

Прогнозування очікуваних результатів на основі інтелектуальної технології ідентифікації

О.М. Левчук

Постановка проблеми та аналіз останніх досліджень. Нині внаслідок швидких змін зовнішнього середовища зростає роль методів прогнозування, кожен з яких вирізняється оцінкою надійності та вживання [1-4]. Оскільки прогнозування має ймовірнісний характер, то пріоритетними в разі його реалізації є статистичні методи прогнозування, які розглядаються такими вченими, як В.Н. Афанасьєв, М.М. Юзбашев, А.А. Френкель, І.Б. Загайтов, Г. Тейл, М. Кендалл та ін. [4-6]. Проте статистичний аналіз не дає змоги працювати з лінгвістичними змінними, тому була запропонована теорія нечітких множин, яка надає можливість працювати з лінгвістичною невизначеністю, а проблема прогнозування очікуваних результатів якраз і повинна базуватися на гнучких технологіях прогнозування, таких як інтелектуальні технології ідентифікації [7-8].

Мета статті. У нижченаведеній методиці пропонується здійснювати прогноз очікуваних результатів на основі нечіткої експертної системи, тому поставлені такі задачі:

- визначити основні вхідні параметри середовища, яке важко формалізується;
- побудувати нечітку експертну систему;

- описати залежність прогнозування очікуваного результату;
- сформувати дерево логічного виводу лінгвістичних змінних;
- визначити функції приналежності лінгвістичним змінним;
- надати прогноз очікуваних результатів, використовуючи базу знань нечіткої експертної системи;
- експериментально апробувати розроблену методику прогнозування очікуваних результатів на прикладі сільського господарства.

Виклад основного матеріалу. Процес побудови нечіткої експертної системи виконується за таким алгоритмом.

1. Визначення характеристик системи.

На цьому етапі збирається інформація про призначення системи, визначається вихідна змінна об'єкта, що ідентифікується.

2. Формування дерева логічного виводу.

Дерево логічного виводу формується шляхом послідовного виконання операцій додавання і/або видалення вузлів. У разі додаванні вузла визначається назва змінної, її позначення, кількість і назви термів для її оцінки, встановлюється діапазон її зміни.

3. Визначення функцій приналежності лінгвістичних термів.

На цьому етапі, який в теорії нечітких множин відповідає етапу фазифікації, визначаються функції приналежності, які використовуються для формалізації термів-оцінок змінних.

4. Формулювання експертних правил «ЯКЩО - ТО», що описують поведінку об'єкта.

Ці правила вносяться до відповідних матриць знань.

5. Налаштування нечіткої експертної системи шляхом вирішення задач оптимізації з використанням навчальної вибірки.

Будь-який прогноз очікуваних результатів визначатимемо за допомогою таких змінних (від нижчого рівня до вищого): d_1 - низький; d_2 - нижче середнього; d_3 - середній; d_4 - вище середнього; d_5 - високий.

Перераховані рівні підлягають розпізнаванню. У разі встановлення рівня для конкретної технології виробництва будь-якого продукту та кліматичних умов до уваги беруть велику кількість різних параметрів, які в тій чи іншій мірі можуть вплинути на прогноз. Параметри позначимо через x_1, \dots, x_n , де $n \in N$.

Відповідно до кожного параметру для його оцінювання ставиться термножина значень. Наприклад, якщо x_1 - середня температура навколишнього середовища в період дослідження, то терми для оцінок даного параметра будуть: низька (1), середня (2), підвищена (3). Параметри x_1, \dots, x_n розглядатимемо як лінгвістичні змінні. Крім того, лінгвістичні змінні можна об'єднувати в групи, класи, тобто розглядати їх як змінні спільної функції. Отримаємо такі вирази:

d - очікуваний прогноз, який вимірюється рівнями $d_1 - d_5$;

$y, z, \alpha, \beta, \gamma$ - функції, що характеризуються:

y - параметрами $\{x_1, x_2, \dots, x_k\}$;

z - параметрами $\{x_{k+1}, \dots, x_l\}$;

α - параметрами $\{x_{l+1}, \dots, x_m\}$;

β - параметрами $\{x_{m+1}, \dots, x_p\}$;

γ – параметрами $\{x_{p+1}, \dots, x_n\}$ і так далі, де $k < l < m < p < n$, $k, l, m, p, n \in \mathbb{N}$.

$$d = f_d(y, z, \alpha, \beta, \gamma, \dots), \quad y = f_y(x_1, x_2, \dots, x_k), \quad z = f_z(x_{k+1}, \dots, x_l), \quad (1)-(3)$$

$$\alpha = f_\alpha(x_{l+1}, \dots, x_m), \quad \beta = f_\beta(x_{m+1}, \dots, x_p), \quad \gamma = f_\gamma(x_{p+1}, \dots, x_n) \quad (4)-(6)$$

Для оцінки значень лінгвістичних змінних x_1, \dots, x_n , а також $y, z, \alpha, \beta, \gamma$ та інших використовуватимемо шкали якісних термів. Кожен з термів є нечіткою множиною, заданою за допомогою відповідної функції приналежності.

