

УДК 622.24:681.3

DOI <https://doi.org/10.32782/2663-5941/2025.3.2/58>**Чигур І.І.**

Івано-Франківський національний технічний університет нафти і газу

Лагойда А.І.

Івано-Франківський національний технічний університет нафти і газу

Чигур Л.Я.

Івано-Франківський національний технічний університет нафти і газу

ІНФОРМАТИВНІ ОЗНАКИ В ЗАВДАННЯХ АВТОМАТИЗОВАНОГО РОЗПІЗНАВАННЯ ТЕХНОЛОГІЧНИХ СИТУАЦІЙ ПРИ БУРІННІ СВЕРДЛОВИН

Стаття присвячена актуальній проблемі автоматизованого розпізнавання та класифікації технологічних ситуацій, що виникають у процесі буріння свердловин, як важливого етапу для прийняття обґрунтованих рішень при проектуванні та управлінні цим складним технологічним процесом.

Зазначено, що в процесі буріння виникає необхідність ідентифікувати різноманітні фактори, такі як зміни буримості та тріщинуватості гірських порід, ступінь зношення бурового інструменту, а також геологічні ускладнення та аварії. Інформація, отримана в результаті розпізнавання, є критично важливою для ефективного керування процесом буріння, визначення оптимального моменту для підйому бурового інструменту та, що особливо важливо, для попередження аварійних ситуацій.

Запропоновано завдання розпізнавання образів звести до виділення елементів, що належать до конкретного класу, серед множини елементів, які можуть належати до різних класів за допомогою дивергенції. Підкреслюється, що наявність певних спільних властивостей або ознак у елементів одного класу є ключовим критерієм для їх ідентифікації та віднесення до цього класу.

Розглядається важлива проблема попередньої обробки та стиснення вихідних даних у задачах розпізнавання технологічних ситуацій з метою зменшення розмірності вектору вимірювань.

Запропоновано два основні підходи до зменшення числа ознак: селекція найбільш інформативних ознак шляхом відсікання менш значущих та створення невеликої кількості нових, комбінованих ознак на основі наявних. Підкреслено, що другий підхід, хоча й більш трудомісткий, потенційно забезпечує вищу інформативність новостворених ознак за рахунок комбінування вихідних.

Наголошено на необхідності мінімізації втрати інформативності при відборі або комбінуванні ознак, пропонуючи використовувати поняття дивергенції як міри інформативності. Запропоновано подальше застосування дивергенції для скорочення кількості ознак на прикладі двох важливих випадків розподілу даних: нормального розподілу та розподілу Бернуллі, з аналізом обох способів зменшення розмірності (відсікання та лінійне перетворення ознак) для кожного з них.

Таким чином, стаття є вступом до важливої науково-технічної проблеми розпізнавання та класифікації технологічних ситуацій у бурінні, окреслює її актуальність, аналізує існуючі підходи та пропонує потенційні інформаційні основи для подальших досліджень у цій галузі.

Ключові слова: автоматизоване розпізнавання, буріння, інформативні ознаки, дивергенція, розподіл Бернуллі.

Постановка проблеми. Прийняттю рішень при проектуванні або управлінні процесом буріння завжди передують процес розпізнавання конкретних ознак, що характеризують умови буріння і визначають вибір найбільш раціональної техніки, технології та організації бурових робіт [1]. Особливо актуальним є розпізнавання технологічних ситуацій при бурінні свердловин. До недавнього часу таке розпізнавання здійснювалося людиною на основі досвіду, знань, інтуїції.

У процесі буріння необхідно розпізнавати гірські породи різної буримості і тріщинуватості, стан зношення породоруйнівного інструменту, геологічні ускладнення і аварії. Результати розпізнавання використовуються для керування процесом буріння, визначення моменту підйому породоруйнівного інструменту, попередження аварій [2]. Природа розпізнавання об'єктів істотно впливає на процес розпізнавання і вирішальні схеми (алгоритми), що дозволяють на

основі аналізу значень ознак об'єкта (ситуацій) зробити висновок про те, до якого класу належить даний об'єкт (ситуація).

Традиційні заходи попередження аварій і ускладнень зводилися в основному до контролю стану обладнання, інструменту, бурового розчину і виконання заданих технологічних параметрів ведення процесу. Попередження аварійних ситуацій в процесі буріння ґрунтувалося на досвіді бурильника, узагальнити і формалізувати який у вигляді рекомендацій без відповідних засобів фіксації процесу досить важко.

