

Application and Development of Bayesian Methods in the Digitalization of Cultural Heritage

Shichen Zhang

*School of Mathematical Sciences, Dalian University of Technology, Dalian, China
zscsimon@mail.dlut.edu.cn*

Abstract. Bayesian methods are prominent in data processing. Based on Bayes' theorem, they dynamically update knowledge and probabilistic judgments about events using prior information and observed data, enabling more precise analysis and prediction. In terms of application, Bayesian methods are widely used in numerous fields such as machine learning, medical diagnosis, and financial risk assessment. This paper focuses on their application and development in this field. Using literature analysis, it explores two research questions: the current application status and future development directions. The study reveals that Bayesian methods are mainly used in parameter optimization and risk modeling/decision - making. They offer advantages like efficient parameter tuning and flexible model selection. However, challenges such as hyperparameter sensitivity exist. Future research could integrate more cross - domain knowledge and develop a probabilistic knowledge base for heritage conservation.

Keywords: Bayesian methods, heritage conservation, parameter processing, numerical optimization

1. Introduction

Currently, the application of Bayesian methods in cultural heritage conservation remains relatively scarce. Despite their prominence in data processing across various fields, their adoption in heritage conservation has not been widespread. This paper aims to explore two key aspects. Firstly, it seeks to understand the current status of Bayesian methods' application in cultural heritage conservation. Secondly, it endeavors to identify the future development directions of these methods within this domain. A comprehensive literature analysis is employed. By thoroughly examining relevant literature, the study summarizes and synthesizes the application and development of Bayesian methods in cultural heritage conservation. The research provides valuable guidance and outlines development directions for the application of digital models in the cultural heritage field. It reveals how Bayesian methods can enhance the efficiency and accuracy of heritage conservation, ultimately contributing to the intelligent decision - making process in preventive conservation.

2. Current status of Bayesian applications in digital heritage conservation

2.1. Updating parameter information using Bayesian algorithms

Bayesian algorithms, grounded in Bayes' theorem, play a pivotal role in updating parameter information. This is accomplished by synthesizing the prior distribution with the currently observed data to construct the posterior distribution of the target function. The detailed operational procedure unfolds as follows:

(1) Constructing the Prior Distribution

At the onset of the optimization process, in the absence of comprehensive knowledge regarding the form of the target function—for instance, the intricate relationship between model performance metrics and hyperparameters—it becomes imperative to establish a prior distribution. This prior distribution encapsulates initial assumptions. Gaussian Processes (GP) are frequently employed as priors due to their ability to inherently model function uncertainty and correlations. A GP is characterized by a mean function and a kernel function. The mean function designates the expected value at every point, while the kernel function delineates the correlation among different points.

(2) Evaluating the Target Function

Within the hyperparameter space, an initial set of hyperparameter combinations is judiciously selected. These combinations serve as the basis for training and subsequently evaluating the model. The observed values of the target function corresponding to these hyperparameters are obtained, forming the foundation for subsequent parameter updates.

(3) Updating the Posterior Distribution

Bayes' theorem serves as the linchpin for amalgamating the prior distribution with the observations, thereby enabling the update of the posterior distribution of the target function. The posterior distribution effectively integrates the prior knowledge with the newly acquired observed data, offering a more precise approximation of the true target function. In the context of Gaussian Process Regression, the posterior mean and variance are computed based on its mean function, covariance function, and the observed data. This computation reflects the updated expected value and the associated uncertainty of the target function.

(4) Sampling Strategy

Guided by the posterior distribution, new hyperparameter combinations are strategically chosen for evaluation. Among the commonly adopted strategies is the Expected Improvement (EI) approach. EI strikes a balance between the predicted value and the uncertainty of hyperparameter combinations, thereby mitigating the risk of getting trapped in local optima. Specifically, EI quantifies the expected enhancement that a new hyperparameter combination can bring about compared to the current best. The hyperparameter combination with the highest EI value is then selected for further exploration.

(5) Iterative Update

The model is evaluated using the newly selected hyperparameter combination. The resultant outcome is incorporated into the existing observation data. Subsequently, the steps of posterior update and sampling are reiterated. As the number of iterations progresses, the posterior distribution converges more closely to the true target function. Concurrently, the sampling strategy becomes increasingly efficient in exploring the hyperparameter space, ultimately leading to the identification of the optimal hyperparameter combination.

