A POISSON PROCESS MODEL FOR ACTIVITY FORECASTING

Tahmida Mahmud[⋆] *Mahmudul Hasan*[⋆] *Anirban Chakraborty*[†] *Amit K. Roy-Chowdhury*[⋆]

[⋆] University of California, Riverside
[†] Nanyang Technological University, Singapore

ABSTRACT

Activity forecasting has recently become an active research area for its importance in critical applications like automated navigation and human-computer interaction. However, for a video observed upto a certain time, all of the existing forecasting works focus on predicting the activity label, i.e., predicting what the next unobserved activity is. To the best of our knowledge, no work has answered the crucial question yet: *when the next unobserved activity will occur*. In this paper, we propose an approach for predicting the starting time of the next unobserved activity without assuming that we know its label. We model activities occurring at a variable rate using a Log-Gaussian Cox Process (LGCP) and learn the rate function from the training data. Then the starting time is predicted using importance sampling algorithm. In our experiments on the challenging MPII-Cooking dataset, we find that both the label of the last observed activity and the label of the activity being predicted affect the time prediction accuracy.

Index Terms— Activity forecasting, Inhomogeneous Poisson Process (IPP), importance sampling.

1. INTRODUCTION

Activity forecasting has a wide range of applications in intelligent video surveillance, robot vision, human-computer interaction, game control, etc. [1]. The problem of activity forecasting can be split into two parts: forecasting the label of the next unobserved activity and forecasting the starting time of that activity. Although there have been few works on activity label prediction, to the best of our knowledge, forecasting the starting time is a novel problem which has not been explored in the video analysis community yet.

For a video which has been observed upto a particular time, our paper aims to determine exactly when an unobserved activity will occur irrespective of its label. Predicting the starting time of the next unobserved activity is crucial for applications like automated navigation, effective surveillance systems and more natural human-computer interfaces. For example, in case of anomaly detection in videos [2], if the starting time of the next activity does not match the predicted time, an anomaly may be detected. For video completion [3, 4], the occurrence time of the activity in the previous frame and the next frame can help to infer about the activities in the missing frame and reconstruct accordingly. In case of saliency detection in videos [5, 6], the predicted time can help determine the most salient location in the video. For autonomous navigation [7], activity time forecasting can help to decide how to maneuver depending on the next predicted activity at the predicted time. In active sensing [8], robots can make decision about when to perform the next action by knowing the timing of the next activity in that region.

In order to predict the starting time of an unobserved activity, a model is required that represents the inter-activity time and the rate at which activities are happening. We propose to model the inter-activity time between activities using Log-Gaussian Cox Process (LGCP) [9]. Our method works without any knowledge of the label of the unobserved activity. This is critical because it will allow the system to anticipate when something will happen, even if what will happen is not known exactly. Moreover, knowledge of the starting time of the next unobserved activity can help to determine its label.

Prior Work: Few works have shown significant accuracy in forecasting the label of the activity well before they are observed such as approaches using semantic scene labeling [10], Probabilistic Suffix tree (PST) [11], augmented- Hidden Conditional Random Field (a-HCRF) [12], Markov Random Field (MRF) [13], kernel-based reinforcement learning [14], and max-margin learning [15]. However, to the best of our knowledge, no work on activity forecasting has been able to predict the starting time of the unobserved activity with or without any information about its label. Thus, forecasting the starting time is a novel problem in the video understanding community. There are a few relevant works in other fields like modeling tweet arrival time [16] and modeling conflict dynamics in war [17]. We leverage upon the ideas presented in these papers but adapt our solution approach for learning the model parameters based on our problem in video analysis.

