总结报告19

论文A convex formulation for learning a shared predictive structure from multiple tasks

论文On better exploring and exploiting task relationships in multitask learning: Joint model and feature learning

（2020.2.29）

1. contents

Chen, J., Tang, L., Liu, J., & Ye, J. (2013). **A convex formulation for learning a shared predictive structure from multiple tasks**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5), 1025–1035.

Li, Y., Tian, X., Liu, T., & Tao, D. (2018). **On better exploring and exploiting task relationships in multitask learning: Joint model and feature learning**. IEEE Transactions on Neural Networks and Learning Systems, 29(5), 1975–1985.

Ando, R. K., & Zhang, T. (2005). **A framework for learning predictive structures from multiple tasks and unlabeled data**. Journal of Machine Learning Research, 6, 1817–1853.

Gonçalves, A. R., Von Zuben, F. J., & Banerjee, A. (2016). **Multi-task sparse structure learning with Gaussian copula models**. Journal of Machine Learning Research, 17, 1–30.

1. 论文1
2. 题目

Chen, J., Tang, L., Liu, J., & Ye, J. (2013). **A convex formulation for learning a shared predictive structure from multiple tasks**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5), 1025–1035.

1. Idea

In ASO, a separate linear classifier is trained for each of the tasks and dimension reduction is applied on the **predictor space**, finding low-dimensional structures with the highest predictive power.

1. 论文脉络

3.1 概述

从

原始Alternating structure optimization (ASO)

到

Improved ASO（iASO）🡪非凸

再到

凸松弛形式，但不适应大规模数据集（due to复杂的约束条件）

最后

优化算法（cASO）convex relaxation

3.2 符号说明

m tasks，

3.3 ASO

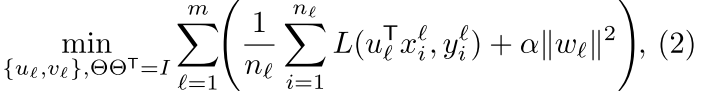
让Θ：orthonormal rowsΘΘT = I

Predictor：



：full feature space ：high-dimensional ：shared low-dimensional

ASO的目标函数：



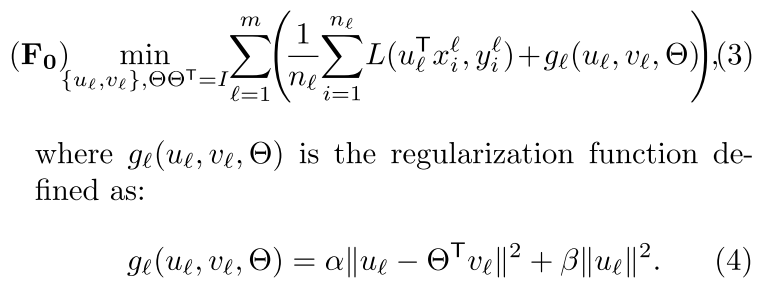
Note：

：控制task relatedness，



3.3 iASO（improved）是非凸的（记为F0、F1）

iASO目标函数：



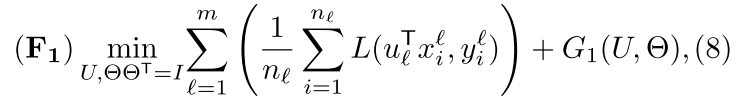
Note:

