

Адаптивное определение параметров алгоритма кластеризации потоков данных, взвешенных по времени.

НИССЕНБАУМ ОЛЬГА ВЛАДИМИРОВНА

ГОУ ВПО "Тюменский государственный университет" (Тюмень), Россия
e-mail: onissenbaum@rambler.ru

ХАРЧЕНКО АНАСТАСИЯ МИХАЙЛОВНА

ГОУ ВПО "Тюменский государственный университет" (Тюмень), Россия

На основании динамического EM-алгоритма [1,2] построен алгоритм кластеризации для потока данных, взвешенных по времени поступления [3]. Алгоритм предназначен для кластеризации данных с нормальным распределением в \mathbf{R}^n , параметры которого изменяются во времени, что соответствует ситуации в реальных динамических системах, таких как компьютерные системы, сети связи и т.п. Хранения обработанных данных не требуется, алгоритм эффективно вычислим, может применяться в системах реального времени.

В качестве весовой функции данных от времени t , прошедшего с момента поступления, выбрана экспоненциальная функция $w(t) = e^{-at}$, $t \geq 0$, обеспечивающая эффективно вычислимый пересчет параметров кластеров в момент поступления новой точки, накопление весов точек в параметрах кластера и, как следствие, отсутствие необходимости хранения обработанных точек.

Получен адаптивный эвристический алгоритм определения параметра весовой функции a .

Проведен ряд экспериментов на имитационной модели потока с нормальной плотностью распределения кластеров с изменяющимися параметрами распределения (среднее и ковариационная матрица), с целью разработанного алгоритма.

Список литературы

1. J. Chandrika, K.R. Ananda Kumar. Dynamic Clustering Of High Speed Data Streams // International J. of Comp. Sc. Issues. Vol. 9, No.1, pp.224-228, 2012.
2. Mingzhou Song, Hongbin Wang. Highly efficient incremental estimation of Gaussian mixture models for online data stream clustering // Proceedings of SPIE 5803. – 2005. – P.174-183.
3. О.В. Ниссенбаум. Алгоритм кластеризации потока данных с изменяющимися параметрами распределения // Вестник ТюмГУ –2013. – №6. – С.180-186.