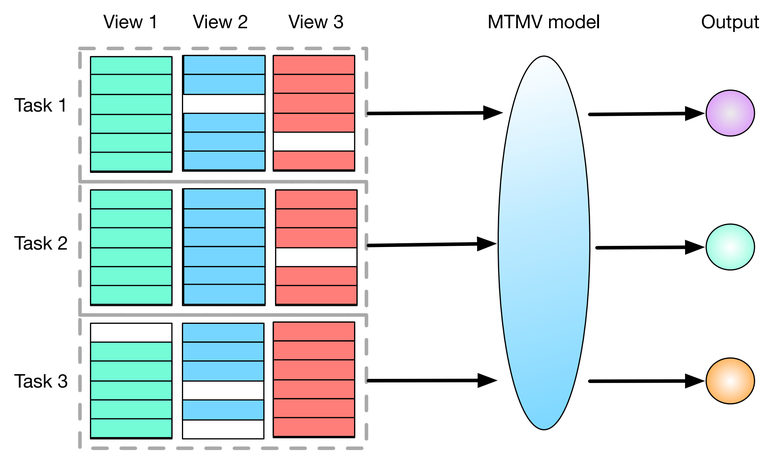
Multilinear Factorization Machines for Multi-Task Multi-View Learning

Chun-Ta Lu (clu29@uic.edu), Lifang He, Weixiang Shao, Bokai Cao and Philip S. Yu

Paper Code

Multi-Task Multi-View Learning

Multi-Task Multi-View (MTMV) Learning: combine different views (e.g., images and texts) to learn multiple related tasks together.



- **Conventional approaches try to minimize the difference of the predictive functions among different views, while different views might not be consistent. How to exploit complementary information from multiple views?**
- **Some instances may not have certain views. How to explore the feature interactions between multiple views to estimate parameter under sparsity?**

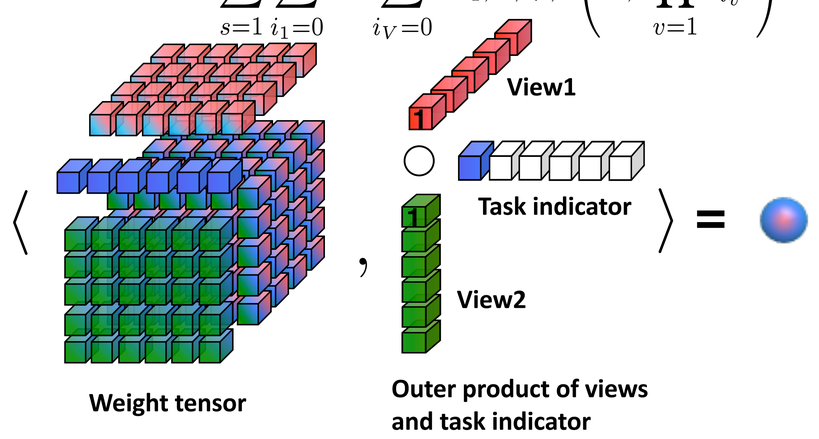
Multilinear Factorization Machines

We propose a multilinear structure for MTMV learning

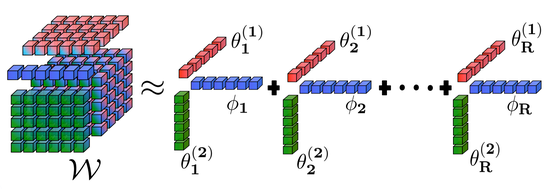
Sum up to full-order interaction by adding an extra feature with constant value 1 to the feature vector $\mathbf{z}^{(v)} = [1; \mathbf{x}^{(v)}] \in \mathbb{R}^{1+I_v}$

$$f_t(\{\mathbf{x}^{(v)}\}) = \left\langle \mathcal{W}, \mathbf{z}^{(1)} \circ \cdots \circ \mathbf{z}^{(V)} \circ \mathbf{e}_t \right\rangle = \langle \mathcal{W}, \mathcal{Z}_t \rangle$$

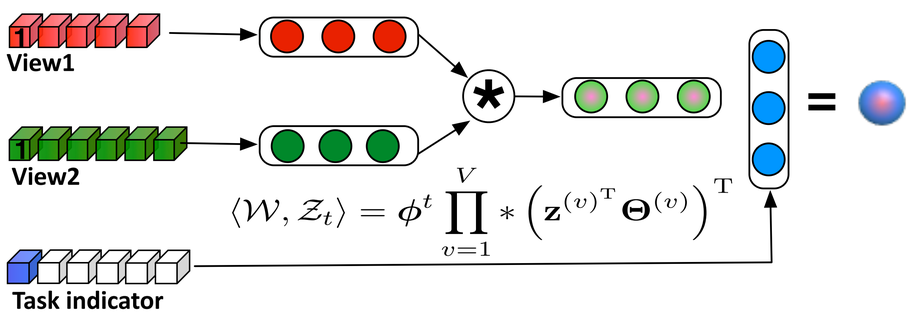
$$= \sum_{s=1}^{T} \sum_{i_1=0}^{I_1} \cdots \sum_{i_V=0}^{I_V} w_{i_1,\ldots,i_V,s} \left(e_{t,s} \prod_{v=1}^{V} z_{i_v}^{(v)} \right)$$



Applying tensor factorization on the weight tensor.