Для співвідношень (1)-(6) складається таблиця знань

Таблиця 1

Знання про співвідношення (1)

y	z	α	β	γ	...	d
r_{11}	r_{12}	r_{13}	r_{14}	...	r_{1j}^*	d_1
r_{21}	r_{22}	r_{2j}	
...	d_2
...	d_3
...	d_4
...	d_5
r_{i1}	r_{i2}	r_{i3}	r_{i4}	...	r_{ij}	

* $r_{ij} \in T$, T - множина термів така, що може виконуватись умова $r_{ij} = r_{ji}$ при $\forall i, j \in \mathbb{N}$, \mathbb{N} - множина натуральних чисел. Тобто терми для $y = f_y(x_1, x_2, \dots, x_k)$ можуть (не обов'язково) бути однаковими як для d_1 , так і для d_2, d_3 і т.д.

Аналогічно формуються таблиці знань про співвідношення (2) -(6).

Використовуючи дані таблиць знань про співвідношення (1) -(6) та операції (I-min) і (Або-max) запишемо систему нечітких логічних рівнянь, які зв'язують

функції очікуваних прогнозів і вхідних змінних:

$$\mu^{\alpha_q}(d) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(y) \cdot \mu^{r_{ij}}(z) \cdot \mu^{r_{ij}}(\alpha) \cdot \mu^{r_{ij}}(\beta) \cdot \mu^{r_{ij}}(\gamma) \cdot \dots \right), \quad (7)$$

$$\mu^{\alpha_t}(y) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(x_1) \cdot \dots \cdot \mu^{r_{ij}}(x_k) \right), \quad \mu^{\alpha_t}(z) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(x_{k+1}) \cdot \dots \cdot \mu^{r_{ij}}(x_l) \right), \quad (8)-(9)$$

$$\mu^{\alpha_t}(\alpha) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(x_{l+1}) \cdot \dots \cdot \mu^{r_{ij}}(x_m) \right), \quad \mu^{\alpha_t}(\beta) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(x_{m+1}) \cdot \dots \cdot \mu^{r_{ij}}(x_p) \right), \quad (10)-(11)$$

$$\mu^{\alpha_t}(\gamma) = \bigvee_i \prod_j \left(\mu^{r_{ij}}(x_{p+1}) \cdot \dots \cdot \mu^{r_{ij}}(x_n) \right), \quad (12)$$

де $q = \overline{1,5}$;

α_t – терми, $\alpha_t \in T$;

$r_{ij} \in T$, $i, j \in \mathbb{N}$;

r_{ij} у співвідношенні (7) – терми для оцінок очікуваного прогнозу d , у співвідношенні (8) – для оцінок лінгвістичних змінних $\{x_1, x_2, \dots, x_k\}$ і т.д.

У загальному випадку частина змінних x_1, x_2, \dots може мати свої функції приналежності нечітким термам $(r_{11}, r_{12}, \dots, r_{1e})$, які використовуються в наведених вище логічних рівняннях. Для цих змінних наведемо інтервали зміни кожній змінній до універсального інтервалу $[0; e-1]$ за допомогою таких співвідношень:

$$\mu^j(x_i) = \tilde{\mu}^j(u), \quad u = (e-1) \frac{x_i - \underline{x}_i}{\underline{x}_i - \overline{x}_i}, \quad j=r_{11}, r_{12}, \dots, r_{1e}, \quad (13)$$

де $[\underline{x}_i - \overline{x}_i]$ – інтервал зміни змінних x_i ;

\underline{x}_i та \overline{x}_i – відповідно найменше та найбільше значення інтервалу.

Аналітична модель функції приналежності має вигляд:

$$\tilde{\mu}^j(u) = \frac{1}{1 + \left(\frac{u-b}{c}\right)^2}, \quad (14)$$

де c задається експертом, b і параметри функції приналежності (14) нижче:

Терм	r_{11}	r_{12}	...	r_{1e}
b	0	1	...	$e-1$
c	0,923	0,923	0,923	0,923

Наведемо алгоритм, за яким нечіткі логічні рівняння разом з функціями приналежності нечітких термів дозволяють обчислити очікуваний прогноз:

1. Фіксуються значення параметрів змінних (дивись табл. 1), що визначають прогноз урожайності: $x^* \left(x_1^*, x_2^*, x_3^*, \dots, x_{12}^* \right)$.

