або складні програми на основі бібліотек програм.

4. Проблемно-орієнтовані системи (пакети прикладних програм). Передбачають організацію проблемного програмного забезпечення, яка допускає природне спілкування із системою на мовах, прийнятих у дисципліні даної предметної області, а також автоматизацію процесів реалізації, модифікації та вдосконалення.

Слід зауважити, що вказані типи програм були розроблені практично для кожного окремого методу як геологічних, так і геофізичних досліджень. Наприклад, існують пакети прикладних програм для обробки даних геохімічних спостережень, сейсмічних, гравіметричних, електрометричних тощо.

Як вже зазначалося, сучасна тенденція розвитку комп'ютеризації процесу розвідки вуглеводнів полягає у створенні таких програмних продуктів, в яких різні алгоритми обробки геолого-геофізичних та геохімічних даних об'єд-

нані у єдиний програмний комплекс засобами інтелектуального інтерфейсу. На відміну від інших типів програмних засобів, такі системи дозволяють обробляти різноманітну геоінформацію на основі інтегрованого аналізу. На сьогоднішній день створені методи такого аналізу, які забезпечують вирішення задач виділення та локалізації перспективних об'єктів, проведення кількісної інтерпретації різноманітної інформації, будувати багатофакторні фізико-геологічні моделі об'єктів, що вивчаються.

Розглянемо технологію інтегрування геологічної інформації [8]. На першому етапі по геологічних, геохімічних даних, по результатах дистанційного зондування, даних магнітометрії та гравіметрії, сейсміки та каротажу будується структурно-тектонічна карта території. Для цього проводиться статистичний аналіз інформації і виявляються аномалії, простежені у різних полях. Потім виконується геологічна інтерпретація виділених аномалій, і за допомогою методів розпізнавання образів здійснюється картування геологічних об'єктів.

На наступному етапі створюється тривимірна модель області пошуку та об'єкта досліджень. На основі сейсмічних розрізів та свердловинних даних формується перше (плоске) наближення такої моделі. Далі на основі рішення прямих та обернених задач геофізики формується взаємопов'язана з усіма експериментальними даними плоска модель. Перехід до просторової моделі здійснюється за допомогою методів районування з подальшим застосуванням на отриманих блоках методів регресійного аналізу.

Далі описується модель досліджуваного об'єкта та формується множина ознак, які характеризують перспективні області, за якими вони і виділяються. Виділені об'єкти ранжуються за ступенем перспективності методами крігінгу, регресійного аналізу і спеціально розробленими евристичними методами в залежності від вихідної інформації та поставленої мети.

Прикладом ГІС-системи, в якій реалізований такий підхід, є геоінформаційна система INTEGR0, розроблена у ВНИИГеосистеми (Росія).

Геолого-геофізичні дослідження [8] проводяться в межах пошуково-розвідувальних робіт для вирішення конкретних задач, а саме визначення якісного та кількісного складу корисних копалин. При проведенні пошуково-розвідувальних робіт на корисні копалини, зокрема на нафту та газ, застосовуються різноманітні методи досліджень – способи вивчення геологічної будови надр, виявлення і оцінки покладів: геологічні, геоморфологічні, геофізичні, геохімічні та буріння свердловин. Робота, виконана в даному дослідженні, базувалась на використанні даних отриманих за допомогою геологічних та геофізичних методів для створення інтелектуальної технології видобування знань на їх основі.

Геологічні методи досліджень базуються на вивченні характеру, складу і віку осадових відкладів і тектонічної будови території шля-

хом виконання різних польових і дистанційних спостережень, лабораторних досліджень і комплексного узагальнення одержаних матеріалів. Результати, отримані за допомогою геологічних методів, характеризуються високою точністю та достовірністю, однак їх застосування є трудомістким і має високу вартість.

Важливе місце серед пошуково-розвідувальних робіт на сьогоднішній день займають геофізичні дослідження свердловин. Геофізичні методи визначення колекторських властивостей і нафтогазонасиченості гірських порід стали основою при підрахунку запасів нафти і газу. В останні роки промислово-геофізична інформація широко використовується при проектуванні розробки родовищ нафти і газу, а також при контролі та аналізі процесу розробки, тобто забезпечує отримання всіх основних параметрів, необхідних для підрахунку запасів.

В комплексі геофізичних методів досліджень при пошуково-розвідувальних роботах на нафту і газ виділяють дві групи:

1) польові методи (розвідувальна геофізика);

2) промислові методи (геофізичні дослідження в свердловинах – каротаж).

Польові методи геофізичних досліджень базуються на вивченні і аналізі розподілу природних або штучно створених фізичних полів – гравітаційного, магнітного, електричного, радіоактивного, теплового і сейсмічного.

Неоднакові фізичні властивості гірських порід зумовлюють неоднорідні фізичні поля, внаслідок чого з'являється можливість виявляти особливості геологічної будови окремих територій. Штучні поля збуджуються спеціальними технічними засобами. На їх використанні базуються такі методи, як сейсмозрозвідка і електророзвідка постійним і змінним струмом, вивчення викликаних теплових полів. Достовірність і детальність методів штучного поля, як правило, вища, але їх застосування, відповідно, складніше і дорожче.

Таким чином, за допомогою методів геофізичного дослідження свердловин (ГДС) вирішується широке коло геологічних задач:

– літологічно-стратиграфічне розділення розрізу і виділення колекторів;

– виявлення змін у присвердловинних зонах і оцінка їхніх параметрів;

– оцінка фільтраційно-ємнісних властивостей та компонентного (мінералогічного) складу порід;

– виявлення продуктивних колекторів та оцінка запасів на основі визначення ефективних товщ; газорідних та водонафтових контактів та їх динаміки; виділення заводнених інтервалів;

– геохронологічна і літологічна кореляція розрізів, побудова просторових моделей родовищ;

– складання проектів розробки родовищ та їх моніторинг.

На завершальних стадіях розробки необхідними є:

– розподіл неоднорідних пластів по розрізу на пропластки;

– оптимізація відстаней між нагнітальними та видобувними свердловинами.

Разом з ускладненням задач, що вирішуються сучасним комплексом ГДС, зростає актуальність проблеми складних колекторів [8]. Складними називають колектори, для яких неможливе визначення та використання граничних значень пористості, глинистості та інших петрофізичних характеристик внаслідок характеру його структури та складу: полімінеральний склад породи; складна структура ємнісного простору; багатоконпонентний склад флюїду; мала ефективна товща. До таких колекторів відносяться також слабопроникні глинисті колектори. Виділення таких колекторів є дуже важливим завданням під час розробки родовищ.

До об'єктивних проблем також можна віднести неоптимальність комплексів ГДС та технологій дослідження, недосконалість метрологічного забезпечення, методик індивідуальної та комплексної інтерпретації даних ГДС.

