Forest-Based Point Process for Event Prediction from Electronic Health Records

Jeremy C. Weiss and David Page

University of Wisconsin-Madison, Madison, WI, USA
jcweiss@cs.wisc.edu, page@biostat.wisc.edu

Abstract. Accurate prediction of future onset of disease from Electronic Health Records (EHRs) has important clinical and economic implications. In this domain the arrival of data comes at semi-irregular intervals and makes the prediction task challenging. We propose a method called multiplicative-forest point processes (MFPPs) that learns the rate of future events based on an event history. MFPPs join previous theory in multiplicative forest continuous-time Bayesian networks and piecewise-continuous conditional intensity models. We analyze the advantages of using MFPPs over previous methods and show that on synthetic and real EHR forecasting of heart attacks, MFPPs outperform earlier methods and augment off-the-shelf machine learning algorithms.

1 Introduction

Ballooning medical costs and an aging population are forcing governments and health organizations to critically examine ways of providing improved care while meeting budgetary constraints. A leading candidate to fulfill this mandate is the advancement of personalized medicine, the field surrounding the customization of healthcare to individuals. Predictive models for future onset of disease are the tools of choice here, though the application of existing models to existing data has had mixed results.

The research into improvements in predictive modeling has manifested in two main areas: better data and better models. Electronic health records (EHRs) now provide rich medical histories on individuals including diagnoses, medications, procedures, family history, genetic information, and so on. The individual may have regular check-ups interspersed with hospitalizations and medical emergencies, and the sequences of semi-irregular events can be considered as timelines.

Unlike timelines, the majority of models incorporating time use a time-series data representation. In these models data are assumed to arrive at regular intervals. Irregular arrivals of events violate this assumption and lead to missing data and/or aggregation, resulting in a loss of information. Experimentally, such methods have been shown to underperform analogous continuous-time models [1].

To address the irregularity of medical event arrivals, we develop a continuous-time model: multiplicative-forest point processes (MFPPs). MFPPs model the

rate of event occurrences and assume that they are dependent on an event history in a piecewise-constant manner. For example, the event of aspirin consumption (or lack thereof) may affect the rate of myocardial infarction, or heart attack, which in turn affects the rate of thrombolytic therapy administration. Our goal is to learn a model that identifies such associations from data.

MFPPs build on previous work in piecewise-constant conditional intensity models (PCIMs) using ideas from multiplicative-forest continuous-time Bayesian networks (mfCTBNs) [2,3]. MFPPs extends the regression tree structure of PCIMs to regression forests. Unlike most forest learning algorithms, which minimize a classification loss through function gradient ascent or ensembling, MFPPs are based on a multiplicative-forest technique developed in CTBNs. Here, a multiplicative assumption for combining regression tree values leads to optimal marginal log likelihood updates with changes in forest structure. The multiplicative representation allows MFPPs to concisely represent composite rates, yet also to have the flexibility to model rates with complicated dependencies. As the multiplicative forest model leads to representational and computational gains in mfCTBNs, we show that similar gains can be achieved in the point process domain. We conduct experiments to test two main hypotheses. First, we test for improvements in learning MFPPs over PCIMs, validating the usefulness of the multiplicative-forest concept. Second, we assess the ability of MFPPs to classify individuals for myocardial infarction from EHR data, compared to PCIMs and off-the-shelf machine learning algorithms.

Specifically we address two modeling scenarios for forecasting: *ex ante* (meaning "from the past") forecasting and supervised forecasting. An *ex ante* forecast is the traditional type of forecasting and occurs if no labels are available in the forecast region. An example of *ex ante* forecasting is the prediction of future disease onset from the present day forwards. Acquiring labels from the future is not possible, and labels from the past may introduce bias through a cohort effect. However, in some cases, labels may be used, and we call such forecasts "supervised". An example of supervised forecasting is the retrospective cohort study to predict the class of unlabeled examples as well as to identify risk factors leading to disease. The application of continuous-time models to the forecasting case is straightforward. When labels are available, however, we choose to apply MFPPs in a cascade learning framework, where the MFPP predictions contribute as features to supervised learning models.

In Section 2, we discuss point processes and contrast them from continuous-time Bayesian networks (CTBNs) noting their matching likelihood formulations given somewhat different problem setups. We show that multiplicative forest methods can be extended to point processes. We also introduce the problem of predicting myocardial infarction, discuss the various approaches to answering medical queries, and introduce our method of analysis. In Section 3, we present results on synthetic timelines and real health records data and show that MFPPs outperform PCIMs on these tasks, and that the timeline analysis approach outperforms other standard machine learning approaches to the problem. We conclude in Section 4.

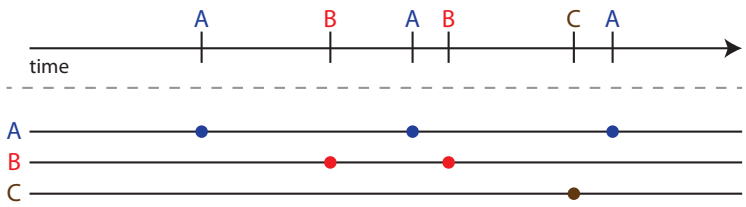

Fig. 1. A timeline (top) deconstructed into point processes (bottom).

