

Evidence Regularization for Multimodal Deep Evidential Regression

Zhimin Shao¹

SZM21@MAILS.TSINGHUA.EDU.CN

¹ *Department of Electronic Engineering, Tsinghua University, Beijing, China*

Weibei Dou¹

DOUWB@TSINGHUA.EDU.CN

Abstract

Uncertainty estimation is crucial in cost-sensitive areas, especially in the medical field, where multimodal information is common and effective. Existing studies have found the zero-confidence issue in unimodal settings, while the analysis in multimodal scenarios is lacking. In this work, we introduce the confidence paradox, where unimodal uncertainty is high but decreases after fusion, and present evidence of regularization to tackle this issue. Initial results on the cubic and CT slice datasets show reduced root mean squared errors and improved detection of out-of-distribution samples, improving predictive reliability and training stability.

Keywords: Uncertainty estimation, Multimodal, Deep learning, Regression

1. Introduction

In medical image analysis, regression tasks predict continuous values y in \mathbb{R}^d using input images or features X , crucial for applications like forensic age estimation (Halabi et al., 2019), disease score regression (Stonnington et al., 2010), and image segmentation by identifying boundaries (Yin et al., 2020). With the increasing prevalence of multimodal data in the medical field and the necessity for reliable predictions, assessing the predictive uncertainty in deep learning models for each data modality and the combined uncertainty following multimodal data integration is of paramount importance (Ma et al., 2021).

Predictive uncertainty is divided into aleatoric (data variability) and epistemic (model knowledge gaps). While Bayesian methods like dropout (Gal and Ghahramani, 2016) and deep ensemble (Lakshminarayanan et al., 2017), infer the posterior distribution of parameters or observed targets to estimate uncertainty, challenges like computational intensity and disentanglement between these uncertainty types. Deep evidential regression (DER) (Amini et al., 2020) and its multimodal version (Ma et al., 2021) have revolutionized uncertainty estimation by probabilistically estimating target distribution parameters and enhancing the separation of uncertainties but faced issues like zero-confidence in training. Specific activation functions were employed to ensure non-negative evidential distribution parameters, which limits the models' learning efficiency from training data (Pandey and Yu, 2023). To bypass the need for selecting specific activation functions, dual activation function fusion was introduced (Shao et al., 2024), yet the zero-confidence issue persists. The issue, identified in classification (Pandey and Yu, 2023) and regression (Ye et al., 2024), leads to the cessation of model updates in zero-confidence areas during training. Although evidence regularization has successfully mitigated the zero-confidence issue in unimodal settings (Pandey and Yu, 2023; Ye et al., 2024), multimodal regression's complexities necessitate extending this approach to improve training stability and uncertainty management.

2. Confidence Paradox in Multimodal Scenario

Deep evidential regression (Amini et al., 2020) assumes that a target value y is i.i.d from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. The unknown mean μ and variance σ are presumed to follow a Normal Inverse-Gamma (NIG) distribution $(\mu, \sigma^2) \sim NIG(\gamma, \nu, \alpha, \beta)$, where $\mu \sim \mathcal{N}(\gamma, \frac{\sigma^2}{\nu})$ and $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$. $\Gamma(\cdot)$ is the gamma function. The parameters of NIG $\mathbf{m} = (\gamma, \nu, \alpha, \beta)$ are determined by the neural network’s output $\mathbf{o} = (o_\gamma, o_\nu, o_\alpha, o_\beta) = f(X|\boldsymbol{\theta})$ with $\boldsymbol{\theta}$ as the network’s trainable parameters. To enforce the constraints on \mathbf{m} , a SoftPlus activation function is used to generate (ν, α, β) , with an additional increment of 1 for α , and a linear activation function for $\gamma \in \mathbb{R}$. For multimodal scenarios (Ma et al., 2021), $y \sim \sum_{m=1}^M \frac{1}{M} NIG(\gamma_m, \nu_m, \alpha_m, \beta_m)$. The parameters $(\gamma_m, \nu_m, \alpha_m, \beta_m)$ are learned from training data in m -th modality by $f_m(\cdot)$. The summation of NIG distributions is

$$NIG(\gamma, \nu, \alpha, \beta) \triangleq NIG(\gamma_1, \nu_1, \alpha_1, \beta_1) \oplus NIG(\gamma_2, \nu_2, \alpha_2, \beta_2) \oplus \dots \oplus NIG(\gamma_M, \nu_M, \alpha_M, \beta_M) \quad (1)$$

