Learning Global Pairwise Interactions with Bayesian Neural Networks

Tianyu Cui   Pekka Marttinen   Samuel Kaski

Finnish Center for Artificial Intelligence; Department of Computer Science, Aalto University

Background

Estimating interactions between features, and the uncertainties of the interactions, is a challenge common to many data science tasks. For a simplest example could be:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e.$$

Existing methods include two following approaches:
1) Conducting tests for each combination, such as ANOVA [2];
2) 'White-box' machine learning models, such as Lasso [1].
The first approach lacks statistical power due to multiple testing. The second approach has to restrict the functional form of interactions.

Question: How to estimate interaction effects with quantified uncertainties without any functional form ?

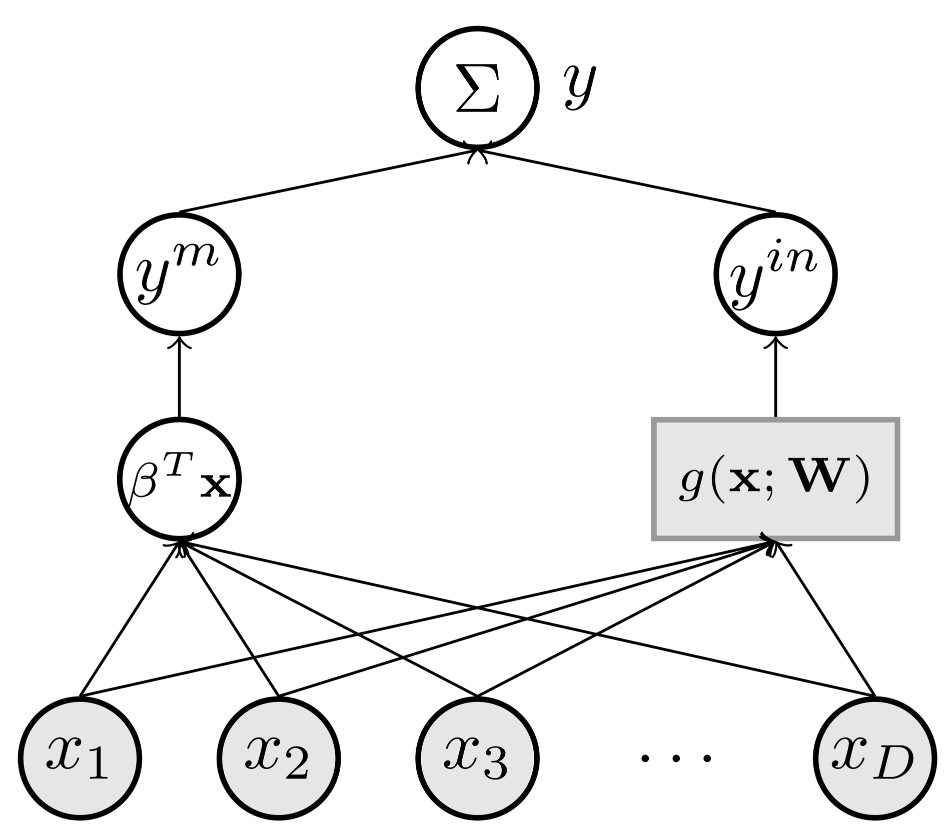
Proposed Approach

unlimited functional forms? $\Rightarrow$ Neural network
uncertainty estimation? $\Rightarrow$ Bayesian
interaction effects? $\Rightarrow$ Hessian

An intuitive approach with two steps (modeling and detecting):
1. Train a Bayesian Neural Network on the data of interests;
2. Find encoded interactions by estimating input Hessian of NN.

Modeling Interactions and their Uncertainty

Instead of using a single Bayesian MLP [3], we model the main effects by a linear regression separately from the interactions.



We use concrete dropout as $g^{\mathbf{W}}(\mathbf{x})$ by maximizing:

$$\int q_\theta(\mathbf{W}) \log p(\mathbf{Y}|\beta^T\mathbf{X} + g^{\mathbf{w}}(\mathbf{X}))\mathrm{d}\mathbf{W} - \mathrm{KL}(q_\theta(\mathbf{W})||p(\mathbf{W})),$$

where $q_\theta(\mathbf{W}) = q_{\mathbf{p},\mathbf{M}}(\mathbf{W}) = \prod_{l=1}^{L}\prod_{k=1}^{K_l} \mathbf{m}_{l,k}\mathrm{Bernoulli}(1-p_{l,k})$, $p_{l,k}$ is the dropout probability for node k in layer l, and $\mathbf{m}_{l,k}$ is a vector of outgoing weights from node k in layer l. We learn dropout probability for *each node* instead of *each layer* to select important features (with low dropout probabilities) as an *ARD* prior.