Main Contribution and Overview of Our Approach: The contribution of our work is a novel framework for predicting the starting time of the next unobserved activity in a video irrespective of its label. We leverage upon a Poisson process for modeling the inter-activity time. Because of the bursty nature of activities in most of the video datasets, their inter-arrival times generally follow an exponential distribution. It is therefore justified to consider them as a part of a Poisson process since the distribution of inter-arrival time for a Poisson

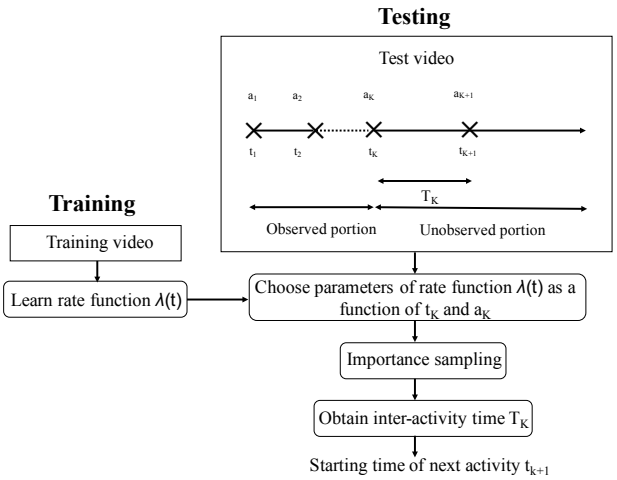

Fig. 1. Overview of our approach. In the test video, the K^{th} activity, a_K, occurred at time, t_K, and the occurrence time of the next unobserved activity, a_{K+1}, is $t_{K+1} = t_K + T_K$. The cross sign represents the occurrence of activities.

process is exponential. The rate parameter associated with such process is the expected number of occurrences per unit of time. As the activities usually occur at different rates at different periods in the videos, we use a special case of Inhomogeneous Poisson Process (IPP) [18], the Log-Gaussian Cox Process (LGCP) for modeling these activities where the rate parameter itself is a function of time. The choice of this model is justified in more details later. We learn the parameters of the rate function in the training phase and then predict the starting time of the next unobserved activity in a video through importance sampling using the starting time and the label of the last observed activity. Detailed overview of our proposed framework is illustrated in Fig. 1.

2. PROPOSED APPROACH

Problem Statement: If a video is observed upto time t with K occurrences of activities, $A = \{a_i\}_{i=1}^{K}$, the K^{th} activity, a_K occurred at time t_K and the occurrence time of the next unobserved activity, a_{K+1}, is $t_K + T_K$, then we want to predict this inter-activity time T_K as shown in Fig. 1. We start by justifying the use of a Poisson process model, then learn the parameters of this model in the training phase, and finally predict the starting time using an importance sampling algorithm.

Motivation for Poisson Process Assumption: A Poisson process is a stochastic process counting the number of events in a given time interval. If the rate at which these events occur is λ, then the time between each pair of consecutive events has an exponential distribution with parameter λ. For a homogeneous Poisson process, the rate parameter λ is the expected number of events which occur per unit time. For a homogeneous Poisson process $N(t)$ with rate

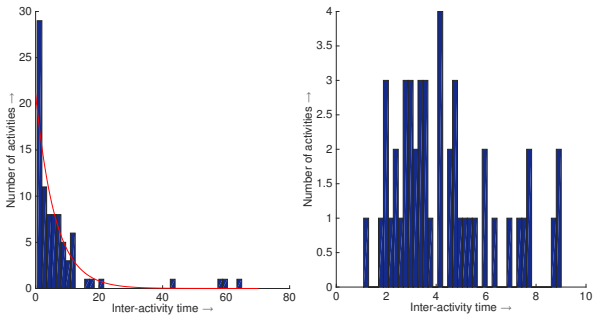

Fig. 2. Histograms of the inter-activity time for all the activities in videoclip 25 from MPII-Cooking dataset [19] (left) and histograms of the average inter-activity time of different types of activities from all the videos of MPII-Cooking dataset [19] (right).

parameter λ, the waiting time until the first arrival is more than t if and only if there is no arrival before time t. So, $p(T_1 > t) = p(N(t) = 0) = e^{-\lambda t}$ and $p(T_1 \leq t) = p(N(t) > 0) = 1 - e^{-\lambda t}$. Here, $1 - e^{-\lambda t}$ is the cumulative distribution function (CDF) for exponential distribution. We verified it numerically that all the videos from our dataset have an exponential inter-activity time. So, we chose Poisson process for modeling the inter-activity time. The distribution of inter-activity times in a video from our dataset is shown in Fig. 2 (left). An Inhomogeneous Poisson Process counts events which occur at a variable rate and the generalized rate function is given as $\lambda(t)$. The histogram of the inter-activity times for different types of activities is shown in Fig. 2 (right). Since the activities occur at a variable rate, we assume an Inhomogeneous Poisson Process in our case where the rate parameter varies with time.

