Recent Works on Feature Interaction

(ingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statisti

.

Models

Tree-based ensembels

Neuron Netwo

Factorization Machines

neuron networks

Summary

Recent Works on Feature Interaction

Xingzhi Sun

https://github.com/xingzhis/XAI

Wednesday 30th September, 2020

Table of Contents

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction Post-hoc statistic

Mode

GA²M Neuron Networks Factorization Machines Hybrid models with

Summai

1 Definitions of feature interaction

- Post-hoc statistic
- Model-specific

2 Models

- Tree-based ensembels
- GA²M
- Neuron Networks
- Factorization Machines
- Hybrid models with neuron networks

Table of Contents

Recent Works on Feature Interaction

Xingzhi Sı

Outlin

Definitions of feature interaction

Post-hoc statistic Model-specific

Mode

Tree-based ensemb

Neuron Networks

Machines
Hybrid models wi

Summa

1 Definitions of feature interaction

- Post-hoc statistic
- Model-specific

2 Models

- Tree-based ensembels
- \blacksquare GA 2 M
- Neuron Networks
- Factorization Machines
- Hybrid models with neuron networks

Non-additiveness

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Xingzhi Sun

S ...

Outline

Definitions of feature interaction

Post-hoc statistic

Models

Tree-based ensem

GA⁻M

Machines

Hybrid models with

Hybrid models wineuron networks

Summary

Non-additiveness

 $F(\mathbf{x})$ cannot be written in the form of $F(\mathbf{x}) = f_j\left(x_j\right) + f_{\backslash j}\left(\mathbf{x}_{\backslash j}\right)$

Non-additiveness

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Xingzhi Sun

Outline

Definitions of feature interaction

Post-hoc statistic Model-specific

Mode

Tree-based ensembels

Neuron Networks
Factorization
Machines
Hybrid models wit

Summary

Non-additiveness

 $F(\mathbf{x})$ cannot be written in the form of $F(\mathbf{x}) = f_j\left(x_j\right) + f_{\backslash j}\left(\mathbf{x}_{\backslash j}\right)$

the effect of both variables

VS

the sum of effects of each variable

Non-additiveness

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statistic Model-specific

Mode

Tree-based ensembe

Neuron Networks
Factorization
Machines
Hybrid models wit

Summar

Non-additiveness

$$F(\mathbf{x})$$
 cannot be written in the form of $F(\mathbf{x}) = f_j\left(x_j\right) + f_{\backslash j}\left(\mathbf{x}_{\backslash j}\right)$

the effect of both variables

VS

the sum of effects of each variable

let
$$F_{s}\left(\mathbf{x}_{s}\right)=\mathbb{E}_{\mathbf{x}_{\backslash s}}\left[F\left(\mathbf{x}_{s},\mathbf{x}_{\backslash s}\right)\right]$$

Predictive learning via rule ensembles

$$H_{jk}^{2} = \sum_{i=1}^{N} \left[\hat{F}_{jk}(x_{ij}, x_{ik}) - \hat{F}_{j}(x_{ij}) - \hat{F}_{k}(x_{ik}) \right]^{2} / \sum_{i=1}^{N} \hat{F}_{jk}^{2}(x_{ij}, x_{ik})$$

Model Expressiveness

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Xingzhi Sur

Outline

Definitions of feature interaction

Post-hoc statistic

Models

Tree-based ensembe

Neuron Networks
Factorization
Machines
Hybrid models with

Summary

the accurate model that contains the interaction

٧S

the accurate model that bans the interaction

Model Expressiveness

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Outlin

Definitions of feature interaction

Post-hoc statistic Model-specific

Mode

Tree-based ensembel

Factorization Machines Hybrid models with neuron networks

Summar

the accurate model that contains the interaction

VS

the accurate model that bans the interaction

 $F^*(x)$: Target function

F(x): Highly accurate model

 $R_{ij}(x)$: Highly accurate but ban interaction between x_i and x_j

 $\mathrm{stRMSE}(F(\mathbf{x})) = \frac{\mathrm{RMSE}(F(\mathbf{x}))}{\mathrm{StD}(F^*(\mathbf{x}))}$,

Detecting statistical interactions with additive groves of trees

 $I_{ij}(F(\mathbf{x})) = \text{stRMSE}(F(\mathbf{x})) - \text{stRMSE}(R_{ij}(\mathbf{x}))$

Hessian - second derivative

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Outline

D. C. 111...

feature interaction

Post-hoc statistic

Models

Tree-based ensemb

Neuron Networks
Factorization
Machines
Hybrid models with

Summary

The prediction has non-zero hessian over the interaction variables.

