IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 7, 2018, accepted February 8, 2018, date of publication February 15, 2018, date of current version March 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2806340

Dynamic Modeling of Failure Events in Preventative Pipe Maintenance

CHAO ZHANG[1,2], HAO WU[1], RONGFANG BIE[1], RASHID MEHMOOD[1], AND ANTON KOS[3], (Senior Member, IEEE)

[1]College of Information Science and Technology, Beijing Normal University, Beijing 100875, China
[2]IBM Research–China, Beijing 100085, China
[3]Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia

Corresponding author: Hao Wu (wuhao@bnu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61571049, Grant 61601033, Grant 61401029, Grant 11401028, and Grant 61472044.

ABSTRACT Urban water supply network is ubiquitous and indispensable to city dwellers, especially in the era of global urbanization. Preventative maintenance of water pipes, especially in urban-scale networks, thus becomes a vital importance. To achieve this goal, failure prediction that aims to pro-actively pinpoint those ''most-risky-to-fail'' pipes becomes critical and has been attracting wide attention from government, academia, and industry. Different from classification-, regression-, or ranking-based methods, this paper adopts a point process-based framework that incorporates both the past failure event data and individual pipe-specific profile including physical, environmental, and operational covariants. In particular, based on a common wisdom of previous work that the failure event sequences typically exhibit temporal clustering distribution, we use mutual-exciting point process to model such triggering effects for different failure types. Our system is deployed as a platform commissioned by the water agency in a metropolitan city in Asia, and achieves state-of-the-art performance on an urban-scale pipe network. Our model is generic and thus can be applied to other industrial scenarios for event prediction.

INDEX TERMS Point process, pipe failure prediction, event series modeling.

I. INTRODUCTION

Modeling the dynamics of event data in continuous time space is challenging yet useful because it generally fits to a wide class of real-world problems in which a set of events are collected. These events are often associated with continuous time stamps and additional information such as event type, event participator, among others. In general, one aims to uncover the dynamic patterns from the observed event data, such that various downstream applications can be benefited.

Point process has been a flexible tool to analyze event data in continuous time space by modeling the intensity function. In particular, the Hawkes process [1] is motivated to compactly mimic the self/mutually triggering effects of the event sequences, whose occurrence patterns do not follow obvious temporal distributions. Demonstrated early on effective earthquake modeling [2], [3], Hawkes process and its derivatives have found successful applications in a wide range of practical problems especially more recently: sales analytics [4], online behavior modeling [5], [6], crime modeling [7], armed conflict analysis [8] and asset management [9], [10]. More details on Hawkes process will be introduced later.

For preventative maintenance such as water supplies network [9] and electrical power systems [10], some common wisdoms by recent studies [9], [10] suggest that: First, the failure intensity for an asset i.e. its occurrence propensity, normally stays at a relatively stable level. This can be regarded as a base component depending on its intrinsic profile covariants. Second, an occurrence of a failure can often lead to an instantaneous rise of its vulnerability. The implicit reasons are two folds: i) the occurrence of failures may suggest the existences of changed external factors contributing the increased failure risk. For instance, an increased traffic load caused by the temporal road re-routing or road excavation in a certain period, can endanger the nearby underground pipes and makes them fail more easily. ii) When the asset fails, it often becomes more fragile to failures due to the fundamental physical damage. Moreover, the vulnerability gradually fades back to the baseline when the external failure factors disappear or the asset recovers by a certain means e.g. receiving a renovation [10]. Third, different types of failures have different triggering effects to each other, e.g. a pipe burst will cause more damage to a leak failure.

The above observation perhaps has already triggered a recent work on electrical power failure modeling [10] using a variant of self-exciting Hawkes model, which is termed as Reactive Point Process (RPP). We will describe this model

in more details later. By the same motivation, we propose a generic method for event dynamics modeling and prediction, and apply our method to the problem of urban-scale water pipe failure prediction to verify the efficacy of our method.

A. WATER PIPE FAILURE PREDICTION

The pipe failure prediction task refers to the so-called preventative maintenance – a widely-adopted industrial practice, especially for those asset-intensive enterprizes and agencies like power grid and petroleum company. One core function of such a system is evaluating the failure risk for each pipe instance by utilizing the information including the profile covariants such as material type, pipe diameter, and the past dynamic failure events. Then, budgeting and maintenance can be performed more cost-effectively.

The pipe failure problem has long been an issue of concerning for municipal agencies as reported in the early studies [11]–[14]. Moreover, the urban water pipe networks are fast growing along with the global urbanization. The structural deterioration has posed challenges, not only to the daily life but also a to sustainable society since water is an essential and precious resource for human beings. Reference [15] estimates that more than $32 billion cubic meters of treated water physically leak annually through distributed network worldwide. The corresponding total annual cost is about $ 14 billion. In particular, the New York City has spent $54.6 million to manage its break pipes from 1997 to 2011 [16]. In addition, as reported by Environmental Protection Agency [17] and American Water Works Association [18], there are over 54,000 drinking water systems in the United States serving over 320 million residents via 2 million miles of water distribution pipes. Many of them are old ones, and the 20-year financing needs estimates range from $280 billion to $1 trillion for rehabilitation and replacement [17].

Due to limited budget and expensive operational cost for inspection as pipes are mostly laid underground, agencies can hardly afford to comprehensively inspect their pipes. Hence, the task of pro-actively pinpointing those most-likely-to-fail pipes, which enables cost-effective replacement and rehabilitation, becomes of vital importance. The current average annual pipeline replacement rate for utilities is usually less than 1% of the total network [9], [19].

One traditional methodology for pipe risk estimation is based on the scoring from the experienced subject matter experts (SME). They devise business rules to decide which pipes are risky and should receive top maintenance priority. However, such a domain knowledge intensive methodology usually involves subjective rules and is not readily adaptable when the setting changes e.g. the geography [20]. Moreover, compared with petroleum transmission pipelines where several manuals have been well developed such as the Pipeline Research Council International and the manual by [21],[1] urban water distribution network is much less standard, which calls for more exploration to advance this area.

Works on water network failure modeling in both industry and academia are reviewed in the surveys [22], [23]. Compared with those well-studied areas like Europe [24] and North America [25], the subtropical areas (where the case study is performed by this paper) are relatively underreported for urban pipe failure study [20]. For statistical methods, many early works focus on descriptive analysis towards pipe failures. Reference [26] calculate the average number of failures on a unit year and unit pipe length. The spatial and temporal patterns of water distribution pipe failure in the city of Winnipeg are examined in [27]. Later, predictive modeling starts more investigation. Reference [28] addresses the problem of forecasting the aggregated number of pipe failures for the network, which is key to beforehand planing. Reference [29] performs survival analysis to predict the evolution of the annual number of pipe breaks and to estimate the impact of different replacement scenarios in real case studies. Reference [30] uses Cox survival analysis as a pilot study for pipe failure prediction. Many other methods formulate the problem by a Logistic regression problem [22], [31]. Artificial Neural Network is also applied to estimate the pipe-level failure count in the water distribution networks of Benghazi city [32]. More recently a rank boosting algorithm is adopted by [19] to rank the pipe break risks. However, recent survey by [23] reveals that only a few large utilities start to employ failure prediction models to their decision support systems for rehabilitation and replacement activities. This calls for more advanced and principled approaches as well as experience sharing of a deployed system, from which they can be benefited.

This paper is organized as follows. In Section II we introduce several basic concepts on point process, and several existing methods related to our approach. Section III highlights the overview of our method and the main contribution. We present the technical details and empirical study in Section IV and in Section V respectively. Section VI concludes this paper.

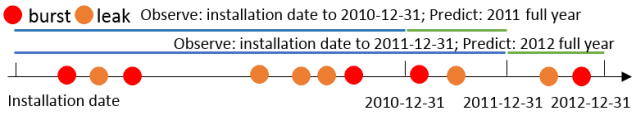


FIGURE 1. Illustration for the protocol of one-year forward failure prediction: past failures collected since the installation date to the current year end 2010 (2011 *resp.*), and prediction time window is set to 2011 (2012 *resp.*). Note the failure events exhibit a temporal clustering pattern but do not obey an obvious parametric distribution, which suggests the suitability of adopting the mutual-exciting model. This observation is also made in [9] and [10].

II. BASIC CONCEPTS

We first present a brief technical description of the problem, based on which closely related concepts and models are then introduced. In general, given 1) a prediction time point or window e.g. the next year as shown in FIGURE 1, 2) the past failure sequence associated with each pipe

[1] In the Pipeline Research Council International, 9 categories are classified and 4 in the manual [21].

instance, where the time-stamp and types of failures are recorded, and 3) the associated profile of each pipe instance i.e. its attributes such as material, diameter, length etc., one aims to predict the failure likelihood, or the expected number of failures of each pipe for each failure type.

