

ESTIMATION OF PARAMETERS OF GAUSSIAN MIXTURE MODELS BY A HYBRID METHOD COMBINING A SELF-ADAPTIVE DIFFERENTIAL EVOLUTION WITH THE EM ALGORITHM

Wojciech Kwedlo

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

Abstract: In the paper the problem of learning of Gaussian mixture models (GMMs) is considered. A new approach based on hybridization of a self-adaptive version of differential evolution (DE) with the classical EM algorithm is described. In this approach, called DE-EM, the EM algorithm is run until convergence to fine-tune each solution obtained by the mutation and crossover operators of DE. To avoid the problem with parameter representation and infeasible solutions we use a method in which the covariance matrices are encoded using their Cholesky factorizations. In a simulation study GMMs were used to cluster synthetic datasets differing by a degree of separation between clusters. The results of experiments indicate that DE-EM outperforms the standard multiple restart expectation-maximization algorithm (MREM). For datasets with high number of features it also outperforms the state-of-the-art random swap EM (RSEM).

Keywords: Gaussian mixture models, differential evolution, expectation maximization, model-based clustering

ESTYMACJA PARAMETRÓW MODELI MIESZANIN ROZKŁADÓW NORMALNYCH PRZY POMOCY METODY HYBRYDOWEJ ŁĄCZĄCEJ SAMOADAPTACYJNĄ EWOLUCJĘ RÓŻNICOWĄ Z ALGORYTMEM EM

Streszczenie: W pracy poruszono problem uczenia modeli mieszanin rozkładów normalnych. Zaproponowano nowe podejście, nazwane DE-EM, oparte na hybrydyzacji samoadaptacyjnego algorytmu ewolucji różnicowej i klasycznego algorytmu EM. W nowej metodzie rozwiązanie otrzymane jako wynik operatorów mutacji i krzyżowania jest poddawane optymalizacji lokalnej, prowadzonej aż do momentu uzyskania zbieżności, przez algorytm EM. Aby uniknąć problemu z reprezentacją macierzy kowariancji i niedopuszczalnością rozwiązań użyto metody, w której macierze kowariancji są kodowane przy pomocy dekompozycji Cholesky'ego. W badaniach symulacyjnych modele mieszanin rozkładów normalnych zastosowano do grupowania danych syntetycznych. Wyniki eksperymentów wskazują, że metoda DE-EM osiąga lepsze wyniki niż standardowa technika wielokrotnego startu algorytmu EM. Dla zbiorów danych z dużą liczbą cech, metoda osiąga lepsze wyniki niż technika losowej wymiany rozwiązań połączona z algorytmem EM.

Słowa kluczowe: Mieszanki rozkładów normalnych, ewolucja różnicowa, algorytm EM, grupowanie danych

Artykuł zrealizowano w ramach pracy badawczej S/WI/2/2013.