

ПРО РЕЗУЛЬТАТИ ДІАГНОСТУВАННЯ НАЯВНОСТІ ЗАХВОРЮВАНЬ З ВИКОРИСТАННЯМ АЛГОРИТМІВ КЛАСИФІКАЦІЇ НА ОСНОВІ НЕЧІТКИХ ПРАВИЛ

Безменова О.М.

*Національний технічний університет
«Харківський політехнічний інститут»,
ТОВ «CloudWorks Ltd.», м. Харків*

Робота присвячена застосуванню методів нечіткої логіки при розв'язанні задач кластеризації.

Було розроблено декілька алгоритмів, а саме, базовий алгоритм і алгоритми, що базуються на декомпозиції задачі на декілька підзадач.

Перевірка роботи алгоритмів здійснювалася на спеціально підібраних модельних даних, після чого була здійснена спроба застосувати алгоритми для обробки реальних даних – обстежень пацієнтів на хвороби серця, що взяті з бази даних медичного центру Клівленда. База даних містить дані по 303 обстеженням. Для побудови матриці нечіткого розбиття та еталонів кластерів використовувалися 227 записів з бази. Решта записів використовувалася для перевірки правильності діагностики.

Серед 227 пацієнтів по яким проводилася побудова класів, 123 – здорові, 104 – мають хвороби серця. Розбиття пацієнтів на 2 класи, за урахуванням діагнозів дало наступні результати:

Базовий алгоритм: 114 пацієнтів, 26 хворих, 88 здорових (23% хворі, 77% – здорові); 2 клас – 113 пацієнтів, 78 хворих, 35 здорових (69% хворі, 31% – здорові). Вдосконалений базовий алгоритм: 116 пацієнтів, 26 хворих, 90 здорових (23% хворі, 77% – здорові); 2 клас – 111 пацієнтів, 78 хворих, 33 здорових (70% хворі, 30% – здорові).

Однорівнева декомпозиція: 113 пацієнтів, 28 хворих, 85 здорових (25% хворі, 75% – здорові); 2 клас – 114 пацієнтів, 76 хворих, 38 здорових (67% хворі, 33% – здорові).

Таким чином, перший клас здебільшого сконцентрував пацієнтів, щодо яких відомо, що вони здорові. Другий клас містить здебільшого пацієнтів з хворобами серця.

Після навчання алгоритму для того, щоб перевірити точність діагностики пацієнтів, на основі обчислених еталонів кластерів проводилася діагностика для 76 обстежень пацієнтів, що не використовувалися при роботі алгоритмів кластеризації, серед них 42 – здорові, 34 – мають хвороби серця.

Результати діагностики наступні: після побудови еталонів класів базовим алгоритмом відсоток правильної діагностики пацієнтів склав 71%; після побудови еталонів класів вдосконаленим базовим алгоритмом відсоток правильної діагностики пацієнтів склав 71%; після побудови еталонів класів декомпозиційним алгоритмом відсоток правильної діагностики пацієнтів склав 69%.