Interactive Multi-Task Relationship Learning

Kaixiang Lin and Jiayu Zhou
ILLIDAN Lab
Computer Science and Engineering
Michigan State University

Outline

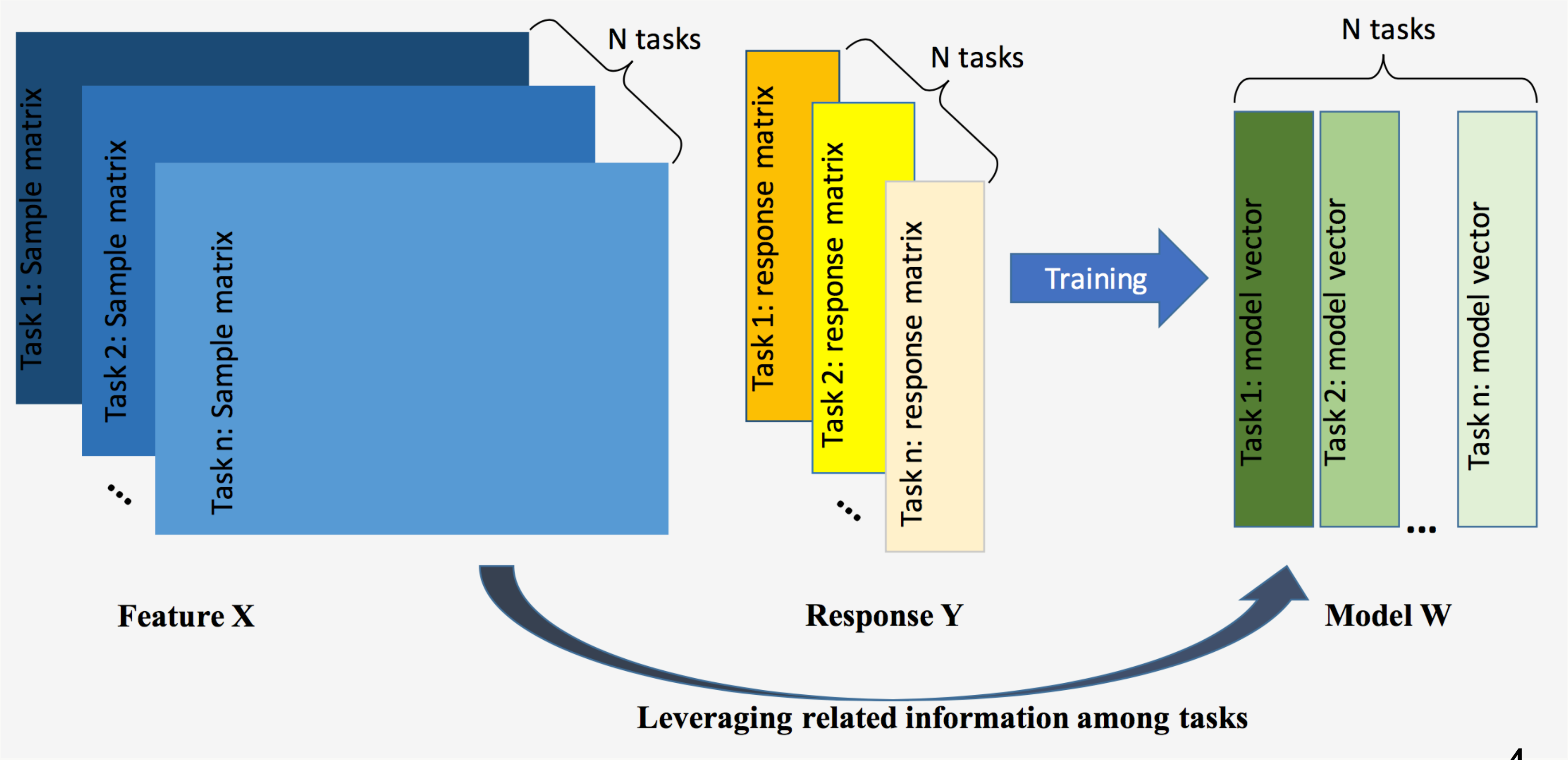
SPARTANS WILL.

- Introduction
  - Multi-task Learning, Active learning, Interactive Learning
- Method & Algorithms
  - Objective function, Algorithm
- Experimental Results