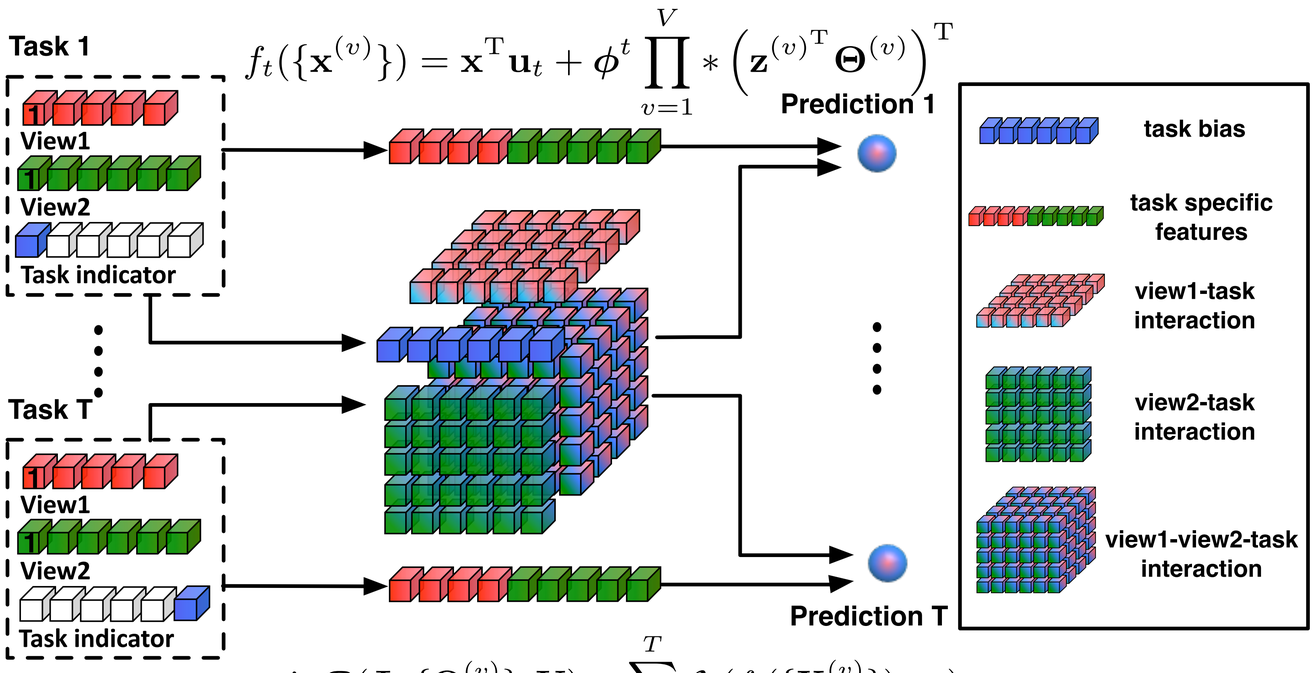
$$\mathcal{W} = \sum_{r=1}^{R} \phi_{\mathbf{r}} \circ \theta_{\mathbf{r}}^{(1)} \circ \theta_{\mathbf{r}}^{(2)}$$



$$\langle \mathcal{W}, \mathcal{Z}_t \rangle = \phi^t \prod_{v=1}^{V} * \left(\mathbf{z}^{(v)\mathrm{T}} \mathbf{\Theta}^{(v)} \right)^{\mathrm{T}}$$

It is too restrictive to constrain all tasks to share the common subspace. We design a predictive function to learn from both the task-specific feature map and the task-view shared multilinear structures.

