**CURE Algorithm**

**Проект по „Извличане на закономерности от данни”**

**на Георги Тончев фн:** **М23449 , Светлин Чобанов фн: М23459**

**и Диана Чалъкова фн: М23337**

**Cure:**

Методът CURE е пример за агломеративна клъстеризация. Основната разлика с останалите такива алгоритми е в „гъвкавото” избиране на представителни точки на всяка стъпка. Като резултат от избора може да бъде получена „неправилна” форма от представителни точки.

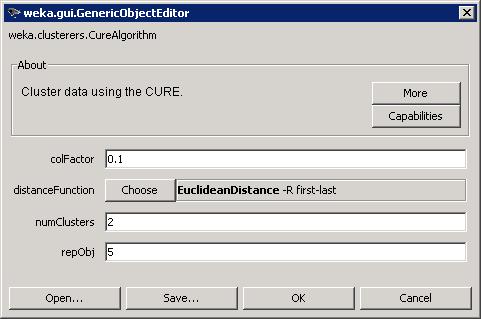
**Начин на използване:**

Алгоритъмът е интегриран във вече съществуващата система Weka и използва нейния интерфейс.

Системата предоставя възможност за избор на голяма част от параметрите в интерфейса си.

Входни параметри:

1. repObj - максимален брой на представителни обекти в клъстера
2. numClusters - желаният брой на изходни клъстери
3. colFactor- фактор на свиване (в интервал от 0 до 1)
4. distanceFunction- функция на разстояние между точките в клъстера.



**Описание на алгоритъма:**

1. Инициализирай клъстерното множество (всяка точка – отделен клъстер).

2. За всеки клъстер Ci намери най-близки клъстер Cj (на този етап разстоянието

между клъстери е просто разстояние между съответните точки от тези

клъстери)

3. Направи сливането на два най-близки клъстера Ck и Cl в клъстер Cm (както

винаги Cm е обединение от точки от клъстери Ck и Cl )

4. За клъстер Cm избери c представителни (добре разпределени в пространството)

точки:

i. Избери първата точка, най-отдалечена от центроида на Cm

ii. Докато c точки не бъдат избрани направи:

Избери точката, която най-далечна по отношение на предишната

5. Премести избраните c точки по направление към центроида на клъстера с

фактор на свиване α: p = p + α (mean - p), където p е координати на

представителната точка, mean - е координати на центроида.

6. Определи клъстер, който е най-близък към Cm, като разстоянието между

клъстери се определя като разстояние между две най-близки представителни

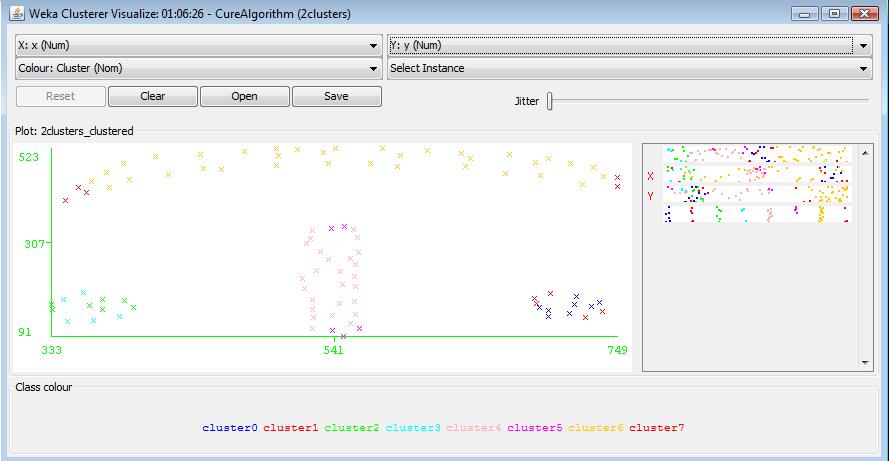
точки на съответните клъстери.

7. Направи сливане на най-близката двойка клъстери

8. Повтори стъпки 3-7 докато не се останат k клъстера.



Фигура (a) представя два клъстера като по-големите кръгчета представляват представителните точки. На тази итерация алгоритъмът ще слее двата клъстера като най-близки (фигура (b)), защото разстоянието между две представителни точки от тях е най-малко.



Разделяне на сто точки на четири клъстера с CURE. През един, започвайки от клъстер номер 1, са изобразени представителните точки на съответните клъстери. Може да се види, че точките не образуват правилна геометрична фигура.

**Особености при имплементацията:**

1. При избора на представителни точки от клъстера, ако има повече от указания брой точки с еднакво разстояние – то се избират тези, които са влезли по-рано в клъстера(на възможно най-ранна итерация).
2. Алгоритъмът работи с многомерни числови данни.