Introduction

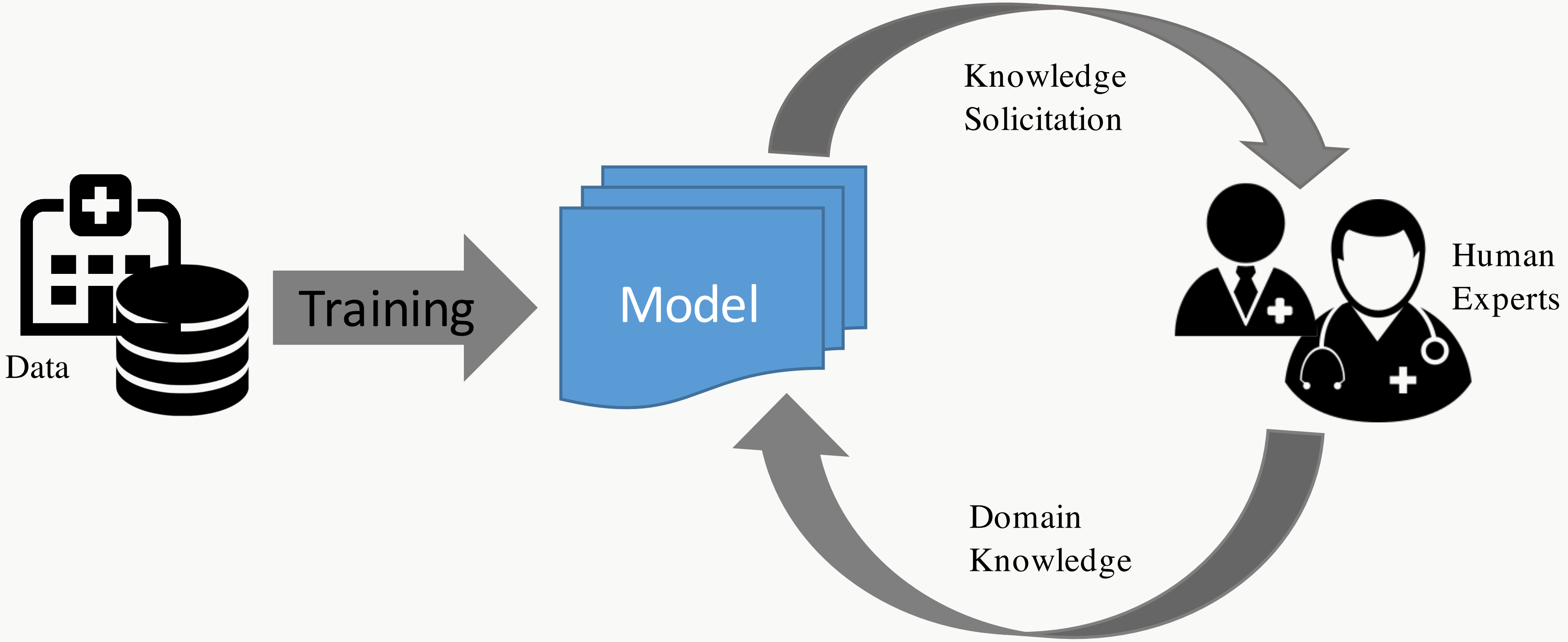
Introduction

- **Multi-task Learning (MTL)**
  - A learning paradigm that try to improve the generalization performance of a set of related machine learning tasks by transferring knowledge among the tasks.



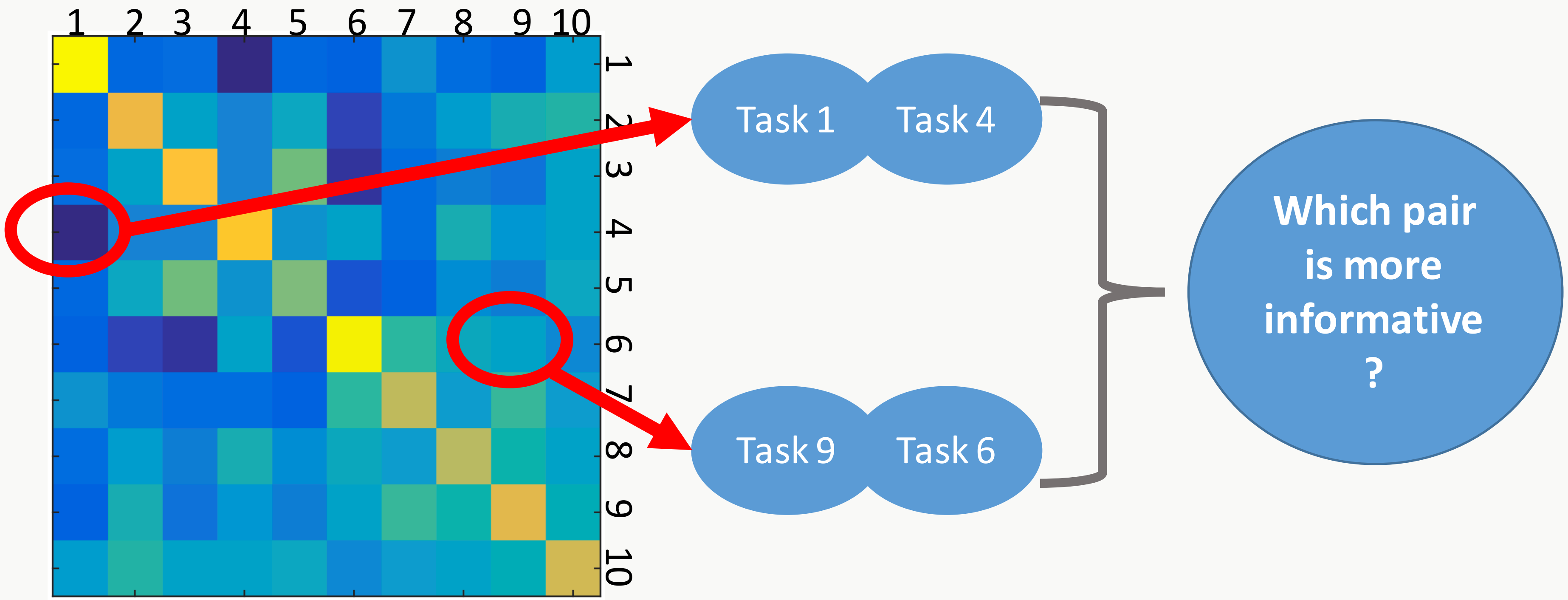
Introduction

- **Interactive Learning**
  - A systematic way to include human in the learning loop, observing the results of learning and providing feedback to improve the generalization performance of learning model.