Naively, it can be solved by a binary supervised learning method by treating the failed pipes as positive instances and non-failure ones as the negative as done by most previous work. However, a more natural way is to model the failure sequences via a point process, and make it allow to incorporate the profile information in a principled way. This paper is an endeavor in this direction.

A. INTENSITY FUNCTION OF POINT PROCESS

Point processes are widely used to model the occurrences of discrete events. In general, a point process is a time series $\{t_s\}_{s=1}^m (0 < t_i \leq T)$ at which an sequence of events $\{c_s\}_{s=1}^m$ occur. Denote $N(t)$ as the number of points (i.e., occurrences of events) in $(-\infty, t]$ and $\mathcal{H}_t = c_s | t_s < t$ as the historical events happening before t. The core concept of point process is the conditional intensity function:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}(N(t+\Delta t) - N(t)|\mathcal{H}_t)}{\Delta t} = \frac{\mathbb{E}(dN(t)|\mathcal{H}_t)}{dt}$$

where $\mathbb{E}(dN(t)|\mathcal{H}_t)$ is the expectation of the number of events happened in the interval $(t, t+dt]$ given the historical observations $\mathcal{H}_t$. The conditional intensity function represents the expected instantaneous rate of future events at time t. Note this is not a probability and can exceed 1.

B. HAWKES PROCESSES AND EXISTING VARIANTS

For one-dimensional, i.e. self-exciting Hawkes process, its conditional intensity denoting the expected instantaneous rate of future events at time t, can be written as [33]:

$$\lambda(t) = \mu + a \sum_{i:t_i<t} g(t - t_i)$$

where μ is the base intensity and t_i the time of events in the process before time t. $g(t)$ is the kernel to control the influence from the previous events. Given an event sequence $\{t_i\}_{i=1}^n$ observed in $[0, T]$, its log-likelihood estimator is

$$\begin{aligned}\mathcal{L} &= \log \prod_{i=1}^{n} \lambda(t_i) \exp\left(-\int_0^T \lambda(t)dt\right) \\ &= \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \lambda(t)dt\end{aligned}$$

We generalize the above one-dimensional formula to the D-dimension case, a multi-dimensional Hawkes process [34] is defined by a D-dimensional mutual-exciting point process. Its conditional intensity of the d-th dimension is:

$$\lambda_d(t) = \mu_d + \sum_{i:t_i<t} a_{dd_i} g_{dd_i}(t - t_i)$$

where λ is added by a base intensity component μ_d and an accumulative mutual-exciting component $\sum_{i:t_i<t} a_{dd_i} g_{dd_i}(t - t_i)$. This is a parsimonious model to incorporate the mutual-exciting effect from different types (i.e. dimensions) of events which is suited to the scenario of the pipe failure modeling problem as will be shown later. Here a_{dd_i} measures the influence from event type d_i to d. Note that the event type can refer to the failure type in this paper.

Recent works [5], [7], [8], [10], [35] start to use a single parameter to model the base intensity for each instance respectively, or universally for all instances. And several generic learning algorithms are proposed [36], [37]. However, they ignore the instance level profile covariants which can otherwise be rather informative to model since the instance level covariants are widely used in classification and regression models for failure prediction [19], [28]. As a result, on one hand, their performance may sacrifice significantly if they adopt a universal base parameter μ_d for all instances $s \in \mathcal{S}$. On the other hand, if we voraciously use many μ_d^s for each instance s, the complexity of the model i.e. its number of parameters will grow quickly and overfitting becomes a major issue which also hurts the prediction performance. We find these two alternatives are less-competitive in our empirical study and will be shown later in this paper.

The above limitation motivates us to devise a more principled model to incorporate the two important information sources: profile covariants and past failure events.

III. OVERVIEW AND CONTRIBUTION

A. OVERVIEW

In particular, there are two data sources for pipe failure modeling: i) the profile data for each pipe including a range of static covariants such as material, diameter, length, among others; ii) the pipe failure tickets recording the historical failure date, and the type of failures associated with each ticket. Ideally, a detailed classification of these failures includes cracks, splits, joint failures, and hydrant valve failures as defined in [25]. In our case, due to the granularity of the available ticket information, the failures are simplified into two coarse categories: leak and burst. Given these two datasets and a future time period window, the problem is scoring the failure risk, or formally the expected number of failures in that period as a function of the static and dynamic information associated with each pipe. The output assessment is critical to decision making for prioritizing the city scale pipe infrastructure inspection and maintenance, especially given a tight budget and limited manpower resource.

Different from the commonly-used binary classification or bipartite rank (or BoostRank) models [19], [38], we formulate the failure prediction problem via a profile-specific mutual-exciting point process model. On one hand, the profile covariants are used to parameterize the base intensity of the failure occurrence likelihood, which reflects the inherent failure propensity dependent on their built-in profiles. On the other hand, the historical failure ticket data serves as the event sequence for the purpose of exploring the reciprocal triggering effect between the two failure types. Hence, the trained model is used to evaluate the risk score of each pipe given their observed historical failure records and profiles.

Before diving into the details in Section IV, we highlight the strengths of our approach and the compromised simplifications we have made in face of a real-world problem with its particular settings. Then we summarize our contributions.

B. TECHNICAL STRENGTH OF THE PROPOSED MODEL

Compared with state-of-the-arts point process based methods for event dynamics modeling and failure prediction [9], [10], the main advantages of our approach are as follows.

1) Profile-specific base intensity modeling We devise a parametric model for the base intensity by the profile covariants and the model complexity is regardless of the instance size but depends on the number of covariants, which is much smaller as listed in TABLE 2. In contrast, [10] sets a base intensity parameter for each instance that increases the model complexity and ignores the profile data. This may introduce more risk for overfitting. While in [9] the base intensity parameter is shared by all instances, which is often unrealistic and prone to underfitting in practice.

2) Failure type modeling via a mutual-exciting point process In [10], self-exciting and self-regulation point process are employed for power grid failure event modeling, based on the assumption that a past event will cause a temporary rise or decline for the invulnerability to future events. However, their models only consider one-dimensional point process, which in fact assumes the different types of failure events have the same impact to the future failure risk. In contrast, our approach considers the impact of different failure types, which is more realistic to the common wisdom that different types of failure have different severities. For instance, a burst failure usually will cause more failures than a leak failure due to its damage, but not vice versa.

C. PRACTICAL AND COMPROMISED SIMPLIFICATION

We also make particular compromises as follows.

1) Information loss of the failure data There are three simplifications we make for utilizing the ticket data. First, the types of failures are reduced to two categories due to the data granularity currently we have, though ideally more fine-grained stratification can be defined [25]; Second, empirical studies [27] report that past failures may only contribute to subsequent failures within 20 meters. However, our accessible ticket data can only tell the time stamp rather than the exact failure location on the pipe. Thus in our model, we treat all failures of the same pipe equally regardless its (unknown) location; Third, the severity of failures is not thoroughly recorded such that only a binary flag is used, which indicates a failure occurs or not.

2) Pure temporal modeling without using spatial information Currently we have no access to, and thus do not consider the spatial location information. To some extent, they are implicitly encoded by the profile covariants such as 'ZoneImp', 'Rainfall', 'HighImp' which show their environmental settings as listed in TABLE 2. As a complementary approach, spatial analytics has been studied [20], [25], [27] to identify the high-risk areas for monitoring purpose. It is interesting to learn more complex spatiotemporal model tailored to pipe failure modeling, which we leave for future work if the fine-grained spatial data is available.[2]

D. MAIN CONTRIBUTIONS

This paper performs failure prediction by a more modernized machine learning paradigm – formulating the problem into a parameterized profile-specific mutual-exciting point process model. In addition, this model can be efficiently learned by a trackable optimization algorithm based on posterior likelihood maximization, by alternative optimization.

1) From a methodology perspective, to our knowledge, this paper is the first work to propose modeling event sequences generated from a set of instances together with their profile covariants via a unified profile-specific mutual-exciting point process model. In contrast, most existing Hawkes models and their related variants ignore the profile covariants and train separate models for different groups of instances or trivially build individual model for each instance [9], [10]. In addition, our model naturally accounts for the triggering effect between different failure types by the mutual-exciting point process kernel. We also propose an alternating optimization algorithm to learn the parameters of the model. Our point process based model is also significantly different from those regression or classification based ones [19], [28]. Our method can be seen as an extension to the two relevant point process based models [9], [10] by incorporating the failure type, as well as the profile covariants into the Hawkes process model. These two factors are not considered in their works.

