

МОДИФІКОВАНА МОДЕЛЬ КОМП'ЮТЕРНИХ ПРИБОРІВ ТА СИСТЕМ ЯК ОБ'ЄКТІВ ПРОГНОЗУВАННЯ

В роботі описано модифіковану модель комп'ютерних пристроїв та систем як об'єктів прогнозування. Запропонована модель дозволяє використати апріорну діагностичну інформацію про об'єкт прогнозування, зокрема, інформацію про динаміку зміни попередніх технічних станів. Застосування модифікованої моделі дало можливість зменшити похибку прогнозування нейромережного методу до 14 відсотків.

It is offered and described modifying model of computer devices and systems as an object of forecasting. New model has allowed to use the aprioristic diagnostic information about process and object of forecasting, for example, information about preceding technical state of computer system. New model allowed decrease mistakes of neural network approach to forecast working capacity of computer systems till 14 percent.

Ключові слова: прогнозування, нейронні мережі.

Вступ

Використання штучних нейронних мереж та інших компонентів теорії штучного інтелекту для розв'язання задач прогнозування у багатьох галузях є достатньо дослідженими. Зокрема, відомі рішення щодо прогнозування рівня споживання електроенергії [1, 2], прогнозування в економіці [3, 4], прогнозування природних явищ [5] тощо. Разом з тим дослідження, з використанням компонентів теорії штучного інтелекту для прогнозування технічного стану, працездатності або надійності КП та С практично відсутні. Частково це можна пояснити неможливістю адаптувати для розв'язання даної задачі підходи, властиві, наприклад, прогнозуванню часових рядів. В даному випадку, об'єктом прогнозування є комп'ютерний пристрій чи система. Незважаючи на те, що вихідним для прогнозу залишається набір певних параметрів, які характеризують технічний стан об'єкта (в даний момент часу або протягом визначеного інтервалу), не можна залишати без уваги особливості комп'ютерних пристроїв та систем як об'єктів прогнозування. Саме внаслідок складних взаємозв'язків між параметрами об'єкта прогнозування (ОП) використання стандартних підходів прогнозування часових рядів є неефективним щодо КП та С. Застосування ж методів прогнозування стану складних технічних об'єктів [6, 7] щодо КС в цілому так і до їх складових компонентів зокрема ускладнюється рядом об'єктивних причин, серед яких найвагомішими є: громіздкість математичного апарату, необхідність розроблення математичних моделей окремо для кожного об'єкта, великі похибки, відсутність доступу до необхідних прогностичних параметрів та неврахування взаємовпливу компонентів.

Постановка задачі

В роботах [7-9] запропоновано використати в якості засобу прогнозування працездатності КП та С апарат штучних нейронних мереж, розроблено та описано нейромережний метод прогнозування, який враховує особливості КП та С як об'єктів прогнозування та на противагу відомим методам дозволяє врахувати взаємовплив компонентів, обмежитись доступними параметрами та формує єдину штучну нейронну мережу не для поодинокого об'єкта, а для класу об'єктів.

У [9] вказано на залежність точності та достовірності нейромережного прогнозування технічного стану комп'ютерних пристроїв та систем від якості та повноти навчальної вибірки, а також від вибору параметрів які стануть вихідними для побудови прогнозу. Тому забезпечення якісного інформаційного середовища нейромережного прогнозування працездатності КС та їх компонентів є актуальною задачею, вирішення якої дозволить підвищити якість прогнозування, а від так і надійність та ефективність функціонування КС та їх компонентів.

В нейромережному методі прогнозування працездатності КП та С інформаційне середовище визначається моделлю ОП та є обмеженим даними про об'єкт прогнозування, які є доступними на етапі експлуатації. Крім того, вихідною точкою для отримання прогнозу про працездатність об'єкта прогнозування є одноразове отримання значень його прогностичних параметрів. Це призводить до втрати такої важливої і цінної інформації, як динаміка зміни технічного стану об'єкта прогнозування. Також не враховується інформація про надійнісні аспекти об'єктів прогнозування, попередні прогнози та інша апріорна діагностична інформація.