 $\frac{\partial F(\mathbf{x})}{\partial x_i \partial x_j} \neq 0$

Hessian - second derivative

Post-hoc statistic for interaction

Recent Works on Feature Interaction

Xingzhi Sur

Outline

Definitions of feature interaction

Post-hoc statistic

Mode

GA²M Neuron Networks Factorization

c

The prediction has non-zero hessian over the interaction variables.

$$\frac{\partial F(\mathbf{x})}{\partial x_i \partial x_j} \neq 0$$

Learning Global Pairwise Interactions with Bayesian Neural Networks

$$\begin{aligned} & \mathrm{EAH}_{g}^{i,j}(\mathbf{W}) = \mathbb{E}_{p(\mathbf{x})} \left[\left| \frac{\partial^{2} g^{\mathbf{W}}(\mathbf{x})}{\partial x_{i} \partial x_{j}} \right| \right] \\ & \mathrm{AEH}_{g}^{i,j}(\mathbf{W}) = \left| \mathbb{E}_{p(\mathbf{x})} \left[\frac{\partial^{2} g^{\mathbf{W}}(\mathbf{x})}{\partial x_{i} \partial x_{j}} \right] \right| \end{aligned}$$

Table of Contents

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statis

Model-specific

Tree-based ensemb

Neuron Networks

Machines

Hybrid models wi

Summa

1 Definitions of feature interaction

- Post-hoc statistic
- Model-specific

2 Models

- Tree-based ensembels
- \blacksquare GA 2 M
- Neuron Networks
- Factorization Machines
- Hybrid models with neuron networks

Model-specific depiction of interaction

Recent Works on Feature Interaction

(ingzhi Su

Outlin

Definitions of feature interaction

Model-specific

Mode

Tree-based ensembels

Factorization Machines Hybrid models wit

Summary

- Explicit: parameters, such as weights of interaction terms.
- Implicit: Neuron networks, embedding vectors, etc.

Table of Contents

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Model-specific

Mode

Tree-based ensembels

Neuron Netwo

Factorization Machines Hybrid models wit

Summa

- 1 Definitions of feature interaction
 - Post-hoc statistic
 - Model-specific
- 2 Models
 - Tree-based ensembels
 - GA²M
 - Neuron Networks
 - Factorization Machines
 - Hybrid models with neuron networks

Tree-based ensembels

Recent Works on Feature Interaction

Xingzhi Sun

Outline

Definitions of feature interaction

Post-hoc statis

Models

Tree-based ensembels

GA²M

Machines
Hybrid models wit

Summary

Main effects: Linear

Interaction: Rules derived from decision trees

Tree-based ensembels

Recent Works on Feature Interaction

Xingzhi Sui

Outline

Definitions o feature interaction Post-hoc statistic Model-specific

Models

Tree-based ensembels

Factorization Machines Hybrid models with

Summary

Main effects: Linear

■ Interaction: Rules derived from decision trees

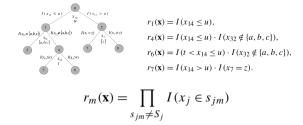


Figure: A desicion tree and its corresponding rule term

Tree-based ensembels

Recent Works on Feature Interaction

Aingzni Si

Outlin

feature interaction

Post-hoc statistic

Mode

Tree-based ensembels

Neuron Networks
Factorization
Machines

Summa

Main effects: Linear

Interaction: Rules derived from decision trees

$$r_{1}(\mathbf{x}) = I(x_{14} \le u),$$

$$r_{2}(\mathbf{x}) = I(x_{14} \le u),$$

$$r_{3}(\mathbf{x}) = I(x_{14} \le u),$$

$$r_{4}(\mathbf{x}) = I(x_{14} \le u) \cdot I(x_{32} \notin \{a, b, c\}),$$

$$r_{6}(\mathbf{x}) = I(t < x_{14} \le u) \cdot I(x_{32} \notin \{a, b, c\}),$$

$$r_{7}(\mathbf{x}) = I(x_{14} \le u) \cdot I(x_{32} \notin \{a, b, c\}),$$

$$r_{7}(\mathbf{x}) = I(x_{14} \le u) \cdot I(x_{7} = z).$$

$$r_m(\mathbf{x}) = \prod_{s_{jm} \neq S_j} I(x_j \in s_{jm})$$

Figure: A desicion tree and its corresponding rule term

RuleFit (Friedman and Popescu, 2008)