Introduction

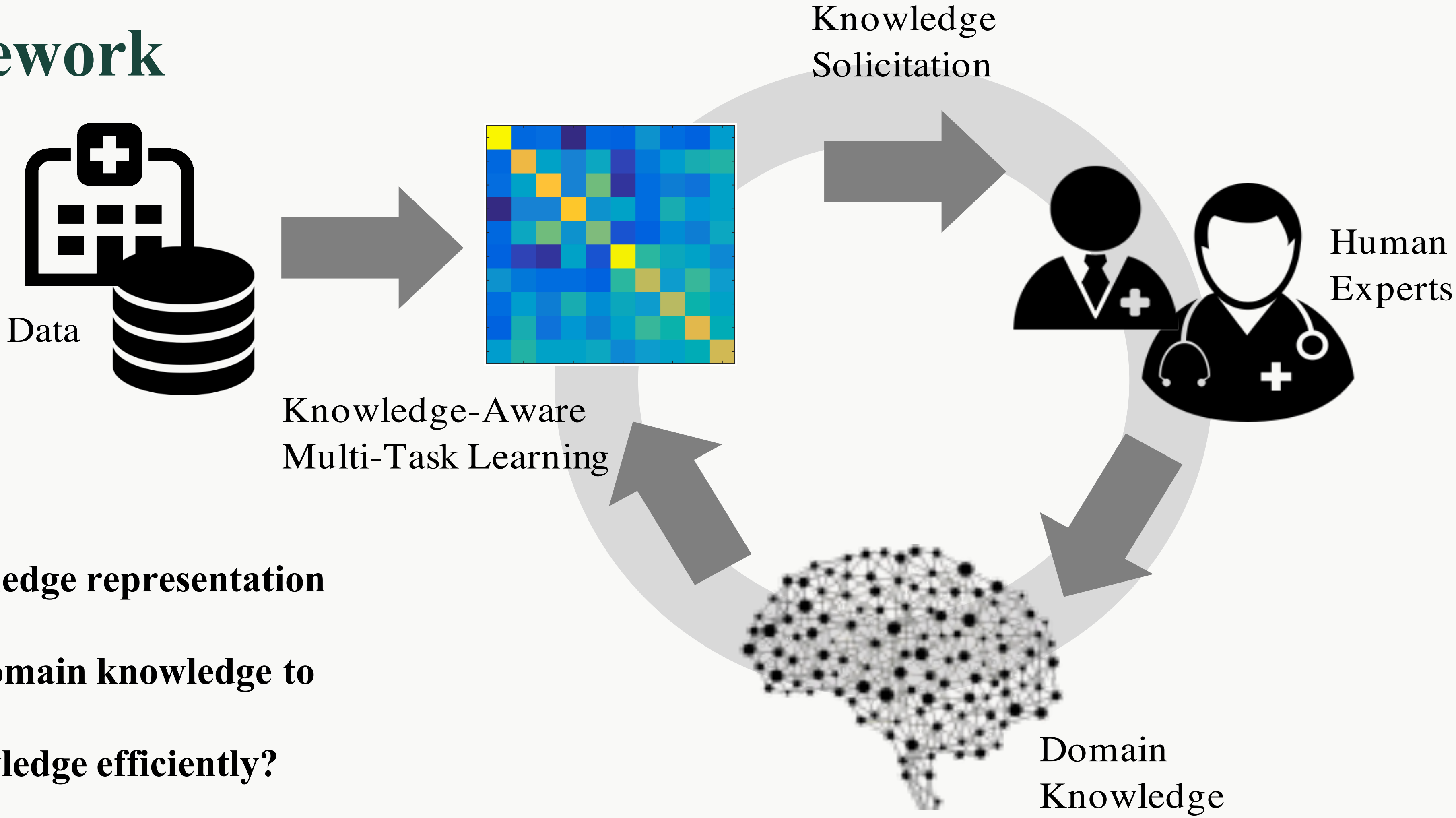
- **Active Learning**
  - A learning algorithm that is allowed to choose data from which it learns, so that achieve higher performance with few labeled data.



Methodology

interactive Multi-Task Relationship Learning

iMTRL Framework



Key Challenges:

1. **What type of knowledge representation should we choose?**
2. **How to integrate domain knowledge to MTL algorithm?**
3. **How to solicit knowledge efficiently?**

Multi-Task Relationship Learning[1]

MTRL – revisit

$$\min_{\mathbf{W},\Omega} \sum_{k=1}^{K} \frac{1}{n_k} \|\mathbf{y}^k - \mathbf{X}^k \mathbf{w}_k - b_k \mathbf{1}_{n_k}\|_F^2 + \frac{\lambda_1}{2} \mathrm{tr}(\mathbf{W}\mathbf{W}^T)$$

$$+ \frac{\lambda_2}{2} \mathrm{tr}(\mathbf{W}\boldsymbol{\Omega}^{-1}\mathbf{W}^T). \text{ s.t. } \boldsymbol{\Omega} \succeq 0, \mathrm{tr}(\boldsymbol{\Omega}) = 1$$

K: number of tasks;
X: data matrix, $n_k * d$;
W: model, d $*$ K matrix;
Ω: task correlation matrix, K $*$ K
$|| * ||_F$: Frobenius norm;

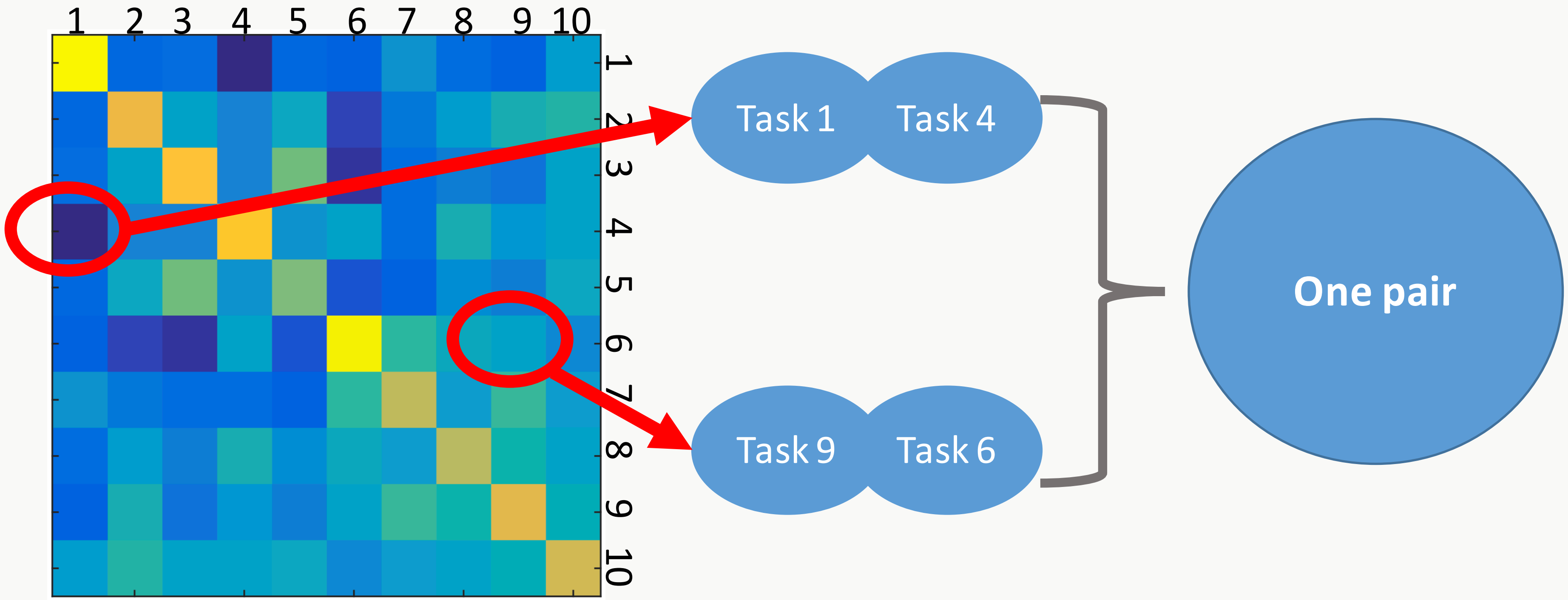
n_k: number of samples of task k;
y: responses, $n_k * 1$ vector
b: bias, a scalar
$tr(*)$: trace of a matrix

[1] Zhang Y, Yeung D Y. A convex formulation for learning task relationships in multi-task learning[J]. arXiv preprint arXiv:1203.3536, 2012.

