Additive tree latent variable models with applications to insurance loss prediction

Guangyuan Gao[1]

Renmin University of China

Insurance Data Science Conference at Bayes Business School, London
2025.06

[1] Joint work with Zhihao Wang from Xinjiang University of Finance and Economics and Yanlin Shi from Macquarie University

Latent variable models in insurance loss prediction

We focus on a specific class of regression models with latent variables. Examples include mixture models (Jacobs et al. 1991) for claims severity (Fung et al. 2019), zero-inflated Poisson (ZIP) model (Lambert 1992) for claims frequency (Yip & Yau 2005) and Tweedie's compound Poisson model (Smyth 1989) for pure premium (Gao 2024).

Since multiple regression functions and latent variable are involved in parametric latent variable regression models, feature engineering, variable selection (Khalili & Chen 2007), and model selection (Kasahara & Shimotsu 2015) become particularly challenging.

A nonparametric method: Additive tree model

We propose additive tree latent variable models and design an Iteratively Re-weighted Gradient Boosting (IRGB) algorithm to efficiently calibrate the trees. The IRGB algorithm combines the EM algorithm with gradient boosting (Friedman 2001).

At each IRGB iteration, only one tree is fitted in a stagewise manner, with the current fitting taking into account the previously fitted trees. The model fitting, feature engineering, and variable selection are performed simultaneously due to the characteristics of recursive binary splitting trees (Breiman et al. 1983).

Additive tree latent variable models

A latent variable model $p(y, z|\boldsymbol{x}; \theta) = p(y, z; \theta(\boldsymbol{x}))$, where z is the latent variable and θ is the parameter of the joint density function depending on the covariate vector $\boldsymbol{x}$.

The joint density has a factorization

$$p(y, z; \theta(\boldsymbol{x})) = p(y|z; \theta_1(\boldsymbol{x}))p(z; \theta_2(\boldsymbol{x})). \tag{1}$$

This decomposition of θ_1 and θ_2 motivates the EM algorithm and also inspires our proposed IRGB algorithm.

Assume additive regression functions as follows:

$$\begin{aligned} g_1(\theta_1(\boldsymbol{x})) &= F(\boldsymbol{x}) = F^{[0]}(\boldsymbol{x}) + v_1 \sum_{m=1}^{M} f^{[m]}(\boldsymbol{x}), \\ g_2(\theta_2(\boldsymbol{x})) &= G(\boldsymbol{x}) = G^{[0]}(\boldsymbol{x}) + v_2 \sum_{m=1}^{M} g^{[m]}(\boldsymbol{x}) \end{aligned} \tag{2}$$

Iteratively re-weighted gradient boosting algorithm

Algorithm 1 The IRGB algorithm.

Data augmentation. Get an augmented complete data with the possible values of latent variable $(y_i, z_s^*, \boldsymbol{x}_i)_{i=1:n, s=1:S}$.

Initialization. Set $\hat{F}^{[0]}(\boldsymbol{x}_i), \hat{G}^{[0]}(\boldsymbol{x}_i), i = 1, \ldots, n$.

for $m = 1$ to M **do**

    Weight update. Calculate the conditional PMF of the latent variable $\hat{Q}_i^{[m-1]}(z)$. Set the weight of sample $(y_i, z_s^*, \boldsymbol{x}_i)$ as $\hat{w}_{i,s}^{[m-1]} = \hat{Q}_i^{[m-1]}(z_s^*)$.

    Gradient boosting. A gradient descent step is used to find a tree $\hat{f}^{[m]}(\boldsymbol{x}) = \sum_{l=1}^{L} \hat{\alpha}_l^{[m]} \mathbb{1}_{\hat{R}_l^{[m]}}(\boldsymbol{x})$. Similar for the tree $\hat{g}^{[m]}$. Set $\hat{F}^{[m]} = \hat{F}^{[m-1]} + v_1 \hat{f}^{[m]}$ and $\hat{G}^{[m]} = \hat{G}^{[m-1]} + v_2 \hat{g}^{[m]}$.

    The MLE of covariate-free parameter ϕ.

end for

return $\hat{F} = \hat{F}^{[M]}, \hat{G} = \hat{G}^{[M]}$ and $\hat{\phi} = \hat{\phi}^{[M]}$.

We prove the monotonically non-decreasing likelihood in the IRGB algorithm.

