ПОБУДОВА ФУНКЦІЙ НАЛЕЖНОСТІ ПРИ ФОРМАЛІЗАЦІЇ ХАРАКТЕРИСТИК ТЕРМОГРАМ МОЛОЧНИХ ЗАЛОЗ

УДК 004.896

ТІТОВА Анастасія Юріївна
аспірант кафедри комп'ютерних наук Донецького національного технічного університету м. Покровськ,
Наукові інтереси: інформаційні технології, нечітке моделювання, моделі та методи обробки даних та зображень.
e-mail: a.titova.wk@gmail.com

ВСТУП

Проведення кількісного та візуального аналізу характеристик термограм виконується в контактній термографії при діагностиці онкологічних захворювань молочної залози (МЗ) жінок. Для пошуку інформативно-значущих характеристик теплового зображення розробляють та досліджують нові інформаційні системи діагностики раку МЗ, в тому числі, з використанням засобів нечіткої логіки. Для застосування нечіткого моделювання необхідно вирішити задачу побудови функції належності термів лінгвістичних змінних, що описують предметну область. Сьогодні існує багато методів побудови функцій належності, що застосовуються при вирішенні поставлених задач [1, 2].

Методи побудови функції належності на основі експертних оцінок поділяються на прямі та непрямі. Запропонована програма для побудови функції належності множин з використанням методів статистичних даних, парних рівнянь та прямого групового методу, де для визначення параметрів функцій належності розраховуються максимальні значення та вагові коефіцієнти матриці експертних оцінок [3]. Для універсальної нечіткої множини великої розмірності реалізовані непрямі методи побудови функції належності, методи парних порівнянь та Ягера [4, 5, 6]. Виконано аналіз різних видів функції належності на чутливість при оцінюванні якості та управління [7]. В роботах [8, 9] реалізовано метод побудови функцій належності нечітких множин, який дозволяє обробляти наближені точкові та інтервальні експертні оцінки для забезпечення адекватних значень параметрів функцій належності. Для побудови функцій належності за розподілом експериментальних даних невеликих обсягів використовуються методи нечіткої кластеризації та метод потенціалів [10].

Аналіз літератури показав, що існуючі методи побудови функцій належності не в повній мірі охоплюють специфіку предметної області. В нечіткій моделі аналізу термограм формалізовано лінгвістичні змінні та за допомогою експертів визначено множини їх термів. Для побудови функцій належності термів необхідно використати експертні оцінки набору термограм, що вимагає розробки нових підходів для вирішення цієї задачі.

ПОСТАНОВКА ЗАДАЧІ

Розробка методу побудови функції належності при формалізації характеристик

термограм в процесі діагностики раку МЗ є **метою дослідження**.

Для досягнення поставленої мети слід вирішити наступні задачі:

- постановка задачі нечіткого моделювання при діагностиці онкологічних захворювань МЗ;
- розробка етапів побудови функцій належності термів лінгвістичних змінних нечіткої моделі аналізу термограм;
- оцінка адекватності вибору типу функцій належності та розрахунку їх параметрів.

Вирішення задачі

Необхідно розробити нечітку модель аналізу характеристик термограм, що основана на функціональній залежності виду:

