

УДК 659.1

ПОБУДОВА ФУНКЦІЙ НАЛЕЖНОСТІ НЕЧІТКИХ МНОЖИН ЗА КЛАСТЕРИЗАЦІЄЮ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

С.Д. Штовба

Вступ

40 років назад західні та радянські науковці майже одночасно дізналися про нечіткі множини [1, 2]. Сьогодні теорія нечітких множин є однією із модних та ефективних інтелектуальних технологій моделювання та проектування складних систем в умовах невизначеності. Її практичне застосування можливе, коли функції належності нечітких множин є відомими.

Функції належності традиційно визначають за експертною інформацією. Методом статистичної обробки експертної інформації [3] будують функції належності, які узагальнюють думки колективу фахівців щодо розподілу елементів по множинам. Методами парних порівнянь [4, 5] функції належності будують за опитуванням одного експерта. При цьому експерт за шкалою Сааті [6] попарно визначає перевагу одного елемента над іншим щодо задоволення властивостей нечіткої множини. Обмеженнями цих методів є використання суб'єктивної інформації та певних припущень при перетворенні її в ступені належності нечітких множин.

Інший підхід до побудови функцій належності базується на параметричній ідентифікації нечітких моделей за експериментальними даними “входи – вихід”. При ідентифікації оптимізують параметри функцій належності з метою мінімізації розбіжності між експериментальними даними та результатами нечіткого моделювання. Особливості навчання нечітких моделей різних типів добре описані в книжках [5, 7, 8, 9]. Застосування такого оптимізаційного підходу зняло суб'єктивізм визначення функцій належності, але тепер потрібно мати нечітку модель та навчальну вибірку “входи – вихід”. Зауважимо, що

функції належності семантично ідентичних нечітких множин, в результаті нечіткої ідентифікації різних залежностей “входи – вихід” будуть різними. Таким чином, функція належності стала чутливою до виду модельованої залежності та навчальної вибірки.

Майже поза увагою науковців залишилась задача побудови функцій належності за розподілом результатів спостережень. Ця задача аналогічна побудові функції розподілу випадкової величини за експериментальними даними. В статистиці для цього використовують метод гістограм. За цим методом будують і функції належності, розглядаючи функцію щільності випадкової величини як функції належності відповідної субнормальної нечіткої множини [10]. **Метою дослідження** є розробка нового підходу до побудови функцій належності за розподілом експериментальних даних, який, на відміну від методу гістограм, не вимагає великих вибірок. Основною ідеєю пропонованого підходу є використання кластеризації експериментальних даних, за якою синтезують функції належності при екстракції нечітких моделей [8, 9].

Постановка задачі

Будемо вважати відомими числові значення деякого показника: (y_1, y_2, \dots, y_v) . Розглядається дві задачі побудови функцій належності за цими даними.

Перша задача – синтез нечітких множин $\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_C$, функції належності яких відповідають розподіленню даних (y_1, y_2, \dots, y_v) . З математичної точки зору, ця задача відповідає відображенню виду:

$$(y_1, y_2, \dots, y_v) \rightarrow (\mu_{ij}), \quad i = \overline{1, v}, \quad j = \overline{1, C}.$$

де μ_{ij} – ступінь належності елемента y_i нечіткій множині \tilde{I}_j .

Друга задача – синтез однієї нечіткої множини \tilde{Y} , функції належності яких відповідають розподіленню даних (y_1, y_2, \dots, y_v) . З математичної точки зору, це відповідає відображенню виду:

$$(y_1, y_2, \dots, y_v) \rightarrow (\mu_1, \mu_2, \dots, \mu_v),$$

де μ_i – ступінь належності елемента y_i нечіткій множині \mathcal{Y} .

Для вирішення першої задачі пропонується використання нечіткої кластеризації за методом нечітких с-середніх [11], а для другої – потенціальної функції з кластеризації за гірським методом [8].

