Computer Science and Information Systems 19(2):1007–1022 https://doi.org/10.2298/CSIS210609016P

A Machine Learning Approach for Learning Temporal Point Process

Andrija Petrović[1], Aleksa Biserčić[2], Boris Delibašić[2], and Dimitrije Milenković[2]

[1] Technical Faculty University Singidunum, Danijelova 32, 11000 Belgrade, Serbia
apetrovic@singidunum.ac.rs
[2] Faculty of Organizational Sciences, POB 52, 11000 Belgrade, Serbia
boris.delibasic@fon.bg.ac.rs

Abstract. Despite a vast application of temporal point processes in infectious disease diffusion forecasting, ecommerce, traffic prediction, preventive maintenance, etc, there is no significant development in improving the simulation and prediction of temporal point processes in real-world environments. With this problem at hand, we propose a novel methodology for learning temporal point processes based on one-dimensional numerical integration techniques. These techniques are used for linearising the negative maximum likelihood (neML) function and enabling backpropagation of the neML derivatives. Our approach is tested on two real-life datasets. Firstly, on high frequency point process data, (prediction of highway traffic) and secondly, on a very low frequency point processes dataset, (prediction of ski injuries in ski resorts). Four different point process baseline models were compared: second-order Polynomial inhomogeneous process, Hawkes process with exponential kernel, Gaussian process, and Poisson process. The results show the ability of the proposed methodology to generalize on different datasets and illustrate how different numerical integration techniques and mathematical models influence the quality of the obtained models. The presented methodology is not limited to these datasets and can be further used to optimize and predict other processes that are based on temporal point processes.

Keywords: temporal point process, Hawkes process, Poisson process, highway traffic prediction, ski injury prediction.

1. Introduction

Nowadays, one of the most popular research areas is focused on modelling event sequences. Event sequencing has become extremely popular in a wide range of applications such as road traffic estimation [1], epidemiology prediction [2], network activities [3], bioinformatics [4], e-commerce, etc. Event data carry information about event occurrence. Additionally, event data can also provide information about classes of events, types of events, participants, etc. This type of point process is known as a marked point process.

A point process is extremely useful in modelling traffic congestion and traffic event occurrences, e.g. arrival of vehicles, pedestrian movement, etc. [8]. Simulating highway traffic and predicting highway congestion is one of the main problems connected with point process modeling [9].

If compared with time series, event occurrences are treated as random variables generated in an asynchronous manner, which makes them fundamentally different from the time series where equal and fixed time intervals are assumed. This property makes them useful in a wide variety of applications where discretizing events to a fixed interval would result in poor prediction performances and high computational cost.

Generally, there are two types of point process models: temporal (univariate) point process and spatial-temporal (multivariate) point process. In the case of the univariate point process, the objective is to model temporally correlated event occurrences, whereas in spatial-temporal point process the event occurrences are correlated in space and time. Generally, multivariate point process is mostly used in the analysis of protein patterns [5] and financial market predictions [6]. The general formulation of the point processes makes them available to model event occurrences, both continuous or discontinuous (with jumps). Additionally, the point process can be further generalized by stochastic differential equations to stochastic point process.

The main idea behind different types of point process models is hidden in modelling a conditional intensity function (CIF). A CIF can be interpreted heuristically as the expected number of events that are going to occur in an infinitely small timestamp (dt). CIF can be modelled as a constant (homogeneous process) or as a function of time (inhomogeneous process). Learning an intensity function from a given dataset presents one of the most popular subjects of research [7].

In this paper, we present a data-driven approach for learning different types of CIFs used in temporal point process models. Our approach is based on the implementation of numerical integration methods for linearization of negative maximum likelihood (neML) in order to backpropagate derivatives of neML.

We tested our methodology on two real-life datasets that consisted of exact timestamps. The first dataset included highway toll passes recordings, and was a high-frequency dataset. The second dataset included timestamps when ski injuries occurred, and was a low-frequency dataset. Our methodology shows that it can be successfully used for various types of CIFs. Furthermore, four different baseline models based on neML scores: second-order Polynomial inhomogeneous process, Hawkes with exponential kernel, Gaussian process, and Poisson process were compared. The proposed methodology was evaluated on several metrics, amongst which is the minimization of negative log likelihood loss for demonstration of how well models fitted conditional intensity functions, Akaike information criteria (AIC), and the mean absolute error (MAE) for evaluating the quality of prediction for future time events.

To summarize, the contributions of this work are as follows:

- We presented a novel framework for learning different type of CIF in temporal point process based on implementation of one-dimensional numerical integration techniques for linearization of neML.
- The proposed method can be used with any kind of one-dimensional numerical integration technique.
- The method is tested on two real world datasets with high frequency and low frequency occurring events.

