

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

ІНФОРМАТИКА, МАТЕМАТИКА,
АВТОМАТИКА

ІМА :: 2016

**МАТЕРІАЛИ
та програма**

НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

(Суми, 18–22 квітня 2016 року)



Суми
Сумський державний університет
2016

Використання SR-дерев у щільнішому методі кластеризації числових просторів DBSCAN

Козлов З.В., *аспірант*; Кулік Є.С., *аспірант*
Сумський державний університет, м. Суми

Алгоритм щільнісного кластерного аналізу DBSCAN при обробці просторових даних дуже часто потребує знаходження ε -околу точки у n -вимірному просторі. Цей крок алгоритму виконується для кожної точки щонайменше один раз [1], отже покращення ефективності знаходження ε -околу матиме значний вплив на результати роботи алгоритму загалом. Тривіальна реалізація цього етапу щільнісної кластеризації має асимптотичну складність $O(n)$, при цьому загальна асимптотична складність DBSCAN складає $O(n^2)$.

Для вирішення цієї проблеми було застосовано структуру даних Sphere/Rectangle Tree (SR-Tree), асимптотична складність пошуку ε -околу в якій складає $O(\log n)$. SR-Tree використовує обмежувальні гіперсфери та гіперкуби для побудови дерева, висота якого складає щонайбільше $\log n$, де n – кількість точок у множині, що підлягає кластерному аналізу [2]. Таким чином, асимптотична складність щільнісного алгоритму DBSCAN набуває значення $O(n * \log n)$.

Для перевірки ефективності SR-Tree було застосовано алгоритм DBSCAN з його використанням для кластерного аналізу двовимірної множини даних, яка не має однорідної щільності точок у просторі. Аналіз результатів показав, що швидкість роботи алгоритму підвищилася в 658 разів, при цьому збереглася точність знаходження кластерів та кількість точок, що були віднесені до шуму. Таким чином, було покращено швидкість роботи щільнісного алгоритму кластерного аналізу DBSCAN, при цьому не порушуючи його ефективність.

Керівник: Довбиш А.С., *професор*

1. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *In Proc. ACM SIGMOD Int. Conf. on Management of Data* **22**, 226 (1996).
2. N. Katayama, S. Satoh, *Proc. ACM SIGMOD Int. Conf. on Management of Data* **23**, 369 (1997).