2 Point Processes

Data that arrive at irregular intervals are aptly modeled with timelines. A timeline is a sequence of {event,time} pairs capturing the relative frequency and ordering of events. This representation arises in many domains, including neuron spike trains [4], high-frequency trading [5], and medical forecasting [6]. We describe and build upon one such model: the point process.

A point process treats each event type individually and specifies that it (re-)occurs according to the intensity (or rate) function $\lambda(t|h)$ over time t given an event history h. Figure 1 shows a sample timeline of events deconstructed into individual point processes. The conditional intensity model (CIM) is a probabilistic model formed by the composition of such processes. Our work will build on piecewise-constant conditional intensity models (PCIMs), which make the assumption that the intensity functions $\lambda(t|h)$ are constant over positive-length intervals. PCIMs represent the piecewise-constant conditional intensity functions with regression trees, and one is shown in Figure 2 (left).

The piecewise-constant intensity assumption is convenient for several reasons. For one, the likelihood can be computed in closed form. We can also compute the sufficient statistics by counting events and computing a weighted sum of constant-intensity durations. With these, we can directly estimate the maximum likelihood model parameters. Finally, we note that with this assumption the likelihood formulation becomes identical to the one used in continuous-time Bayesian networks (CTBNs). The shared likelihood formula lets us apply a recent advance in learning CTBNs: the use of multiplicative forests. Multiplicative forests produce intensities by taking the product of the regression values in active leaves. For example, a multiplicative forest equivalent to the tree described above is shown in Figure 2 (right). These models were shown to have large empirical gains for parameter and structure learning similar to those seen in the transition from tree models to random forests or boosted trees [3]. Our first goal is to show that a similar learning framework can be applied to point processes. We describe the model in fuller detail below.

2.1 Piecewise-Continuous Conditional Intensity Models (PCIMs)

Let us consider the finite set of event types $l \in \mathcal{L}$. An event sequence or trajectory x is an ordered set of {time, event} pairs $(t, l)_{i=1}^{n}$. A history h at time t is the

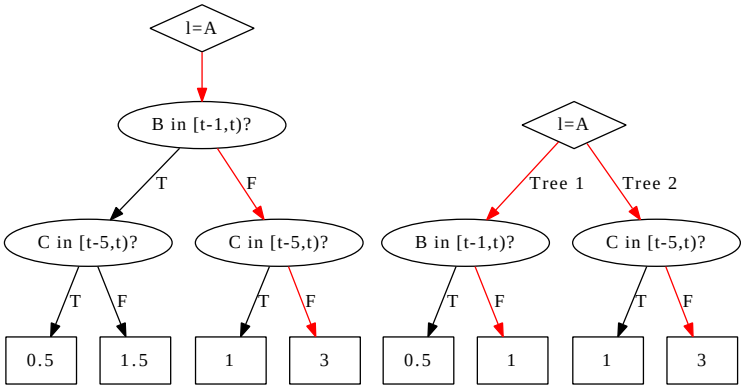

Fig. 2. A piecewise-constant conditional intensity tree for determining the rate of event type A (left). An equivalent multiplicative intensity forest (right). An example of active paths are shown in red. The active path in the tree corresponds to the intersection of active paths in the forest, and the output intensity is the same ($3 = 1 \times 3$).

subset of x whose times are less than t. Let l_0 denote the null event type, and use the null event pairs (l_0, t_0) and (l_0, t_{end}) to denote the start and end times of the trajectory. Then the likelihood of the trajectory given the CIM θ is:

$$p(x|\theta) = \prod_{l \in \mathcal{L}} \prod_{i=1}^{n} \lambda_l(t_i|h_i, \theta)^{\mathbb{1}(l=l_i)} e^{\int_{-\infty}^{t} \lambda_l(\tau|x,\theta)d\tau}$$

PCIMs introduce the assumption that the intensity functions are constant over intervals. As described in [2], let Σ_l be a set of discrete states so that we obtain the set of parameters λ_{ls} for $s \in \Sigma_l$. The active state s is determined by a mapping $\sigma_l(t, x)$ from time and trajectory to s. Let S_l hold the pair $(\Sigma_l, \sigma_l(t, x))$ and let $S = \{S_l\}_{l \in \mathcal{L}}$. Then the PCIM likelihood simplifies to:

$$p(x|S, \theta) = \prod_{l \in \mathcal{L}} \prod_{s \in \Sigma_l} \lambda_{ls}^{M_{ls}(x)} e^{-\lambda_{ls} T_{ls}(x)} \tag{1}$$

where $M_{ls}(x)$ is the count of events of type l while s is active in trajectory x, and $T_{ls}(x)$ is the total duration that s, for event type l, is active.

2.2 Continuous-Time Bayesian Networks (CTBNs)

Continuous-time Bayesian networks model a set of discrete random variables $x_1, x_2, \ldots, x_d = \mathcal{X}$ over continuous time, each with s_i number of discrete states for i in $\{1, \ldots, d\}$ [7]. CTBNs make the assumption that the probability of transition for variable x out of state x^j at time t is given by the exponential distribution $\lambda_{x^j|u} e^{-\lambda_{x^j|u} t}$ with rate parameter (intensity) $\lambda_{x^j|u}$ given parents setting u. The variable transitions from state x^j to x^k with probability $\Theta_{x^j x^k|u}$, where $\Theta_{x^j x^k|u}$

is an entry in state transition matrix Θ_u. The parents setting u is an element of the joint state U_x over parent variables of x, and the parent dependencies are provided in a directed possibly-cyclic graph. A complete CTBN model can be described by two components: a distribution $\mathcal{B}$ over the initial joint state, typically represented by a Bayesian network, and a directed graph over variables $\mathcal{X}$ with corresponding conditional intensity matrices (CIMs). The CIMs hold the intensities $\lambda_{x^j|u}$ and state transition probability matrices Θ_u.