Multi-Task Relationship Learning[1]

MTRL – illustration of Ω

$$\Omega$$



[1] Zhang Y, Yeung D Y. A convex formulation for learning task relationships in multi-task learning[J]. arXiv preprint arXiv:1203.3536, 2012.

Multi-Task Relationship Learning[1]

An alternating algorithm

1. Optimizing w.r.t. W and b when Ω is fixed
2. Optimizing w.r.t. Ω when W and b are fixed

The closed-form solution of task correlation matrix:

$$\mathbf{\Omega} = (\mathbf{W}^T\mathbf{W})^{1/2}/\mathrm{tr}((\mathbf{W}^T\mathbf{W})^{1/2}).$$

[1] Zhang Y, Yeung D Y. A convex formulation for learning task relationships in multi-task learning[J]. arXiv preprint arXiv:1203.3536, 2012.

knowledge-aware MTRL

kMTRL formulation

$$\min_{\mathbf{W},\mathbf{b},\boldsymbol{\Omega}} \mathcal{F}(\mathbf{W},\mathbf{b},\boldsymbol{\Omega}) = \sum_{k=1}^{K} \frac{1}{n_k} \|\mathbf{y}^k - \mathbf{X}^k \mathbf{w}_k - b_k \mathbf{1}_{n_k}\|_F^2$$

$$+\frac{\lambda_1}{2}\mathrm{tr}(\mathbf{W}\mathbf{W}^T) + \frac{\lambda_2}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}(\mathbf{W}\mathbf{W}^T + \epsilon\mathbf{I})),$$

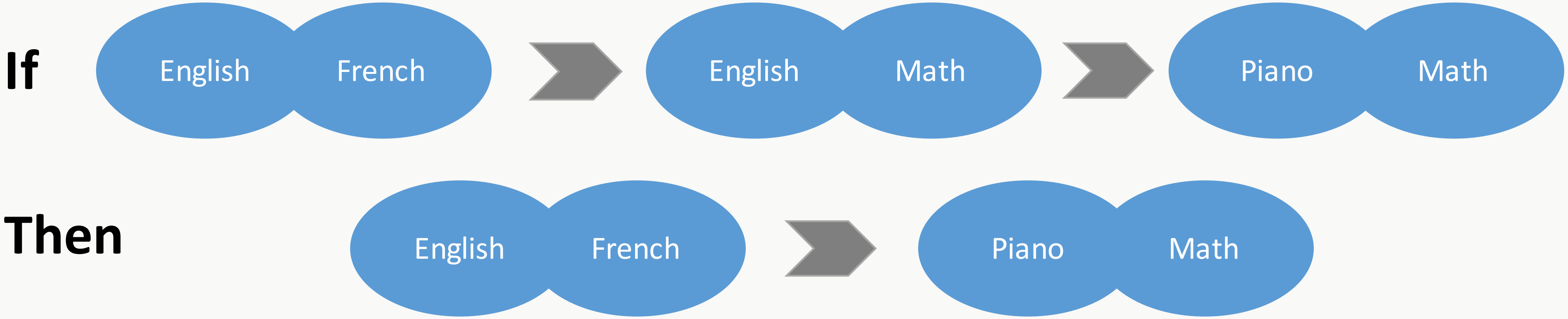
$$\text{s.t.} \qquad \boldsymbol{\Omega} \succeq 0,\ \mathrm{tr}(\boldsymbol{\Omega}) = 1,\ \boldsymbol{\Omega} \in \mathcal{T}$$

$$\mathcal{T} = \{\boldsymbol{\Omega} : \boldsymbol{\Omega}_{i_1,j_1} \geq \boldsymbol{\Omega}_{i_2,j_2}\ \forall (i_1, j_1, i_2, j_2) \in S\}$$

kMTRL Assumptions

- **A1: The domain knowledge acquired from human expert is accurate. The expert may choose not to label if he/she is not confident.**

- **A2: The acquired partial orders are compatible, i.e. when $\Omega_{i,j} > \Omega_{i,k}$ and $\Omega_{i,k} > \Omega_{k,p}$ are established, the $\Omega_{i,j} < \Omega_{k,p}$ cannot be included.**



kMTRL Optimization

- **Step 1: Optimizing w.r.t. W and b when Ω is fixed**
- **Step 2: Optimizing w.r.t. Ω when W and b are fixed**
- **Step 3: Project Ω to the feasible set[2]**
  - **Theorem 1.** *Suppose that* $\mathcal{T} = \{\mathbf{\Omega} : \mathbf{\Omega}_{i1,j1} \geq \mathbf{\Omega}_{i2,j2} + c\}$, *then, for any* $\mathbf{\Omega} \in \mathbb{R}^{K\times K}$, *the projection of* $\mathbf{\Omega}$ *to the convex set* $\mathcal{T}$ *is given by:*

$$\mathrm{Proj}(\mathbf{\Omega}) = \mathbf{\Omega} \ \textit{if}\ \mathbf{\Omega} \in \mathcal{T},$$