– The obtained results showed satisfactory performances on both datasets with respect to MAE, log likelihood and AIC.

The remainder of the paper is structured as follows. In section 2 the related work is reviewed. Background methodology, point process and Ogata's modified thinning algorithm are presented in section 3. A novel methodology for learning point process is presented in section 4. Experimental setup and results of real-world applications are presented in sections 5 and 6, respectively. The conclusions are drawn in section 7.

2. Related Work

We structure the discussion of the related work into two broad, previously mentioned, categories: intensity-based approaches and intensity-free approaches. The intensity-based approaches present methods where a point process is modelled by different functional forms of CIFs [10]. Intensity-free approaches present methods where a point process is modelled with some type of unsupervised machine learning algorithms.

Intensity-based approaches present the oldest approaches in point process modelling. They rely on a functional form that completely depends on the CIF. The Poisson process presents the simplest point process where conditional intensity function has a constant value [11]. The more complicated variant of this process is observed when the CIF is modelled as a product of kernels [12]. Recent research proposed different variants of modelling CIF by deep neural networks [7, 13]. Xiao et al. [13] presented an interesting approach of modelling CIF by a recurrent neural network. However, in this paper authors assume that integral in negative maximum likelihood is correlated only with the current timestamp. Even though this strong assumption cannot be justified by theoretical properties of point process models, the obtained results were significantly better compared to well-known baseline models. Chen et al. [14] and Zhang et al. [15] presented an interesting approach for modelling dynamics by deep neural networks. Moreover, the authors presented an interesting example where the point process was modelled by a differential equation and solved using the Euler method. Besides, the authors implemented the backpropagation technique for reducing memory complexity during the training phase.

Intensity-free approaches are based on modelling point processes by unsupervised learning techniques [16]. When compared to intensity-based approaches these methods can obtain better results, but they are more prone to overfitting due to smaller datasets or large expressive powers of the model. Variational autoencoders (VAE) present unsupervised machine learning algorithms that are mostly used for point process modelling. The Action Point Process variational autoencoder (APP-VAE) presents a variational auto-encoder that can capture the distribution over the times and categories of action sequences [17]. The APP-VAE obtained state-of-the-art results on the MultiTHUMOS and Breakfast datasets. A declustering based hidden variable model that leads to an efficient inference procedure via a variational autoencoder for solving multivariate highly correlated point process is presented by [18]. Besides VAE, generative adversarial networks (GANs) have recently been proposed as a method for describing event occurrences [19]. The authors proposed an intensity-free approach for point process modelling that transforms nuisance processes to a true underlying distribution by using Wasserstein GANs. Experiments on various synthetic and real-world data substantiate the superiority of the proposed point process model over conventional ones. Compared to intensity-free approaches such as the GANs and VAEs, intensity approaches can provide information that is more explainable and interpretable in the case when CIF function is in linear form.

Applications of intensity-based point processes can be found in a wide variety of areas [35-37]. Point process are useful in medical care and health care. Liu et al. [28] presented an EM (expectation maximization) based point process for modelling of drug overdoses with heterogeneous and missing data. Additionally, a daily living activity prediction via combination of temporal point process and neural networks is presented in [29]. Besides medical care, point processes are also used in a wide variety of problems related to traffic [32-34]. A novel framework for modelling traffic congestion events over road networks based on spatio-temporal point process in combination with attention mechanism is presented in [30]. Motagi et al. [31] developed a self-exciting temporal point process to analyse crash events data and classify it into primary and secondary crashes. This model uses a self-exciting function to describe secondary crashes while primary crashes are modelled using a background rate function.

The model presented in this paper belongs to the class of intensity-based approaches. Compared with the standard intensity-based approaches, our model has more expressive power, whereas compared with intensity-free approaches it is less prone to overfitting. Moreover, our model can be easily applied to any type of conditional intensity functions, and linearization term can be also applied to very deep neural networks.

3. Framework

Based on the methodology proposed in this paper, we implemented a general framework for learning point processes. The framework consists of three distinct parts: Data cleaning and processing, model and hyperparameter selection, evaluation and simulation part. The framework is presented in Fig. 2.

