

DOCTORAL THESIS

Studies on probabilistic tensor subspace learning

Zhou, Yang

Date of Award:
2019

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

Most real-world data such as images and videos are naturally organized as tensors, and often have high dimensionality. Tensor subspace learning is a fundamental problem that aims at finding low-dimensional representations from tensors while preserving their intrinsic characteristics. By dealing with tensors in the learned subspace, subsequent tasks such as clustering, classification, visualization, and interpretation can be greatly facilitated. This thesis studies the tensor subspace learning problem from a generative perspective, and proposes four probabilistic methods that generalize the ideas of classical subspace learning techniques for tensor analysis.

Probabilistic Rank-One Tensor Analysis (PROTA) generalizes probabilistic principle component analysis. It is flexible in capturing data characteristics, and avoids rotational ambiguity. For robustness against overfitting, concurrent regularizations are further proposed to concurrently and coherently penalize the whole subspace, so that unnecessary scale restrictions can be relaxed in regularizing PROTA.

Probabilistic Rank-One Discriminant Analysis (PRODA) is a bilinear generalization of probabilistic linear discriminant analysis. It learns a discriminative subspace by representing each observation as a linear combination of collective and individual rank-one matrices. This provides PRODA with both the expressiveness of capturing discriminative features and non-discriminative noise, and the capability of exploiting the (2D) tensor structures.

Bilinear Probabilistic Canonical Correlation Analysis (BPCCA) generalizes probabilistic canonical correlation analysis for learning correlations between two sets of matrices. It is built on a hybrid Tucker model in which the two-view matrices are

combined in two stages via matrix-based and vector-based concatenations, respectively. This enables BPCCA to capture two-view correlations without breaking the matrix structures.

Bayesian Low-Tubal-Rank Tensor Factorization (BTRTF) is a fully Bayesian treatment of robust principle component analysis for recovering tensors corrupted with gross outliers. It is based on the recently proposed tensor-SVD model, and has more expressive modeling power in characterizing tensors with certain orientation such as images and videos. A novel sparsity-inducing prior is also proposed to provide BTRTF with automatic determination of the tensor rank (subspace dimensionality).

Comprehensive validations and evaluations are carried out on both synthetic and real-world datasets. Empirical studies on parameter sensitivities and convergence properties are also provided. Experimental results show that the proposed methods achieve the best overall performance in various applications such as face recognition, photograph-sketch match, and background modeling.

Keywords: Tensor subspace learning, probabilistic models, Bayesian inference, tensor decomposition.

Table of Contents

DECLARATION	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	xi
Chapter 1 Introduction	1
1.1 Background	1
1.2 Tensor Subspace Learning	3
1.3 Challenges: Probabilistic Approaches for Tensor Subspace Learning	4
1.4 Contributions	7
1.5 Organizations	8
Chapter 2 Fundamentals of Tensor Analysis	11
2.1 Mathematical Background	11
2.1.1 Notations	11
2.1.2 Multilinear Basics	12
2.1.3 Matrix-Variate Normal Distribution	12
2.1.4 Discrete Fourier Transformation	13

2.1.5	Tensor Product	14
2.2	Tensor Factorization Models	15
2.2.1	Tucker Decomposition	16
2.2.2	CP Decomposition	16
2.2.3	Tensor Singular Value Decomposition	16
Chapter 3 Probabilistic Rank-One Tensor Analysis		18
3.1	Introduction	18
3.2	Related Works	22
3.2.1	Probabilistic PCA	22
3.2.2	Tucker-Based Multilinear PPCAs	22
3.2.3	CP-Based Multilinear PPCAs	23
3.3	Probabilistic Rank-One Tensor Analysis	23
3.3.1	The PROTA Model	23
3.3.2	Connections with Existing PPCAs	25
3.3.3	ECM Algorithm for PROTA	28
3.3.4	Concurrent Regularizations for CP-Based PPCAs	29
3.3.5	PROTA with Bayesian CR	32
3.3.6	Algorithmic Issues	36
3.4	Experiments	37
3.4.1	Subspace Estimation on Synthetic Data	37
3.4.2	Classification on 2D Images	39
3.4.3	Classification on 3D Sequences	43
3.4.4	Parameter Sensitivity and Convergence Study	45
3.5	Summary	49
Chapter 4 Probabilistic Rank-One Discriminant Analysis		50
4.1	Introduction	50
4.2	Related Works	53
4.2.1	Linear Discriminant Analysis	53

4.2.2	Probabilistic LDA	54
4.2.3	Multilinear LDA	55
4.3	The Proposed Method	58
4.3.1	PRODA Model	59
4.3.2	Parameter Estimation for PRODA	61
4.3.3	Initialization and Prediction	64
4.3.4	Time Complexity Analysis	65
4.4	Experiments	66
4.4.1	Experimental Settings	67
4.4.2	Face Recognition	68
4.4.3	Facial Photograph-Sketch Matching	71
4.4.4	Convergence and Parameter Sensitivity	73
4.5	Summary	76
Chapter 5 Bilinear Probabilistic Canonical Correlation Analysis		77
5.1	Introduction	77
5.2	Related Works	79
5.3	Bilinear Probabilistic CCA	81
5.4	Experiments	88
5.5	Summary	94
Chapter 6 Bayesian Low-Tubal-Rank Tensor Factorization		95
6.1	Introduction	95
6.2	Bayesian Low-Tubal-Rank Robust Tensor Factorization	101
6.2.1	Model Specification	101
6.2.2	Variational Inference	103
6.2.3	Efficient Updates in Frequency Domain	107
6.2.4	Multi-Rank Prior	108
6.2.5	Initialization	112
6.3	Experiments	113

6.3.1	Validation on Synthetic Data	113
6.3.2	Image Denoising	115
6.3.3	Background Modeling	118
6.4	Summary	123
Chapter 7 Conclusion		124
7.1	Major Contributions	124
7.2	Future Work	125
Bibliography		127
CURRICULUM VITAE		144