2) From an application perspective, our empirical study and developed solution is a trade-off between algorithmic tractability and practical feasibility compromised to the available real-world data. Its performance is empirically verified on the urban-scale water pipe data – a metropolitan city which is comprised of hundreds of thousands pipes for both fresh and salt water systems. In contrast, the average failure count is very few per pipe.

Beyond the machine learning and data mining community, this paper is expected to be also of use to city dwellers, urban planners, environmental scientists and civil engineers.

IV. MODEL AND LEARNING ALGORITHM

Suppose we have m instances i.e. event sequences $\{c_s\}_{s=1}^m$ such that each pipe s is associated with a failure sequence c_s. For each sequence, it consists of n_s events and every event is comprised of the occurrence time stamp t_i^s and failure type i.e. dimension d_i^s: $c_s = \{(t_i^s, d_i^s)\}_{i=1}^{n_s}$, observed from the time window $[0, T_s]$. The window is instance-specific such that the starting point is the installation date of the pipe t_s^a, and the ending date is the right-censored point t_s^b, e.g. the current time-stamp. We let $T_s = t_s^b - t_s^a$.

[2]As a matter of fact, in our engagement with the water supply agency, the location information of each pipe is removed or de-identified for its sensitivity. Disclosure of location and topology information may lead to malicious physical attack.

For interpreting the proposed method clearly, the key notations used in this paper are summarized in TABLE 1.

TABLE 1. Key notations used in our model.

s	pipe instance i.e. sample
m	total number of pipe instances
n_s	total number of failures for pipe s
t_s^a	observation starting time of pipe s
t_s^b	observation ending time of pipe s
c_s	event sequence of pipe s
t_i^s	time-stamp for failure i of pipe s
d_i^s	type for failure i of pipe s
$g_{dd'}$	kernel of triggering effect from type d to d'
D	number of dimensions i.e. failure types

By specifying the multi-dimensional Hawkes model for the intensity function, we obtain the following log-likelihood objective function [34] where $G_{dd_j^s}(t) = \int_0^t g_{dd_j^s}(\tau)d\tau$.

$$\mathcal{L} = \sum_{s=1}^{m}\left(\sum_{i=1}^{n_s}\log\left(\mu_{d_i^s} + \sum_{t_j^s<t_i^s} a_{d_i^s d_j^s} g_{d_i^s d_j^s}(t_i^s - t_j^s)\right) - T_s\sum_{d=1}^{D}\mu_{d^s} - \sum_{d=1}^{D}\sum_{j=1}^{n_s} a_{dd_j^s} G_{dd_j^s}(T_s - t_j^s)\right),$$

Here μ_{d^s} denotes the base intensity at dimension d for instance s which we assume is time-constant, while depending on the profile covariants in this paper, and $a_{dd'}$ denotes the exciting impact for d affected by d'. Specifically, we consider $D = 2$ in our case: $d = 1$ for leak failure and $d = 2$ for burst. An exponential decaying kernel is used to model the mutual-exciting effects between different types of failures: $g_{ij}(t - t_0) = w_{ij}e^{-w_{ij}(t-t_0)}$ which is in line with [9], [10].

A. PROFILE-SPECIFIC BASE INTENSITY MODELING

The base intensity incorporates the inherent profile attributes associated with a pipe, irrespective how the past failure occurred. For instance, a pipe equipped with a special coating material would be more invulnerable.

Specifically, we adopt the similar profiling specific modeling techniques presented in [39] to connect the profile covariants with the base term, we model it via a Logistic function and rewrite μ_{d^s} to μ_d^s:

$$\mu_d^s = \beta_d / \left(1 + \exp(-\theta_d^T x^s)\right) \tag{1}$$

This parameterization incorporates the K profile covariates where $x^s = [1, x_1^s, x_2^s, \ldots, x_K^s]^T$ refers to the covariants in the top half in TABLE 2 and $\theta_d = [\theta_{d0}, \theta_{d1}, \ldots, \theta_{dK}]^T$ is the encoding coefficients for x^s. β_d is the scalar parameter governing the weight between base term and the exciting term. Note for the tractability of model learning, we assume each pipes has its own base intensities dependent on its associated profiles, but the base intensity is constant over time. We leave the mathematically challenging time-variant base intensity modeling for future work.

TABLE 2. Factors for pipe failure prediction. Failure ticket data is complete while the profile information is partially missing for a subset of pipes. The factors in the top half are used as profile covariants.

Factor	Available	Description
Material	partial	such as DI, UPVC, GI etc.
Length	partial	length of pipe instance
Diameter	partial	pipe diameter
ZoneImp	partial	supply zone impact
Rainfall	partial	nearby average rainfall
WCNo	partial	water crossing number
HighImp	partial	nearby highway impact
Excavate	partial	road excavation impact
Failure type	full	two types: leak and burst
$T_{\text{leak}}^{\text{last}}$	full	period since last leak failure
$T_{\text{burst}}^{\text{last}}$	full	period since last burst failure
$T_{\text{failure}}^{\text{last}}$	full	period since last failure
$F_{\text{leak}}^{\text{last3}}$	full	leak frequency in last 3 years
$F_{\text{burst}}^{\text{last3}}$	full	burst frequency in last 3 years
$F_{\text{failure}}^{\text{last3}}$	full	failure frequency in last 3 years

We further rewrite the objective regarding with the profile coefficient θ_d for failure type d. Since profile x^s is known we can rewrite the objective into the following form:

$$\mathcal{L} = \sum_{s=1}^{m}\left(\sum_{i=1}^{n_s}\log\left(\beta_{d_i^s} h^s(\theta_{d_i^s}) + \sum_{t_j^s<t_i^s} a_{d_i^s d_j^s} g_{d_i^s d_j^s}(t_i^s - t_j^s)\right) - \sum_{d=1}^{2} T_s\beta_d h^s(\theta_d) - \sum_{d=1}^{2}\sum_{j=1}^{n_s} a_{dd_j^s} G_{dd_j^s}(T_s - t_j^s)\right) \tag{2}$$

where $h^s(\theta_d) \triangleq 1/\left(1 + \exp(-\theta_d^T x^s)\right)$, and thus $\mu_d^s = \beta_d h^s(\theta_d)$.

We believe that using the profile to encode the base term is effective compared with the alternatives in [9] and [10] as shown in FIGURE 2. On one hand, if we voraciously use m independent μ_d^s to model the m pipes as in [10], we are at the risk of overfitting since the model complexity is very high. On the other hand, if we use a shared parameter μ_d for all m pipe instances as done in [9], the base intensity will be inappropriately modeled because it is unreasonable to assume all pipes have a same level of base intensity regardless their profile diversity regarding with material, length, diameter etc. This leads to underfitting. In contrast, we model the base intensity with relatively small number of parameters i.e. θ whose size is regardless of the whole pipe instance set size.

B. MUTUAL-EXCITING EFFECT MODELING

Since it is commonly believed that the failures are modulated by an reciprocal exciting effect rather than an inhibiting one, thus the coefficients $a_{ij} \geq 0$. This leads to an important trick that $\mathcal{L}$ can be surrogated by its tight lower bound $\overline{\mathcal{L}}$ according to the Jensen's inequality – see more details in Appendix. This lower bound helps break down the summation in the log term. As a result, we optimize this lower bound $\overline{\mathcal{L}}$ instead of the original one since it is much easier to solve by closed form in

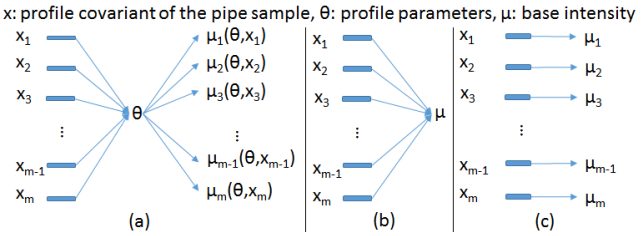


FIGURE 2. Illustration for the three modeling strategies used in three work. (a) Our method – Profile x specific base intensity μ modeling via parameter θ; (b) Method [9] – Universal base intensity μ shared by all instances from 1 to m regardless profile x. (c) Method [10] – Individual base intensity μ_i of each instance without encoding x, which is purely learned from the failure data of each single pipe i.

