UDK 519.245

ESTIMATING THE INEFFICIENCY OF MACHINERY AND ITS OPERATORS USING HIERARCHICAL BAYESIAN MODELS

N.V. Guletskii
Systems analyst at HiQo Solutions,
master's student at BSU

Ya. M. Radyna
Statistical analyst at HiQo Solutions,
associate professor at BSU

HiQo Solutions, Ltd., Republic of Belarus
E-mail: nikolay.guletskiy@hiqo-solutions.com

N. Guletskii

Graduated from the Belarusian State University. Master's student at BSU. Works as a systems analyst at HiQo Solutions, Ltd, Republic of Belarus, Minsk. Fields of research include functional analysis, abstract harmonic analysis, optimization, artificial neural networks, computer vision, and Bayesian modeling.

Ya. Radyna

PhD in Mathematics. Associate professor at Belarusian State University. Works as a statistical analyst at HiQo Solutions, Ltd, Republic of Belarus, Minsk. Fields of research include Functional Analysis, Statistics and Machine Learning.

Abstract. Inefficient operation and deterioration of machinery leads to increased costs and energy consumption. Here we describe a method for disentangling operator inefficiency from the inefficiency of their equipment by building a hierarchical Bayesian model to model the fuel consumption of each operation.

Keywords: Hierarchical Bayesian models, Bayesian Analysis, Markov Chain Monte Carlo, Statistical Modeling.

Introduction. The goal of this paper is to detect and attribute increased fuel consumption in an industrial process involving a set of machines, grouped by model, and a set of operators, which operate the machinery. Each operation involves a single piece of machinery and a single operator but may consist of multiple sub-operations which have different fuel consumption dynamics. The fuel consumption is measured on a per-operation basis, meaning that we do not have fuel consumption measurements for individual sub-operations. We assume that each set of machines used by a given operator and vice versa is sufficiently diverse to be able to draw conclusions about the overall (marginal) efficiency of that operator/machine.

As a baseline one might consider linear and log-linear models that predict the fuel consumption from operation duration, operator id, machine id and model id (encoded using dummy variables). The parameters corresponding to operators and machines can then be reinterpreted as absolute inefficiencies. Unfortunately, linear models only allow additive inefficiency, while log-linear models require a log-transform of fuel consumption. Both restrictions are sub-optimal since they do not accurately reflect the dynamics of fuel consumption: operation inefficiency should be multiplicative, but the fuel consumption should be almost linear in operation duration. Moreover, such approaches do not allow us to incorporate clustering by model and are unlikely to produce interpretable results for quantities such as the ideal fuel consumption for a given operation.

We propose using non-linear hierarchical Bayesian models to model the dynamics of fuel consumption. Unlike basic frequentist methods, like lasso regression [1], Bayesian hierarchical modeling allows us to express arbitrary physical processes involved in the industrial process and infer their latent parameters, which can then be used to perform comparisons with the machinery's reference documentation.

The dataset. Our proprietary dataset consists of approximately 700000 rows with 128 operators and 21 machines, grouped into 3 machine models. The dataset has many outliers, exhibits multi-modality, and likely suffers from dataset shift. Moreover, we can't distinguish between multiple sub-operation subtypes within the dataset: there are multiple tasks involving the same sub-operations which result in slightly different fuel consumption.

Our model. For each machine model and each sub-operation type, we introduce an ideal fuel consumption coefficient which determines the minimal fuel consumption per unit of time that can be achieved with a perfectly maintained machine and a perfect operator. For each machine and operator, we introduce an inefficiency coefficient, which, when added to 1, acts as a multiplier for a sub-operation's ideal fuel usage. The inefficiency is either taken into account or ignored depending on the nature of the sub-operation, but we expect at least one sub-operation type to include both machine and operator inefficiency. The expected fuel consumption for an operation is declared to be the sum of expected fuel consumption for sub-operations, which may be modeled as non-linear functions of both latent and observed parameters. We then assume that the actual fuel consumption is sampled from a log-normal distribution located at the expected fuel consumption. The scale of the log-normal distribution is controlled by a global scale parameter σ. The general formula for the fuel consumption φ_i during operation i for m subtasks given vectors of measurements $x_{i,l}$ and physical models $f_s(x_{i,s})$ for all subtasks s, as well as the operator's and vehicle's inefficiency coefficients θ_{op} and θ_{mc}, may be described as follows:

$$\tilde{\varphi}_i = \sum_{i=1}^{m} \left(1+\theta_{op}\right)^{q_{op,s}} (1+\theta_{mc})^{q_{mc,s}} f_s\left(x_{i,s}\right),$$

$$q_{op,s}, q_{mc,s} \in \{0,1\},$$

$$\varphi_i \sim Lognormal\left(\log(\tilde{\varphi}_i) - \frac{\sigma^2}{2}, \sigma^2\right).$$