Learning Rate Function $\lambda(\mathbf{t})$: For an IPP with rate function $\lambda(t)$, the expected number of events in the time interval $[t_1, t_2]$ is $N_{t_1,t_2} = \int_{t_1}^{t_2} \lambda(t)dt$ and the probability that exactly x activities occur in the time interval $[t_1, t_2]$ is,

$$p(N(t_2) - N(t_1) = x) = \frac{\left(\int_{t_1}^{t_2} \lambda(t)dt\right)^x \exp\left(-\int_{t_1}^{t_2} \lambda(t)dt\right)}{x!} \tag{1}$$

As mentioned earlier, the K^{th} activity, a_K occurred at time t_K and the occurrence time of the next unobserved activity a_{K+1}, is $t_K + T_K$. The probability that another activity occurs by time $t_K + \tau$ which is the cumulative distribution for T_K is given by,

$$\begin{aligned} p(T_K \leq \tau) &= 1 - p(T_K > \tau) \\ &= 1 - p(\text{No activity occurred in } [t_K, t_K + \tau]) \\ &= 1 - \exp\left(-\int_{t_K}^{t_K+\tau} \lambda(t)dt\right) \\ &= 1 - \exp\left(-\int_{0}^{\tau} \lambda(t_K + t)dt\right) \end{aligned} \tag{2}$$

By taking the derivative of Eqn. (2), we obtain the probability density function of T_K as

$$p(T_K = \tau) = \lambda(t_K + \tau) \exp\left(-\int_0^{\tau} \lambda(t_K + t)dt\right) \quad (3)$$

We assume $\lambda(t)$ to be constant in an interval to get rid of the integration complexity [9,20] and the approximate density of T_K would be

$$p(T_K = \tau) = \lambda(t_K + \tau) \exp\left(-\tau\lambda\left(t_K + \frac{\tau}{2}\right)\right) \quad (4)$$

We propose to use Log-Gaussian Cox Process (LGCP) where the rate parameter $\lambda(t)$ has a non-parametric form and is considered to be a function of time, so the model complexity depends only on the data. In LGCP, $\lambda(t)$ is assumed to be stochastic. We assume the occurrences of all the activities as a single Poisson process and associate a specific intentisy function $\lambda(t) = \exp(f(t))$ with it. As it is a Log-Gaussian Cox process, where taking the logarithm of the rate function should yield a Gaussian distribution, $f(t)$ is defined through a Gaussian Process (GP) prior [21]. So, finally the approximate density of T_K becomes

$$\begin{aligned} p(T_K = \tau) &= \exp(f(t_K + \tau)) \exp\left(-\tau \exp\left(f\left(t_K + \frac{\tau}{2}\right)\right)\right) \\ &= \exp\left(\frac{\exp(-(t_K + \tau - \mu)^2/2\sigma^2)}{\sqrt{(2\pi\sigma^2)}}\right) \\ &\quad \exp\left(-\tau \exp\left(\frac{\exp(-(t_K + \frac{\tau}{2} - \mu)^2/2\sigma^2)}{\sqrt{(2\pi\sigma^2)}}\right)\right) \end{aligned} \quad (5)$$

From the training videos, we learn the inter-activity times for all the activity labels irrespective of the next activities and compute their mean and variance. Then for predicting the starting time of the next unobserved activity, given the label of the last observed activity, we choose the parameters μ and σ from the mean and variance computed from the training data based on the label of that activity.

