High order data analytics through Hypergraphs and Tensor Decomposition

Ahmed Elgammal
I/UCRC Center for Dynamic Data Analytics
Rutgers University

• Traditional data analysis is dominated by pairwise metrics between data entries
  – Clustering
  – Classification
  – Regression
  – ...

• In many applications, besides pairwise metrics, relations between triplets, quadruplets, and higher order subsets of data can be measured.

• Information in terms of groups of data
High order data analytics denotes the analysis of data where metric relations between tuples or groups of data can be measured.

“Information in terms of groups of data”

How can we do:
- Clustering
- Classification
- Regression
- ...

What models we can use for this data

Important role for Hypergraphs in high order data analytics.

Most known algorithms maps the problem to a traditional pairwise metric problem.
• In 1762 Joseph Louis Lagrange formulated what is now known as the eigenvalue-eigenvector problem.
• Significant importance in data analytics in all domains.

• High order data relations (and Hypergraphs) can be modeled using high order matrices (tensors)

• How to generalize the eigenvalue problem to tensors?
• Spectral theory for tensors?
• Singular Value Decomposition for tensors?
• Tensor Decomposition?

Previous Related Work
• Several Tensor decomposition framework exists:
  – Canonical and Parallel Factor decomposition [1970]
  – Tucker Decomposition [Tucker, L.R. 1966]
  – High order SVD (HOSVD) [Lathauwer, L. et al 2000]
• HOSVD: of interest in many domains recently in data mining, web search, DNA microarray analysis, computer vision, and graphics.
• Not really a generalization of the spectral theory
• Many Limitations: optimality, uniqueness, ...

• Generalized Spectral Theory for Tensors: [in collaboration with Edinah Gnang and Vladimir Retakh]
• Factorize a tensor into a “product” of a diagonal and orthogonal tensors

• Consistent multilinear algebra
• High order product of tensors (ternary products,...)
• Tensor orthogonality, diagonality, transpose, determinant, trace, ...
• Generalization to any order tensor
Proposed Work

Study different high order data analytics problems and what tensor spectrum means for them

Umbrella of projects:
• Image and Video Ensemble analysis.
• Object recognition from 2D and 3D data
• Web search

Implication of Project

• Many Potential Applications
• Implication to Industry:
  A new fundamental tool for high order data analytics
Project Name: High order data analytics through Hypergraphs and Tensor Decomposition
Project Investigators: Dr. Elgammal (PI), Edinah Gnang (students)

Description:
The project investigates the application of a novel mathematical model for high order data analytics. Traditional data analysis and learning theory is heavily dominated by pairwise metrics between data entries. In many important applications, besides pairwise metrics, relations between triplets, quadruplets, and higher order subsets of data can be measured. High order data analysis denotes the analysis of data where metric relations between tuples or groups of data can be measured. Such high order data relations can be mathematically modeled through high order matrices, also known as tensors. In this project we will investigate the application of a novel tensor decomposition framework that we developed in different data analytics domains. Our proposed factorization decomposes a symmetric tensor into a product of an orthogonal and a diagonal tensor. We will investigate applications in image and video ensemble analysis and object recognition from 2D and 3D data.

Experimental Plan:
- Fall 2010: Data collection, algorithm development
- Spring 2011: Testing of algorithms
- Summer 2011: Performance evaluation

Related Work Elsewhere:
- Most known algorithms that deal with high order data mainly try to map the problem to a traditional pairwise metric problem. Available tensor decomposition frameworks do not offer a factorization in terms of products of tensors.

How Ours is Different:
- We can deal with high order relations without losing information by mapping to pairwise metrics.

Related Work in Center:
- Tensor decomposition for hypergraph problems
- Application to human motion analysis

Milestones:
-2010-2011: High order data analytics for image ensemble analytics
-2011:

Deliverables:
- Technical demonstration along with a technical report resulting in a publication.

Budget: $60,000

Potential Benefits to Member Companies:
- A novel fundamental tool to handle high order data analytics in many domains.