

Горобченко Олександр¹

¹ Професор кафедри електромеханіки і рухомого складу залізниць Державного університету інфраструктури та технологій, м. Київ, Україна, ORCID: <http://orcid.org/0000-0002-9868-3852>

*Автор відповідальний за листування: gorobchenko.a.n@gmail.com

ТЕОРЕТИЧНІ ОСНОВИ ОЦІНКИ ІНФОРМАТИВНОСТІ ОЗНАК ПОЇЗНОЇ СИТУАЦІЇ ДЛЯ РОБОТИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ КЕРУВАННЯ ЛОКОМОТИВОМ

Стаття присвячена проблемі впровадження інтелектуальних систем керування на транспорті. Важливим завданням є оцінка інформаційних параметрів роботи систем керування. В існуючих роботах не розглянуто питання визначення одного з основних параметрів функціонування локомотивних систем керування - інформаційна цінність окремих ознак поїзної ситуації. Це не дає можливості визначити порядок обробки сигналів на вході та оцінити їх внесок у прийняття того або іншого керуючого рішення. Причому інформативність є величиною відносною, що виражається в різній інформаційній цінності окремої ознаки для класифікації різних поїзних ситуацій. Також інформативність ознаки може залежати від типу рішючих правил в процедурі класифікації. Якість розпізнавання тієї або іншої поїзної ситуації, в якій знаходиться локомотивна бригада, залежить від якості ознак, що використовуються системою класифікації. Вирішальним критерієм інформативності ознак в задачі розпізнавання образів пропонується величина втрат від помилок. Для визначення кола найбільш інформативних ознак поїзних ситуацій використано метод випадкового пошуку з адаптацією. Результати роботи дають можливість оптимізувати роботу автоматизованих та інтелектуальних систем керування поїздом за рахунок зменшення обсягів обчислень та спрощення їх алгоритму.

Ключові слова: локомотив, інформаційна ознака, інтелектуальна система, керування, поїзна ситуація, керування поїздом.

Вступ. Використання інтелектуальних систем керування в транспортній галузі є одним з перспективних завдань, що дозволить значно скоротити ризики негативного впливу людського фактору на безпеку та якість перевезень. Інтелектуалізація процесу керування потребує проведення додаткових досліджень за такими напрямками: розробка оптимізованих структур баз знань; удосконалення систем збирання поточної інформації та дистанційної передачі сигналів та даних [1]; визначення параметрів та обсягів інформаційних потоків, що обробляються системою керування; вдосконалення систем розпізнавання поточних транспортних ситуацій з визначенням найбільш небезпечних умов експлуатації та розробки алгоритмів виходу з них, і таке інше. Основним об'єктом, з яким працює інтелектуальна система керування, є інформаційний потік. Від його обсягу та якісних параметрів залежить обґрунтованість та своєчасність прийняття керуючих рішень, що безпосередньо впливає на безпеку руху та економічні показники перевізного процесу. Розмаїтість експлуатаційних умов, під якими звичайно розуміють вагу і довжину поїзда, його опір руху, порядок пропуску поїзда по перегонах, погодні умови, установлені попередженнями обмеження швидкості руху та ін., ставить перед машиністом у кожній поїзді завдання вибору і реалізації раціонального режиму ведення поїзда, що відповідає саме даним умовам [2]. Таким чином ці експлуатаційні умови необхідно постійно відстежувати, класифікувати та надавати їм належну оцінку перед використанням в системах керування.

Аналіз останніх досліджень. Підходи до обробки технічної інформації, отриманої в автоматичному режимі в теперішній час достатньо розроблені. В [3] розглянуто проблеми підвищення точності параметрів, що вимірюються, у великомасштабних інформаційних системах управління. Показано, що використання додаткових каналів виміру та додаткових непрямих показників дозволяє підвищити точність вимірюваних значень та оцінок параметрів математичних моделей досліджуваної системи. Використання додаткових непрямих індикаторів дозволяє зменшити розмір системи та кількість вимірювального обладнання, не знижуючи точності вимірюваних параметрів та техніко-економічних показників. В роботі [4] застосовано теорія нечітких множин та нова методологія обробки інформації при визначенні надійності та безпеки складних систем керування. В роботі [5] Представлено структуру інформаційно-логічної моделі та наведено характеристику засобів і методів збору та обробки даних при оцінці функціональної надійності оператора авіаційних ергатичних систем управління.