each iteration:

$$\overline{\mathcal{L}} = \sum_{s=1}^{m}\Bigg(\sum_{i=1}^{n_s}\Big(p_{ii}^{s}\log\frac{\beta_{d_i^s}h^s(\theta_{d_i^s})}{p_{ii}^{s}} + \sum_{j=1}^{i-1}p_{ij}^{s}\log\frac{a_{d_i^s d_j^s}g_{d_i^s d_j^s}(t_i^s-t_j^s)}{p_{ij}^{s}}\Big) - \sum_{d=1}^{2}T_s\beta_d h^s(\theta_d) - \sum_{d=1}^{2}\sum_{j=1}^{n_s}a_{dd_j^s}G_{dd_j^s}(T_s-t_j^s)\Bigg) \quad (3)$$

According to Jensen's inequality, if and only if p_{ii}^s, p_{ij}^s are set as follows in each iteration, the equation between $\mathcal{L}$ and its lower bound $\overline{\mathcal{L}}$ holds and this makes the relaxation tighter.

$$p_{ii}^{s(l)} = \frac{\beta_{d_i^s}^{(l)}h^s(\theta_{d_i^s}^{(l)})}{\beta_{d_i^s}^{(l)}h^s(\theta_{d_i^s}^{(l)}) + \sum_{j=1}^{i-1}a_{d_i^s d_j^s}^{(l)}g_{d_i^s d_j^s}(t_i^s-t_j^s)} \quad (4)$$

$$p_{ij}^{s(l)} = \frac{a_{d_i^s d_j^s}^{(l)}g_{d_i^s d_j^s}(t_i^s-t_j^s)}{\beta_{d_i^s}^{(l)}h^s(\theta_{d_i^s}^{(l)}) + \sum_{j=1}^{i-1}a_{d_i^s d_j^s}^{(l)}g_{d_i^s d_j^s}(t_i^s-t_j^s)} \quad (5)$$

For a pipe s, p_{ij}^s can be interpreted as the likelihood that the i-th failure (d_i^s, t_i^s) is affected by the previous j-th one (d_j^s, t_j^s) of the event sequence for pipe s. p_{ii}^s is the probability that i-th event is sampled from the base intensity.

Zeroing the partial derivatives $\frac{\partial\overline{\mathcal{L}}}{\partial\beta_1}$, $\frac{\partial\overline{\mathcal{L}}}{\partial\beta_2}$ and $\frac{\partial\overline{\mathcal{L}}}{\partial a_{11}}$, $\frac{\partial\overline{\mathcal{L}}}{\partial a_{12}}$, $\frac{\partial\overline{\mathcal{L}}}{\partial a_{21}}$, $\frac{\partial\overline{\mathcal{L}}}{\partial a_{22}}$ leads to the following equations for updating β_u $(u=1,2)$ and a_{uv} $(u,v=1,2)$ respectively:

$$\beta_u^{(l+1)} = \frac{\sum_{s=1}^{m}\sum_{i=1,d_i^s=u}^{n_s}p_{ii}^{s(l)}}{\sum_{s=1}^{m}h^s(\theta_u^{(l)})T_s} \quad (6)$$

$$a_{uv}^{(l+1)} = \frac{\sum_{s=1}^{m}\sum_{i=1,d_i^s=u}^{n_s}\sum_{j=1,d_j^s=v}^{i-1}p_{ij}^{s(l)}}{\sum_{s=1}^{m}\sum_{j=1,d_j^s=v}^{n_s}G_{uv}(T_s-t_j^s)} \quad (7)$$

We estimate the exciting kernel parameter $w_{uv}(u,v=1,2)$ in $g_{uv}(t-t_j)=w_{uv}e^{-w_{uv}(t-t_j)}$ by zeroing $\frac{\partial\overline{\mathcal{L}}}{\partial w_{uv}}$.

Note updating $w_{uv}(u,v=1,2)$ by zeroing $\frac{\partial\overline{\mathcal{L}}}{\partial w_{uv}}$ is implicit since it involves solving a non-linear equation. We instead replace $w_{uv}^{(l+1)}$ on the right side by $w_{uv}^{(l)}$ for efficiency:

$$w_{uv}^{(l+1)} = \frac{\sum_{s=1}^{m}\sum_{i=1,d_i^s=u}^{n_s}\sum_{j=1,d_j^s=v}^{i-1}p_{ij}^{s(l)}}{M_{uv}} \quad (8)$$

where

$$M_{uv} = \sum_{s=1}^{m}\sum_{j=1,d_j^s=v}^{n_s}a_{uv}(T_s-t_j^s)e^{-w_{uv}^{(l)}(T_s-t_j^s)} + \sum_{s=1}^{m}\sum_{i=1,d_i^s=u}^{n_s}\sum_{j=1,d_j^s=v}^{i-1}(t_i-t_j)p_{ij}^{s(l)}$$

By dropping off constant $\sum_{d=1}^{2}\sum_{j=1}^{n_s}a_{dd_j^s}G_{dd_j^s}(T_s-t_j^s)$ in $\overline{\mathcal{L}}$ w.r.t. θ_1,θ_2, we obtain two objective functions w.r.t. θ_u

$$\sum_{s=1}^{m}\Bigg(\sum_{i=1,d_i^s=u}^{n_s}\log\left(\beta_u h^s(\theta_u)+C_{ui}^s\right) - T_s\beta_u h^s(\theta_u)\Bigg) \quad (9)$$

where

$$C_{ui}^s = \sum_{t_j^s<t_i^s}a_{ud_j^s}g_{ud_j^s}(t_i^s-t_j^s), \quad u=1,2$$

Since there is no closed form solution, we apply gradient descent to update θ_{uk} $(u=1,2;k=0,1,\ldots,K)$.

The overall learning method is presented in Algorithm 1.

Algorithm 1 Learning Profile-Specific Mutual-Exciting Point Process (**PMP**) for Failure Prediction

1: **Input:** m instances i.e. pipes: $\{c_s\}_{s=1}^m$ where each pipe is with profile x^s and failure event sequence $\{t_i, d_i\}_{i=1}^{n_s}$ with failure date t_i and failure type d_i;
2: Profile $x^s=[1,x_1^s,\ldots,x_K^s]^T$ in TABLE 2 for pipe s;
3: Random initialization for $\{\beta_u,\theta_u,a_{uv},w_{uv}\}_{u,v=1}^2$;
4: Iteration stop threshold L, gradient descent step-size α;
5: **Output:**Learned parameters $\{\beta_u,\theta_u,a_{uv},w_{uv}\}_{u,v=1}^2$;
6: **for** $l=1:L$ **do**
7: Compute $\{p_{ii}^{s(l+1)}\}_{i=1}^2$ by Eq.4;
8: Compute $\{p_{ij}^{s(l+1)}\}_{i,j=1}^2$ by Eq.5;
9: Update $\{\beta_i^{(l+1)}\}_{i=1}^2$ by Eq.6;
10: Update $\{a_{ij}^{(l+1)}\}_{i,j=1}^2$ by Eq.7;
11: Update $\{w_{ij}^{(l+1)}\}_{i,j=1}^2$ by Eq.8;
12: Update $\{\theta_i^{(l+1)}\}_{i=1}^2$ by Eq.9;
13: **end for**

C. FURTHER COMPARISON WITH PEER MODELS

Besides the baseline self-exciting Hawkes process model without using the profile information [9], we find [10], [19], [40] are mostly related to this paper. The work [19] is more similar regarding with the application scenario since it addresses the same urban-scale water pipe failure prediction problem. The studies [10], [40] deal with a closely related industrial scenario – electrical power grid failure analytics while the authors in [10] also propose a variant of the self-exciting Hawkes process model for grid failure

dynamics modeling. Note in [19] and [40], they both adopt a learning-to-rank model [41] widely used in the field of machine learning and information retrieval. Compared with a point process based approach, the learning-to-rank paradigm is more similar to classification models used in [22], [31], and [32] for water pipe failure modeling. Therefore we focus on the two-fold differences to the Reactive Point Processes (RPP) model [10]:

TABLE 3. Categorization by quality of the preprocessed profile data and the denoting term of each sub-dataset. The number in bracket is the average failure rate per year from 2000 to 2011.

data quality	fresh water system	salt water system
complete	D_f^c|457,232(0.75%)	D_s^c|81,998(0.99%)
repopulated	D_f^r|156,209(0.78%)	D_s^r|18,305(1.19%)
fully missing	D_f^m|834(1.39%)	D_s^m|228(1.63%)

1) Profile-specific mutual-exciting vs. Instance-specific self-exciting model RPP [10] trains separate models for each instance (a manhole in the grid application). In our case, as already discussed in Section IV-A, a unified model is trained by using all instances and their events. The personalization is achieved by encoding the base intensity term via the individual profile covariates. By collaboratively exploring the event data across all instances it is supposed to improve the robustness of the trained model, especially given a small ratio of failure per year – see TABLE 3. Also, we model the triggering effect of event types by mutually-exciting terms. Comparatively, their method considers self-exciting effect and treats all failure types equally though different manhole outage e.g. fires, smoking. The profile-specific modeling is also used in [39] in point process learning. However, that work incorporates the profile information using an additional prior term for optimization, while in our approach, it is directly optimized via gradient descent as a whole in the overall loss function.

