

ДОДАТОК А

Графічний матеріал атестаційної роботи

Харківський національний університет радіоелектроніки
Кафедра ЕОМ

**НА ТЕМУ: «МЕТОДИ ТА АЛГОРИТМИ УЗАГАЛЬНЕННЯ
ЗНАНЬ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ
ДЛЯ РЕАЛЬНОГО ЧАСУ»**

АТЕСТАЦІЙНА РОБОТА
ДРУГИЙ (МАГІСТЕРСЬКИЙ РІВЕНЬ)

ВИКОНАВ:
студ гр. СПм-19-1
Слюсар О.В.

КЕРІВНИК
доц. Мартовицький В.О.

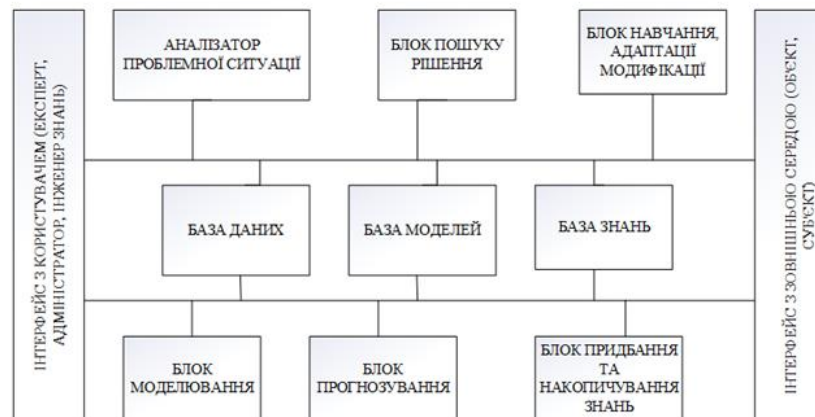
Харків 2020 рік

Метою роботи – є дослідження та розробка методів і алгоритмів узагальнення знань, що дозволяють отримувати узагальнений опис класів ситуацій, що змінюються з часом.

У ході виконання атестаційної роботи було вирішено такі завдання:

- дослідження існуючих методів і алгоритмів уявлення та узагальнення знань в інтелектуальних системах підтримки прийняття рішень;
- розробка методів і алгоритмів узагальнення знань, що дозволяють отримувати загальний опис класів ситуацій (об'єктів), що змінюються з часом;
- вивчення можливості використання методів і алгоритмів узагальнення знань в інтелектуальних системах підтримки прийняття рішень реального часу;
- розширення понятійного апарату: введення понять, які враховують динамічну природу об'єктів узагальнення; формалізація завдання узагальнення для роботи з динамічними даними;
- розробка методів і алгоритмів узагальнення знань для динамічних об'єктів узагальнення;
- проектування і розробка програмного комплексу, що реалізує розглянуті в роботі методи і алгоритми.

Базова структура ІСППР



Показання наявних датчиків в певний момент часу i можна представити у вигляді вектору

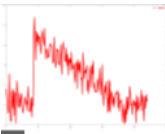
$$s_i = \langle x_1(t = i), x_2(t = i), \dots, x_q(t = i), t = i \rangle$$

Динамічний об'єкт узагальнення

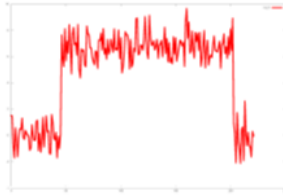
	Параметр ₁	Параметр ₂	...	Параметр _q	Час (t)
(s _i)	$x_1(t = i)$	$x_2(t = i)$...	$x_q(t = i)$	i
(s _{i+1})	$x_1(t = i + 1)$	$x_2(t = i + 1)$...	$x_q(t = i + 1)$	$i + 1$
(s _{i+2})	$x_1(t = i + 2)$	$x_2(t = i + 2)$...	$x_q(t = i + 2)$	$i + 2$
...
(s _{i+r-1})	$x_1(t = i + r - 1)$	$x_2(t = i + r - 1)$...	$x_q(t = i + r - 1)$	$i + r - 1$

Методи пошуку аномалій

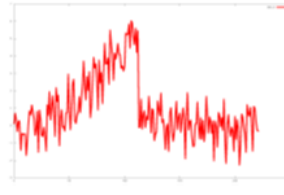
- нормальні об'єкти належать кластеру, аномальні – ні;
- нормальні об'єкти лежать близько до центру кластера, аномальні – далеко від центру;
- нормальні об'єкти належать великим, щільним кластерам, в той час як аномалії належать невеликим і розрідженим.