Але у вказаних роботах не розглянуто питання визначення одного з основних параметрів інформації – інформаційна цінність окремих ознак. В цих роботах всі канали інформації, що надходять до систем керування, вважаються рівнозначними та підходи до обробки інформації з кожного з них є однаковими. Це дещо знижує якість подальшого використання вхідних даних та збільшує час на обробку вхідної інформації за рахунок послідовного розгляду всіх сигналів, в незалежності від поточного стану транспортного засобу та умов його експлуатації.

Роботи з визначення цінності інформації [6-9] дозволяють формалізувати параметр обсягу та важливості інформації. Це є підставою для визначення даних, що потребують першочергової обробки. Але при наявності сотен каналів, по яких інформація надається до системи керування поїздом, необхідно розробити удосконалену методологію, за якою буду виконано ранжування отриманих сигналів та їх внесок у прийняте рішення з керування локомотивом [10].

Мета дослідження. Якість розпізнавання тієї або іншої поїзної ситуації, в якій знаходиться локомотивна бригада, залежить від якості ознак, що використовуються системою класифікації. Основною характеристикою ознаки є її інформативність. Причому інформативність є величиною відносною, що виражається в різній інформаційній цінності окремої ознаки для класифікації різних поїзних ситуацій. Також інформативність ознаки може залежати від типу рішальних правил в процедурі класифікації.

Сучасні локомотиви обладнані потужними системами діагностики, контролю поточних параметрів руху та вхідних сигналів. Тому в розпорядженні систем розпізнавання образів поїзних ситуацій знаходиться множина ознак, що може налічувати декілька сот членів. Використання всієї сукупності ознак може бути недоцільне з наступних причин. Кожна ознака потребує окремого інформаційного каналу для її контролю, що ускладнює інтерфейсну частину системи, що розробляється, та призводить до її здороження. Другий фактор – це використання обчислювальних ресурсів бортової ЕОМ, які на даний момент є хоч і достатньо потужними, але на локомотивів бажано мати максимальну швидкість роботи системи, що неможливо при обробці і аналізі такого широкого спектру даних. Таким чином завдання оцінки та вибору інформативних ознак при розпізнаванні поїзних ситуацій є актуальним та дозволить підвищити ефективність локомотивної системи розпізнавання образів.

Параметри, що визначають умови руху поїзда, змінюються в результаті дії багатьох факторів. Деякі з них взаємозалежні, окремі є незалежними, але їхній вплив може викликати зміну в широких межах параметрів, що визначають характер руху поїзда. Все це значно ускладнює регулювання потужності локомотива й вибір раціонального режиму ведення поїзда, вимагає від машиністів специфічних професійних навичок, обумовлених теоретичною підготовкою і певним практичним досвідом.

Вибір інформативного набору ознак для розпізнавання поїзних ситуацій. Попередній склад ознак задається неформалізованим шляхом, на основі досвіду та особистих переваг спеціаліста. Формальні методи застосовуються до навчаючої вибірки A для перевірки цієї вихідної системи на достатність і необхідність. Серед усіх можливих систем ознак достатньою вважаємо систему, яка при заданих S і D забезпечує витрати N , що не перевищують певного

порогу N_0 . Під витратами N тут розуміється вартість вимірювання ознак (N_x) і вартість втрат, викликаних помилками розпізнавання (N_r):

$$N = N_x + N_r. \quad (1)$$

Необхідною є достатня система мінімальної складності (вартості). Так що фактично на навчальній вибірці A вирішується переборне завдання типу

$$\beta = \arg \min_{\beta \in B} N(X_\beta) / C, D, A, N_0, \quad (2)$$

де $C = \langle c_1, c_2, \dots, c_k \rangle$ – перелік поїзних ситуацій, що розглядається

D – тип рішальної функції;

B – множина всіх можливих систем ознак;

X – інформативна множина описуючих ознак;

N – витрати;

N_0 – порогове значення витрат.

Ця задача одночасної мінімізації N_x і N_r вперше була сформульована в [11]. Витрати на вимірювання залежать від того, скільки і яких ознак потрібно вимірювати і яке число розрядів потрібно для представлення результатів вимірювань. Зі зрозумілих причин основна увага приділяється зменшенню кількості вимірюваних ознак, тобто пошуку інформативної підсистеми з n ознак (X_n) серед g ознак вихідної системи (X_g).