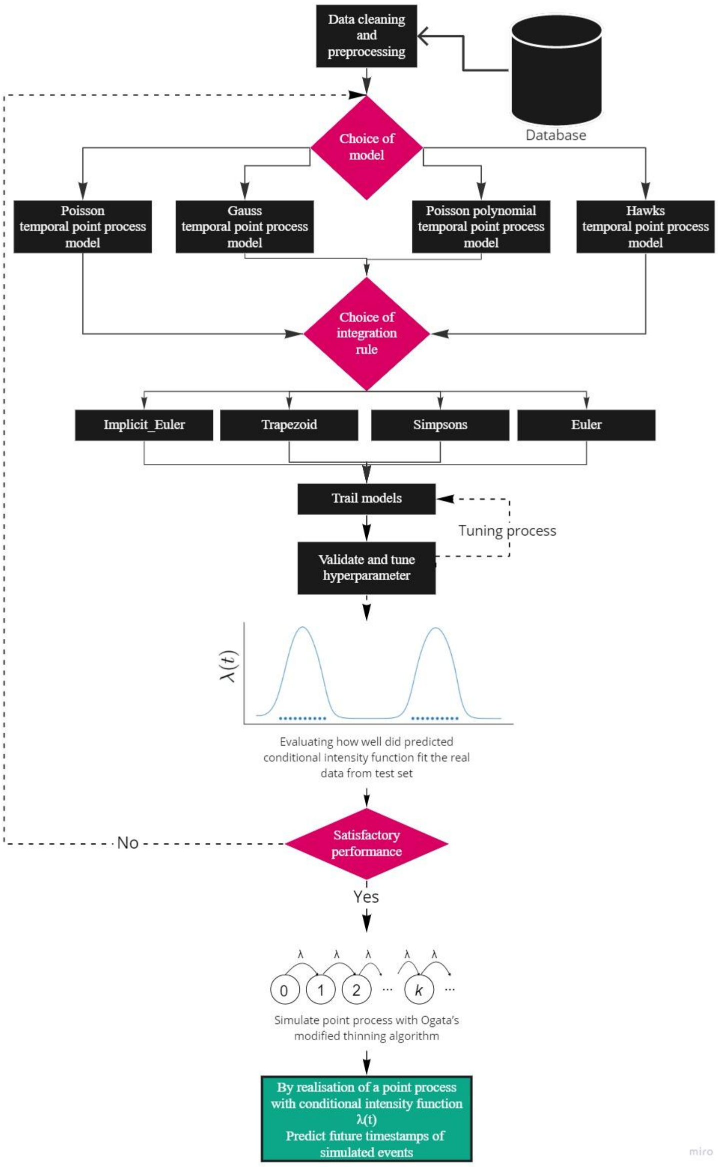

Fig 2. Framework for learning point processes

Data cleaning and preprocessing part consists of the methods that are used for cleaning and transforming raw data prior to processing and analysis. It is an important step that involves reformatting data, making corrections to data and making time sequences of occurred events. The transformation used in this part depends on problem formulations and raw data formats. Additionally, during this part the dataset is split on training, validation, and test set.

The first step in model selection and hyperparameter tuning step is to choose the point process model. In this framework the decision maker must choose between four different kinds of models: The Poisson temporal point process, the Gaussian point process, the Poisson polynomial process and the Hawkes process. The choice of a model primarily depends on the way the events are generated. Therefore, it is advisable to plot approximations of CIF with respect to moving windows and choose the model that best fits to it.

After model selection, the integration step and integration method must be chosen. In this framework, three different kinds of integration rules are presented: Implicit Euler, Trapezoidal, Simpsons and Gaussian quadrature method. Based on the integration rule, the integration step must be finely tuned in order to reduce the approximation error of the integral. Depending on the frequency of events, the integration step must be small and sufficiently large when event frequency is high and low, respectively. After model selection, the model is trained on training set.

The validation of integration step is applied with respect to log likelihood metric on the validation set. If the decision makers are satisfied with the obtained performances of the selected and validated models, they can proceed to the evaluation and simulation phase.

In evaluation and simulation part, the model is simulated and the results of simulations are used for testing the model on test set. Besides, log likelihood metric, mean absolute error is also used to evaluate model performances. Additionally, in combination with simulation, the obtained models can be further used in order to predict occurrence of the following events or to summarize some important statistical measures that can provide useful information to the decision maker.

4. Experiments

4.1. Experimental setup

In this section, we briefly present experimental setup, along with a detailed description of datasets and procedure for training and evaluation.

Datasets. The presented methodology was tested on two different datasets: a high frequency events dataset - traffic prediction on highway toll dataset, and on low frequency events dataset - prediction of ski injuries in ski resort Kopaonik. Therefore, in experiments we provided the methodology performance to learn event generation from two different types of datasets.

In the case of high frequency events dataset, the sequence of cars arriving at the ramp toll on the E 75 highway was taken as a concrete example of interest. Highway European Route E 75 is part of the International E-road network. The observed part connects two large Serbian cities - Belgrade and Niš. More precisely, the goal was to model the process of arrivals on the busiest ramp toll located at Niš from the Belgrade direction. The average time between two passes in one day is about 20 seconds, with a caveat that the time between two passes is highly dependent on the time of the observed day. Standard 70/10/20 train, validation, and test splits were chosen respectively.

