INTERACTION MODELING OF HIGH-DIMENSIONAL DATA WITH THE BIAS AND HIGH CORRELATION- SPARSE FACTORIZATION MACHINES APPROACH

Haruki Ozawa, Taiki Ito, Takuya Matsuzawa, Sara Hoshino, Kyo Watanabe, Sumika Arima(Tsukuba University)

Introduction

・Treasure hunting is needed as industry data becomes diverse and voluminous, and their combinations of features (interactions).

・ The key is to efficient selection and estimation for interactions.

・FMs (Factorization Machines): highly efficient interaction modeling of a sparse data originally for recommendations. It is also easy to use.

Aim

Our previous study [3] has been clarified FMs [1], Sparse FMs (SFM (L2)) [2], and our extended SFM (L1) [3] are applicable to numerical-categorical mixed data. Its wider industrial uses can be expected.

However, High in False-Positive, the accuracy can deteriorate particularly for **highly correlated, high-dimensional, biased data**.

(1) Uncorrelated high-dimensional mixed data consisting of half numerical and half categorical variables.

→**Adaptive SFM**

Compared to the False-Positive of interaction (FP_{int}) of previous method, it is reduced by **96.96%**

(2) Highly correlated data consisting only of numerical data with three levels of strength of correlated parts in the data: weak, medium, and strong.

→**CHANOL-SFM**

We obtained a model that is better at estimating interactions than previous models.

[1] Rendle, S (2010). Factorization Machines, IEEE International conference data mining 2010.
[2] Atarashi, K., Oyama, S., and Kurihara, M. (2021). Factorization Machines with Regularization for Sparse Feature Interactions, Journal of Machine Learning Research 22.153. 1-50.
[3]. Hoshino, S., Watanabe, K., and Arima, S. (2023). Interaction Modelling of High Dimensional Production Data, Procedia Computer Science, 225. Elsevier. 4055-4064.

Methods

<Previous method>

Factorization Machines(FMs)

FMs expresses the weight matrix of an interaction by the inner product of a matrix and its transpose. It is characterized by its high efficiency and adaptability to categorical data.

$$\begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_1, \boldsymbol{v}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{v}_n, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_n, \boldsymbol{v}_n \rangle \end{bmatrix}$$

(Shape of the interaction matrix P)

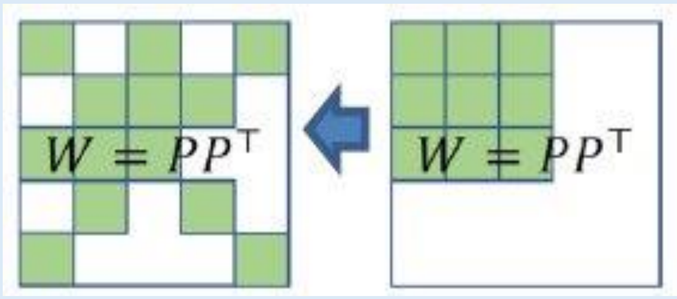
$$\hat{y}(x) \coloneqq \omega + \sum_{i=1}^{n} w x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle V_i, V_j \rangle x_i x_j$$

$$where: \omega \in \mathbb{R}, w \in \mathbb{R}^n, V \in \mathbb{R}^{n \times k}$$

Sparse Factorization Machines(SFM) [2]

SFM realizes sparse estimation capabilities to Factorization Machines.

Triangular Inequality (TI) term $(\gamma\ \Omega_{TI}(P))$ realizes unstructured interaction matrices efficiently



(Unstructured Matrix)

$$L(w, P) \coloneqq \frac{1}{N} \sum_{n=1}^{N} l(y_n, \hat{y}_n) + \frac{\alpha}{2} \|w\|_2^2 + \frac{\beta}{2} \|P\|_2 + \gamma \Omega_{TI}(P)$$

$$\Omega_{TI}(P) = \sum_{i=1}^{d} \sum_{j=i+1}^{d} \sum_{s=1}^{k} |p_{j,s}| |p_{i,s}| \geq \sum_{i=1}^{d} \sum_{j=i+1}^{d} \sum_{s=1}^{k} |p_{j,s}| |p_{i,s}| = \Omega_*(P)$$

Proposal methods

SFM with L_1norm $(TI - SFM(L_1))$ [3] *Our Lab's study

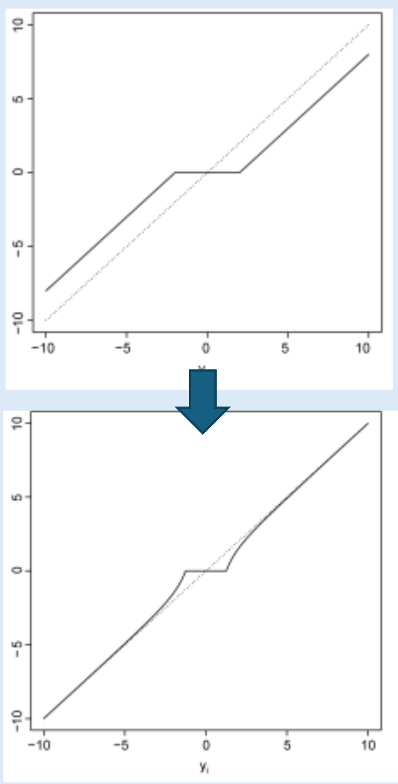
There is an SFM that performs sparse estimation for main effects, which we have independently improved$\left(TI - SFM(L_1)\right)$, and the proposed method is an application of $TI - SFM(L_1)$.

