总结报告10

（2019.11.25——2019.12.1）

**一、多任务学习（论文：A survey on Multi-Task Learning）**

**1、为什么要研究多任务学习？**

个人理解：通常机器学习需要大量的带标签的数据或者无标签数据来进行训练，但是，在现实生活中我们通常难以获得大量的数据或者获得大量数据的成本较高。所以，在数据较少的情况下，就迫使人们逐渐地去探索更多的方法。受到生活经验的启发——类似一个会骑自行车的人，在学习开三轮车时通常会用到骑自行车的经验。于是，人们开始通过发掘任务之间的信息来弥补数据不足的缺陷，多任务学习应运而生。

**2、多任务学习的分类**

a. feature learning approach

1)feature transformation approach

2)feature selection approach

b. low-rank approach

c. task clustering approach

d. task relation learning approach

e. decomposition approach

**3、为了描绘这种关联性（relateness）有三个注意项：**

①when to share：To make choices between single-task and multi-task models for a multi-task problem.

②what to share：to determine the form through which knowledge sharing among all the tasks could occur.

share通常有三种形式：

1. **Feature-based MTL**

aims to learn **common features** among different tasks as a way to share knowledge.

1. **Instance-based MTL**

wants to identify useful data instances in a task for other tasks and then shares knowledge via the identified instances.

1. **Parameter-based MTL**

uses **model parameters** (e.g., coefficients in linear models) in a task to help learn model parameters in other tasks in some ways, for example, the regularization.

现有的模型通常都是**Feature-based MTL、Parameter-based MTL.**

1. how to share: specifies concrete ways to share knowledge among tasks.

**\*\*feature-based MTL:**

\*feature learning approach

**\*\*parameter-based MTL:**

\*low-rank approach

\*task clustering approach

\*task relation learning approach

\*decomposition approach

**4、feature learning approach**

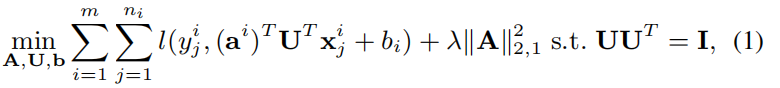
focuses on learning common **feature representations** for multiple tasks based on shallow or deep models.

**4.1 feature transformation approach**

——a transformation of the original feature representation

(如 多层前馈神经网络，output layer有m个输出)

**4.1.1 [11][12]的方法——MTFL：**



Note:

L( , ): loss function

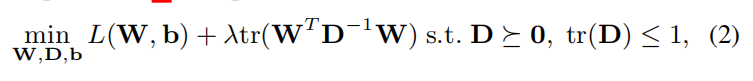
U: square transformation matrix 平方了的变换矩阵

A: 它的每一列表示每一个任务变换后的模型参数

2，1范数：保证A是稀疏的

s．t． ：保证U是正交的，以便U没有冗余，而不像多层前馈神经网络那样隐藏层可能冗余

上述问题（1）等价于



Note:

基于上述公式（2），to learn the feature covariance D for all the tasks.



 是任务i的模型参数



Given D, the learning of different tasks can be decoupled.

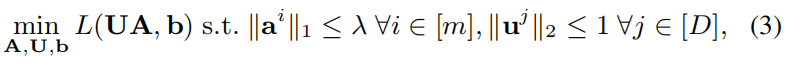
s．t.部分：D是半正定的

注：迹范数：矩阵特征值的和

核范数：矩阵奇异值的和

**4.1.2 [14] multi-task sparse coding method** ：

to learn a linear transformation on features

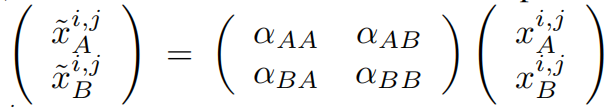


注意：Recently deep learning becomes popular due to its capacity to learn **nonlinear features** in many applications.