Тому метою даної роботи є підвищення ефективності нейромережного методу прогнозування працездатності КС за рахунок внесення змін в модель КП та С, які дозволять використати апріорну діагностичну інформацію.

Модель комп'ютерних пристроїв та систем як об'єктів прогнозування

Нейромережний метод прогнозування працездатності КП та С обумовлює узагальнену модель об'єкта прогнозування:

$$M_{OP} = \langle O_G, O_D \rangle, \quad (1)$$

де M_{OP} – модель об'єкта прогнозування; O_G – множина його головних ознак; O_D – множина допоміжних ознак. Виділення в моделі двох складових частин дозволило розмежувати поняття класу об'єктів прогнозування, для яких розробляється система та окремого об'єкта (КП чи С).

В свою чергу, множина головних ознак визначається набором:

$$O_G = \langle M_{B\Phi}, M_{VE}, M_K \rangle, \quad (2)$$

де $M_{B\Phi}$ – множина виконуваних функцій; M_{VE} – множина показників умов експлуатації; M_K – множина компонентів, з яких складається ОП. З точки зору модифікації моделі ОП, множина головних ознак є достатньо зручною, оскільки дозволяє внести апріорну діагностичну інформацію (надійнісні характеристики, особливості впливу виконання різних функцій на технічний стан, оптимальність конфігурацій та інші). При цьому вигляд моделі не змінюється, що дозволяє залишити всі принципи та особливості вже розробленого та апробованого методу без змін. Змінюється лише наповнення моделі, а не її структура.

Множина допоміжних ознак складається з двох матриць:

$$O_D = \langle P_{zag}, P_{konkr} \rangle, \quad (3)$$

де P_{zag} – матриця загальних та системних характеристик ОП; P_{konkr} – матриця конкретних характеристик та значень параметрів ОП [7-10]. Можливості внесення додаткових інформаційних ознак в множину допоміжних ознак значно обмежені в порівнянні з головними. Матриця P_{konkr} має чітко визначену структуру та чітко обумовлює інформацію яка може міститися в ній. Тому внесення змін в цю частину моделі можливе тільки на структурному рівні та вимагатиме модифікації методу. Матриця P_{zag} містить частину, вміст якої може бути змінений без зміни методу. Це системні характеристики. Згідно поданого у [10] опису мікропроцесорних пристроїв та систем як об'єктів прогнозування під системними характеристиками розуміють опис ОП з точки зору єдності компонентів системи. Тобто, в якості параметрів розглядаються такі, що не можуть бути явно віднесені до одного з компонентів системи. Подібний опис допускає деяку варіативність, тому дозволяє внести обмежену кількість апріорної інформації щодо узагальнення впливу одних і тих же характеристик на стан об'єкта прогнозування в цілому.

Отже, беручи до уваги вищезгадане, слід зазначити, що при аналізі запропонованої моделі ОП не виявлено можливості використати найбільш вагомому апріорну діагностичну інформацію про ОП, а саме, інформацію щодо попередніх прогнозів та динаміки зміни технічного стану ОП. Без цієї інформації нейромережний метод є статичним і не дозволяє динамічно реагувати на погіршення чи поліпшення технічного стану ОП. Тому було вирішено внести зміни у модель КП та С як об'єктів прогнозування та дослідити ефективність прогнозування з використанням модифікованої моделі.

Місце апріорної діагностичної інформації в нейромережному методі прогнозування

Під апріорною діагностичною інформацією будемо розуміти діагностичну інформацію про ОП, процес та метод прогнозування, яка є доступною ще до початку проведення прогнозування, і може бути використана для підвищення його ефективності.

При розгляді прогнозування працездатності КС та їх компонентів апріорну інформацію доцільно розділити на чотири класи:

- інформація про мету прогнозування;
- інформація про предметну область;
- інформація про попередні прогнози;
- інформація про технологію формування навчальної вибірки та отримання кінцевих результатів.

Розглянемо кожен з виділених класів.

