

# НАВЧАННЯ НЕЧІТКОГО КЛАСИФІКАТОРА ЗА ПРИНЦИПОМ ГОЛОВНИХ КОНКУРЕНТІВ

<sup>1</sup>Вінницький національний технічний університет

## Анотація

Останнім часом все більш популярним інструментом прийняття рішень стають нечіткі класифікатори. В статті порівнюються результати комп'ютерних експериментів з навчання нечіткого класифікатора для розпізнавання вин за 5 критеріями.

**Ключові слова:** класифікація, нечітка база знань, навчання, критерії навчання, головні конкуренти.

## Abstract

Fuzzy classifiers become a popular tool for decision making in recent years. The results of computer experiments on learning fuzzy classifier for wine recognition by 5 criteria are compared in this article.

**Key words:** classification, fuzzy knowledge base, learning, learning criteria, main competitors.

## Вступ

Задача класифікації полягає в віднесенні об'єкта за деякими ознаками до одного з наперед визначених класів. Вона здійснюється на основі аналізу атрибутів об'єкта, що класифікується. В нечітких класифікаторах прийняття рішень відбувається за лінгвістичними правилами <Якщо – тоді>, антецеденти яких містять нечіткі терми “низький”, “середній”, “високий” тощо [1]. Кожне правило задає область факторного простору, в межах якої об'єкти належать одному класу. Границі цих областей нечіткі, тому один і той же об'єкт може одночасно належати декільком класам, але з різним ступенем.

Для підвищення безпомилковості нечіткий класифікатор навчають за експериментальними даними. Для цього ітераційно змінюють його параметри, щоб мінімізувати відстань між експериментальними даними та результатами нечіткого виведення. Ця відстань називається критерієм навчання і може визначатись різними способами.

Метою статті є тестування 5 критеріїв навчання нечітких класифікаторів - трьох традиційних з [2, 3] та двох нових, що враховують різницю належностей нечіткого висновку лише до головних конкурентів з [4].

## 1. Нечіткий класифікатор

Нечіткий класифікатор на основі бази нечітких правил являє собою відображення:

$$\mathbf{X} = (x_1, x_2, \dots, x_n) \rightarrow y \in \{l_1, l_2, \dots, l_m\},$$

де  $\mathbf{X} = (x_1, x_2, \dots, x_n)$  – вектор інформативних ознак об'єкта класифікації (вектор вхідних атрибутів);  $l_1, l_2, \dots, l_m$  – класи рішень;  $y$  – результат класифікації.

Ґрунтуючись на [2] база правил нечіткого класифікатора має вигляд:

$$\text{Якщо } (x_1 = \tilde{a}_{1j} \text{ та } x_2 = \tilde{a}_{2j} \text{ та } \dots \text{ та } x_n = \tilde{a}_{nj} \text{ з вагою } w_j), \text{ тоді } y = d_j, \quad j = \overline{1, k}, \quad (1)$$

де  $k$  – кількість правил;  $d_j \in \{l_1, l_2, \dots, l_m\}$  – категоріальне значення консеквента  $j$ -го правила;  $w_j \in [0, 1]$  – ваговий коефіцієнт, який задає достовірність  $j$ -го правила,  $j = \overline{1, k}$ ;  $\tilde{a}_{ij}$  – нечіткий терм, яким оцінюється ознака  $x_i$  в  $j$ -му правилі,  $i = \overline{1, n}$ ,  $j = \overline{1, k}$ .

Класифікація поточного об'єкта з атрибутами  $\mathbf{X}^* = (x_1^*, x_2^*, \dots, x_n^*)$  здійснюється таким чином. Спочатку розраховується ступінь виконання  $j$ -го правила з бази (1):

$$\mu_j(X^*) = w_j \cdot \left( \mu_j(x_1^*) \wedge \mu_j(x_2^*) \wedge \dots \wedge \mu_j(x_n^*) \right), \quad j = \overline{1, k}, \quad (2)$$

де  $\mu_j(x_i^*)$  – ступінь належності значення  $x_i^*$  нечіткому терму  $\tilde{a}_{ij}$ ;  $\wedge$  – t-норма, яку зазвичай реалізують операцією мінімуму або добутком.