$$f_t(\{\mathbf{x}^{(v)}\}) = \mathbf{x}^{\mathrm{T}} \mathbf{u}_t + \phi^t \prod_{v=1}^{V} * \left(\mathbf{z}^{(v)\mathrm{T}} \mathbf{\Theta}^{(v)} \right)^{\mathrm{T}}$$



$$\min \mathcal{R}(\mathbf{\Phi}, \{\mathbf{\Theta}^{(v)}\}, \mathbf{U}) = \sum_{t=1}^{T} \mathcal{L}_t(f_t(\{\mathbf{X}_t^{(v)}\}), \mathbf{y}_t) + \lambda \Omega_\lambda(\mathbf{\Phi}, \{\mathbf{\Theta}^{(v)}\}) + \gamma \Omega_\gamma(\mathbf{U})$$

Solved by alternating block coordinate descent w/ AdaGrad

- **Proposed MFMs can efficiently learn the task-specific feature map and the task-view shared multilinear structures from full-order interactions, without physically building the tensor.**
- **MFMs jointly factorize the interaction parameters for different orders to allow accurate parameter estimation under sparsity and renders the model with the capacity to avoid overfitting.**
- **The model complexity of MFMs is linear in the feature dimensionality, making it applicable to large-scale applications.**

Experiment Results

Datasets

Classification	#Feature	T	N_p	N_n
FOX	image(996), text(2,711)	4	178∼635	888∼1,345
DBLP	linkage(4,638), text(687)	6	635∼1,950	2,688∼3,985
Regression	**#Feature**	**T**	**N**	**Density**
MovieLens	users(943), movies(1,599), tags(1,065)	10	758∼39,895	6.3%
Amazon	users(1,805,364), items(192,978), text(83,143)	5	349,038∼1,015,189	0.001%

T: number of tasks
N: total number of instances
Np (Nn): number of positive (negative) instances in each task
Density: the density of the user-item matrix

Average results of 10 times of random sampling: n% labeled instances as training set (n=10,20, and 30), 10% as validation set, 40% as testing set

Comparison Methods

- **rMTFL:** robust multi-task feature learning algorithm
- **IteM²:** transductive MTMV classification algorithm
- **CSL-MTMV:** state-of-the-art inductive MTMV learning algorithm
- **Factorization Machine (FM)**: state-of-the-art factorization model
- **Tensor Factorization (TF):** factorize highest-order weight tensor
- **Multilinear Tensor Factorization (MFM):** proposed method
  - **MFM-T:** only using tensor part (**U** is fixed as a zero matrix)
  - **MFM-F:** using F-norm regularizers for all parameters
  - **MFM-F-S:** using $\ell_{2,1}$-norm on **U** for joint feature selection
    F-norm on the rest parameters

FOX

Training Ratio	Measure	rMTFL	FM	TF	IteM²	CSL-MTMV	MFM-T	MFM-F	MFM-F-S
10%	ACC	0.8816±0.011	0.7883±0.011	0.8460±0.035	0.4052±0.076	0.8986±0.011	0.9259±0.019	**0.9343±0.012**	**0.9364±0.011**
	F1	0.6911±0.035	0.2930±0.046	0.6362±0.044	0.3598±0.030	0.7335±0.029	0.7799±0.053	**0.8076±0.038**	**0.8119±0.027**
	AUC	0.9109±0.013	0.7764±0.018	0.8681±0.038	0.5326±0.036	0.9342±0.011	0.9678±0.015	**0.9763±0.008**	**0.9777±0.009**
20%	ACC	0.9039±0.013	0.8087±0.011	0.8546±0.025	0.5091±0.078	0.9264±0.005	0.9551±0.005	**0.9569±0.010**	**0.9612±0.005**
	F1	0.7654±0.026	0.3764±0.050	0.6632±0.051	0.3306±0.068	0.8004±0.012	0.8721±0.012	**0.8769±0.027**	**0.8882±0.014**
	AUC	0.9353±0.016	0.8260±0.012	0.8751±0.029	0.4954±0.043	0.9705±0.003	0.9883±0.003	**0.9885±0.006**	**0.9922±0.002**
30%	ACC	0.9314±0.005	0.8255±0.007	0.8767±0.082	0.4289±0.134	0.9390±0.004	0.9641±0.007	**0.9709±0.003**	**0.9697±0.004**
	F1	0.8051±0.015	0.4448±0.026	0.7302±0.132	0.3314±0.056	0.8341±0.012	0.9000±0.018	**0.9185±0.010**	**0.9149±0.010**
	AUC	0.9709±0.005	0.8393±0.012	0.9010±0.091	0.5365±0.039	0.9812±0.003	0.9916±0.003	**0.9949±0.001**	**0.9949±0.001**