：控制task relatedness

：控制the complexity of the predictor functions

形式化简：

由，最终推导出



其中，



3.4 凸松弛

把F1的可行域转化为一个凸集，

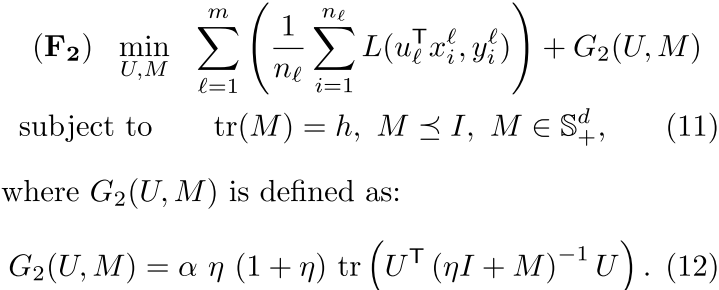
Me的凸包是Mc





将正交条件转换成关于M的约束，令ΘΘT=M

凸松弛目标函数：

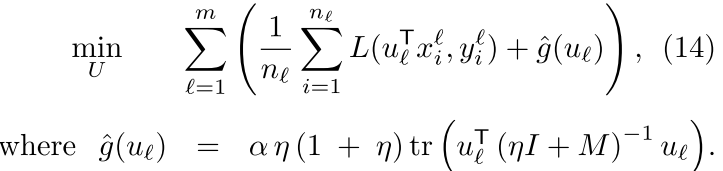


It is not scalable to high-dimensional data (with a large ) due to its positive semidefinite constraints.

3.5 cASO：Convex Alternating Structure Optimization

1) Computation of U for a Given M

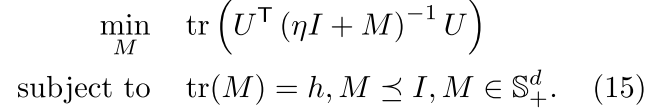
给定M，问题转化为，



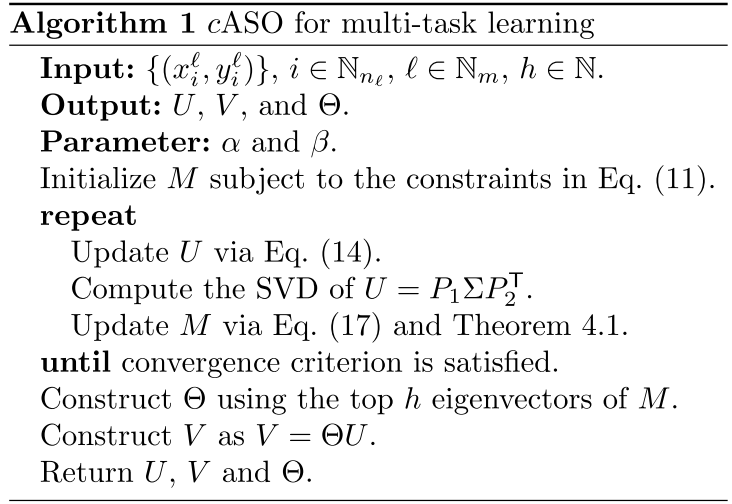
这是一个强凸问题。

2) Computation of M for a Given U

给定U，问题转化为，



论文中通过解决一个特征值优化问题（eigenvalue optimization problem），解决了（15）求得M的一个最优解。（具体详见论文4.2.1）



三、论文2

1. 题目

Li, Y., Tian, X., Liu, T., & Tao, D. (2018). **On better exploring and exploiting task relationships in multitask learning: Joint model and feature learning**. IEEE Transactions on Neural Networks and Learning Systems, 29(5), 1975–1985.