Інформація про мету прогнозування. Загальною метою прогнозування є отримання прогнозу про майбутній стан об'єкта, в кожній предметній галузі, однак, для кожної окремої системи можна конкретизувати поставлену задачу, визначивши особливості мети прогнозування. Наприклад, це може бути значення певного параметра пристрою або системи (або границі значення), на заданому інтервалі часу; імовірність знаходження об'єкта в працездатному стані на заданому інтервалі часу; час, протягом якого значення параметрів об'єкта будуть знаходитись у заданих межах і таке інше. Цей вид апріорної інформації визначає саму постановку задачі на прогнозування. В нейромережному методі, вказана інформація суттєво впливає на архітектуру прогнозуючої ШНМ. Зокрема, мета прогнозування визначає сенс значень на виході штучної нейронної мережі.

Інформація про предметну область. Це найбільш багатий та різномірний клас апріорної інформації. З одного боку він містить інформацію про основні сутності предметної області, їх характеристики, взаємовплив і таке інше. З іншого боку клас містить інформацію про відображення сутностей предметної області на конкретні типи та класи об'єктів. Це достатньо добре корелюється з узагальненою моделлю об'єкта прогнозування. В нейромережному методі запропонований параметричний опис ОП, в якому

відокремлені параметри, які описують поточний технічний стан та параметри, які визначають клас об'єкта прогнозування та його місце у цьому класі.

Інформація про попередні прогнози є достатньо корисною та утворює третій клас апіорної діагностичної інформації. Якщо мова йде про конкретний акт прогнозування, то він починається з формування вихідних даних про поточний технічний стан ОП та закінчується отриманням прогнозу. Очевидним є те, що не ефективно починати новий акт, без врахування результатів попередніх прогнозувань. В свою чергу, особливістю нейромережного методу прогнозування працездатності КС [4, 5] є те, що при його розробленні автори намагались мінімізувати необхідність тривалого спостереження за технічним станом ОП. Це дає можливість прогнозувати працездатність нових об'єктів (моніторинг яких раніше не проводився) навіть при одноразовому визначенні технічного стану. Такий підхід є достатньо ефективним для одноразового прогнозування. Однак, в дійсності, для більшості компонентів КС, в обслуговуючих системах проводиться постійний моніторинг їх технічного стану. Від так з'являється можливість накопичувати такі дані та отримувати фактично додаткову діагностичну апіорну інформацію.

Отримана таким чином інформація про конкретний об'єкт прогнозування є достатньо вагомою як з точки зору прогнозування стану цього об'єкта, так і з точки зору корегування системи прогнозування в цілому (внесення додаткових даних про поведінку об'єктів класу). В нейромережному методі прогнозування працездатності мікропроцесорних засобі запропоновано використовувати отриману інформацію про ОП для донавчання нейронної мережі, що повинно забезпечити підвищення достовірності і точності прогнозу для даного конкретного об'єкта.

Однак технологія проведення донавчання штучної нейронної мережі вимагає від користувача формування векторів у формі: $\langle X_i^j, Y_i \rangle$, де $X_i = (X_{заг\ i}, X_{конкр\ i}^j, T_{пр\ i})$ - вектор «причин», який складається з $X_{заг\ i}$ - значення загальних параметрів, $X_{конкр\ i}^j$ - значення конкретних параметрів в момент часу t_j , $T_{пр\ i}$ - інтервал часу на який робиться прогноз, а Y_i - ймовірність знаходження відповідного ОП в працездатному стані через час $- T_{пр\ i}$. Якщо вектор X_i формується достатньо легко, то визначення Y_i може призвести до внесення додаткових похибок, що навпаки призведе до погіршення якості прогнозу. Адже про ймовірність Y_i абсолютно точно можна вести мову тільки в момент виходу з ладу ОП, тоді $Y_i = 0$, в решті випадків оцінити ймовірність Y_i з необхідною точністю є достатньо складною задачею. Отже отримаємо ситуацію, коли накопичена інформація може бути використана тільки для корегування поведінки об'єктів класу в цілому, і не забезпечує підвищення якості прогнозу для конкретного об'єкта. Тому необхідно провести адаптацію моделі ОП та методу прогнозування для врахування накопичуваної апіорної діагностичної інформації.