2) Coordinate descent learning vs. stochastic sampling method Our model is simplified in two-folds: a) the self-regulation effect (by repair) is not explicitly taken care of because we assume each failure will be followed by a repair while such repair information in fact is unavailable from the ticket data. In addition, the replacement for failed pipe can be trivially treated as the pipe database will remove the replaced one and add a new one for the replacing pipe, and assign its completion date as up-to-date; b) nor the saturation effect of the exciting effect, in order to avoid additional nonlinearity. Thanks to such simplifications, we are able to devise an efficient optimization approach (mostly closed form updating). In contrast, [10] applies the Approximate Bayesian Computation (ABC) sampling method [42].

V. EXPERIMENTS AND CASE STUDY

Mathematically being a generic machine learning paradigm for failure prediction, while our method is specifically driven by the requirement from a water agency of a metropolitan city in Asia for cost-effective and preventative urban-scale water pipe maintenance. We will briefly describe the deployed platform and report the evaluation results compared with several peer methods SVM$^{\text{stru}}$ (termed by SVM$^{\text{s}}$ for short) [43], RankBoost [19], RPP [10] and two Hawkes process baselines. Note that as reported in [19], the RankBoost model has shown superior performance against many other classical models such as NaiveBayes, Cox, Logistic Regression and Artificial Neural Network. For brevity, we exclude these methods in our evaluation. In fact we find the listed models in most cases outperform these methods on our dataset notably which is consistent with the result in [19].

A. DEPLOYED PLATFORM AND EXPERIENCE

Our pipe management platform consists of two modules: descriptive module and predictive module. The back end of the descriptive module is a data cube supporting for ad-hoc query at different aggregation levels. The predictive module is built on the proposed model. An interactive interface is provided such that user can pinpoint the pipe either by GIS (in the left panel) or by the list (in the right panel). However, the pipe location information is only used in online browsing while they are not shown in our accessible database.

The main users of this platform are the pipe network utilities such as the water agency, who need this platform to plan their budget and prioritize the preventative maintenance and rehabilitation in advance.

The dataset consists of over 600,000 pipes for fresh water system and nearly 100,000 for salt in the year when we are engaged with the agency. To construct our dataset, there are two raw data sources from the agency: i) failure ticket data including pipe ID and the associated failure type (leak/burst), failure date of each event. From the ticket data, one can extract aggregated indicators as listed in the bottom half of TABLE 2. The number of failures ranges from zero to decades for a pipe; ii) static profile database including installation date, pipe ID and other static attributes as listed in the top half of TABLE 2. The pipe ID is the universal key for matching the identical pipe from these two data sources, which constitutes the whole life period of a pipe.

The failure ticket data is well kept hence we assume them intact. For the profile related data, the pipe instances can be divided into three categories according to their quality as structured in TABLE 3: i) the major part of pipes with complete profile values; ii) a part of pipes being attribute-incomplete and the missing attributes (i.e. profile covariants) are repopulated by business rules provided by the engaged water agency based on their specific domain knowledge; iii) a small portion of pipes that are of fully profile-missing, and cannot be recovered by any means.

The installation date of all pipes is available. This allows for trivially adopting a point process model by setting the starting point of the observation window as the installation date. Evaluation are performed by setting 2010 and 2011 as the one-forward year respectively, as exemplified in FIGURE 1.[3]

[3]Our engagement with the water agency is from the year 2010 to 2012, thus we only collect the data censored by 2012.

Note the standard self-exciting Hawkes process model [9] can be regarded a baseline to our model, because neither instance-specific profile nor failure type is taken into account. In the following experiments, we will show the notable superiority over this baseline by our model.

B. EVALUATION PROTOCOL

1) CENSORED TIME WINDOW

In line with the recent pipe failure prediction work [19], the protocol follows a one year forward practice. Specifically, we use the historical failure event data up to the end of censored year as the right-censored date, together with the profile attributes, as the training data to train our model. The failure events occurring during the next year is used to validate the performance. Evaluations are performed for predicting the failure score for year 2010 and 2011, given the observation window censored by the end of 2009 and 2010 respectively. For pipe s from training dataset $\mathcal{D}$, it is associated with the profile x^s and failure tickets $\{c_s\}_{s=1}^m$ during the observation time window.

2) PERFORMANCE METRICS

The area under the receiver operating characteristic curve (ROC) [44] as adopted by related works [19], and Average Precision (AP) are both used as the performance metrics. ROC is widely used for classification and AP is often used for information retrieval. Note ROC cannot guarantee the optimal (mean) AP performance [45], as ROC assigns equal penalty to each misordering of a relevant/non-relevant instance. In contrast, AP assigns greater penalties to misorderings higher up in the predicted ranking. This behavior we believe is more suited in our case as very few failures occur each year compared with the whole instance set. The metrics are either for a specific failure type by comparing the failure-specific risk score with the actual certain type of failure, or at the 'overall' level by comparing the actual failure occurrence regardless of the types.

3) PEER METHODS

The comparing methods can be grouped into two categories: i) point process based models including RPP [10], HP^{self}, HP^{hom} and our method termed by PMP as in Algorithm 1. Note HP^{self}, HP^{hom} (we use HP^{s}, HP^{h} for short in the paper) are two simplified Hawkes Process (HP) variants to our models, by replacing the mutual-exciting term with a self-exciting term (treating different failure types equally) and building a mutual-exciting Hawkes model with a homogenous[4] base term μ (also learned from the event data) ignoring individual profile covariants. TABLE 4 shows the relation of the compared point process models by two dimensions. ii) scoring based models SVM^{stru} (termed by SVM^{s} for short) and RankBoost (RB), the former is suited in our case because it allows the outputs have more than one dimension.

[4]Here the term 'homogenous' emphasizes homogenous across instances, although throughout the paper we assume the base is homogenous over time.

TABLE 4. Comparison of different methods.

	self-exciting	mutual-exciting
profile-specific μ_d	HP^s	PMP (Ours)
instance-specific μ_d	RPP [10]	—
universal μ_d	HP [9]	HP^h

This structured learning model is expected able to better capture the relation between the two dimensions (failure types) of outputs. In contrast, the RB model is a one-dimensional output model. We select the RankBoost.B [41] in line with [19].

4) MODEL BUILDING PROTOCOL

Raw event data is used by the four point process models. For our model and HP^{s}, profile covariants are encoded in the base term of the intensity, while for HP^{h} and the RPP model, profile information is not used. For SVM^{s}, two failure types are modeled under a unified structured learning paradigm. While for the RB model, limited by its single-dimension output mechanism, we train two models for two types respectively in addition with an 'overall' model that treats all failure types equally. The input features for SVM^{s} and RB are the profile features in addition with the aggregated past failure statistics as shown in the bottom part of TABLE 2. These ad-hoc frequency statistics are used to encode the dynamic cues which is a common practice in feature engineering for training a classification/regression/ranking model. In contrast, a point process based model can naturally handle the event data.

5) RISK COMPUTING AND EVALUATION

For each pipe instance, given the observed past failure records up to the censored year $\{c_s\}_{s=1}^{m_0}, t_{m_0}^s$. The observations serve two roles: a) training an up-to-date model for next-year risk prediction; b) computing the failure score related to the next year by the trained model. Then the ROC AUC and AP can be calculated by three cases: i) the risk score for leak failure over the next year; ii) the risk for burst over the next year; iii) the 'overall' risk score regardless of the failure types. For the first two cases, when a point process model is employed, the risk score is defined by the integrated risk intensity over the whole prediction time window $r_s^k = \int_{t_a}^{t_b} \lambda_k(t)dt (k_s = 1, 2)$ and the integration protocol follows [46]; for the other two peer methods, the model output scores are used. SVM^{s} is a structured learning model such that it can issue failure-specific scores by one model, and for the RB model, scores for different failures are generated by two different models. For the 'overall' risk regardless of failure types, we add up the individual failure scores $r_s^k (k_s = 1, 2)$ as they physically refer to intensity functions. But for the scoring models, we additionally train an SVM^{s} and an RB model and use their output as the overall risk. Different output scores from two failure-specific scores cannot be added up due to the fact that for the scoring model SVM^{s} and RB.