The CTBN likelihood is defined as follows. A trajectory, or a timeline, is broken down into independent intervals of fixed state. For each interval $[t_0, t_{\text{end}})$, the duration $t = t_{\text{end}} - t_0$ passes and a variable x transitions at t_{end} from state x^j to x^k. All other variables $x_i \neq x$ rest during this interval in their active states x_i'. Then, the interval density is given by:

$$\underbrace{\lambda_{x^j|u} e^{-\lambda_{x^j|u} t}}_{x \text{ transitions}} \underbrace{\Theta_{x^j x^k|u}}_{\text{to state } x^k} \underbrace{\prod_{x_i' : x_i \neq x} e^{-\lambda_{x_i'|u} t}}_{\text{while } x_i\text{'s rest}}$$

The trajectory likelihood is given by the product of intervals:

$$\prod_{x \in \mathcal{X}} \prod_{x^j \in x} \prod_{u \in U_x} \lambda_{x^j|u}^{M_{x^j|u}} e^{-\lambda_{x^j|u} T_{x^j|u}} \prod_{x^k \neq x^j} \Theta_{x^j x^k|u}^{M_{x^j x^k|u}} \tag{2}$$

where the $M_{x^j|u}$ (and $M_{x^j x^k|u}$) are the numbers of transitions out of state x^j (to state x^k), and where the $T_{x^j|u}$ are the amounts of time spent in x^j given parents settings u. Defining rate parameter $\lambda_{x^i x^j|u} = \lambda_{x^i|u} \Theta_{x^i x^j|u}$ and set element $p = x^j \times u$ (as in [3]), Equation 2 can be rewritten as:

$$\prod_{x \in \mathcal{X}} \prod_{x' \in x} \prod_{p} \lambda_{x'|p}^{M_{x'|p}} e^{-\lambda_{x'|p} T_p} \tag{3}$$

Note how the form of the likelihood in Equation 1 is identical to Equation 3.

2.3 Contrasting PCIMs and CTBNs

Despite the similarity in form, PCIMs and CTBNs model distinctly different types of continuous-time processes. Table 1 contrasts the two models. The primary difference is that, unlike point processes, CTBNs model a persistent, joint state over time. That is, a CTBN provides a distribution over the joint state for any time t. Additionally, CTBN variables must possess a 1-of-s_i state representation for $s_i > 1$ whereas point processes typically assume non-complementary event types. Furthermore, in CTBNs, observations are typically not of changes in state at particular times but instead probes of the state at a time point or interval. With persistent states, CTBNs can be used to answer interpolative queries, whereas CIMs are designed specifically for forecasting. Another notable difference is that CTBNs are Markovian: the intensities are determined entirely by the current state of the system. While more restrictive, this assumption allows for

Table 1. Contrasting piecewise-constant continuous intensity models (PCIMs) and multiplicative-forest continuous-time Bayesian networks (mfCTBNs). Key similarities are highlighted in blue.

	PCIM	mfCTBN
Model of:	event sequence	persistent state
Intensities	piecewise-constant	network-dependent constant
Dependence	event history	joint state (Markovian)
Labels	event types	variables
Emissions	events	states (x', 1 of s_i)
Structure	regression tree	multiplicative forest
Evidence	events	(partial) observations of states
Likelihood	$\prod_l \prod_s \lambda_{ls}^{M_{ls}} e^{-\lambda_{ls} T_{ls}}$	$\prod_{x'} \prod_p \lambda_{x'\|p}^{M_{x'\|p}} e^{-\lambda_{x'\|p} T_p}$

variational and MCMC methods to be applied. On the other hand, PCIMs lend themselves to forecasting because the potentially prohibitive inference about the persistent state that CTBNs require is no longer necessary. This is because the rate of event occurrences depends on the event history instead of the current state.

2.4 Multiplicative-Forest Point Processes (MFPPs)

The similar likelihood forms allow us to extend the multiplicative-forest concept [3] to PCIMs. Following [2], we define the state Σ_l and mapping $\sigma_l(t, x)$ according to regression trees. Let $\mathcal{B}_l$ be the set of basis state functions $f(t, x)$ that maps to a basis state set Σ_f, akin to $\sigma(t, x)$ that maps to a single element s. As in [3], we can view the basis functions as set partitions of the space over $\Sigma = \Sigma_{l_1} \times \Sigma_{l_2} \times \ldots \Sigma_{l_{|\mathcal{L}|}}$. Each interior node in the regression tree is associated with a basis function f. Each leaf holds a non-negative real value: the intensity. Thus one path ρ through the regression tree for event type l corresponds to a recursive subpartition resulting in a set Σ_ρ, and every $(l, s) \in \Sigma_\rho$ corresponds to leaf intensity $\lambda_{l\rho}$, i.e., we set $\lambda_{ls} = \lambda_{l\rho}$. Figure 2 shows an example of the active path providing the intensity ($\lambda_{ls} = \lambda_{l\rho} = 3$).