2. Використовуючи модель (14) і параметри b і c для відповідних співвідношень (13), визначаються функції приналежності $\tilde{\mu}^j(x_i^*)$ за фіксованих параметрів $x_i^* \left(i = \overline{1, n} \right)$.

3. Використовуючи логічні рівняння, обчислюють значення функцій приналежності $\mu^{d_j} \left(x_1^*, x_2^*, x_3^*, \dots, x_n^* \right)$ за умови, що вектор стану $x^* = \left(x_1^*, x_2^*, x_3^*, \dots, x_n^* \right)$ для прогнозів d_1, d_2, d_3, d_4, d_5 .

4. Визначається рішення, для якого

$$\mu^{d_j} \left(x_1, x_2, x_3, \dots, x_n \right) = \max_{j=1,5} \left[\mu^{d_j} \left(x_1^*, x_2^*, x_3^*, \dots, x_n^* \right) \right] \quad (15)$$

Прогнозування очікуваного результату d здійснюється з використанням узагальненого елемента нечіткого логічного виводу, який описує залежність

$d = f(y, z, \alpha, \beta, \gamma, \dots)$ за допомогою системи висловів «ЯКЩО – ТО». Послідовне застосування цієї моделі до дерева виводу дає змогу обчислити прогноз у заданій точці простору чинника:

$$d = (y, z, \alpha, \beta, \gamma) = \left[\frac{\mu^{d_1}(d)}{d_1} + \frac{\mu^{d_2}(d)}{d_2} + \dots + \frac{\mu^{d_5}(d)}{d_5} \right],$$

де l – кількість непарних термів для оцінки змінної;

d_i – назва i -го терму, $i = \overline{1, l}$;

$\mu^{d_i}(d)$ – ступінь приналежності змінної d для терму d_i .

Практична реалізація. Прогноз очікуваних результатів розглянемо на прикладі врожайності сільськогосподарських культур. Введемо лінгвістичні змінні і відповідні їм терм-множини (табл. 2).

Лінгвістичні змінні

Таблиця 2

Позначення і назва змінної	Інтервали можливих значень	Терми для оцінок
Потенційна (природна) родючість ґрунту – x_1	(5-50) у.о.	Низька (1), нижче середнього, середня (3), вище середнього, висока (5)
Внесення органічних добрив – x_2	(0-10) т/га	Не вносилися (1), чверть норми (2), половина норми (3), норма (4)
Внесення мінеральних добрив – x_3	(0-50) кг/га	Не вносилися (1), чверть норми (2), половина норми (3), норма (4)
Середній вміст вологи в ґрунті в I періоді вегетації – x_4	(10-200) мл/дм ³	Низький (1), нижче середнього, середній (3), вище середнього, високий (5)
Середній вміст вологи в ґрунті в II періоді вегетації – x_5	(10-200) мл/дм ³	Низький (1), нижче середнього, середній (3), вище середнього, високий (5)
Середній вміст вологи в ґрунті в III періоді вегетації – x_6	(10-200) мл/дм ³	Низький (1), нижче середнього, середній (3), вище середнього, високий (5)