$$F(\mathbf{x}) = \hat{a}_0 + \sum_{k=1}^{K} \hat{a}_k r_k(\mathbf{x}) + \sum_{j=1}^{n} \hat{b}_j l_j(x_j)$$

Learn the model with regularized regression.

Tree-based ensembels

Recent Works on Feature Interaction

(ingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statist

Models

Tree-based ensembels

GA²M Neuron Networks Factorization Machines

Summan

RuleFit (Friedman and Popescu, 2008)

$$F(\mathbf{x}) = \hat{a}_0 + \sum_{k=1}^{K} \hat{a}_k r_k(\mathbf{x}) + \sum_{j=1}^{n} \hat{b}_j l_j(x_j)$$

Tree-based ensembels

Recent Works on Feature Interaction

(ingzhi S

Outlin

Definitions of feature interaction Post-hoc statistic

Mode

Tree-based ensembels

Factorization
Machines
Hybrid models with

Summar

RuleFit (Friedman and Popescu, 2008)

$$F(\mathbf{x}) = \hat{a}_0 + \sum_{k=1}^{K} \hat{a}_k r_k(\mathbf{x}) + \sum_{j=1}^{n} \hat{b}_j l_j(x_j)$$

statistic for interaction

$$H_{jk}^{2} = \sum_{i=1}^{N} \left[\hat{F}_{jk} (x_{ij}, x_{ik}) - \hat{F}_{j} (x_{ij}) - \hat{F}_{k} (x_{ik}) \right]^{2} / \sum_{i=1}^{N} \hat{F}_{jk}^{2} (x_{ij}, x_{ik})$$

Detect interaction with significant H_{jk} , whose distribution is obtained by bootstrapping.

Detecting statistical interactions with additive groves of trees

Tree-based ensembels

Recent Works on Feature Interaction

Tree-based ensembels

the **accurate** model that contains the interaction

VS

the accurate model that bans the interaction

statistic for interaction

$$I_{ij}(F(\mathbf{x})) = \text{stRMSE}(F(\mathbf{x})) - \text{stRMSE}(R_{ij}(\mathbf{x}))$$

Detecting statistical interactions with additive groves of trees

Tree-based ensembels

Recent Works on Feature Interaction

Outline

Definitions of feature

interaction

Post-hoc statistic

Model-specific

Mode

GA² M Neuron Networks Factorization Machines

Tree-based ensembels

Summarv

the **accurate** model that contains the interaction

VS

the **accurate** model that bans the interaction

statistic for interaction

$$I_{ij}(F(\mathbf{x})) = \text{stRMSE}(F(\mathbf{x})) - \text{stRMSE}(R_{ij}(\mathbf{x}))$$

- Obtain a highly accurate model through a tree ensemble that is later bagged: $F_0(x) = \sum_{i=1}^K T_i(x)$
- Restrict interaction by forbidding one of the interacting variables when growing a tree.

Detecting statistical interactions with additive groves of trees Tree-based ensembels

Recent Works on Feature Interaction

Xingzhi Sun

Outlin

Definitions of feature interaction

Post-hoc statistic

Mode

GA²M Neuron Networks Factorization Machines Hybrid models with

Tree-based ensembels

Summa

the **accurate** model that contains the interaction

VS

the accurate model that bans the interaction

statistic for interaction

$$I_{ij}(F(\mathbf{x})) = \text{stRMSE}(F(\mathbf{x})) - \text{stRMSE}(R_{ij}(\mathbf{x}))$$

- Obtain a highly accurate model through a tree ensemble that is later bagged: $F_0(x) = \sum_{i=1}^K T_i(x)$
- Restrict interaction by forbidding one of the interacting variables when growing a tree.

Beats the Rulefit statistic in avoiding spurious interaction at sparse regions.