MFPPs replace these trees with random forests. Given that each tree represents a partition, the intersection of trees, *i.e.* a forest, forms a finer partition. The subpartition corresponding to a single intensity is given by the intersection $\Sigma_\rho = \bigcap_{j=1}^{k} \Sigma_{\rho,j}$ of sets corresponding to the active paths through trees $1 \ldots k$. The intensity $\lambda_{l\rho}$ is given by the product of leaf intensities. Figure 2 (right) shows an example of the active paths in a tree, producing the forest intensity ($\lambda_{ls} = \lambda_{l\rho} = 1 \times 3$).

MFPPs use the PCIM generative framework. Forecasting is performed by forward sampling or importance sampling to generate an approximation to the dis-

tribution at future times. Learning MFPPs is analogous to learning mfCTBNs. A tree is learned iteratively by replacing a leaf with a branch with corresponding leaves. As in forest CTBNs, MFPPs have (1) a closed form marginal log likelihood update and (2) a simple maximum likelihood calculation for modification proposals. The intensities for the modification are the ratios between observed (M_{ls}) divided by expected ($\lambda_{ls} T_{ls}$) number of events prior to modification and while Σ_ρ is active. These two properties together provide the best greedy update to the forest model.

The use of multiplicative forest point processes has several advantages over previous methods.

- Compared to trees, forest models can represent more intensities per parameter, which is equal to the number of leaves in the model. For example, if a ground truth model has k stumps, that is, k single-split binary trees, then the forest can represent the model with $2k$ parameters. An equivalent tree would require 2^k parameters. This example arises whenever two risk factors are independent, i.e., their risks multiply.
- While forests can represent these independences when needed, they also can represent non-linear processes by increasing the depth of the tree beyond one. This advantage was established in previous work comparing trees to Poisson Networks [2,8], and forests possess advantages of both approaches.
- Unlike most forest models, multiplicative-forest trees may be learned in an order that is neither sequential nor simultaneous. The forest appends a stump to the end of its tree list when that modification improves the marginal likelihood the most. Otherwise it increases the depth of one tree. The data determines which expansion is selected.
- Multiplicative forests in CTBNs are restricted to learning from the current state (the Markovian assumption), whereas MFPPs learn from a basis set over some combination of the event history, deterministic, and constant features.
- Compared to the application of supervised classification methods to temporal data, the point process model identifies patterns of event sequences over time and uses them for forecasting. Figure 3 shows an example of the supervised forecasting setup. In this case, it may be harder to predict event B without using recurrent patterns of event sequences.

We hypothesize that these advantages will result in improved performance at forecasting, particularly in domains where risk factors are independent. As many established risk factors for cardiovascular disease are believed to contribute to the overall risk independently, we believe that MFPPs should outperform tree methods at this task. Because of their facility in modeling irregular series of events, we also believe that MFPPs should also outperform off-the-shelf machine learning methods.

2.5 Related Work

A rich literature exists on point processes focusing predominantly on spatial forecasting. In spatial domains, the point process is the temporal component of

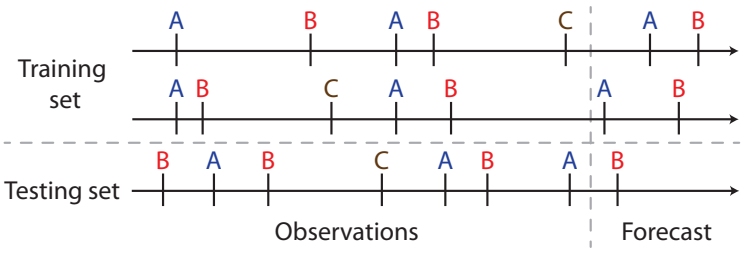

Fig. 3. Supervised forecasting. Labels are provided by the binary classification outcome: whether at least one event occurs in the forecasted region.

a model used to predict spatiotemporal patterns in data. The analysis of multivariate, spatial point processes is related to our work in its attempt to characterize the joint behavior of variables, for example, using Ripley's K function test for spatial homogeneity [9]. However, these methods do not learn dependency structures among variables; instead they seek to characterize cross-correlations observed in data. Generalized linear models for simple point processes are more closely related to our work. Here, a linear assumption for the intensity function is made, seen for example in Poisson networks [8]. PCIMs adopt a non-parametric approach and was shown to substantially improve upon previous methods in terms of model accuracy and learning time [2]. Our method builds on upon the PCIM framework.

Risk assessment for cardiovascular disease is also well studied. The primary outcome of most studies is the identification of one or a few risk factors and the quantification of the attributable risk. Our task is slightly different; we seek to predict from data the onset of future myocardial infarctions. The prediction task is closely related to risk stratification. For cardiovascular disease, the Framingham Heart Study is the landmark study for risk assessment [10]. They provide a 10-year risk of cardiovascular disease based on age, cholesterol (total and HDL), smoking status, and blood pressure. A number of studies have been since conducted purporting significant improvements over the Framingham Risk Score using different models or by collecting additional information [11]. In particular, the use of EHR data to predict heart attacks was previously addressed in [12]. However, in that work the temporal dependence of the outcome and its predictors was strictly logical and limited the success of their approach. We seek to show that, compared to standard approaches learning from features segmented in time, a point process naturally models timeline data and results in improved risk prediction.