6) DATA PARTITION FOR TESTING

The pipes are divided into 6 groups w.r.t. two systems: fresh and salt water, in addition with the three categories w.r.t.

TABLE 5. Four model categories trained by different sub-datasets segmented by system: fresh(f)/salt(s) – subscript, and by profile data quality – superscript: complete(c)/repopulated(r)/missing(m). The last column contains the estimated values of our mutual-exciting model hyper-parameters a where the subscript 'l' and 'b' refers to 'leak' and 'burst'. Specifically, a_{lb} refers to the coefficient for the exciting term from 'burst' to 'leak' and similar for a_{bl}, a_{ll}, a_{bb}. For profile-missing instances, our model is trained for each instance thus we provide the mean, standard deviations (all less than 0.15) are omitted due to space limitation. In addition, the estimated parameters by the profile-homogenous model HP^h is listed in the bracket, whereby one can see the coefficients are a little wield and counter-intuitive. We think this is because HP^h causes model bias since it neglects the profile covariants. This further suggests the importance for incorporating profile covariants to uncover the unbiased mutual relation of failures.

model	$M_f^{c\cup r}$	$M_s^{c\cup r}$	M_f^m	M_s^m
system	fresh	salt	fresh	salt
train data	$\mathcal{D}_f^c \cup \mathcal{D}_f^r$	$\mathcal{D}_s^c \cup \mathcal{D}_s^r$	$\mathcal{D}_f^m$	$\mathcal{D}_s^m$
test data	$\mathcal{D}_f^c, \mathcal{D}_f^r$	$\mathcal{D}_s^c, \mathcal{D}_s^r$	$\mathcal{D}_f^m$	$\mathcal{D}_s^m$
a_{lb}	0.82(0.56)	0.74(0.53)	0.52(—)	0.61(—)
a_{bl}	0.13(0.45)	0.21(0.48)	0.26(—)	0.48(—)
a_{ll}	0.24(0.52)	0.18(0.39)	0.31(—)	0.29(—)
a_{bb}	0.57(0.49)	0.63(0.41)	0.67(—)	0.59(—)

the quality and availability of the profile, as classified in TABLE 3. These 6 datasets are termed as $\mathcal{D}_s^c, \mathcal{D}_s^r, \mathcal{D}_s^m, \mathcal{D}_f^c, \mathcal{D}_f^r, \mathcal{D}_f^m$ where the subscripts denote for 'salt' and 'fresh' water systems, and the superscripts for 'complete', 'repopulated' and 'missing' w.r.t. the profile information respectively. Four categories of models $M_f^{c\cup r}$, $M_s^{c\cup r}$, M_f^m, M_s^m are trained and tested as depicted in TABLE 5. Note for dataset $\mathcal{D}_f^c$ and $\mathcal{D}_f^r$ (similar for the salt system), the model is trained by the union of these two datasets, while for prediction, we are interested in evaluating their performances separately since the latter $\mathcal{D}_f^r$ is assumed to contain more noises. Note that for model M_f^m and M_s^m, since profile covariants are fully missing, we only use the event data for building the point process model, and their statistics listed in the bottom of TABLE 2 to train the two scoring models. When M_f^m and M_s^m refer to our model, we adopt a similar strategy as RPP that build separate models per instance which we find better than using a constant base term for all instances.

C. RESULTS AND DISCUSSION

We first point out the methodological limitation of our study. It shall be noted that our deployed platform only serves as one of the references for the client's decision making especially in the early stage. The client's maintenance action is influenced by additional facts and considerations such as public complaints, their past working arrangement, and other constraints that make the road not readily allowed to be excavated for pipe maintenance, e.g. business regulation or the land is privately owned by the residents. Thus the prediction is supposed to not intervene the failure result notably thus rendering our test roughly as a 'blind test'.

Meanwhile, the net economic impact is out of the scope of this research paper. Because the final business impact not only depends on the deployment of the model but also i) how the prediction results are accepted by the decision maker; ii) other factors e.g. the pressure for replacing a few less-risk pipes drawing more attentions from the local citizens or media; iii) the cost of replacing each pipe is diverse and is not considered (and unavailable). It would be valuable to use the economic impact for an end-to-end decision making system, and academically it involves a stochastic optimization model to take the failure predictions, the replacement/repair cost or constraint, and other ad-hoc factors into account. However, at current stage, our predictions serve as only one of the decision-support information sources. We leave the economic evaluation in future work if more data is available.

Now we present our discussion based on the results.

1) The proposed model achieves the best performance in most cases The performance in terms of ROC and AP are reported in TABLES 6 (2010) and 7 (2011) where different failure types are evaluated on partitioned datasets.[5] Elegantly, our mutual-exciting model can handle these two types in a unified fashion thus only one model is trained to output the risk score for all failure types. Our model almost outperforms state-of-the-arts on all datasets: profile-complete data $\mathcal{D}^c$, profile-repopulated $\mathcal{D}^r$ and profile-missing $\mathcal{D}^m$ for both fresh and salt water systems. We regard this is due to that our model i) effectively incorporates all available information, especially capturing the temporal clustering pattern of failures at the fine-grained level while the peer method RankBoost [19] build separate models for each failure type and the aggregated statistics (in the bottom of TABLE 2) used by this method may not well capture the clustering pattern of failures – See FIGURE 1 for illustration; ii) collectively trains the model by all instances via covariants parametrization to improve its robustness instead of training separate models per instance as done by RPP [10]. In this spirit, our method is akin to the idea of multi-task learning here the individual task is learning the failure dynamics of each pipe. We devise a parametric model for the base intensity by the profile covariants and the model complexity is regardless of the instance size but depends on the number of covariants. Note we also test a more general model by parameterizing the profile covariants not only in the base term but also in the mutual-effect parameters a_{ij}, while we found the performance gain is very limited and even hurts the present model. This suggests our current model effectively captures the problem and keeps the model as simple as possible.

2) Separate models for individual instances are welcomed when profile is fully unknown Another interesting observation is that for dataset $\mathcal{D}_f^m$ and $\mathcal{D}_s^m$, whose covariants are missing, the RPP model performs better than HP^h, which suggests trivially modeling the personalized base term (one model per instance as no covariants can be used) is more useful than modeling the fine-grained failure types. Our approach considers the impact of different failure types, which is more realistic to the common wisdom that different types of failure have different severities. Moreover, HP^h

[5]In this paper, due to data limitation of our business contract, we only disclose the results for year 2010 and 2011.

TABLE 6. ROC (top half) and AP (bottom half) of one-year forward prediction for year 2010.

	All failures						Leak failures						Burst failures					
	SVM^s	RPP	HP^s	HP^h	RB	**PMP**	SVM^s	RPP	HP^s	HP^h	RB	**PMP**	SVM^s	RPP	HP^s	HP^h	RB	**PMP**
D_f^c	.827	.818	.841	.655	.862	.871	.814	.802	.829	.629	.847	.856	.805	.789	.813	.626	.811	.823
D_s^c	.845	.821	.849	.648	.854	.867	.832	.811	.827	.654	.844	.849	.814	.798	.810	.635	.808	.818
D_f^r	.835	.824	.837	.637	.857	.863	.828	.792	.831	.641	.843	.843	.807	.781	.805	.625	.821	.828
D_s^r	.828	.831	.829	.632	.847	.861	.834	.808	.845	.687	.848	.855	.817	.792	.807	.634	.814	.836
D_f^m	.628	.629	.607	.614	.634	.634	.621	.632	.621	.632	.638	.638	.605	.623	.614	.622	.613	.627
D_s^m	.616	.638	.626	.630	.630	.640	.617	.631	.618	.622	.627	.634	.608	.614	.601	.615	.614	.618
D_f^c	.817	.809	.837	.658	.872	.881	.808	.798	.822	.622	.856	.852	.798	.775	.816	.626	.794	.835
D_s^c	.829	.826	.844	.653	.868	.875	.819	.802	.814	.656	.859	.853	.803	.794	.812	.635	.802	.839
D_f^r	.827	.828	.832	.642	.869	.867	.811	.810	.838	.634	.829	.849	.801	.792	.811	.625	.808	.833
D_s^r	.831	.822	.839	.645	.852	.855	.824	.804	.848	.679	.835	.865	.802	.787	.812	.634	.815	.841
D_f^m	.615	.626	.609	.624	.636	.654	.621	.622	.619	.624	.638	.645	.623	.630	.607	.624	.609	.635
D_s^m	.618	.648	.633	.638	.632	.655	.617	.617	.623	.628	.627	.641	.615	.621	.610	.618	.613	.632

TABLE 7. ROC (top half) and AP (bottom haf) of one-year forward prediction for year 2011.