Prediction of Starting Time using Importance Sampling: For predicting the starting time of the next unobserved activity, we define our nominal distribution $p(T_K = \tau)$. As we know the label of the last observed activity, we choose the parameters μ and σ of this distribution as the mean and variance of the inter-activity time for that activity label from the training data. We also have the observed starting time, t_K of the last observed activity, a_K. The inter-activity time T_K is then obtained through importance sampling algorithm with an exponential proposal distribution according to [22]. The starting time of the next unobserved activity, a_K is then found as $t_{K+1} = t_K + T_K$. The entire method is described in Algorithm 1.

Algorithm 1 Starting Time Prediction

1: **Input** Starting time of last observed activity t_K, label of last observed activity, proposal distribution $q(\tau)$, number of samples N
2: Learn rate function $\lambda(\tau)$ from training based on the label of the last observed activity.
3: **for** $n = 1$ **to** N **do**
4: Sample $t_n \sim q(\tau)$
5: Obtain weights $w_n = \frac{p(t_n)}{q(t_n)}$
6: **end for**
7: Evaluate inter-activity time as $T_K = \frac{\sum_{n=1}^{N} t_n w_n}{\sum_{m=1}^{N} w_m}$
8: Predict the next activity starting time as $t_{K+1} = t_K + T_K$
9: **return** t_{K+1}

3. EXPERIMENTAL RESULTS

The objective of the experiments is to analyze the performance of our framework for predicting the starting time of future activities. We conduct our experiments on MPII-Cooking dataset [19]. Activity numbers we use are same as the original dataset. We have excluded 17 types of activities $(1, 8, 11, 12, 18, 24, 28, 30, 37, 43\text{-}46, 53, 61, 63, 64)$ from our experiments. These activities have very low sample points in the video (occurred infrequently) and the standard deviations of their durations are much higher compared to other activities. Hence, it is not possible to learn a reliable model for these activities. We worked with 44 videos having 48 types of activities and the training-testing videoclip ratio is $15 : 7$. Five random combinations of training and testing videos are chosen, and results are averaged over these combinations.

Prediction Error: We analyze the prediction error in two ways: as a function of the last observed activity and as a function of the activity being predicted. For the 48 types of activities we used in the dataset, the Root-Mean-Square Error (RMSE) values of the predicted starting time based on the label of the last observed activity and the label of the activity being predicted are shown in Fig. 3 and Fig. 4 respectively. For the first case, the label of the last observed activity is fixed and we predict the starting time of the next activity irrespective of its label. For the latter case, the label of the activity being predicted is fixed and we predict its starting time irrespective of the label of the last observed activity. We observe that activities with fewer examples in the testing set produce higher RMSE for prediction. We find an average RMSE of 3.6489 seconds for the predicted starting time based on the label of the last observed activity and an average RMSE of 3.6785 seconds based on the label of the activity being predicted. The normalized RMSE of the predicted starting times for all the possible activity sequnces in the videos are shown in Fig. 5. As can be seen from the figure, in most cases the RMSE values are small. The activity sequences which occurred fewer times produce higher RMSE as it is difficult to learn a reliable model for those sequences. The actual RMSE

values are provided in the supplementary material. We obtain an overall RMSE of 3.9431 seconds for all of our predicted starting times.

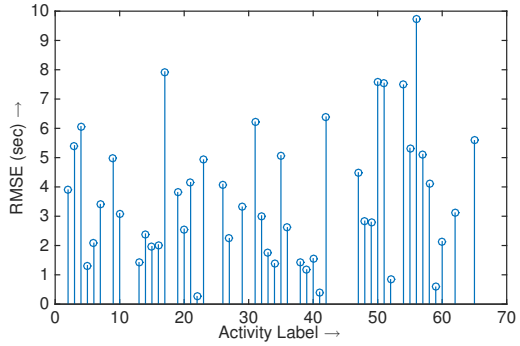

Fig. 3. RMSE of the predicted starting time based on the label of the last observed activity. The gaps in the figure are due to the activities we removed in the beginning.

