

ЗВЕДЕННЯ ЗАДАЧІ АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ ДО ТРАНСПОРТНОЇ ЗАДАЧІ

Матрична постановка транспортної задачі має вигляд [1]:

– цільова функція
$$\sum_{i=1}^m \sum_{j=1}^n X_{ij} C_{ij} \rightarrow \min .$$

– обмеження

$$\left\{ \begin{array}{l} \sum_{i=1}^m X_{ij} \Theta B_i, \text{ де } \left\{ \begin{array}{l} \Theta \Rightarrow "=" \text{ в закритій та в задачі з перевищенням ресурсів} \\ \Theta \Rightarrow "<" \text{ в задачі з перевищенням потреб} \end{array} \right. , \\ \sum_{j=1}^n X_{ij} \Theta A_j, \text{ де } \left\{ \begin{array}{l} \Theta \Rightarrow "=" \text{ в закритій та в задачі з перевищенням потреб} \\ \Theta \Rightarrow "<" \text{ в задачі з перевищенням ресурсів} \end{array} \right. , \\ X_{ij} \geq 0, \end{array} \right.$$

де A_i – ресурс i -го постачальника (запас продукції або план відвантаження з поточного виробництва); B_j – потреби в тій же продукції в пунктах j ; C_{ij} – відстань або вартість перевезення з i в j ; $j=1,2, \dots, n$ – кількість постачальників, $i=1,2, \dots, m$ – кількість споживачів; X_{ij} – змінні параметри задачі в яких вираховується оптимальний обсяг перевезення продукції від i -го постачальника j -му споживачеві.

Автоматична класифікація об'єктів нагадує цю задачу, якщо відстань між об'єктами класифікації (D_{ij}) представити у вигляді діагонально-симетричної квадратної матриці, в головній діагоналі якої стоять нулі, що очевидно, адже відстань кожного об'єкта до самого себе дорівнює нулю.

Тоді загальна сумарна відстань між об'єктами має прагнути до максимуму, щоб утворені кластери якомога менше перетиналися

$$\sum_{i=1}^m \sum_{j=1}^n X_{ij} D_{ij} \rightarrow \max ,$$

де X_{ij} – квадратна матриця бінарного типу, тобто, така, що приймає значення 0 або 1.

В ній також здійснюється «перевезення» кожного i -го об'єкту до свого j -го кластеру. На початку розрахунку кількість кластерів дорівнює кількості об'єктів. Очевидно, що деякі кластери будуть пустими, тобто в них не попаде жоден об'єкт. Аналогом цього є транспортна задача з перевищенням потреб. Тоді маємо наступні обмеження

$$\left\{ \begin{array}{l} \sum_{i=1}^m X_{ij} = 1, \\ \sum_{j=1}^n X_{ij} \leq n, \\ X_{ij} - \text{бінарні } [0,1], \end{array} \right.$$

де $i, j=1,2, \dots, n$ – кількість об'єктів. Подібна постановка задачі дозволяє робити класифікацію за допомогою функції Solver електронних таблиць Excel.

В сучасних версіях цієї програми дозволяється використовувати до 200 змінних факторів. З урахуванням того, що матриця X_{ij} – квадратна – це означає можливість класифікувати до 14 об'єктів.

Наведемо приклад. Нехай матриця відстаней (D_{ij}) між п'ятьма об'єктами має наступні значення

		Об'єкти				
		1	2	3	4	5
Об'єкти	1	0	0,89	0,41	0,74	0,46
	2	0,89	0	0,86	0,87	0,61
	3	0,41	0,86	0	0,96	0,66
	4	0,74	0,87	0,96	0	0,62
	5	0,46	0,61	0,66	0,62	0

Тоді вирішення цієї задачі 0-1 програмування дає нам наступний результат, в якому цільова функція прийме значення 4,36. Як видно з рішення, перший об'єкт потрапив у другий кластер, другий – в перший, 4 та 5 об'єкти – у третій кластер, четвертий – у третій. П'ятий кластер виявився пустим.

		Кластери					Сума по об'єктам
		1	2	3	4	5	
Об'єкти	1	0	1	0	0	0	1
	2	1	0	0	0	0	1
	3	0	0	0	1	0	1
	4	0	0	1	0	0	1
	5	0	0	1	0	0	1
Сума по кластерам		1	0	2	1	0	

Список використаних джерел

1. Пістунів І.М., Антонюк О.П., Турчанинова І.Ю. Кластерний аналіз в економіці: Навч. посібник – Дніпропетровськ: Національний гірничий університет, 2008.– 84 с.