Визначення інформативності окремих ознак. Вирішальним критерієм інформативності ознак в задачі розпізнавання образів пропонується величина втрат від помилок R . Навіть якщо розподіли генеральної сукупності відомі, обчислення втрат пов'язано з дуже великими витратами машинного часу. У зв'язку з цим робляться спроби знайти критерії, більш просто обчислювані і разом з тим жорстко корелюючі з оцінкою втрат R .

Якщо розподіл реалізацій кожного образу підпорядковується нормальному закону з діагональними матрицями коваріацій (при цьому поверхні рівної щільності являють собою сфери однакового радіуса), то мірою труднощі розпізнавання D , назад пропорційної очікуваних втрат, може служити середнє значення евклідової відстані між математичними очікуваннями всіх пар образів [12]:

$$D = (1/C_k^2) \sum_{i,j=1}^k \rho(ij), \quad (3)$$

де $\rho(ij)$ – евклідова відстань між математичними очікуваннями i -го та j -го образів.

В термінах теорії інформації мірою труднощі розпізнавання служить ентропія H розподілів щільності ймовірності образів. Нехай розподіли образів k спроектовані на одну вісь x , що вимірюється з точністю t до градацій (рис. 1). Ймовірність попадання реалізацій i -го образу j -ю градацією дорівнює $P(j/i)$. Підсумувавши для j -ї градації ймовірності всіх образів, ми отримуємо величину $H_x = \sum_{j=1}^t H_j P_j$. Внесок i -го образу в цю суму, так що ентропія j -ї градації виражається таким значенням

$$H_j = -(r_1 \log r_1 + r_2 \log r_2 + \dots + r_i \log r_i + \dots + r_k \log r_k). \quad (4)$$

З принципу адитивності ентропії випливає, що загальна невизначеність при розпізнаванні образів за ознакою x має вигляд $H_x = \sum_{j=1}^t H_j P_j$. Якщо початкова невизначеність ситуації дорівнювала $\log k$, то кількість інформації I_x , одержуваної в результаті вимірювання ознаки x , дорівнює $H_0 - H_x$.

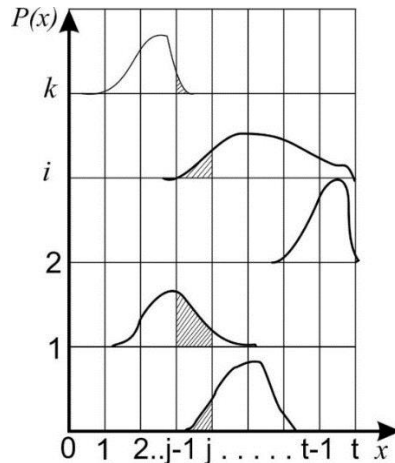


Рис. 1. Розподіл образів

При визначенні поїзної ситуації немає можливості отримати закони розподілів образів. Це пов'язано з невеликим обсягом статистичних даних, які можна достовірно отримати під час руху. У цих умовах доцільно використовувати методи, які не вимагають побудови моделей розподілу і спираються на конкретні об'єкти, наявні в навчальній вибірці. За цими прецедентами будується вирішальна функція (наприклад, правило k найближчих сусідів), розпізнається контрольна послідовність, і за кількістю отриманих помилок виноситься оцінка інформативності окремої ознаки або їх системи.

Можливі й інші способи оцінки інформативності. Гіпотеза компактності дає нам основу для оцінки інформативності простору ознак через прояв характеристик компактності. З неї випливає, що для хорошого розпізнавання образів бажано, щоб відстані між своїми точками кожного образу були малими, а відстані до точок інших образів по можливості великими. А якщо опуклі оболонки різних образів накладаються одна на одну, то бажано, щоб вони якомога більше відрізнялися за своїми розмірами. Компактність (щільність) W_i образу i , представленого в навчальній вибірці m_i точками $1, 2, \dots, t, \dots, 1, \dots, m_i$, можна характеризувати середньою довжиною ребер поєднуючого їх повного графа:

$$W_i = (1/C_{m_i}^2) \sum_{t,l=1}^{m_i} r(t,l). \quad (5)$$

Також компактність W_j точок $1, 2, \dots, s, \dots, v, \dots, m_j$, що представляють образ j , має вигляд

$$W_j = (1/C_{m_j}^2) \sum_{s,v=1}^{m_j} r(s,v). \quad (6)$$

Рознесеність образів у просторі характеристик можна оцінювати через середню відстань між усіма парами точок з різних образів:

$$W(i, j) = (1/m_i m_j) \sum r(t, s) \text{ для } t = 1 \div m_i, s = 1 \div m_j \quad (7)$$