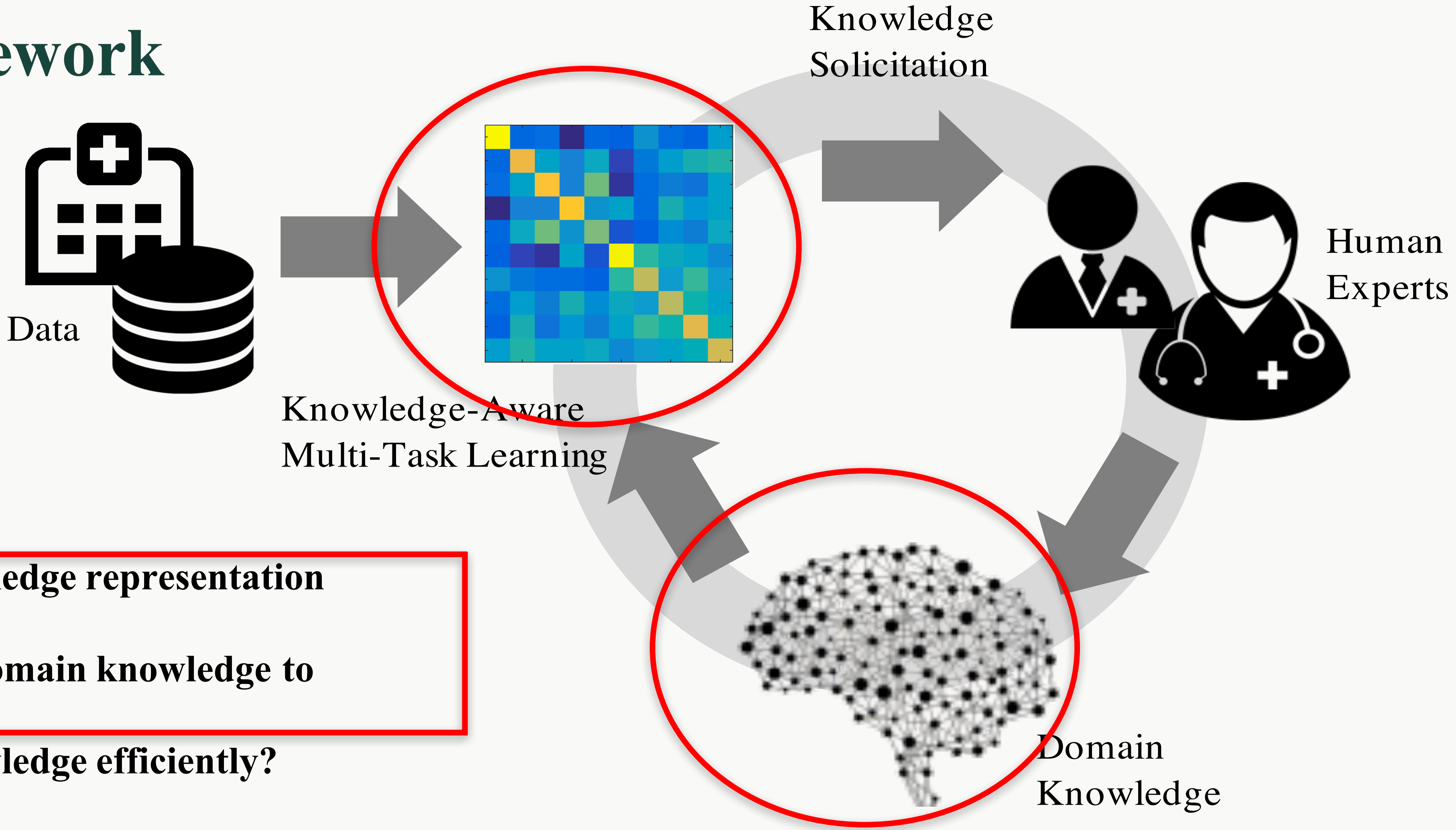
otherwise

$$\mathrm{Proj}(\mathbf{\Omega}) = \mathbf{\Omega}^* = \begin{cases} \mathbf{\Omega}^*_{i1,j1} = \frac{1}{2}(\mathbf{\Omega}_{i1,j1} + \mathbf{\Omega}_{i2,j2} + c) \\ \mathbf{\Omega}^*_{i2,j2} = \frac{1}{2}(\mathbf{\Omega}_{i1,j1} + \mathbf{\Omega}_{i2,j2} - c) \\ \mathbf{\Omega}^*_{p,q} = \mathbf{\Omega}_{p,q},\ \forall (p,q) \neq (i1,j1)\ \textit{and}\ (i2,j2) \end{cases}$$

[2] Chang S, Qi G J, Aggarwal C C, et al. Factorized similarity learning in networks[C]//2014 IEEE International Conference on Data Mining. IEEE, 2014: 60-69.

interactive Multi-Task Relationship Learning

iMTRL Framework



Key Challenges:

1. **What type of knowledge representation should we choose?**
2. **How to integrate domain knowledge to MTL algorithm?**
3. **How to solicit knowledge efficiently?**

Query Strategy

- $(\mathbf{W}, \mathbf{b}, \Omega) \longleftarrow$ **kMTRL**
- $\widehat{\Omega} \longleftarrow$ **Calculate by analytical solution using W**

  The closed-form solution of task correlation matrix:

  $$\hat{\mathbf{\Omega}} = (\mathbf{W}^T\mathbf{W})^{1/2}/\mathrm{tr}((\mathbf{W}^T\mathbf{W})^{1/2}).$$

- $\Omega \longleftarrow$ **Returned from kMTRL**
- **Query based on the discrepancy between Ω and $\widehat{\Omega}$**

Inconsistency

SPARTANS WILL.

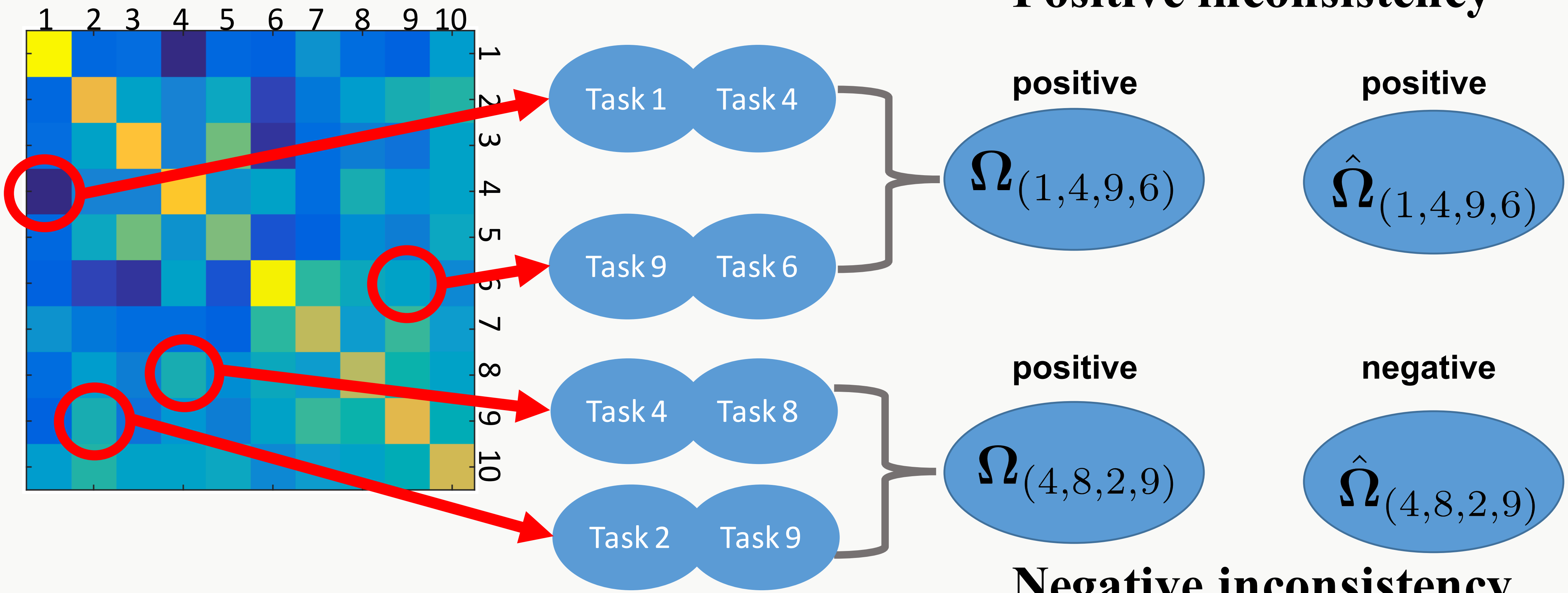
Definition 1 *Inconsistency is defined as:*