Аналіз останніх досліджень і публікацій. Першими дослідженнями по формалізації і систематизації, інформаційних моделей у вигляді аналогових записів витрат потужності при алмазному бурінні з'явилися роботи О. І. Шерстюка, Н. Н. Міхеєва та ін. Інформаційною основою для розпізнавання технологічних ситуацій служили спеціальні форми запису витрат потужності для різних вибійних процесів (припалювання коронки, заповірювання алмазів і ін.). Однак аналіз аналогових записів самописного ваттметра показав велику різноманітність і, що особливо важливо, неможливість однозначного тлумачення форм запису, а отже, і розпізнавання ситуацій.

Дослідженнями інформаційних ознак процесу буріння в тому числі з метою ідентифікації його станів займалися такі вчені як Г. Н. Семенцов, М. І. Горбійчук, М. А. Мислюк, А. І. Волобуєв, В. Б. Кропивницька, А. А. Погарський, Р. М. Ейгелес, Е. А. Козловський, Г. Д. Бревдо, Н. Б. Ситніков, Ю. А. Пешалов, Ю. Ф. Алексєєв, А. І. Співак, В. А. Бражніков, В. К. Маурер (США), І. М. Галлі (США), Г. Б. Вудс (США), А. Т. Боургойн (США), Ф. С. Йоунг (США) та ін.

На основі проведеного аналізу публікацій встановлено, що інформаційною основою розпізнавання можуть служити параметри і показники технологічного процесу буріння свердловин (осьове навантаження, крутний момент, витрати потужності, витрата очисного агента, швидкість буріння, інтенсивність зміни цих параметрів і ін.), Енергоємність руйнування породи, параметри коливань бурильної колони (частота, амплітуда і ін.), властивості промивальної рідини (щільність, в'язкість, ступінь аерації) на вході і виході зі свердловини, статистичні характеристики перерахованих ознак (дисперсія, спектральна густина і ін.) [3].

Постановка завдання. Метою статті є розробка методики оцінки інформативності ознак, що характеризують процес буріння свердловин

для забезпечення ефективного розпізнавання технологічних ситуацій, які виникають у процесі роботи.

Виклад основного матеріалу. Загальне завдання розпізнавання полягає у виділенні елементів, що належать конкретному класу, серед безлічі елементів, що відносяться до декількох класів. Елементи, що відносяться до одного класу, мають ряд загальних властивостей або ознак. Наявність цих ознак і служить тим критерієм, який дозволяє відносити конкретні елементи до даного класу.

Однак об'єкти, що підлягають розпізнаванню, можуть мати дуже велику кількість ознак, інформативність яких може бути різною. Для формалізованої оцінки інформативності ознак використовують характеристики розсіювання випадкових величин і різні ентропійні заходи, отримані із статистичної теорії інформації. У таких завданнях доцільно використати дивергенцію в якості міри «відстані» або відмінності між класами. Дивергенція – це повна середня інформація для розрізнення класів. У разі неперервних розподілів випадкової величини x вона задається формулою

$$I = \int_{-\infty}^{+\infty} (p_1(x) - p_2(x)) \ln \frac{p_1(x)}{p_2(x)} dx, \quad (1)$$

де $p_1(x)$ і $p_2(x)$ – щільності першого і другого розподілів.

Для дискретних розподілів існує аналогічна формула:

$$I = \sum_a (p_1(a) - p_2(a)) \ln \frac{p_1(a)}{p_2(a)}, \quad (2)$$

де випадкова величина x може приймати N різних значень a з вірогідністю, рівними $p_1(a)$ і $p_2(a)$ для першого і другого розподілів відповідно.

Надалі для спрощення викладу вважатимемо, що є два класи об'єктів, перехід до більшої кількості класів здійснюється по індукції. Для кожного з класів розподіл об'єктів за ознаками є деяким імовірнісним розподілом. Нехай загальна кількість ознак становить n , а $x_i^{(j)}$ – значення i -ї ознаки у j -го об'єкта, $i = 1, 2, \dots, n$. Тоді об'єкт можна представити як n -вимірний вектор $x^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})$ або точку в n -вимірному просторі R^n , а сукупність всіх ознак утворює n -вимірний простір R^n .

