

ПІДХІД ДЛЯ ПОБУДОВИ НЕЧІТКИХ ЛОГІЧНИХ ПРАВИЛ ДЛЯ ВЕЛИКИХ ДАНИХ

Ляшенко А.В.

*Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна
E-mail: andrey.lyashenko44@gmail.com*

The approach to building fuzzy logical rules for big data

The article describes how to build fuzzy logic rules after processing and clustering large amounts of data.

У статті наводиться опис побудови нечітких логічних правил після обробки та кластеризації великих обсягів даних.

При проектуванні й розробці реальних смарт-систем з'являється проблема – формування складної математичної моделі, яка буде включати різні технічні особливості обладнання і фізичні характеристики середовища (складні обчислення можуть потребувати додаткові обчислювані ресурси). В ході подальшої розробки, може виникнути потреба врахувати нові умови, що може привести до повної переробки існуючої моделі. В такому випадку, різні умови можна можна визначати, як деякі правила (ЯКЩО ... ТО), використовуючи нечітку логіку. Для отримання цих правил, можна використовувати експертів, які самі задають умови. А можна навчити систему самій будувати ці правила, використовуючи різні алгоритми кластеризації [1].

Алгоритми кластеризації оперують з об'єктами. З кожним об'єктом X ототожнюється вектор характеристик $X_i = (x_1, \dots, x_d)$. Компоненти X_i , $i = 1, \dots, d$ є окремими характеристиками об'єкта. Розмірність простору характеристик визначає кількість характеристик d . Що складається з усіх векторів характеристик об'єктів, безліч позначається $M = (X_1, \dots, X_n)$. Підмножина «близьких один до одного» об'єктів з M являє собою кластер. Відстань $D(X_i, X_j)$ між об'єктами X_i і X_j визначається в просторі характеристик на основі обраної метрики.

Опис підходу. Суть підходу полягає в тому, що вхідні дані моделюють N -вимірний простір. Якщо змоделювати кожну характеристику об'єкта вхідних даних, як окрему вісь простору, то отримаємо модель в якій, кожен об'єкт цієї вибірки буде знаходитися в цьому просторі. Характеристика об'єкта – це лінгвістична змінна правила. Лінгвістична змінна – це сукупність всіх терм-множин. Терм - опис значення параметра (наприклад температура: холодно, тепло, жарко). Якщо на вхід подається вибірка, яка має тільки дві лінгвістичні змінні, тобто має тільки два стовпця, то вимірний простір – це двовимірна площина. На етапі не ієрархічної кластеризації, обчислюються значення центрів кластерів. Кожен кластер являє собою терм-множину.

P_j	1	2	3	4	5	6	7	8	9
$x=x_1$	0.7	1.4	1.2	1.8	2.5	2.3	2.4	3.0	2.1
$y=x_2$	0.8	0.9	1.3	0.8	0.9	1.6	2.4	2.7	3.0

Рис. 1. Вхідна вибірка.

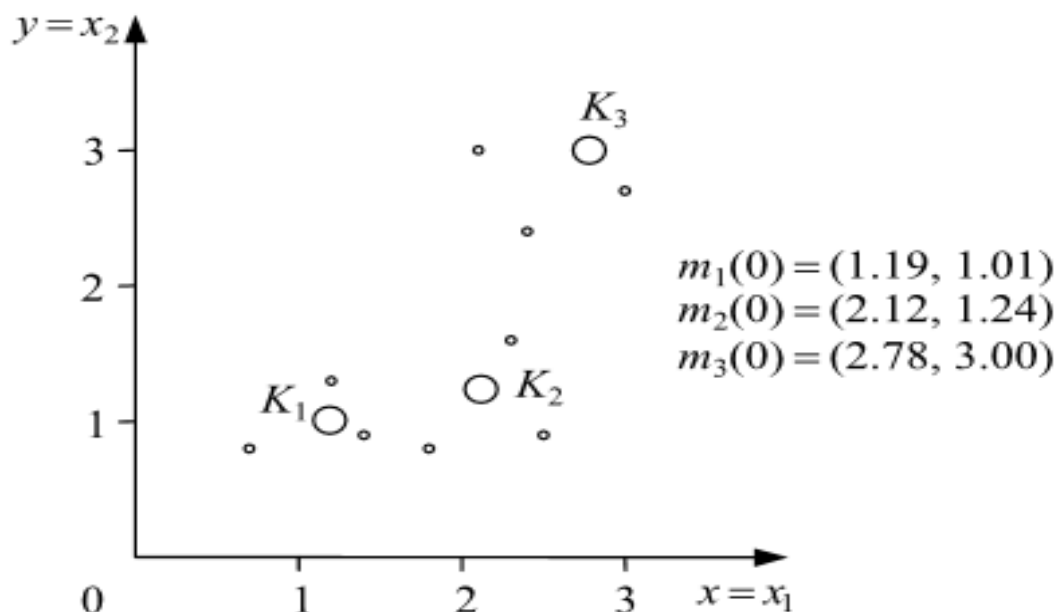


Рис. 2. Двох-вимірний простір, побудований із вхідної вибірки.

При побудові нечітких логічних правил, потрібно визначити функції приналежності. Функції приналежності – це функція, яка по осі ординат має значення від 0 (однозначно належить) до 1 (однозначно не належить), а по осі абсцис – кількісні значення даного терма. Ці функції бувають різних видів: у формі трапеції, трикутна форма, L-функція, R-функція, функція Гауса. Найлегше можна побудувати функцію Гауса, а вже потім переходити до інших видів.