Table of Contents

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statistic

Mode

Tree-based ensembe

GA⁺M

Factorization Machines Hybrid models with neuron networks

Summai

- 1 Definitions of feature interaction
 - Post-hoc statistic
 - Model-specific
- 2 Models
 - Tree-based ensembels
 - GA²M
 - Neuron Networks
 - Factorization Machines
 - Hybrid models with neuron networks

Accurate Intelligible Models with Pairwise Interactions $_{\mathsf{GA}^2\mathsf{M}}$

Recent Works on Feature Interaction

Outline

Definitions of feature

Post-hoc statis

Model-specific

Models

GA²M

Factorization Machines Hybrid models with

Summary

Main effects: GAM.

■ Interaction: 2D bin functions f_{ij} on the residual of GAM.

Accurate Intelligible Models with Pairwise Interactions GA²M

Recent Works on Feature Interaction

Xingzhi Sur

Outline

Definitions of feature interaction

Post-hoc statistic

Model-specific

Mode

Tree-based ensemb

Neuron Networks
Factorization
Machines
Hybrid models with
neuron networks

Summar

Main effects: GAM.

■ Interaction: 2D bin functions f_{ij} on the residual of GAM.

FAST

Speed up the calculation of bin averages: pre-caluclate a CDF lookup table for reusing.

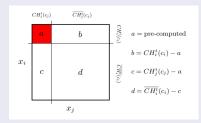


Table of Contents

Recent Works on Feature Interaction

Xingzhi Sui

Outlin

Definitions of feature interaction

Post-hoc statistic

Mode

GA²M

Neuron Networks

Machines

Hybrid models wit

Summai

- 1 Definitions of feature interaction
 - Post-hoc statistic
 - Model-specific
- 2 Models
 - Tree-based ensembels
 - GA²M
 - Neuron Networks
 - Factorization Machines
 - Hybrid models with neuron networks

Learning Global Pairwise Interactions with Bayesian Neural Networks Neuron Networks

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statistic

Mode

Tree-based ensembe

Neuron Networks

Machines
Hybrid models wi

Summary

Definition of feature interaction for smooth models

The prediction has non-zero hessian over the interaction variables.

$$\frac{\partial F(\mathbf{x})}{\partial x_i \partial x_j} \neq 0$$

Learning Global Pairwise Interactions with Bayesian Neural Networks

Recent Works on Feature Interaction Neuron Networks

Xingzhi Su

Outlin

Definitions o feature interaction Post-hoc statistic Model-specific

Mode

Tree-based ensembel

Neuron Networks

Factorization Machines Hybrid models wit

Summar

Definition of feature interaction for smooth models

The prediction has non-zero hessian over the interaction variables.

$$\frac{\partial F(\mathbf{x})}{\partial x_i \partial x_j} \neq 0$$

In practice, we take the expectation of the hessian over the distribution of x

Recent Works on Feature Interaction

Aingzni S

Outlin

Definitions o feature interaction Post-hoc statistic

Mode

GA²M

Neuron Networks

Factorization

Machines

c......

Two ways of taking expectation

$$\begin{split} & \mathrm{EAH}_g^{i,j}(\mathbf{W}) = \mathbb{E}_{p(\mathbf{x})} \left[\left| \frac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j} \right| \right] \text{ lowest FNR, highest FPR.} \\ & \mathrm{AEH}_g^{i,j}(\mathbf{W}) = \left| \mathbb{E}_{p(\mathbf{x})} \left[\frac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j} \right] \right| \text{ lowest FPR, highest FNR.} \end{split}$$

- \blacksquare EAH avoids (+,-) noise to cancel and could capture spurious interactions.
- AEH could have true interactions cancel out and fail to capture true interactions.

Learning Global Pairwise Interactions with Bayesian Neural Networks

Recent Works on Feature Interaction Neuron Networks

Xingzhi Sı

Outlin

Definitions of feature interaction

Model-specific

Mode

Tree-based ensemble

Neuron Networks

Factorization Machines Hybrid models wit

Summar

strike a balance: Group Expected Hessian (GEH)

$$M ext{-}\mathsf{GEH}_g^{i,j}(\mathbf{W}) = \sum_{m=1}^M rac{|A_m|}{\sum_{k=1}^M |A_k|} \left| \mathbb{E}_{p(\mathbf{x}|\mathbf{x}\in A_m)} \left[rac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j}
ight]
ight|$$

Partition the datapoints into M clusters, and expect the interaction is simmilar within each cluster, where only the noise is canceled out.