Побудова функцій належності за нечіткою кластеризацією

При нечіткій кластеризації кожен об'єкт може належати усім кластерам, але з різним ступенем. Якщо об'єкт має лише одну ознаку, тоді пропонується співставити нечіткий кластер одній нечіткій множині. При цьому функція належності кластера буде відповідати шуканій функції належності нечіткої числа. Знайдені нечіткі числа потрібно апроксимувати бажаними параметричними функціями належності для забезпечення властивостей нормальності та випуклості. Нижче описується основні положення нечіткої кластеризації [11] та наводиться приклад побудови функції належності за методом нечітких с-середніх.

Нечіткі кластера описують такою матрицею нечіткого розбиття:

$$\mathbf{F} = [\mu_{ij}] \quad \mu_{ij} \in [0, 1], \quad i = \overline{1, v}, \quad j = \overline{1, C},$$

в якій i -тий рядок містить ступені належності об'єкта y_i нечітким кластерам $\tilde{\mathcal{I}}_1, \tilde{\mathcal{I}}_2, \dots, \tilde{\mathcal{I}}_C$.

Матриця нечіткого розбиття має задовольняти такі умови:

$$\mathbf{F} = [\mu_{ij}] \quad \mu_{ij} \in [0, 1], \quad i = \overline{1, v}, \quad j = \overline{1, C},$$

$$\sum_{j=\overline{1, C}} \mu_{ij} = 1, \quad i = \overline{1, v};$$

$$0 < \sum_{i=\overline{1, M}} \mu_{ij} < v, \quad j = \overline{1, C}.$$

Для оцінки якості нечіткого розбиття використовується такий критерій [11]:

$$\sum_{j=\overline{1, C}} \sum_{i=\overline{1, v}} (\mu_{ij})^M \cdot (g_j - y_i)^2, \quad (1)$$

де $g_j = \frac{\sum_{i=\overline{1, v}} (\mu_{ij})^M \cdot y_i}{\sum_{i=\overline{1, v}} (\mu_{ij})^M}$ – центри нечітких кластерів $\tilde{\mathcal{I}}_1, \tilde{\mathcal{I}}_2, \dots, \tilde{\mathcal{I}}_C$;

$M \in [1, \infty)$ – експоненційна вага.

Найбільш популярним методом мінімізації критерія (1) є алгоритм нечітких с-середніх. Він базується на методі невизначених множників Лагранжа.

Приклад 1. Дані 392 вимірювань часу виконання операції “розігнання автомобілю до швидкості 60 міль за годину” наведені в [12]. В результаті кластеризації цих даних за методом нечітких с-середніх отримаємо 3 нечітких множини, функції належності яких зображено на рис 1. Нечіткі множини можна інтерпретувати лінгвістичними термами «Малий», «Середній» та «Значний».

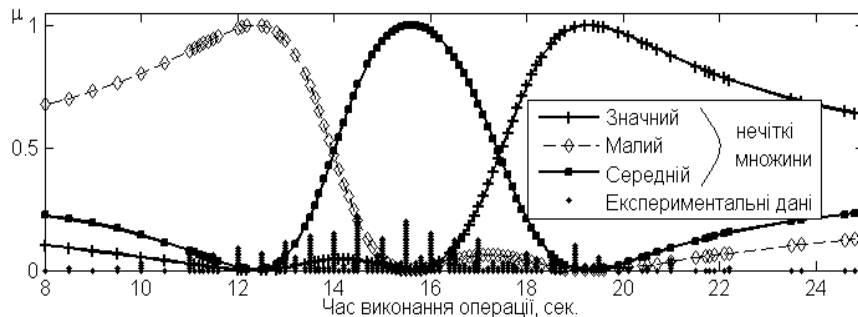


Рис. 1 Функції належності з прикладу 1, що побудовані за нечіткою кластеризацією

Побудова функції належності за кластеризацією методом потенціалів

При побудові функцій належності будемо використовувати ідеї потенціалу точки, на якому базується гірська кластеризація [8]. Потенціал точки – це число, яке показує наскільки щільно розташовані в її околі експериментальні дані. Чим вище потенціал точки, тим ближче вона до центру кластера. Потенціал точки y_i ($i = \overline{1, v}$) розраховують таким чином [8]:

$$pot_i = \sum_{j=\overline{1, v}} \exp(-4\alpha^2(y_i - y_j)^2), \quad (2)$$

де $\alpha > 0$ – коефіцієнт, що визначає розмазаність кластера за гірською кластеризацією.