Kernel trick. To reduce the heteroskedasticity caused by the presence of multiple operation sub-types which can't be distinguished directly, we introduce a fuel consumption adjustment coefficient $b_m(x_{i,s,dur})$, which depends on the duration of a specific sub-task and acts as an additional non-linearity inside the physical model. To apply the kernel trick, we associate fixed-length weight, center and scale vectors w_m, y_m, and γ_m with each machine model m, and calculate the coefficient as a dot product of the weight vector with a vector of RBF kernels with the corresponding centers and scales. To ensure that different MCMC chains converge to the same posterior distribution, we restrict the vector of centers to be a monotonically increasing sequence of numbers, which is acceptable in our case because the kernel trick is applied to a 1-dimensional space (representing a sub-operation's duration). The monotonicity is achieved by reparametrizing the vector of centers to be a cumulative sum of positive offsets β_m, which define how far each following center is from the previous one:

$$b_m\left(x_{i,s,dur}\right) = \sum_{k=1}^{n} w_{m,k} \exp\left(-\gamma_{m,k} \left| x_{i,s,dur} - \sum_{j=1}^{k} \beta_{m,j} \right|^2\right).$$

Priors. For ideal fuel consumption, we use a fixed truncated normal prior with fixed parameters determined from domain experience. For other parameters of the physical model, we also use normal and truncated normal distributions with weakly-informative hyperpriors. For inefficiency coefficients, we use half-normal priors with a weakly-informative (Half-Cauchy(5)) hyperprior for each group (i.e. two hyperpriors, one for operators and one for machines). For the observation noise scale parameter, we also use a weakly-informative (Half-Cauchy(5)) hyperprior. For the kernel trick's weights and center offsets, we use Half-Normal(1) priors. For the kernel scales, we use a truncated normal distribution with a unit mean and scale.

Parametrization. Despite the relatively large size of the dataset, we use non-centered parametrizations for all truncated normal and half-normal distributions. To sample from a truncated normal distribution $\mathcal{TN}(\mu, \sigma^2, l)$, where l is the minimum value for numbers sampled from the distribution, we multiply the samples from a truncated normal distribution $\mathcal{TN}(0, 1, l')$ by the desired scale σ and add the desired mean μ. The value of l' corresponding to the desired minimum value l is given by

$$l' = \sigma^{-1}\left[l - \left(\mu - \sigma \frac{-p_{\mathcal{N}(0,1)}\big((l-\mu)/\sigma\big)}{1 - cdf_{\mathcal{N}(0,1)}\big((l-\mu)/\sigma\big)}\right)\right].$$

Inference. Our implementation is based on numpyro[2]. We fit the model using the No-U-Turn Sampler (NUTS) [3], an adaptive sampler based on Hamiltonian Monte Carlo. We set the target acceptance probability to 0.99 to make sampling more robust to the high curvature introduced by the non-linear physical model. We achieve convergence despite the complex geometry of the typical set (Gelman-Rubin statistic = 1 ± 1e-2 for all trained models).

Preprocessing. To speed up convergence, we normalize the operation durations, fuel consumption and other positive observations to have a mean of 1.

Evaluation. To validate whether our model is capable of predicting the fuel usage for individual operations, we perform a 1%-to-99% stratified train-test split and compare the mean absolute error of our model to four 5-fold cross-validated lasso regression models with 3rd degree polynomial features: two linear models, two log-linear models, two models with categorical variables and two models without categorical variables. We also compare our hierarchical model to a model obtained by removing the inefficiency coefficients from the main model. The small size of the training set is due to the computational complexity of training the Bayesian model.

To validate whether the model is capable of determining the inefficiencies of various parties, we build an auxiliary dataset by randomly selecting 10 operators and 5 machines and artificially adding inefficiencies to their operations by multiplying fuel usage by a per-operator/per-machine random number which we call jitter. We then train two models: one on a 5% subsample of the original dataset, and one on the auxiliary dataset constructed from the subsample. The relative increase in fuel consumption between datasets can then be derived from the physical model's definition and compared to the random jitter applied during the creation of the auxiliary dataset.

Results on the prediction task. Our model has the largest coefficient of determination (R^2) and the lowest mean average error (MAE) on the hold-out set. Table 1 holds the results of an ablation study performed on a 1%-to-99% split of the dataset. For all Bayesian models, we ran NUTS for 1000 iterations, the first 500 of which were used for adaptation.

