

Міністерство освіти і науки України

Національна академія наук України

Мала академія наук України

Національний університет «Полтавська політехніка імені Юрія Кондратюка»

Секція
«Академічна й університетська наука»

Збірник наукових праць
за матеріалами

Всеукраїнської науково-практичної конференції
«Сучасні рецепції світоглядно-ціннісних
орієнтирів Григорія Сковороди»

02 грудня 2022 року

Том 2

Полтава 2022

УДОСКОНАЛЕННЯ ТЕХНОЛОГІЇ АНАЛІЗУ, КЛАСИФІКАЦІЇ ТА ПРОГНОЗУВАННЯ ДІЯЛЬНОСТІ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ

1. Актуальність. Існуючі технології, котрі вирішують завдання аналізу, класифікації, прогнозування діяльності складних технічних систем зорієнтовані на конкретні досліджувані об'єкти та розглядаються окремо. У якості інструментарію досліджень використовують класи лінійних, нелінійних моделей та різні різновиди бібліотек [1, 2]. Великі масиви ненормалізованих оцінок ускладнюють процес побудови моделей. Недостатнє використання нормованих інтегрованих показників, індексів актуалізує необхідність дослідження.

2. Мета. Підвищити ефективність аналізу складних технічних систем за допомогою інтелектуальної технології класифікації та прогнозування рівня якості взаємодії елементів підсистем складних технічних систем на основі методу визначення узагальненого показника якості взаємодії елементів підсистем системи.

3. Методика та організація дослідження

3.1. Метод визначення узагальненого показника якості взаємодії елементів підсистем системи. Дано. Складна технічна система (далі система). Система складається з чотирьох елементів підсистеми EP_1, EP_2, EP_3, EP_4 диференційованих на п'ять станів. Елементи першої, другої й четвертої підсистем мають максимальне значення оцінки 5, коли $X_1=X_2=X_3=X_4=5$, а третьої $EP_3 = 5^{0,2}+5^{0,2}+5^{0,2}+5^{0,2} = 5,98$. Тому до елементів підсистем використовують наступні методи нормалізації змінних у межах $[0; 1]$, що дозволяє досягти уніфікації. Якість взаємодії елементів підсистем системи (ЯВЕП) описується формулою (1)

$$ЯВЕП = \lambda_1 \cdot (EP_1/X_{imax}) + \lambda_2 \cdot (EP_2/X_{imax}) + \lambda_3 \cdot (EP_3/X_{imax}) + \lambda_4 \cdot (EP_4/X_{imax}) \quad (1)$$

де $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ – вагові коефіцієнти, знаходяться за технікою [4];

EP_i – елементи підсистем, що визначаються

$$EP_1 = X_1 + X_2 + X_3 + X_4 / 4;$$

$$EP_2 = \sqrt[4]{X_1 X_2 X_3 X_4};$$

$$EP_3 = X_1^{\lambda_1} X_2^{\lambda_2} X_3^{\lambda_3} X_4^{\lambda_4};$$

$$EP_4 = 4 / (1/X_1) + (1/X_2) + (1/X_3) + (1/X_4).$$

Запропонований підхід визначення якості взаємодії елементів підсистем системи пристосований до діагностики складних людино-машинних систем різної природи. Однією з особливостей підходу є автоматичне визначення вагових коефіцієнтів за технікою з роботи [3]. За необхідності кількість станів елементів підсистем може бути збільшена (другий сценарій дій).

3.2. Опис запропонованої технології аналізу, класифікації та прогнозування діяльності складних технічних систем

Оцінки визначені за формулою (1) накопичуються у датасет та використовуються у моделях машинного навчання за відомими технологіями [4 – 5]. Моделі машинного навчання дозволяють вирішити завдання прогнозування та класифікації систем.

Блок-схема технології аналізу, класифікації та прогнозування діяльності складних технічних систем складається з чотирнадцяти блоків інструменти котрих обрані з [4 – 5].

Блок 1. Здійснюється завантаження датасету з оцінками елементів підсистем та узагальненою оцінкою якості взаємодії елементів підсистем.

Блок 2. Передбачає створення та аналіз описової статистики [5]: кількість даних, середнє значення, стандартне відхилення, мінімальне й максимальне значення та процентилі.

Блок 3. Фіксація пропусків оцінок та за наявності пропусків їх очищення.

Блок 4. Очищені оцінки піддаються дослідженню на предмет вивчення викидів, кореляції.

Блок 5. Здійснюється вибір досліджуваного завдання (прогнозування чи класифікація).

Блок 6. Здійснюється дослідження регресійних моделей лінійних та нелінійних класів.

Блок 7. Виконується оцінка точності побудованих моделей та визначення існування процесу перенавчання, недонавчання чи компромісу.

Блок 8. Якщо отримані результати у блоці 7 не задовольняють людину, котра приймає рішення здійснюється перехід у блок 6, інакше перехід у блок 9.

Блок 9. Здійснюється використання побудованої моделі для вирішення завдання прогнозування та формуються звіти.