$$f : \overline{X} = \{x_1, x_2, ...x_{12}\} \to Y, . \quad (1)$$

де $\overline{X}$ – вектор вхідних лінгвістичних змінних, які містять діагностичні ознаки захворювань МЗ: x_1 – біль в МЗ, терми якої: «немає», «слабка», «рідко», «локалізована», «періодична», «сильна», «дуже сильна»; x_2 – температура тіла пацієнта з термами: «норма», «підвищена», «висока»; x_3 – зміни шкіри МЗ, терми якої: «відсутня», «локальне почервоніння», «почервоніння», «велике почервоніння», «почервоніння та лущення»; x_4 – виділення з сосків МЗ, де терми: «нормального характеру», «зеленуваті», «кров'яні», «жовті»; x_5 – форма новоутворення з термами: «не пальпується», «вузлова», «неправильна», «гладка кулька», «значна та неправильна», «майданчик»; x_6 – середня різниця температур МЗ, термами якої: «в межах норми», «ледь вище норми», «вище норми», «дуже вище норми», «перевищення норми», «велике перевищення норми»; x_7 – різниця температур симетричних позицій, терми котрої: «граничні», «ледь вище граничних», «вище граничних», «дуже вище граничних», «перевищення граничних», «максимальне перевищення»; x_8 – максимальне локальне перевищення температури правої МЗ, де терми: «нормальне», «незначне», «незначно більше», «більше», «значно більше»; x_9 – максимальне локальне перевищення температури лівої МЗ, з термами: «нормальне», «незначне», «незначно більше», «більше», «значно більше»; x_{10} – максимальна локальна асиметрія: вимірюються в ˚С, її терми: «не зазначена», «дуже маленька», «маленька», «середня», «велика», «дуже велика»; x_{11} – площа гіпертермії, має значення кількості пікселів області гіпертермії, де терми: «межова», «більше межової», «більша», «найбільша», «менш максимальна», «максимальна»; x_{12} – колір гіпертермії з термами: «зелений», «жовтий», «помаранчевий», «червоний».

Y – вихідна змінна, значення якої відповідає конкретному захворюванню МЗ, що містить наступні терми: «Нормотермограма», «Вузлова форма раку», «Дифузна форма раку», «Кіста», «Мастопатія», «рак Паджета», «Мастиподібна форма раку», «Мастит».

Вихідними даними для формування вхідного вектора є відомості про стан здоров'я пацієнта, результати розрахунку кількісних та візуальних характеристик термограми.

Для побудови функції належності термів лінгвістичних змінних нечіткої моделі аналізу термограм необхідно виконати наступні етапи:

- оцінка експертами набору термограм, під час якої необхідно для кожної термограми обрати відповідні значення з терм-множин лінгвістичних змінних;
- розрахунок значень кількісних змінних для кожної термограми, що відповідають лінгвістичним змінним;

- формування вибірки значень кількісних даних для кожної терм-множини відповідної лінгвістичної змінної по набору термограм;
- побудова таблично-заданої функції належності для отриманої вибірки даних терм-множини методом потенціалів;
- апроксимація таблично-заданої функції належності терм-множини класичними функціями належності;
- оцінка точності апроксимації для кожного виду функцій належності;
- вибір функції належності, що забезпечує мінімальне значення критерію апроксимації.

Вихідними даними для побудови функцій належності є набір термограм, відібраний експертами. Експерти оцінюють кожну термограму та визначають значення лінгвістичних змінних, обираючи по одному терму з їх терм-множин. Для кожної термограми з набору розраховується значення кількісних змінних, що відповідають лінгвістичним змінним. До вибірки значень кількісної змінної для конкретного терму обираються ті значення з термограм, для яких більшість експертів визначило цей терм в якості значення лінгвістичної змінної.

Розглянемо формування вибірок даних по 51 термограмі з відомими діагнозами. Дванадцять експертів повинні оцінити реальні діагностичні ознаки за відповідними терм-множинами лінгвістичних змінних.

Для вхідних лінгвістичних змінних x_6 з термами: MN – «в межах норми», LVN – «ледь вище норми», VN – «вище норми», DVN – «дуже вище норми», PN – «перевищення норми», VPN – «велике перевищення норми» та x_7 з термами: G – «граничні», LVG – «ледь вище граничних», VG – «вище граничних», DVG – «дуже вище граничних», PG – «перевищення граничних», MP – «максимальне перевищення» експерти оцінюють статистичні дані термограм. Потім вибираються максимально схожі оцінки та дані з термограм заносяться в вибірки для кожної терм-множини середньої різниці температур M3 та різниці температур симетричних позицій, що представлені в таблиці 1.