Методика проведення ГДС характеризується великою кількістю методів, що використовують всі види фізичних полів (електричні, електромагнітні, теплові, ядерне випромінювання, гравітаційне, механічних напружень). Жодна геологічна задача не може бути вирішена тільки одним з методів ГДС. Звідси випливає принципова комплексність застосування методів ГДС і інтерпретації отриманих результатів.

Геологічна інформативність комплексу ГДС визначається багатьма факторами:

- геолого-геохімічними умовами;
- технічними умовами;
- промисловими умовами;
- термодинамічними умовами залягання порід;
- комплексом ГДС (його оптимальністю);
- метрологічними характеристиками інформаційно-вимірювальних систем;
- якістю інтерпретаційно-алгоритмічного забезпечення.

Під останніми розуміють рівень формального обґрунтування алгоритмів індивідуальної (одновимірної) інтерпретації і геологічна інтелектуальність системи комплексної (багатовимірної) інтерпретації. Чим вища якість інформаційного забезпечення, тим менше залежить інформативність ГДС від геолого-технологічних умов.

Якість індивідуальної інтерпретації визначається мірою використання фізичних законів, що застосовуються для теоретичної побудови інтерпретаційної моделі, алгоритмів розв'язання прямих і обернених задач методів ГДС. Ці закони відображають фізичну природу методів і кількісно відображаються диференціальними і інтегро-диференціальними рівняннями математичної фізики (рівняння Максвелла, Лапласа, Нав'є-Стокса, переносу випромінювання Больцмана, гідродинаміки, дифузії, теплопровідності та ін.). Ці рівняння лінійні. Ви-

вчення фізичних властивостей гірських порід в інтерференційних полях різної природи (сейсмоелектричний ефект в нейтронних та гамма-полях, нейтронно-акустичний ефект та ін.) вимагає дослідження та опису нелінійних ефектів.

Розробка способів кількісної інтерпретації даних кожного окремо взятого метода ГДС включає в себе вирішення трьох взаємопов'язаних проблем: вирішення прямої задачі геофізики; обґрунтування петрофізичної моделі та вирішення оберненої задачі.

Під *прямою задачею* [8] розуміють розрахунок показників приладів в системі «свердловина-пласт» при фіксованих геолого-технічних умовах вимірювань. Вирішення прямої задачі включає вивчення впливу на показники свердловинних приладів радіальної неоднорідності системи «свердловина-пласт» (сюди відноситься також проблема врахування впливу змін фізичних властивостей порід в присвердловинних зонах) і вертикальної роздільної здатності методу.

Вирішення прямої задачі дозволяє вивчити закономірності фізичного поля в системі «свердловина-пласт», вплив на показники радіальної неоднорідності цієї системи; такі спеціальні характеристики поля (апаратури), як глибинність дослідження – геометричну та інформаційну, вертикальну роздільну здатність.

Вирішення прямої задачі дозволяє вивчити інтерпретаційні залежності, виявити метрологічні характеристики апаратури і сформулювати правила (процедури і технічні засоби) їх визначення, знайти критерії та технічні можливості для оптимізації апаратури.

Під *цільовою функцією* [8], що реалізує критерій оптимальності інформаційно-вимірювальної системи за її метрологічними характеристиками, розуміють результуючу похибку визначення шуканого інтерпретаційного (петрофізичного) параметра. Величина цієї похибки залежить від чутливості показів не лише до визначеного параметра, а і до всіх параметрів-перешкод, а також від похибок вимірювання параметрів-перешкод.

Можливі три шляхи вирішення прямих задач: емпіричний (*апроксимаційний*), феноменологічний і строгий аналітичний. При очевидній перевазі останнього, він має ряд принципових обмежень. При дуже низькій симетричності задачі і складності конструкції свердловинного приладу строго вирішення класичними методами математичної фізики неможливо отримати в аналітичній формі з врахуванням великої кількості параметрів, що характеризують конструктивні особливості апаратури і технічні умови вимірювань. Можливе поєднання всіх трьох підходів. Якщо вирішення прямої задачі вдається отримати (наприклад, з допомогою феноменологічного підходу) в аналітичній формі, яка допускає обернення відносно шуканого петрофізичного (інтерпретаційного) параметра, то таке рішення називається *інтерпретаційною моделлю*. Якщо рішення прямої задачі вдається отримати у вигляді інтерпретаційної моделі, то це не лише значно полегшує побудову алгори-

тму інтерпретації, але і суттєво підвищує точність результатів (в порівнянні з емпіричними алгоритмами). Алгоритми, які отримані оберненням інтерпретаційної моделі, реалізують пряму безпоправкову, безпалеткову і безкернову (алгоритмічну) інтерпретацію. За наявності строго обґрунтованих петрофізичних моделей такі моделі не вимагають петрофізичних налаштувань.

Під петрофізичною моделлю метода розуміють виражений в аналітичній формі зв'язок результуючої петрофізичної характеристики породи в цілому з об'ємним вмістом і петрофізичними характеристиками її мінеральних компонентів (включаючи газ та рідкі мінерали – нафта, газ).

Знання закономірностей фізичного поля в системі «свердловина-пласт», наявність інтерпретаційної та строгої петрофізичної моделі забезпечують загалом безкернову, безпалеткову і безпоправкову індивідуальну інтерпретацію.

Однак, для деяких методів ГДС (наприклад, електрометрії та геоакустики) гірські породи є настільки складними, що замість строгих петрофізичних моделей використовують з певною мірою ефективності модельні підходи. Прикладами останніх можуть слугувати рівняння інтервального часу в акустиці та рівняння Арчі-Дахнова в електрометрії [8]. Для побудови таких моделей використовують принципи «вкладення» та «зміщення», а в якості критеріїв відбору найбільш реалістичних моделей – спеціальні фільтри. Зокрема метод власної поляризації інтерпретується до цього часу виключно на емпіричній основі.

Інтерпретація геофізичної інформації базується на вирішенні оберненої задачі, коли за просторово-часовим розподілом геофізичного поля визначають будову і властивості об'єкта, що вивчається. Найбільш відповідальним і трудомістким етапом при цьому є геологічна інтерпретація, при проведенні якої геофізична інформація використовується в поєднанні з даними інших видів спостережень (буріння, геохімія та інші дослідження).

Під *оберненою задачею* [8] розуміють розрахунок шуканої петрофізичної характеристики (або відомим способом зв'язаного з нею інтерпретаційного параметра) за показниками свердловинного приладу при певних умовах вимірювань. Таким чином, розв'язання оберненої задачі базується на розв'язках прямих задач. Проте наявність потужних обчислювальних засобів вирішення прямих задач методів ГДС (методи математичного моделювання) не гарантує успішного розв'язання обернених задач. Як парадоксальний приклад можна вважати відомий метод природної радіоактивності в інтегральній модифікації (гамма-метод, ГМ). Даний метод є першим методом ядерної геофізики, що отримав широке застосування в нафтогазовій промисловості. Йому присвячена велика кількість теоретичних та експериментальних робіт вітчизняних та зарубіжних дослідників. Незважаючи на це, ГК до цього часу інтерпретується як метод глинистості виключно на ем-

піричній основі. Широко застосовний інтерпретаційний параметр даного метода не має петрофізичного змісту, для нього не може бути обґрунтована строга петрофізична модель. Саме тому для кількісної інтерпретації даних ГК замість петрофізичної моделі використовують емпіричні кореляційні залежності (якщо можливо їх знайти). Останні встановлюють за допомогою досліджень на зразках керну.