By using such trick, we can significantly reduce the size of BNN, which improves the training.

Detecting Interactions

Input Hessian (Hessian of $g^{\mathbf{W}}(\mathbf{x})$ w.r.t. the input) is only a **local** analogy to β_{12} for non-multiplicative interaction. **Global** interaction effects can be estimated by averaging local effects.

Existing approaches:

$$\mathrm{EAH}_g^{i,j}(\mathbf{W}) = \mathbb{E}_{p(\mathbf{x})}\left[\left|\frac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j}\right|\right], \mathrm{AEH}_g^{i,j}(\mathbf{W}) = \left|\mathbb{E}_{p(\mathbf{x})}\left[\frac{\partial^2 g^{\mathbf{W}}(\mathbf{x})}{\partial x_i \partial x_j}\right]\right|.$$

where $p(\mathbf{x})$ is the empirical distribution of $\mathbf{x}$.
EAH aggregates both signal and noise $\Rightarrow$ High FPR, but low FNR;
AEH averages both signal and noise $\Rightarrow$ High FNR, but low FPR.

Our approach: Group Expected Hessian (GEH):
We cluster dom($\mathbf{x}$) into M *subregions*, and calculate AEH for each subregion, then compute their weighted average. By tuning M, we trade-off between EAH and AEH. For M groups, M-$\mathrm{GEH}_g^{i,j}$ is

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

When $M \to 1$, M-GEH $\to$ AEH; $M \to N$, M-GEH $\to$ EAH.

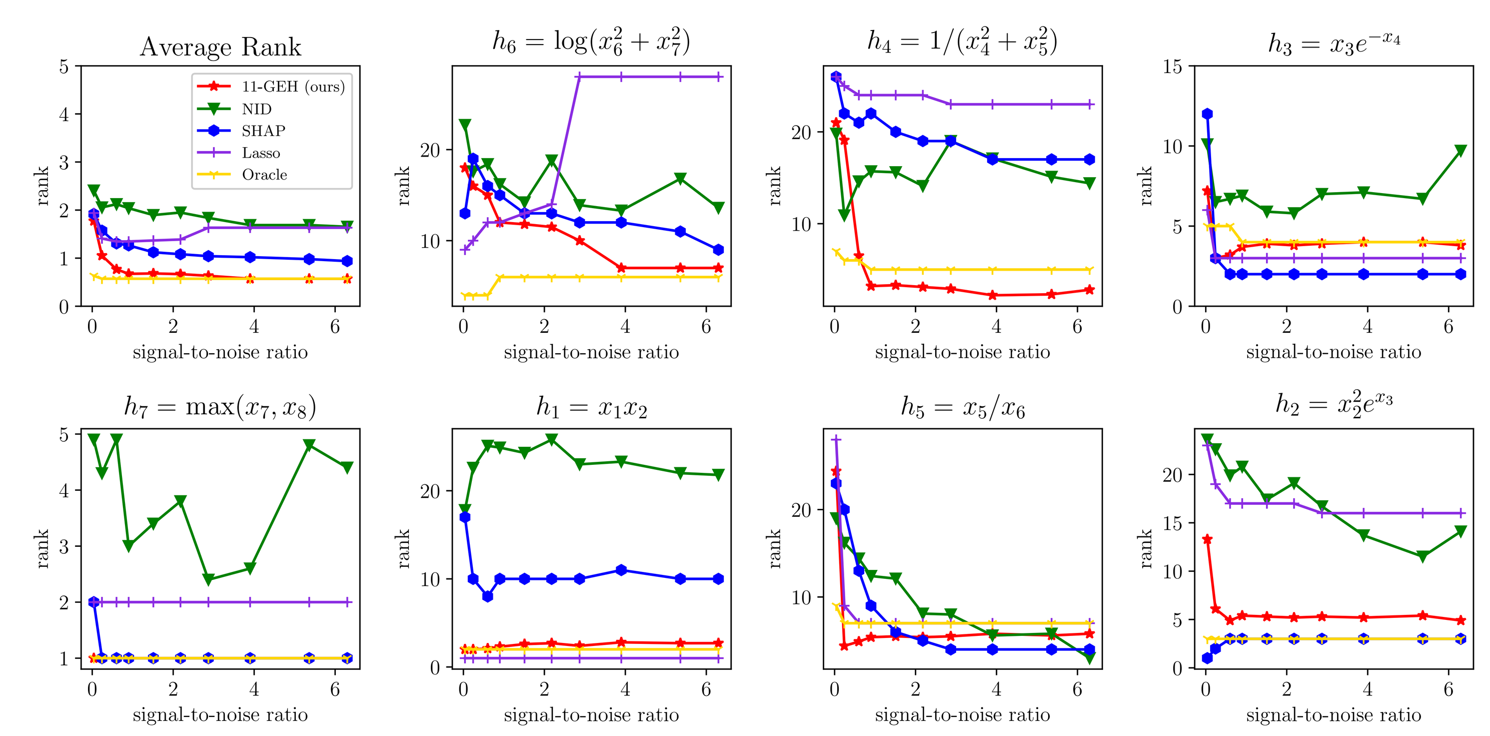
Optimal M: the smallest that can capture rich enough interactions.