Application 1: zero-inflated Poisson regression

$$f_{\mathrm{ZIP}}(N|\boldsymbol{x};\lambda,\pi) = \begin{cases} \pi(\boldsymbol{x}) + (1-\pi(\boldsymbol{x}))e^{-\lambda(\boldsymbol{x})} & \text{for} \quad N = 0; \\ (1-\pi(\boldsymbol{x}))\,\frac{e^{-\lambda(\boldsymbol{x})}\lambda(\boldsymbol{x})^N}{N!} & \text{for} \quad N \in \mathbb{N}_+, \end{cases}$$

Table 1: The test loss for the three calibration approaches: first approach (Model BST), second approach (Models BST$-\lambda$ and BST$-\lambda-\pi$) and third approach (Models BST$-\pi$ and BST$-\pi-\lambda$).

models	BST	BST$-\lambda$	BST$-\lambda-\pi$	BST$-\pi$	BST$-\pi-\lambda$
test loss	0.6766	0.7791	0.7035	**0.6736**	0.6733

Table 2: Comparison of different ZIP models and a Poisson GBDT in terms of the test loss, the single Poisson loss and the predicted proportion of zeros.

models	test loss	single Poisson loss	zeros %
BST$-\pi$	**0.6736**	**0.8057**	0.6217
GLM$-\lambda$ (Yip & Yau 2005)	0.7965	0.8995	0.5949
GLM$-\lambda^*$	0.7831	0.8328	0.5821
GLM$-\pi^*$	0.6765	0.8078	0.6157
NULL	0.8256	0.9779	0.6067
GBDT	-	0.8229	0.5243

Application 2: Tweedie's compound Poisson regression

$$f_{Y,N}(y,k) = f_{Y|N=k}(y) f_N(k) = \begin{cases} f_{Poi}(0; e_i\lambda_i) & \text{for } k = 0; \\ f_{Poi}(k; e_i\lambda_i) f_{gam}(y; k\alpha, k\tau_i/e_i) & \text{for } k > 0, \end{cases}$$

Table 3: The test loss for the three calibration approaches.

models	BST	BST−λ	BST−$\lambda-\tau$	BST−τ	BST−$\tau-\lambda$
test loss	**0.6496**	0.6509	**0.6501**	0.6622	**0.6502**

Table 4: Comparison of the latent variable models (BSTs and GLM) and the Poisson-gamma models for both claim counts and amounts (BST0, GLM0 and NULL) in terms of test loss on five mutually exclusive subsets of data.

models	test loss 1	test loss 2	test loss 3	test loss 4	test loss 5
BST	0.6515	0.6513	0.6526	0.6542	**0.6496**
BST−$\lambda-\tau$	**0.6506**	**0.6509**	**0.6514**	**0.6537**	0.6501
BST−$\tau-\lambda$	**0.6506**	0.6516	0.6515	0.6538	0.6502
GLM	0.6538	0.6526	0.6556	0.6577	0.6513
BST0	0.6415	0.6416	0.6437	0.6419	0.6471
GLM0	0.6538	0.6526	0.6558	0.6577	0.6513
NULL	0.6607	0.6597	0.6613	0.6649	0.6607

Thank you!

Q & A

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. (1983), *Classification and Regression Trees*, CRC Press.

Friedman, J. H. (2001), 'Greedy function approximation: a gradient boosting machine', *The Annals of statistics* **29**, 1189–1232.

Fung, T. C., Badescu, A. L. & Lin, X. S. (2019), 'A class of mixture of experts models for general insurance: Theoretical developments', *Insurance: Mathematics and Economics* **89**, 111–127.

Gao, G. (2024), 'Fitting Tweedie's compound poisson model to pure premium with the EM algorithm', *Insurance: Mathematics and Economics* **114**, 29–42.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J. & Hinton, G. E. (1991), 'Adaptive mixtures of local experts', *Neural Computation* **3**(1), 79–87.

Kasahara, H. & Shimotsu, K. (2015), 'Testing the number of components in normal mixture regression models', *Journal of the American Statistical Association* **110**(512), 1632–1645.

Khalili, A. & Chen, J. (2007), 'Variable selection in finite mixture of regression models', *Journal of the American Statistical Association* **102**(479), 1025–1038.

Lambert, D. (1992), 'Zero-inflated poisson regression, with an application to defects in manufacturing', *Technometrics* **34**(1), 1–14.

Smyth, G. K. (1989), 'Generalized linear models with varying dispersion', *Journal of the Royal Statistical Society: Series B (Methodological)* **51**(1), 47–60.

Yip, K. C. & Yau, K. K. (2005), 'On modeling claim frequency data in general insurance with extra zeros', *Insurance: Mathematics and Economics* **36**(2), 153–163.