Перечислені методичні недоліки мають принциповий характер. Щоб усунути їх, методика інтерпретації повинна забезпечувати хоча б приблизне автоматичне налаштування інтерпретаційних залежностей відповідно до характеру середовища. Це можливо лише в тому випадку, коли методика оперує не з фіксованими палетками, а з інтерпретаційною моделлю, що дозволяє генерувати інтерпретаційні залежності і вирішувати прямі та обернені задачі для будь-яких умов вимірювань.

Основну проблему комплексної інтерпретації можна сформулювати як проблему створення інтерпретаційної системи, що володіє «геологічним інтелектом». В такій системі характеристики пластів визначаються не лише сукупністю амплітуд показів різних методів ГДС, але і фактичним впливом пластів у визначених циклах осадконакопичення і характеристикою їх складу. З цієї точки зору існуючі системи не володіють геологічним інтелектом.

Проблеми інтерпретації обумовлені принципово неоднозначністю переходу від вимірюваної фізичних полів до шуканих характеристик порід. Вимірювання фізичних полів різної природи в системі «свердловина-пласт» не створюють нових фізичних сутностей, але вводять додаткові взаємодії. Складність інтерпретації, подібно до складності інженерної конструкції, різко зростає із збільшенням числа елементів і зв'язків між ними. Проте, на відміну від розрахунків інженерної конструкції, при комплексній інтерпретації даних ГДС для подолання неоднозначності рішення системи петрофізичних рівнянь вимагається залучення і використання нових законів, які не залежать від характеру породи та в силу вищевказаних складностей вимагають застосування специфікованих парадигм аналізу даних.

На сьогоднішній день одним з перспективних напрямків у побудові нових методик інтелектуального аналізу ГД є використання методів штучного інтелекту, зокрема використання апарату теорії розпізнавання образів.

Серед методів розпізнавання образів [9, 10] виділяють параметричні, непараметричні та евристичні. Методи розпізнавання образів класифікуються наступним чином:

- методи на основі принципу розподілу;
- статистичні методи;
- методи типу «потенціальних функцій»;
- методи обчислення оцінок (голосування);
- методи, що базуються на апараті обчислення висловлювань.

Крім того, істотним для методу, що базується на теорії розпізнавання образів, є спосіб

представлення знань. Виділяють два основних способи:

- інтенціональні представлення – схеми зв'язків між атрибутами (ознаками);
- екстенціональні представлення – конкретні факти (об'єкти, приклади).

Група інтенціональних методів розпізнавання образів включає наступні підкласи:

1. Методи, що базуються на оцінках густин розподілу значень ознак (методи непараметричної статистики).

2. Методи, що базуються на припущенні про клас вирішальних функцій (методи, що використовують як вирішальний алгоритм мінімізації функціоналу ризику чи помилки).

3. Логічні методи, що базуються на апараті алгебри логіки, дають змогу оперувати інформацією, що міститься не лише в окремих ознаках, але і в сполученні їх значень.

4. Лінгвістичні (структурні) методи, які ґрунтуються на використанні спеціальних граматики та нотацій, за допомогою яких можуть бути описані множини властивостей об'єктів, що розпізнаються.

Група екстенціональних методів включає наступні:

1. Метод порівняння з прототипом, що застосовується тоді, коли класи, що розпізнаються, відображаються в просторі ознак компактними геометричними групами.

2. Метод k -найближчих сусідів, в якому рішення про віднесення об'єкта до деякого класу, приймається на основі інформації про належність k його найближчих сусідів.

3. Алгоритм обчислення оцінок (голосування), що полягає в обчисленні пріоритетів (оцінок подібності), що характеризують наближеність еталонного та розпізнавального об'єктів за системою ансамблів ознак, що є системою підмножин заданої множини ознак.

При порівнянні екстенціональних та інтенціональних методів розпізнавання образів використовується аналогія з мозком людини: інтенціональні методи відповідають способу мислення лівої півкулі, що базується на знаннях про статичні і динамічні закономірності структури інформації, що сприймається; екстенціональні методи відповідають способу мислення правої півкулі, що базується на цілісному відображенні об'єктів навколишнього світу. Розглянемо методи побудови методик аналізу, побудованих на базі інтенціональних методів, які ґрунтуються на припущеннях про клас вирішальних функцій (decision functions).

Основною перевагою методів, що базуються на припущенні про клас вирішальних функцій, є чіткість математичної постановки задачі розпізнавання як пошуку екстремуму. Різноманітність методів даної групи пояснюється широким спектром використовуваних функціоналів якості вирішального правила і алгоритмів пошуку екстремуму. Узагальненням даного класу алгоритмів є метод стохастичної апроксимації.

В даному класі алгоритмів розпізнавання образів, змістовне формулювання задачі здійс-

нюється так: нехай задана деяка множина спостережень, які належать до p різних класів. Необхідно, використовуючи інформацію про ці спостереження та їхню класифікацію, знайти таке правило, за допомогою якого можна було б з мінімальною кількістю помилок класифікувати нові спостереження [11, 12, 13]. Спостереження задають вектором x , а його класифікацію – числом ω , $\omega \in [0, 1, \dots, p-1]$. Таким чином, використовуючи послідовність з l спостережень та класифікацій $(x_1, \omega_1), (x_2, \omega_2), \dots, (x_l, \omega_l)$, необхідно побудувати таке вирішальне правило $\omega = F(x)$, використовуючи яке можна класифікувати нові спостереження з найменшою кількістю помилок.

Для формалізації поняття «помилка» приймається припущення про те, що існує деяке правило Φ , що визначає для кожного правила x класифікацію $\omega = \Phi(x)$, яка називається «істинною». Помилкою класифікації вектора x за допомогою правила $F(x)$ називається така класифікація, при якій $F(x)$ і $\Phi(x)$ не співпадають.

Припускається, що в просторі векторів x існує невідома ймовірнісна міра $P(x)$, у відповідності з якою випадково і незалежно виникають ситуації x , що класифікуються з допомогою правила $P(\omega|x)$. Таким чином визначається навчальна послідовність $(x_1, \omega_1), (x_2, \omega_2), \dots, (x_l, \omega_l)$.

Якість вирішального правила $F(x)$ можна подати у вигляді

$$I(F) = \sum_{i=0}^p \int \theta(F(x) - \omega_i) P(\omega_i|x) P(x) dx, \quad (1)$$

де

$$\theta(z) = \begin{cases} 0, & \text{якщо } z = 0 \\ 1, & \text{якщо } z \neq 0 \end{cases}, \quad (2)$$

Звідси видно, що проблема полягатиме в побудові вирішального правила $F(x)$ таким чином, щоб мінімізувати функціонал якості $I(F)$.