Перед застосуванням формули (2) дані слід відмасштабувати на одиничний відрізок. Ступені належності нечіткої множини \mathcal{Y} пропонується розраховувати з потенціалів (2) таким чином:

$$\mu_{\mathcal{Y}}(y_i) = \frac{pot_i}{\max_{j=\overline{1, v}}(pot_j)},$$

При потребі отриману нечітку множину \mathcal{Y} можна апроксимувати бажаною параметричною функцією належності.

Приклад 2. Дані 392 вимірювань часу виконання операції “розігнання автомобілю до швидкості 60 миль за годину” наведені в [12]. Функції належності часу виконання цієї операції на автомобілях з різною кількістю циліндрів зображені на рис. 2. Функції належності побудовано за запропонованим методом потенціалів для двох значень коефіцієнта розмазаності: $\alpha = 1$ та $\alpha = 2$. Чим більше цей коефіцієнт, тим вища сконцентрованість нечітких множин. На відміну від методу гістограм, запропонований метод добре працює і при невеликому числі експериментальних точок, про це свідчить функція належності часу розгону на п’ятициліндрових автомобілях, яка побудована за трьома вимірними значеннями. Розраховані ступені належності добре апроксимуються двохсторонніми гаусовими функціями належності, графіки яких теж наведені на рис. 2.

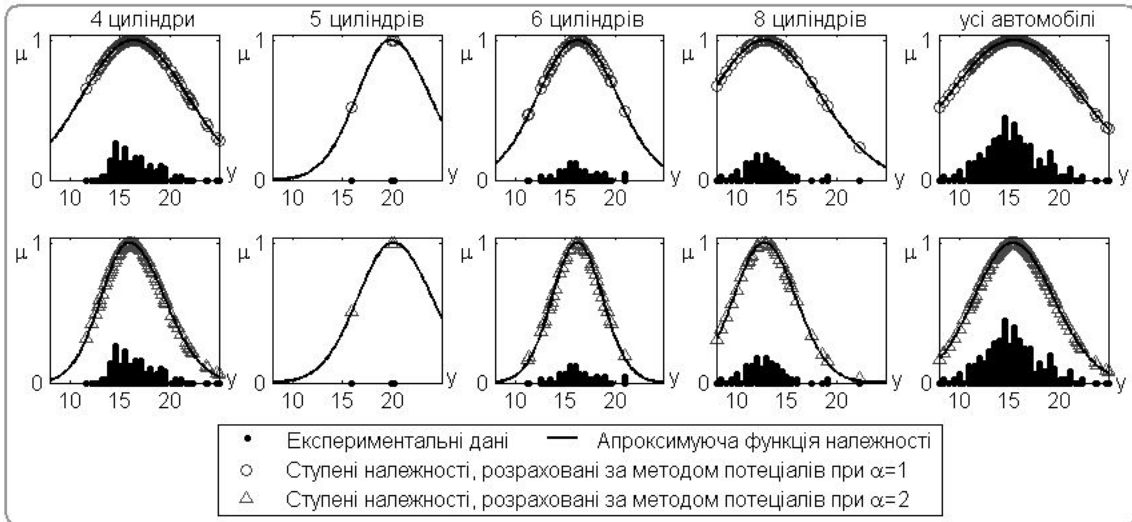


Рис. 2. Функції належності часу розгону автомобіля до 60 миль за годину, які побудовані в прикладі 2 за методом потенціалів

Висновки

Запропоновано два методи побудови функцій належності за розподілом експериментальних даних, які, на відміну від методу гістограм, не потребують великих вибірок. Основною ідеєю запропонованих методів є використання кластеризації експериментальних даних, за якою синтезують функції належності при екстракції нечітких моделей. Запропоновані методи будуть корисними для отримання початкових даних для моделювання за узагальненими на нечіткий випадок аналітичними моделями, наприклад, при прогнозуванні надійності алгоритмічних процесів [13, 14]. Іншим перспективним напрямком застосування запропонованих методів побудови функцій належності є формування нечітких навчальних вибірок для ідентифікації нелінійних залежностей нечіткими базами знань [15, 16].