Таблиця 1 –

Вибірки даних за термограмами для терм-множин лінгвістичних змінних x_6 та x_7

Змінна	Терм-множина	Дані за термограмами	min	max
x_6	MN	0,1; 0,2; 0,3; 0,2; 0,5	0.1	0.5
	LVN	0,2; 0,6; 0,8; 0,6; 0,3; 0; 0,6; 0,7; 0,5	0	0.8
	VN	0,5; 0,6; 0,6; 0,5; 0,8; 0,8; 1,1; 0,5; 0,6; 1,3	0.5	1.3
	DVN	0,7; 1,4; 1,3; 1,8; 1,0; 1,3	0.7	1.8
	PN	1,6; 2,1; 0,6; 2,0; 2,4; 2,0; 2,1; 2,4; 2,9; 2,1; 3	0.6	3
	VP	2,4; 1,6; 2,1; 2,4; 1,8; 2,9; 3,0; 2,4; 2,9; 3,4	1.6	3.4
x_7	G	0,6; 0; 0,8; 0,2; 0,6	0	0,8
	LVG	1,4; 1,1; 0,3;1,8;1,8; 1,1;0,8;0,3	0,3	1,8
	VG	0,8; 1,1;2; 1,9; 1,3;1,8;2; 2	0,8	2
	DVG	2,7; 2,4; 0,5;2,9;1;2,6; 1,5;0,6;2,4	0,5	2,9
	PG	1,1; 2; 4,5;4,3;1,9;4,4;	0,5	4,5
	MP	1,1; 1,4; 1,5; 3,5; 1,9; 4,6; 3,6; 3,2; 2,9;1,1; 2,3;4,6;4	1,1	4,6

Також в таблиці 1 представлені максимальний та мінімальний елементи відповідних вибірок, що використовуються для подальшої обробки.

Для вибору виду функцій належності, що максимально буде відповідати реальним симптомам захворювань МЗ та статистичним характеристикам термограм, необхідно побудувати таблично-задані функції належності на основі методу потенціалів.

При побудові таблично-заданих функцій належності використовується поняття потенціалу точки – це число, яке показує, наскільки щільно в її околиці розташовані експериментальні дані за термограмами. Чим вище потенціал точки, тим ближче вона до центральної точки (потенціалу). Потенціал точки розраховується таким чином [10]:

$$pot_i = \sum_{j=\overline{1,k}} \exp\left(-4\beta^2 (u_i - u_j)^2\right), \quad (2)$$

де pot_i – потенціал точки u_i, $\beta \succ 0$ – коефіцієнт, що задає ступінь компактності вибірки даних, k– кількість точок експериментальних даних u_j.

Для розрахунку значень таблично-заданих функції належності експериментальні дані необхідно масштабувати на відрізок [0,1]. Ступені належності даних нечіткій множині слід розрахувати через потенціали наступним чином:

$$\mu_i(u_i) = \frac{pot_i}{\max\limits_{j=\overline{1,k}} (pot_j)}, \quad (3)$$

де $\mu_i(u_i)$ – значення таблично-заданої функції належності.

Після побудови таблично-заданих функцій належності, необхідно провести їх апроксимацію класичними функціями належностями: трикутною, трапецієподібною, симетричною функцією Гауса для вибору виду функцій належності кожної терм-множини вхідних лінгвістичних змінних.

В аналітичному вигляді трикутну функцію належності нечіткої терм-множини можна представити наступним чином:

$$\mu^f(u) = \begin{Bmatrix} 0, u < a, \\ \dfrac{u-a}{b-a}, a \le u \le b, \\ \dfrac{c-u}{c-b}, b \le u \le c, \\ 0, u > c. \end{Bmatrix}, \quad (4)$$

де $\mu^f(u)$ – функція належності, a та c – параметри функції належності, що задають основу трикутника, b – параметр функції належності, що відповідає координаті максимуму функції та вершині трикутника.

Для визначення значень параметрів трикутної функцій належності слід використати формули:

$$\begin{cases} a = \min(u_i), \forall i = \overline{1,k} \\ c = \max(u_i), \forall i = \overline{1,k} \\ b = \min\limits_{pot_i=1} (u_i). \end{cases}, \quad (5)$$

Наступну трапецієподібну функцію належності нечіткої терм-множини можна представити в такому аналітичному вигляді:

$$\mu^f(u) = \begin{Bmatrix} 0, u \le a, \\ \dfrac{u-a}{b-a}, a \le u \le b, \\ 1, b \le u \le c, \\ \dfrac{d-u}{d-c}, c \le u \le d, \\ 0, u \ge d. \end{Bmatrix}, \quad (6)$$

де $\mu^f(u)$ – функція належності; a та d – параметри функції належності, що задають нижню основу трапеції та песимістичну оцінку значень лінгвістичної змінної, b та c – параметри функції належності, що відповідають верхній основі трапеції

та оптимістичній оцінці значень лінгвістичної змінної.