As for the low frequency point processes dataset, the observations of ski injuries in ski resort Kopaonik were taken as a concrete events of interest. Ski resort Kopanik is the biggest ski resort in Serbia. The dataset consists of records of ski injuries for the period from 2005 to 2020. Training and validation were done for period prior to 2020, and the test and evaluation were done for the year 2020.

Models. Four different well-known point process models were compared: Poisson temporal point process, Gaussian point process, Poisson, Polynomial point process, and Hawkes process. Additionally, each of these processes was combined with three distinct integration rules: Implicit Euler, Trapezoidal, and Simpsons.

Implementation: All **defined** models were implemented in Pytorch, an optimized tensor library for deep learning using GPUs and CPUs implemented in Python [22]. To run our experiments, we used a PC with the following configuration: Intel i9 CPU 9900K: 16 threads, 3.60GHz, 64 GB DDR4-2133, GPU RTX 3070 GPU 8GB GDDR6. Additionally, we provided the public repository with available implementation of the presented machine learning framework for learning point process.

Training. Due to heterogeneous nature of our benchmark datasets, both in the number of samples and **frequency** of the events, it was observed that we could get better results by fine-tuning the number of epochs (training time) and framework architecture (base model selection, integration rule, integration step) independently for both datasets. Validation dataset was used to choose hyperparameters (integration step, learning rate, etc.). Additionally, early stopping procedure evaluated on validation set was applied for obtaining best generalization performances of trained models.

Optimization. The Adam optimizer was used in order to fit the parameters of point processes. After hyperparameter tuning it was showed that all the models should be trained by 200 epochs, with a constant learning rate of 0.001, and integration step of 30. Each model was trained with the purpose to minimize negative log likelihood in order to reconstruct the true underlying event generation process.

Evaluation metrics. The models were evaluated on two key benchmark tasks. Firstly, we presented how well the models fitted real conditional intensity functions on the test set, or in the other words - how well the models performed minimization of negative log likelihood loss. Moreover, we evaluated Akaike information criteria (AIC) [27] to take in consideration model complexity. Secondly, using Ogata's modified thinning algorithm and conditional intensity function learned during training, we evaluated models prediction performances. The predicted time events were separated into bins of 3 different sizes: 5 minutes, 10 minutes, and 15 minutes for highway toll dataset, and the bins of 5 days, 10 days, and 15 days for ski injuries dataset, respectively. Then, for each binned period the mean absolute error (MAE) was calculated between the number of the predicted events and real number of events in that period. The visualization of sliding window approach for point process performance evaluation is presented in Fig. 3. In the case of high frequency dataset, due to dense event generation process, events were slid by 1 minute ($\sigma(t) = 1$ min), whereas in the case of low frequency dataset events were slid by 1 day period ($\sigma(t) = 1$ day).

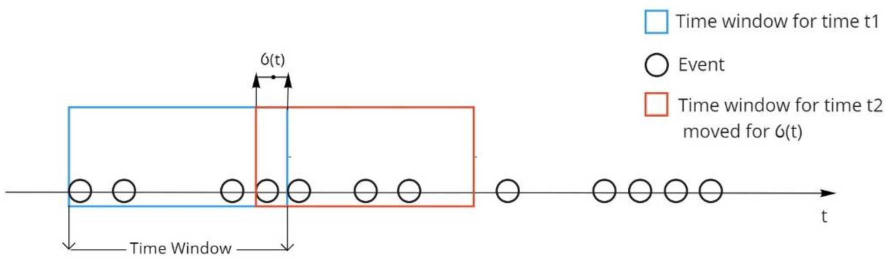

Fig. 3. Sliding window algorithm

Statistical tests. One way to prove that the obtained results are statistically significant is to apply two-sided “Welch” t-test [26]. “Welch” t-test is a two-sample location test used to test the hypothesis that two populations have equal means. Compared to standard Student’s t-test, it is more reliable when two samples have unequal variances. One of the main conditions for applying t-test is samples independence assumption. Bearing in mind that the samples obtained by simulating point process are independent, first the time period is split in bins with fixed size. In each of these bins, the number of occurred events sampled from a point process model is counted and compared to the ground truth number of occurred events (absolute error (AE) is calculated). For each sample, the MAE error is calculated and samples obtained in this manner are completely independent from other samples, hence they can be used as inputs for two-sided “Welch” t-test. If the p-value in two-sided “Welch” t-test are less than 1% threshold, it can be stated that the means of the two groups (in this case MAE or two models) are unequal. If this is true, the results obtained by evaluation metrics are statistically significant.

4.2. Results and discussion

The prediction performances obtained by fitting different types of point process models with three distinct integration methods on highway car arrivals dataset are presented in Table 1.