На підставі сказаного інформативність простору ознак тим більше, чим більше величина

$$J = W(i, j) / (W_i + W_j). \quad (8)$$

Оцінку інформативності ознак поїзної ситуації можна отримати і безпосередньо в процесі побудови вирішального правила у вигляді дерева дихотомічних поділів вибірки за окремими ознаками [13]. Припустимо, що є можливість розділити ознаку X тільки на дві градації: $x \leq l$ і $x > l$. Подивимося склад реалізацій, що потрапили в ці градації. Якщо в першій градації виявиться m_{il} реалізацій i -го образу і m_{vl} реалізацій v -го образу,

то неоднорідність складу цієї градації можна оцінити величиною

$$R_1 = \sum_{\substack{i=1 \\ v=i+1}}^k m_{ii} m_{vl}, \quad (9)$$

то зменшення невизначеності після отримання інформації від ознаки x , тобто інформативність ознаки x , можна оцінити величиною $J_x = (R_0 - R') / R_0$. Якщо $R' = 0$, то інформативність ознаки J_x буде максимальною і дорівнювати одиниці. Якщо R' не зменшило початкової невизначеності, то $J_x = 0$ й ознака x природно вважати неінформативним.

Якщо відомо, що ознаки не залежать одна від одної, то можна з допомогою одного з описаних методів оцінити інформативність всіх g ознак вихідної системи і потім вибрати з них n найбільш інформативних. Але в реальних таблицях даних залежність між ознаками спостерігається дуже часто. А якщо ознаки залежні, то при виборі найбільш інформативною підсистеми оцінками їх індивідуальної інформативності керуватися не можна.

Складність використання наведеного підходу полягає в отриманні дуже великої кількості комбінацій для умов роботи локомотивної бригади, що приблизно складає більше 10^{22} . Виконання перебору цих комбінацій займе роки при поточній потужності обчислювальних засобів. Тому необхідно розробити алгоритм оцінювання інформативності ознак, прийнятний для реалізації на рухомому складі.

Серед алгоритмів, що здобули найбільшу розповсюдженість [14,15], можна виділити метод послідовного скорочення (алгоритм Del), метод послідовного додавання ознак (алгоритм Add), метод випадкового пошуку з адаптацією (алгоритм ВПА).

Алгоритм Del полягає в такому:

1. Виконується оцінка похибки розпізнавання при використанні всіх W ознак.
2. Виключається з системи перша ознака і знаходиться помилка, яку дають решта $(W-1)$ ознак.
3. Перша ознака повертається в вихідні дані, а друга виключається і знаходиться помилка в новому $(W-1)$ -мірному просторі.
4. Цю операцію почергового виключення однієї ознаки проводять W разів. Серед отриманих величин визначають саму малу. Вона вкаже на ознаку, виключення якої з системи було найменш відчутним.
5. Виключають цю ознаку з системи і приступають до випробування $W-1$ ознак, що залишились. Їх почергове виключення з системи дозволить знайти найбільш інформативну і знизити розмірність простору до $W-2$. Ці процедури повторюються стільки раз, поки в системі не залишиться задане число ознак.

Алгоритм Add відрізняється від попереднього лише тим, що порядок перевірки підсистем ознак починається не з W -мірного простору, а з одновимірних просторів. Спочатку всі W ознак перевіряються на інформативність. Для цього робиться розпізнавання контрольної послідовності по кожній з ознак окремо і в інформаційну підсистему включається ознака, що дала найменше число помилок. Потім до неї по черзі додаються всі ознаки по одній. Отримані двовимірні підпростори оцінюються за кількістю помилок розпізнавання. Вибирається найбільш інформативна пара ознак. До неї таким же шляхом підбирається найкраща третя ознака з тих, що залишились, і так продовжується до отримання системи з потрібною кількістю ознак.

Трудомісткість цього алгоритму приблизно така ж, як і алгоритму Del, однак результати, одержувані алгоритмом Add, зазвичай краще, ніж у Del. Пояснюється цей факт впливом малої показності навчальної вибірки: при одному і тому ж обсязі вибірки чим вище розмірність простору ознак, тим менше обґрунтованість одержуваних статистичних висновків (у нашому випадку – оцінки інформативності).