При побудові трапецієподібної функції належності слід розрахувати параметри наступним чином:

$$\begin{cases} a = \min\,(u_i), \\ b = \min\limits_{pot_i=1}\,(u_i), \\ c = \max\limits_{pot_i=1}\,(u_i), \\ d = \max\,(u_i), \end{cases} \quad (7)$$

Для побудови функції належності можна використати симетричну функцію Гауса, яка добре апроксимує експертні оцінки:

$$\mu^f(u) = e^{-\frac{(u-b)^2}{2c^2}}, \quad (8)$$

де $\mu^f(u)$ – функція належності, b – параметр функції належності, що відповідає координаті максимуму; c – параметр ширини «дзвону».

Розрахунок параметрів симетричної функції Гауса:

$$\begin{cases} b = \overline{(u_i)}, \forall i : pot_i = 1 \\ c = \sqrt{\frac{1}{k}\sum\limits_{i=1,k}(u_i - \bar{u})^2}, \end{cases} \quad (9)$$

де $\overline{(u_i)}$ – середнє значення серед даних, в яких потенціал дорівнює 1, $\overline{u_i}$ – середнє значення серед вибірки експериментальних даних терм-множини.

Для терм-множин PN – «перевищення норми» лінгвістичної змінної середньої різниці температур та MP – «максимальне перевищення» різниці температур симетричних позицій результати розрахунку значень таблично-заданої функції належності методом потенціалів та апроксимація класичними функціями належності зображені відповідно на рисунку 1-2.

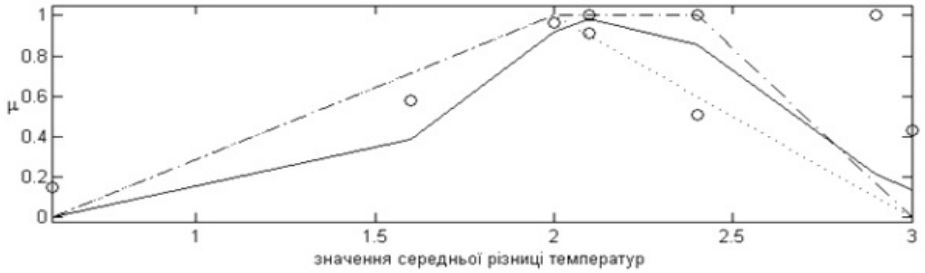

Рисунок 1 – Функції належності нечіткої терм-множини «перевищення норми» середньої різниці температур

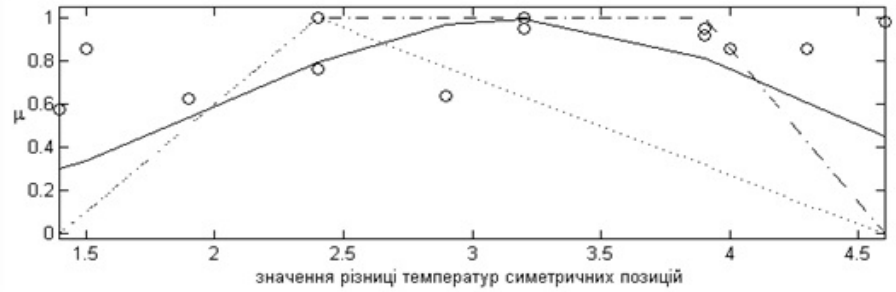

Рисунок 2 – Функції належності нечіткої терм-множини «максимальне перевищення» різниці температур симетричних позицій

На рисунках 1-2 значення таблично-заданої функції належності нечіткої терм-множини відображені кільцями, трикутна – подвійним пунктиром, трапецієподібна

– штрих пунктиром, симетрична функція Гауса – лінією.

Для оцінки точності апроксимації таблично-заданої функції належності обраними класичними функціями належності необхідно використати критерій:

$$\varepsilon = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\mu_i - \mu^f(u_i))^2}\ , \qquad (10)$$

де ε – значення критерію точності апроксимації.

