

**姓 名 ：高崇铭**

**指导老师：姓名 邵俊明 职称 教授**

**毕业去向：本校计算机攻读硕士学位**

**社会活动：英才1班学习委员，带领班级拿获得2014学年“校优秀班集体”荣誉，并获得2014~2015年“先进团支部”荣誉。**

**主要获奖情况：**

**全国大学生数学建模竞赛省一****等奖，美国大学生数学建模竞赛****M奖(一等奖)，ACM西南地区高校邀请赛三等奖(2次)，****电子科技大学2014~2015年度优秀团员**

**科技项目：于数据挖掘实验室完成“基于Lucene的全文搜索引擎开发”。**

**毕业论文题目： 基于同步原理的多边聚类算法研究**

**中文简介：**

近年来，双边聚类成为了生物信息学、文本挖掘和推荐系统里广受关注的问题，越来越多的研究工作围绕其展开。目前存在的相关算法大多以启发式搜索的方式来找寻数据矩阵中的联合簇，搜索指标的选取决定了这类算法的性能。在本文中，我们提出了一种全新的基于同步思想的双边聚类算法CoSync，其工作方式为用动态交互的方式来自动第发掘嵌入数据矩阵中的联合簇。CoSync算法的核心思想是将数据矩阵视为一个动态系统，矩阵中每一个元素的值都将受到其邻居元素值的加权影响，随着迭代属于同一联合簇子矩阵的元素值终将同步到一起，构成一个常数值子块。我们将展示CoSync吸引人的优势：(a) CoSync可以根据数据内在的结果，自动识别出高质量的联合簇。(b) CoSync对联合簇在数据矩阵中的位置分布没有限制，其可以发掘现实数据中分布复杂的联合簇。(c) 加入非负矩阵分解的模块后，CoSync可以对任意高维数据矩阵进行分析。最后，实验证明我们的算法能够成功发掘出数据集中高质量的联合簇，且性能超过国际上有代表性的其他算法。

关键词：双边聚类，同步，基因表达数据

**英文简介：**

Co-clustering has gained growing attentions recently due to its wide practical applications in biological data analysis, text mining and recommender systems. Most existing co-clustering algorithms usually search co-clusters by heuristic searching algorithm, the performance depends on the choosing of criteria. In this paper, we propose a new synchronization-inspired co-clustering algorithm by dynamic simulation, called CoSync, which aims at discovering all co-clusters embedding in a given gene expression data matrix. The basic idea is to view the data matrix as a dynamical system, and the weighted two-sided interactions are imposed on each entry of the matrix from both aspects of rows and columns, resulting in the values of all entries in a co-cluster synchronizing together. We demonstrate that our new co-clustering approach has several attractive benefits: (a) CoSync is capable of identifying co-clusters with high-quality, driven by the intrinsic data structure. (b) Without any co-cluster structure assumption, CoSync supports finding co-clusters of arbitrary size, not limited to disjoint co-clusters. (c) In conjunction with non-negative matrix factorization, CoSync allows analyzing large-scale data. Experiments show that our algorithm faithfully uncovers co-clusters embedded in data sets and has good performance compared to state-of-the-art algorithms.

Keywords: Co-clustering, Synchronization, Gene Expression Data