3 Experiments

We evaluate MFPPs in two experiments. The first uses a model of myocardial infarction and stroke, and the goal is to learn MFPPs to recover the ground truth

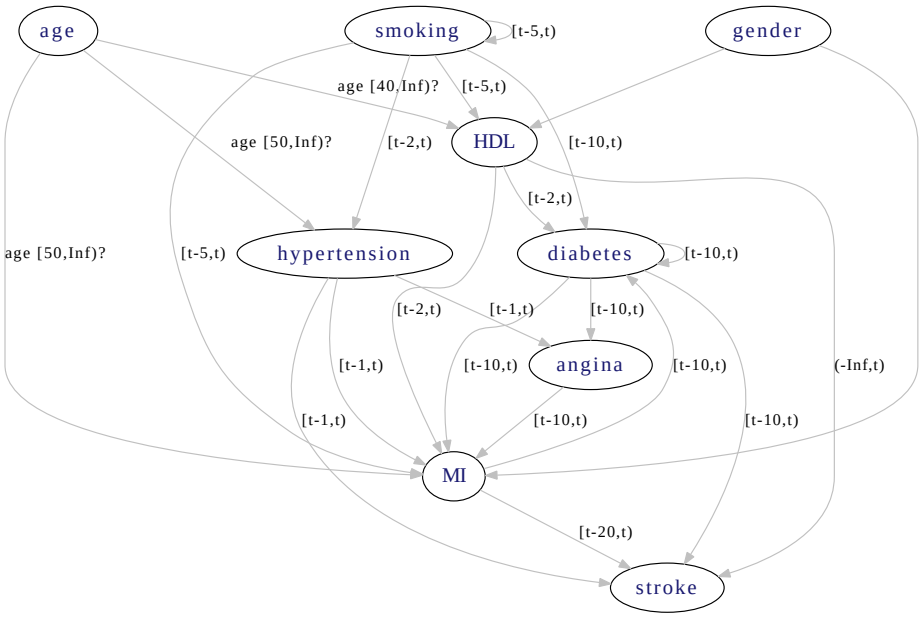

Fig. 4. Ground truth dependency structure of heart attack and stroke model. Labels on the edges determine the active duration of the dependency. Omitted in the graph is the age dependency for all non-deterministic nodes if the subject is older than 18.

model from sampled data. The second is an evaluation of MFPPs in predicting myocardial infarction diagnoses from real EHR data.

3.1 Model Experiment: Myocardial Infarction and Stroke

We introduce a ground truth PCIM model of myocardial infarction and stroke. The dependency structure of the model is shown in Figure 4. To compare MFPPs with PCIMs, we sample k trajectories from time 0 to 80 for $k = \{50, 100, 500, 1000, 5000, 10000\}$. We train each model with these samples and calculate the average log likelihood on a testing set of 1000 sampled trajectories. Each model used a BIC penalty to determine when to terminate learning. For features, we constructed a feature generator that uniformly at random selects an event type trigger and an active duration of one of $\{t-1, t-5, t-10, t-20, t-50\}$ to t. Note that the feature durations do not have a direct overlap with the dependency intervals shown in Figure 4. Our goal was to show that, even without being able to recover the exact ground truth model, we could get close with surrogate features. MFPPs were allowed to learn up to 10 trees each with 10 splitting features; PCIMs were allowed 1 tree with 100 splitting features. We also performed a two-tailed paired t-test to test for significant differences in MFPP and PCIM log likelihood. We ran each algorithm 250 times for each value of k.

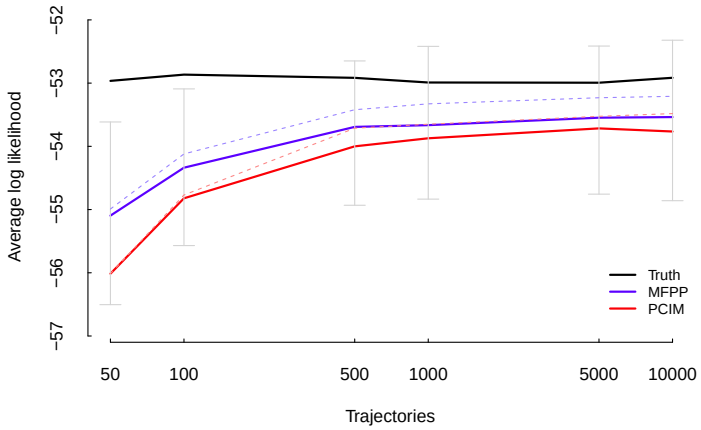

Fig. 5. Average log likelihoods for the {ground truth, MFPP, PCIM} model by the number of training set trajectories. Error bars in gray indicate the 95 percent confidence interval (omitted for the ground truth and PCIM models). Paired t-tests comparing MFPPs and PCIMs were significant at a p-value of 1-e20. Dotted lines show the likelihoods when ground truth features were made available to the models.

Figure 5 shows the average log likelihood results. Both MFPPs and PCIMs appear to converge to close to the ground truth model with increasing training set sizes. The lack of complete convergence is likely due to the mismatch in ground truth dependencies and the features available for learning. Error bars indicating the empirical 95 percent confidence intervals are also shown for MFPP. Similar error bars were observed for the ground truth and PCIM models but were omitted for clarity. The width of the interval is due to the variance in testing set log likelihoods. If we look at level average log likelihood lines in Figure 5, we observe that we only need a fraction of the data to learn a MFPP model equally good as the PCIM model. Both models completed all runs in under 15 minutes each.

