A Survey of Partial Least Squares

Yun-Hao Yuan, Jin Li, Yun Li, Jipeng Qiang, Yi Zhu

Abstract—Recently, partial least squares (PLS) has become a widely concerned technique for multiview data analysis. PLS is a classic tool for analyzing relationship between two sets of variables, which has been widely used in fields of chemometrics, bioinformatics and pattern recognition. During the last few decades, numerous extensions of PLS model have been proposed to deal with different kinds of problems. However, the literature still needs a most recent and comprehensive overview for PLS and its related variants. In this paper, we provide a detailed and well-organized overview of PLS models following such taxonomy, *i.e.*, 1) basic PLS, 2) orthonormalized PLS, 3) nonlinear PLS, 4) multiset PLS, 5) sparse PLS, and 6) probabilistic PLS. The introduction of each group of PLS is demonstrated with several representative methods, followed by discussion about characteristics and limitations. In addition, we introduce popular applications of PLS variants and gather some quantitative results of representative methods to demonstrate their effectiveness. Several potential research directions for future development of PLS are provided in the end.

Index Terms—Partial Least Squares, Survey, Machine Learning.

1 Introduction

In recent years, data collected from different sources, also known as multiview data, are increasingly available in many real-world applications. Different from single view data, multiview data are usually associated with complementary information between different views. Hence, it is promising to explore comprehensive relationship from multiview data for better learning performance. However, associations between different views in practice are usually entangled in high-dimensional spaces, thus making direct relation measurement impractical. As a classic yet powerful multivariate statistical approach, partial least squares (PLS) overcomes this obstacle by extracting latent representation and modeling relations in a reduced feature space. The resulting promising performance makes PLS a popular solution for multiview data analysis (MDA).

PLS was first proposed by Herman Wold [1], [2] in the 1970s to reveal relations between different sets of latent variables during econometric path-modeling. In the 1980s, it was subsequently developed as a regression tool [3], as we called PLS regression (PLSR) [4], [5], to estimate efficient and stable predictive relations between two sets of data. Specifically, PLSR projects both original predictors and responses onto some uncorrelated components, called latent features, and then applies regression onto these components. In this way, PLSR shows advantages over other classic regression methods (*e.g.*, ordinary least squares) in robustness to the collinearity problem [3] or the singularity problem brought by small sample size (SSS). In the 1990s, another popular form of PLS, termed simultaneous PLS (sPLS) in this paper, has been introduced [6], [7] for more flexibility in analyzing brain imaging data. Owing to high expressive power and robustness, PLS has achieved great success in a wide range of applications, such as, chemometrics [8], [9], [10], neural image analysis [11], [12], [13], face recognition [14], [15], object detection [16] and object tracking [17].

Apart from the ordinary PLS, numerous variants of PLS have been developed during the past decades. In order to well-organize and introduce the panorama of the whole family of PLS, in this paper, we categorize them with the following taxonomy, *i.e.*, orthonormalized PLS, nonlinear PLS, multiset PLS, sparse PLS and probabilistic PLS.

Orthonormalized PLS (OPLS) was first developed by Worsley *et al.* [18] to preserve the invariance to arbitrary linear transformations on predictors. Later in 1999 [19], OPLS has been proven to be optimum for linear regression with bottleneck imposed on feature dimension. Such property leads to a wider range of attention [20], [21], [22], [23]. Nevertheless, conventional OPLS only seeks the projections of one set of data, which limits the applicability in ordinary MDA. In 2021, A more generalized OPLS [24] with projections on both two-view data is introduced for face hallucination. The orthogonality and uncorrelation of high-resolution projections are of great help in minimizing redundancy, and lead to competitive performance of face super-resolution. In addition, OPLS has a close connection to canonical correlation analysis (CCA), which is another popular statistical technique for correlation analysis. Sun *et al.* [25] investigated the underlying equivalence relationship between OPLS and CCA with detailed mathematical proof.

Nonlinear PLS is an important branch to overcome the handicap of linear PLS to reveal underlying nonlinear relations. Two kinds of nonlinear PLS are introduced in this paper, polynomial PLS and kernel PLS (KPLS). In 1989, Wold *et al.* developed quadratic PLS [26] to model the quadratic polynomial relation between latent features. Subsequently, a spline version [27] and a fuzzy version [28] of PLS were further proposed for more flexibility in practice. But still, polynomial PLS methods have limited nonlinear expressive power for their fixed models. In 2001, Rosipal successfully introduced kernel method into PLS [29]. With the kernel

- *Yun-Hao Yuan is with the School of Information Engineering, Yangzhou University, Yangzhou 225127, China and he is also with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: yhyuan@yzu.edu.cn).*
- *Jin Li, Yun Li, Jipeng Qiang, and Yi Zhu are with the School of Information Engineering, Yangzhou University, Yangzhou 225127, China (e-mail: cvjinli@outlook.com; liyun@yzu.edu.cn; jpqiang@yzu.edu.cn; z8d1177@126.com).*