Hawkes model with trapezoid numerical integration techniques, had the smallest log likelihood loss, AIC and the smallest MAE obtained in the case of all bin sizes. Furthermore, despite small training data and stochastic nature of data generation (i.e., dependency on part of day) it can be concluded that on average, the error of Hawkes model is less than half car per minute compared to the real car arrivals events. In addition, compared to the Hawkes process, Polynomial process obtained the worst results, whereas the results of Poission process are satisfactory, bearing in mind that the conditional intensity function is constant. Moreover, in Table 2, the results of two-sided “Welch” t-test are presented. In all presented models, trapezoid integration rule was used. It can be observed that p-values for each pair of models are less than 1% (0.01) threshold. Therefore, it can be concluded that the prediction means of each pair of models are unequal and results presented in Table 1 are statistically significant.

Table 1. Results of models performances with three distinct integration methods on highway car arrivals dataset

Bin_size	Model	Integration_	MAE	NLL (test_set)	AIC
5	Hawkes	Trapezoid	4.9	112.56	229.12
		Implicit_Euler	5.6	116.03	236.06
		Simpson	5.9	126.78	257.56
10		Trapezoid	8.8	112.56	229.12
		Implicit_Euler	10.3	116.03	236.06
		Simpson	9.7	126.78	257.56
15		Trapezoid	12.3	112.56	229.12
		Implicit_Euler	14.07	116.03	236.06
		Simpson	14	126.78	257.56
5	Gaussian PP	Trapezoid	6	178.85	363.7
		Implicit_Euler	5.9	216.44	438.88
		Simpson	5.7	156.25	318.5
10		Trapezoid	11.5	178.85	363.7
		Implicit_Euler	11.3	216.44	438.88
		Simpson	11.1	156.25	318.5
15		Trapezoid	17.2	178.85	363.7
		Implicit_Euler	16.6	216.44	438.88
		Simpson	15.4	156.25	318.5
5	Polynomial	Trapezoid	30.8	670.65	1347.3
		Implicit_Euler	63.4	756.04	1518.08
		Simpson	364	1206.34	2418.68
10		Trapezoid	61.7	670.65	1347.3
		Implicit_Euler	185.3	756.04	1518.08
		Simpson	729.3	1206.34	2418.68
15		Trapezoid	92.6	670.65	1347.3
		Implicit_Euler	336	756.04	1518.08
		Simpson	1094.3	1206.34	2418.68
5	Poisson	-	9.1	142.12	287
10	Poisson		17.7	142.12	287
15	Poisson		26.6	142.12	287

Table 2. Results of two-sided "Welch" t-test

Bin_size	Model 1	Model 2	t-statistic	p-value
5	Hawkes	Gaussian PP	-11.68	0.002
	Hawkes	Poisson	-78.19	0
	Gaussian PP	Poisson	-66.6	0.001
10	Hawkes	Gaussian PP	-45.04	0.001
	Hawkes	Poisson	-325.95	0
	Gaussian PP	Poisson	-291.207	0
15	Hawkes	Gaussian PP	-52.47	0.001
	Hawkes	Poisson	-347.21	0
	Gaussian PP	Poisson	-303.74	0

The results obtained by Hawkes model on highway car arrivals dataset are visualized in Fig 4. Firstly, it can be observed that the real and predicted conditional intensity functions are plotted with blue and green lines, respectively, by varying length of sliding window (Fig. 4a – 1 min, Fig. 4b – 5 min, Fig. 4c – 10 min). Additionally, the real and simulated timestamps of car arrivals were visualized as red dots. It can be concluded that despite being trained on just 70% of data, the model was pretty successful in predicting the real conditional intensity function. Moreover, the simulated car arrival events can imitate the real world application in the same manner.

In Table 3, the results of models performances on ski injuries dataset are presented. In the same manner, four different point process models performances with respect to three different numerical integration methods are showed. Gaussian point process model with Implicit Euler numerical integration techniques, had the smallest log likelihood loss, AIC, and MAE. The results of two-sided "Welch" t-test are presented in Table 4. The trapezoid integration rule was used in Hawkes model, whereas Implicit Euler rule was used in Gaussian point process. Based on the small p-values, it can be concluded that the prediction means of each pair of models are unequal and these results are presented in Table 1 and are statistically significant.

Again, the Polynomial process obtained the worst results. The Gaussian point process on average achieved MAE of 1.6 in the period of 5 days, which means that the Gaussian point process is going to predict on average 1.6 injuries more or less compared to the real number of injuries. Based on this, it can be emphasized that it is necessary to fit different point process models in order to find the one that best explains the true underlying distribution of event generation.

In addition, the results obtained by Gaussian point process on ski injuries dataset are visualized in Fig 5. Firstly, it can be observed that the real and predicted conditional intensity functions are plotted with blue and green lines, respectively, by varying length of sliding window (Fig. 5a – 5 days, Fig. 5b – 10 days, Fig. 5c – 15 days). Additionally, the real and simulated timestamps of ski injuries were visualized as red dots. It can be observed that real and predicted conditional intensity functions look very similar, and timestamps of the simulated events correspond to the timestamps of the real events.