Подібною до задачі розпізнавання образів є задача відновлення регресії, передумови до якої формулюються наступним чином: дві множини елементів зв'язані функціональною залежністю, якщо кожному елементу x може бути поставлений у відповідність елемент y . Ця залежність називається функцією, якщо множина x – вектори, а множина y – скаляри. Проте існують і такі залежності, де кожному вектору x ставиться у відповідність число y , що отримане за допомогою випадкового випробування, відповідно до умовної густини $P(y|x)$. Іншими словами, кожному x ставиться у відповідність закон $P(y|x)$, відповідно до якого у випадковому випробуванні здійснюється вибір y .

Існування таких зв'язків відображає наявність стохастичних залежностей між вектором x і скаляром y . Повне знання стохастичної залежності вимагає відновлення умовної густини $P(y|x)$, проте дана задача є досить складною і на практиці (наприклад, при обробці даних вимірювань) може бути зведена до задачі визначення умовного математичного очікування. Ця задача формулюється наступним чином: визначити функцію умовного математичного очікування, тобто функцію, яка кожному x ставить у відповідність число $y(x)$, рівне математичному очікуванню скаляра y : $y(x) = \int yP(y|x)dy$. Функція $y(x)$ в такому випадку називається функцією регресії, а задача відновлення функції умовного математичного очікування – задачею відновлення регресії.

Строга постановка даної задачі є наступною: в деякому середовищі, що характеризується густиною розподілу ймовірності $P(x)$, випадково і незалежно виникають ситуації x . В цьому середовищі функціонує перетворювач, який кожному вектору x ставить у відповідність число y , отримане в результаті реалізації випадкового випробування, відповідно до закону $P(y|x)$. Властивість середовища $P(x)$ і закон $P(y|x)$ невідомі, проте відомо, що існує регресія $y = y(x)$. Необхідно за випадковою незалежною вибіркою пар $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ відновити регресію, тобто в класі функцій $F(x, \alpha)$ відшукати функцію $F(x, \alpha^*)$, найбільш близьку до регресії $y(x)$.

Задача відновлення регресії є однією з основних задач прикладної статистики. До неї зводиться проблема інтерпретації даних прямих вимірювань.

Вирішення задачі базується на наступних припущеннях:

- шукана закономірність зв'язує функціональною залежністю величину y з вектором x : $y = y(x)$;

- метою дослідження є визначення залежності $y = y(x)$ у випадку, коли в будь-якій точці x може бути проведений прямий експеримент для визначення даної залежності, тобто існують проведені прямі вимірювання величини $y^* = y(x^*)$. Однак внаслідок недосконалості експерименту результат вимірювання визначає істинну величину з деякою випадковою похибкою, тобто в кожній точці x визначається не величина $y(x)$, а величина $y = y_x$, де $\xi = y_x - y(x)$ – похибка експерименту, $M\xi^2 < D$;

- в жодній точці x умови проведення експерименту не допускають систематичної похибки, тобто математичне очікування вимірюван-

ня y_x функції в кожній фіксованій точці рівне значенню функції $y(x)$ в цій точці: $M y_x = y(x)$;

- випадкові величини незалежні.

В цих умовах необхідно за скінченною кількістю експериментів відновити функцію $y = y(x)$. Шукана залежність є регресією, а суть проблеми полягає в знаходженні регресії за послідовністю пар $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$. Задачу відновлення регресії прийнято зводити до проблеми мінімізації функціоналу, відповідно з (1.1): $I(\alpha) = \int (y - F(x, \alpha))^2 P(y|x)P(x)dx dy$ на множині $F(x, \alpha) \subset L_p^2$ (L_p^2 – множина інтегрованих з квадратом у міру $P(x)$ функцій) у випадку, коли густина $P(x, y) = P(y|x)P(x)$ невідома, проте задана випадкова та незалежна вибірка пар $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$.

За аналогією до загальних діагностичних задач геолого-геофізичні дослідження можна також віднести до цього класу. Основним завданням при цьому є точне визначення властивостей об'єкта, що досліджується, тобто визначення характеристики, що трактується як своєрідне вираження і конкретизація загальнонаукового поняття «стан», що відображає домінуючий спосіб зміни і розвитку системи при даних умовах, у визначеному місці та часі.

Під геолого-геофізичним діагностуванням будемо розуміти процес, який спрямований на задоволення потреб суб'єкта шляхом безпосереднього впливу на об'єкт через використання законів його існування.

Суттєвою складністю у випадку геолого-геофізичного діагностування є відмінність між теоретичним, вимірним та реальним станом об'єкта.

Комплексна геологічна інтерпретація ГДС має велику кількість методичних підходів, які базуються як на теоретичних так і емпіричних знаннях. Важливою рисою цих методик є їх придатність до формалізації, що дозволяє в свою чергу розробляти та застосовувати комп'ютерну інформаційну технологію. Поняття формалізації конкретизується, як розбиття на самостійні елементи впливів на геологічний об'єкт та відповідей на ці впливи і операції з інформацією, що породжена реакцією об'єкта.

Важливим елементом, який часто не враховується в процесі формалізації методик комплексної інтерпретації даних ГДС, але в той же час є досить важливим та суттєво може впливати на кінцеві результати – це інтуїція та досвід експерта.

Інтуїція [14, 15, 16] – знання, що виникає без усвідомлення шляхів і умов його отримання, в силу чого суб'єкт володіє ним в результаті безпосереднього умовиводу. Інтуїція трактується і як специфічна здатність і як цілісне охоплення умов проблемної ситуації (інтелектуальна інтуїція).

Інтуїцію будемо розглядати як момент виходу за межі логічних програм пошуку рішення

певних задач. Відповідно інтуїція – евристичний процес, який полягає в знаходженні рішення задачі на основі орієнтирів пошуку, не пов'язаних логічно або недостатніх для отримання логічного висновку. Для інтуїції характерна швидкість формулювання гіпотез і прийняття рішень, а також недостатня усвідомленість його логічних основ.

Інтуїція проявляється в умовах суб'єктивно або об'єктивно неповної інформації і є складовою здатності людини до екстраполяції.

Механізм інтуїції складається з симультанного об'єднання декількох інформативних ознак різних модальностей в комплексні орієнтири, що направляють пошук рішення. На такому одночасному врахуванні якісно різної інформації базується різниця інтуїтивних процесів – дискурсивних (логічних), в яких в одному логічному кроці може враховуватися лиш деяка одна модифікація ознак задачі, що зв'язуються між собою.

Орієнтири пошуку в інтуїтивних і дискурсивних процесах не мають принципової різниці за складом вхідної інформації. Логічні принципи, в тому числі формальні, включаються в сформований інтуїтивно інформативний комплекс і, будучи недостатніми для отримання рішення, в сполученні з іншими інформаційними зв'язками визначають напрямок пошуку. Основну роль в інтуїції відіграють семантичні узагальнення, що відносяться до даної проблемної області.

