

UNSUPERVISED CLASSIFICATION AND PARTICLE SWARM OPTIMIZATION

Adam Truszkowski, Magdalena Topczewska

Faculty of Computer Science, Białystok University of Technology, Białystok, Poland

Abstract: This article considers three algorithms of unsupervised classification - *K-means*, *Gbest* and the *Hybrid* method, the last two have been proposed in [D.W. van der Merwe and A.P. Engelbrecht, Data clustering using particle swarm optimization, The 2003 Congress on Evolutionary Computation (CEC'03), vol. 1, 215- 220, Canbella, Australia, 2003]. All three algorithms belong to the class of non-hierarchical methods. At first, the initial split of objects into known in advance number of classes is performed. If it is necessary, some objects are then moved into other clusters to achieve better split - between cluster variation should be much larger than within cluster variation. The first algorithm described in this paper (*K-means*) is well-known classical method. The second one (*Gbest*) is based on the particle swarm intelligence idea. While the third is a hybrid of two mentioned algorithms. Several indices assessing the quality of obtained clusters are calculated.

Keywords: unsupervised classification, clustering, particle swarm optimization

KLASYFIKACJA NIENADZOROWANA I OPTYMALIZACJA ROJEM CZĄSTEK

Streszczenie: W niniejszym artykule porównywane są trzy algorytmy analizy skupień - metoda k-średnich, algorytm gbest oraz metoda hybrydowa. Algorytmy gbest oraz hybrydowy zostały zaproponowane w publikacji [D.W. van der Merwe and A.P. Engelbrecht, Data clustering using particle swarm optimization, The 2003 Congress on Evolutionary Computation (CEC'03), vol. 1, 215- 220, Canbella, Australia, 2003]. Wszystkie trzy metody należą do rodziny metod niehierarchicznych, w których na początku tworzony jest podział obiektów na znaną z góry liczbę klastrów. Następnie, niektóre obiekty przenoszone są pomiędzy klastrami, by uzyskać jak najlepszy podział - wariancja pomiędzy skupieniami powinna być znacznie większa niż wariancja wewnątrz skupień. Pierwszy algorytm (k-means) jest znaną,

klasyczną metodą. Drugi oparty jest na idei inteligencji roju cząstek. Natomiast trzeci jest metodą hybrydową łączącą dwa wymienione wcześniej algorytmy. Do porównania uzyskanych skupień wykorzystano kilka różnych indeksów szacujących jakość otrzymanych skupień.

Słowa kluczowe: klasyfikacja nienadzorowana, analiza skupień, optymalizacja rojem cząstek