Principal Investigator: Zha, Hongyuan
Organization: GA Tech Res Corp - GIT
Submitted By:

Title:
Matrix Algorithms for Data Clustering and Nonlinear Dimension Reduction

Project Participants

Senior Personnel
Name: Zha, Hongyuan
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student
Name: Zhang, Ming
Worked for more than 160 Hours: Yes
Contribution to Project:
Ming Zhang carried out numerical experiments for the project

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Activities and Findings

Research and Education Activities:
Major research and education activities:
Our research focuses on simultaneous clustering algorithms
and their applications in bioinformatics and information retrieval.
In particular, we proposed methods for learning ranking functions
by exploring the difference in query distributions (ref. 1). We also proposed robust algorithms for PCA and clustering by incorporating
rotational invariance (ref. 2). We applied simultaneous clustering methods for document segmentation (ref. 3, ref. 4).

During this period, I moved to College of Computing at Georgia Tech where I developed a new course on 'Web search and text mining'
graduate course. I also further investigate applications of clustering and dimension reduction methods for citation analysis and document recommendations (ref. 5 and ref. 6).


Findings:
Implicitly clustering queries into different types can significantly affect the performance of the learned ranking. But at which granularity to cluster the queries is still an open problem that deserves more in-depth research.

Training and Development:
Jiang Bian is a graduate student supported by this grant. Hongyuan Zha also developed a graduate course 'Web search and text mining' at Georgia Tech.

Outreach Activities:

Journal Publications


Books or Other One-time Publications

Editor(s): ICDM
Collection: IEEE International Conference on Data Mining
Bibliography: IEEE Press

Editor(s): World Wide Web
Collection: Proceedings of World Wide Web Conference
Bibliography: World Wide Web

Web/Internet Site

URL(s):
http://www.cse.psu.edu/~zha/papers.html
Description:

Other Specific Products

Contributions within Discipline:

Implicitly clustering queries into different types can significantly affect the performance of the learned ranking. But at which granularity to cluster the queries is still an open problem that deserves more in-depth research.

Contributions to Other Disciplines:
The applications of clustering to better organize and improve retrieval relevance.

Contributions to Human Resource Development:
Jiang Bian is supported by the grant as a graduate research assistant.

Contributions to Resources for Research and Education:
Our research results are being incorporated into a graduate course on web search and text mining
Contributions Beyond Science and Engineering:

Categories for which nothing is reported:

Organizational Partners
Activities and Findings: Any Outreach Activities
Any Product
Contributions: To Any Beyond Science and Engineering