В якості функції належності термножин обирається та класична функція належності, для якої помилка апроксимації буде мінімальною.

Оцінка точності апроксимації для терму PN - «перевищення норми» та MP – «максимальне перевищення» була проведена на 51 термограмі. Значення критерію точності апроксимації для трикутної, трапецієподібної та симетричної функції Гауса для терму PN дорівнюють відповідно 0,0989; 0,1004; 0,0873.

Значення критерію для трикутної, трапецієподібної та симетричної функції Гауса для терму MP дорівнюють відповідно 0,1005; 0,1157; 0,0723.

На основі наведених вище розрахунків для даних терм-множин у якості функції належності обрана симетрична функція Гауса.

ВИСНОВКИ

В ході дослідження розроблено метод визначення виду та параметрів функцій належності для формалізації лінгвістичних змінних нечіткої моделі аналізу термограм МЗ при діагностиці онкологічних захворювань. Метод поєднує експертні оцінки термограм, формування таблично-заданих функцій належності на основі методу потенціалів та їх апроксимацію класичними функціями належності. Застосування розробленого методу дозволяє автоматизувати побудову функції належності термів лінгвістичних змінних на основі набору термограм.

ЛІТЕРАТУРА:

1. Shushura O.M. Rozrobka modeli analizu termohram molochnykh zaloz zhinok na zasadakh nechitkoi lohiky. // Visnyk NTU "KhPI". Seriia: Informatyka ta modeliuvannia. – Kharkiv: NTU "KhPI". – 2016. – № 21. – S.184–192.
2. Berezkiy O.M. Nechitka baza znan intelektualnoyi sistemi diagnostuvannya vidiv raku molochnoyi zalozi. // Visnik Hmelnitskogo natsionalnogo universitetu – 2013. – №6 – S. 284-291.
3. Kaid V.A.A. Metody pobudovy funktsii nalezhnosti nechitkykh mnozhyn. // Zvistky Pivdennoho federalnoho universytetu. – 2013. – № 2 (139). – S. 144-153.
4. Melkumova E. M. Metody pobudovy funktsii nalezhnosti nechitkii mnozhyni. // Visnyk VDU. Seriia: Systemnyi analiz ta informatsiini tekhnolohii. – 2009. – № 2. – S. 13-18.
5. Kravets V.O. Analiz metodiv pobudovy funktsii prynalezhnosti pry obrobtsi ekspertnykh znan. // Visnyk NTU «KhPI». Seriia: Tekhnika ta elektrofizyka vysokykh napruh. – Kh.: NTU «KhPI», 2012. – № 52 (958). – S. 126-132.
6. Korolyuk N.O. Osoblivosti formalizatsiyi lingvistichnih zminnih, yaki vikoristovuyutsya pri opisi protsesu viboru parametriv zaplanovanogo perehvatu pri priznachenni vpliviv vinishchuvachami na povitryani tsili. // Viyskovo-tehnichni problemi. – 2006. – № 3. – S. 39-41.
7. Kondrashov C.I. Vplyv vydu funktsii nalezhnosti na chutlyvist pry otsiniuvanni yakosti // Systemy obrobky informatsii. – 2015. – № 6 – S.99-102.
8. Kuprinenko O.M. Pobudova funktsii nalezhnosti nechitkykh mnozhyn, yaki vidpovidaiut kilkisnym ekspertnym otsinkam. // Systemy obrobky informatsii. – 2007. – №. 5. – S. 142-143.
9. Patrakeev I. M. GIS v upravlinni teritoriyami: Konspekt lektsiy (dlya studentiv 5 kursu dennoyi formi navchannya spetsialnostey 7.070908, 8.070908 «Geoinformatsiyni sistemi i tehnologiyi») // Hark. nats. akad. misk. gosp-va. – H.: HNAMG, 2011. - 115 s.
10. Shtovba S.D. Pobudova funktsii nalezhnosti nechitkykh mnozhyn za klasteryzatsiieiu eksperymentalnykh danykh // Informatsiini tekhnolohii ta komp'iuterna inzheneriia. – 2006. – №2. – S. 92–95.

Рецензент: д.т.н., проф. Федоров Є.Є.,
зав. каф. КН, ДонНТУ