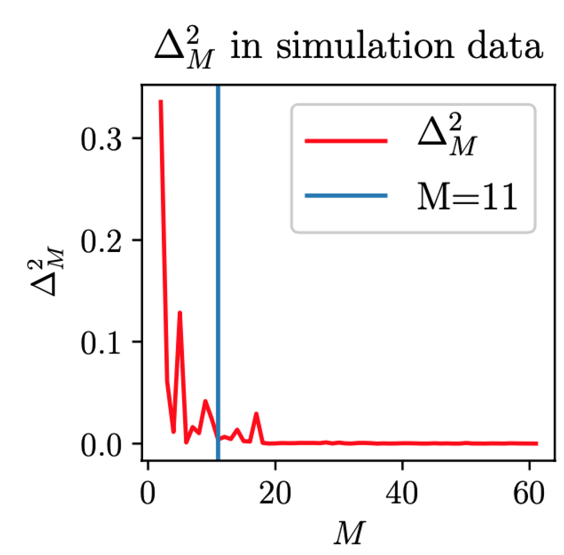
$$\Delta_M^2 = \sum_{i=1}^{L}(w_M(i) - w_{M-1}(i))^2(\pi_M(i) - \pi_{M-1}(i))^2.$$

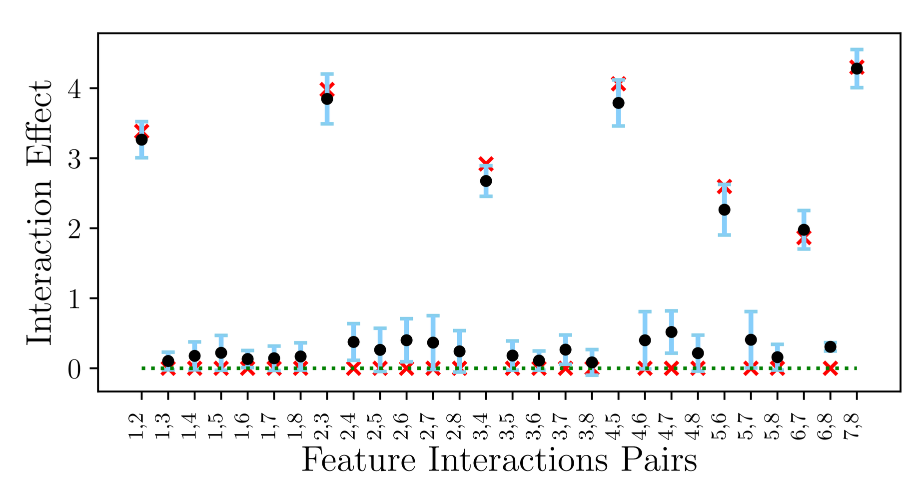
We compare two interaction effect vectors corresponding to consecutive numbers of clusters. We plotted values of Δ_M^2 as a function of M, and choose M when Δ_M^2 approximately converges to 0.

Experiments on Simulated Data

We use simulator: $y_i = \sum_{j=1}^{8}\beta_j^m x_j + \sum_{k=1}^{7}\beta_k^i h_k(x_k, x_{k+1}) + \epsilon$.







Aalto University
School of Science