Одним із підходів до вирішення даної задачі є застосування теорії штучних нейронних мереж як потужний засіб формалізації евристик та видобування знань на основі емпіричних даних.

Питання відображення роботи мозку людини в обчислювальних системах є основним напрямком роботи вчених зі створення і дослідження штучних нейронних мереж (ШНМ). В роботі [17] була висловлена ідея про нейрон як структурну одиницю мозку. Проте нейрон має на 5-6 порядків меншу швидкість реакції, ніж напівпровідниковий елемент. Як засвідчили більш пізніші дослідження, секрет високої продуктивності мозку базується на великій кількості нейронів та взаємозв'язків між ними.

Мережа нейронів, що утворюють людський мозок, представляють собою високоефективну, комплексну, нелінійну, паралельну систему обробки інформації. Така система здатна організовувати свої нейрони так, щоб реалізувати сприйняття образу, його розпізнавання чи здійснювати керування набагато швидше, ніж це здатні робити сучасні обчислювальні системи.

ШНМ є спрощеною моделлю мозку. Вона будується на основі штучних нейронів, які володіють тією ж основною властивістю, що і біологічні – пластичністю. Використання структури мозку і пластичності нейронів робить ШНМ універсальною системою обробки інформації. В загальному випадку ШНМ – це машина, що моделює роботу мозку. Зазвичай ШНМ реалізуються у вигляді електронних при-

строїв чи комп'ютерних програм. Серед багатьох визначень нейронних мереж можна виділити ті, що характеризують ШНМ як адаптивну систему, зокрема в роботі [17]: штучна нейронна мережа – це паралельний розподілений процесор, що володіє природною схильністю до збереження дослідного знання і можливістю його надання користувачу. Подібність ШНМ та мозку можна розглядати з двох точок зору: знання здобуваються мережею в процесі навчання; для збереження знання використовуються сили міжнейронних з'єднань, які також називаються синаптичними вагами.

Процедура, яка використовується для здійснення процесу навчання, називається алгоритмом навчання. Її функція полягає в модифікації визначеним чином синаптичних ваг ШНМ так, щоб вона набула необхідних властивостей [7].

Модифікація вагових коефіцієнтів є традиційним методом навчання ШНМ. Для прикладу, такий підхід є близьким до теорії адаптивних лінійних фільтрів, що досить давно застосовується в задачах керування. Проте, для ШНМ існує також можливість модифікації власної топології, що базується на біологічних особливостях нейронів головного мозку – вони можуть виникати, вмирати і змінювати свої зв'язки з іншими нейронами.

З наведеного вище, зрозуміло, що ШНМ реалізують свою обчислювальну потужність на основі двох основних властивостей: 1) суттєво розпаралелено-розподілена архітектура; 2) здатності до навчання та узагальнення отриманих знань. Під властивістю узагальнення розуміється здатність генерувати правильні виходи для вхідних сигналів, які не були враховані в процесі навчання. Таким чином штучні нейронні мережі можна вважати інтелектуальними системами аналізу даних та видобування знань, які здатні вирішувати складні багатомірні задачі на основі знань здобутих з навчальної вибірки даних.

З роботи [17] розпочався сучасний етап розвитку ШНМ, в цій роботі було введено поняття «порогового логічного нейрона» і описано логічну модель ШНМ. Ранні етапи досліджень були пов'язані з синтезом автоматичних систем керування на основі ШНМ. Було описано ряд важливих концепцій керування, а пізніше додано розділи, присвячені навчанню, самоорганізації та нейрокібернетиці. Отримані результати показали, що не лише активація, але і гальмування повинні бути присутні в процесі навчання. Показано, що ШНМ з модифікованими синапсами може бути натренована для вирішення задачі класифікації простих наборів бінарних шаблонів. Висунуто гіпотезу про те, що ефективність змінного синапсу в нервовій системі залежить від статистичних взаємозалежностей між станами, що змінюються з обох сторін від синапса, тим самим проводячи зв'язок з теорією інформації Шенона. В роботі була висловлена ідея про те, що неспадкоємність адаптивної поведінки живих систем і здатність до адаптації проявляється в процесі навчання. В роботі були розглянуті задачі пошу-

ку, розпізнавання образів, навчання, планування та індукції і робота містила огляд успішних евристик їх вирішення, зокрема і за допомогою ШНМ. Необхідно відзначити важливість робіт фон Неймана по створенню цифрових комп'ютерів для всього кібернетичного напрямку. Ці теоретичні розробки створили базу для створення першої електронної моделі штучної нейронної мережі – перцептрона, яка була призначена для вирішення задачі розпізнавання шаблонів і за допомогою якої було проведено успішні експерименти з навчання розпізнаванню простих зображень. В роботі також була введена теорема про збіжність перцептрона. Запропоновано було також метод найменших квадратів і його використання для опису ШНМ Adaline. Основна різниця між перцептроном і Adaline полягала в процедурах їхнього навчання. Одна з перших ШНМ із шаровою структурою була запропонована архітектура Madaline. Не менш важливим було введення сигмоїдальної активаційної функції для моделі логічного нейрона.

Проте нейромережеві обчислення мають і ряд недоліків. Зокрема, в роботі [17] строго математично було доведено існування фундаментальних обмежень можливостей одношарового перцептрона. Також показано, що одношарова мережа довільної розмірності не здатна вирішувати задачі, що доступні для багатшарових конфігурацій. Іншим фактором гальмування як теорії так і практики нейротехнологій була відсутність персональних комп'ютерів та низька ефективність існуючих на той час обчислювальних систем. З іншого боку були отримані значні результати в розвитку карт самоорганізації на основі конкурентного навчання.

З початку 80-х років розпочалося відновлення досліджень питань ШНМ – з однієї сторони були розроблені нові моделі ШНМ та здійснено більш ґрунтовне формальне представлення методів їх навчання, а з іншої – були з'ясовані фундаментальні обмеження формальних методів при моделюванні функцій інтелекту [7]. Зокрема Хопфілдом була розроблена ШНМ з повністю зв'язаною архітектурою і оригінальним алгоритмом налаштування вагових коефіцієнтів. Проривом в застосуванні ШНМ до практичних задач є опис алгоритму зворотного поширення. Це був перший ефективний алгоритм навчання багатшарового перцептрона довільної структури.

Теорія ШНМ РБФ-мереж, на сьогоднішній день, є останнім значним відкриттям у даній області. РБФ-мережі є альтернативною багатшаровою перцептронною схемою, в якій використовуються приховані нейрони з радіально-базисною функцією активації. Основою цієї ідеї є метод потенціальних функцій.

Загалом поняття «штучна нейронна мережа» охоплює ансамблі будь-якої структури (ensemble structure), однак практичне застосування знайшли лише деякі з них. Це пояснюється тим, що архітектура ШНМ безпосередньо зв'язана з методом її навчання, що чітко видно з етапів розвитку ШНМ.