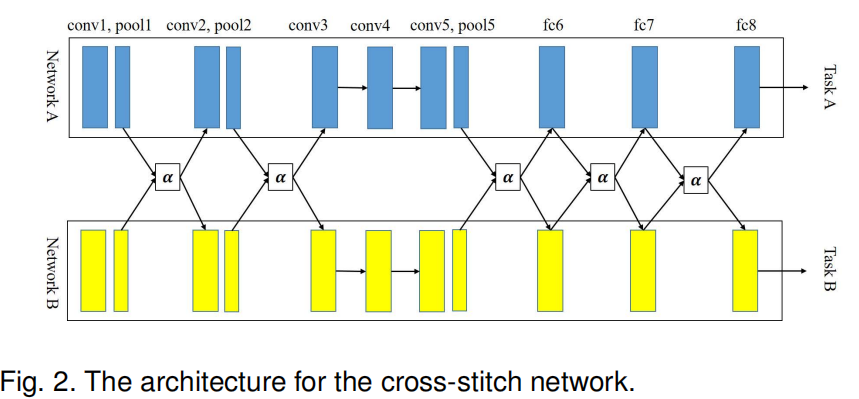
**4.1.3 [22] cross-stitch network** (deep model)

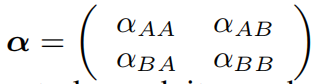
A and B: two tasks with an identical network architecture.

** :**the hidden feature outputted by the jth unit of the ith hidden layer for task A(B).

** :**cross-stitch operation on 

Note:andare new hidden features after learning the two tasks jointly.当 αAB 和 αBA 等于0, 联合训练两个网络相当于分别独立训练他们



Note:两个任务之间的联系，可以通过反向传播学得。

**4.2 feature selection approach**

——a subset of the original feature representation.

**5、low-rank approach**

任务之间有联系，就暗示着参数矩阵的低秩。

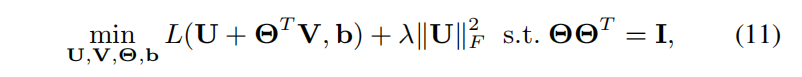
interprets the relatedness of multiple tasks as the **low rankness of the parameter matrix** **of these tasks.**

**5.1 [44] the model parameters of different tasks share a low-rank subspace**



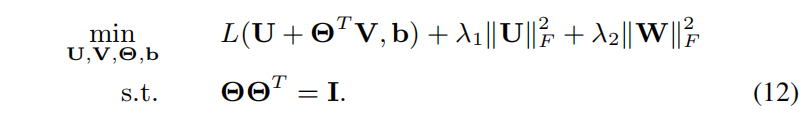
Note: 低秩子空间

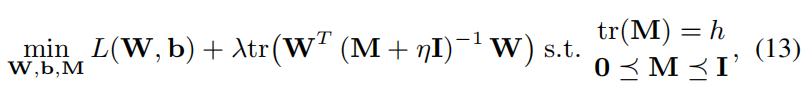
目标函数：



Note: 正交部分保证U无冗余

**5.2 [45]**

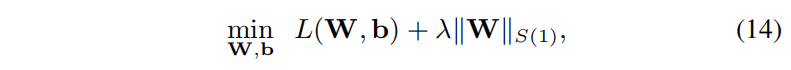
非凸：

凸：

Note：

注意：using the trace norm as a regularizer can make a matrix have low rank.

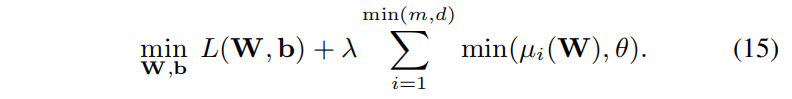
**5.3 [49]使用迹范数**



Note:代表W的第i个最小单值。那么，就代表迹范数

迹范数：矩阵特征值的和

**5.4 [50] capped（有上限的）trace regularizer**



Note:设置了上限，>θ的不惩罚，<θ的才惩罚，**为了人为决定W的秩为多少。**

当θ足够大，（15）退化到（14）

**6、 Task clustering approach**

all the tasks form a few clusters where tasks ina cluster are related to each other.

