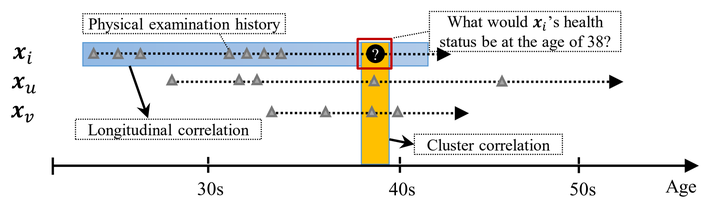
LMLFM: Longitudinal Multi-Level Factorization Machine

Junjie Liang,[1,2] Dongkuan Xu,[2] Yiwei Sun,[1] Vasant Honavar[1,2]

[1] Artificial Intelligence Research Laboratory, Pennsylvania State University
[2]College of Information Sciences and Technology, Pennsylvania State University

Longitudinal Data

Longitudinal data consists of irregularly repeated, sparse observation from a set of individuals over time. Such data often exhibits longitudinal correlation (LC, correlation among observations centered on the same individual over time), cluster correlation (CC, correlation among individuals that have similar characteristics), or multilevel correlation (MLC, combination of LC and CC). Therefore, longitudinal data are not independent and identically distributed (i.i.d.). Analysis that does not account for such correlations can lead to misleading statistical inferences.



Motivation

Mixed effects models (MEM) are popular in longitudinal data analysis due to their simplicity and effective way of handling data correlation through the design of random effects. However, existing MEM bear many limitations.

- They rely on expert input to decide which variables are subject to random effects as opposed to fixed effects. This typical process of trial-and-error does not work for high-dimensional data.
- They assume the type of data correlation is a priori known. A wrong correlation assumption leads to poor performance.
- They are computationally intensive with cubic time cost w.r.t. the size of random effects.

Problem Definition

RQ1: Is a_i predictive or redundant? If predictive, which effect does a_i have, fixed or random?

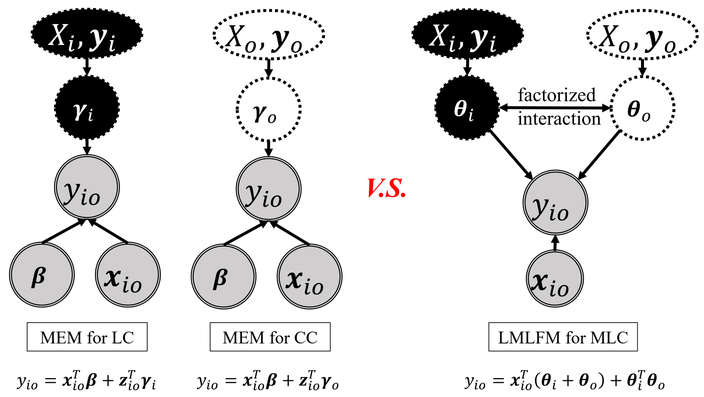
	a_1	a_2	a_3	y
x_{11}	1	0.2	3	1
x_{12}	1	0.22	2	2.3
x_{13}	1	0.24	1	?
x_{21}	0	0.1	5	4
x_{22}	------ unobserved ------			
x_{23}	1	0.1	3	?

RQ2: Given X, how do we predict the missing outcomes while accounting for *unknown* data correlation?

Specifically, RQ1 is a joint variable selection problem unique to longitudinal data. For any given variable, we simultaneously select between *predictive vs. redundant* and *random vs. fixed*; RQ2 is a special regression problem for longitudinal (non-i.i.d.) data.

Prediction Model

Existing MEM account for LC (or CC) by introducing individual (or observation) specific random effects as proxies that subsume the association within the relevant information. Inspired by MEM, our model incorporates both individual factors $\boldsymbol{\theta}_i$ and observation factors $\boldsymbol{\theta}_o$, as proxies for the relevant information to accommodate MLC.