$$Inc_{(i_1,j_1,i_2,j_2)} = sign(i_1, j_1, i_2, j_2)|\mathbf{\Omega}_{(i_1,j_1,i_2,j_2)} - \hat{\mathbf{\Omega}}_{(i_1,j_1,i_2,j_2)}|,$$

$$where\ sign(i_1, j_1, i_2, j_2) = \frac{\mathbf{\Omega}_{(i_1,j_1,i_2,j_2)}\hat{\mathbf{\Omega}}_{(i_1,j_1,i_2,j_2)}}{|\mathbf{\Omega}_{(i_1,j_1,i_2,j_2)}\hat{\mathbf{\Omega}}_{(i_1,j_1,i_2,j_2)}|},\ \mathbf{\Omega}_{(i_1,j_1,i_2,j_2)} = \mathbf{\Omega}_{i_1,j_1} - \mathbf{\Omega}_{i_2,j_2}$$

Inconsistency

Positive inconsistency



Negative inconsistency

iMTRL algorithm

Algorithm $(\mathbf{\Omega}, \mathbf{W}, \mathbf{b}) = \text{iMTRL}(\text{data, parameters})$

Require: Training sets $\{\mathbf{X}^k, \mathbf{y}^k\}_k^K$, number of selected queries $\mathbf{q}$. regularization parameters λ_1, λ_2, positive number c, $\mathcal{T}^0 = \emptyset$

```
for i = 1, ..., n do
    (Ω^i, W^i, b^i) = kMTIL({X^k, y^k}_k^K, T^{i-1}, λ1, λ2, c)
    T^i = query(W^i, Ω^i, q_i)
    T^i = T^i ∪ T^{i-1}
end for
Ω = Ω^i, W = W^i, b = b^i
return Ω, W, b
```