We used a two-sided paired t-test to test for significant differences in the average log likelihood. For all numbers of trajectories k, the p-value was smaller than 1e-20. We conclude that the MFPP algorithm significantly outperformed the PCIM algorithm at recovering the ground truth model from data of this size.

3.2 EHR Prediction: Myocardial Infarction

In this section we describe the experiment on real EHR data. We define the task to be forecasting future onset of myocardial infarction between the years 2005 and 2010 given event data prior to 2005. We propose two forms of this

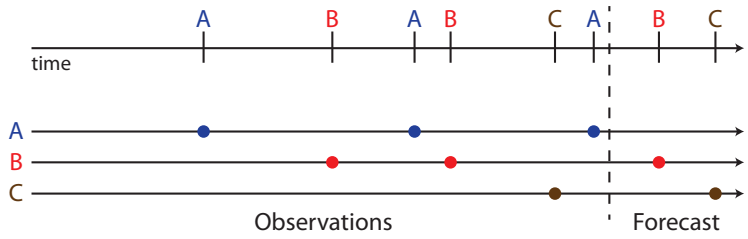

Fig. 6. *Ex ante* (traditional) forecasting. No labels for any example are available in the forecast region. The goal is to recover the events (B and C) from observations in the past.

experiment: *ex ante* and supervised forecasting. First, we test the ability of MFPP to forecast events between 2005 and 2010 in all patients given the data leading up to 2005. Figure 6 depicts the *ex ante* forecasting setup.

Second, we split our data into training and testing sets to test MFPP in its ability to perform supervised forecasting. In this setup, we provide data between 2005 and 2010 for the training set in addition to all data prior to 2005 for both training and testing sets. We choose to focus on the outcome of whether a subject has at least one myocardial infarction event between the 2005 and 2010. Figure 3 shows the supervised forecasting setup.

We use EHR data from the Personalized Medicine Research Project (PMRP) cohort study run at the Marshfield Clinic Research Foundation [13]. The Marshfield Clinic has followed a patient population residing in northern Wisconsin and the outlying areas starting in the early 1960s up to the present. From this cohort, we include all subjects with at least one event between 1970 and 2005, and with at least one event after 2010 or a death record after 2005. Filtering with these inclusion criteria resulted in a study population of 15,446, with 428 identified individuals with a myocardial infarction event between 2005 and 2010.

To make learning and inference tractable, we selected additional event types from the EHR corresponding to risk factors identified in the Framingham Heart Study[10]: age, date, gender, LDL (critical low, low, normal, high, critical high, abnormal), blood pressure (normal, high), obesity, statin use, diabetes, stroke, angina, and bypass surgery. Because the level of detail specified in EHR event codes is fine, we use the above terms that represent aggregates over the terms in our database, *i.e.*, we map the event codes to one of the coarse terms. For example, an embolism lodged in the basilar artery is one type of stroke, and we code it simply as "stroke". The features we selected produced an event list with over 1.8 million events. As MFPPs require selecting active duration windows to learn, we used durations of size $\{0.25, 1, 2, 5, 10, 100 \text{ (ever)}\}$, with more features focused on the recent past. Our intuition suggests that events occurring in the recent past are more informative than more distant events.

We compare MFPP against two sets of machine learning algorithms based on the experimental setup. For *ex ante* forecasting, we test against PCIMs [2]

and homogeneous Poisson point processes, which assume independent and constant event rates. We assess their performance using the average log likelihood of the true events in the forecast region and precision-recall curves for our target event of interest: myocardial infarction. For supervised forecasting, we test against random forests and logistic regression [2,14]. As MFPP is not an inherently supervised learning algorithm, we also include a random forest learner using features corresponding to the intensity estimates based on the *ex ante* forecasting setup. We call this method MFPP-RF. We use modified bootstrapping to generate non-overlapping training and testing sets, and we train on 80 percent of the entire data. We compare the supervised forecasting methods only in terms of precision-recall due to the non-correspondence of the methods' likelihoods.

We also make a small modification to the MFPP and PCIM learning procedure when learning for modeling myocardial infarction, i.e., rare, events. On each iteration we expand one node in the forest of every event type instead of the forest of a single event type. The reason for this is that low intensity variables contribute less to the likelihood, so choosing the largest change in marginal log likelihood will tend to ignore modeling low intensity variables. By selecting an expansion for every event type each iteration, we ensure a rich modeling of myocardial infarction in the face of high frequency events such as blood pressure measurements and prescription refills. We note that because of the independence of likelihood components for each event type, this type of round-robin expansion is still guaranteed to increase the model likelihood. This statement would not hold, for example, in CTBNs, where a change in a variable intensity may change its latent state distribution, affecting the likelihood of another variable. Finally, for ease of implementation and sampling, we learn trees sequentially and limit the forest size to 40 total splits.

***Ex Ante* Forecasting Results** Table 2 shows the average log likelihood results for *ex ante* forecasting for the MFPP, PCIM and homogeneous Poisson point process models. Both MFPPs and PCIMs perform much better than the baseline homogeneous model. MFPPs outperform PCIMs by a similar margin observed in the synthetic data set.

