Short message
DOI: https://doi.org/10.18721/JCSTCS.17305
UDC 004.855.5

PREDICTIVE MODELS AND DYNAMICS OF ESTIMATES OF APPLIED TASKS CHARACTERISTICS USING MACHINE LEARNING METHODS

A.V. Konstantinov ✉

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ andrue.konst@gmail.com

Abstract. The paper considers the machine learning problem of simultaneous estimation of the conditional survival distribution and dynamic characteristics of computational tasks. The problem arises in cluster workload management and is extremely relevant for optimal scheduling. To solve the problem, a new method is proposed, based on the combination of the attention mechanism and the random survival forest. The key feature is the use of a tree structure derived from a random survival forest. The forest construction algorithm uses only the survival dataset. Each leaf uses the unconditional Kaplan-Meier estimate, which is a serious limitation of the forest, especially for rare events in some parts of the feature space. Moreover, the random survival forest does not allow estimating the dynamic parameters of the task. The proposed method solves these problems by extending the already constructed random survival forest with the attention mechanism inside each leaf of the tree. The Beran estimator is used to model survival distribution, and the Nadaraya-Watson regression with the same parameters is used to predict the dynamic characteristics of tasks. To do this, subsets of training data corresponding to the same leaf as the input vector are used. As a result, the joint model is obtained that allows us to estimate the survival function more accurately and at the same time to predict the dynamic characteristics of the task. The developed model combines the advantages of smooth models based on the attention mechanism and stepwise decision trees.

Keywords: machine learning, survival analysis, attention mechanism, random survival forest, Beran estimator

Acknowledgements: The research was partially financially supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the state assignment "Development and research of machine learning models for solving fundamental problems of artificial intelligence in the fuel and energy complex" (FSEG-2024-0027).

Citation: Konstantinov A.V. Predictive models and dynamics of estimates of applied tasks characteristics using machine learning methods. Computing, Telecommunications and Control, 2024, Vol. 17, No. 3, Pp. 54–60. DOI: 10.18721/JCSTCS.17305

Краткое сообщение
DOI: https://doi.org/10.18721/JCSTCS.17305
УДК 004.855.5

МОДЕЛИ ПРЕДСКАЗАНИЯ И ДИНАМИКА ОЦЕНОК ХАРАКТЕРИСТИК ПРИКЛАДНЫХ ЗАДАЧ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

А.В. Константинов ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ andrue.konst@gmail.com

Аннотация. В статье рассматривается задача машинного обучения, заключающаяся в одновременной оценке условного распределения выживаемости и динамических характеристик вычислительных задач. Проблема возникает при управлении рабочей нагрузкой кластера, и крайне актуальна для оптимального планирования. Для решения задачи предложен новый метод, основанный на комбинации механизма внимания и случайном лесе выживаемости. Ключевой особенностью является использование древовидной структуры, полученной случайным лесом выживания. Алгоритм построения леса опирается только на данные задачи выживаемости. В каждом листе используется безусловная оценка Каплана-Мейера, что является серьезным ограничением леса, особенно в случае редких событий в некоторых частях пространства признаков. Более того, случайный лес выживаемости не позволяет оценить динамические параметры задачи. Предлагаемый метод решает данные проблемы, дополняя уже построенный случайный лес выживаемости механизмом внимания внутри каждого листа дерева. Для моделирования выживаемости применяется оценка Берана, а для предсказания динамических характеристик задач – регрессия Надарая-Ватсона с теми же параметрами. Для этого используются подмножества обучающих данных, соответствующие тому же листу, что и входной вектор. В результате получена совместная модель, позволяющая более точно оценить функцию выживаемости и одновременно предсказать динамические характеристики задачи. Разработанная модель сочетает в себе преимущества гладких моделей, основанных на механизме внимания, и ступенчатых деревьев решений.

Ключевые слова: машинное обучение, анализ выживаемости, механизм внимания, случайный лес выживаемости, оценка Берана

Финансирование: Исследование выполнено при частичной финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания «Разработка и исследование моделей машинного обучения для решения фундаментальных задач искусственного интеллекта в топливно-энергетическом комплексе» (FSEG-2024-0027).

