Midwest Machine Learning Symposium (MMLS) | June 2026

Missing Value Uncertainty: Could Collecting Missing Values Change the Prediction?

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

SAAB TECHNOLOGIES SAAB

DEVCOM ARMY RESEARCH LABORATORY

Purdue University
Department of Electrical and Computer Engineering
David Burnett, Surojit Ganguli, David Inouye

US Army Research Lab
Lance Kaplan

Purdue University,, Saab Inc.
Devesh Upadhyay

OVERVIEW

We address the challenge of estimating whether it is worth collecting the cost of missing inputs based on information at inference time. We developed missing value uncertainty (MVU) to estimate the hard confidence over a distribution of missing values, allowing an operator to estimate whether collecting missing values might change the prediction. We then developed the MVCE metric to evaluate estimates of MVU and demonstrate our work on real world datasets.

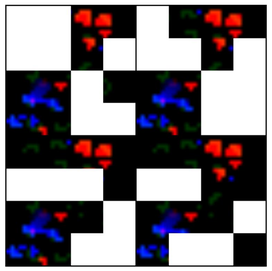

Figure 1: The StarCraft dataset is used to simulate a battlefield scenario, where lost sensors lead to uncertain about unit positions, but environment is easy to estimate.

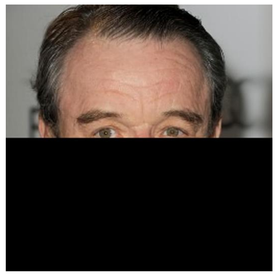

Figure 2: When the bottom half of the CelebA face is missing, it can be difficult to estimate attributes such as smiling, while attributes like skin color may be easier.

Problem Setup

In high-stakes domains such as healthcare and security, decisions are often made with incomplete information. This raises a critical and practical question for a human operator: **Is it worth the cost and effort to collect missing input values for a specific instance at inference time?** For example, a security analysis observing a potential threat with data from a partially failed sensor network (figure 1) must determine whether to act immediately or deploy resources to gather more information.

We start by asking whether collecting missing values is likely to change the prediction. If additional information is unlikely to alter course of action, an operator can proceed without the missing values. However, if the missing values could significantly change the decision, the most prudent action is to collect them first. This paper centers on developing a framework that, given a partially observed input at inference time, can help an operator decide whether to collecting missing input feature values.

Prior work has addressed aspects of this problem but fails to quantifying when more information may change the prediction. Missing value literature primarily focuses on robustness to make the best possible prediction given the observed values (Little & Rubin, 2019; Azur et al., 2011). While useful, this does not quantify uncertainty introduced by the missing values, leaving the operator unsure of how the prediction might change if they were revealed.

Missing Value Uncertainty (MVU)

MVU is the distribution of predictions induced by incomplete inputs at inference time, represented as $P(\Phi|X_o = x_o)$, where Φ is our prediction probabilities and X_o is our partial observation. Given this distribution, traditional methods that are simply robust to missing values can be thought of taking the mean of this distribution, giving us *soft confidence* of the prediction. This notably ignores the variance of the prediction which makes it difficult to distinguish between irreducible uncertainty and uncertainty reducible by collecting missing values. Inspired by ensemble model hard voting, we recommend using hard confidence, which averages over the argmax of Φ. This notably is influence by the variance of the MVU distribution and makes the estimate more robust to miscalibrated models. If the hard confidence is high, gathering more information is unlikely to change the prediction. If the hard confidence is low, more information is likely to change the prediction.

To use hard confidence, we must estimate the MVU distribution. Traditional approaches for handling missing values would give us $\mathbb{E}[\Phi|x_o]$, requiring a heuristic to estimate uncertainty. If we have an imputation model that can produce many samples of missing values x_m from $p(X_m|x_o)$, we could leverage a Monte Carlo approach to estimate MVU, but this is too expensive in practice. Thus, we propose a novel direct estimator of MVU distribution. The training loss function is developed from minimizing the KL divergence between the true and estimated distributions:

$$\min_{\theta,\psi} \mathbb{E}\left[\ell_{CE}(\hat{\pi}_\theta(X), Y)\right],$$
$$s.t. \psi \in \arg\min_{\bar{\psi}} \mathbb{E}\left[-\log p_{\bar{\psi}}\left(\hat{\pi}_\theta(X_o, X_m)\right)|X_o\right]$$

Since simplifying KL divergence to negative log-likelihood requires $\hat{\pi}_\theta$ to be constant with respect to θ, this objective is run in two stages: first we train $\hat{\pi}_\theta$ using complete data, then we train the uncertainty model $p_\psi(\Phi|X_o)$ assuming that $\hat{\pi}_\theta$

Evaluating via Missing Value Calibration Error (MVCE)