Neuron Networks

Summa

strike a balance: Group Expected Hessian (GEH)

$$M$$
-GEH $_g^{i,j}(\mathbf{W}) = \sum_{m=1}^{M} rac{|A_m|}{\sum_{k=1}^{M} |A_k|} \left| \mathbb{E}_{p(\mathbf{x}|\mathbf{x} \in A_m)} \left[rac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j}
ight] \right|$

Partition the datapoints into M clusters, and expect the interaction is simmilar within each cluster, where only the noise is canceled out.

Bayesian NN allows for the distribution of the M-GEH statistic and thus mean, std, confidence intervals, etc.

Detecting Statistical Interactions from Neural Network Weights

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction Post-hoc statistic Model-specific

Mode

Tree-based ensemble GA²M

Neuron Networks

Factorization Machines Hybrid models with neuron networks

Summary

How does a neuron network capture an interaction?

- Features share units of the first hidden layer.
- The shared units are passed to the output through descendent edges.

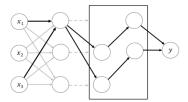


Figure: Interaction in an NN

Detecting Statistical Interactions from Neural Network Weights

Recent Works on Feature Interaction

Outline

Definitions of feature interaction

Modo

Tree-based ensembels $\mathsf{GA}^2\mathsf{M}$

Neuron Networks

Machines
Hybrid models wit

Summary

Write in the form of NN weights:

NID (Tsang et al., 2018)

$$\mathbf{z}^{(\ell)} = |\mathbf{w}^y|^\top \left| \mathbf{W}^{(L)} \right| \cdot \left| \mathbf{W}^{(L-1)} \right| \dots \left| \mathbf{W}^{(\ell+1)} \right|$$
$$\omega_i(\mathcal{I}) = z_i^{(1)} \mu \left(\left| \mathbf{W}_{i,\mathcal{I}}^{(1)} \right| \right)$$

Detecting Statistical Interactions from Neural Network Weights

Recent Works on Feature Interaction

Outlin

Definitions of feature interaction Post-hoc statistic Model-specific

Mode

GA² M

Neuron Networks

Factorization

Summai

Write in the form of NN weights:

NID (Tsang et al., 2018)

$$\mathbf{z}^{(\ell)} = |\mathbf{w}^{y}|^{\top} \left| \mathbf{W}^{(L)} \right| \cdot \left| \mathbf{W}^{(L-1)} \right| \dots \left| \mathbf{W}^{(\ell+1)} \right|$$
$$\omega_{i}(\mathcal{I}) = z_{i}^{(1)} \mu \left(\left| \mathbf{W}_{i,\mathcal{I}}^{(1)} \right| \right)$$

The algorithm is fast because only the features with top weights is considered in each iteration.

Neural Interaction Transparency: Disentangling Learned Interactions for Improved Interpretability Neuron Networks

Recent Works on Feature Interaction

Amgziii Jui

Outline

Definitions of feature interaction Post-hoc statistic Model-specific

Model

Tree-based ensemb

Neuron Networks

Factorization Machines Hybrid models wineuron networks

Summary

- The first hidden layer captures first-order interactions.
- The following layers capture higher-order interactions.
- But they entangle and could contain spurious interaction: $x_1x_2 + x_3x_4 \rightarrow x_1, x_2, x_3, x_4$ instead of x_1, x_2 and x_3, x_4



Figure: Entangled vs disentangeled

Neural Interaction Transparency: Disentangling Learned Interactions for Improved Interpretability Neuron Networks

Recent Works on Feature Interaction

Alligziii Ju

Outlin

Definitions of feature interaction Post-hoc statistic Model-specific

Model

GA²M

Neuron Networks

Factorization
Machines

Hybrid models with

Summar

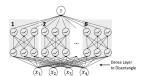
■ The first hidden layer captures first-order interactions.

- The following layers capture higher-order interactions.
- But they entangle and could contain spurious interaction: $x_1x_2 + x_3x_4 \rightarrow x_1, x_2, x_3, x_4$ instead of x_1, x_2 and x_3, x_4



Figure: Entangled vs disentangeled

Solution: Add a penalty on the weight matrix to control maximum allowed order of interaction.