1. Idea

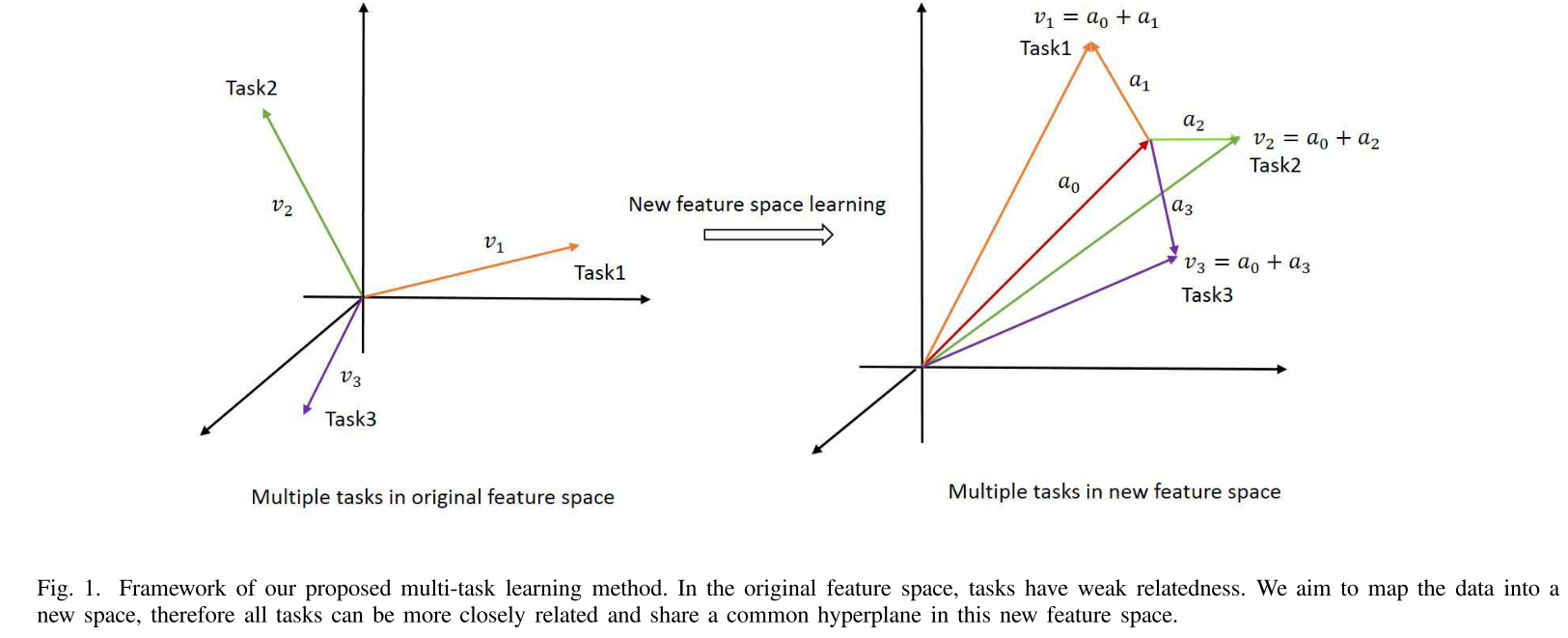
作者分类：测量多任务之间的关系（task relationship）有两种方法：