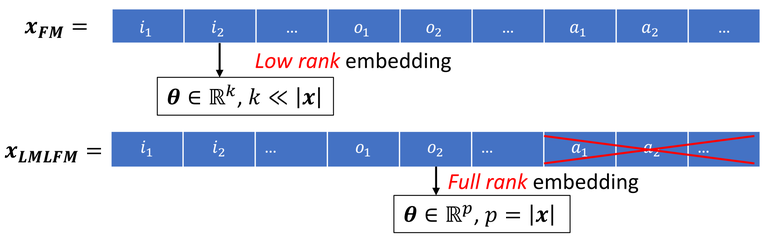


$y_{io} = x_{io}^T\boldsymbol{\beta} + z_{io}^T\boldsymbol{\gamma}_i$ $\quad y_{io} = x_{io}^T\boldsymbol{\beta} + z_{io}^T\boldsymbol{\gamma}_o$ $\quad y_{io} = x_{io}^T(\boldsymbol{\theta}_i + \boldsymbol{\theta}_o) + \boldsymbol{\theta}_i^T\boldsymbol{\theta}_o$

Notation:

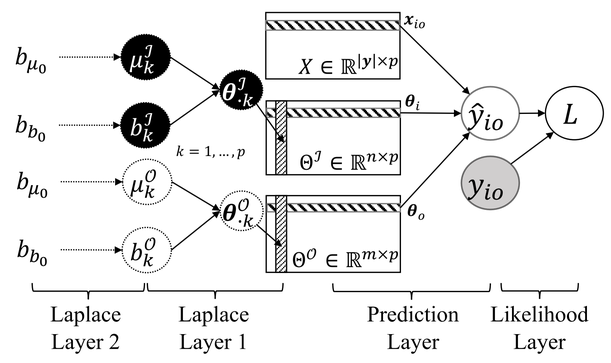
x_{io}	Covariate vector for individual i and observation o
z_{io}	Subset of covariates that are subject to random effects.
$\boldsymbol{\beta}$	Fixed effects.
$\boldsymbol{\gamma}$	Random effects.
$\boldsymbol{\theta}$	Latent factors. Could be further used to estimate $\boldsymbol{\gamma}$ (see Effects Estimation)

LMLFM vs. FM



We initially allow latent dimension to be as large as feature dimension, and then apply variable selection to identify the relevant subset of latent factors (see below). This makes our model almost as interpretable as a simple linear model.

Hierarchical Bayesian Model



We enforce sparsity on $\boldsymbol{\mu}$ and $\boldsymbol{b}$ to identify fixed and random effects (Laplace layer 2), and enforce sparsity on Θ to select relevant subset of latent factors (Laplace layer 1). More formally, our generative model is defined as

$(y_{io}|\boldsymbol{x}_{io}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_o, \alpha) \sim N(y_{io}|\hat{y}_{io}, \alpha^{-1})$ $\quad (\alpha|\alpha_0, \beta_0) \sim Gamma(\alpha|\alpha_0, \beta_0)$

$(\theta_{ik}|\mu_k^{\mathcal{I}}, b_k^{\mathcal{I}}) \sim Laplace(\theta_{ik}|\mu_k^{\mathcal{I}}, b_k^{\mathcal{I}})$ $\quad (\theta_{ok}|\mu_k^{\mathcal{O}}, b_k^{\mathcal{O}}) \sim Laplace(\theta_{ok}|\mu_k^{\mathcal{O}}, b_k^{\mathcal{O}})$

$(\mu_k^{\mathcal{I}}|b_{\mu_0}) \sim Laplace(\mu_k^{\mathcal{I}}|0, b_{\mu_0})$ $\quad (\mu_k^{\mathcal{O}}|b_{\mu_0}) \sim Laplace(\mu_k^{\mathcal{O}}|0, b_{\mu_0})$

$(b_k^{\mathcal{I}}|b_{b_0}) \sim Laplace(b_k^{\mathcal{I}}|0, b_{b_0})$ $\quad (b_k^{\mathcal{O}}|b_{b_0}) \sim Laplace(b_k^{\mathcal{O}}|0, b_{b_0})$

Inference

We consider the Maximum A Posteriori (MAP) formulation

$$\hat{\Theta} = \arg\max_{\Theta} \pi(\Theta|\boldsymbol{y}, X, \Theta_0)$$

We solve the objective function using iterated conditional modes (ICM) algorithm. ICM updates blocks of parameters with the modes of their conditional posterior while keeping the remaining parameters fixed. Our choice of priors permits closed-form update and sparsity estimation. We theoretically prove the linear computational complexity, ascent property and convergence of our model.