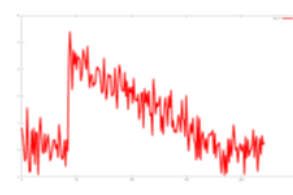
Представники тимчасових рядів



а) «циліндр»



б) «дзвін»



в) «воронка»

Набір даних «циліндр-дзвін-воронка»

тимчасові ряди для даного набору можна отримати штучно за такими формулами:

- 1) «циліндр»: $c(t) = (6 + \zeta) * \chi_{[a,b]}(t) + \epsilon(t), 1 \leq t \leq M,$
- 2) «дзвін»: $b(t) = (6 + \zeta) * \chi_{[a,b]}(t) * \frac{(t-a)}{(b-a)} + \epsilon(t), 1 \leq t \leq M,$
- 3) «воронка»: $f(t) = (6 + \zeta) * \chi_{[a,b]}(t) * \frac{(b-t)}{(b-a)} + \epsilon(t), 1 \leq t \leq M,$

де:

1) M – довжина часового ряду;

$$2) \chi_{[a,b]} = \begin{cases} 0, t < a \\ 1, a \leq t \leq b \\ 0, t > b \end{cases}$$

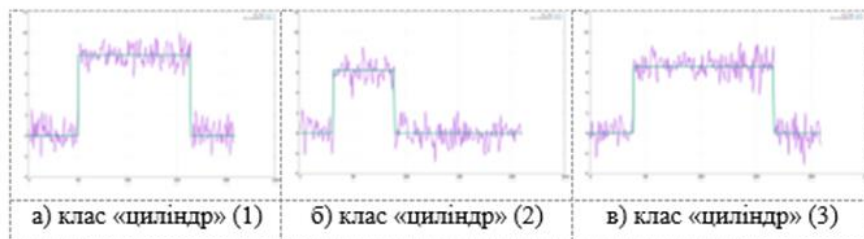
3) ζ – випадкова величина, що підкоряється стандартному нормальному розподілу $N(0,1)$;

4) $\epsilon(t)$ – випадкові величини, що підкоряються стандартному нормальному розподілу $N(0,1)$;

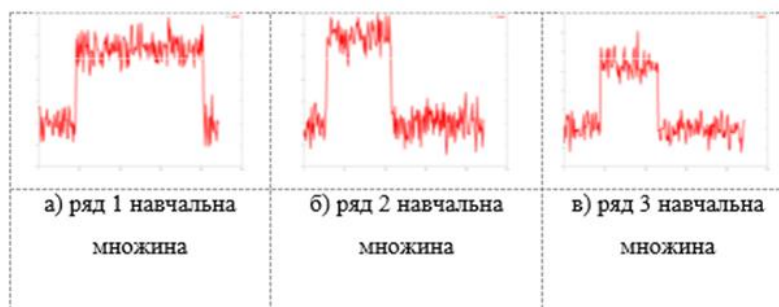
5) a – випадкова величина, що підкоряється рівномірному розподілу на відрізьку;

6) b – випадкова величина, що підкоряється рівномірному розподілу на відрізьку.

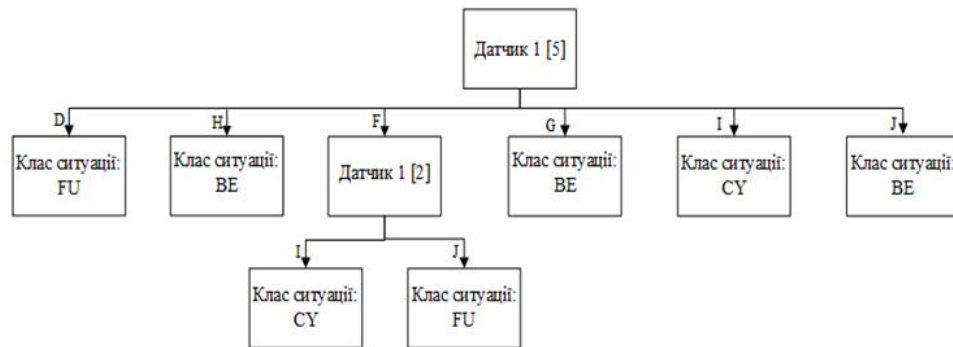
Приклади штучно побудованих формул тимчасових рядів класу «циліндр»



