

ДЕЯКІ МІРИ СХОЖОСТІ ОБ'ЄКТІВ

Мета кластерного аналізу – виявити групи об'єктів (індивідів), що мають максимальну кількість спільних рис (схожих) між собою і в той же час мінімальну схожість з іншими групами. Оскільки техніки кластерного аналізу дозволяють визначати схожість і відмінність, вони дуже корисні в сенсі виявлення в сукупності даних деяких базових взірців (представників кластерів). Внаслідок цього вони часто використовуються при попередньому аналізі даних.

Мірою подібності, при кластеризації даних, як правило виступає відстань між об'єктами, на основі якої і побудовані різні види метрик. Але існує багато класів практичних задач, із фізичного змісту яких слідує, що потрібно виявити схожість об'єктів $O_i, i = \overline{1, m}$ за їх векторами ознак $\overline{c_i}, i = \overline{1, m}$ використавши інші (альтернативні) міри подібності, зокрема кутову та довжинну [1]. Крім того, в багатьох задачах кластерного аналізу групування об'єктів необхідно проводити тільки в групах однотипних векторних ознак.

З метою здійснення попереднього аналізу вхідних об'єктів, для виявлення їх однотипних груп, пропонується використати функції-фільтри, які, в певному сенсі, визначають міру знакової однотипності об'єктів. Для їх математичної формалізації використано математичний апарат теорії нечітких бінарних відношень.

Нечітке бінарне відношення R^{RS} із функцією належності $\varphi_{R^{RS}} : \{\overline{c_i} | i = \overline{1, m}\}^2 \rightarrow [0, 1]$:

$$\varphi_{R^{RS}}(\overline{c_i}, \overline{c_j}) = \frac{2n - \left| \sum_{k=1}^n \text{sign}(c_i^k) - \sum_{l=1}^n \text{sign}(c_j^l) \right|}{2n}$$

буде визначати відносну відсоткову однотипність (Relative Similarity) об'єктів. Тобто, чим більше вектори ознак $\overline{c_i}$ та $\overline{c_j}$ двох об'єктів O_i і O_j будуть мати подібну сумарну знакову структуру часткових ознак, тим ближче $\varphi_{R^{RS}}$ буде ближче до одиниці. І навпаки. Крім того, числове значення $\varphi_{R^{RS}}$ буде відповідати відсотку їх схожості.

Нечітке бінарне відношення R^{AS} із функцією належності $\varphi_{R^{AS}} : \{\overline{c_i} | i = \overline{1, m}\}^2 \rightarrow [0, 1]$:

$$\varphi_{R^{AS}}(\overline{c_i}, \overline{c_j}) = \frac{2n - \sum_{k=1}^n |\text{sign}(c_i^k) - \text{sign}(c_j^k)|}{2n}$$

буде визначати відносну відсоткову однотипність (Absolute Similarity) об'єктів. Чим більша буде знакова схожість відповідних часткових ознак векторів $\overline{c_i}$ та $\overline{c_j}$ двох об'єктів O_i і O_j , тим ближче $\varphi_{R^{AS}}$ буде ближче до одиниці. Як і в попередньому випадку, числове значення $\varphi_{R^{AS}}$ буде відповідати відсотку їх схожості.

Література

1. Кондрук, Н. Е. Деякі методи автоматичного групування об'єктів / Н. Е. Кондрук // Південно-Європейський журнал передових технологій. – 2014. – № 2(4). – С.20–24.
2. Кондрук, Н. Е. Застосування багатокритеріальних моделей для задач збалансованого харчування / Н. Е. Кондрук, М. М. Маляр // Вісник Черкаського державного технологічного університету. Серія: технічні науки. 2010. – №1. – Вип. 1. – С.3–7.