Feature Interaction Interpretability: A Case for Explaining

Ad-Recommendation Systems via Neural Interaction Detection Neuron Networks

Recent Works on Feature Interaction

Outline

Definitions of feature interaction Post-hoc statistic

Models

Tree-based ensemb

Neuron Networks

Factorization Machines Hybrid models wineuron networks

Summarv

Given a black-box model, how to interprete global interaction?

- detect local interaction
- count the local interactions.
- interactions with many occurances are interpreted as global.

Feature Interaction Interpretability: A Case for Explaining

Ad-Recommendation Systems via Neural Interaction Detection Neuron Networks

Recent Works on Feature Interaction

Amgziii 3u

Outline

Definitions of feature interaction Post-hoc statistic Model-specific

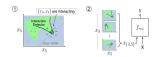
Tree-based ensembe

Neuron Networks
Factorization
Machines
Hybrid models with

Summai

Given a black-box model, how to interprete global interaction?

- detect local interaction
- count the local interactions.
- interactions with many occurances are interpreted as global.



GLIDER (Tsang et al., 2020)

- 1 For each data instance x, purturb to get a local dataset, and predict on that dataset.
- 2 on the local dataset, detect interaction with NID.
- 3 count and rank occurances of each interaction.

Table of Contents

Recent Works on Feature Interaction

Xingzhi Sui

Outlin

Definitions of feature interaction

Post-hoc statist

Mode

GA²M

Factorization

Hybrid models wit

Summa

- 1 Definitions of feature interaction
 - Post-hoc statistic
 - Model-specific
- 2 Models
 - Tree-based ensembels
 - GA²M
 - Neuron Networks
 - Factorization Machines
 - Hybrid models with neuron networks

Factorization machines

Factorization Machines

Recent Works on Feature Interaction

Outline

Definitions of feature

Post-hoc statis

Model-specific

Models

Tree-based ensem

GA W

Factorization

Hybrid models wit neuron networks

Summary

Background: Highly-sparse categorical data with one-hot coding. e.g. Shopping history, movie reviews.

Factorization machines

Factorization Machines

Recent Works on Feature Interaction

Factorization

Background: Highly-sparse categorical data with one-hot coding. e.g. Shopping history, movie reviews.

FM (Rendle, 2010)

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$
 learn the parameters with gradient descent

Each feature x_i corresponds to an *embedding vector* \mathbf{v}_i . Interaction: the inner products of the vectors $\langle \mathbf{v}_i, \mathbf{v}_i \rangle$.

Factorization machines

Factorization Machines

Recent Works on Feature Interaction

ingzhi Su

Outline

Definitions o feature interaction Post-hoc statistic

Mode

GA²M

Factorization Machines Hybrid models with

Summa

Background: Highly-sparse categorical data with one-hot coding. e.g. Shopping history, movie reviews.

FM (Rendle, 2010)

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$
 learn the parameters with gradient descent

Each feature x_i corresponds to an *embedding vector* \mathbf{v}_i . Interaction: the inner products of the vectors $\langle \mathbf{v}_i, \mathbf{v}_i \rangle$.

Advantages over regression (polynomial kernel SVM):

- Generalizes to instances that do not appear in the training set.
- linear complexity.

Variants of FM

Factorization Machines

Recent Works on Feature Interaction

Xingzhi Sι

Outlin

Definitions of feature interaction

Post-hoc statistic

Mode

Tree-based ensembel $\mathsf{GA}^2\mathsf{M}$

Factorization Machines Hybrid models wit neuron networks

Summa

Field-aware Factorization Machines for CTR Prediction

- Categorize features into fields: Clothes, Food, Electronics....
- 2 For each feature, instead of one embedding vector, learn a vector for each field.
- 3 When calculating interaction, use the vectors matching each other's field to take inner product.

$$\phi_{\text{FFM}}(\boldsymbol{w}, \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\boldsymbol{w}_{i, f_2} \cdot \boldsymbol{w}_j, f_1) x_i x_j$$

Attentional Factorization Machines

Instead of taking the inner product of the embedding vectors, take weighed outer product with weights learned by an Attention Neuron Network.