Ступінь належності вхідного вектору  $\mathbf{X}^*$  до класів  $l_1, l_2, \dots, l_m$  розраховується так:

$$\mu_{l_s}(y^*) = \frac{\sum_{\forall j: d_j=l_s, j=1, k} \mu_j(X^*)}{\max_{s=1, m} \left( \sum_{\forall j: d_j=l_s, j=1, k} \mu_j(X^*) \right)}, \quad s = \overline{1, m}. \quad (3)$$

Нечітким рішенням задачі класифікації буде нечітка множина

$$\tilde{y}^* = \left( \frac{\mu_{l_1}(y^*)}{l_1}, \frac{\mu_{l_2}(y^*)}{l_2}, \dots, \frac{\mu_{l_m}(y^*)}{l_m} \right). \quad (4)$$

Результатом виведення оберемо ядро нечіткої множини (4), тобто клас з максимальною сумою ступенів належності:

$$y^* = \arg \max_{\{l_1, l_2, \dots, l_m\}} \left( \mu_{l_s}(y^*) \right).$$

Можлива ситуація, коли в ядро нечіткої множини (4) входять кілька елементів. Тоді об'єкт одночасно належить кільком класам з однаковими ступенями, значення яких дорівнює  $\max_{s=1, m} \left( \mu_{l_s}(y^*) \right)$ . Для вибору

одного з цих конкурентних класів застосуємо схему на основі єдиного правила-переможця. За нею серед правил, що стосуються цих конкурентних класів, виберемо правило з максимальним ступенем виконання.

## 2. Критерії навчання нечіткого класифікатора

Навчальна вибірка з  $M$  пар “входи – вихід” має вигляд:

$$(\mathbf{X}_r, y_r), \quad r = \overline{1, M}, \quad (5)$$

де  $\mathbf{X}_r = (x_{r1}, x_{r2}, \dots, x_{rm})$  - вхідні атрибути  $r$ -го об'єкту;  $y_r \in \{l_1, l_2, \dots, l_m\}$  - клас  $r$ -го об'єкту.

Введемо такі позначення:

$\mathbf{P}$  – вектор параметрів функцій належності термів з бази знань (1);  $\mathbf{W}$  – вектор вагових коефіцієнтів правил бази знань (1);  $F(\mathbf{K}, \mathbf{X}_r) \in \{l_1, l_2, \dots, l_m\}$  – результат класифікації за базою знань з параметрами  $\mathbf{K} = (\mathbf{P}, \mathbf{W})$  для вхідного вектору  $\mathbf{X}_r$  з  $r$ -го рядка вибірки (5).

Навчання нечіткого класифікатора полягає в знаходженні такого вектору  $\mathbf{K}$ , який мінімізує частоту помилок класифікації на тестовій вибірці. При цьому для настроювання параметрів  $\mathbf{K}$  використовується лише навчальна вибірка (5). Навчання розглядається як задача оптимізації. Вона полягає в знаходженні таких керованих змінних  $\mathbf{K}$ , які мінімізують відстань між результатами логічного виведення та експериментальними даними з вибірки (5).

**Критерій 1** – частота помилок класифікації [3]:

$$Crit_1 = \frac{1}{M} \sum_{r=1, M} \Delta_r(\mathbf{K}), \quad (6)$$

$$\Delta_r(\mathbf{K}) = \begin{cases} 1, & \text{якщо } y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & \text{якщо } y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}.$$

Переваги критерію полягають в його простоті та ясній змістовній інтерпретації. Але цільова функція в задачі оптимізації за цим критерієм приймає дискретні значення, що ускладнює застосування швидких градієнтних методів оптимізації, особливо за малих вибірок даних.

