

ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ КРИТЕРІЇВ ЯКОСТІ НЕЧІТКОЇ КЛАСТЕРИЗАЦІЇ

**Н. Р. Кондратенко, к.т.н., доцент,
О. О. Снігур, аспірант
Вінницький національний технічний університет
sleeeey@meta.ua**

Основними проблемами, характерними для переважної більшості сучасних методів нечіткого кластерного аналізу, є визначення оптимального числа кластерів та вибір рівня нечіткості. Крім того, прикладне застосування будь-якого методу суттєво ускладнюється за наявності в експериментальних даних шумів та аномалій.

Під аномаліями, або аномальними спостереженнями, слід розуміти спостереження, що належать до кластерів, число представників яких у досліджуваній сукупності є суттєво малим порівняно з числом представників основних кластерів.

Зашумленість та значна кількість аномалій породжують невизначеність, яка не може не перенестися на остаточний результат кластеризації. Ці невизначеності можуть настільки спотворити результат, що висновок про приховані закономірності, що існують у вихідних даних, буде хибним.

Зважаючи на це, в задачі кластеризації пропонується використовувати математичний апарат інтервальних нечітких множин типу 2. Вони дають змогу ефективно враховувати невизначеності, присутні у вихідних даних, представляючи результат кластеризації у формі інтервалу, а не єдиного числа. Інтервал несе в собі значно більше

інформації про характер досліджуваних даних, зокрема його ширина може говорити про якість отриманого розбиття.

Границі інтервалів пропонується визначати, виходячи з різних значень рівня нечіткості, що є ключовим параметром переважної більшості алгоритмів нечіткої кластеризації. Проблема оптимального задання рівня нечіткості тісно пов'язана з визначенням оптимального числа кластерів. Для наближеної оцінки якості кластеризації при різних значеннях цих двох параметрів існує ряд критеріїв.

Слід зауважити, що загального, універсального критерію, який дав би змогу прийняти остаточне рішення про якість того чи іншого розбиття, не існує. Виходячи з цього, було проведено дослідження властивостей різних критеріїв якості та їх поведінки при різних значеннях рівня нечіткості та кількості кластерів. Сформульовано підхід на основі критеріїв якості для інтервальної нечіткої кластеризації на основі.

В основу запропонованого підходу покладено метод робастної кластеризації, що дає змогу принаймні частково уникнути впливу аномалій на остаточний результат.

Такий підхід дасть змогу більш повною мірою зрозуміти особливості та закономірності, характерні для досліджуваної множини об'єктів.