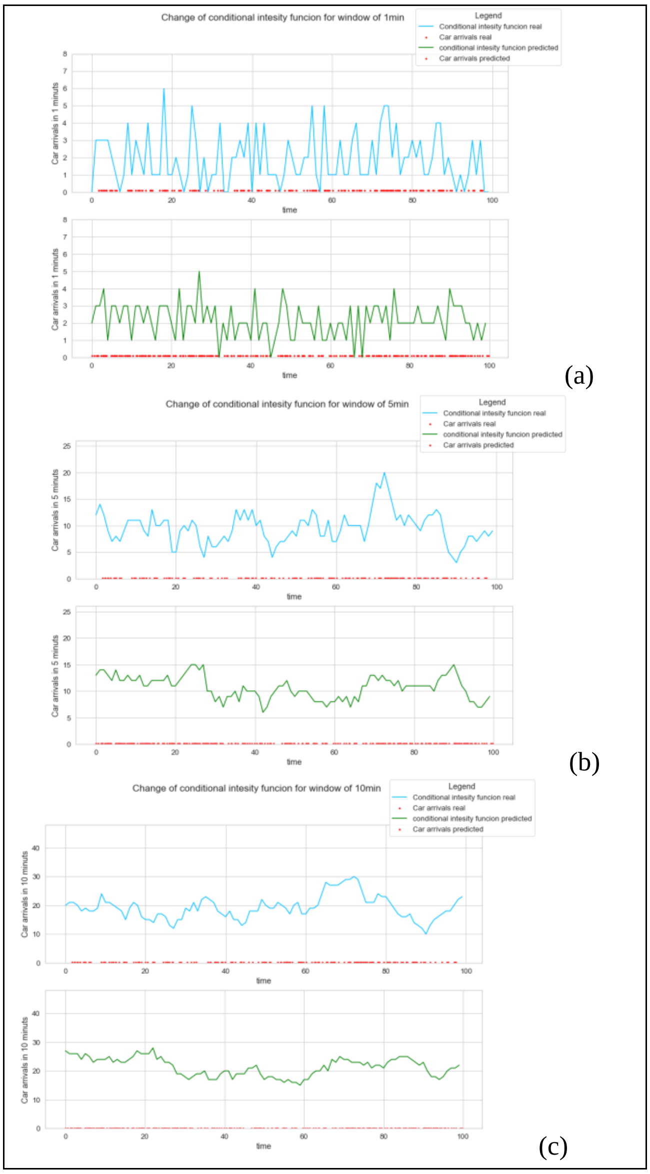

Fig 4. Visualization of how conditional intensity function predicted by Hawk's point process model fitted the real intensity function

Table 3. Experimental results - ski injuries dataset

Bin_size	Model	Integration	MAE	NLL (test_set)	AIC
5	Hawkes	Trapezoid	3.2	76.20	156.4
		Implicit_Euler	3.3	78.81	161.62
		Simpson	4	77.82	159.64
10		Trapezoid	4.8	76.20	156.4
		Implicit_Euler	6.2	78.81	161.62
		Simpson	8	77.82	159.64
15		Trapezoid	7.6	76.20	156.4
		Implicit_Euler	10	78.81	161.62
		Simpson	10	77.82	159.64
5	Gaussian PP	Trapezoid	3.8	99.64	205.28
		Implicit_Euler	1.6	75.42	156.84
		Simpson	2.8	88.56	183.12
10		Trapezoid	5.8	99.64	205.28
		Implicit_Euler	3.8	75.42	156.84
		Simpson	5.4	88.56	183.12
15		Trapezoid	9.6	99.64	205.28
		Implicit_Euler	4.9	75.42	156.84
		Simpson	7.3	88.56	183.12
5	Polynomial	Trapezoid	5.1	68223.36	136452.72
		Implicit_Euler	4.6	22480.22	44966.44
		Simpson	46	4262.03	8530.06
10		Trapezoid	9.8	68223.36	136452.72
		Implicit_Euler	6.6	22480.22	44966.44
		Simpson	102.4	4262.03	8530.06
15		Trapezoid	14	68223.36	136452.72
		Implicit_Euler	8	22480.22	44966.44
		Simpson	134	4262.03	8530.06
5	Poisson	-	3.1	142.00	286
10	Poisson		6.4	142.00	286
15	Poisson		12.3	142.00	286