Table 1: – Prediction error statistics on the hold-out set.

Model	MAE	R^2
Linear without categorical	0.128	0.641
Linear with categorical	0.128	0.641
Log-linear without categorical	0.129	0.624
Log-linear with categorical	0.129	0.625
Bayesian without inefficiencies	0.119	0.643
Bayesian	0.119	0.655
Bayesian with a kernel trick	**0.116**	**0.661**

This demonstrates that linear models do not benefit from operator/machine id information, while introducing inefficiency coefficients into the Bayesian model noticeably reduces the unexplained variance.

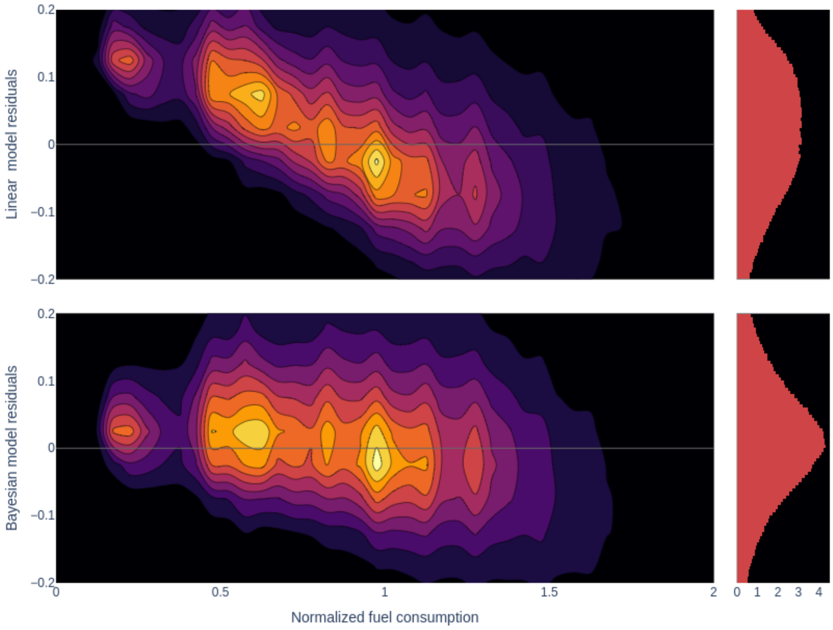

Figure *1*: – Residual plots for the best linear model and the best Bayesian model

Linear models and basic Bayesian models exhibit major heteroskedasticity, while the best Bayesian model is relatively homoscedastic.

Results on the synthetic inefficiency detection task. To fit the Bayesian model to both the original and the auxiliary datasets, we ran NUTS for 2000 iterations, the first 1500 of which were used for adaptation. To recover the jitter multiplier γ from the original predicted inefficiency θ_{orig} and the predicted inefficiency on the jittered dataset θ_{jitt}, we use the following relationships:

$$(1+\theta_{orig})(1+\gamma) \approx (1+\theta_{jitt}),$$

$$\gamma \approx \frac{\theta_{jitt}-\theta_{orig}}{1+\theta_{orig}}.$$

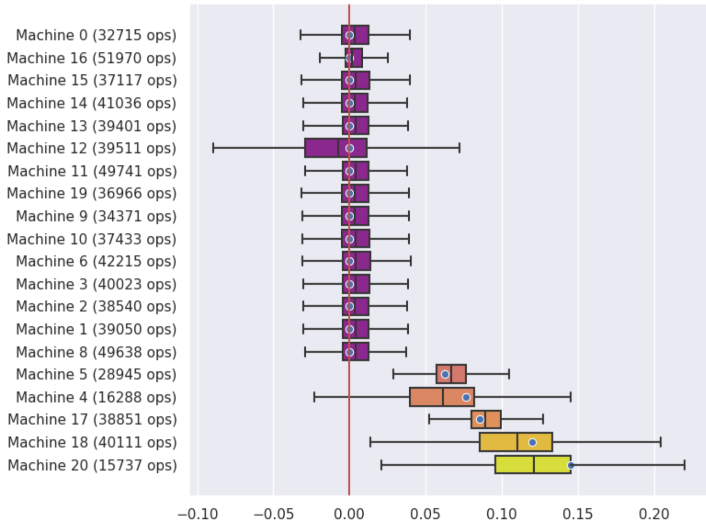

Figure *2*: – A comparison of the predicted relative machine inefficiency jitter with the actual jitter used to construct the auxiliary dataset. The box plot summarizes the posterior distribution of the jitter multiplier, while the dots represent the actual fuel usage jitter multipliers. The closer the median line is to the blue dot, the better.