Table of Contents

Recent Works on Feature Interaction

Xingzhi Sui

Outlin

Definitions of feature interaction

Post-hoc statist

Mode

Tree-based ensembel GA²M

Neuron Networ

Hybrid models with neuron networks

Summa

- 1 Definitions of feature interaction
 - Post-hoc statistic
 - Model-specific
- 2 Models
 - Tree-based ensembels
 - GA²M
 - Neuron Networks
 - Factorization Machines
 - Hybrid models with neuron networks

Wide & Deep Learning for Recommender Systems

Hybrid models with neuron networks

Recent Works on Feature Interaction

Outline

Definitions of feature interaction Post-hoc statistic

Model

Tree-based ensemble

Neuron Netw Factorization

Hybrid models with neuron networks

Summai

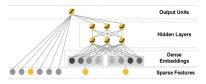


Figure: Wide & Deep

- Linear crossing model: *manual* low-order interactions.
- Deep model: *automatic* high-order interactions.

Deep & Cross Network for Ad Click Predictions

Hybrid models with neuron networks

Recent Works on Feature Interaction Xingzhi Sun

Outline

Definitions of feature interaction

Post-hoc statist Model-specific

Mode

GA²M

Neuron Netwo

Hybrid models with neuron networks

Summary

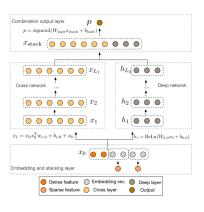


Figure: Deep & Cross

- Crossing network: *automatic* low-order interactions.
- Deep network: *automatic* high-order interactions.



DeepFM: A Factorization-Machine based Neural Network for CTR Prediction Hybrid models with neuron networks

Recent Works on Feature Interaction

Xingzhi Su

Outline

Definitions (
feature
interaction

Post-hoc statistic

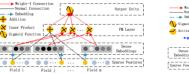
Mode

Tree-based ensemb

Neuron Netwo

Hybrid models with neuron networks

Summa



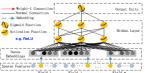


Figure: DeepFM

- FM: automatic low-order interactions.
- Deep model: *automatic* high-order interactions.

xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems Hybrid models with neuron networks

Recent Works on Feature Interaction

Outline

Definitions of feature interaction

Model

Tree-based ensembe

Factorization Machines Hybrid models with neuron networks

Summar

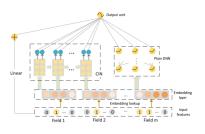


Figure: xDeepFM

- Linear model: main effects.
- Compressed Interaction Network: automatic low-order interactions: Interaction of each order goes to the output.
- Deep neuron network: automatic high-order interactions.



Deep Interest Network for Click-Through Rate Prediction

Attention Neuron Networks

Recent Works on Feature Interaction

Outline

Definitions of feature interaction Post-hoc statistic

Mode

Tree-based ensemb

Neuron Networks
Factorization
Machines
Hybrid models with

neuron networks Summary

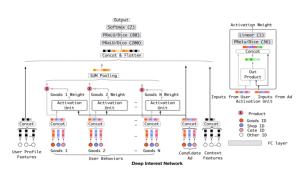


Figure: Deep Interest Network

The network comprises Activation Units, an attention network that leverages the user preference history and identify the true interest from diverse interests.

AutoInt: Automatic feature interaction learning via self-attentive neural networks Attention Neuron Networks

Recent Works on Feature Interaction

Xingzhi Su

Outlin

Definitions of feature interaction

Post-hoc statis

Mode

GA²M

Neuron Network Factorization

Hybrid models with neuron networks

Summai

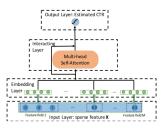


Figure: AutoInt

Each *multi-head self-attentive* layer captures interaction by learning interaction weights.

The model is a black-box with all the interactions entangled.

Summary

Models for capturing feature interactions

Recent Works on Feature Interaction

lingzhi Si

Outlin

Definitions of feature interaction Post-hoc statistic

Mode

Tree-based ensembels $GA^{2}M$ Neuron Networks

Factorization Machines Hybrid models with neuron networks

Summary

Models for capturing feature interactions:

- post-hoc
 - Tree Ensembles: non-additiveness, expressiveness
 - Neuron Networks: hessian, weights
- ad-hoc
 - Regularized NN: additive terms
 - GA2M: simple, fast
 - FM: sparse data, generalizaton, fast
 - Non-Deep and Deep hybrids: low-order and high-order