The answer to whether it is worth collecting missing values depends heavily on the specific real-world problem context such as the cost of revealing missing values and the costs of being wrong. Thus, instead of focusing on a particular scenario, we aim for a problem-agnostic evaluation of MVU models by asking an uncertainty calibration question w.r.t. hard confidence: "When my model predicts a hard confidence of $c\%$ on a partially observed input x_o , is the prediction on the fully observed x the same $c\%$ of the time?" This is like the standard uncertainty calibration question corresponding to soft confidence which is: "When my model predicts a soft confidence of $c\%$, does it match the true class $c\%$ of the time?" Specifically, the hard confidence calibration gives the probability that the prediction will not change, while the soft confidence calibration gives the probability the class is correct. While soft confidence values can be naturally evaluated using the well-known Expected Calibration Error (ECE) (Naeini et al., 2015; Guo et al., 2017), evaluating hard confidence values for MVU requires some adaptation.

Because hard confidence aims to quantify the probability that the prediction will stay the same, we will simulate the phenomena of revealing all missing values using complete training data. Intuitively, we simulate a partially observed input x_o by dropping features from the full input x and then compare with the prediction on the complete x. Like ECE, we first partition the dataset $\mathcal{D}$ into a set of bins $\mathcal{B}$ based on hard confidence values. MVCE is then defined as:

$$MVCE = \sum_{B\in\mathcal{B}} \frac{|B|}{|\mathcal{D}|} |cons(B) - \bar{c}^o_{hard}(B)|$$

where the consistency $cons(B) = 1/(|B|)\sum_{i\in B} \mathrm{I}(\hat{y}_i^o = \hat{y}_i)$ compares the prediction under partial inputs to the prediction under clean data, and $\bar{c}^o_{hard}(B)$ is the average hard confidence on partial inputs in the bin B.

Since consistency only evaluates whether the prediction changed compared to clean data, it does not evaluate whether the model is accurate. For best results, MVCE should be employed alongside accuracy and ECE.

Evaluating MVU

CelebA Method \ Missing	Time (sec.)	Feature: Smiling Acc C	Acc MV	MVCE	MV Calib.	
Diffusion - 1 Sample, 0 Variance	182	91.9%	84.1%	0.1357	-	-
Diffusion - 1 Sample, 0.5 Max Variance	182	91.9%	84.1%	0.1028	0.0990	-4%
Diffusion - 1 Sample, Scale Probability	182	91.9%	84.1%	0.1226	0.1028	-16%
Diffusion - 3 Samples - Empirical Variance	547	91.9%	85.6%	0.0555	0.0530	-5%
Diffusion - 30 Samples - Empirical Variance	5465	91.9%	**86.4%**	**0.0163**	**0.0144**	-12%
Mean Imputation, 0 Variance	0.547	91.9%	70.5%	0.2633	-	-
Mean Imputation, 0.5 Max Variance	0.538	91.9%	70.5%	0.2335	0.2281	-2%
Mean Imputation, Scale Probability	0.619	91.9%	70.5%	0.2548	0.2335	-8%
Direct Missing Value (DMV)	0.901	92.8%	**77.1%**	**0.0711**	**0.0695**	-2%

Method \ Missing	StarCraftCIFAR10 Acc C	Acc MV	MVCE	MV Calib.	
Mean Imputation, 0 Variance	79.9%	70.5%	0.2231	-	-
Mean Imputation, 0.5 Max Variance	79.9%	70.6%	0.0493	0.0214	-57%
Mean Imputation, Scale Probability	79.9%	70.6%	0.0989	0.0233	-76%
Missing Robust, 0 Variance	71.8%	71.3%	0.1278	-	-
Missing Robust, 0.5 Max Variance	71.8%	71.2%	0.0809	0.0582	-28%
Missing Robust, Scale Probability	71.8%	71.2%	0.0937	0.0571	-39%
Direct Missing Value	79.8%	**78.5%**	**0.0201**	**0.0113**	-44%

Discussion

On CelebA data, DMV consistently produced MVCE comparable to Monte Carlo using a diffusion model for imputation, but in a fraction of the time per sample. It also has a notably lower drop in accuracy than the mean imputation baseline, showing the speed up does not come at a significant performance cost. On StarCraftCIFAR10, we demonstrated DMV's performance generalizes to a multiclass dataset, comparing against a classifier simply trained to be robust to missingness. For both experiments, we consider both the original MVCE values, and values calibrated using a grid search over validation MVCE.

Future Work

For our work, we only considered simulated missingness in data, allowing both model training and our metric to leverage complete data. While often reasonable, some types of data such as medical rarely have complete data, so it would be useful to be able to train and evaluate on incomplete data. Additionally, we only considered evaluating whether missing values might change the prediction; to fully answer the question of if its worth collecting missing values, we should also develop a method to estimate which feature has the best value relative to the cost of collecting it.