Четвертий клас містить інформацію про безпосередній метод проведення прогнозування. Висуваються вимоги до штучних нейронних мереж, даних, якості навчальної вибірки, класу використовуваних параметрів і таке інше. Ця інформація є корисною та використовуються на етапі проектування інтелектуальних систем прогнозування. На сьогодні, немає моделей використання та формалізованого опису цієї інформації щодо прогнозування працездатності КС. Задача вимагає додаткових досліджень, що виходять за межі даної роботи.

Модифікована модель КП та С як об'єктів прогнозування

Нейромережний метод прогнозування працездатності КС містить положення, що частково дозволили врахувати апіорну діагностичну інформацію, зокрема, першого та другого класів.

Другий клас, а саме апіорна діагностична інформація про предметну область, враховується при формуванні вигляду та значень матриці $P_{заг}$, яка містить значення загальних параметрів, що визначають місце конкретної КС у своєму класі. Також інформація про предметну область враховується при описі класу об'єктів, для яких розробляється інтелектуальна система прогнозування. Згідно з моделлю ОП, обов'язковою умовою при включенні об'єкта до класу є відповідність його множини виконуваних функцій, множини умов експлуатації та множини складових компонентів відповідним множинам класу.

Важливим аспектом використання апіорної діагностичної інформації другого класу (інформація про предметну область) є пошук оптимальних «загальних» параметрів, значення яких будуть складати матрицю $P_{заг}$. Як свідчать проведені дослідження, перенасичення матриці $P_{заг}$ призводить до «приглушення» чутливості прогнозуючої штучної нейронної мережі. На рис. 1 наведено залежність похибки навчання ШНМ від кількості використаних загальних параметрів КС. Як видно з графіку, суттєве зменшення похибки навчання, при конкретній реалізації, відбулося при зменшенні кількості загальних параметрів з 14 до 13. Тенденція зменшення похибки навчання при зменшенні кількості загальних параметрів відобразилася і при проведенні верифікації роботи навченої прогнозуючої мережі. Зрозуміло, що вплив кількості загальних параметрів на похибку навчання та похибку прогнозування ШНМ буде змінюватись залежно від загальної кількості параметрів, особливостей предметної області та ОП, інформативності кінцевих параметрів, розмірності класу і таке інше. Однак загальна тенденція зберігається. Тому важливим є застосування апіорної діагностичної інформації саме на етапі формування матриці загальних параметрів.



Рис. 1. Залежність отриманої похибки навчання від кількості загальних параметрів

Місцем для врахування апріорної діагностичної інформації третього класу, а саме дані про результати попередніх прогнозів, на нашу думку є матриця конкретних параметрів, яка для КС має вигляд:

$$P_{\text{конкр}}^j = \begin{array}{c|ccccccc} & 1 & 2 & \dots & i & i+1 & \dots & n \\ \hline name_1 & p_{11}^j & p_{12}^j & \dots & p_{1i}^j & p_{1(i+1)}^j & \dots & p_{1n}^j \\ \hline name_2 & p_{21}^j & p_{22}^j & \dots & p_{2i}^j & p_{2(i+1)}^j & \dots & - \\ \hline \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \hline name_k & p_{k1}^j & p_{k2}^j & \dots & p_{ki}^j & - & - & - \end{array}$$

де $P_{\text{конкр}}^j$ – матриця параметрів конкретного ОП, визначених у час t_j ; $name_k$ – вказівник на k -й компонент системи. Кількість компонентів в обох матрицях ($P_{\text{заг}}$ та $P_{\text{конкр}}^j$) однакова. Кожному рядку загальних параметрів матриці $P_{\text{заг}}$ відповідає рядок конкретних параметрів матриці $P_{\text{конкр}}^j$; p_{ki}^j – значення i -го параметра компонента k , в момент часу t_j . Кількість параметрів для кожного компонента відрізняється від інших та від кількості загальних параметрів того самого компонента в матриці $P_{\text{заг}}$, вона визначається конкретною моделлю компонента та можливостями, які надані для збору значень параметрів.

