

## عنوان مقاله:

Repeated Record Ordering for Constrained Size Clustering

## محل انتشار:

ماهنامه بین المللی مهندسی، دوره 33، شماره 7 (سال: 1399)

تعداد صفحات اصل مقاله: 8

## نویسنده:

R. Mortazavi - School of Engineering, Damghan University, Damghan, Iran

## خلاصه مقاله:

One of the main techniques used in data mining is data clustering, which has many applications in computer science, biology and social sciences. Constrained clustering is a type of clustering in which side information provided by the user is incorporated into current clustering algorithms. One of the well researched constrained clustering algorithms is called microaggregation. In a microaggregation technique, the algorithm divides the dataset into groups containing at least  $k$  members, where  $k$  is a user-defined parameter. The main application of microaggregation is in Statistical Disclosure Control (SDC) for privacy preserving data publishing. A microaggregation algorithm is qualified based on the sum of within-group squared error,  $W$ . Unfortunately, it has been proven that the optimal microaggregation problem is NP-Hard in general, but the special univariate case can be solved optimally in polynomial time. Many heuristics exist for the general case of the problem that are founded on the univariate case. These techniques order multivariate records in a sequence. This paper proposes a novel method for record ordering. Starting from a conventional clustering algorithm, the proposed method repeatedly puts multivariate records into a sequence and then clusters them again. The process is repeated until no improvement is achieved. Extensive experiments have been conducted in this research to confirm the effectiveness of the proposed method for different parameters and datasets.

## کلمات کلیدی:

Constrained Clustering, Microaggregation, data privacy

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1040948>