**Критерій 2** – квадратична нев’язка між двома нечіткими множинами – бажаними та реальними результатами класифікації [2]. Для її розрахунку значення вихідної змінної  $y$  в навчальній вибірці фазифікують таким чином:

$$\left. \begin{aligned} \tilde{y} &= \left( \frac{1}{l_1}, \frac{0}{l_2}, \dots, \frac{0}{l_m} \right), & \text{якщо } y = l_1 \\ \tilde{y} &= \left( \frac{0}{l_1}, \frac{1}{l_2}, \dots, \frac{0}{l_m} \right), & \text{якщо } y = l_2 \\ &\vdots \\ \tilde{y} &= \left( \frac{0}{l_1}, \frac{0}{l_2}, \dots, \frac{1}{l_m} \right), & \text{якщо } y = l_m \end{aligned} \right\}. \quad (7)$$

Критерій навчання враховує відстань між логічним висновком у формі нечіткої множини (4) та бажаним нечітким значенням вихідної змінної (7):

$$Crit_2 = \sum_{r=1, M} D_r(\mathbf{K}), \quad (8)$$

де  $D_r(\mathbf{K})$  – відстань між бажаною та дійсною вихідними нечіткими множинами при класифікації  $r$ -го об’єкту з навчальної вибірки (5).

Для розрахунку  $D_r(\mathbf{K})$  використовується евклідова метрика:

$$D_r(\mathbf{K}) = \sum_{s=1, m} (\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r))^2, \quad (9)$$

де  $\mu_{l_s}(y_r)$  – ступінь належності  $r$ -го об’єкту навчальної вибірки до класу  $l_s$  згідно до (7);  $\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$  – розрахований за формулою (3) ступінь належності висновку за нечіткою моделлю з параметрами  $\mathbf{K}$  до класу  $l_s$  для вхідного вектору  $\mathbf{X}_r$ .

Перевага критерію  $Crit_2$  полягає в урахуванні міри впевненості в прийнятому рішенні на основі ступенів належності об’єкту різним класам. В критерії  $Crit_1$  вважається, що результат класифікації об’єкту є абсолютно достовірним, тобто неважливо наскільки ступінь належності у рішення більший, ніж у інших альтернатив – на 0.0001 або на 1. Крім того, цільова функція в задачі навчання за критерієм (8) не має довгих плато, тому вона придатна до оптимізації градієнтними методами. Але близькі до границь класів об’єкти вносять майже однаковий вклад в критерій навчання (8) як за правильної, так і за помилкової класифікації, тому навчання може бути нерезультативним.

**Критерій 3** – квадратична нев’язка між нечіткими бажаними та реальними результатами класифікації з додатковим штрафом за помилкове рішення [5 - 7]. Цей критерій успадковує переваги двох попередніх. Ідея полягає в збільшенні відстані  $D$  для помилково класифікованих об’єктів:

$$Crit_3 = \sum_{r=1, M} (\Delta_r(\mathbf{K}) \cdot p + 1) \cdot D_r(\mathbf{K}), \quad (10)$$

де  $p > 0$  – штрафний коефіцієнт.

**Критерій 4** – відстань між головними конкурентами з штрафом за помилкове рішення. Ідея цього критерію полягає у врахуванні різниці належностей нечіткого висновку лише до головних конкурентів. За алгоритмом логічного виведення рішенням обирається клас з максимальним ступенем належності. Клас-переможець позначимо через  $win$  та присвоїмо йому перший ранг. У випадку правильної класифікації головним конкурентом прийнятого рішення є  $vicewin$  – клас з другим рангом, тобто клас з другим за величиною ступенем належності. Чим більша різниця між ступенями належності до класів  $win$  та  $vicewin$ , тим більша впевненість у логічному висновку, і тим далі об’єкт знаходиться від границі розділу класів. Позначимо через  $smax$  - операцію знаходження другого за величиною елемента множини. Тоді, для  $r$ -го об’єкту з вибірки (5)  $\mu_{win}(\mathbf{X}_r) = \max_{s=1, m} (\mu_{l_s}(\mathbf{X}_r))$  та  $\mu_{vicewin}(\mathbf{X}_r) = smax_{s=1, m} (\mu_{l_s}(\mathbf{X}_r))$ . Відповідно, різниця між головними конкурентами дорівнює  $\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)$ .