where the summation \oplus for any two NIG distributions is defined as

$$\begin{aligned} \gamma &= (\nu_1 + \nu_2)^{-1} (\nu_1 \gamma_1 + \nu_2 \gamma_2), & \alpha &= \alpha_1 + \alpha_2 + \frac{1}{2} \\ \nu &= \nu_1 + \nu_2, & \beta &= \beta_1 + \beta_2 + \frac{1}{2} \nu_1 (\gamma_1 - \gamma)^2 + \frac{1}{2} \nu_2 (\gamma_2 - \gamma)^2 \end{aligned} \quad (2)$$

Based on NIG distribution, we use $\mathbb{E}[\mu] = \gamma$ as the prediction, $\mathbb{E}[\sigma^2] = \frac{\beta}{(\alpha-1)}$ as the aleatoric uncertainty, and $\text{Var}[\mu] = \frac{\beta}{v(\alpha-1)}$ as the epistemic uncertainty.

In the zero-confidence area, characterized by the parameter $\alpha = 1$, both the aleatoric and epistemic uncertainties diverge to infinity due to the term $\alpha - 1 = 0$ in the denominator. Consequently, $\text{SoftPlus}(o_\alpha) = 0$ and $\frac{\partial \alpha}{\partial o_\alpha} = 0$. Therefore, the gradient of loss $\frac{\partial \mathcal{L}}{\partial o_\alpha} = \frac{\partial \mathcal{L}}{\partial \alpha} \frac{\partial \alpha}{\partial o_\alpha}$ is also zero. The model would stick in the zero-confidence area once it falls into it, leading to infinite values for both aleatoric and epistemic uncertainties in unimodality. However, this scenario presents a paradox where the aggregation of two highly uncertain NIG distributions results in a significant reduction of the overall uncertainty, which contradicts intuition, shown in Table 1.

Commonly, we would anticipate that incorporating more comprehensive multimodal data would reduce uncertainty. However, deriving a certain decision based on completely uncertain evidence from each modality is illogical and highlights a critical flaw in the uncertainty modeling approach. Traditional regularization on predictive-level fused α cannot propagate to every sub-modality.

Table 1: An Numeric Example of Confidence Paradox

	NIG($\gamma, \nu, \alpha, \beta$)	AU	EU
Mod 1	NIG(0, 1, 1, 1)	Inf.	Inf.
Mod 2	NIG(1, 1, 1, 1)	Inf.	Inf.
Sum	NIG(0.5, 2, 2.5, 2.25)	1.5	0.75

3. Evidence Regularization for Multimodal DER (ER-MDER)

Similar to (Ma et al., 2021), the overall loss function for multimodal learning is the sum of losses of multiple modalities (m denotes the m -th modality), the pseudo modality (obtained by features concatenation or concatenation after representation learning, denoted by P), and

the predictive-level fused distribution (denoted by F), $\mathcal{L} = \sum_{m=1}^M \mathcal{L}_m(\mathbf{w}) + \mathcal{L}_P(\mathbf{w}) + \mathcal{L}_F(\mathbf{w})$. To avoid the confidence paradox, we put constraints directly on o_α for each modality.

$$\begin{aligned} \mathcal{L}(\mathbf{w}) &= \mathcal{L}^{\text{NLL}}(\mathbf{w}) + \lambda_1 \mathcal{L}^{\text{R}}(\mathbf{w}) + \lambda_2 \mathcal{L}^{\text{O}}(\mathbf{w}) \\ &= \frac{1}{2} \log\left(\frac{\pi}{v}\right) - \alpha \log(\Omega) + \left(\alpha + \frac{1}{2}\right) \log\left((y_i - \gamma)^2 v + \Omega\right) + \log\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})}\right) \\ &\quad + \lambda_1 (|y - \gamma| \cdot (2\nu + \alpha)) - \lambda_2 (|y - \gamma| \cdot o_\alpha) \end{aligned} \quad (3)$$

where $\Omega = 2\beta(1 + \nu)$. While there is no o_α for the fused distribution, $\mathcal{L}_F(\mathbf{w})$ only contains two items $\mathcal{L}^{\text{NLL}}(\mathbf{w}) + \lambda_1 \mathcal{L}^{\text{R}}(\mathbf{w})$.