Обидва описаних алгоритму дають оптимальне рішення на кожному кроці, але це не забезпечує глобального оптимуму. Для ослаблення впливу помилок на перших кроках алгоритму застосовується релаксаційний метод. В алгоритмі Add набирається певна кількість інформативних ознак і потім частина з них виключається методом Del. Можлива і зворотна стратегія: спочатку працює алгоритм Del, після скорочення вихідної системи на декілька ознак включається алгоритм

Add, який повертає в систему помилково виключені з неї ознаки. Повторення цих процедур (алгоритм DelAdd) продовжується до отримання системи з найбільш інформативних ознак. В [16] вказано, що аналіз наведених методів дозволяє зробити висновок про перевагу алгоритму ВПА над алгоритмами Add, Del та DelAdd.

Використання методу випадкового пошуку з адаптацією для визначення кола найбільш інформативних ознак поїзних ситуацій. Перш за все визначається перелік поїзних ситуацій, що підлягають розпізнаванню при керуванні поїздом. Для різних видів руху (вантажний рух, швидкісний рух, станційна робота локомотива, тощо) він може відрізнятися. Далі необхідно отримати множину W ознак поїзної ситуації. Вона будується на основі даних, що доступні бортовій системі діагностики локомотива та машиністу під час поїздки. В результаті аналізу інформаційних потоків, що виникають при керуванні поїздом, отримана множина ознак, що складається з таких членів:

- w_1 – швидкість руху;
- w_2 – відстань до сигналу;
- w_3 – значення сигналу попереду;
- w_4 – маса поїзду;
- w_5 – положення контролеру машиніста;
- w_6 – положення гальмівного крану машиніста;
- w_7 – тиск в гальмівній магістралі поїзду;
- w_8 – поточний профіль колії;
- w_9 – наступний профіль колії;
- w_{10} – сигнал буксування;
- w_{11} – вільність колії попереду;
- w_{12} – прослідкування станції (або роз'їзду);
- w_{13} – струм навантаження тягових електродвигунів;
- w_{14} – поточне значення обмеження швидкості;
- w_{15} – різниця між поточним та графіковим часом слідування;
- w_{16} – загальна оцінка стану екіпажної частини локомотива;
- w_{17} – загальна оцінка стану системи передачі потужності локомотива;
- w_{18} – психофізіологічні характеристики локомотивної бригади;
- w_{19} – час доби;
- w_{20} – час від початку роботи локомотивної бригади;
- w_{21} – метеорологічні умови;
- w_{22} – загальна оцінка стану гальм поїзду (за даними пробних гальмувань);
- w_{23} – наявність в поїзді особливих та розрядних вантажів, а також людських вагонів;
- w_{24} – відхилення витрати енергоресурсів на тягу від норми;
- w_{25} – оперативний план роботи локомотивної бригади.

Кожна ознака нормалізується за відомими алгоритмами в залежності від того, якого типу величина – якісна або кількісна. В результаті на вхід системи розпізнавання подаються нормалізовані сигнали, що знаходяться в інтервалі від 0 до 1.

Розіб'ємо умовний відрізок (0-1) на W однакових ділянок і зіставимо кожну ділянку зі своєю ознакою: 1-ша ділянка відповідає першій ознаці, 2-га – другій і т. д. Кожна ділянка має ширину $1/W$ (рис. 2).

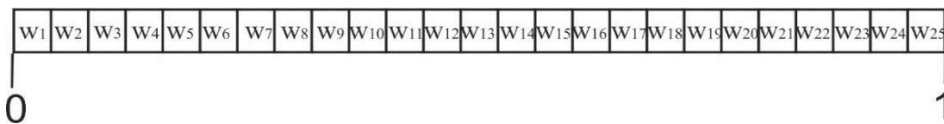


Рис. 2. Початковий розподіл ділянок, що відповідають ознакам поїзних ситуацій

Другим етапом у випадковому порядку обирається низка значень в одиничному інтервалі. Ті ділянки, в які потрапили ці випадкові значення, використовуються для визначення образу поїзної ситуації. Якість цієї випадково обраної підсистеми оцінюється за одним з критеріїв,

наприклад, за кількістю одержуваних помилок розпізнавання α_i .