Отже в явному вигляді, інформація про попередні стани КС неврахована. Для врахування результатів попередніх прогнозів, матрицю конкретних параметрів пропонується модифікувати наступним чином: доповнити кожен рядок матриці ще двома рядками. Наприклад, перший рядок слід замінити на такі:

$$name_1 \rightarrow \begin{array}{c|ccccccc} & 1 & 2 & \dots & i & i+1 & \dots & n \\ \hline value & p_{11}^j & p_{12}^j & \dots & p_{1i}^j & p_{1(i+1)}^j & \dots & p_{1n}^j \\ \hline worst & p_{11}^w & p_{12}^w & \dots & p_{1i}^w & p_{1(i+1)}^w & \dots & p_{1n}^w \\ \hline reliab & r_{11}^w & r_{12}^w & \dots & r_{1i}^w & r_{1(i+1)}^w & \dots & r_{1n}^w \end{array}$$

В такому випадку, рядок позначений як *value* містить поточні значення параметрів технічного стану КС, рядок *worst* містить найгірші з зафіксованих значень відповідних параметрів під час проведення попередніх прогнозів, рядок *reliab* значення імовірності знаходження даної КС в працездатному стані на інтервалі, що відповідає поточному часу, але визначене у момент найгіршого значення відповідного параметра.

У випадку, якщо поточне значення параметру є найгіршим із зафіксованих, воно дублюється у другому рядку, третій рядок при цьому залишається пустим. Така ситуація відповідає першому формуванню матриці конкретних параметрів для отримання першого прогнозу.

Модифікована матриця конкретних параметрів вирішує декілька проблем. По-перше, визначається та враховується під час прогнозування динаміка зміни значень параметрів. По-друге, враховується прогноз, який теж може мати динамічний характер. Крім того, матриця зручна для співставлення та пошуку залежності між значенням конкретного параметру та кінцевим прогнозом.

Розроблені в процесі досліджень програмні засоби моніторингу та накопичення інформації про технічний стан КС, дозволяють накопичувати всю необхідну інформацію. Тому при формуванні навчальної вибірки не виникло проблем у зв'язку з модифікацією моделі ОП.

Проведені дослідження (моделювання роботи прогнозуючої штучної нейронної мережі) щодо порівняння використання існуючої та модифікованої моделі показали підвищення точності та достовірності прогнозування на одній і тій же навчальній вибірці.

Для формування навчальної вибірки V використані дані накопичені під час спостереження за

персональними комп'ютерами спеціалізованих лабораторій кафедри системного програмування Хмельницького національного університету. Формат збереження даних, обумовлений моделлю КС як об'єкту прогнозування, передбачав накопичення інформації окремо по кожному компоненту робочої станції, який входить до складу матриць $P_{заг}$ та $P_{конкр}^j$.

Під час навчання вся наявна вибірка V була розділена на дві частини: V_1 – навчальна вибірка та V_2 – вибірка для верифікації процесу навчання. Показник достовірності навчання D , що визначається на етапі верифікації, обраховувався як відносна кількість збігань значень істинного вектора Y_i^j , P_i -ї ситуації протоколу R , $i = 1..M$ та прогнозованого вектора Y_i^{net} – вихідних сигналів ШНМ:

$$D = \frac{1}{M} \sum_{i=1}^M \xi_i = \frac{1}{Mz_3} \sum_{i=1}^M \sum_{j=1}^{z_3} g_{ji}, \quad (4)$$

де M – кількість ситуацій вибірки верифікації V_2 ; z_3 - кількість компонентів системи, стан яких прогнозується (розмірність вектора Y^{net}).