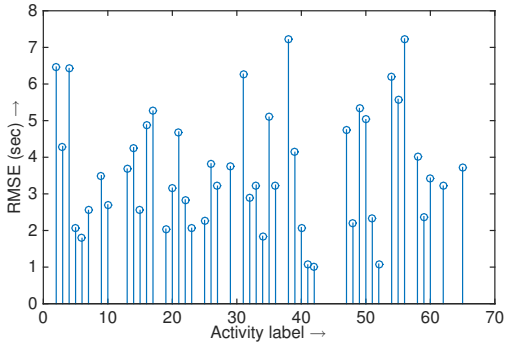

Fig. 4. RMSE of the predicted starting time based on the label of the activity being predicted. The gaps in the figure are due to the activities we removed in the beginning.

Increasing the Prediction Horizon: If we do a multi-step prediction, i.e., increase the forecasting horizon then the RMSE value for prediction increases. For one step prediction, we use the observed starting time of the last activity. But, in case of multi-step prediction, we use the predicted value of the starting time of the last activity instead of their observed starting time and so the error accumulates. This gradual increase in RMSE for predicted starting time is shown in Fig. 6 upto a forecasting horizon of 5 activities.

Comparison with a Baseline Model: We Use a Homogeneous Poisson process (HPP) as a baseline method which has an exponential distribution for the inter-activity times. The rate parameter is constant and is the reciprocal of the average inter-arrival times for all the activities in the training data. Using this baseline model, we obtain an overall RMSE value of 4.7972 seconds for predicted starting times compared to 3.9431 seconds for our Inhomogeneous Poisson Process (IPP) leading to a 21.66% increase in error.

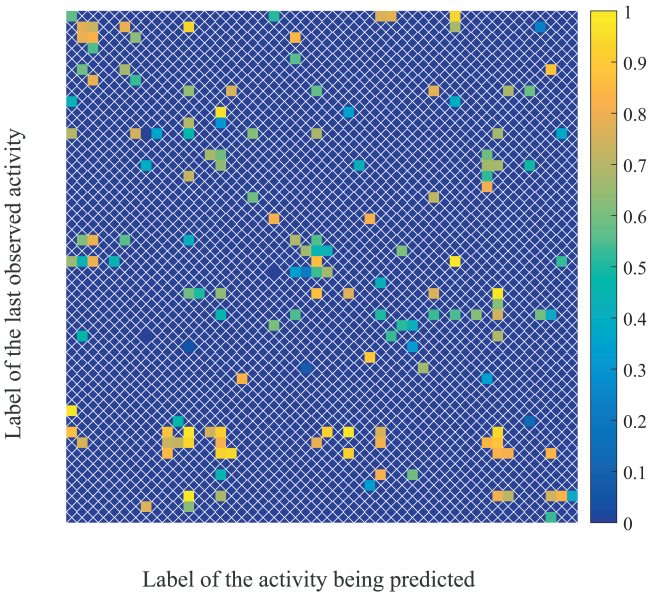

Fig. 5. A graphical representation of the matrix where the rows represent the last observed activity, the columns represent the next predicted activity and each entry represents the corresponding normalized RMSE for starting time prediction. The white cross markers represent the activity sequences which never occurred. This is best viewed in color.

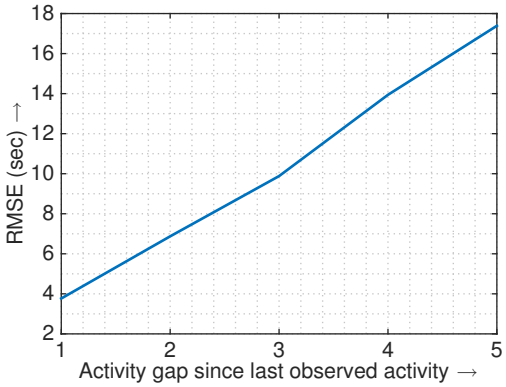

Fig. 6. RMSE of the predicted starting time for multi-step prediction with increasing forecasting horizon.