В критерії навчання враховуватимемо відносні показники, розділивши різницю на ступінь належності класу-переможцю. За правильної класифікації відносна різниця дорівнює  $D_r^1 = \frac{\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$ , а

за неправильної –  $D_r^0 = \frac{\mu_{win}(\mathbf{X}_r) - \mu_{y_r}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$ . Крім того, аналогічно до критерія 3, за помилкової класифікації зважимо різницю штрафним коефіцієнтом. Математично критерій навчання запишемо таким чином:

$$Crit_4 = p \cdot \sum_{r=1, M} \sum_{y_r \neq F(K, X_r)} D_r^0(K) - \sum_{r=1, M} \sum_{y_r = F(K, X_r)} D_r^1(K), \quad (11)$$

де  $p \geq 1$  – штрафний коефіцієнт.

**Критерій 5** – квадратична відстань між головними конкурентами з штрафом за помилкове рішення. Цей критерій є модифікацією попереднього. Відмінність полягає у використанні не абсолютних відстаней, а їх квадратів:

$$Crit_5 = p \cdot \sum_{r=1, M} \sum_{y_r \neq F(K, X_r)} D_r^0(K) - \sum_{r=1, M} \sum_{y_r = F(K, X_r)} D_r^1(K)^2.$$

Піднесення до квадрату в  $Crit_5$  дозволяє збільшити внесок в критерії навчання великих різниць та знецілювати внесок малих різниць.

### 3. Комп'ютерні експерименти

Метою експериментів є визначення критерію навчання, який забезпечує найкращу безпомилковість. Розглядається тестова задача Wine Dataset з UCI Machine Learning Repository. Вона полягає у виявленні сорту винограду ( $y$ ), з якого виготовлено вино. База даних містить результати лабораторних аналізів по 13-ти показникам 178 зразків італійських вин, виготовлених в одному регіоні. Для кожного зразка вказано 1 з трьох сортів винограду, з якого виготовлено вино.

Навчальну вибірку сформуємо з рядків бази даних з граничними значеннями кожного із 13 атрибутів. Додатково в навчальну вибірку включимо всі непарні рядки бази даних. Всі інші дані занесемо в тестову вибірку. В результаті отримаємо навчальну вибірку з 100 рядків і тестову – з 78. Побудуємо нечіткий класифікатор вин за трьома ознаками:  $x_7$  – flavanoids,  $x_{10}$  – color intensity та  $x_{13}$  – proline. Експерименти проведемо для нечіткого класифікатора з базою знань з табл. 1.

Таблиця 1 – Нечітка база знань класифікатора вин

№	$x_7$	$x_{10}$	$x_{13}$	$y$
1	–	–	Високий	Сорт 1
2	Високий	Високий	Середній	Сорт 1
3	–	Низький	Низький	Сорт 2
4	Низький	Низький	Середній	Сорт 2
5	Низький	Високий	–	Сорт 3

Для кожного критерію проведемо 1000 експериментів із навчання нечіткої бази знань на основі квазіньютонівського алгоритму. Після навчання кожний класифікатор перевіримо на тестовій вибірці за частотою помилок (критерій  $Crit_1$ ).

Проведемо 2 серії експериментів. Першу серію для нечіткого класифікатора з реалізацією t-норми мінімумом (min), а другу – з реалізацією t-норми добутком (prod). В експериментах з критеріями  $Crit_3$ ,  $Crit_4$  та  $Crit_5$  спочатку визначимо прийнятний рівень штрафного коефіцієнту. Для цього проведемо по 200 оціночних експериментів для  $p = 1, 3, \dots, 9$ . Результати цих експериментів наведено в табл. 2.