4. Experiments

Cubic Regression We evaluate our proposed ER-MDER in comparison with the multimodal DER (MDER) on the cubic regression dataset, particularly within zero-confidence areas. Following (Amini et al., 2020), models were trained on $y = x^3 + \epsilon$ with $\epsilon \sim \mathcal{N}(0, 3)$ over the interval $x \in [-4, 4]$ and tested in the range $x \in [-6, -4] \cup (4, 6]$. The m -th modality input x_m equals to $x + \epsilon_x$, $\epsilon_x \sim \mathcal{N}(0, 0.01)$. After evidential regularization, our model can update faster than the original one within the zero-confidence area and generate more reliable predictions, shown in Figure 1.

CT Slices Location We evaluated our methods on CT slices dataset (Graf et al., 2011). Each image includes two modalities: one with 240 attributes for bone structure and another with 144 attributes for air inclusions. The objective is to predict the image’s axial axis position, ranging from 0 to 180. Similar settings were used as (Ma et al., 2021). We introduce noise ($\epsilon = 0.1$) to half of the test samples to create out-of-distribution (OOD) samples and distinguish them using uncertainty. According to Table 2, our method achieves lower Root Mean Squared Error (RMSE) and enhanced performance in OOD detection, as indicated by a higher AUROC score.

Future work will focus on more experiments on public medical image datasets, and application to in-house data.

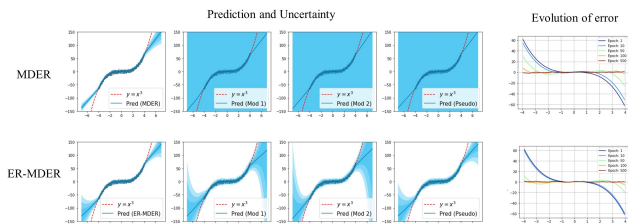


Figure 1: Prediction, uncertainty estimation and error evolution in MDER and ER-MDER. The blue shade represents prediction uncertainty.

Table 2: Results of prediction and OOD detection in CT Slices.

	RMSE(↓)	AUROC(↑)
MDER	0.79	0.615
ER-MDER (Ours)	0.67	0.956

References

- Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in neural information processing systems*, 33:14927–14937, 2020.
- Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.
- F. Graf, H.-P. Kriegel, S. Schubert, M. Poelsterl, and A. Cavallaro. Relative location of CT slices on axial axis. UCI Machine Learning Repository, 2011. DOI: <https://doi.org/10.24432/C5CP6G>.
- Safwan S Halabi, Luciano M Prevedello, Jayashree Kalpathy-Cramer, Artem B Mamonov, Alexander Bilbily, Mark Cicero, Ian Pan, Lucas Araújo Pereira, Rafael Teixeira Sousa, Nitamar Abdala, et al. The rsna pediatric bone age machine learning challenge. *Radiology*, 290(2):498–503, 2019.
- Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.
- Huan Ma, Zongbo Han, Changqing Zhang, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. Trustworthy multimodal regression with mixture of normal-inverse gamma distributions. *Advances in Neural Information Processing Systems*, 34:6881–6893, 2021.
- Deep Shankar Pandey and Qi Yu. Learn to accumulate evidence from all training samples: theory and practice. In *International Conference on Machine Learning*, pages 26963–26989. PMLR, 2023.
- Zhimin Shao, Weibei Dou, and Yu Pan. Dual-level deep evidential fusion: Integrating multimodal information for enhanced reliable decision-making in deep learning. *Information Fusion*, 103:102113, 2024.
- Cynthia M Stonnington, Carlton Chu, Stefan Klöppel, Clifford R Jack Jr, John Ashburner, Richard SJ Frackowiak, Alzheimer Disease Neuroimaging Initiative, et al. Predicting clinical scores from magnetic resonance scans in alzheimer’s disease. *Neuroimage*, 51(4):1405–1413, 2010.
- Kai Ye, Tiejun Chen, Hua Wei, and Liang Zhan. Uncertainty Regularized Evidential Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15):16460–16468, March 2024. doi: 10.1609/aaai.v38i15.29583. URL <https://ojs.aaai.org/index.php/AAAI/article/view/29583>.
- Shi Yin, Qinmu Peng, Hongming Li, Zhengqiang Zhang, Xinge You, Katherine Fischer, Susan L Furth, Gregory E Tasian, and Yong Fan. Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. *Medical image analysis*, 60:101602, 2020.