Змінна g_{ji} визначалась як:

$$\begin{cases} 1, & \text{якщо } \zeta_{ji} \leq \varepsilon_{ji}; \\ 0, & \text{якщо } \zeta_{ji} > \varepsilon_{ji}, \end{cases} \quad (5)$$

де ζ_{ji} – абсолютна, приведена до шкали, похибка прогнозування працездатного стану j -го компонента об'єкта прогнозування при подаванні на вхід навченої ШНМ i -го вектора X_i верифікаційної вибірки V_2 ; ε_{ji} – граничне значення абсолютної, приведеної до шкали похибки прогнозування ζ_{ji} .

Отримавши достатній рівень достовірності навчання штучної нейронної мережі прогнозування працездатності КС було проведено оцінку похибки прогнозування. Похибка прогнозування оцінювалась на вибірці V_3 , яка не містила векторів з V_1 та V_2 . В результаті, нейромережний метод дав похибку прогнозування 14 % при використанні модифікованої моделі на противагу 19 % при використанні моделі запропонованої у [10]. Разом з тим, було відмічено незначне збільшення похибки навчання ШНМ. Таке положення вимагає проведення подальших досліджень, зокрема, необхідно перевірити результати прогнозування з іншими навчальними вибірками інших класів об'єктів прогнозування.

Висновки

Підвищення точності та достовірності прогнозування технічного стану компонентів КС є актуальною науково-технічною задачею, що вимагає пошуку нових підходів та методів, які враховують особливості сучасних комп'ютерних пристроїв. Одним із перспективних напрямів проведення досліджень щодо підвищення точності та достовірності прогнозування є використання компонентів теорії штучного інтелекту, зокрема штучних нейронних мереж. Однак ефективність використання інтелектуальних методів прогнозування залежна від якості, кількості та виду діагностичної інформації. Тому, на сьогодні, в окресленому напрямі актуальними та пріоритетними є дослідження щодо уточнення складу апіорної діагностичної інформації, аналізу способів використання різної апіорної інформації на різних етапах процесу прогнозування, визначення методів представлення апіорної інформації.

Проведені у роботі дослідження свідчать, що модифікація моделі КП та С як ОП призвела до поліпшення результатів прогнозування. Насамперед, це відбулося завдяки отриманій можливості значно розширити інформаційне середовище процесу прогнозування та використати накопичену апіорну діагностичну інформацію. Важливим є той факт, що використання модифікованої моделі не примушує до збору додаткової інформації про ОП, а лише дозволяє використати вже накопичені дані.

Роботу виконано за підтримки Державного фонду фундаментальних досліджень (Міністерства освіти і науки України) в межах гранту Ф25.1/121 «Дослідження методів внесення апіорної діагностичної інформації в структуру штучних нейронних мереж для реалізації процесу діагностування комп'ютерних систем».

Література

1. Connors J., Martin D., and Atlas L. Recurrent neural networks and robust time series prediction – IEEE Trans. Neural Networks, vol. 5, 1994. – pp. 240– 254.
2. Park D., M.El-Sharkawi, Marks R., Atlas L., and Damborg M. Electric load forecasting using an artificial neural network – IEEE Trans. Power Syst., vol. 6, 2001. – pp. 442– 449.
3. P. K. H. Phua, X. Zhu, C. H. Koh “Forecasting stock index increments using neural networks with trust

region methods, " Proc. Int. Joint Conf. Neural Networks 2003, vol. 1. – pp. 260-265.

4. Castiglione F. Forecasting price increments using an artificial Neural Network // Adv. Complex Systems, vol. 1, 2002. – pp. 1-12.

5. Weigend and N. Gerschenfeld, Eds., Time Series Prediction: Forecasting the Future and Understanding the Past. Reading, MA: Addison-Wesley, 1994. – 184 p.

6. Гаскаров Д.В., Голинкевич Т.А., Мозгалеvский А.В. Прогнозирование технического состояния и надежности радиоэлектронной аппаратуры. – М.: Сов. Радио, 1974. – 224 с.

7. Медзатий Д.М. Прогнозування технічного стану мікропроцесорних пристроїв та систем // Вісник національного технічного університету «Львівська політехніка». – 2005. № 555. – С. 158-165.