Напівжирним шрифтом виділено за яких значень штрафного коефіцієнту навчання відбувається краще. Саме за таких значень штрафного коефіцієнту проведемо решту експериментів.

Таблиця 2 – Вплив штрафного коефіцієнта в критерії навчання на безпомилковість класифікатора на тестовій вибірці (статистика 200 експериментів)

t-норма	Критерій	Середня безпомилковість				
		$p = 1$	$p = 3$	$p = 5$	$p = 7$	$p = 9$
min	$Crit_3$	<b>0.0856</b>	0.0915	0.0912	0.0910	0.0960
	$Crit_4$	0.0732	<b>0.0640</b>	0.0662	0.0736	0.0770
	$Crit_5$	0.0690	0.0604	<b>0.0602</b>	0.0688	0.0657
prod	$Crit_3$	0.0727	0.0656	0.0665	0.0681	<b>0.0649</b>
	$Crit_4$	0.0565	0.0512	<b>0.0493</b>	0.0499	0.0527
	$Crit_5$	0.0448	<b>0.0443</b>	0.0456	0.0486	0.0446

На рис. 1 наведено результати експериментів, що вказують на корельованість значень критеріїв  $Crit_1 - Crit_5$  на навчальній вибірці з частотою помилок на тестовій вибірці. Відповідно ці критерії можна застосовувати для навчання нечіткого класифікатора. Щодо результативності навчання (табл. 3), то вона суттєво краща при використанні критеріїв  $Crit_4$  та  $Crit_5$ .

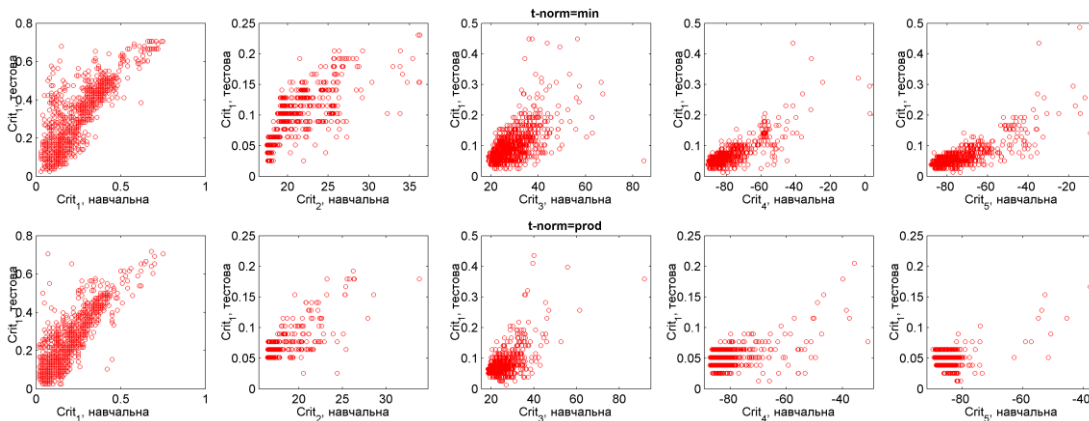


Рис. 1. Розподіл результатів навчання нечіткого класифікатора

Таблиця 3 – Статистика навчання нечітких класифікаторів (напівжирним виділено найкращі результати)

t-норма	Критерій навчання	Частота помилок ( $Crit_1$ ) на тестовій вибірці			
		мінімальне	середнє	медіанне	максимальне
min	$Crit_1$	0.0256	0.3194	0.3333	0.7179
	$Crit_2$	0.0256	0.0900	0.0641	0.4487
	$Crit_3$	0.0256	0.0921	0.0769	0.3462
	$Crit_4$	0.0128	<b>0.0680</b>	0.0513	0.4615
	$Crit_5$	0.0128	<b>0.0680</b>	0.0513	0.4744
prod	$Crit_1$	0.0128	0.2459	0.2308	0.7179
	$Crit_2$	0.0385	0.0631	0.0641	0.1795