Table 4. Results of two-sided "Welch" t-test

Bin size	Model 1	Model 2	t-statistic	p-value
5	Hawkes	Gaussian PP	-18.98	0.003
	Hawkes	Poisson	-87.71	0.001
	Gaussian PP	Poisson	-103.91	0
10	Hawkes	Gaussian PP	-15.87	0.002
	Hawkes	Poisson	-70.51	0.001
	Gaussian PP	Poisson	-88.08	0.001
15	Hawkes	Gaussian PP	-39.28	0.002
	Hawkes	Poisson	-103.03	0
	Gaussian PP	Poisson	-153.47	0

5. Conclusion

In this paper, we propose a new machine learning approach methodology for learning temporal point process based on the implementation of one-dimensional numerical integration techniques. The likelihood function of the point process has an integral of the CIF given in the limits of data observation. Bearing in mind that the CIF can take any kind of mathematical form, in many cases this integral is analytically intractable. Due to this, in this paper, we present an approach to linearize this integral with standard numerical techniques and to backpropagate the derivative through it. The presented approach was successfully tested on real-life data. The main disadvantage of this approach lies in high computational cost that is connected with backpropagation of derivative through each integration step. Therefore, this approach should be used only in the cases when point processes with analytically tractable integrals cannot obtain satisfactory prediction performances.

Furthermore, the methodology was evaluated on four different well-known point process models. In addition, we presented that different numerical techniques for integration can be successfully implemented in this framework. Moreover, we successfully simulated the obtained CIFs and compared them with the observed intensity functions.

Further studies should address using deep neural networks (feed-forward and recurrent networks) as a CIF to better capture dependencies between event occurrences.

Acknowledgment. This work was supported in part by the ONR/ONR Global under Grant N62909-19-1-2008. The authors would like to express their gratitude to the company Saga NFG d.o.o. Belgrade, for supporting this research. The authors would also like to thank public enterprises Roads of Serbia and Ski resorts of Serbia for providing data for this research.

References

1. Ryu, B., Steven, B. L. (1998). Point process models for self-similar network traffic, with applications. Communications in statistics. Stochastic models, 14(3), 735-761.
2. Zahrieh, D. (2017). Bayesian point process modeling to quantify excess risk in spatial epidemiology: an analysis of stillbirths with a maternal contextual effect.
3. Liu, S., & Wu, W. (2017). Generalized Mahalanobis depth in point process and its application in neural coding. The Annals of Applied Statistics, 11(2), 992-1010.
4. Farajtabar, M., Wang, Y., Gomez-Rodriguez, M., Li, S., Zha, H., & Song, L. (2017). Coevolve: A joint point process model for information diffusion and network evolution. The Journal of Machine Learning Research, 18(1), 1305-1353.
5. Jacobsen, S., Grove, H., Nedenskov Jensen, K., Sørensen, H. A., Jessen, F., Hollung, K., Søndergaard, I. (2007). Multivariate analysis of 2-DE protein patterns–Practical approaches. Electrophoresis, 28(8), 1289-1299.
6. Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics, 141(2), 876-912.
7. Mei, H., Eisner, J. M. (2017). The neural hawkes process: A neurally self-modulating multivariate point process. In Advances in Neural Information Processing Systems (pp. 6754-6764).
8. Jia, R., Jiang, P., Liu, L., Cui, L., Shi, Y. (2018). Data driven congestion trends prediction of urban transportation. IEEE Internet of Things Journal, 5(2), 581-591.
9. Nguyen, T. T., Krishnakumari, P., Calvert, S. C., Vu, H. L., & Van Lint, H. (2019). Feature extraction and clustering analysis of highway congestion. Transportation Research Part C: Emerging Technologies, 100, 238-258.
10. Rasmussen, J. G. (2011). Temporal point processes: the conditional intensity function. Lecture Notes, Jan.
11. Last, G., Penrose, M. (2017). Lectures on the Poisson process (Vol. 7). Cambridge University Press.
12. Kirchner, M. (2017). An estimation procedure for the Hawkes process. Quantitative Finance, 17(4), 571-595.
13. Xiao, S., Yan, J., Yang, X., Zha, H., & Chu, S. M. (2017, February). Modeling the intensity function of point process via recurrent neural networks. In Thirty-first aaai conference on artificial intelligence.
14. Chen, R. T., Rubanova, Y., Bettencourt, J., Duvenaud, D. K. (2018). Neural ordinary differential equations. In Advances in neural information processing systems (pp. 6571-6583).
15. Zhang, T., Yao, Z., Gholami, A., Gonzalez, J. E., Keutzer, K., Mahoney, M. W., Biros, G. (2019). ANODEV2: A Coupled Neural ODE Framework. In Advances in Neural Information Processing Systems (pp. 5151-5161).
16. Ghahramani, Z. (2003, February). Unsupervised learning. In Summer School on Machine Learning (pp. 72-112). Springer, Berlin, Heidelberg.
17. Mehrasa, N., Jyothi, A. A., Durand, T., He, J., Sigal, L., Mori, G. (2019). A variational auto-encoder model for stochastic point processes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3165-3174).
18. Yuan, B., Wang, X., Ma, J., Zhou, C., Bertozzi, A. L., Yang, H. (2019, September). Variational Autoencoders for Highly Multivariate Spatial Point Processes Intensities. In International Conference on Learning Representations.
19. Xiao, S., Farajtabar, M., Ye, X., Yan, J., Song, L., Zha, H. (2017). Wasserstein learning of deep generative point process models. In Advances in Neural Information Processing Systems (pp. 3247-3257).
20. Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58(1), 83-90.