DBLP

Training Ratio	Measure	rMTFL	FM	TF	IteM²	CSL-MTMV	MFM-T	MFM-F	MFM-F-S
10%	ACC	0.8057±0.004	0.7264±0.004	0.7471±0.011	0.6223±0.004	0.7290±0.005	0.8008±0.004	**0.8058±0.004**	**0.8062±0.005**
	F1	0.5395±0.015	0.0732±0.019	0.5606±0.011	0.3176±0.007	0.4402±0.004	0.5278±0.018	**0.5469±0.014**	**0.5471±0.015**
	AUC	0.7888±0.007	0.6264±0.023	0.7723±0.009	0.5310±0.007	0.6890±0.006	0.8039±0.010	**0.8113±0.010**	**0.8120±0.009**
20%	ACC	0.8319±0.004	0.7628±0.007	0.7878±0.007	0.6309±0.003	0.7760±0.002	0.8346±0.004	**0.8374±0.004**	**0.8371±0.004**
	F1	0.6447±0.008	0.2680±0.038	0.6247±0.014	0.3494±0.006	0.5295±0.007	0.6274±0.013	**0.6499±0.012**	**0.6508±0.012**
	AUC	0.8374±0.005	0.7548±0.022	0.8200±0.010	0.5550±0.006	0.7655±0.005	0.8531±0.006	**0.8658±0.005**	**0.8632±0.005**
30%	ACC	0.8412±0.004	0.7978±0.005	0.8191±0.008	0.6256±0.003	0.8037±0.003	0.8501±0.004	**0.8527±0.004**	**0.8535±0.004**
	F1	0.6796±0.010	0.4312±0.021	0.6670±0.021	0.3569±0.009	0.5869±0.007	0.6800±0.013	**0.6891±0.012**	**0.6892±0.009**
	AUC	0.8590±0.005	0.8351±0.010	0.8498±0.009	0.5563±0.006	0.8083±0.006	0.8757±0.005	**0.8866±0.006**	**0.8866±0.006**

MovieLens

Training Ratio	Measure	rMTFL	FM	TF	CSL-MTMV	MFM-T	MFM-F	MFM-F-S
10%	RMSE	1.1861±0.008	1.0251±0.003	1.5679±0.099	1.05013±0.005	1.0078±0.005	**1.0069±0.005**	**0.9976±0.004**
	MAE	0.8516±0.004	0.8422±0.004	1.2497±0.088	0.8516±0.004	0.8142±0.005	**0.8082±0.005**	**0.8022±0.004**
20%	RMSE	1.0631±0.005	0.9898±0.003	1.2519±0.069	1.0214±0.004	**0.9877±0.003**	0.9977±0.003	**0.9857±0.003**
	MAE	0.8539±0.005	0.7997±0.004	0.9801±0.053	0.8294±0.004	**0.7987±0.003**	0.8023±0.003	**0.7927±0.004**
30%	RMSE	0.9917±0.003	**0.9765±0.003**	1.2066±0.061	1.0082±0.003	0.9795±0.003	0.9887±0.004	**0.9785±0.003**
	MAE	0.8159±0.003	**0.7815±0.003**	0.9380±0.045	0.8189±0.003	0.7885±0.002	0.7823±0.004	**0.7789±0.004**

Amazon

Training Ratio	Measure	FM	TF	MFM-T	MFM-F	MFM-F-S
10%	RMSE	0.9834±0.001	3.6044±0.003	**0.9775±0.001**	0.9857±0.001	**0.9825±0.002**
	MAE	0.7420±0.001	3.4574±0.005	0.7249±0.001	**0.7158±0.002**	**0.7129±0.001**
20%	RMSE	0.9814±0.001	3.5611±0.018	**0.9764±0.001**	0.9845±0.001	**0.9775±0.001**
	MAE	0.7343±0.002	3.3965±0.030	0.7255±0.001	**0.7112±0.001**	**0.7086±0.001**
30%	RMSE	0.9782±0.002	3.4962±0.018	**0.9705±0.002**	0.9841±0.001	**0.9733±0.001**
	MAE	0.7257±0.002	3.2945±0.034	**0.7001±0.001**	0.7115±0.001	**0.7078±0.001**

Due to the memory overhead, rMTFL and CSL-MTMV are not compared.

- **Due to view inconsistency, previous MTMV methods such as IteM2, and CSL-MTMV perform worse than rMTFL.**
- **MFMs consistently outperform compared methods, and improve 6~10% over best compared methods.**
- **MFM-F-S performs the best, and task-specific linear feature map is more important for classification than for regression.**

Acknowledgments

This work is supported in part by NSF through grants IIS-1526499, CNS-1626432, NSFC (61472089, 61503253, 61672357), NSFC- Guangdong Joint Found (U1501254), and the Science Foundation of Guangdong Province (2014A030313556). We gratefully acknowledge the support of SIGIR Student Travel Grant and NVIDIA Corporation with the donation of the Titan X GPU used for this research.