——shared parameters

——shared feature representations

作者认为之前的论文大多是上面两种方法中的一种来测量task relationship，但是这篇论文的创新之处就在于作者将**这两种方法结合起来**。

如下图所示：下图的核心在于右图的: a common hyperplane



1. 论文脉络
   1. 符号说明

T：different tasks

：the dataset of task t，每个task的examples为个，x是d维的

* 1. 概述

这篇论文的脉络思路为：

提出非凸目标函数🡪凸松弛为凸的目标函数🡪提出解决凸函数的算法

* 1. 非凸目标函数

Predictor：



令，



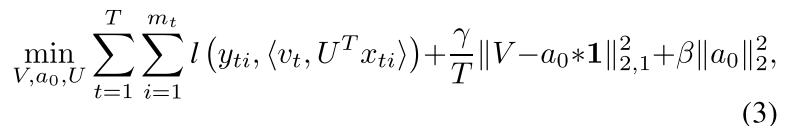
Note：

U：a feature mapping matrix，将x映射到另外一个超空间，在这个空间里，可能才能保证task之间是related。

a0：the shared central hyperplane in the new feature space

at： the offset of task t to maintain its own characteristic

非凸目标函数：

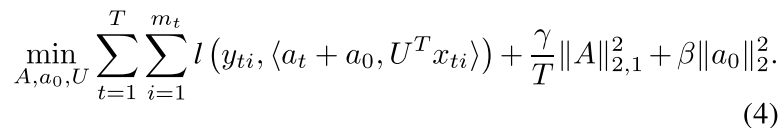


其中，



因为，，并且令，

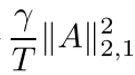
非凸目标函数最终表达形式：



Note：

：测量中心超平面的平滑度和复杂度

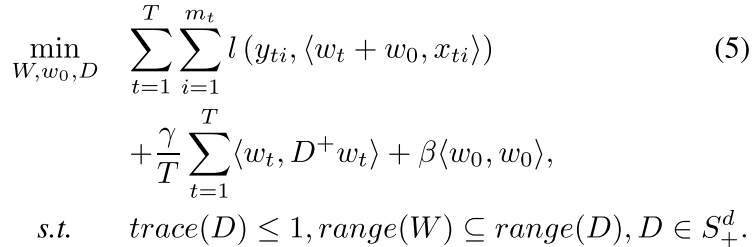


：保证all tasks share a subset of common features and the sparsity of shared features

这里的U是正交矩阵

* 1. 凸目标函数

目标函数的凸版本：



Note：

：D是半正定阵

：D的伪逆



：代表一个向量的集

Solution：

（5）凸解为  （4）非凸解为，

（4）非凸解为（5）凸解为表示一个对角阵，对角元素由a0的元素组成

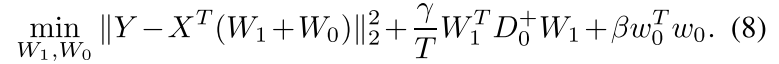
注：的列是一组标准正交基，的特征向量就是由这些正交基组成的。

* 1. 优化算法：交替更新和D

1. Optimate  by given D

·第一次符号的简化，

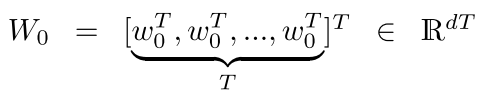
从（5）得到（8），注意此时min下面已经没有了D



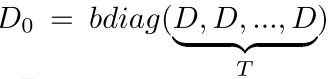
Note: 下面的符号就是将所有的矩阵都块对角化。

X所有T tasks的data matrix组成对角阵

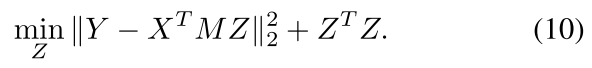
Y=



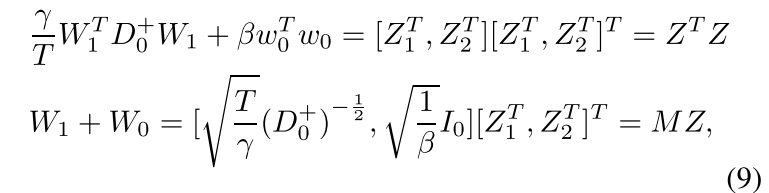




·第二次符号简化得到，



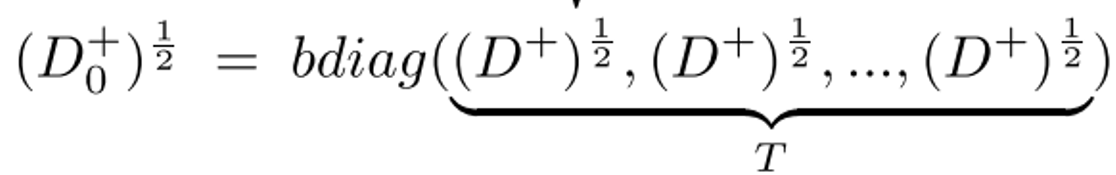
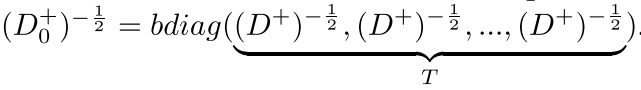
其中，





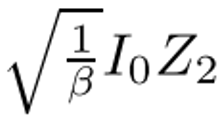




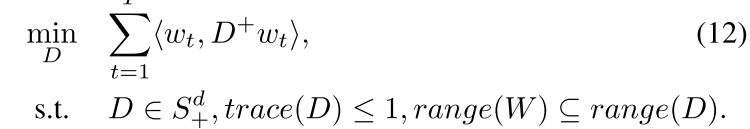
 其中，

Solution：



再利用，求解

1. Optimate D by given 



Solution：具体细节在[11]