Figure 7 shows the precision-recall curve for predicting a myocardial infarction event between 2005 and 2010 given data on subjects prior to 2005. MFPPs and PCIMs perform similarly at this task. The high-recall region is of particular interest in the medical domain because it is more costly to miss a false negative (e.g. undiagnosed heart attack) than a false positive (false alarm). Simply put, clinical practice follows the "better safe than sorry" paradigm, so performance high-recall region is of highest concern. We plot the precision-recall curves between re-

Table 2. Log likelihood of {MFPP, PCIM, independent homogeneous Poisson processes} for forecasting patient medical events between 2005 and 2010.

Method	Log likelihood
MFPP	12.1
PCIM	10.3
Poisson	-54.8

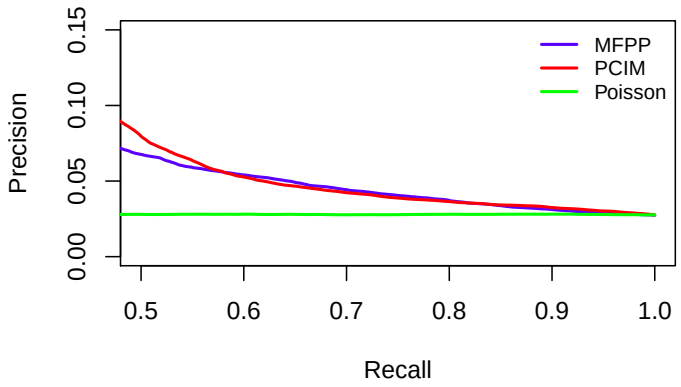

Fig. 7. Precision-recall curves for *ex ante* forecasting. MFPPs are compared against PCIMs and homogeneous Poisson point processes.

calls of 0.5 and 1.0 for this reason. The absolute precision for all methods remains low and might exhibit the challenging nature of *ex ante* forecasting. Alternatively, the low precision results could be a result of potential incompatibility of the exponential waiting time assumption and medical event data. Since forecasting can be considered a type of extrapolative prediction, a violation of the model assumptions could lead to suboptimal predictions. Despite these limitations, compared to the baseline precision of $428/15{,}446 = 0.028$, the trained methods do provide utility in forecasting future MI events nonetheless.

Supervised Forecasting Results Figure 8 provides the precision-recall curve for the supervised forecasting experiment predicting at least one myocardial infarction event between 2005 and 2010. As we see, MFPP underperforms compared to all supervised learning methods. However, the MFPP predicted intensities features boosts the MFPP-RF performance compared to the other classifiers. This suggests that while MFPP is a valuable model but may not be optimized for classification.

MFPPs also provide insight into the temporal progression of events. Figure 9 shows the first two trees of the forest learned for the rate of myocardial infarction. We observe the effects on increased risk: history of heart attack, elevated LDL cholesterol levels, abnormal blood pressure, and history of bypass surgery. While the whole forest is not shown (see `http://cs.wisc.edu/~jcweiss/ecml2013/`), the first two trees provide the main effects on the rate. As you progress through the forest, the range over intensity factors narrows towards 1. The tapering effect of relative tree "importance" is a consequence of experimental decision to learn the forest sequentially, and it provides for nice interpretation: the first few trees

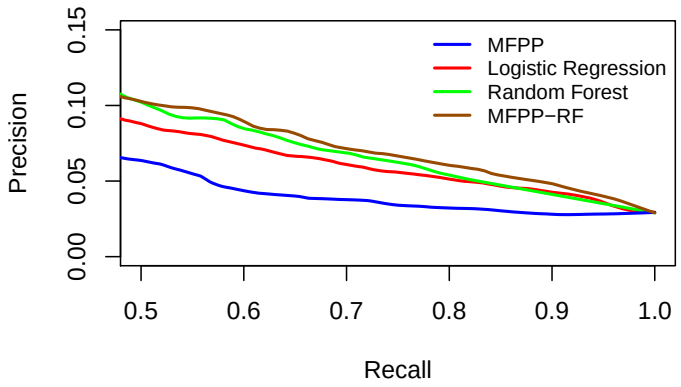

Fig. 8. Precision-recall curves for supervised forecasting. MFPPs are compared against random forests, logistic regression, and random forests augmented with MFPP intensity features.

identify the main effects, and subsequent trees make fine adjustments for the contribution of additional risk factors.

As Figure 9 shows, the dominating factor of the rate is whether a recent myocardial infarction event was observed. In part, this may be due to an increased risk of recurrent disease, but also because some EHR events are "treated for" events, meaning that the diagnosis is documented because care is provided. Care for incident heart attacks occurs over the following weeks, and so-called myocardial infarction events may recur over that time frame.

Despite the recurrence effect, the MFPP model provides an interpretable representation of risk factors and their interactions with other events. For example, Tree 1 shows that elevated cholesterol levels increase the rate of heart attack recurrence while normotensive blood pressure measurements decrease it. The findings corroborate established risk factors and their trends.