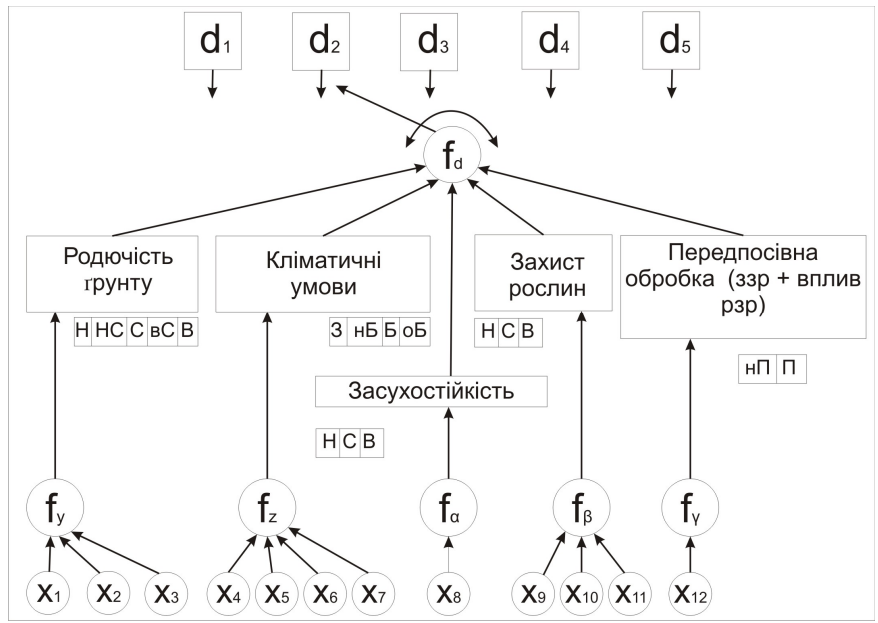
Середня температура навколишнього середовища в період зростання і дозрівання сільськогосподарської культури – x_7	(15-25) °С	Низька (1), середня (2), підвищена (3)
Засухостійкість с/г культури – x_8	(5-50) у.о.	Низька (1), середня (2), висока (3)
Ефективність боротьби з хворобами с/г культури – x_9	(10-40) у.о.	Не проводилася (1), низька (2), середня (3), висока (4)
Ефективність боротьби з шкідниками с/г культури – x_{10}	(10-40) у.о.	Не проводилася (1), низька (2), середня (3), висока (4)
Ефективність боротьби з бур'янами – x_{11}	(10-20) у.о.	Не проводилася (1), низька (2), середня (3), висока (4)
Передпосівна обробка насіння (знезараження насіння + дія регулювальником зростання рослин) – x_{12}	(0-10) у.о.	Не проводилася (1), проводилася (2)
Врожайність с/г культури – $d_1 - d_5$	(10-50) ц/га	Низька (d_1), нижче середнього (d_2), середня (d_3), вище середнього (d_4), висока (d_5).

Структура моделі для диференційного прогнозу врожайності показана на рисунку у вигляді дерева логічного виводу, що відповідає співвідношенням (1)-(6):

$$d = f_d(y, z, \alpha, \beta, \gamma), \quad y = f_y(x_1, x_2, x_3), \quad z = f_z(x_4, x_5, x_6, x_7), \quad \alpha = f_\alpha(x_8),$$

$$\beta = f_\beta(x_9, x_{10}, x_{11}), \quad \gamma = f_\gamma(x_{12}). \quad (15)$$

Для співвідношень (15) складаються таблиці знань (Табл. 3, 4).



Дерево логічного виводу

Таблиця 3

Знання про співвідношення

y	z	α	β	γ	d
Н	нБ	В	С	П	d_1
Н	нБ	С	В	П	
НС	нБ	С	Н	П	
С	нБ	С	нП	П	
вС	З	В	В	П	
В	З	С	С	П	
НС	нБ	В	С	П	d_2
НС	Б	С	Н	П	
С	нБ	С	С	П	
В	З	В	В	П	
С	нБ	В	В	П	d_3
С	Б	С	С	П	
вС	нБ	Н	В	П	
вС	нБ	С	нП	П	
В	нБ	Н	С	П	
вС	нБ	В	В	П	d_4
В	Б	С	В	П	
вС	ОБ	Н	Н	нП	
вС	ОБ	С	нП	П	
В	нБ	В	С	нП	
вС	ОБ	С	В	нП	d_5
В	Б	В	В	П	
В	ОБ	С	В	П	
В	ОБ	Н	Н	П	

Знання про співвідношення

x_1	x_2	x_3	$У$
Н	НВ	НВ	H
Н	чН	НВ	
Н	НВ	чН	
НС	НВ	НВ	HC
Н	Н	НВ	
Н	НВ	Н	
Н	ПН	ПН	C
С	НВ	НВ	
НС	Н	НВ	
НС	НВ	Н	
НС	ПН	ПН	BC
вС	НВ	НВ	
С	Н	НВ	
С	НВ	Н	
С	ПН	ПН	B
В	НВ	НВ	
ВС	Н	НВ	
ВС	НВ	Н	
ВС	ПН	ПН	

Аналогічно будуються таблиці для вихідних і лінгвістичних змінних $z, \alpha, \beta, \gamma, x_4 - x_{12}$. Варто відзначити, що терми для відповідних вихідних змінних визначаються експертним шляхом, щоб виключити велику кількість перестановок і пов'язаних з ними обчислень.

