



## Conversation Topics

## Current Topics

## Clustering

— Curated by: Ian Davidson

Cluster analysis or clustering aims at grouping a number of different groups such that objects within each group are similar to each other and dis-similar to those in other groups. Domains including image analysis and social network analysis traditionally inherently explore clusters to uncover the underlying structure of the data. Supervised (semi-supervised) clustering extensions to handle complex data.

[Read More](#)

## Frequent Pattern Mining

— Curated by: Xifeng Yan

Finding frequent patterns plays a central role in many other interesting relations such as classification, clustering, and other data mining tasks. An important data mining task and a large literature has been dedicated to mining frequent patterns.

## Clustering x +



## Title &amp; Authors ORGANIZERS

## Efficient Frequent Directions Algorithm

Author(s): Mina Ghashami\*, University of Utah

## AnyDBC: An Efficient Anytime Dense Clustering

Author(s): Son Mai\*, Aarhus University; Ira Averbach

## City-Scale Map Creation and Update

Author(s): Chen Chen\*, Stanford University; Christos Gunopulos; Leonidas Guibas, Stanford University

## A Text Clustering Algorithm Using a

Author(s): Jianhua Yin\*, Tsinghua University;

## Batch model for batched timestamps

Author(s): Qingqi Yue\*, NIH; Ao Yuan, NIH; Xifeng Yan

## Structured Doubly Stochastic Matrix

Author(s): Xiaoqian Wang, Univ. of Texas at Arlington

## Infinite Ensemble for Image Clustering

Author(s): Hongfu Liu\*, Northeastern University; Yun Fu, Northeastern University

## Data-driven Automatic Treatment Recommendation

Author(s): Leilei Sun\*, Dalian University of Technology; Yanming Xie,

[submit](#)

ORGANIZERS

[Home](#) / [Topics](#) / [Clustering](#)

## Efficient Frequent Directions Algorithm for Sparse Matrices

Mina Ghashami\*, University of Utah; Edo Liberty, Yahoo; Jeff Phillips, School of Computing, University of Utah

## Abstract

This paper describes Sparse Frequent Directions, a variant of Frequent Directions for sketching sparse matrices. It resembles the original algorithm in many ways: both receive the rows of an input matrix  $A \in \mathbb{R}^{n \times d}$  one by one in the streaming setting and compute a small sketch  $B \in \mathbb{R}^{n \times k}$ . Both share the same strong (provably optimal) asymptotic guarantees with respect to the space-accuracy tradeoff in the streaming setting. However, unlike Frequent Directions which runs in  $O(nd^2)$  time regardless of the sparsity of the input matrix  $A$ , Sparse Frequent Directions runs in  $O(n \cdot \text{nnz}(A) + n^2 k)$  time. Our analysis loosens the dependence on computing the Singular Value Decomposition (SVD) as a black box within the Frequent Directions algorithm. Our bounds require recent results on the properties of fast approximate SVD computations. Finally, we empirically demonstrate that these asymptotic improvements are practical and significant on real and synthetic data.

[Download PDF](#)Filed under: [Clustering](#) | [Frequent Pattern Mining](#) | [Big Data](#) | [Time Series and Stream Mining](#)

Texto

Texto

Tópicos