Вираз для визначення дивергенції залежить від закону розподілу ознак. У бурінні вимірювані ознаки і показники мають, як правило, неперерв-

ний розподіл. Багато з них підпорядковуються нормальному або близькому до нього закону розподілу.

У деяких задачах розпізнавання аварійних ситуацій зручніше розглядати ознаки, які приймають лише два значення: +1 при перевищенні порогового значення деяким параметром і -1 при неперевіщенні цього значення. Оцінки ймовірностей перевищення порогу, тобто того, що ознаки приймуть значення +1, визначаються на основі збору статистичних даних в конкретних умовах. У цьому випадку розподіл ознак є дискретним, причому ознаки x_i , приймають лише два значення. У цьому випадку кількість різних векторів із цих ознак становить 2^n , де n – кількість ознак. Якщо ж ознаки незалежні, то випадкова величина x має розподіл Бернуллі.

При виведенні формул дивергенції для двох найбільш важливих для буріння законів розподілу використовувалася властивість адитивності дивергенції:

$$I(\xi^{(1)}, \xi^{(2)}) = \sum_{i=1}^n I(\xi_i^{(1)}, \xi_i^{(2)}). \quad (3)$$

Зауважимо, що ознаки $\xi_i^{(1)}, \xi_i^{(2)}, i=1, \dots, n; j=1, \dots, n$ також є векторними величинами.

Знайдемо спочатку дивергенцію в разі нормального розподілу. Нехай є два нормальних розподіли з однаковими матрицями коваріацій $C^{(1)} = C^{(2)} = C$. У цьому випадку щільності нормальних розподілів представляють собою наступні функції:

$$p_1(x) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp \left\{ -\frac{1}{2} [C^{-1}(x - m^{(1)})] \right\}, \quad (4)$$

$$p_2(x) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp \left\{ -\frac{1}{2} [C^{-1}(x - m^{(2)})] \right\}, \quad (5)$$

де $x = (x_1, \dots, x_n)$ – вектор ознак; $m^{(1)} = (m_1^{(1)}, \dots, m_n^{(1)})$ і $m^{(2)} = (m_1^{(2)}, \dots, m_n^{(2)})$ – вектори математичних очікувань першого і другого розподілу відповідно.

В роботі [6] доведено, що

$$I = (C^{-1}\delta, \delta), \delta = m^{(1)} - m^{(2)}. \quad (6)$$

Візьмемо в якості базису сукупність власних векторів матриці C . У цій системі координат матриця C має вигляд:

$$C = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

У зв'язку з цим вираз (6) перетворюється в наступний:

$$I = \sum_{i=1}^n \frac{\delta_i^2}{\sigma_i^2}. \quad (7)$$

Зауважимо, що умова рівності коваріаційних матриць в задачах буріння не є жорсткою. Дисперсія параметрів і показників буріння протягом рейсу визначається в основному властивостями гірських порід – зміна твердості, тріщинуватості і ін. Статистична обробка матеріалів показує, що для багатьох гірських порід (монолітні, мономінеральні і ін.) характерною особливістю є однакова в межах точності вимірювань дисперсія. Тому правильним відбором породоруйнівного інструменту і параметрів режиму буріння можна зменшити дисперсію показників в породах які різко відрізняються властивостями, наблизивши її до дисперсії показників в породах з однаковими властивостями.

Тепер знайдемо дивергенцію між двома розподілами Бернуллі. Ці розподіли визначаються параметрами $p_1^{(1)}, \dots, p_n^{(1)}$ і $p_1^{(2)}, \dots, p_n^{(2)}$, де

$$1, \dots, p_i^{(1)} = p\{\xi_i^{(1)} = 1\}, p_i^{(2)} = p\{\xi_i^{(2)} = 1\}, i=1, \dots, n;$$

$$1 - p_i^{(1)} = p\{\xi_i^{(1)} = -1\}, 1 - p_i^{(2)} = p\{\xi_i^{(2)} = -1\}, i=1, \dots, n;$$

$$\xi^{(1)} = (\xi_1^{(1)}, \dots, \xi_n^{(1)}), \xi^{(2)} = (\xi_1^{(2)}, \dots, \xi_n^{(2)}).$$

Величини $\xi_i^{(1)}, \xi_j^{(1)}$ можуть приймати лише два значення: 1 або -1; $i=1, \dots, n; j=1, \dots, n$.