8. Локазюк В.М., Поморова О.В., Медзатий Д.М. Метод прогнозування технічного стану комп'ютерних систем // Вісник Хмельницького національного університету – 2005. – № 4. – Ч.1. – Т.1. – С.81-85.

9. Victor Lokazyuk, Dmitro Medzatiy. Neural Network Approach to Forecast Working Capacity of Microprocessor and Digital Devices. – Збірник праць за результатами роботи МНПК "ACSN-2007", 1, № 1, 2007. – С.94-99.

10. Медзатий Д. М. Мікропроцесорні пристрої та системи як об'єкти прогнозування // Оптико-електронні інформаційно-енергетичні технології. – № 2 (10). – 2005. – С.119-123.

Надійшла 28.9.2009 р.

УДК 004

В.М ДЖУЛІЙ, К.В. ІВАНОВ
Хмельницький національний університет

МЕТОД КЛАСТЕРИЗАЦІЇ СТИСНЕННЯ ВЕКТОРА ХАРАКТЕРНИХ ОЗНАК

У даній статті представлений метод кластеризації стиснення вектора характерних ознак. Базуючись на запропонованій методиці, розроблено алгоритм рандомізованої кластеризації з метою безпечного хешування, пропонується нова функція затрат для стиснення вектора характерних ознак при застосуванні перцептивного хешування.

In this article the presented method clusterisation compression vector of characteristic signs. Based on the offered method the algorithm of random clusterisation developed with the purpose of safe hashing, the new function of expenses is offered for the compression vector characteristic signs at application of perception hashing.

Ключові слова: алгоритм рандомізованої кластеризації.

Вступ. Проблему отримання хеш зображення розділено на два кроки [1]. Завдання першого кроку полягає в отриманні вектора характерних ознак зображення, тоді як на другому кроці необхідно стиснути цей вектор до остаточної хеш величини.

На першому кроці отримання характерних ознак, двомірне зображення відображається як одномірний вектор характерних ознак. Даний вектор повинен вловити перцептивні якості зображення. Тобто, два зображення, що є однаковими для людської зорової системи, повинні мати близькі на певній метричній відстані вектори характерних ознак. Два зображення, що є точно різними у своєму вигляді, повинні мати вектори характерних ознак, які відрізняються великою відстанню. На другому кроці необхідно стиснути перехідний хеш вектор до остаточної хеш величини. Це включатиме кластеризацію між перехідним хеш вектором введеного джерела (зображення) та перехідним хеш вектором перцептивно ідентичних версій. Вектор характерних ознак зображення формується сумуванням величин хвильових коефіцієнтів в обраних точках характерних ознак [1, 2].

Визначення проблеми. Представимо нотацію, що використовуватиметься у даній статті. Нехай V позначає метричний простір векторів перехідного хешу, отриманих на стадії хеш алгоритму [1, 2]. Нехай $L \in V$ позначає кінцевий ряд векторів $\{l_i\}_{i=1}^n$, до яких застосовується метод кластеризації/стиснення. Нехай $D: V \times V \rightarrow R_+$ буде метричною відстанню, визначеною на області перетину. І нарешті, нехай $C: L \rightarrow \{1, 2, \dots, k\}$ визначає область кластеризації. Необхідно знайти всі зображення, що візуально не можливо відрізнити, із однаковою хеш величиною та із високою вірогідністю. В даному разі хеш функція зображення подібна до квантування вектора (VQ) або до схеми кластеризації. Необхідно згрупувати зображення, вектори перехідного хешу яких метрично близькі, і з високою вірогідністю:

$$\text{if } D(l_i, l_j) < \varepsilon \text{ then } C(l_i) = C(l_j), \quad (1)$$

$$\text{if } D(l_i, l_j) > \delta \text{ then } C(l_i) \neq C(l_j), \quad (2)$$

де $0 < \varepsilon < \delta$. Нехай l_i, l_j позначають довільні вектори на l (відповідно розподілу перехідного хешу) і