**ВПЛИВ РОЗМІРУ НАВЧАЛЬНОЇ ВИБІРКИ НА ЯКІСТЬ КЛАСИФІКАЦІЇ НЕЧІТКОЇ ЕКСПЕРТНОЇ СИСТЕМИ**

**Д. І. Єгошкін, Н. А. Гук**

K[nightDanila@i.ua](mailto:KnightDanila@i.ua), [NatalyGuk29@gmail.com](mailto:NatalyGuk29@gmail.com)

*Дніпровський національний університет імені Олеся Гончара*

Задачі, що пов'язані з багатофакторними залежностями, актуальні в різних галузях знань. Для їх розв'язання використовують методи штучного інтелекту, які розкривають приховані зв'язки між великою кількістю вхідних ознак та вихідними результатами. Використання нечіткої логіки та нечітких баз правил у експертних системах дозволяє відобразити ці залежності у вигляді правил "ЯКЩО - ТО", що спрощує обробку та аналіз здобутих знань експертами та вченими.

Існує два основні підходи для здобуття знань: дедуктивний та індуктивний. Дедуктивний зосереджується на формалізації експертних знань в певній області задач та використовує правила, що були надані експертами. В той час як індуктивний підхід дозволяє проводити процес навчання автоматично за допомогою тренувальних наборів даних шляхом виявлення загальних закономірностей.

В роботах [1,2,3] було запропоновано та досліджено декілька методів індуктивного моделювання для генерації моделей з тренувальних вхідних даних без втручання людини. Висока точність згенерованих моделей і наближена кореляція між помилками моделювання та помилками моделі дозволяють припустити, що можливо виявити причинно-наслідкові зв'язки за допомогою навчання нечітких когнітивних карт [1]. Алгоритм навчання на основі Хебба для створення вагових матриць [2] дозволяє наблизити адекватність моделі до прийнятливого рівня точності.

Індуктивний підхід дозволяє автоматизувати процес отримання знань та надає переваги в обробці масивів даних та виявленні прихованих взаємозв'язків. Але залишається актуальним питання впливу розміру навчальної вибірки на якість отриманої експертної системи. В роботі для побудови нечіткої бази правил для класифікації об’єктів за допомогою нечіткої логіки та елементів навчальної вибірки пропонується представлення правил у матричному вигляді та визначення консеквентів правил за допомогою процедури навчання.

Для оцінки якості розв'язання задачі мультикласової класифікації Iris Data Set застосовано метрики: accuracy, precision (точність), recall (повнота) та f1-score (f-міра). Вхідну вибірку було поділено на дві частини: навчальну та тестову. Точність розв'язання задачі для різного розміру вибірок, що застосовувались під час навчання та тестування моделі, наведено у таблиці:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Розмір навчальної вибірки (навчальна/ тестова) | accuracy | precision | recall | f1- score |
| 130/20 | 0,95 | 0,91 | 0,91 | 0,94 |
| 120/30 | 0,94 | 0,88 | 0,89 | 0,92 |
| 100/50 | 0,92 | 0,85 | 0,83 | 0,90 |
| 75/75 | 0,85 | 0,83 | 0,81 | 0,85 |

Додатково здійснювався аналіз репрезентативності навчальної вибірки та діаграми розсіювання класів, проводилось кілька циклів відбору навчальної вибірки. Встановлено, що розмір навчальної вибірки впливає на результат класифікації, а саме зі зменшенням обсягу навчальної вибірки якість класифікації погіршується.

**Бібліографічні посилання:**

1. Ye Chen, Lawrence J. Mazlack, Ali A. Minai, Long J. Lu, Inferring causal networks using fuzzy cognitive maps and evolutionary algorithms with application to gene regulatory network reconstruction, Applied Soft Computing, Volume 37, 2015, Pages 667-679, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2015.08.039>.

2. Papageorgiou, Elpiniki I., and Arthi Kannappan. “Fuzzy Cognitive Map Ensemble Learning Paradigm to Solve Classification Problems: Application to Autism Identification.” Applied Soft Computing 12, no. 12 (2012): 3798–3809. doi:10.1016/J.ASOC.2012.03.064.

3. Wojciech Stach, Lukasz Kurgan, Witold Pedrycz, Marek Reformat, Genetic learning of fuzzy cognitive maps, Fuzzy Sets and Systems, Volume 153, Issue 3, 2005, Pages 371-401, ISSN 0165-0114, https://doi.org/10.1016/j.fss.2005.01.009.