	All failures						Leak failures						Burst failures					
	SVM^s	RPP	HP^s	HP^h	RB	**PMP**	SVM^s	RPP	HP^s	HP^h	RB	**PMP**	SVM^s	RPP	HP^s	HP^h	RB	**PMP**
D_f^c	.834	.824	.850	.658	.858	.874	.818	.815	.829	.629	.843	.865	.805	.789	.813	.626	.811	.823
D_s^c	.841	.828	.853	.642	.851	.870	.825	.818	.827	.654	.848	.852	.814	.798	.810	.635	.808	.818
D_f^r	.829	.812	.839	.649	.864	.860	.815	.790	.831	.641	.828	.841	.807	.781	.805	.625	.821	.828
D_s^r	.836	.838	.841	.638	.852	.863	.829	.795	.845	.687	.825	.849	.817	.792	.807	.634	.814	.836
D_f^m	.632	.636	.614	.619	.629	.634	.611	.640	.614	.638	.641	.641	.617	.621	.617	.621	.616	.634
D_s^m	.621	.629	.621	.632	.638	.640	.619	.641	.622	.627	.623	.638	.621	.619	.604	.602	.619	.623
D_f^c	.839	.809	.848	.641	.863	.881	.818	.815	.829	.629	.843	.876	.805	.789	.813	.626	.811	.843
D_s^c	.859	.823	.858	.649	.862	.890	.825	.818	.827	.654	.858	.855	.814	.798	.810	.635	.808	.829
D_f^r	.834	.818	.835	.655	.856	.852	.815	.790	.831	.641	.828	.847	.807	.781	.805	.625	.821	.836
D_s^r	.846	.821	.828	.631	.832	.868	.829	.795	.845	.687	.825	.852	.817	.792	.807	.634	.814	.838
D_f^m	.636	.645	.624	.626	.634	.648	.601	.644	.636	.634	.627	.647	.612	.624	.617	.614	.602	.638
D_s^m	.615	.653	.627	.622	.645	.651	.597	.647	.625	.626	.614	.642	.625	.615	.592	.592	.605	.624

performs slightly better than HP^s in these cases which suggests capturing the fine-grained failure types is better than treating them equally when the base term in both methods is assumed homogenous across instances.

3) The mutual effects between different failure types are quantitatively captured by our method Our model depicts the mutual effect between the leak and burst failures by the hyper-parameters a_{ij} listed in the bottom 4 rows in TABLE 5, especially in the first two columns where the model is trained via profile data: burst raises higher vulnerability than leak. This reiterates the importance of fine-grained failure types modeling. From the business perspective, it urges the owners to take more effective measurements to suppress the burst failure as they are more damaging compared with the leak break, in a quantitative fashion.

VI. CONCLUSION AND FUTURE WORK

This paper proposes a point process model which synergetically incorporates both the profile covariants, and the dynamic failure events of different types. Based on a relaxation using the lower-bound of the raw log-likelihood function, an efficient coordinate descent learning algorithm is devised to learn the model parameters. Our approach performs competitively against several state-of-the-arts, as empirically verified on the real-world urban-scale pipe data. For future work, it is an inverse problem to infer the underlying attributes via the observed failure events for the pipes whose missing profiles are currently repopulated by the ad-hoc rules. It is also attractive to incorporate the spatial dimension in the point process modeling. Moreover deep learning models [47] for point process learning is another way to improve the current method.

APPENDIX
FEASIBILITY OF THE SURROGATE FUNCTION

First, according to Jensen's inequality, we have

$$\beta_{d_i^s} h^s(\theta_{d_i^s}) + \sum_{t_j^s < t_i^s} a_{d_i^s d_j^s} g_{d_i^s d_j^s}(t_i^s - t_j^s)$$
$$\geq p_{ii}^s \log \frac{\beta_{d_i^s} h^s(\theta_{d_i^s})}{p_{ii}^s} + \sum_{j=1}^{i-1} p_{ij}^s \log \frac{a_{d_i^s d_j^s} g_{d_i^s d_j^s}(t_i^s - t_j^s)}{p_{ij}^s}$$

The equation holds if and if only: $\beta = \beta^{(l)}$, $a = a^{(l)}$ and $w = w^{(l)}$ due to p_{ii} and p_{ij} are a function w.r.t. $\beta^{(l)}, a^{(l)}, w^{(l)}$ in the $l+1$th iteration. As a result, we have the following relation where the term on the right is $\overline{\mathcal{L}}$ in Eq.3:

$$\mathcal{L}(\beta, a, w) \geq \overline{\mathcal{L}}(\beta, a, w | \beta^{(l)}, a^{(l)}, w^{(l)})$$
$$\mathcal{L}(\beta^{(l)}, a^{(l)}, w^{(l)}) = \overline{\mathcal{L}}(\beta^{(l)}, a^{(l)}, w^{(l)} | \beta^{(l)}, a^{(l)}, w^{(l)})$$

Therefore,

$$\begin{aligned}
&\mathcal{L}(\beta^{(l)}, a^{(l)}, w^{(l)}) \\
&= \overline{\mathcal{L}}(\beta^{(l)}, a^{(l)}, w^{(l)} | \beta^{(l)}, a^{(l)}, w^{(l)}) \\
&\leq \overline{\mathcal{L}}(\beta^{(l+1)}, a^{(l+1)}, w^{(l+1)} | \beta^{(l)}, a^{(l)}, w^{(l)}) \\
&\leq \overline{\mathcal{L}}(\beta^{(l+1)}, a^{(l+1)}, w^{(l+1)} | \beta^{(l+1)}, a^{(l+1)}, w^{(l+1)}) \\
&= \mathcal{L}(\beta^{(l+1)}, a^{(l+1)}, w^{(l+1)})
\end{aligned}$$

This implies maximizing $\overline{\mathcal{L}}$ with respect to β, a, w at each iteration ensures that the value of $\mathcal{L}$ increase monotonically.

REFERENCES

[1] A. G. Hawkes and D. Oakes, "A cluster process representation of a self-exciting process," *J. Appl. Probab.*, vol. 11, no. 3, pp. 493–503, 1974.

[2] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *J. Amer. Stat. Assoc.*, vol. 83, no. 401, pp. 9–27, 1988.

[3] Y. Ogata, "Space-time point-process models for earthquake occurrences," *Ann. Inst. Stat. Math.*, vol. 50, no. 2, pp. 379–402, 1998.

[4] J. C. Yan *et al.*, "On machine learning towards predictive sales pipeline analytics," in *Proc. AAAI*, 2015, pp. 1945–1951.

[5] L. Li, H. Deng, A. Dong, Y. Chang, and H. Zha, "Identifying and labeling search tasks via query-based Hawkes processes," in *Proc. KDD*, 2014, pp. 731–740.

[6] Y. Zhang, Y. Wei, and J. Ren, "Multi-touch attribution in online advertising with survival theory," in *Proc. ICDM*, 2014, pp. 687–696.

[7] A. Stomakhin, M. B. Short, and A. L. Bertozzi, "Reconstruction of missing data in social networks based on temporal patterns of interactions," *Inverse Problems*, vol. 27, no. 11, p. 115013, 2011.

[8] A. Zammit-Mangion, M. Dewar, V. Kadirkamanathan, and G. Sanguinetti, "Point process modelling of the Afghan War Diary," *Proc. PNAS*, 2012, pp. 12414–12419.

[9] J. C. Yan *et al.*, "Towards effective prioritizing water pipe replacement and rehabilitation," in *Proc. IJCAI*, 2013, pp. 2931–2937.

[10] Ş. Ertekin, C. Rudin, and T. H. McCormick, "Reactive point processes: A new approach to predicting power failures in underground electrical systems," *Ann. Appl. Stat.*, vol. 9, no. 1, pp. 122–144, 2015.

[11] G. E. Arnold, "Experience with main breaks in four large cities-Philadelphia," *J. Amer. Water Works Assoc.*, vol. 53, no. 8, pp. 1041–1044, 1960.

[12] E. J. Clark, "Experience with main breaks in four large cities-New York," *J. Amer. Water Works Assoc.*, vol. 53, no. 8, pp. 1045–1048, 1960.