4. CONCLUSION

We predicted the starting time of an unobserved activity by modeling all the activities using Log-Gaussian Cox Process (LGCP). The method does not require any knowledge about the label of the unobserved activities. In future, we intend to use this starting time information to improve the accuracy of the label forecasting algorithms, and study the accuracy of the forecasting methodology in various applications.

Acknowledgment: The work was partially supported through a research grant from the US Dept. of Defense.

5. REFERENCES

[1] K. Cheng, Y. Chen, and W. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," in *CVPR*, 2015, pp. 2909–2917.

[2] Y. Zhu, N. M. Nayak, and A. K. Roy-Chowdhury, "Context-aware activity recognition and anomaly detection in video," *IEEE Journal of Selected Topics in Signal Processing.*, vol. 7, no. 1, pp. 91–101, 2013.

[3] A. Mosleh, N. Bouguila, and A. Ben Hamza, "Video completion using bandlet transform," *IEEE Transactions on Multimedia*, vol. 14, no. 6, pp. 1591–1601, 2012.

[4] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez, "Video inpainting of complex scenes," *SIAM Journal on Imaging Sciences.*, vol. 7, no. 4, pp. 1993–2019, 2014.

[5] D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in *CVPR*, 2013, pp. 1147–1154.

[6] L. Junling, F. Meng, and J. Mao, "Saliency detection on videos with scene change," in *ICALIP*, 2014, pp. 506–510.

[7] V. Sezer, T. Bandyopadhyay, D. Rus, E. Frazzoli, and D. Hsu, "Towards autonomous navigation of unsignalized intersections under uncertainty of human driver intent," in *IROS*, 2015, pp. 3578–3585.

[8] M. Herbert, "Active and passive range sensing for robotics," in *ICRA*, 2000, vol. 1, pp. 102–110.

[9] J. Møller, A. R. Syversveen, and R. P. Waagepetersen, "Log Gaussian Cox processes," *Scandinavian Journal of Statistics*, vol. 25, no. 3, pp. 451–482, 1998.

[10] K. M. Kitani, B. D. Zeibart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *ECCV*, 2012, pp. 201–214.

[11] K. Li and Y. Fu, "Prediction of human activity by discovering temporal sequence patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1644–1657, 2014.

[12] X. Wei, P. Lucey, S. Vidas, S. Morgan, and S. Sridharan, "Forecasting events using an augmented hidden conditional random field," in *ACCV*, 2014, pp. 569–582.

[13] A. Chakraborty and A. K. Roy-Chowdhury, "Context-aware activity forecasting," in *ACCV*, 2014, pp. 21–36.

[14] D. Huang and K. M. Kitani, "Action-reaction: Forecasting the dynamics of human interaction," in *ECCV*, 2014, pp. 489–504.

[15] T. Lan, T. Chen, and S. Savarese, "A hierarchical representation for future action prediction," in *ECCV*, 2014, pp. 689–704.

[16] M. Lukasik, P. K. Srijith, T. Cohn, and K. Bontcheva, "Modeling tweet arrival times using log-Gaussian Cox processes," in *Empirical Methods in Natural Language Processing(EMNLP), Proceedings of the 2015 Conference on*, 2015, pp. 250–255.

[17] A. Zammit-Mangion, M. Dewar, V. Kadirkamanathan, and G Sanguinetti, "Point process modelling of the afghan war diary," *Proceedings of the National Academy of Sciences*, vol. 109, no. 31, pp. 12414–12419, 2012.

[18] S. Lee, J. R. Wilson, and M. M. Crawford, "Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior," *Communications in Statistics-Simulation and Computation*, vol. 20, no. 2-3, pp. 777–809, 1991.

[19] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, "A database for fine grained activity detection of cooking activities," in *CVPR*, 2012, pp. 1194–1201.

[20] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari, "Gpstuff: Bayesian modeling with Gaussian processes," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1175–1179, 2013.

[21] C. K. Williams and C. E. Rasmussen, "Gaussian processes for machine learning," *the MIT Press*, vol. 2, no. 3, pp. 4, 2006.

[22] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, "Bayesian data analysis," vol. 2, 2014.