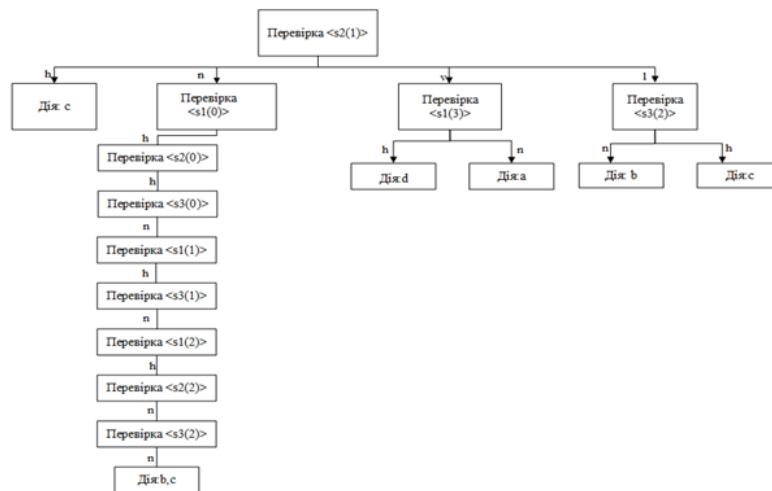
Навчальна вибірка TS_STUDY з трьох тимчасових рядів



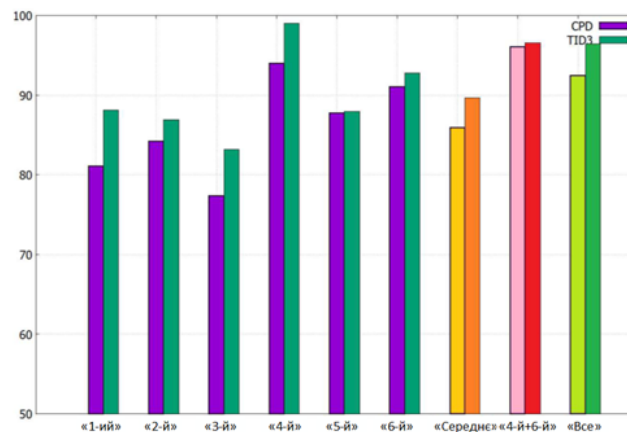
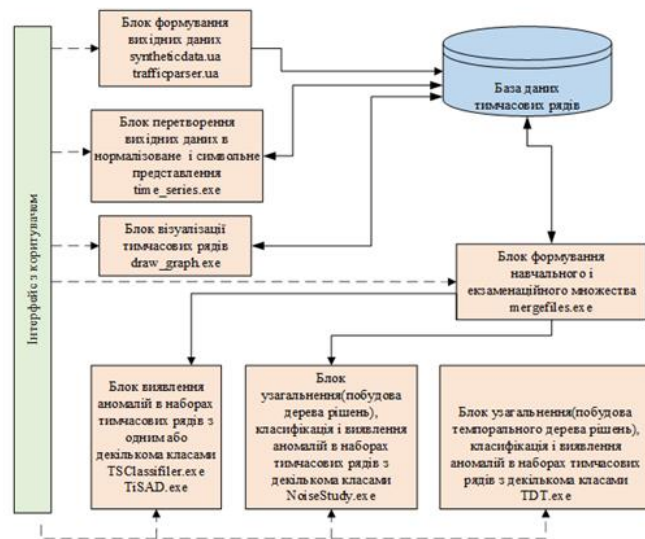
Дерево рішень



Дерево рішень, побудоване з використанням алгоритму «Темпоральний ID3»



Архітектура програмного комплексу



Точність класифікації (%) – набір даних wafer. Класифікація по одному і кількома ознаками. CPD – алгоритм «CPD»; TID3 – алгоритм «Темпоральний ID3»

ВИСНОВОК:

- 1. Проведено огляд методів представлення знань в сучасних інтелектуальних системах і розглянута проблема роботи з даними, явно залежать від часу - темпоральними даними.
- 2. Розглянуто проблему виявлення аномалій в разі, коли стан складної технічної системи представимо тимчасовим поруч.
- 3. На підставі аналізу підходів до вирішення завдання виявлення аномалій в наборах часових рядів запропоновані методи і розроблені алгоритми «TS-ADEEP», «TS-ADEEP-Multi» виявлення аномалій для наборів часових рядів з одним і декількома класами.
- 4. Проведено аналіз різних способів подання знань в інтелектуальних системах. Виділено клас динамічних об'єктів, які представлені кількома часовими рядами.

