

# Weighted K-Means for Density-Biased Clustering

Kittisak Kerdprasop<sup>1</sup>, Nittaya Kerdprasop<sup>1</sup>, and Pairote Sattayatham<sup>2</sup>

<sup>1</sup> Data Engineering and Knowledge Discovery Research Unit,  
School of Computer Engineering, Suranaree University of Technology,  
111 University Avenue, Nakhon Ratchasima 30000, Thailand  
{kerdpras, nittaya}@sut.ac.th

<http://www.sut.ac.th/engineering/computer/faculty/nittaya>

<sup>2</sup> School of Mathematics, Suranaree University of Technology  
111 University Avenue, Nakhon Ratchasima 30000, Thailand  
pairote@sut.ac.th

**Abstract.** Clustering is a task of grouping data based on similarity. A popular k-means algorithm groups data by firstly assigning all data points to the closest clusters, then determining the cluster means. The algorithm repeats these two steps until it has converged. We propose a variation called weighted k-means to improve the clustering scalability. To speed up the clustering process, we develop the reservoir-biased sampling as an efficient data reduction technique since it performs a single scan over a data set. Our algorithm has been designed to group data of mixture models. We present an experimental evaluation of the proposed method.