Штучний нейрон (чи просто нейрон) є елементарною функціональною одиницею, з множини яких формуються ШНМ. Він представляє собою модель живого нейрона, проте відображає лише здійснювані ним перетворення, а не спосіб функціонування існуючих логічних, неперервних та імпульсних моделей нейронів. Формальний нейрон містить наступні елементи:

1. Адаптивний суматор обчислює скалярний добуток вектора вхідного сигналу на вектор параметрів. Адаптивним він називається через наявність векторів параметрів, що налаштовуються – синаптичні ваги та порогових елементів.

2. Функція активації – нелінійний перетворювач сигналу – отримує скалярний вхідний сигнал і відповідним чином його перетворює.

Формальний нейрон реалізує передавальну (активаційну) функцію загального виду:

$$N(x_1, x_2, \dots, x_n) = Act \left(\sum_{i=1}^n w_i \cdot x_i + b \right). \quad (3)$$

З формальних нейронів можна утворювати шари нейронів, які, в свою чергу, можна об'єднувати в багатшарові мережі, зокрема багатшарові перцептрони.

Нейрони вхідного шару отримують сигнали, перетворюють їх і через точки розгалуження передають до нейронів прихованого шару. Далше відбувається спрацьовування наступного шару і т.д. Вихідний шар активується останнім і видає результат роботи мережі користувачу (системі). Кожний вихід нейронів будь-якого шару подається на вхід всіх нейронів наступного шару. Кількість нейронів в шарі є теоретично довільною величиною.

Основним етапом видобування знань за допомогою ШНМ є процес навчання. ШНМ навчаються, шляхом покращення свого функціонування через налаштування своїх параметрів під впливом зовнішнього середовища. Існує велика кількість визначень терміну «навчання», однак стосовно ШНМ, таким, що найбільш підходить, є наступне: навчання – це процес, при якому вільні параметри нейронної мережі адаптуються в результаті її неперервної стимуляції зовнішнім середовищем. Тип навчання визначається тим способом, за допомогою якого відбувається зміна параметрів.

Поряд з терміном «навчання» досить часто використовуються рівноправні поняття «тренування мережі» і «налаштування параметрів мережі».

При вирішенні задач аналізу даних, використовуються два підходи до навчання – контрольований, тобто такий, що базується на наявності вчителя, та неконтрольований – без вчителя. Існує два види контрольованого навчання – пряме контрольоване навчання та навчання із стимуляцією.

Враховуючи досвід впровадження інформаційних технологій на підприємствах нафтогазової промисловості як в Україні, так і в світі, можна зробити висновок, що розробка інтелектуальних інформаційних систем та систем під-

тримки прийняття рішень для використання на підприємствах даної галузі є одним із способів підвищення ефективності використання природних ресурсів та зменшення витрат на їх видобування. При розробці таких систем необхідно враховувати специфіку підприємств. Загалом, підприємство нафтогазової промисловості можна представити як одну комплексну систему, яка складається з підсистем, кожна з яких забезпечує певний фактор нафтогазового виробництва, зокрема виділимо основні – дослідження та буріння свердловин, розробка родовищ, транспортування нафти і газу, переробка та реалізація продукту. Кожна підсистема, відповідно може бути розділена на певну кількість підкласів.

При створенні інтелектуальної технології, першочерговим є виділення кола завдань і визначення класів задач на вирішення яких вона має бути зорієнтована першочергово. Враховуючи значні витрати, які необхідні на дослідження родовищ та ризики, пов'язані з цим, наше дослідження зосереджується на застосуванні інтелектуальних інформаційних технологій для здійснення прогнозування та виявлення покладів вуглеводнів з метою забезпечення процесу техніко-економічного обґрунтування, а саме початкової геолого-економічної оцінки та попередньої геолого-економічної оцінки.

Проаналізувавши методи зберігання промислової інформації на нафтогазових підприємствах виявлено такі особливості:

1) значна кількість даних, часто представлена в агрегованому або підсумованому (за рік) вигляді;

2) дані зберігаються в різних форматах і часто дублюються;

3) велика кількість інформації по одному об'єкту (свердловині, родовищу) розміщена в різних базах даних;

4) значна кількість окремих баз даних.

Дані, збережені в такому вигляді, важко піддаються аналізу як ручному, так і автоматизованому. Розробка деякої нової структури бази даних не вирішить даних проблем, тому для забезпечення ефективної аналітичної обробки, провівши попередній аналіз, вирішено основну увагу приділити концепції сховищ даних (data warehouse) та технології здобуття знань (data mining) як основному інструменту інтелектуалізації існуючих інформаційних систем на основі баз даних та знань [7].

В основу сучасної технології Data Mining (discovery-driven data mining) [18] покладена концепція шаблонів (паттернів), що відображають фрагменти багатоаспектних залежностей і зв'язків між даними. Ці шаблони являють собою закономірності, які властиві підвибіркам даних і можуть бути компактно виражені в зрозумілій людині формі. Пошук шаблонів здійснюється методами, не обмеженими рамками апіорних припущень про структуру вибірки та види розподілів значень аналізованих показників.

Традиційна математична статистика та методи математичного моделювання тривалий час

претендувала на роль основного інструмента аналізу ГДД, проте на сьогоднішній день вона не забезпечує вирішення задач аналітичної обробки даних. Головна причина – концепція усереднення по вибірці, що вимагає операцій над фіктивними величинами (типу середньої температури пласта, середньої пористості пласта, який складається як з мікро- так і макрокаверн). Методи математичної статистики виявилися корисними насамперед для перевірки задалегідь сформульованих гіпотез (verification-driven data mining) і для “грубого” аналізу, що становить основу оперативної аналітичної обробки даних (online analytical processing, OLAP). Методи математичного моделювання є ефективними при аналізі даних окремо взятого методу ГДС, проте вони не забезпечують задовільного вирішення проблем комплексного аналізу як методів ГДС, так і геолого-геофізичних даних.

Вибір алгоритму здобуття знань залежить, насамперед, від природи поставленого завдання, тому виділимо три основні: прогнозування (класифікація або навчання з тьютором), кластеризація (сегментація або навчання без тьютора) та виділення шаблонів. Модель даних для цих алгоритмів є досить простою – дані розглядаються як колекція записів, де кожен запис є колекцією полів.

Важливим аспектом при створенні інтелектуальних технологій є побудова онтології предметної області, зокрема нафтогазової, яка є основою для розробки сховища даних та ефективного здобуття знань.