ВИСНОВОК:

- 5. Проведено огляд методів і алгоритмів побудови темпоральних дерев рішень. Запропоновано новий алгоритм «Темпоральний ID3» побудови темпоральних дерев рішень, що використовує в якості критерію вибору спостережень для розбиття величину «приріст інформативності».
- 6. Для алгоритмів TS-ADEEP і TS-ADEEP-Multi показано, що на відомих наборах даних точність виявлення аномалій порівнянна з точністю виявлення аномалій низкою відомих алгоритмів (метод опорних векторів, алгоритм C4_5, байєсовські мережі, алгоритм Random Forest і ін.).
- 7. Проведено порівняння точності класифікації часових рядів з використанням алгоритму «Темпоральний ID3» з відомими алгоритмами класифікації (метод K найближчих сусідів, C4_5, байєсовські мережі, Random Forest і ін.).
- 8. Проведено порівняння точності класифікації динамічних об'єктів, представлених наборами часових рядів, з алгоритмом «CPD», найбільш «близьким» до алгоритму «Темпоральний ID3».
- 9. Результати експерименту дозволяють зробити висновок про ефективність використання алгоритму «Темпоральний ID3» для роботи з динамічними об'єктами, які представлені наборами часових рядів.

ДОДАТОК Б

ПРИКЛАД РОБОТИ З ПРОГРАМНИМ КОМПЛЕКСОМ

Розглянемо роботу з програмним комплексом на прикладі. Припустимо, що потрібно провести моделювання для набору даних «циліндр-дзвін-воронка». Утиліта `syntheticdata.py` генерує необхідну кількість тимчасових рядів, що відносяться до класів «циліндр», «дзвін», «воронка». Будемо вважати, що дана утиліта згенерувала файли `cyl_0.txt`, `cyl_1.txt`, .. для класу «циліндр», `bel_0.txt`, `bel_1.txt`, .. для класу «дзвін», `fun_0.txt`, `fun_1.txt`, .. для класу «воронка».

Далі необхідно перетворити вихідні дані в нормалізований і символічний (SAX) вид. Для цього потрібно сформувати bat-файл наступного виду:

```
process_wparams & (
for %%i in (cyl_, bel_, fun_) do for /l %%j in (0,1,9) do
process_serie_total cbf/%%i%%j.txt
) & (
5 %FULLPATH%MERGEFILEPATH% -d %FULLPATH%scripts/> -m «Sax_» -o «OUT
«
)

```

За допомогою `TSClassifier.exe` також можна отримати дані в форматі, необхідному для алгоритмів побудови дерев рішень.

```
SET MYPROGDIR=«../.././Release/tsclassifier.exe»
for %%j in (1, 5, 10, 15, 20, 23, 25, 30, 40, 50) do (
for %%s in (5, 10, 15, 20, 25, 30, 40, 50) do (
5 %MYPROGDIR% --classifier eepmulti --mode s --studysset «
sax_%%j_points_%%s_letters_study.txt» --testset «sax_%%
j_points_%%s_letters_test.txt» --lettddist «dist_%%s» --
alphabetsize %%s --dumppid3 --dumpclasses «#class:
CYLINDER,#class:BELL,#class:FUNNEL.»
)
)
)

```

Та темпоральних дерев рішень

```
SET MYPROGDIR=«../.././Release/tsclassifier.exe»
for %%j in (1, 5, 10, 15, 20, 23, 25, 30, 40, 50) do (
for %%s in (5, 10, 15, 20, 25, 30, 40, 50) do (
5 %MYPROGDIR% --classifier eepmulti --mode s --studysset «
sax_%%j_points_%%s_letters_study.txt» --testset «sax_%%
j_points_%%s_letters_test.txt» --lettddist «dist_%%s» --

```

```
alphabetsize %%s --dumptdt --dumpclasses «#class:
CYLINDER,#class:BELL,#class:FUNNEL.»
)
)
```

Для перевірки алгоритма побудови дерева рішень потрібно запустити

bat-файл наступного виду

```
SET MYPROGDIR=«noisestudy.exe»
for %%j in (1, 5, 10, 15, 20, 23, 25, 30, 40, 50) do (
for %%s in (5, 10, 15, 20, 25, 30, 40, 50) do (
5 %MYPROGDIR% --classifier eepmulti --mode s --studysset «
sax_%%j_points_%%s_letters_id3.data» --testset «sax_%%
j_points_%%s_letters_id3.test» --lettdist «dist_%%s» --
alphabetsize %%s
)
)ь
```