Field-aware Factorization Machines

YuChin Juan, Yong Zhuang, and Wei-Sheng Chin

NTU CSIE MLGroup

Recently, field-aware factorization machines (FFM) have been used to win two click-through rate prediction competitions hosted by Criteo[1] and Avazu[2]. In these slides we introduce the formulation of FFM together with well known linear model, degree-2 polynomial model, and factorization machines.

To use this model, please download LIBFFM at:

`http://www.csie.ntu.edu.tw/~r01922136/libffm`

[1] https://www.kaggle.com/c/criteo-display-ad-challenge
[2] https://www.kaggle.com/c/avazu-ctr-prediction

Linear Model

The formulation of linear model is:

$$\phi(\mathbf{w}, \mathbf{x}) = \mathbf{w}^T \mathbf{x} = \sum_{j \in C_1} w_j x_j, [3]$$

where $\mathbf{w}$ is the model, $\mathbf{x}$ is a data instance, and C_1 is the non-zero elements in $\mathbf{x}$.

[3] The bias term is not included in these slides.

Degree-2 Polynomial Model (Poly2)

The formulation of Poly2 is:

$$\phi(\mathbf{w}, \mathbf{x}) = \sum_{j_1, j_2 \in C_2} w_{j_1, j_2} x_{j_1} x_{j_2},^{4}$$

where C_2 is the 2-combination of non-zero elements in $\mathbf{x}$.

[4] The linear terms and the bias term are not included in these slides.

Factorization Machines[6] (FM)

The formulation of FM is:

$$\phi(\mathbf{w}, \mathbf{x}) = \sum_{j_1, j_2 \in C_2} \langle \mathbf{w}_{j_1}, \mathbf{w}_{j_2} \rangle x_{j_1} x_{j_2},^5$$

where $\mathbf{w}_{j_1}$ and $\mathbf{w}_{j_2}$ are two vectors with length k, and k is a user-defined parameter.

---

[5] The linear terms and the bias term are not included in these slides.

[6] This model is proposed in [Rendle, 2010].

Field-aware Factorization Machines[8] (FFM)

The formulation of FFM is:

$$\phi(\mathbf{w}, \mathbf{x}) = \sum_{j_1, j_2 \in C_2} \langle \mathbf{w}_{j_1, f_2}, \mathbf{w}_{j_2, f_1} \rangle x_{j_1} x_{j_2}, \text{[7]}$$

where f_1 and f_2 are respectively the fields of j_1 and j_2, and $\mathbf{w}_{j_1,f_2}$ and $\mathbf{w}_{j_2,f_1}$ are two vectors with length k.

[7] The linear terms and the bias term are not included in these slides.

[8] This model is used in [Jahrer et al., 2012]; a similar model is proposed in [Rendle and Schmidt-Thieme, 2010].

FFM for Logistic Loss

The optimization problem is:

$$\min_{\mathbf{w}} \quad \sum_{i=1}^{L} \Big(\log \big(1 + \exp(-y_i \phi(\mathbf{w}, \mathbf{x}_i)\big) + \frac{\lambda}{2}\|\mathbf{w}\|^2 \Big),$$

where

$$\phi(\mathbf{w}, \mathbf{x}) = \sum_{j_1, j_2 \in C_2} \langle \mathbf{w}_{j_1, f_2}, \mathbf{w}_{j_2, f_1} \rangle x_{j_1} x_{j_2},$$

L is the number of instances, and λ is regularization parameter.

A Concrete Example

Consider the following example:

User (Us)	Movie (Mo)	Genre (Ge)	Pr (Pr)
YuChin (YC)	3Idiots (3I)	Comedy,Drama (Co,Dr)	$9.99

Note that “User,” “Movie,” and “Genre” are categorical variables, and “Price” is a numerical variable.