Література

1. Zadeh L. Fuzzy Sets // Information and Control. – 1965. – Vol. 8. – P. 338–353.
2. Заде Л. Тени нечетких множеств // Проблемы передачи информации. – 1966. – Т.2. – С.37–44.
3. Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей: примеры использования. - Рига: Зинатне. - 1990.- 184 с.
4. Yager R.R. Fuzzy Decision Making Including Unequal Objectives // Fuzzy Sets and Systems. No.1.- 1978.- P. 87-95.
5. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети. - Винница: УНІВЕРСУМ-Вінниця, 1999. - 320 с.
6. Саати Т.Л. Взаимодействие в иерархических системах // Техническая кибернетика. - 1979.- №1.- С. 68-84.
7. Митюшкин Ю.И., Мокин Б.И., Ротштейн А.П. Soft-Computing: идентификация закономерностей нечеткими базами знаний. – Винница: УНІВЕРСУМ-Вінниця, 2002.– 145 с.
8. Yager R., Filev D. Essentials of Fuzzy Modeling and Control. USA: John Wiley & Sons.– 1994.– 387p.
9. Babuska R. Fuzzy Modeling for Control. Boston: Kluwer Academic Publishers.– 1998.
10. Дюбуа Д., Прад А. Теория возможностей. Приложения к представлению знаний в информатике. – М: Радио и связь, 1990. – 288 с.
11. Bezdek J.C. Pattern Recognition with Fuzzy Objective Function. New York: Plenum Press.- 1981.
12. MPG data base of UCI Machine Learning Repository (<http://www.ics.uci.edu/~mllearn/MLRepository.html>).
13. Ротштейн А.П., Штовба С.Д. Нечеткая надежность алгоритмических процессов.– Винница: Континент – ПРИМ, 1997.– 142 с.
14. Ротштейн А.П., Штовба С.Д. Прогнозирование надежности алгоритмических процессов при нечетких исходных данных // Кибернетика и системный анализ. – 1998. – №4. – С.85-93.
15. Штовба С.Д., Ротштейн О.П., Козачко О.М. Навчання нечіткої моделі по нечіткій навчальній виборці // Вісник Вінницького політехнічного інституту. – 2001. – №6. – С.18–23.
16. Ротштейн А.П., Штовба С.Д. Идентификация нелинейной зависимости нечеткой базой знаний с нечеткой обучающей выборкой // Кибернетика и системный анализ. – 2006. – №2. – С. 17–24.

Штовба Сергій Дмитрович, к.т.н., доцент,
докторант кафедри комп'ютерних систем управління,
Вінницький національний технічний університет,
Хмельницьке шосе, 95, Вінниця, 21021
shtovba@ksu.vstu.vinnica.ua www.vinnitsa.com/shtovba
Тел.: (0432)-598430, 598222.

Побудова функцій належності нечітких множин за кластеризацією експериментальних даних

С.Д. Штовба

Анотація. Запропоновано два методи побудови функцій належності нечітких множин за розподілом експериментальних даних. Ці методи, на відміну від методу гістограм, не потребують великих вибірок даних. Основною ідеєю запропонованих методів є використання кластеризації експериментальних даних, за якою синтезують функції належності при екстракції нечітких моделей.

Ключові слова: нечіткі множини, функція належності, кластеризація.

Построение функций принадлежности нечетких множеств с помощью кластеризации экспериментальных данных

С.Д.Штовба

Аннотация. Предложено два метода построения функций принадлежности нечетких множеств за распределением экспериментальных данных. Эти методы, в отличие от метода гистограмм, не требуют больших выборок данных. Основной идеей предложенных методов является использование кластеризации экспериментальных данных, по которой синтезируются функции принадлежности при экстракции нечетких моделей.

Ключевые слова: нечеткие множества, функции принадлежности, кластеризация.

Fuzzy set membership function construction via experimental data clustering

Serhiy Shtovba

Abstract. Two methods for membership function construction via experimental data clustering are proposed. These methods need less data set than histogram method. A main idea of the proposed methods is an usage the clustering procedure, that produces membership functions during extraction of fuzzy models from experimental data sets.

Keywords: fuzzy sets, membership function, clustering.