Для цитирования: Konstantinov A.V. Predictive models and dynamics of estimates of applied tasks characteristics using machine learning methods // Computing, Telecommunications and Control. 2024. Т. 17, № 3. С. 54–60. DOI: 10.18721/JCSTCS.17305

Introduction

Computational clusters are widely used to solve problems that require significant computing power, allowing many computational tasks to be performed simultaneously. One of the key aspects of effective cluster resource allocation planning is estimating the parameters of computing tasks, such as execution time, and individual characteristics that are unknown at the time the task is launched. Each computational task is characterized by a feature vector that is a set of input parameters, including user-specified characteristics and parameters determined by the state of the system at the time the task is queued, including its execution time. After starting a task, two outcomes are possible: the task is completed within the allotted time, or the task is terminated by the control system after the specified time has elapsed, in other words, censoring occurs. Thus, the task of estimating execution time is to determine the expected time to an event (completion of a task) under censoring conditions and is the task of survival analysis.

The i-th training observation consists of input feature vector x_i, time to event t_i, censoring indicator δ_i and target vector of computational task parameters y_i. The event in the current problem is task completion or interruption. The time to event t_i corresponds to the time between the task launch and completion or interruption, depending on δ_i. The censoring indicator $\delta_i = 1$, if the task has finished normally. Otherwise, the task execution has interrupted. The interruption can be caused by time limit violation or some program error. In this case, the observation is called censored in terms of survival analysis. Finally, the training data set D is composed of N labeled training observations:

$$D = \left\{\left(x_i, t_i, \delta_i, y_i\right)\right\}_{i=1}^{N}.$$

Since the execution time does not depend on the input feature vector deterministically, we introduce T, a random variable corresponding to the execution time. The main goal is to estimate the parameters of an applied task $y(x)$, including, but not limited to, the expected execution time of the task, conditioned on the input feature vector $x : \mathbb{E}\left[T \middle| X = x\right]$. More generally, the survival distribution $S(t|x)$ is of interest.

Survival Models

In this paper, we consider two base models for survival function estimation. The first is the Random Survival Forest (RSF) [1]. It is a machine learning algorithm that does not impose any assumptions on the data distribution, which makes it different from classical survival analysis methods for conditional distribution estimation, for example Cox Proportional Hazards [2]. Instead, it partitions the data using feature vector x, and then estimates distribution shape based on unconditional non-parametric statistics. The second model is the Beran estimator [3]. It can be considered as the kernel-based extension of Kaplan-Meier unconditional estimator to the conditional case. Given the weight function W, the method estimates conditional survival function as:

$$\hat{S}(t|x) = \prod_{t_i \le t} \left(1 - \frac{W(x, x_i)}{1 - \sum_{j=1}^{i-1} W(x, x_j)}\right)^{\delta_i},$$

where W is normalized over dataset points, and training observations are ordered such that time t_i increases by the index i. Specifically, in the original Beran the weights are obtained by normalization of kernels:

$$W(x, x_i) = \frac{K(x, x_i)}{\sum_{j=1}^{N} K(x, x_j)}.$$

Let us describe the RSF construction and prediction algorithms. Like in classical Random Forest, the trees of the forest are built independently of each other, using different random dataset subsamples and feature subsets (random subspace method). Each tree is built by a recursive algorithm. At each step the algorithm considers a tree node and tries to make a split, resulting in two child nodes, connected to the node. The data point falls to the left child node if some selected feature value is less than the specific threshold, and otherwise it falls to the opposite, right node. When splitting a node, at the training stage, the feature and threshold values are determined by optimizing goodness of split criterion. The main goal of splitting is to divide the sample received into it in such a way that the survival distributions for the left and right subtrees are as different as possible. For this purpose, the logrank test is used as a criterion. The training algorithm stops splitting a node, when the number of training data points falling to the node is less than some predefined number. The resulting decision trees can be used to estimate the conditional survival function for a new observed input feature vector: at each leaf of each tree, a nonparametric Kaplan-Meier estimator is constructed based on the data falling to the leaf. Note that even though such an estimation is unconditional, each leaf of the tree corresponds to a strictly defined region of space for which this estimation is valid.

Attention mechanism

The attention mechanism is the main element of the most successful method for processing sequences, including natural language, the Transformer [4]. It is implemented as a convex combination of vectors called value vectors, where the weights are obtained by kernel applied to the query and key vector pairs. This mechanism allows the model to focus on the most important parts of the input data when making predictions. In recent years, attention mechanisms have been successfully applied to problems from other domains, such as computer vision, speech recognition, and regression and classification [5]. Despite successful application of attention mechanism to many machine learning problems, attention mechanisms and their combinations with decision trees have not previously been used to estimate survival distributions, and simultaneously solve survival analysis and regression problems.