Описана процедура випадкового вибору підсистем ознак і оцінки їх якості повторюється k раз. Розгляд отриманого списку оцінок $\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_k$, дозволяє вибрати найкращу і найгіршу з підсистем. На цій підставі робиться процедура «заохочення» і «покарання»: ділянки, відповідні ознакам, які потрапили в найкращу підсистему, заохочуються шляхом розширення їх меж на величину h , а ділянки, відповідні ознаками з самої неінформативної підсистеми, караються тим, що їх ширина зменшується на величину h , при чому $h < 1/W$. Сумарна довжина всіх ділянок так і залишається рівною одиниці. Отримаємо деякий новий розподіл ділянок ознак поїзних ситуацій (рис. 3).

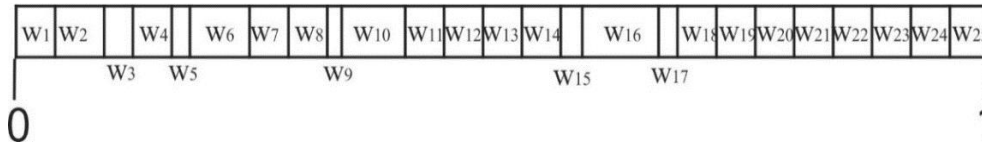


Рис. 3. Розподіл ділянок, що відповідають ознакам поїзних ситуацій, після процедури «заохочення-покарання»

З рис. 3 видно, що ознаки $w_3, w_5, w_9, w_{15}, w_{17} \in$ «покараними», тобто вони найгірше вплинули на якість розпізнавання образу ситуації.

Після цього випадковим чином вибираються і випробовуються нові підсистеми. Але тепер ймовірність попадання ознак у ці підсистеми не однакова: заохочені ознаки, представлені більш широкими смужками, мають більше шансів увійти в чергову підсистему, ніж покарані. За результатами випробування цієї партії підсистем процедура адаптації (покарання та заохочення) повторюється. Якщо певна ознака випадково потрапляє і в найкращу і найгіршу підсистеми, то довжина її ділянки залишається незмінною. Якщо ж вона регулярно опиняється у складі самої інформативної підсистеми, то довжина його ділянки зростає з кожним кроком адаптації. Також ознака, що систематично потрапляє в саму неінформативну підсистему, з кожним кроком скорочує довжину своєї ділянки і тим самим зменшує ймовірність включення в випробуванні підмножини ознак.

Після певної кількості циклів пошуку та адаптації процес стабілізується: ділянки ознак, що мають найбільшу інформативність, займають практично весь відрізок (0-1) і випробувану підсистему вибираються одні й ті ж ознаки. Цей факт служить сигналом до закінчення процесу вибору підсистеми найбільш інформативних ознак.

Для реалізації описаного методу визначення інформативності окремих ознак ведення поїзду було розроблено комп'ютерну програму. До вхідних даних цієї програми було занесено всі двадцять п'ять ознак керування поїздом, що наведені вище. Алгоритм роботи програми передбачає декілька десятків ітерацій. Критерієм зупинки роботи програми є досягнення однією з ознак мінімальної ймовірності попадання в підсистему оцінки поточної поїзної ситуації, тобто досягнення максимальної величини «покарання».

Висновки. Якість розпізнавання тієї або іншої поїзної ситуації, в якій знаходиться локомотивна бригада, залежить від якості ознак, що використовуються системою класифікації. Вирішальним критерієм інформативності ознак в задачі розпізнавання образів пропонується величина втрат від помилок. Для визначення кола найбільш інформативних ознак поїзних ситуацій використано метод випадкового пошуку з адаптацією. В результаті аналізу інформаційних потоків, що виникають при керуванні поїздом, отримана множина ознак, що складається з 25 членів (w_1 – швидкість руху; w_2 – відстань до сигналу; w_3 – значення сигналу попереду; w_4 – маса поїзду; w_5 – положення контролеру машиніста т. д.). Після певної кількості циклів пошуку та адаптації процес стабілізується: ділянки ознак, що мають найбільшу інформативність, займають практично весь відрізок (0-1) і випробувану підсистему вибираються одні й ті ж ознаки. Цей факт служить сигналом до закінчення процесу вибору підсистеми найбільш інформативних ознак.