Незважаючи на велику кількість методів Data Mining, найбільшого застосування набули логічні алгоритми пошуку в даних правил типу if-then. Зокрема, використовуючи їх, вирішуються завдання прогнозування, класифікації, розпізнавання образів, сегментації БД, видобування з даних “прихованих” знань, інтерпретації даних, встановлення асоціацій в БД і ін. Результати таких алгоритмів ефективні та легко інтерпретуються. Головною проблемою логічних методів виявлення закономірностей є проблема перебору варіантів за прийнятний час. Відомі методи або штучно обмежують такий перебір (алгоритми KOPA, WizWhy [18]), або будують дерева рішень (алгоритми CART, CHAID, ID3, See5, Sipina [18]), що мають принципові обмеження ефективності пошуку if-then правил. Інші проблеми пов'язані з тим, що відомі методи пошуку логічних правил не підтримують функцію узагальнення знайдених правил і функцію пошуку оптимальної композиції таких правил.

Таким чином, проведений аналіз підходів, методів та засобів інтерпретації дозволяє виділити формальну структуру очікуваних знань, що видобуваються в результаті інтерпретації даних ГДС в формі базового подання, розвинутого в роботі [19].

Означення 1. Знаннями про нафтогазовий об'єкт будемо вважати інтерпретацію в термінах штучного інтелекту сукупності даних, понять, фактів та закономірностей нафтогазової

предметної області, що об'єднуються певною ознакою, яка уможливорює представлення цілісної структури. В такому розумінні предметні знання можна розглядати як моделі інтелектуальної поведінки експертів предметної області (геологів, фахівців з проблем видобутку та транспортування нафти і газу і т. д.).

Початкову формалізацію знань предметної області (*SDK* – *Subject Domain Knowledge's*) представимо у вигляді залежності множини задач предметної області (*SDT* – *Subject Domain Tasks*) і доступних ресурсів предметної області (*SDR* – *Subject Domain Resources*) у вигляді множини правил (*Rules* – *Subject Domain Rules*), а саме:

$$Knowledges^{set} = (SDT^{set}, SDR^{set}, Rules^{set}). \quad (4)$$

Проте, незалежно від способу представлення та функціональності, наявність входження знань в певній інформаційній системі не можуть характеризувати її як систему інтелектуального рівня. Вирішення задачі інтелектуалізації в даному випадку полягає у створенні баз знань як якісної систематизації та класифікації зв'язків та відношень між об'єктами предметної області.

Означення 2. Базою знань нафтогазової предметної області (*KB* – *Knowledge Base*) будемо називати структуру представлення та управління наборами фактів, правил та процедур логічного висновку на множині нафтогазових об'єктів. Класичне рішення задачі розробки баз знань вимагає: розробки програмних структур та засобів для представлення та зберігання знань; збір та розміщення знань предметної області в базі.

За допомогою чіткої математичної основи категорій та функторів можливо досліджувати семантичний зміст правильно з математичної точки зору (шляхом побудови різноманітних моделей представлення знань), що необхідно для формалізації знань, побудови баз знань та інтелектуальної підтримки прийняття рішень.

Логічна модель призначена для вирішення простих однорідних задач. Вона базується на алгебрі тверджень та предикатів, системах, аксіомах та методах висновку. Відповідно у кожній логіці наявний синтаксис, що визначає правила побудови синтаксично та семантично коректних висловлювань, і семантики як правил інтерпретації логічних тверджень.

В основі логічного представлення знань (*LKP* – *Logical Knowledge's Presentation*) лежить формальна теорія, що виражається кортежем:

$$LKP = (BS^{set}, (F : bs \in BS^{set} \rightarrow Statements^{set})^{set}, TF, FR), \quad (5)$$

де BS^{set} – скінченна множина базових символів;

F^{set} – множина формул у вигляді синтаксичних правил, за допомогою яких з елементів BS^{set} формуються коректні твердження;
 TF – множина істинних формул;

FR – множина відношень між формулами (правилами).

Крім того, для множини BS^{set} повинні існувати процедури верифікації щодо належності елементів та входжень синтаксично правильних сутностей.

Основні проблеми імплементації логічної моделі полягають в усуненні суперечностей даних та наявності процедур (правил) оновлення входжень кортежу логічного представлення.

Логіка тверджень розглядає судження як єдине ціле, що має лише властивість істинності. За необхідності розробити логічну модель знань зручно використовувати логіку предикатів. Логіка предикатів більше підходить для відтворення логічних міркувань, оскільки дозволяє врахувати будову і зміст простих висловлювань.

Означення 3. Численням предикатів вважають формальну мову подання логічних та інших видів відношень, результатом яких є булева величина.

Центральним твердженням у численні предикатів є те, що кожне просте висловлювання містить в собі інформацію про деякий об'єкт і його властивості. Числення предикатів має потужний і зрозумілий механізм рекурсивного висновку. Важливою властивістю є також широкий набір виконуваних операцій, зокрема квантор існування (\exists) і квантор узагальнення (\forall), що дозволяє описувати загальні властивості множини об'єктів.

Відповідно базу знань на основі предикатів можна представити у вигляді:

$$KB = (A, SR^{set}, TF^{set}, IR^{set}), \quad (6)$$

де A – алфавіт (сукупність використовуваних символів);

SR^{set} – множина синтаксичних правил;

TF^{set} – початкова множина істинних формул;

IR^{set} – множина правил висновку для побудови похідних формул.

Інтерпретація логіки предикатів наступна. Нехай $Object^{set}$ – деяка множина об'єктів і O_1, O_2, O_3 – елементи цієї множини, відповідно логічні висловлювання про ці об'єкти позначатимемо як $State(O_1), State(O_2), State(O_3), State(O_4)$. Згідно з початковим припущенням, дані висловлювання будуть істинними або хибними.

Нехай x – випадковий елемент з $Object^{set}$, тобто $x \in Object^{set}$, тоді $State(x)$ позначатиме висловлювання, що стає визначеним, коли x замінюється конкретним елементом з $Object^{set}$. З іншого боку $State(x)$ представлятиме булеву функцію на множині $Object^{set}$. Відповідно невизначені функції будемо розглядати як предикатні, причому в таких предикатах змінні можуть бути як з одного алфавіту, так і з різних:

$$[Object^{set}]^n \rightarrow Bool^{set}, \quad (7)$$

$$[Object^{set}]^1 \times [Object^{set}]^2 \times \dots \times [Object^{set}]^n \rightarrow Bool^{set}, \quad (8)$$

де $[Object^{set}]^1, [Object^{set}]^2, \dots, [Object^{set}]^n$ – алфавіти;

$Bool^{set}$ – булева множина.

Застосування семантичних моделей в інженерії знань бере свій початок від методів семантичного аналізу природньомовних текстів, що з розвитком мов програмування призвело до створення їх семантик, а на сьогоднішній день концепція семантизації WEB розглядається як найбільш актуальний тренд розвитку WEB та Internet технологій.

Означення 4. Семантичною мережею (*Semantic Network*) вважають спосіб представлення інформації, яка базується на графовій інтерпретації. Вважається, що предметну область можна представити у вигляді семантичної мережі якщо її модель можна звести до набору двомісних предикатів, які згідно свого означення дозволяють подання у вигляді графів, де ребрам відповідають предикати, а вершинам аргументи. Таким чином, семантичну мережу можна розглядати як орієнтований граф з таких складових:

1) множина вершин (*Vertices Set*), яка використовується для опису та представлення даних та об'єктів.