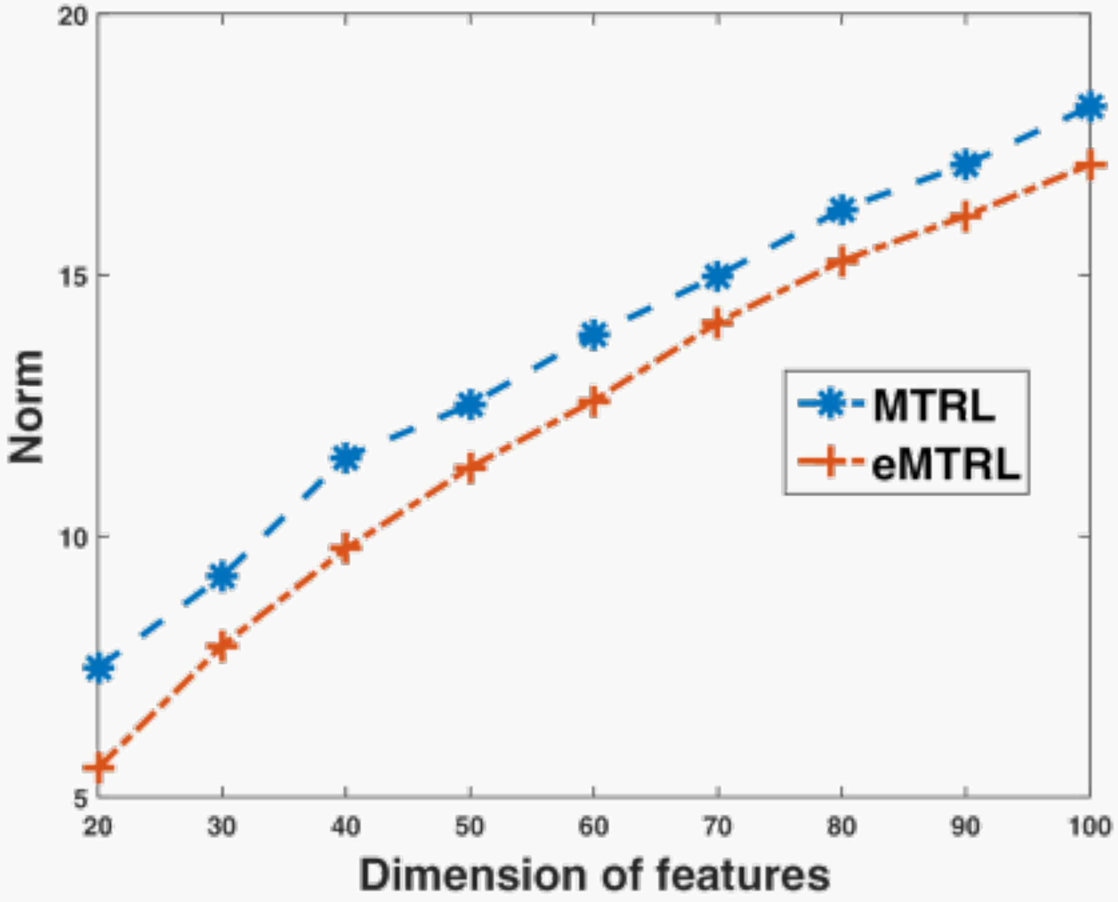
Experimental Results

Importance of High-Quality Task Relationship

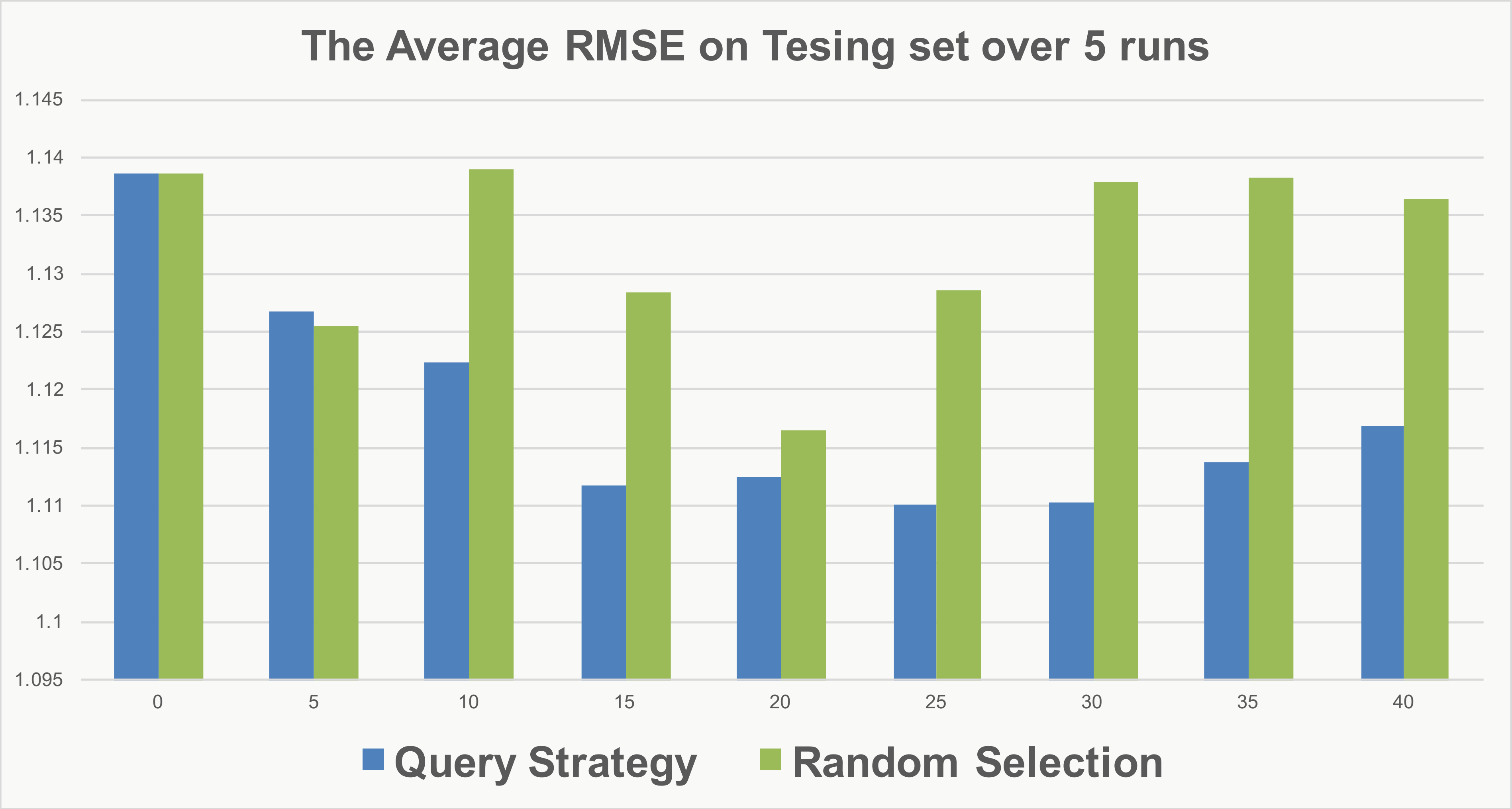
Synthetic dataset

- $y = wx + b + e$
- 10 tasks, 25 samples, x ~ Gaussian(0, 10), e~Gaussian(0,5)
- w and b ~ Normal(0,1)
- 20% are for training, 30% for validation and 50% for testing.
- Change the number of features from 20 to 100
- truth-encoded multi-task relationship learning (eMTRL)

Importance of High-Quality Task Relationship



Effectiveness of Query Strategy



Effectiveness of kMTRL

SPARTANS WILL.

TABLE II

THE AVERAGE RMSE COMPARISON OF COMPETING METHODS ON THE SCHOOL DATASET AND MMSE DATASET. THE FIRST COLUMN IS THE PERCENTAGE OF TRAINING SAMPLES IN EACH TASK. THE KMTRL METHODS OUTPERFORMS ALL OTHER METHODS

School	RR	MTL-L	MTL-l21	MTRL	kMTRL-20	kMTRL-40	kMTRL-60	kMTRL-80
5%	1.1737±0.0041	1.1799± 0.0047	1.176± 0.0043	1.0615 ± 0.0167	1.0584 ± 0.0128	1.0553 ± 0.0155	**1.0551** ± 0.0158	1.0551 ± 0.0159
10%	1.1428±0.0306	1.1485 ± 0.0293	1.1477 ± 0.0282	0.9872 ± 0.0057	0.9823 ± 0.0030	0.9805 ± 0.0014	**0.9803** ± 0.0018	0.9803 ± 0.0018
15%	1.0665±0.0395	1.0699 ± 0.0405	1.0700 ± 0.0399	0.9491 ± 0.0060	0.9334 ± 0.0057	**0.9321** ± 0.0081	0.9322 ± 0.0083	0.9323 ± 0.0082
20%	0.9756±0.0157	0.9774 ± 0.0153	0.9776 ± 0.0149	0.9047 ± 0.0031	0.8966 ± 0.0123	0.8906 ± 0.0123	0.8844 ± 0.0022	**0.8843** ± 0.0019
MMSE	RR	MTL-L	MTL-l21	MTRL	kMTRL-5	kMTRL-10	kMTRL-15	kMTRL-20
2%	0.9503± 0.1467	0.9319±0.1497	0.9314±0.1693	0.9106 ± 0.0976	0.9113 ± 0.0982	**0.9058** ± 0.0926	0.9058 ± 0.0926	0.9058 ± 0.0926

Case Study: Alzheimer's Disease

Brain Atrophy and Alzheimer's Disease

- Alzheimer's disease (AD) is a type of dementia that causes problems with memory, thinking and behavior.[3]
- AD has become one of the leading causes of death in the United States.
- AD is characterized by structural atrophy in the brain, we want to understand how the brain atrophy is related to the progression of the disease.