Attention mechanism can be formalized as follows. Let q be a "query" vector, $\{(k_i, v_i)\}_{i=1}^{K}$ be a set of "key-value" pairs, and "score" be a function mapping pair (q, k_i) to relative score or relevance of the query to the key. Attention of "q" to the given set of pairs is the convex combination of values v_i:

$$A\left(q, \{(k_i, v_i)\}_{i=1}^{K}\right) = \sum_{i=1}^{K} \alpha_i v_i,$$

where coefficients α_i are defined as:

$$\alpha_i = \left(softmax\left[\left(score\left(q, k_j\right) \right)_{j=1}^{K} \right] \right)_i = \frac{\exp\left(score\left(q, k_i\right)\right)}{\sum_{j=1}^{K} \exp\left(score\left(q, k_j\right)\right)}.$$

Therefore, attention is a function of vector and a set of pairs of arbitrary size, which maps them the one vector, characterizing the set for the query linearly. By using this property, attention was successfully applied for improving Random Forest performance [6]. It should be noted that attention resembles well-known kernel regression algorithm, called the Nadaraya-Watson regression [7]. Indeed, if score is defined as:

$$score(q, k) = -\frac{1}{2}\|q - k\|^2,$$

and as pairs (x_i, y_i) are considered, then attention is equivalent to the Nadaraya-Watson regression with the Gaussian kernel. However, the "score" function can be more complex, reflecting complex structure of the dataset, for example, it can be implemented as a neural network and trained in an end-to-end manner [8].

Attention-based Random Survival Forest

We propose a new approach based on incorporation of the attention mechanism into RSF for estimating the parameters of applied tasks, called Attention-based Random Survival Forest (ABRSF). The key idea of ABRSF is to leverage the same attention weights to improve quality of survival estimation and to approximate the task parameter vector.

The algorithm consists of two steps. At the first step, a classical RSF is constructed using the survival dataset. Its leaf nodes estimate unconditional survival distribution by the Kaplan-Meier estimator. These estimators can be replaced by the Beran models, based on attention weights, instead of classical kernels. At the same time, the same attention model can also be considered as kernel regression and applied with the same kernel to solve the task parameter vector estimation problem [8]. Formally, in each tree leaf attention weights are calculated based on the train data points which fall into the same leaf as "keys", and an input vector as "query". Then, for survival function estimation Beran estimator is applied, where attention weights are as used instead of kernels. For the task parameter vector estimation simply the attention mechanism is used, where "values" are training dataset task parameter vectors. It is important, that the attention in the proposed model is applied only locally, where neighboring points are determined by the RSF structure, which was optimized for the survival problem. The obtained model is smooth inside each region defined by leaf, and has discontinuities at separating hyperplanes, defined by internal nodes of the forest trees.

The described one-step approach allows us to solve the formulated problem but has the following drawback: different random forest trees, depending on the training subsample and feature subspace, have different accuracy. In addition, some trees may be accurate in the context of a survival problem and less accurate in the context of a regression task of estimating target parameters. To eliminate this drawback, we modify the proposed approach by adding tree weights. In addition to estimating the target parameters, each leaf of each tree also estimates the input feature vector using the same mechanism, where feature vectors are used as "values", as well as "keys":

$$\hat{x} = A\left(x, \left\{ \left(x_{l_i}, x_{l_i} \right) \right\}_{i=1}^{K_l} \right),$$

where l represents indices of train dataset points, falling into the same leaf as x, and K_l is the number of such points. The negative distance or "score" between x and its reconstruction $\hat{x}$ can be used as a measure of attention weights quality. By how close the input feature vector is to its estimate, one can judge about closeness of the target parameter vector estimate to the true value. So, next, the feature vector reconstructions obtained from different trees act as keys for the global attention.

Let the $\hat{x}(j)$ be a reconstruction of the input feature vector x by the j-th tree, and the $\hat{y}(j)$ be the task parameter vector estimation by the same tree. Then the final estimation is defined by the attention:

$$\tilde{y} = A\left(x, \left\{ \left(\hat{x}(j), \hat{y}(j) \right) \right\}_{j=1}^{\tau} \right),$$

where τ is the number of trees. The "query" is the original feature vector, the "keys" are reconstructed feature vectors, and the "values" are the estimates of the task parameter vector. The same technique is applied to combine tree survival function estimations to the final one.