Effects/Outcome Estimation

Effect type	Estimator
Temporal Individual-specific Random Effects	$\boldsymbol{\gamma}_{io} = \boldsymbol{\theta}_i + \boldsymbol{\theta}_o$
Averaged Individual-specific Random Effects	$\boldsymbol{\gamma}_i = \mathbb{E}_{\pi(\boldsymbol{\theta}_o\|\boldsymbol{y},X,\boldsymbol{\Theta}_0)}[\boldsymbol{\gamma}_{io}]$
Temporal Population-averaged Random Effects	$\boldsymbol{\gamma}_o = \mathbb{E}_{\pi(\boldsymbol{\theta}_i\|\boldsymbol{y},X,\boldsymbol{\Theta}_0)}[\boldsymbol{\gamma}_{io}]$
Fixed Effects	$\beta_k = \mu_k^{\mathcal{I}} + \mu_k^{\mathcal{O}}$ iff. $b_k^{\mathcal{I}} = b_k^{\mathcal{O}} = 0$
Outcome for seen individual and observation	$x_{io}^T(\boldsymbol{\theta}_i + \boldsymbol{\theta}_o) + \boldsymbol{\theta}_i^T\boldsymbol{\theta}_o$
Outcome for unseen individual or observation	$\mathbb{E}_{\pi(\boldsymbol{\theta}\|\boldsymbol{y},X,\boldsymbol{\Theta}_0)}[x_{io}^T(\boldsymbol{\theta}_i + \boldsymbol{\theta}_o) + \boldsymbol{\theta}_i^T\boldsymbol{\theta}_o]$

Experimental Results

Results on simulated data

Method	$p = 100$			$p = 5000$		
	R^2 (%)	f.p.	f.n.	R^2 (%)	f.p.	f.n.
LMLFM	**92±1**	**0.2**	0.8	**88±2**	2.2	4.2
rPQL	88±2	20.6	**0**	-	-	-
M-LMM	90±1	92	**0**	-	-	-
GLMMLASSO	83±4	91	**0**	-	-	-
LMMLASSO	88±2	92	**0**	-	-	-
LASSO	88±1	42.4	**0**	84±4	415.8	0.4
MLLASSO	40±8	23.8	0.8	1±1	**0**	6.2

Remark. (i) All longitudinal baselines fail when the number of variables exceeds 100. (ii) Applying the wrong correlation to data generally leads to poor regression accuracy. LMLFM is the only model that can detect the correlation type.

Results on real-life data

Method	SWAN			GSS			
	R^2_- (%)	f.p.$_-$	f.n.$_-$	R^2_+ (%)	R^2_- (%)	f.p.$_-$	f.n.$_-$	R^2_+ (%)
LMLFM	**30±4**	**0**	**0**	**49±2**	**17±2**	2	**0**	**55±2**
M-LMM	29±2	1	5	-	16±1	2	1	-
GLMMLASSO	19±3	**0**	2	-	-	-	-	-
LMMLASSO	26±3	2	2	-	-	-	-	-
PGEE	25±4	2	5	-	3±1	**1**	**0**	-
LASSO	21±4	1	2	47±1	0±1	**1**	**0**	47±1
FM	29±3	**0**	5	45±2	12±4	4	2	31±2
RF	23±5	10	**0**	47±1	4±1	10	**0**	**55±2**
MLLASSO	0±2	10	**0**	20±2	-42±10	4	**0**	-2±1

Remark. Variables selected by our model largely coincide with the findings in related domain literature.

Contact

Email: jul672@ist.psu.edu

WeChat QR Code

LinkedIn QR Code