In the realm of machine learning model parameter optimization, Bayesian algorithms, armed with tools such as Bayes' theorem and Gaussian Processes, adeptly update parameter information through a series of well-defined steps, including the construction of the posterior distribution and intelligent

sampling. In contrast to traditional methods like grid search or random search, Bayesian algorithms leverage historical information to steer subsequent searches, thereby enhancing both the efficiency and effectiveness of parameter optimization.

To illustrate, Bayesian methods can be harnessed to optimize a risk prediction model for the surrounding environment of cultural heritage sites. Hyperparameters, such as regularization parameters, the number of decision tree leaf nodes, the maximum depth of decision trees, the learning rate, and the boosting strategy, are meticulously tuned. By precisely defining the parameter space and implementing Bayesian optimization, the optimal parameter combination is derived to construct a LightGBM model. When applied to real - world historical district data, 5 - fold cross - validation yielded a remarkable model prediction accuracy of up to 98% for fire risk prediction within the historical district dataset [1].

2.2. Model selection for Bayesian applications in practice

Fundamentally, the Beta - Bernoulli Model [2] holds significance. This model is particularly applicable to protection - related problems that involve binary outcomes. For instance, it can be used to determine whether an artifact deteriorates under a specific environmental condition (such as when humidity exceeds a certain threshold, with the outcome being either 'yes' or 'no'), whether a particular conservation treatment effectively halts the deterioration process (resulting in either 'success' or 'failure'), or whether a structural component fails under a given load (again, a 'yes' or 'no' outcome).

In the context of Bayesian applications, the Beta distribution is employed as the conjugate prior for the success probability p . For example, based on expert experience, the prior parameters α and β are set. After observing Bernoulli - distributed data, which consists of a sequence of success - failure events, the Beta posterior distribution for p is derived. This approach offers several advantages: it is computationally straightforward, provides an analytical solution, and the results are easily interpretable, as the posterior distribution of p directly reveals the success probability along with its associated uncertainty. When the core problem can be clearly formulated as a binary event and the data are representative of independent repeated trials, this model becomes the preferred choice. An illustrative example is the assessment of the effectiveness of a waterproof coating in safeguarding artifacts from water - stain formation under specific conditions.

However, in practical scenarios, the assumptions and conditions may not be as ideal as those in theoretical models. Outcomes may not always be strictly binary, and the conditions may not be independent. In such situations, Bayesian Linear Regression emerges as a viable alternative. This method is used to explore the linear relationships between a continuous dependent variable and one or more independent variables within the realm of conservation problems.

For example, it can be applied to study the impact of ambient temperature (an independent variable) on the weathering rate of stone artifacts (the dependent variable), the effect of humidity fluctuations (independent variable) on the crack propagation rate in wooden artifacts (dependent variable), or the influence of pollutant concentration (independent variable) on the fading degree of mural pigments (dependent variable).

In the application of Bayesian Linear Regression, expert experience and knowledge are utilized to establish prior distributions for the regression coefficients β and the noise variance σ^2 . These prior distributions can be, for example, Normal, Gamma distributions, or weakly - informative priors. Subsequently, Bayes' theorem is employed to calculate the joint posterior distribution of the parameters β and σ^2 . Additionally, the predictive

distribution for forecasted values is obtained, taking into account both parameter uncertainty and observation noise.

This model structure is relatively uncomplicated and interpretable. It allows for the quantification of the significance of coefficients and the uncertainty of predictions. Moreover, it can be extended to Generalized Linear Models to handle non - normal responses. For example, Poisson regression can be used for analyzing the count of cracks.

When considering the application of this model in conservation, it is suitable when there is a well - defined hypothesis of a linear (or linearizable) relationship between variables and the data volume is of a moderate size. For instance, it can be used to construct a quantitative relationship model between temperature/humidity and the strength loss rate of a specific type of silk. Nevertheless, its performance deteriorates significantly when the relationship between variables is clearly non - linear.