即：将任务聚类成几个类，信息只在同一个类之间进行分享，不同的类之间不进行信息的交换。

**7、Task relation learning approach**

aims to learn quantitative relations between tasks from **data automatically**. （早前的研究都把任务关系（task similarity）当作一个先验信息，Task relation learning 就是通过数据本身来得到这种task similarity）

**7.1 [79] multi-task Gaussian process (MTGP)**

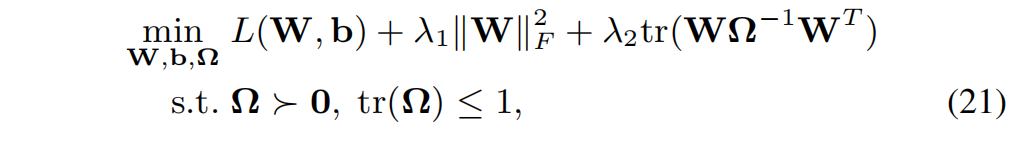
*  ：the functional values for all the training data
* ：其中，the covariance matrix, defines the covariance between

的每一项为，

其中k(·, ·) denotes a kernel function

表示Task i和task p之间的协方差

**7.2 [81] multi-task relationship learning (MTRL) ——Bayesian models**



Note:矩阵正态分布（均值，行向量协方差，列向量协方差）

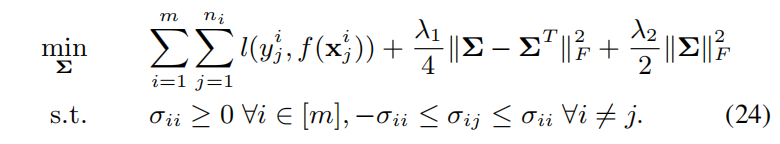
项：penalize the complexity of **W**

项：due to the matrix-variate normal prior

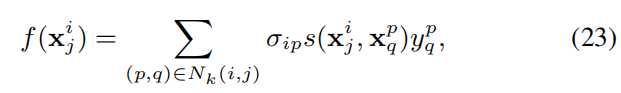
the constraints：control the complexity of the positive definite covariance matrix **Ω**

**7.3 [98]** **learn the task relations in local learning methods（之前的都是全局的）**

**使用了k-nearest-neighbor (kNN)**



Note:

loss function：其中a weighted voting of neighbors： 表示任务Ti和Tp之间的similarity，表示对的贡献，即数据点之间的similarity，求和表示的所有kNN点

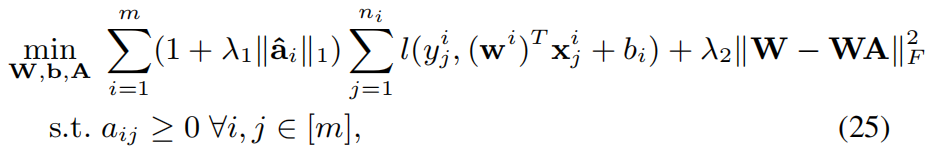
第一个正则化项：保证Σ接近对称矩阵

第二个正则化项：惩罚Σ的复杂性

约束项：任务自身之间的similarity 正的且最大

**7.4 [99]任务之间的关系是非对称的**

之前的模型都任务任务之间的关系是对称的，即Task(A,B)=Task(B,A)，但是[99]则认为Task(A,B)≠Task(B,A)



Note: 表示删除的A的第i行

：保证A稀疏，而且使得不对称信息从简单任务传到复杂的任务

正则化项：使得W接近于WA

**8、Decomposition approach**

decomposes the model parameters of all the tasks into two or more components, which are penalized by different regularizers.

二、问题

1、好多方法都用了这一项，W×一个矩阵×他的转置是什么意思呢？