Функція Гауса має вид: $\mu(x) = e^{-\frac{(m_{xi}-x)^2}{2\sigma^2}}$, де x – центр кластера, σ – середньо-квадратичне відхилення.

Для знаходження ширини σ скористаємося формулою $\frac{|m_{xi} - m_{x(i+1)}|}{N_x}$, де m_{xi} – центр кластера, $m_{x(i+1)}$ – центр наступного кластера, N_x – кількість кластерів.

Тепер потрібно спроектувати та побудувати на кожну вісь N-вимірного простору, для кожного кластера (терма) функції приналежності Гауссового виду [2].

Після побудови функцій приналежності, потрібно перетворити числові дані, які поступали на вхід, на значення термів лінгвістичних змінних. Кожне числове значення належить якомусь кластеру відповідної лінгвістичної змінної. Відповідно кожен об'єкт із вхідної вибірки, створить окреме нове правило вигляду *Якщо .. Та ... То*. Кожне правило такого типу можна унікально ідентифікувати за набором умов *Якщо*. Під час «мапінгу» числових даних в

терми лінгвістичних змінних, можуть виникати такі умови, коли правила будуть повторюватися. Цей підхід дозволяє уникати таких ситуацій, пропускаючи повторююче правило.

Таким чином, будуються нечіткі логічні правила, для великих обсягів даних.

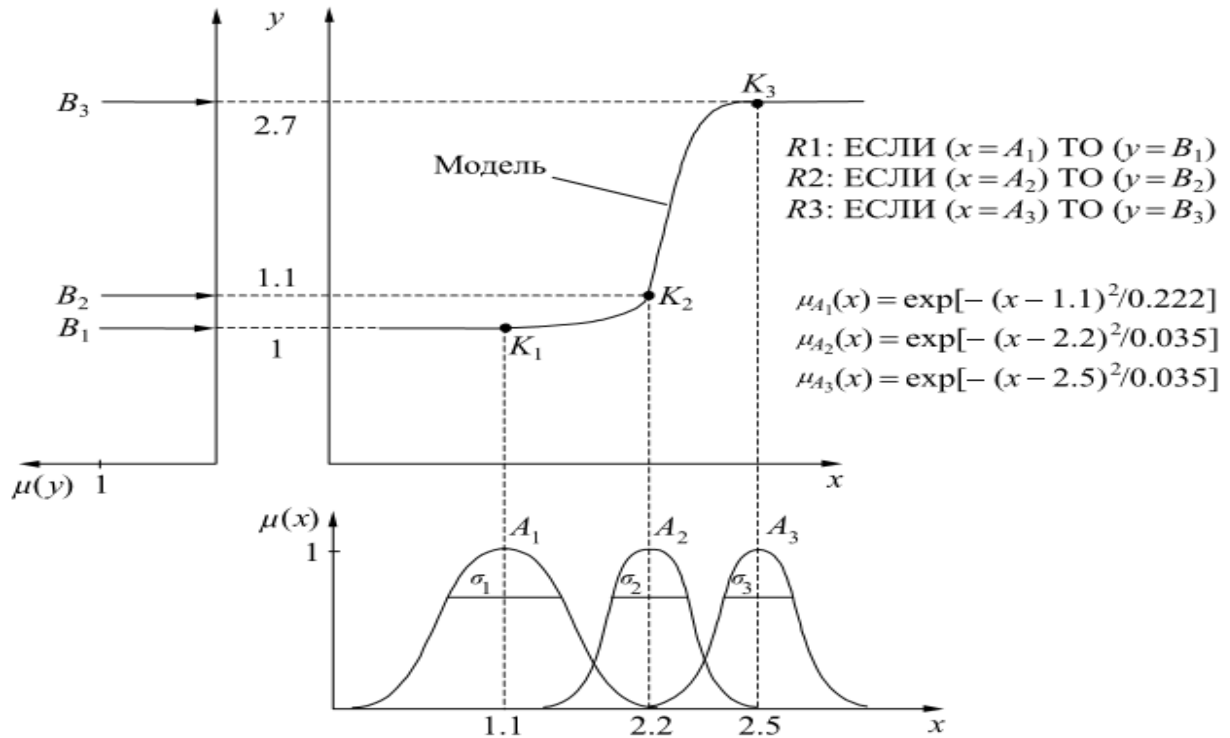


Рис. 3. Побудова функцій приналежності на кожен лінгвістичну змінну.

Висновки. В даній статті був запропонований підхід для побудови нечітких логічних правил для великих даних. Цей підхід базується на моделюванні вхідної відкластеризованої вибірки даних в N-вимірний простір. Кожна вісь цього простору являє собою лінгвістичну змінну, на яку проектується і будується функція приналежності Гауссового виду. Цей підхід допомагає віднести кожне значення об'єкта вхідної вибірки, до відповідного йому терма лінгвістичної змінної і, таким чином, побудувати нечіткі логічні правила, з яких можна створити нечітку базу знань.

В майбутньому планується розробити систему, яка дозволить будувати нечіткі логічні правила, використовуючи даний підхід.

Література

1. Analysis of clustering algorithms for use in the universal data processing system. Larysa Globa, Yurii Buhaienko, Andrii Liashenko, Maria Grebinichenko National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" Kyiv.
2. Нечеткое моделирование и управление. А. Пегат с. [506-520].