	<i>Crit</i> <sub>3</sub>	0.0128	0.0663	0.0641	0.4487
	<i>Crit</i> <sub>4</sub>	0.0128	0.0503	0.0513	0.1667
	<i>Crit</i> <sub>5</sub>	0.0128	<b>0.0454</b>	0.0385	0.2821

### Висновки

Реалізовано навчання нечіткого класифікатора за 5 критеріями, а саме: частота помилок, квадратична нев'язка між двома нечіткими множинами – бажаними та реальними результатами класифікації; квадратична нев'язка між нечіткими бажаними та реальними результатами класифікації з додатковим штрафом за помилкове рішення; відстань між головними конкурентами з штрафом за помилкове рішення; квадратична відстань між головними конкурентами з штрафом за помилкове рішення.

Проведені експерименти із навчання нечіткого класифікатора для UCI-задачі із розпізнавання італійських вин засвідчили, що кращі показники навчання забезпечують критерії - *Crit*<sub>4</sub> та *Crit*<sub>5</sub>, які враховують різниці належностей нечіткого висновку лише до головних конкурентів. Серед них помірну перевагу має критерій *Crit*<sub>5</sub> на основі квадратичної відстані між головними конкурентами з штрафом за помилкове рішення.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Kuncheva L. I. Fuzzy classifier design / Studies in Fuzziness and Soft Computing. Vol. 49. – Berlin – Heidelberg: Springer-Verlag, 2000. – 314 p.
2. Rotshtein A. Design and tuning of fuzzy rule-based system for medical diagnosis. In “Fuzzy and Neuro-Fuzzy Systems in Medicine” (Eds.: Teodorescu N. H., Kandel A., and Jain L. C.). Boca-Raton: CRC-Press, 1998. - P. 243–289.
3. Ishibuchi H. Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining / H. Ishibuchi, T. Nakashima, M. Nii – Berlin – Heidelberg: Springer-Verlag, 2005. – 307 p.
4. Штовба С. Д. Навчання нечіткого класифікатора з урахуванням лише головних конкурентів / С. Д. Штовба, А. В. Галушчак // Вісник Вінницького політехнічного інституту.– 2016. –№1 С. 124-132.
5. Shtovba S. Tuning the fuzzy classification models with various learning criteria: the case of credit data classification / S. Shtovba, O. Pankevich, G. Dounias // Proc. of Inter. Conference on Fuzzy Sets and Soft Computing in Economics and Finance. St. Petersburg (Russia), 2004. Vol. 1. – St. Petersburg: Russian Fuzzy Systems Association, 2004. P. 103–110.
6. Штовба С. Д. Проектирование нечетких систем средствами MATLAB. М.: Горячая линия – Телеком, 2007. – 288 с.
7. Штовба С. Д. Анализ критериев обучения нечеткого классификатора / С. Д. Штовба, О. Д. Панкевич, А. В. Нагорна // Автоматика и вычислительная техника. - 2015. - №3. - С. 5-16.

*Анастасія Володимирівна Галушчак* – асистент кафедри комп’ютерних систем управління, Вінницький національний технічний університет, Вінниця, E-mail: a.v.galushchak@vntu.edu.ua.

Науковий керівник: *Сергій Дмитрович Штовба* – д.т.н., професор, професор кафедри комп’ютерних систем управління, Вінницький національний технічний університет.

*Anastasiia Volodymyrivna Galushchak* – Assistant, Department of Computer Control Systems, Vinnytsia National Technical University, Vinnytsia.

Supervisor: *Serhiy Dmytrovych Shtovba* – Doctor of Sc., Professor, Department of Computer Control Systems, Vinnytsia National Technical University.