Нехай спочатку $n=1$. Тоді з визначення дивергенції (2) випливає, що

$$\begin{aligned} I(\xi^{(1)}, \xi^{(2)}) &= \sum_{a=1, -1} (p^{(1)}(a) - p^{(2)}(a)) \ln \frac{p^{(1)}(a)}{p^{(2)}(a)} = \\ &= (p^{(1)} - p^{(2)}) \ln \frac{p^{(1)}}{p^{(2)}} + (1 - p^{(1)} - 1 + p^{(2)}) \ln \frac{1 - p^{(1)}}{1 - p^{(2)}} = \\ &= (p^{(1)} - p^{(2)}) \ln \frac{p^{(1)}(1 - p^{(2)})}{p^{(2)}(1 - p^{(1)})}. \end{aligned} \quad (8)$$

Звідси, користуючись властивістю адитивності дивергенції і виразом (8), отримуємо формулу для дивергенції між двома розподілами Бернуллі для будь-якої кількості ознак n :

$$I(\xi^{(1)}, \xi^{(2)}) = \sum_{i=1}^n (p_i^{(1)} - p_i^{(2)}) \ln \frac{p_i^{(1)}(1 - p_i^{(2)})}{p_i^{(2)}(1 - p_i^{(1)})}. \quad (9)$$

Таким чином, при визначенні за допомогою дивергенції інформативності ознак, які приймають неперервні значення (швидкість буріння, витрати потужності і ін.), слід користу-

ватися виразом (7). У той же час в деяких задачах розпізнавання можна вважати, що будь-яка з ознак приймає всього два значення, тобто при їх незалежності – підпорядковується розподілу Бернуллі. У цьому випадку для визначення інформативності ознак слід користуватися виразом (9). Очевидно, що результати оцінки інформативності одних і тих же ознак будуть різними.

Оскільки виміряти всі ознаки не завжди можливо і не завжди доцільно з технічних або економічних причин, то необхідно зі всієї множини ознак вибрати найбільш інформативні. Іншими словами, перш ніж приступити до вирішення завдання розпізнавання ситуацій, необхідно провести попередню обробку і стиснення вихідних даних з метою зниження розмірності вектору вимірювань. Завдання зменшення числа ознак можна вирішувати двома способами. Перший полягає у виборі найбільш інформативних ознак і відсікання інших. Інший підхід полягає у створенні з великої кількості наявних ознак невеликої кількості комбінованих ознак. При цьому кожна нова ознака залежить від початкових, наприклад, є їх лінійною комбінацією або іншою функцією. При другому підході інформативність новостворених ознак є вищою оскільки комбінуючи вихідні ознаки можна обмежувати зниження інформативності, хоча цей підхід є більш трудомістким.

Відбираючи ознаки способом відсікання або комбінування, слід обмежити зменшення інформативності ознак, що залишилися тобто дивергенцію, якщо вона використовується в якості міри інформативності.

Перейдемо до використання поняття дивергенції для скорочення кількості ознак при вирішенні завдань з розпізнавання ситуацій. З врахуванням вищенаведеного, необхідно прагнути щоб зменшення дивергенції, тобто інформативності було мінімальним.

Розглянемо, як найважливіші, два випадки: нормальний розподіл і розподіл Бернуллі. Для кожного з них розберемо обидва способи скорочення кількості ознак: відсікання і лінійне перетворення ознак.

Якщо в разі нормальних розподілів з однаковою матрицею коваріацій $C = C^{(1)} = C^{(2)}$ скористатися методом відсікання ознак, то очевидно, що нова матриця коваріацій C^* буде являти собою підматрицю колишньої коваріаційної матриці C . Розподіл при цьому як і раніше залишиться нормальним, і тому вираз для дивергенції «усічених» розподілів залишиться аналогічною виразу (6):

$$I^*(\xi^{(1)}, \xi^{(2)}) = I(C^*, \delta^*) = (C^{*-1} \delta^*, \delta^*), \quad (10)$$

де δ^* вектор-стовбець різниці математичних очікувань

$$\delta^* = \begin{pmatrix} \delta_{i1} \\ \vdots \\ \delta_{im} \end{pmatrix} = \begin{pmatrix} m_{i1}^{(1)} - m_{i1}^{(2)} \\ \vdots \\ m_{im}^{(1)} - m_{im}^{(2)} \end{pmatrix};$$

$i = 1, \dots, m$ – номери змінних, які не брали участі у процесі відсікання.