21. Xu, H., Farajtabar, M., Zha, H. (2016, June). Learning granger causality for hawkes processes. In International conference on machine learning (pp. 1717-1726).
22. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Desmaison, A. (2019). Pytorch: An imperative style, high-performance deep learning library. In Advances in neural information processing systems (pp. 8026-8037).
23. Palm, C. (1988). Intensity variations in telephone traffic. North-Holland.
24. Radunović, D, Numeričke metode (2004). Akademska misao, Beograd.
25. Euler, L. (1768). Institutionum calculi integralis, vol. 1. imp. Acad. imp. Saent.
26. West, R. M. (2021). Best practice in statistics: Use the Welch t-test when testing the difference between two groups. Annals of Clinical Biochemistry,
27. Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. Infectious Disease Modelling, 5, 111-128.
28. Liu, X., Carter, J., Ray, B., Mohler, G. (2021). Point process modeling of drug overdoses with heterogeneous and missing data. The Annals of Applied Statistics, 15(1), 88-101.
29. Fortino, G., Guzzo, A., Ianni, M., Leotta, F., Mecella, M. (2021). Predicting activities of daily living via temporal point processes: Approaches and experimental results. Computers & Electrical Engineering, 107567.
30. Zhu, S., Ding, R., Zhang, M., Van Hentenryck, P., Xie, Y. (2021). Spatio-temporal point processes with attention for traffic congestion event modeling. IEEE Transactions on Intelligent Transportation Systems.
31. Motagi, Samarth, et al. Point-Process modeling of Secondary Crashes. 2021.
32. Saha, A., Ganguly, N., Chakraborty, S., & De, A. (2019, April). Learning network traffic dynamics using temporal point process. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications (pp. 1927-1935). IEEE.
33. Ganguly, N., & Saha, A. (2021). Modeling Inter-process Dynamics in Competitive Temporal Point Processes. Journal of the Indian Institute of Science, 101(3), 455-484.
34. D'Angelo, N., Adelfio, G., & Jorge, M. (2021, April). Some properties and applications of local second-order characteristics for spatio-temporal point processes on networks. In GRASPA 2021.
35. Tang, X., Li, L. (2021). Multivariate temporal point process regression. Journal of the American Statistical Association, 1-16.
36. González, J. A., Rodríguez-Cortés, F. J., Cronie, O., Mateu, J. (2016). Spatio-temporal point process statistics: a review. Spatial Statistics, 18, 505-544.
37. Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. Statistical Science, 33(3), 299-318.

Andrija Petrović is an assistant professor at the Technical faculty, Singidunum University. He received his first PhD in Decision making at the Faculty of Organizational Sciences, University of Belgrade and second PhD in Process engineering Faculty of Mechanical engineering, University of Belgrade. His research interests lie in machine learning, fairness in AI, and applications of machine learning in Chemical engineering.

Aleksa Biserčić is a PhD student at the University of Belgrade, Faculty of Organizational Sciences, with focus on Data Science and Deep learning. He has received a bachelor's degree in mechanical engineering and master degree in industrial engineering from University of Belgrade. His current work involves working as Data Scientist with US based software company Jaggaer and part time internship in laboratory

at University of Belgrade. His interests are in time series forecasting, natural language processing and recommendation systems.

Boris Delibašić is full professor at the University of Belgrade - Faculty of Organizational Sciences, Republic of Serbia. His research interests lie in data science, machine learning, business intelligence, multicriteria decision analysis, and decision support systems.

Dimitrije Milenković has received the B.S. degree in information systems and technologies in 2019 and the M.S. degree in information systems in 2020, both from the University of Belgrade. He worked on projects related to building analytical data platforms and data-driven products at TX Group (2019-2021) and Netcast (2017-2018). His research interests are data engineering, machine learning, decision-support systems, integration and interoperability of enterprise services and applications. Starting from 2020, he has been involved in the research projects of the Center for Business Decision Making at the Faculty of Organizational Sciences. Starting from 2021, he is a Research Associate in the Process Engineering Group of the National Institute of Standards and Technology and actively contributing to the development of Score, an open-source project for data exchange standards life cycle management.

Received: September 06, 2021; Accepted: February 16, 2022.