A Concrete Example

Conceptually, for linear model, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$w_{\text{Us-Yu}} \cdot x_{\text{Us-Yu}} + w_{\text{Mo-3I}} \cdot x_{\text{Mo-3I}} + w_{\text{Ge-Co}} \cdot x_{\text{Ge-Co}} + w_{\text{Ge-Dr}} \cdot x_{\text{Ge-Dr}} + w_{\text{Pr}} \cdot x_{\text{Pr}},$$

where $x_{\text{Us-Yu}} = x_{\text{Mo-3I}} = x_{\text{Ge-Co}} = x_{\text{Ge-Dr}} = 1$ and $x_{\text{Pr}} = 9.99$. Note that because "User," "Movie," and "Genre" are categorical variables, the values are all ones.[9]

[9] If preprocessing such as instances-wise normalization is conducted, the values may not be ones.

A Concrete Example

For Poly2, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
& w_{\text{Us-Yu-Mo-3I}} \cdot x_{\text{Us-Yu}} \cdot x_{\text{Mo-3I}} + w_{\text{Us-Yu-Ge-Co}} \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Co}} + w_{\text{Us-Yu-Ge-Dr}} \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Dr}} + w_{\text{Us-Yu-Pr}} \cdot x_{\text{Us-Yu}} \cdot x_{\text{Pr}} \\
& + w_{\text{Mo-3I-Ge-Co}} \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Co}} + w_{\text{Mo-3I-Ge-Dr}} \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Dr}} + w_{\text{Mo-3I-Pr}} \cdot x_{\text{Mo-3I}} \cdot x_{\text{Pr}} \\
& + w_{\text{Ge-Co-Ge-Dr}} \cdot x_{\text{Ge-Co}} \cdot x_{\text{Ge-Dr}} + w_{\text{Ge-Co-Pr}} \cdot x_{\text{Ge-Co}} \cdot x_{\text{Pr}} \\
& + w_{\text{Ge-Dr-Pr}} \cdot x_{\text{Ge-Dr}} \cdot x_{\text{Pr}}
\end{aligned}
$$

A Concrete Example

For FM, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
&\langle \mathbf{w}_{\text{Us-Yu}}, \mathbf{w}_{\text{Mo-3I}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Mo-3I}} + \langle \mathbf{w}_{\text{Us-Yu}}, \mathbf{w}_{\text{Ge-Co}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Co}} + \langle \mathbf{w}_{\text{Us-Yu}}, \mathbf{w}_{\text{Ge-Dr}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Us-Yu}}, \mathbf{w}_{\text{Pr}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Mo-3I}}, \mathbf{w}_{\text{Ge-Co}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Co}} + \langle \mathbf{w}_{\text{Mo-3I}}, \mathbf{w}_{\text{Ge-Dr}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Mo-3I}}, \mathbf{w}_{\text{Pr}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Ge-Co}}, \mathbf{w}_{\text{Ge-Dr}} \rangle \cdot x_{\text{Ge-Co}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Ge-Co}}, \mathbf{w}_{\text{Pr}} \rangle \cdot x_{\text{Ge-Co}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Ge-Dr}}, \mathbf{w}_{\text{Pr}} \rangle \cdot x_{\text{Ge-Dr}} \cdot x_{\text{Pr}}
\end{aligned}
$$

A Concrete Example

For FFM, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
&\langle \mathbf{w}_{\text{Us-Yu,Mo}}, \mathbf{w}_{\text{Mo-3I,Us}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Mo-3I}} + \langle \mathbf{w}_{\text{Us-Yu,Ge}}, \mathbf{w}_{\text{Ge-Co,Us}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Co}} + \langle \mathbf{w}_{\text{Us-Yu,Ge}}, \mathbf{w}_{\text{Ge-Dr,Us}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Us-Yu,Pr}}, \mathbf{w}_{\text{Pr,Us}} \rangle \cdot x_{\text{Us-Yu}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Mo-3I,Ge}}, \mathbf{w}_{\text{Ge-Co,Mo}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Co}} + \langle \mathbf{w}_{\text{Mo-3I,Ge}}, \mathbf{w}_{\text{Ge-Dr,Mo}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Mo-3I,Pr}}, \mathbf{w}_{\text{Pr,Mo}} \rangle \cdot x_{\text{Mo-3I}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Ge-Co,Ge}}, \mathbf{w}_{\text{Ge-Dr,Ge}} \rangle \cdot x_{\text{Ge-Co}} \cdot x_{\text{Ge-Dr}} + \langle \mathbf{w}_{\text{Ge-Co,Pr}}, \mathbf{w}_{\text{Pr,Ge}} \rangle \cdot x_{\text{Ge-Co}} \cdot x_{\text{Pr}} \\
&+ \langle \mathbf{w}_{\text{Ge-Dr,Pr}}, \mathbf{w}_{\text{Pr,Ge}} \rangle \cdot x_{\text{Ge-Dr}} \cdot x_{\text{Pr}}
\end{aligned}
$$

A Concrete Example

In practice we need to map these features into numbers. Say we have the following mapping.