3. Challenges in application, solutions, future development directions

3.1. Application challenges and optimization

The values of hyperparameters exert a profound influence on the prediction outcomes of models. To address this, point - estimation methods can be employed to re - estimate hyperparameters following multiple iterations, such as 20 iterations. Herein, several relatively uncomplicated methods are presented [2].

(1) Assign a prior $p(\theta)$ to the hyperparameters, then obtain the Maximum A Posteriori (MAP) estimate by maximizing the formula via "Bayes' theorem":

$$P(\theta|D_{1:t}) = \frac{P(D_{1:t}|\theta)P(\theta)}{P(D_{1:t})}, \text{yielding } \hat{\theta}_t^{MAP} \quad (1)$$

(2) Use Type II Maximum Likelihood (ML), also known as Empirical Bayes or Marginal Likelihood Maximization, to obtain. $\hat{\theta}_t^{ML}$

$$p(y|X, \theta) = \int p(y|f)p(f|X, \theta)df = N(0, \Sigma + \sigma^2 I) \quad (2)$$

In addition to hyperparameter issues, the high - dimensionality challenge, often known as the curse of dimensionality, may emerge during practical operations. This can be mitigated by implementing the Adaptive Lasso penalty method [3]. By allocating distinct penalty coefficients to different variables, the Adaptive Lasso Bayesian Quantile Regression (ALBQR) model can be derived.

3.2. Integration with other fields / development of specialized data structures for heritage conservation

Bayesian methods exhibit extensive applicability across various domains. For example:

Path Analysis: In the context of a petrochemical enterprise, Bayesian - based path analysis can be conducted to explore the influencing factors and multiple mediation pathways contributing to insomnia among male employees. By leveraging the posterior probability distributions of variables, it was determined that daily noise exposure time (99.8%), self - reported occupational noise exposure (98.0%), and occupational stress (100%) are significant predictors of elevated insomnia

scores. Subsequently, a chain multiple mediation model was constructed using Bayesian methods [4].

Precision Estimation: Bayesian methods can be employed to estimate missile hit accuracy. Bayesian estimation comprehensively incorporates sample, prior, and population information regarding missile hit accuracy to estimate the hit parameter θ . Through an MCMC - based model, the posterior distribution and posterior expectation of the parameter are obtained [5].

Regarding heritage conservation, as which is shown in Figure 1, an attempt can be made to integrate Bayesian methods into heritage - specific data structures. This integration aims to enable dynamic risk assessment and uncertainty quantification. The underlying concept involves integrating multi - source heterogeneous data via probabilistic models, thereby facilitating continuous updates to heritage state assessments.

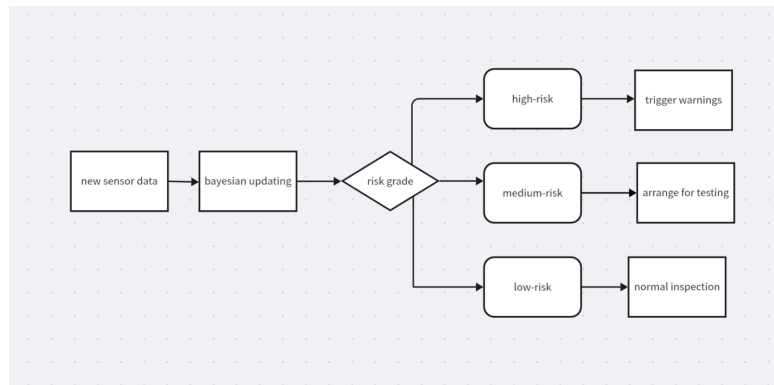


Figure 1. Special data structure for heritage protection

Python Implementation:

```

class HeritageBayesianModel:
    def __init__(self, heritage_id):
        self.id = heritage_id # Unique heritage identifier
        self.prior = {} # Prior probability distributions
        self.evidence = [] # Sequence of observed evidence
        self.network = { # Bayesian Network Structure
            'nodes': ['Environmental Erosion', 'Human Impact', 'Structural
Stability'],
            'edges': [
                ('Environmental Erosion', 'Structural Stability'),
                ('Human Impact', 'Structural Stability')
            ]
        }
        self.cpt = { # Conditional Probability Tables (CPT)
            'Environmental Erosion': {'low': 0.7, 'high': 0.3},
            'Human Impact': {'low': 0.6, 'high': 0.4},
            'Structural Stability': {