Емпірично нескладно вибрати ті m невідсічених ознак, при яких дивергенція (10) буде максимальна.

Перейдемо тепер до вибору комбінованих ознак. Зробимо лінійне відображення простору ознак $(x_1, \dots, x_m)$ в простір ознак $(y_1, \dots, y_m)$:

$$y_i = \sum_{j=1}^n a_{ij} x_j, i = 1, \dots, m,$$

або $y = AX$, где X – це n -вимірний вектор; y – це m -вимірний вектор; A – матриця розмірності $m \times n$.

Обчислимо дивергенцію I^* для цих нових ознак:

$$I^* = I^*(C, \delta, A) = (C^{*-1} \delta^*, \delta^*), \quad (11)$$

де $\delta^* = A\delta$ – це m -вимірний вектор; $C^* = ACA'$ – це m -вимірна матриця.

При будь-якому m і будь-якій матриці A виконується наступна нерівність:

$$I^*(C, \delta, A) \leq I(C, \delta),$$

в якому існує така матриця A , для якої

$$I^*(C, \delta, A) = I(C, \delta).$$

Слід зауважити насамперед, що якщо $m=n$ і A – невироджена матриця (тобто простір всіх ознак залишається незмінним, змінюються тільки самі ознаки), то з урахуванням виразу (11):

$$\begin{aligned} I^*(C, \delta, A) &= (C^{*-1} \delta^*, \delta^*) = (A'^{-1} C^{-1} A^{-1} A\delta, A\delta) = \\ &= (C^{-1} \delta, \delta) = I(C, \delta). \end{aligned} \quad (12)$$

Це означає, що дивергенція залежить лише від простору ознак, але не залежить від вибору самих ознак. Звідси, очевидним є те, що і для m -вимірних ознак дивергенція також залежить тільки від багатовимірного простору $a_i = (a_{i1}, \dots, a_{in})$, $i = 1, \dots, m$, але не від самих цих векторів.

Доведено, що існує таке лінійне перетворення A простору ознак, яке дозволяє отримати вектор нових ознак меншої розмірності, причому

нові ознаки є лінійною комбінацією попередніх, а дивергенція між двома новими розподілами, тобто міра відмінності наявних двох класів об'єктів, дорівнює дивергенції, що базується на попередніх ознаках.

У разі розподілу Бернуллі функціональне відображення ознак є недоцільним, оскільки якщо при відображенні 2^n точок $a = (a_1, \dots, a_n)$, $a_i = \pm 1$ переходять в різні точки, то дивергенція не змінюється. У цьому випадку доцільним є здійснити відсічення ознак.

Розглянемо вираз для дивергенції (9) у для розподілу Бернуллі. У цьому випадку для будь-яких значень $P_i^{(1)}, P_i^{(2)}$, $i = 1, \dots, n$ кожен доданок суми є додатним. Звідси випливає, що якщо в разі розподілу Бернуллі зменшувати кількість ознак за допомогою відсікання, то відсікати слід ті ознаки, внесок яких в вираз (9) для дивергенції мінімальний. Тоді при фіксованій кількості нових ознак дивергенція зменшиться мінімально.

Нехай з n ознак з номерами $i = 1, \dots, n$ необхідно залишити лише 5 найбільш інформативних ознак. Оскільки мірою інформативності ми вважаємо дивергенцію, то залишити слід ті ознаки з номерами $m = 1, \dots, s$ для яких вираз

$$I = (p_{i_m}^{(1)} - p_{i_m}^{(2)}) \ln \frac{p_{i_m}^{(1)}(1 - p_{i_m}^{(2)})}{p_{i_m}^{(2)}(1 - p_{i_m}^{(1)})}, m = 1, \dots, s, \quad (13)$$

тобто вклад в дивергенцію приймає найбільші значення. При цьому нова випадкова величина $y = (x_{i_1}, \dots, x_{i_s})$ як і раніше буде мати розподіл Бернуллі з тими ж параметрами $p_{i_1}^{(1)}, \dots, p_{i_s}^{(1)}, p_{i_1}^{(2)}, \dots, p_{i_s}^{(2)}$. Нова дивергенція I' буде наступною:

$$I' = \sum_{m=1}^s (p_{i_m}^{(1)} - p_{i_m}^{(2)}) \ln \frac{p_{i_m}^{(1)}(1 - p_{i_m}^{(2)})}{p_{i_m}^{(2)}(1 - p_{i_m}^{(1)})}. \quad (14)$$

Оскільки всі складові в виразі (9) є додатними, то існує нерівність $I' \leq I$. Таким чином можна зробити висновок, що у випадку розподілу Бернуллі для скорочення кількості ознак краще користуватися методом відсікання, а в разі нормального розподілу – методом лінійного комбінування ознак, який дозволяє зберегти відмінність між класами (дивергенцію) на початковому рівні.