Далі, використовуючи наведені таблиці знань про співвідношення і вирази (7)-(12), запишемо систему нечітких логічних рівнянь, що зв'язують функції прогнозів урожайності та вхідних змінних:

$$\begin{aligned} \mu^{\alpha_1}(d) = & \left[\mu^H(y) \cdot \mu^{НВ}(z) \cdot \mu^B(\alpha) \cdot \mu^C(\beta) \cdot \mu^\Pi(\gamma) \right] \vee \\ & \vee \left[\mu^H(y) \cdot \mu^{НВ}(z) \cdot \mu^C(\alpha) \cdot \mu^B(\beta) \cdot \mu^\Pi(\gamma) \right] \vee \left[\mu^{HC}(y) \cdot \mu^{НВ}(z) \cdot \mu^C(\alpha) \cdot \mu^H(\beta) \cdot \mu^\Pi(\gamma) \right] \vee \\ & \vee \left[\mu^C(y) \cdot \mu^{НВ}(z) \cdot \mu^C(\alpha) \cdot \mu^H(\beta) \cdot \mu^\Pi(\gamma) \right] \vee \left[\mu^{вC}(y) \cdot \mu^3(z) \cdot \mu^B(\alpha) \cdot \mu^B(\beta) \cdot \mu^\Pi(\gamma) \right] \vee \\ & \vee \left[\mu^B(y) \cdot \mu^3(z) \cdot \mu^C(\alpha) \cdot \mu^C(\beta) \cdot \mu^\Pi(\gamma) \right]. \end{aligned}$$

Аналогічно описуємо $\mu^{\alpha_2}(d)$, $\mu^{\alpha_3}(d)$, $\mu^{\alpha_4}(d)$, $\mu^{\alpha_5}(d)$, $\mu^{HC}(y)$, $\mu^C(y)$, $\mu^{eC}(y)$, $\mu^B(y)$, $\mu^Z(z)$, $\mu^{HB}(z)$, $\mu^B(z)$, $\mu^{OB}(z)$, $\mu^H(\beta)$, $\mu^C(\beta)$, $\mu^B(\beta)$, $\mu^H(\alpha)$, $\mu^B(\alpha)$, $\mu^C(\alpha)$, $\mu^{\Pi}(\gamma)$, $\mu^{H\Pi}(\gamma)$.

Для змінних x_1, x_4, x_5, x_6 зведемо інтервали зміни кожної змінної до універсального інтервалу $[0;4]$ за допомогою таких співвідношень:

$$\mu^j(x_i) = \tilde{\mu}^j(u), \quad u = 4 \frac{x_i - \underline{x}_i}{\overline{x}_i - \underline{x}_i}, \quad \text{де } j = H, HC, C, eC, B, \quad [\underline{x}_i, \overline{x}_i] - \text{інтервал зміни}$$

змінних x_1, x_4, x_5, x_6 .

Змінні $x_2, x_3, x_9, x_{10}, x_{11}$ мають нечіткі терми (НВ(НП); чН(Н); ПН(С); Н(В)).

Для цих змінних зведемо інтервал зміни кожної змінної до універсального

інтервалу $[0;3]$ за допомогою співвідношення $u = 3 \frac{x_i - \underline{x}_i}{\overline{x}_i - \underline{x}_i}$.

Терм – множини нечітких змінних x_7 і $x(H, C, B)$. Для цих змінних зведемо інтервал зміни кожної змінної до універсального інтервалу $[0;2]$ за допомогою

співвідношення $u = 2 \frac{x_i - \underline{x}_i}{\overline{x}_i - \underline{x}_i}$, $j = \text{НП, П}$.

Терм – множина змінної $x_{12}(\text{НП, П})$. Інтервал зміни цієї змінної $(0,1)$ дорівнює універсальному інтервалу $[0;1]$. Співвідношення (12) набирає вигляду:

$$u = \frac{x_i - \underline{x}_i}{\overline{x}_i - \underline{x}_i}, \quad j = \text{НП, Н}$$

Далі, дотримуючись алгоритму прийняття рішення, знаходимо

дефазифікацію нечіткої множини за принципом центру тяжіння, яка дає оцінку d^* прогнозу врожайності з урахуванням заданих значень вхідних змінних:

$$d^* = f(y^*, z^*, \alpha^*, \beta^*, \gamma^*) = \frac{\sum_{i=1}^l \left[\underline{d} + (i-1) \frac{\bar{d} - \underline{d}}{l-1} \right] * \mu^{d_i}}{\sum_{i=1}^l \mu^{d_i}(d)}.$$

Висновки. Реалізація даної методики дає змогу вирішити низку задач, проаналізувати прогноз очікуваних результатів, отриманих за допомогою вищеписаних рівнянь, з середньостатистичними даними. Практична реалізація прогнозу на основі інтелектуальної інформаційної технології ідентифікації здійснюється під час вирішення конкретних завдань прикладної спрямованості для таких видів діяльності, як сільське господарство, спорт вищих досягнень та інше. Наша методика потребує подальших досліджень.