2) множина ребер (*Edges Set*) дозволяє моделювати відношення між об'єктами.

Таким чином, знання, що базуються на семантичній мережі, можна формально представити кортежем:

$$K^S = \langle VS, A^{VS}, A.VS, RTS, PS[VS] \rangle, \quad (9)$$

де K^S (*Knowledges^{set}*) – множина знань;

VS (*VerticesSet*) – множина вершин графа;

A^{VS} (*Attributes^{VerticesSet}*) – множина атрибутів, приписаних вершинам;

$A.VS$ (*Attributes.ValuesSet*) – множини значень атрибутів;

RTS (*RelationsTypesSet*) – множина типів зв'язків, введених в мережі;

$PS[VS]$ (*ProjectionsSet[VerticesSet]*) – множина проєкцій для вершин за видами зв'язків.

Визначимо типовий алгоритм імплементації для розробки інтелектуальної технології видобування знань на основі ГГД, який складатиметься з наступних кроків:

1) визначення моделі здобуття знань нафтогазової предметної області; ідентифікація наборів атрибутів даних, які будуть спрогнозовані, і наборів атрибутів, які будуть використані для прогнозування – визначення інформаційних потоків; визначення ефективної послідовності кроків для здобуття знань;

2) заповнення моделі даними, використовуючи вибрану методику;

3) прогнозування атрибутів для нових даних, використовуючи модель прогнозування;

4) збереження отриманих даних та їхнє представлення за допомогою засобів візуалізації або формування звітів.

Висновки

З проведених досліджень задач, методів та проблем аналізу геолого-геофізичних даних можна бачити, що розвиток комп'ютерних технологій комплексної інтерпретації даних ГДС призвів до розвитку теорії методів ГДС та вдосконалення їх інтерпретаційних та петрофізичних моделей. Інтерпретаційно-алгоритмічне забезпечення методів ГДС розвивається за двома основними напрямками – розвитку методів індивідуальної та комплексної інтерпретації, що реалізуються у вигляді програмно-методичних комплексів. Все більш актуальною стає проблема розробки та реалізації таких інтерпретаційних процедур, які б максимально використовували можливості обчислювальної техніки та були здатні вирішувати виділений клас задач, які є слабоформалізованими і складними в імплементації існуючими засобами. Визначено, що найбільш актуальною задачею є розробка технології, яка б повною мірою могла використовувати та узагальнювати геолого-геофізичну інформацію та імплементувати концепцію геологічного інтелекту. Вирішення даної проблеми відноситься до теорії та практики штучного інтелекту, а саме теорії розпізнавання образів. Одним із ефективних, новітніх та малодосліджених підходів вирішення даної задачі є застосування теорії штучних нейронних мереж як потужного засобу формалізації евристик та видобування знань на основі емпіричних даних. Особливістю розвитку технології штучних нейронних мереж є їхня висока дослідженість в питаннях застосування до задач автоматизованого керування, проте набагато менше прикладів їх використання для вирішення задач інтелектуального аналізу даних, в рамках концепції Data Mining.

Література

1 Історія розвитку та сучасний стан інтерпретаційних технологій ГДС / М.Д. Красножон – Київ: Видавничий центр УкрДГРІ, 2001. – 72 с.

2 Элланский М.М. Использование многомерных связей в нефтегазовой геологии / М.М. Элланский, Б.Н. Еникеев. – М.: Недра, 1991. – 205 с.

3 Добрынин В.М. Петрофизика / В.М. Добрынин, Б.Ю. Вендельштейн, Д.А. Кожевников. – М.: Недра, 1991.

4 Кузнецов О.Л. Геоинформатика и геоинформационные системы / О.Л. Кузнецов, А.А. Никитин, Е.Е. Черемисина. – М.: ВНИИ-геосистем, 2005. – 350 с.

- 5 Поспелов Д.А. Данные и знания. Представление знаний // Искусственный интеллект. Кн.2: Модели и методы: справочник. – М.: Радио и связь. – С.7-13.
- 6 Юрчишин В. М. Інформаційне моделювання нафтогазових об'єктів : монографія / В. М. Юрчишин, В. І. Шекета, О. В. Юрчишин – Івано-Франківськ: Вид-во Івано-Франківського нац. техн. ун-ту нафти і газу, 2010. – 196 с.
- 7 Формально-логічні методи створення інтелектуальної технології видобування знань на основі даних геофізичних досліджень: магістерська робота: 8.080403 / М.Б.Горбаль / наук. кер. – проф., д.т.н. Юрчишин В.М / Нац. техн. ун-т нафти і газу. – Івано-Франківськ, 2010. – 152 с.
- 8 Кожевников Д.А. Проблемы интерпретаций данных ГИС / Д.А. Кожевников // Геофизическое общество «Каротажник». – 2001. – С. 20.
- 9 Журавлев Ю.И. Распознавание образов и анализ изображений / Ю.И. Журавлев, И.Б. Гуревич; в 3-х кн. Кн. 2. – М.: Радио и связь, 1990. – 304 с. – Сер.: Искусственный интеллект
- 10 Журавлев Ю.И. Непараметрические задачи распознавания образов / Ю.И. Журавлев // Кибернетика. – 1976. – №6. – С. 93-103.
- 11 Прикладная статистика. Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, С.И. Енюков, Л.Д. Мешалкин – М.: Финансы и статистика, 1989. – 607 с.
- 12 Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА / И.С. Енюков. – М.: Финансы и статистика, 1986. – С. 232.
- 13 Алгоритмы и программы восстановления зависимостей / В.Н. Вапник, Т.Г. Глазкова, В.А. Кошечев, А.И. Михальский, А.Я. Червоненкис. – М.: Наука, 1984. – С. 8-11, 27-32, 42-55.
- 14 Логико-гносеологические и методологические проблемы прогноза. – М.: Наука, 1986.
- 15 Дюк В.А. Компьютерная психодиагностика / В.А. Дюк. – Санкт-Петербург: Братство, 1994.
- 16 Цыпкин Я.З. Адаптация и обучение в автоматических системах / Я.З. Цыпкин. – М.: Наука, 1968. – 400 с.
- 17 Aleksander I., Morton H. An Introduction to Neural Computing. – London: Chapman & Hall, 1990. – 218 p.
- 18 Deroski S. Multi-relational data mining: an introduction // SIGKD D Explor. Newsl. – 2003. – Vol.5(1) – P. 1-16.
- 19 Демчина М.М. Формальні методи інтерпретації даних та знань про нафтогазові об'єкти / М.М. Демчина, В.Р. Процюк, В.І. Шекета // Науковий вісник Івано-Франківського національного технічного університету нафти і газу. – 2011. – №1. – С. 100-108.

*Стаття надійшла до редакційної колегії
04.09.12
Рекомендована до друку професором
Д.Ф. Тимківим*