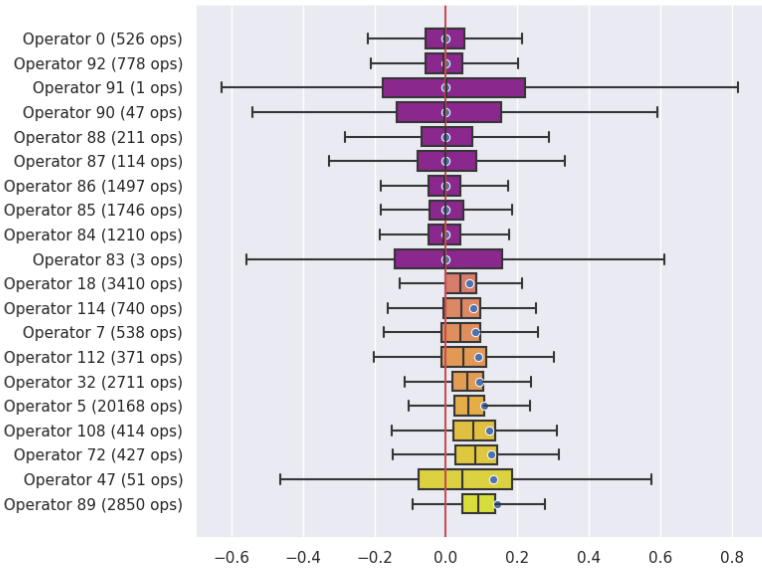

Figure 3: – A comparison of the predicted relative operator inefficiency jitter with the actual jitter used to construct the auxiliary dataset. The box plot summarizes the posterior distribution of the jitter multiplier, while the dots represent the actual fuel usage jitter multipliers. The closer the median line is to the blue dot, the better.

Our model successfully identifies the inefficiencies caused by the machines. For operators, the estimates are significantly more conservative, but this may be explained by the model's low confidence in its predictions for operators who are underrepresented in the sample. Nevertheless, the estimates for operators which were not affected during the creation of the modified dataset are near zero, while the estimates for operators which were affected are almost always significantly larger, and for most operators, the actual jitter is within the inter-quartile range. This indicates that the model gives conservative estimates, which is ethically desirable.

Conclusion. The proposed Bayesian model is superior to the proposed baselines in multiple ways. Firstly, it provides credible intervals for parameters describing the inefficiencies of operators and machines, allowing us to judge whether it is reasonable and fair to make judgements from the inferred parameters. Secondly, it allows us to integrate non-linear physical relationships into our model, giving us interpretable parameters, which can be inspected by a domain expert to determine whether the obtained estimates are reasonable. Thirdly, it outperforms the baselines on challenging predictive power benchmarks and is capable of dealing with various problems present within the dataset, such as multimodality and the presence of outliers. Our experiments demonstrate that the proposed model detects artificially added inefficiencies and that it is not overconfident about its predictions.

References

[1] Tibshirani R. Regression Shrinkage and Selection via the Lasso // Journal of the Royal Statistical Society. Series B (Methodological). 1996. Vol. 58, № 1. P. 267–288.

[2] Phan D., Pradhan N., Jankowiak M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro // arXiv preprint arXiv:1912.11554. 2019.

[3] Hoffman M.D., Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. // Journal of Machine Learning Research. 2014. Vol. 15, № 1. P. 1593–1623.

ОЦЕНКА НЕЭФФЕКТИВНОСТИ ОБОРУДОВАНИЯ И ЕГО ОПЕРАТОРОВ С ПОМОЩЬЮ ИЕРАРХИЧЕСКИХ БАЙЕСОВСКИХ МОДЕЛЕЙ

Н.В. Гулецкий
Системный аналитик, СООО “ХайКво Солюшенс”, магистрант БГУ

Е.М. Радыно
Системный аналитик в СООО "ХайКво Солюшенс", доцент Белорусского государственного университета, канд. физ.-мат. наук

СООО “ХайКво Солюшенс”, Республика Беларусь, Минск
E-mail: nikolay.guletskiy@hiqo-solutions.com

Аннотация. Ненадлежащее состояние оборудование и его неэффективное использование приводит к завышеному потреблению энергоресурсов и повышению затрат. В данной статье описывается метод распутывания неэффективностей операторов оборудования от неэффективностей самого оборудования, основывающийся на иерархической Байесовской модели, моделирующей использование топлива.

Ключевые слова: Иерархические Байесовские модели, Байесовское моделирование, Методы Монте-Карло, Статистическое Моделирование.