4 Conclusion

In this work we introduce an efficient multiplicative forest learning algorithm to the point process community. We developed this algorithm by combining elements of two continuous-time models taking advantage of their similar likelihood forms. We contrasted the differences between the two models and observed that the multiplicative forest extension of the CTBN framework would integrate cleanly into the PCIM framework. We showed that unlike CTBNs, MFPP forests can be learned independently because of the PCIM likelihood decomposition and intensity dependence on event history. We applied this model to two

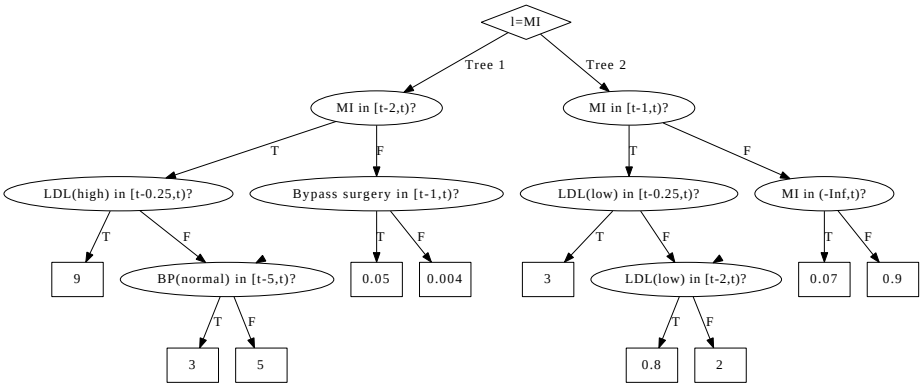

Fig. 9. First two trees in the MFPP forest. The model shows the rate predictions for myocardial infarction (MI) based on cholesterol (LDL), blood pressure (BP), previous MI, and bypass surgery. Time is in years; for example, [t-1,t) means "within the last year", and (-Inf, t) means "ever before".

data sets: a synthetic model, where we showed significant improvements over the original PCIM model, and a cohort study, where we observed that MFPP-RFs outperformed standard machine learning algorithms at predicting future onset of myocardial infarctions. We provide multiplicative-forest point process code at `http://cs.wisc.edu/~jcweiss/ecml2013/`.

While our work has shown improved performance in two different comparisons, it would also be worthwhile to consider extensions of this framework to marked point processes. Marked point processes are ones where events contain additional information. The learning framework could leverage the information about the events to make better predictions. For example, this could mean the difference between reporting that a lab test was ordered and knowing the value of the lab test. The drawback of immediate extension to marked point processes is that the learning algorithm needs to be paired with a generative model of events in order to conduct accurate forecasting. Without the generative ability, sampled events would lack the information required for continued sampling. The integration of these methods with continuous-state representations would also help allow modeling of clinical events such as blood pressure to be more precise. Finally, we would like to be able to scale our methods and apply MFPPs to any disease. Because EHR systems are constantly updated, we can acquire new up-to-date information on both phenotype and risk factors. To fully automate the process in the present framework, we need to develop a way to address the scope of the EHR, selecting and aggregating the pertinent features for each disease of interest and identifying the meaningful time frames of interest.

Acknowledgments

We acknowledge Michael Caldwell at the Marshfield Clinic for his assistance and helpful comments, the anonymous reviewers for their insightful comments, and we are deeply grateful for the support provided by the CIBM Training Program grant 5T15LM007359, NIGMS grant R01GM097618, and NLM grant R01LM011028.

References

1. U. Nodelman, C. R. Shelton, and D. Koller, "Learning continuous time Bayesian networks," in *Uncertainty in Artificial Intelligence*, 2003.
2. A. Gunawardana, C. Meek, and P. Xu, "A model for temporal dependencies in event streams," in *Advances in Neural Information Processing Systems*, 2011.
3. J. Weiss, S. Natarajan, and D. Page, "Multiplicative forests for continuous-time processes," in *Advances in Neural Information Processing Systems 25*, pp. 467–475, 2012.
4. E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature Neuroscience*, vol. 7, no. 5, pp. 456–461, 2004.
5. R. Engle, "The econometrics of ultra-high-frequency data," *Econometrica*, vol. 68, no. 1, pp. 1–22, 2000.
6. P. Diggle and B. Rowlingson, "A conditional approach to point process modelling of elevated risk," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 433–440, 1994.
7. U. Nodelman, C. Shelton, and D. Koller, "Continuous time Bayesian networks," in *Uncertainty in Artificial Intelligence*, pp. 378–387, Morgan Kaufmann Publishers Inc., 2002.
8. S. Rajaram, T. Graepel, and R. Herbrich, "Poisson-networks: A model for structured point processes," in *AI and Statistics*, vol. 10, 2005.
9. B. Ripley, "The second-order analysis of stationary point processes," *Journal of Applied Probability*, pp. 255–266, 1976.
10. P. Wilson, R. D'Agostino, D. Levy, A. Belanger, H. Silbershatz, and W. Kannel, "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.
11. I. Tzoulaki, G. Liberopoulos, and J. Ioannidis, "Assessment of claims of improved prediction beyond the Framingham risk score," *JAMA*, vol. 302, no. 21, p. 2345, 2009.
12. J. Weiss, S. Natarajan, P. Peissig, C. McCarty, and D. Page, "Machine learning for personalized medicine: predicting primary myocardial infarction from electronic health records," *AI Magazine*, vol. 33, no. 4, p. 33, 2012.
13. C. A. McCarty, R. A. Wilke, P. F. Giampietro, S. D. Wesbrook, and M. D. Caldwell, "Marshfield clinic personalized medicine research project (pmrp): design, methods and recruitment for a large population-based biobank," *Personalized Medicine*, vol. 2, no. 1, pp. 49–79, 2005.
14. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.