Блок 10. Побудова класифікаційних моделей розпочинається з вивчення існування збалансованої вибірки. Якщо вибірка незбалансована використовуються відповідні підходи для балансування класів. За допомогою класифікаційної моделі вирішується завдання класифікації систем за узагальненим показником якості взаємодії елементів підсистем.

Блок 11. Здійснюється дослідження класифікаційних моделей різних класів.

Блок 12. Виконується оцінка точності побудованих класифікаційних моделей та визначення існування процесу перенавчання, недонавчання чи компромісу.

Блок 13. Якщо отримані результати у блоці 12 не задовольняють людину, котра приймає рішення здійснюється перехід у блок 10, інакше перехід у блок 14.

Блок 14. Здійснюється використання побудованої моделі класифікації для вирішення завдання класифікації елементів підсистем та формуються звіти.

За результатами роботи запропонованої технології формуються рекомендації щодо обрання оптимального елемента підсистеми для поліпшення

4. Результати дослідження технології аналізу, класифікації та прогнозування діяльності складних технічних систем. У експериментальному дослідженні технології аналізу, класифікації та прогнозування діяльності складних технічних систем використано набір даних 625 систем (клас 0 – 300 систем, клас 1 – 325 систем). Інтервали класів оцінок (перший клас – оцінки систем у межах [0; 0,6], другий – [0,601; 1,0]). Завдання регресії вирішується за допомогою моделей Gradient Boosting Regressor, Random Forest Regressor, KNeighbors Regressor, Bagging Regressor. У якості критеріїв порівняння моделей використано coefficient of determination (R^2), Mean Absolute Error (MAE), Max Error (MAX), Root Mean Square Error (RMSE). Порівняльний аналіз отриманих результатів роботи досліджуваних моделей показані у табл. 1.

Таблиця 1

Результати дослідження моделей Gradient Boosting Regressor, Random Forest Regressor, KNeighbors Regressor, Bagging Regressor за критеріями (R^2), (MAE), (MAX), (RMSE)

Модель	R^2 train/test	MAE	MAX	RMSE
GradientBoostingRegressor	1.0/1.0	0.00045	0.0373	0.0028
RandomForestRegressor	1.0/1.0	0.00044	0.0391	0.0029
KNeighborsRegressor	0.99/0.99	0.00248	0.0532	0.0062
BaggingRegressor	1.0/0.99	0.00047	0.0397	0.0029

Для вирішення завдання класифікації використано моделі Gradient Boosting Classifier, Random Forest Classifier, KNeighbors Classifier, Bagging Classifier, Support Vector Machine Classifier, LGBM Classifier. Критеріями якості слугували accuracy, precision, recall, f1-score.

При дослідженні усі моделі продемонстрували високий показник точності. Однією з найліпших моделей була Support Vector Machine Classifier. Результати дослідження побудованої класифікаційної моделі Support Vector Machine Classifier подані у табл. 2.

Таблиця 2

Результати дослідження моделі Support Vector Machine Classifier

Назва класу системи	Precision	Recall	F1-score	Всього
0	0.96	0.93	0.94	99
1	0.94	0.96	0.95	108
Accuracy	0,95			207
Accuracy train/test	0.95/0.94			

За результатами дослідження висока точність регресійних та класифікаційних моделей пов'язана з особливостями вхідних оцінок, кількістю класів тощо. Якщо збільшити кількість класів оцінок узагальненого показника якості, то виникне проблема із балансуванням класів, що відповідно вплине на кінцеву точність моделей.

Висновки. 1. Аналізуючи проблематику технології прогнозування та класифікації складних технічних систем запропоновано використання методу визначення узагальненого показника якості взаємодії елементів підсистем системи.

2. Запропоновано технологію аналізу, класифікації та прогнозування діяльності складних технічних систем та проведено її експериментальну апробацію.

3. У наступних дослідженнях варто ретельніше дослідити використання відомих моделей машинного навчання на основі узагальненого показника якості взаємодії елементів підсистем системи.

Література

1. *Жирикова І. Структурно-Параметрична Ідентифікація Правил Бази Модельних Знань [Електронний ресурс] / І. Жирикова, С. Голуб – Режим доступу до ресурсу: <http://itcm.comp-sc.if.ua/2017/Zhyriakova.pdf>.*

2. *Павлишенко Б. М. Методи інтелектуального аналізу консолідованих даних для підтримки прийняття рішень : автореф. дис. на здобуття наук. ступеня докт. техн. наук : спец. 05.13.23 «системи та засоби штучного інтелекту» / Павлишенко Б. М. – Харків, 2021. – 41 с.*

3. *Слабченко О. О. Методи, моделі й інформаційна технологія відновлення пропущених даних із акаунтів соціальних мереж : дис. канд. техн. наук : 05.13.06 / Слабченко О. О. – Кременчук, 2017. – 229 с.*

4. *Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.*

5. *Stefanie M. Hands-on data analysis with Pandas / Molin Stefanie. – Birmingham: Packt Publishing Ltd., 2019. – 707 с.*