Finally, the ABRSF model consists of two layers: tree-level estimations and forest-level weighted combination. The scheme of the ABRSF model is shown in Fig. 1. At the first layer, each tree estimates three parameters: the survival function, encoded as a vector, the task parameter vector, and the input vector reconstruction. At the second layer, tree-level estimates are combined by using attention weights, obtained by collating the input vector reconstructions with the given input feature vector. After passing these two layers, the final estimates are locally smooth and more precise than piecewise constant RSF ones. Moreover, the task parameter vector is calculated by using the RSF structure, because only points in the same leaf are considered in each tree, which lead to more accurate results when the survival data is correlated with the estimated vector.

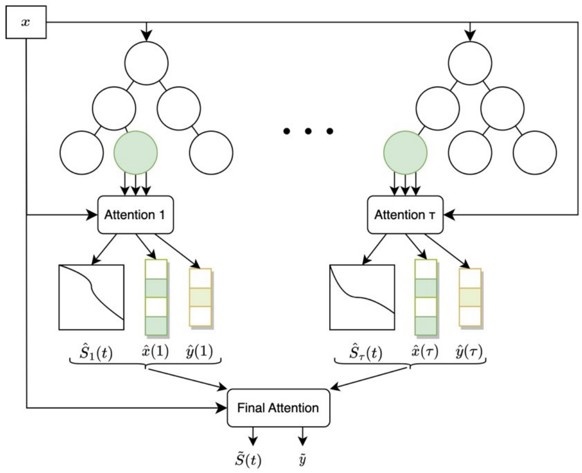

Fig. 1. ABRSF model scheme

Conclusion

The problem of joint target parameter vector and survival distribution estimation has been considered. The novel method, based on combination of Random Survival Forest and Attention mechanism, and called ABRSF, is proposed. The developed method has advantages in comparison to classical forest: it builds piecewise smooth prediction models and leverages the survival tree structure, when estimating the task parameter vector. As a further direction, this approach can be expanded by using multilayer neural networks in the attention mechanism and training the model on regression and survival analysis problems simultaneously, using the backpropagation algorithm, as well as adapting the approach for correctly processing missing features in the input data.

REFERENCES

1. **Ishwaran H., Kogalur U.B., Blackstone E.H., Lauer M.S.** Random Survival Forests. The Annals of applied statistics, 2008, Vol. 2, No. 3, Pp. 841–860. DOI: 10.1214/08-AOAS169

2. **Fox J.** Cox proportional-hazards regression for survival data. An R and S-PLUS Companion to Applied Regression, 2002, Pp. 1–18.

3. **Beran R.** Nonparametric regression with randomly censored survival data, 1981.

4. **Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I.** Attention is all you need. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2023, pp. 1–15. DOI: 10.48550/arXiv.1706.03762

5. **Lin T., Wang Y., Liu X., Qiu X.** A survey of transformers. AI open, 2022, Vol. 3, Pp. 111–132. DOI: 10.1016/j.aiopen.2022.10.001

6. **Bierens H.J.** The Nadaraya–Watson kernel regression function estimator. Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models. Cambridge: Cambridge University Press; 1994.

7. **Utkin L.V., Konstantinov A.V.** Attention-based random forest and contamination model. Neural Networks, 2022, Vol. 154, Pp. 346–359. DOI: 10.48550/arXiv.2201.02880

8. **Konstantinov A.V., Utkin L.V., Lukashin A.A., Muliukha V.A.** Neural attention forests: Transformer-based forest improvement. International Conference on Intelligent Information Technologies for Industry, 2023, Pp. 158–167. DOI: 10.48550/arXiv.2304.05980

INFORMATION ABOUT AUTHOR / СВЕДЕНИЯ ОБ АВТОРЕ

Konstantinov Andrei V.
Константинов Андрей Владимирович
E-mail: andrue.konst@gmail.com
ORCID: https://orcid.org/0000-0002-1542-6480

Submitted: 15.06.2024; Approved: 26.08.2024; Accepted: 13.09.2024.
Поступила: 15.06.2024; Одобрена: 26.08.2024; Принята: 13.09.2024.