$$L(w, P) \coloneqq \frac{1}{N} \sum_{n=1}^{N} l(y_n, \hat{y}_n) + \frac{\alpha}{2} \|w\|_1 + \frac{\beta}{2} \|P\|_2 + \gamma \Omega_{TI}(P)$$

Adaptive SFM (L1)

A new adaptive approach to introduce an additional dynamic weighting ω_j to the TI term to estimate the combinatorial features of FP_{int} are expected to be automatically screened by assigning a dominance as an additional penalty based on preprocessed estimates (ω) for Y and X or P, since the estimates for the false features are relatively close to zero.

Ref. In previous works, Adaptive Lasso or so introduces the weight just for main factors' estimations (Right Fig).

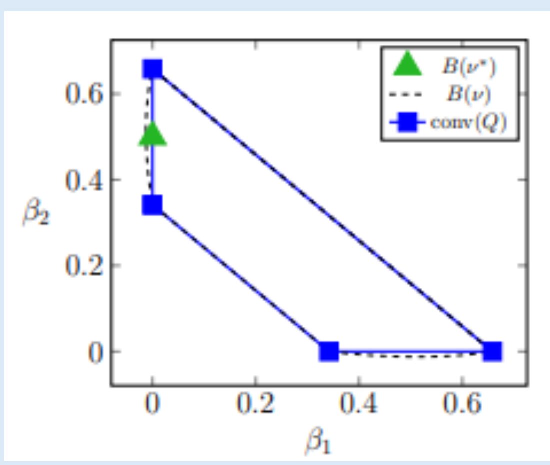


(Adaptive weightings' effects Lasso → Adaptive Lasso)

$$L(w, P) \coloneqq \frac{1}{N} \sum_{n=1}^{N} l(y_n, \hat{y}_n) + \frac{\alpha}{2} \|w\|_1 + \frac{\beta}{2} \|P\|_2 + \omega_j \gamma \Omega_{TI}(P)$$

CHANOL-SFM (L1)

CHANOL is an advanced method of *LASSO** regression and good for a data with bias and highly correlated variables. Multiple relaxed solutions (models) are generated to obtain an unbiased solution instead of applying an ensemble method. The solution set is used for the best representation by collective intelligence.



(Green: Lasso solution, Blue: Relaxation Solutions)

LASSO: Least Absolute Shrinkage and Selection Operator

Result

A comparison was made between the proposed method and the existing method.

(1) **Adaptive-SFM**

Number of samples (n=1000)

Number of dimensions (p=5000).

The number of correct answers is shown in "Oracle" below

Weighting Pattern	*Oracle*	$TI - SFM$ (L_2) [2]	$TI - SFM$ (L_1)	*Adaptive* $-SFM$
TP_{main}	30	30	30	30
FP_{main}	0	9959 -99.99% →	4	31
FN_{main}	0	0	0	0
TP_{int}	5	5	5	5
FP_{int}	0	1600359	1576145 → -96.96%	47965
FN_{int}	0	0	0	0

(2) **CHANOL-SFM**

Number of samples (n=1000)

Number of dimensions (p=2500)

The number of correct answers is shown in "Oracle" below

Weighting Pattern	*Oracle*	$TI - SFM$ (L_2) [2]	$TI - SFM$ (L_1)	*CHANOL* $- SFM$
TP_{main}	12	12	10	10
FP_{main}	0	2488 → -90.23%	243	232
FN_{main}	0	0	2	2
TP_{int}	30	6	8	9
FP_{int}	0	55603	48561 → -19.96%	38866
FN_{int}	0	24	22	21

Discussion and Future work

(1) $TI\_SFM(L_1)$ with L_1 norm **reduce 99% of FP_{main}** while keeping **TP_{main} and TP_{int}at 100%. Furthermore, Adaptive SFM** can automatically improve **FP_{int} by 96.96%**

・Benchmark of parameter γ and simultaneous optimization of w_j as a new dynamic adaptive approach of interactions.

(2) **CHANOL-SFM** method improves $\boldsymbol{FP}$, compared to $TI - SFM\ (L_1)$

・$CHANOL - SFM$ yet does not lead to a dramatic level of $\boldsymbol{FP}$ of the interaction effect

Contact Information Ozawa Haruki : s2420417@u.tsukuba.ac.jp Sumika Arima : arima@sk.tsukuba.ac.jp