Field name		Field index	Feature name		Feature index
User	→	field 1	User-YuChin	→	feature 1
Movie	→	field 2	Movie-3Idiots	→	feature 2
Genre	→	field 3	Genre-Comedy	→	feature 3
Price	→	field 4	Genre-Drama	→	feature 4
			Price	→	feature 5

After transforming to the LIBFFM format, the data becomes:

1:1:1 2:2:1 3:3:1 3:4:1 4:5:9.99

Here a red number is an index of field, a blue number is an index of feature, and a green number is the value of the corresponding feature.

A Concrete Example

Now, for linear model, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$w_1 \cdot 1 + w_2 \cdot 1 + w_3 \cdot 1 + w_4 \cdot 1 + w_5 \cdot 9.99$$

A Concrete Example

For Poly2, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
w_{1,2} \cdot 1 \cdot 1 + w_{1,3} \cdot 1 \cdot 1 + w_{1,4} \cdot 1 \cdot 1 + w_{1,5} \cdot 1 \cdot 9.99 \\
+ w_{2,3} \cdot 1 \cdot 1 + w_{2,4} \cdot 1 \cdot 1 + w_{2,5} \cdot 1 \cdot 9.99 \\
+ w_{3,4} \cdot 1 \cdot 1 + w_{3,5} \cdot 1 \cdot 9.99 \\
+ w_{4,5} \cdot 1 \cdot 9.99
\end{aligned}
$$

A Concrete Example

For FM, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
&\langle \mathbf{w}_1, \mathbf{w}_2 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_1, \mathbf{w}_3 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_1, \mathbf{w}_4 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_1, \mathbf{w}_5 \rangle \cdot 1 \cdot 9.99 \\
&+ \langle \mathbf{w}_2, \mathbf{w}_3 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_2, \mathbf{w}_4 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_2, \mathbf{w}_5 \rangle \cdot 1 \cdot 9.99 \\
&+ \langle \mathbf{w}_3, \mathbf{w}_4 \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_3, \mathbf{w}_5 \rangle \cdot 1 \cdot 9.99 \\
&+ \langle \mathbf{w}_4, \mathbf{w}_5 \rangle \cdot 1 \cdot 9.99
\end{aligned}
$$

A Concrete Example

For FFM, $\phi(\mathbf{w}, \mathbf{x})$ is:

$$
\begin{aligned}
\langle \mathbf{w}_{1,2}, \mathbf{w}_{2,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{1,3}, \mathbf{w}_{3,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{1,3}, \mathbf{w}_{4,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{1,4}, \mathbf{w}_{5,1} \rangle \cdot 1 \cdot 9.99 \\
+ \langle \mathbf{w}_{2,3}, \mathbf{w}_{3,2} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{2,3}, \mathbf{w}_{4,2} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{2,4}, \mathbf{w}_{5,2} \rangle \cdot 1 \cdot 9.99 \\
+ \langle \mathbf{w}_{3,3}, \mathbf{w}_{4,3} \rangle \cdot 1 \cdot 1 + \langle \mathbf{w}_{3,4}, \mathbf{w}_{5,3} \rangle \cdot 1 \cdot 9.99 \\
+ \langle \mathbf{w}_{4,4}, \mathbf{w}_{5,3} \rangle \cdot 1 \cdot 9.99
\end{aligned}
$$

Jahrer, M., Tscher, A., Lee, J.-Y., Deng, J., Zhang, H., and Spoelstra, J. (2012).
Ensemble of collaborative filtering and feature engineered models for click through rate prediction.

Rendle, S. (2010).
Factorization machines.

Rendle, S. and Schmidt-Thieme, L. (2010).
Pairwise interaction tensor factorization for personalized tag recommendation.