Final Report for Period: 09/2009 - 08/2010
Submitted on: 11/01/2010
Principal Investigator: Zha, Hongyuan .
Award ID: 0736328
Organization: GA Tech Res Corp - GIT
Submitted By:
Zha, Hongyuan - Principal Investigator

Title:
Computational Methods for Nonlinear Dimension Reduction

Project Participants

Senior Personnel

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

Name: Park, Haesun
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Yang, Shuang-Hong
Worked for more than 160 Hours: No
Contribution to Project:

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Activities and Findings

Research and Education Activities:
In this project we focus on computational and statistical methods for nonlinear dimension reduction problems. Our emphasis is placed on numerical stability analysis as well as statistical consistency of manifold learning algorithms, visualization problems involving nonlinear dimension reduction, and applications of nonlinear dimension methods in

several challenging application areas.

We organized seminar talks and discussion groups for graduate students to gain hands-on experience in problem formulation, data collection, research methodology and experimental confirmation. We have also encouraged students to work as interns in leading companies to work on challenging real-world problems.

Three graduate students, Ming Zhang, Andrew Smith and Shuang-Hong Yang has been supported at various time by this grant.

Findings: (See PDF version submitted by PI at the end of the report)

In a CVPR 2008 paper, we formulate semi-supervised manifold learning as an eigenvalue problem, and proposed numerical algorithm for its solutions. We especially investigated various parameter settings that determine the trade-offs between preserving the manifold structures and a good fit to the labeled data.

In a NIPS 2008 paper, we discuss the statistical consistency problem of LTSA and provide perturbation bounds that can be used to derive
consistency and convergence rate. One key problem is still open, i.e., the estimate of the smallest nonzero eigenvalue of the alignment matrix.

In a SIAM Review 2009 paper, we analyze the numerical stability issues of manifold learning algorithms. We propose geometric notions to characterize conditions under which the underlying nonlinear parametrization can be recovered.

In a SIGMM 2009 paper, we propose how to learn a good distance metric for the input data

Training and Development:

We emphasize hands-on research experience for advanced graduate students. We organized research seminars and discussions groups to help graduate students to gain experience in research problem formulation, data collection, research methodology, experimental confirmation, and the actual writing of research papers. We encourage students to pay special attention to feedbacks from reviewers of research papers, and careful consideration of reviewer feedbacks.

The three students supported all have done internships at various companies. Particularly, Shuang-Hong Yang did internship with Yahoo! and Microsoft.

Outreach Activities:

Journal Publications

Hongyuan Zha and Zhenyue Zhang, "Spectral Properties of the Alignment Matrices in Manifold Learning", SIAM Review, p. 545, vol. 51, (2009). Published,

Shuang-Hong Yang, Hongyuan Zha, S. Kevin Zhou and Bao-Gang Hu, "Variational Graph Embedding for Globally and Locally Consistent Feature Extraction", Proceedings of ECML/PKDD, p. , vol. , (2009). Published,

Bo Xiao, Xiaokang Yang, Yi Xu and Hongyuan Zha, "Learning Distance Metric for Regression by Semidefinite Programming with Application to Human Age Estimation", Proceedings of ACM Multimedia Conference, p. , vol. , (2009). Published,

Books or Other One-time Publications

Zhengyue Zhang, Hongyuan Zha, and Ming Zhang, "Spectral Methods for Semi-supervised Manifold Learning", (2008). Conference Proceedings, Published
Collection: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008
Bibliography: 2008

Andrew Smith, Xiaoming Huo, and Hongyuan Zha, "Convergence and Rate of Convergence of A Manifold-Based Dimension Reduction Method", (2008). Book, Accepted
Collection: Proceedings of Neural Information Processing Systems, NIPS, 2008
Bibliography: 2008

Web/Internet Site

Other Specific Products

Contributions

Contributions within Discipline:

We developed new methods for analyzing numerical stability and statistical consistency for manifold learning algorithms. We developed new algorithms for metric learning and applied it to human age estimation. We also developed algorithm for feature selection using graph embedding. co-PI Haesun Park investigated algorithms for visualization using Isomap.

Contributions to Other Disciplines:

Algorithms we developed have been applied to multimedia information processing, machine learning, and visualization.

Contributions to Human Resource Development:

Three graduate students have been supported by the project. One student graduated with a MS degree in CSE and is now working for a software company. One student graduated with PhD in ISYE and is now working for a national bank. The third student is still pursuing hist Ph.D degree.

Contributions to Resources for Research and Education:

Contributions Beyond Science and Engineering:

Conference Proceedings

Categories for which nothing is reported:

Organizational Partners

Activities and Findings: Any Outreach Activities

Any Web/Internet Site

Any Product

Contributions: To Any Resources for Research and Education

Contributions: To Any Beyond Science and Engineering

Any Conference

In a CVPR 2008 paper, we formulate semi-supervised manifold learning as an eigenvalue problem, and proposed numerical algorithm for its solutions. We especially investigated various parameter settings that determine the trade-offs between preserving the manifold structures and a good fit to the labeled data.

In a NIPS 2008 paper, we discuss the statistical consistency problem of LTSA and provide perturbation bounds that can be used to derive consistency and convergence rate. One key problem is still open, i.e., the estimation of the smallest nonzero eigenvalue of the alignment matrix.

In a SIAM Review 2009 paper, we analyze the numerical stability issues of manifold learning algorithms. We propose geometric notions to characterize conditions under which the underlying nonlinear parametrization can be recovered.

In a SIGMM 2009 paper, we propose how to learn a good distance metric for the input data1 that is crucial in many pattern recognition and machine learning applications. The novelty of our work is investigating the problem of learning a distance metric that measures the semantic similarity of input data for regression problems. The particular application we consider is human age estimation.

In an ECML/PKDD 2009 paper, we observed that most of the existing feature extraction methods explore either global statistical or local geometric information underlying the data. In our work, we propose an approach that combines both for constructing features based on convex optimization of nonparametric learning criteria. Using mutual information and Bayes error rate as example criteria, we show that our approach can be compactly formalized as linear embedding of well-defined data graphs and efficiently solved by spectral analysis, leading to supervised locality-preserving feature extractors with several appealing properties. Experiments on benchmark face recognition data sets confirm the effectiveness of our proposed algorithms.

In co-PI Haesun Park's recent work, ISOMAP, one of the widely-used nonlinear data embedding methods based on manifold learning, is discussed from the viewpoint of visualization. It has a parameter that controls the number of edges in a neighborhood graph. If the parameter is chosen to be too small, a neighborhood graph fails to contain all the data points. If it is too large, so called a short-circuiting problem occurs. However, it is often difficult to find a suitable parameter value for the data at hand. One might want to optimize it based on certain quantitative criteria, but it may take significant time to repetitively run ISOMAP with different parameter values. Furthermore, when using ISOMAP for data visualization in 2D or 3D, users might want to try ISOMAP with different parameter values to see if they can obtain various insights about the data. Interactions between human and such visualizations require reasonably efficient updating even for1 the large-scale data. To tackle these problems, we proposed an efficient updating algorithm for ISOMAP due to parameter changes, which we call p-ISOMAP. We performed complexity analysis and empirical running times of p-ISOMAP, and also showed interesting

visualization applications of p- ISOMAP to real-world data such as image and text data.