[3] http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp

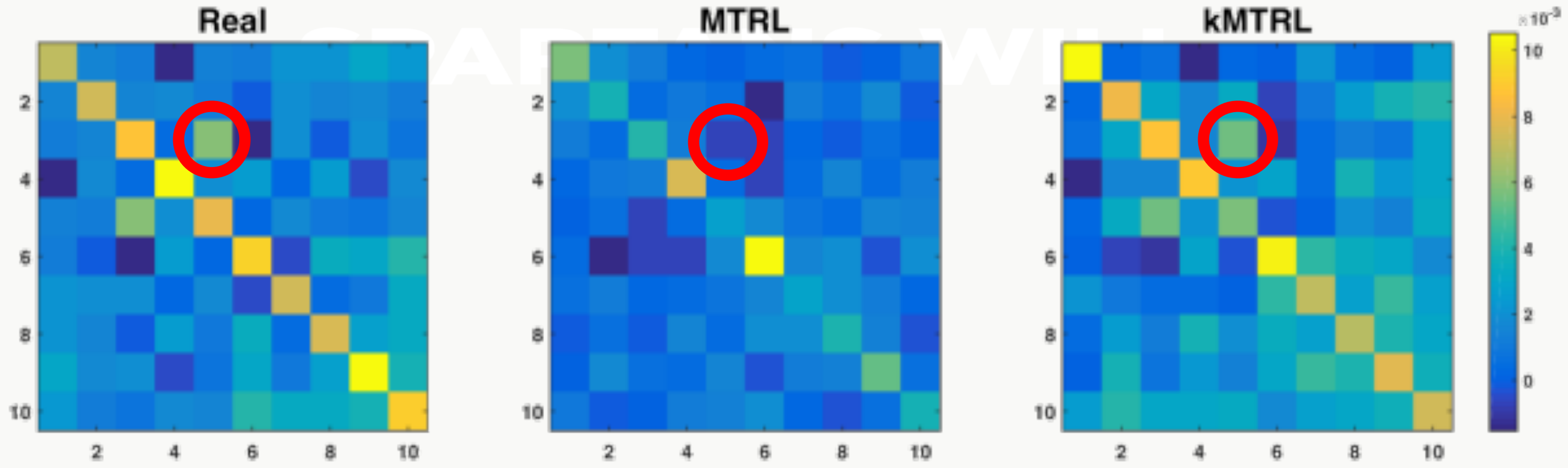
Case Study: Alzheimer’s Disease

Experimental settings

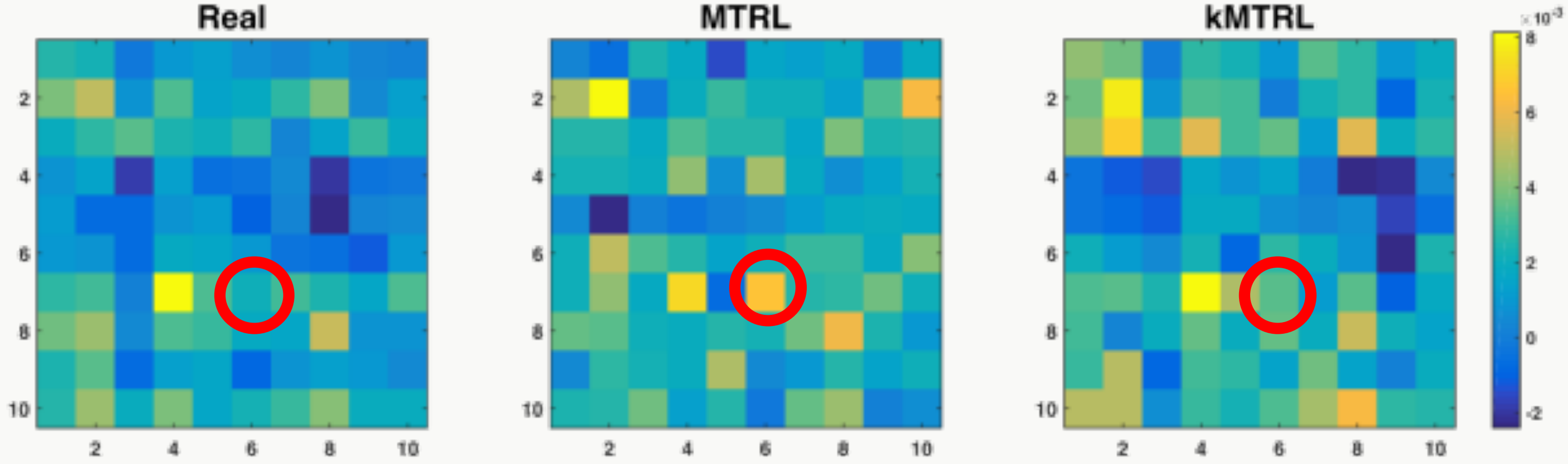
- Predict 99 brain volumes of each regions of interest (ROIs)
- 648 samples, 51 cognitive markers as features.

- Ω_{real}: learned from 90% training samples by MTRL
- Ω_{MTRL}: the covariance matrix learned via MTRL on 10% data
- Ω_{kMTRL}: the covariance matrix learned via kMTRL on 10% data

Case Study: Alzheimer's Disease - Visualization



(a) Intra-region covariance



(b) Inter-region covariance

Conclusion

- **What type of knowledge representation should we choose?**

  Pairwise encoding:

  The domain knowledge of task relationship is represented as partial orders, and can be encoded in the learning as pairwise constraints.

- **How to integrate domain knowledge to MTL algorithm?**

  Knowledge-Aware Multi-Task Learning:

  We propose a novel MTL algorithm that infers models and task relationship from data and conform the solicited knowledge.

- **How to solicit knowledge efficiently?**

  Active Learning based Knowledge Query:

  To maximize the usefulness of solicited knowledge, we propose a knowledge query strategy based on active learning.

SPARTANS WILL.

Q & A

Email: linkaixi@msu.edu

ACKNOWLEDGEMENT: This material is based in part upon work supported by the National Science Foundation under Grant IIS-1565596, IIS- 1615597 and Office of Naval Research N00014-14-1-0631.