Розглянуті ймовірнісні методи оцінки інформативності не можуть дати однозначну для всіх умов оцінку тих чи інших параметрів і показників технологічного процесу. Залежно від співвідношення різноманітних факторів (геологічний розріз, властивості розбурюваних порід, що впливають на розподіл показників, конструкція свердловини, спосіб буріння та інші), можуть бути більш інформативними ті чи інші ознаки в різних завданнях розпізнавання, тому інформативність ознак необхідно оцінювати для кожної ситуації окремо.

Висновки. Розроблено методику оцінки інформативності ознак за допомогою дивергенції. Сформовано залежності, які дозволяють оцінити інформативність ознак, що підпорядковуються нормальному закону розподілу і закону розподілу Бернуллі.

З метою скорочення кількості ознак при вирішенні завдань розпізнавання ситуацій, запропоновано методику, що ґрунтується на методах відсікання і лінійного перетворення ознак.

Список літератури:

1. Чигур І.І. Особливості контролю технічного стану алмазних доліт в процесі буріння свердловин на нафту і газ. Вимірювальна та обчислювальна техніка в технологічних процесах. 2000. №4. С. 182–184.
2. Чигур І.І. Застосування нейро-фаззі моделювання для розпізнавання нештатних ситуацій, що виникають в процесі буріння свердловин на нафту і газ. Академічний вісник Криворізького територіального відділення Міжнародної Академії комп'ютерних наук і систем. Кривий ріг. 2004 № 13. С. 78–80.
3. Семенцов Г.Н., Чигур І.І. Основні концепції створення автоматизованої системи контролю за технічним станом породоруйнівного інструменту. Нафтогазова енергетика. 2007. № 1. С. 61–63.
4. Molinari A, Moufki A. The Merchant's model of orthogonal cutting revisited: A new insight into the modeling of chip formation. International Journal of Mechanical Sciences. 2008. № 50. P. 124–131.

Chyhur I.I., Lahoida A.I., Chyhur L.Ya. INFORMATIVE SIGNS IN THE TASKS OF AUTOMATED RECOGNITION OF TECHNOLOGICAL SITUATIONS DURING WELL DRILLING

The article is devoted to the urgent problem of automated recognition and classification of technological situations that arise during the drilling process, as an important stage for making informed decisions in the design and management of this complex technological process.

It is noted that during the drilling process there is a need to identify various factors, such as changes in the drillability and fracturing of rocks, the degree of wear of the drilling tool, as well as geological complications

and accidents. The information obtained as a result of recognition is critically important for effective control of the drilling process, determining the optimal moment for lifting the drilling tool and, what is especially important, for preventing emergency situations.

It is proposed to reduce the task of pattern recognition to the selection of elements belonging to a specific class among a set of elements that may belong to different classes using divergence. It is emphasized that the presence of certain common properties or features in elements of the same class is a key criterion for their identification and assignment to this class.

The important problem of pre-processing and compression of output data in the tasks of recognizing technological situations in order to reduce the dimensionality of the measurement vector is considered.

Two main approaches to reducing the number of features are proposed: selection of the most informative features by cutting off less significant ones and creating a small number of new, combined features based on existing ones. It is emphasized that the second approach, although more laborious, potentially provides higher informativeness of newly created features due to combining the original ones.

The need to minimize the loss of informativeness when selecting or combining features is emphasized, proposing to use the concept of divergence as a measure of informativeness. Further application of divergence to reduce the number of features is proposed using the example of two important cases of data distribution: normal distribution and Bernoulli distribution, with an analysis of both methods of dimensionality reduction (cut-off and linear transformation of features) for each of them.

Thus, the article is an introduction to the important scientific and technical problem of recognition and classification of technological situations in drilling, outlines its relevance, analyzes existing approaches and offers potential information bases for further research in this area.

Key words: *automated recognition, drilling, informative features, divergence, Bernoulli distribution.*