

Анализ на задача 4 – Глобус

*Борис Караджов
Development Manager
i:FAO*

Предложената в четвърти кръг задача може да се разглежда като конкретен проблем на клъстерния анализ [1,2] със наложени специфични условия:

- Твърдо асоцииране на всеки град към точно един клъстер (за разлика от т.нар. fuzzy clustering[3]).
- Клъстери с формата на кръгове и с еднакъв радиус.
- Центърът на всеки клъстер трябва да съвпада с някой от зададените градове.

Съответно един възможен подход към решаване на проблема е адаптирането на някой от съществуващите алгоритми [4]. Например:

- Групиране в малки клъстери на базата на локална близост и след това проверяване на възможностите за сливането им в по-големи. Такова първоначалното групиране ще намали броя комбинации на втория етап, но може да доведе до изпускане на някои решения с малък брой области.
- Пълно обхождане на оставащите валидни кандидати за центрове с връщане (backtracking) при достигане на нерешима конфигурация. Например избира се център и радиус R на първата област. Всички оставащи градове извън областта, които са по-близо от R до град от тази област, не могат да бъдат кандидати за центрове на друга област. Ако освен това един такъв все още неасоцииран с област град се намира и по-далеч от R от всички оставащи кандидати за областни центрове, решение с направеното до този момент разбиване не може да бъде намерено и е необходимо връщане.

Заслужава да се отбележи, че избраната повърхност (сфера) играе роля само до момента, в който бъде изчислена симетричната матрица на разстоянията между всички двойки градове. Най-лесно това може да стане с изчисляване на декартовите координати на градовете при единичен радиус на сферата и определяне на разстоянието (или квадрата на разстоянието, което спестява операцията извличане на квадратен корен) по Питагоровата теорема.

Алгоритъмът, който след това ги групира в области, може да се окаже приложим със същия успех към обекти разположени върху равнина, цилиндър, тор и т.н.

Във всички случаи задачата има тривиално решение, което се състои в обособяването на всеки град в самостоятелна област. Следователно, добра стратегия е малко преди изтичане на лимита от 10 секунди програмата да генерира такова тривиално решение, ако по-добро не е било намерено.

Тестовите се състояха от три групи:

- Координати на реални градове – тестове 1-4 и 16.
- Синтетични тестове – градове равномерно разположени върху паралели, меридиани и др. – тестове 5-10. Тук заслужава да се отбележи тест №9, който демонстрира

една „патологична” конфигурация – градовете са равномерно разположени върху окръжност. В такъв случай максималният брой градове в една е област е най-големия нечетен делител на техния брой. За конкретния тест със 720 града минималният брой области е $720 / 45 = 16$.

- Полусинтетични тестове – градовете са случайно разположени във вътрешността на правилни геометрични фигури – тестове 11-15.

Линкове:

1. http://en.wikipedia.org/wiki/Cluster_analysis
2. <http://www.fmi.uni-sofia.bg/fmi/statist/Personal/Vandev/lectures/applstat2.pdf>
3. http://en.wikipedia.org/wiki/Fuzzy_clustering
4. <http://cs.joensuu.fi/pages/franti/vq/algorithms.htm>