```

```
        ('low', 'low'): 0.95,
        ('low', 'high'): 0.6,
        ('high', 'low'): 0.4,
        ('high', 'high'): 0.1
    }
}

self.posterior = None # Posterior probability distributions
class HeritageDataRepository:
    def __init__(self):
        self.heritages = {} # heritage_id -> HeritageBayesianModel
        self.sensor_data = { # Multi-source sensor data pool
            'environment': [], # Temp, Humidity, etc.
            'structural': [], # Vibration/Displacement
            'human_impact': [] # Visitor traffic, etc.
        }
```

Key Component Description

First, Bayesian Network Layer

- * Nodes: Key risk factors (Environmental Erosion, Human Impact, Material Aging, etc.)
- * Edges: Causal relationships between risk factors
- * CPT: Quantifies conditional probabilities

```
python # Example CPT: Probability of Structural Stability is only 10% when Env.
Erosion is high AND Human Impact is high cpt = {('high', 'high'): 0.1}
```

Second, Dynamic Update Mechanism

```
python def update_posterior(model, new_evidence): # Update posterior probability
using Bayes' theorem #  $P(AB) = P(BA) * P(A) / P(B)$  for node in
model.network['nodes']: model.posterior[node] = calculate_updated_belief(
model.prior[node], new_evidence, model.cpt ) model.evidence.append(new_evidence)
```

Third, Multi-source Data Fusion

```
python def fuse_sensor_data(repo): # Fuse physical sensor data + manual
inspection data env_score =
calc_environment_score(repo.sensor_data['environment']) human_score =
calc_human_impact(repo.sensor_data['human_impact']) return {'Environmental
Erosion': env_score, 'Human Impact': human_score}
```

4. Conclusion

In conclusion, Bayesian methods are highly valuable in heritage conservation digitalization, thanks to their dynamic updating and uncertainty quantification.

Their applications mainly fall into two areas. Parameter optimization uses Bayesian Optimization algorithms to fine-tune machine-learning hyperparameters, enhancing prediction accuracy, like in historical-district fire-risk models. Risk modeling and decision-making offer flexibility in model selection, from Beta-Bernoulli for binary events to Bayesian Linear Regression for continuous variables, and Bayesian Networks for multi-source risk-factor integration.

However, challenges exist. Hyperparameter sensitivity can be handled with iterative optimization, and high - dimensional data can be reduced using penalty methods. Cross - domain integration and DBN development are necessary for better application.

The future direction is to create a Probabilistic Knowledge Base for Heritage Conservation. By integrating expert knowledge and multi - source data via a Bayesian continuous - learning mechanism, an “Assessment - Update - Warning” closed - loop can be formed for preventive conservation. This paper emphasized Bayesian methods in heritage conservation, but future research should explore their use in other domains for more comprehensive development.

References

- [1] Du, T., Liu, Z., Tian, J., & Wang, J. (2024). Fire risk prediction for historical and cultural districts based on Bayesian optimized LightGBM model. *Digital Technology & Application*, 42(11), 94-96.
- [2] Cui, J., & Yang, B. (2018). Bayesian optimization methods and applications: A review. *Journal of Software*, 29(10), 3068-3090. doi: 10.13328/j.cnki.jos.005607.
- [3] Guo, J. (2016). Application of variable penalty effect in Bayesian quantile regression model. *Statistics and Decision*, (19), 20-22. doi: 10.13546/j.cnki.tjyj.2016.19.006.
- [4] Zeng, K., Ma, J., Lan, Y., Huang, L., Wang, H., & Wang, J. (2025). Analysis of influencing factors and multiple mediation pathways of insomnia among male employees in a petrochemical enterprise based on Bayesian method. *Occupational Health and Emergency Rescue*, 43(3), 316-323. doi: 10.16369/j.oher.issn.1007-1326.2025.240262.
- [5] Liu, H., Hu, T., Chen, T., Liang, J., & Li, M. (2025). Bayesian estimation of missile hit accuracy based on Dirichlet distribution. *Journal of National University of Defense Technology*, 47(3), 119-127.