Результатом використання методу випадкового пошуку з адаптацією є отримання найбільш інформативних ознак поїзної ситуації, та відсікання найменш інформативних за розробленим критерієм. Це дає можливість оптимізувати роботу автоматизованих та інтелектуальних систем керування поїздом за рахунок зменшення обсягів обчислень та спрощення їх алгоритму.

ЛІТЕРАТУРА

1. T. Wen, G. Xie, Y. Cao and B. Cai, "A DNN-Based Channel Model for Network Planning in Train Control Systems," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2021.3093025.
2. Горобченко О., Неведров, О., Незліна, О., Ткаченко, В. Розробка методу кластеризації поїзних ситуацій. Транспортні системи і технології, 2021, (37), 187-195. <https://doi.org/10.32703/2617-9040-2021-37-18>
3. I. S. Durgaryan, A. F. Pashchenko, Y. S. Rodomanova, H. H. Do and T. A. Pham, "Improving the Accuracy of Measuring and Evaluation of Parameters of Large-scale Information Control Systems," 2018 Eleventh International Conference "Management of large-scale system development" (MLSD, 2018, pp. 1-4, doi: 10.1109/MLSD.2018.8551781.
4. L. S. Zvyagin, "Process of information processing when realizing the concept of "soft" measurements," 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM), 2017, pp. 70-73, doi: 10.1109/SCM.2017.7970498.
5. A. V. Bogomolov, G. A. Sviridyuk, A. V. Keller, V. N. Zinkin and M. D. Alekhin, "Information-logical Modeling of Information Collection and Processing at the Evaluation of the Functional Reliability of the Aviation Ergate Control System Operator," 2018 Third International Conference on Human Factors in Complex Technical Systems and Environments (ERGO)s and Environments (ERGO), 2018, pp. 106-110, doi: 10.1109/ERGO.2018.8443849.
6. Rothery, C., Strong, M., Koffijberg, H. E., Basu, A., Ghabri, S., Knies, S., ... & Fenwick, E. Value of information analytical methods: report 2 of the ISPOR value of information analysis emerging good practices task force. Value in health, 2020, 23(3), 277-286.
7. Ayan, O., Vilgelm, M., Klügel, M., Hirche, S., & Kellerer, W. (2019, April). Age-of-information vs. value-of-information scheduling for cellular networked control systems. In Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems, 2019, pp. 109-117.
8. Viet, N. Q., Behdani, B., & Bloemhof, J. The value of information in supply chain decisions: A review of the literature and research agenda. Computers & Industrial Engineering, 2018 120, 68-82.,
9. Kuric, I., Gorobchenko, O., Litikova, O., Gritsuk, I., Mateichyk, V., Bulgakov, M., & Klackova, I. Research of vehicle control informative functioning capacity. Paper presented at the IOP Conference Series: Materials Science and Engineering, 2020 , 776(1) doi:10.1088/1757-899X/776/1/012036.
10. Бабанін, О. Б., Горобченко, О. М. Визначення цільової функції для оптимізації процесу керування в ергатичній системі Машиніст–СППР–Поїзд на підставі критерію корисності. Збірник наукових праць Державного економіко-технологічного університету транспорту. Серія: Транспортні системи і технології, 2014, (25), с. 92-98.
11. King-Sun Fu The Optimal sequential decisions. Lafayette: Purdue Univ. Press, 1967.
12. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний. Издательство: Институт математики, 1999, 270 с. ISBN:5-86134-060-9.
13. Kaufman, L., & Rousseeuw, P. J. Finding groups in data: an introduction to cluster analysis, 2009, (Vol. 344). John Wiley & Sons.
14. Alpaydin, E. Introduction to machine learning. MIT press., 2020.
15. Minelli, M., Chambers, M., & Dhiraj, A. (2013). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses., John Wiley & Sons., 2013, Vol. 578
16. Загоруйко Н.Г. Когнитивный анализ данных. Новосибирск: Академическое издательство ГЕО, 2013, 186 с