[13] H. W. Niemeyer, "Experience with main breaks in four large cities-Indianapolis," *J. Amer. Water Works Assoc.*, vol. 53, no. 8, pp. 1041–1058, 1960.

[14] G. J. Remus, "Experience with main breaks in four large cities-Detroit," *J. Amer. Water Works Assoc.*, vol. 53, no. 8, pp. 1041–1058, 1960.

[15] B. Kingdom, R. Liemberger, and P. Marin, "The challenge of reducing non-revenue water (NRW) in developing countries—How the private sector can help: A look at performance-based service contracting (English)," in *Water Supply and Sanitation Sector Board Discussion Paper Series*. Washington, DC, USA: World Bank, 2006.

[16] S. Carter and P. V. Rush, "Filtration avoidance determination (FAD) reports," Dept. Environ. Protection, New York, NY, USA, Tech. Rep., 2012.

[17] *Drinking Water Infrastructure Needs Survey and Assessment*, Environ. Protection Agency, Washington, DC, USA, 2013.

[18] *Buried No Longer: Confronting America's Infrastructure Challenges*, Amer. Water Works Assoc., Denver, CO, USA, 2012.

[19] R. Wang, W. Dong, Y. Wang, K. Tang, and X. Yao, "Pipe break prediction: A data mining method," in *Proc. ICDE*, 2013, pp. 1208–1218.

[20] W.-Z. Shi, A.-S. Zhang, and O.-K. Ho, "Spatial analysis of water mains failure clusters and factors: A Hong Kong case study," *Ann. GIS*, vol. 19, no. 2, pp. 89–97, 2013.

[21] W. K. Muhlbauer, *Pipeline Risk Management Manual*, 3rd ed. New York, NY, USA: Elsevier, 2004.

[22] S. Yamijala, S. D. Guikema, and K. Brumbelow, "Statistical models for the analysis of water distribution system pipe break data," *Rel. Eng. Syst. Safety*, vol. 94, no. 2, pp. 282–293, 2009.

[23] A. M. S. Clair and S. Sinha, "State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models," *Urban Water J.*, vol. 9, no. 2, pp. 85–112, 2012.

[24] S. E. Christodoulou, A. Gagatsis, A. Agathokleous, S. Xanthos, and S. Kranioti, "Urban water distribution network asset management using spatio-temporal analysis of pipe-failure data," in *Proc. ICCCBE*, 2012, pp. 27–29.

[25] D. P. de Oliveira, D. B. Neill, J. H. Garrett, Jr., and L. Soibelman, "Detection of patterns in water distribution pipe breakage using spatial scan statistics for point events in a physical network," *J. Comput. Civil Eng.*, no. 21, no. 1, pp. 21–30, 2010.

[26] U. Shamir and C. Howard, "An analytical approach to scheduling pipe replacement," *J. Amer. Water Works Assoc.*, vol. 71, no. 5, pp. 248–258, 1979.

[27] I. C. Goulter and A. Kazem, "Spatial and temporal groupings of water main pipe breakage in Winnipeg," *Can. J. Civil Eng.*, vol. 15, no. 1, pp. 91–97, 1998.

[28] Y. Kleiner and B. Rajani, "Using limited data to assess future needs," *J. Amer. Water Works Assoc.*, vol. 91, no. 7, pp. 47–61, 1999.

[29] G. Pelletier, A. Mailhot, and J.-P. Villeneuve, "Modeling water pipe breaks—Three case studies," *J. Water Res. Planning Manage.*, vol. 129, no. 2, pp. 115–123, 2003.

[30] C. Tian, J. Xiao, J. Huang, and F. Albertao, "Pipe failure prediction," in *Proc. SOLI*, Beijing, China, 2011, pp. 121–125.

[31] A. Debón, A. Carrión, E. Cabrera, and H. Solano, "Comparing risk of failure models in water supply networks using ROC curves," *Rel. Eng. Syst. Safety*, vol. 95, no. 1, pp. 43–48, 2010.

[32] A. M. Bubtiena, A. H. Elshafie, and O. Jafaar, "Application of artificial neural networks in modeling water networks," in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2011, pp. 50–57.

[33] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971.

[34] T. J. Liniger, "Multivariate Hawkes processes," Ph.D. dissertation, Swiss Federal Inst. Technol., Zürich, Switzerland, 2009.

[35] L. Li and H. Zha, "Energy usage behavior modeling in energy disaggregation via marked Hawkes process," in *Proc. AAAI*, 2015, pp. 672–678.

[36] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes," in *Proc. AISTATS*, 2013, pp. 641–649.

[37] K. Zhou, H. Zha, and L. Song, "Learning triggering kernels for multi-dimensional Hawkes processes," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1301–1309.

[38] C. Rudin, R. J. Passonneau, A. Radeva, H. Dutta, S. Ierome, and D. Isaac, "A process for predicting manhole events in manhattan," *Mach. Learn.*, vol. 80, no. 1, pp. 1–31, 2010.

[39] J. Yan *et al.*, "Modeling contagious merger and acquisition via point processes with a profile regression prior," in *Proc. IJCAI*, 2016, pp. 2690–2696.

[40] C. Rudin *et al.*, "Machine learning for the New York City power grid," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 328–345, Feb. 2012.

[41] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *J. Mach. Learn. Res.*, vol. 4, pp. 933–969, Nov. 2003.

[42] P. J. Diggle and R. J. Gratton, "Monte Carlo methods of inference for implicit statistical models," *J. Roy. Stat. Soc. B, Methodol.*, vol. 46, no. 2, pp. 193–227, 1984.

[43] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," in *Proc. ICML*, 2004, pp. 104–112.

[44] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

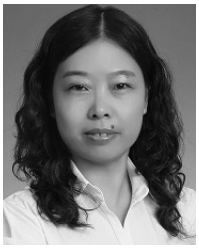
[45] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. ICML*, 2006, pp. 233–240.

[46] A. Dassios and H. Zhao, "Exact simulation of Hawkes process with exponentially decaying intensity," *Electron. Commun. Probab.*, vol. 18, no. 62, pp. 1–13, 2013.

[47] S. Xiao, J. Yan, X. Yang, H. Zha, and S. Chu, "Modeling the intensity function of point process via recurrent neural networks," in *Proc. AAAI*, 2017, pp. 1597–1603.

CHAO ZHANG received the M.S. and B.E. degrees from the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. He is currently pursuing the Ph.D. degree with the College of Information Science and Technology, Beijing Normal University, Beijing, China. He is a Staff Researcher with IBM Research–China, and has been a IBM Master Inventor for his contribution to the patent portfolio of IBM. His research interests include computer vision and machine learning.

HAO WU received the B.E. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively. From 2013 to 2015, he was a Research Associate with the Lawrence Berkeley National Laboratory. Since 2015, he has been taking charge of some related research projects with the Lawrence Berkeley National Laboratory. In 2015, he joined the Center for Big Data Mining and Knowledge Engineering. He is currently a Post-Doctoral Research Fellow with the College of Information Science and Technology, Beijing Normal University. He has published about 20 papers including many international journals or magazines, such as *Neurocomputing*, *Visual Computer*, *Multimedia Tools and Applications*, the *Journal of Visual Communication and Image Representation*, *IET Computer Vision*, and so on.



RONGFANG BIE received the M.S. and Ph.D. degrees from the College of Information Science and Technology, Beijing Normal University. She is currently a Professor with the College of Information Science and Technology, Beijing Normal University. She was a Visiting Faculty member with the Computer Laboratory, University of Cambridge, from 2003 to 2004. She is the author or co-author of over 100 papers. Her current research interests include knowledge representation and acquisition for the Internet of Things, dynamic spectrum allocation, big data analysis and application, and so on.

RASHID MEHMOOD received the master's degree from COMSITS, Islamabad, Pakistan, in 2012, and the Ph.D. degree from the School of Information Sciences and Technology, Beijing Normal University, Beijing, China, in 2017. His current research interests include clustering and DNA barcode analysis.

ANTON KOS (SM'17) received the Ph.D. degree in electrical engineering from the University of Ljubljana, Slovenia, in 2006. He is a member of the Laboratory of Information Technologies at the Department of Communication and Information Technologies. He is currently an Assistant Professor with the Faculty of Electrical Engineering, University of Ljubljana. He is the co-author of 27 papers which have appeared in international engineering journals and of over 50 papers for international conferences. His teaching and research work includes communication networks and protocols, quality of service, dataflow computing and applications, usage of sensors in biofeedback systems and applications, signal processing, and information systems. Since 2018, he has been an IEEE Slovenia ComSoc Chapter Chair.