REFERENCES

1. T. Wen, G. Xie, Y. Cao and B. Cai, "A DNN-Based Channel Model for Network Planning in Train Control Systems," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2021.3093025.
2. Gorobchenko, O., Nevedrov, O., Nezlina, O., & Tkachenko, V. (2021). Development of a method for clustering train situations. Transport systems and technologies, (37), 187-195. <https://doi.org/10.32703/2617-9040-2021-37-18>
3. I. S. Durgaryan, A. F. Pashchenko, Y. S. Rodomanova, H. H. Do and T. A. Pham, (2018) "Improving the Accuracy of Measuring and Evaluation of Parameters of Large-scale Information Control Systems," 2018 Eleventh International Conference "Management of large-scale system development" (MLSD, 2018, pp. 1-4, doi: 10.1109/MLSD.2018.8551781.
4. L. S. Zvyagin, (2017) "Process of information processing when realizing the concept of "soft" measurements," 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM), 70-73, doi: 10.1109/SCM.2017.7970498.
5. A. V. Bogomolov, G. A. Sviridyuk, A. V. Keller, V. N. Zinkin and M. D. Alekhin. (2018) "Information-logical Modeling of Information Collection and Processing at the Evaluation of the Functional Reliability of the Aviation Ergate Control System Operator," 2018 Third International Conference on Human Factors in Complex Technical Systems and Environments (ERGO)s and Environments (ERGO), 106-110, doi: 10.1109/ERGO.2018.8443849.

6. Rothery, C., Strong, M., Koffijberg, H. E., Basu, A., Ghabri, S., Knies, S., ... & Fenwick, E. (2020). Value of information analytical methods: report 2 of the ISPOR value of information analysis emerging good practices task force. *Value in health*, 23(3), 277-286.
7. Ayan, O., Vilgelm, M., Klügel, M., Hirche, S., & Kellerer, W. (2019, April). Age-of-information vs. value-of-information scheduling for cellular networked control systems. In *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems* (pp. 109-117).
8. Viet, N. Q., Behdani, B., & Bloemhof, J. (2018). The value of information in supply chain decisions: A review of the literature and research agenda. *Computers & Industrial Engineering*, 120, 68-82.,
9. Kuric, I., Gorobchenko, O., Litikova, O., Gritsuk, I., Mateichyk, V., Bulgakov, M., & Klackova, I. (2020). Research of vehicle control informative functioning capacity. Paper presented at the IOP Conference Series: Materials Science and Engineering, , 776(1) doi:10.1088/1757-899X/776/1/012036.
10. Babanin, O., Gorobchenko, O. (2014). Determination of the objective function for optimization of the control process in the ergatic system Machinist – DSS – Train on the basis of the utility criterion. *Collection of scientific works of the State Economic and Technological University of Transport. Series: Transport systems and technologies*, (25), p. 92-98.
11. King-Sun Fu (1967) *The Optimal secventional decisions*. Lafayette: Purdue Univ. Press, 1967.
12. Zagoruiko, N.G. (1999). *Applied methods of data and knowledge analysis*. Publisher: Institute of Mathematics. 270 p. ISBN:5-86134-060-9.
13. Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.
14. Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
15. Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses* (Vol. 578). John Wiley & Sons.
16. Zagoruiko N.G (2013). *Cognitive data analysis*. Novosibirsk: Academic publishing house GEO ., 186 p.

Oleksandr Gorobchenko¹

¹ Professor, Department of Electromechanics and rolling stock of railways, State University of Infrastructure and Technologies, 9, Kyrylivska str., Kyiv, 04071, Ukraine, ORCID: <http://orcid.org/0000-0002-9868-3852>

THEORETICAL FUNDAMENTALS OF ESTIMATABILITY ASSESSMENT OF TRAIN SITUATION SIGNS FOR WORK OF INTELLECTUAL LOCOMOTIVE CONTROL SYSTEMS

The article is devoted to the problem of implementation of intelligent control systems in transport. An important task is to assess the information parameters of the control systems. In the existing works the question of definition of one of the basic parameters of functioning of locomotive control systems - information value of separate signs of a train situation is not considered. This does not make it possible to determine the order of signal processing at the input and assess their contribution to the adoption of a control decision. Moreover, informativeness is a relative value, which is expressed in the different information value of a particular feature for the classification of different train situations. Also, the informativeness of the feature may depend on the type of decisive rules in the classification procedure. The quality of recognition of a train situation in which the locomotive crew is, depends on the quality of the features used by the classification system. The decisive criterion for the informativeness of the features in the problem of pattern recognition is the magnitude of losses from errors. To determine the range of the most informative features of train situations, the method of random search with adaptation was used. The results of the work make it possible to optimize the operation of automated and intelligent train control systems by reducing the amount of calculations and simplifying their algorithm.

Keywords: locomotive, information sign, intelligent system, management, train situation, train management.