

Article

# Failure Prognosis of High Voltage Circuit Breakers with Temporal Latent Dirichlet Allocation <sup>†</sup>

Gaoyang Li , Xiaohua Wang \*, Aijun Yang, Mingzhe Rong \* and Kang Yang

State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; ligaoyang@stu.xjtu.edu.cn (G.L.); yangaijun@mail.xjtu.edu.cn (A.Y.); yk3115160030@stu.xjtu.edu.cn (K.Y.)

\* Correspondence: xhw@mail.xjtu.edu.cn (X.W.); mzrong@mail.xjtu.edu.cn (M.R.)

<sup>†</sup> This paper is an extended version of our paper published in Guo, C., Li, G., Zhang, H., Ju, X., Zhang, Y., Wang, X. Defect Distribution Prognosis of High Voltage Circuit Breakers with Enhanced Latent Dirichlet Allocation. In Proceedings of the International Conference on Prognostics and Health Management (PHM-Harbin 2017), Harbin, China, 9–12 July 2017.

Received: 15 October 2017; Accepted: 15 November 2017; Published: 20 November 2017

**Abstract:** The continual accumulation of power grid failure logs provides a valuable but rarely used source for data mining. Sequential analysis, aiming at exploiting the temporal evolution and exploring the future trend in power grid failures, is an increasingly promising alternative for predictive scheduling and decision-making. In this paper, a temporal Latent Dirichlet Allocation (TLDA) framework is proposed to proactively reduce the cardinality of the event categories and estimate the future failure distributions by automatically uncovering the hidden patterns. The aim was to model the failure sequence as a mixture of several failure patterns, each of which was characterized by an infinite mixture of failures with certain probabilities. This state space dependency was captured by a hierarchical Bayesian framework. The model was temporally extended by establishing the long-term dependency with new co-occurrence patterns. Evaluation of the high voltage circuit breakers (HVCBs) demonstrated that the TLDA model had higher fidelities of 51.13%, 73.86%, and 92.93% in the Top-1, Top-5, and Top-10 failure prediction tasks over the baselines, respectively. In addition to the quantitative results, we showed that the TLDA can be successfully used for extracting the time-varying failure patterns and capture the failure association with a cluster coalition method.

**Keywords:** failure prognosis; Latent Dirichlet Allocation; high voltage circuit breakers

## 1. Introduction

With the increasing and unprecedented scale and complexity of power grids, component failures are becoming the norm instead of exceptions [1–3]. High voltage circuit breakers (HVCBs) are directly linked to the reliability of the electricity supply, and a failure or a small problem with them may lead to the collapse of a power network through chain reactions. Previous studies have shown that traditional breakdown maintenance and periodic checks are not effective for handling emergency situations [4]. Therefore, condition-based maintenance (CBM) is proposed as a more efficient maintenance approach for scheduling action and allocating resources [5–7].

CBM attempts to limit consequences by performing maintenance actions only when evidence is present of abnormal behaviors of a physical asset. Selection of the monitoring parameters is critical to its success. Degradation of the HVCB is caused by several types of stress and aging, such as mechanical maladjustment, switching arcs erosion, and insulation level decline. The existing literature covers a wide range of specific countermeasures, including mechanism dynamic features [8–10], dynamic contact resistance [11], partial discharge signal [12,13], decomposition gas [14], vibration [15], and spectroscopic monitoring [16]. Furthermore, numerous studies applied neural networks [8], support vector machine

(SVM) [17], fuzzy logic [18], and other methods [19], to introduce more automation and intelligence into the signal analysis. However, these efforts were often limited to one specific aspect in their diagnosis of the failure conditions. In addition, the requirements for dedicated devices and expertise restrict their ability to be implemented on a larger scale. Outside laboratory settings, field recordings, including execution traces, failures, and warning messages, offer another easily accessible data source with broad failure category coverage. The International Council on Large Electric Systems (CIGRE) recognizes the value of event data and has conducted three world-wide surveys on the reliability data of circuit breakers since the 1970s [20–22]. Survival analysis, aiming at reliability evaluation and end-of-life assessment, also relies on the failure records [2,23].

Traditionally, the event log is not considered as an independent component in the CBM framework, as the statistical methodologies were thought to be useful only for average behavior predictions or comparative analysis. In contrast, Salfner [24] viewed failure tracking as being of equal importance to symptom monitoring in online prediction. In other fields, such as transactional data [25], large distributed computer systems [26], healthcare [27], and educational systems [28], the event that occurs first is identified as an important predictor of the future dynamics of the system. The classic Apriori-based sequence mining methods [29], and new developments in nonlinear machine learning [27,30] have had great success in their respective fields. However, directly applying these predictive algorithms is not appropriate for HVCB logs for three unique reasons: weak correlation, complexity, and sparsity.

- (1) Weak correlation. The underlying hypothesis behind association-based sequence mining, especially for the rule-based methods, is the strong correlation between events. In contrast, the dependency of the failures on HVCBs is much weaker and probabilistic.
- (2) Complexity. The primary objective of most existing applications is a binary decision: whether a failure will happen or not. However, accurate life-cycle management requires information about which failure might occur. The increasing complexity of encoding categories into sequential values can impose serious challenges on the analysis method design, which is called the “curse of cardinality”.
- (3) Sparsity. Despite the cardinality problem, the types of failure occurring on an individual basis is relatively small. Some events in a single case may have never been observed before, which makes establishing a statistical significance challenging. The inevitable truncation also aggravates the sparsity problem to a higher degree.

The attempts to construct semantic features of events, by transforming categorical events into numerical vectors, provide a fresh perspective for understanding event data [31,32]. Among the latent space methods, the Latent Dirichlet Allocation (LDA) method [33], which represents each document as mixtures of topics that ejects each word with certain probabilities, offers a scalable and effective alternative to standard latent space methods. In our preliminary work, we introduced the LDA into failure distribution prediction [34]. In this paper, we further extended the LDA model with a temporal association by establishing new time-attenuation co-occurrence patterns, and developed a temporal Latent Dirichlet Allocation (TLDA) framework. The techniques were validated against the data collected in a large regional power grid with regular records over a period of 10 years. The Top-N recalls and failure pattern visualization were used to assess the effectiveness. To the best of our knowledge, we are the first to introduce the advanced sequential mining technique into the area of HVCB log data analysis.

The rest of this paper is organized as follows. The necessary process to transfer raw text data into chronological sequences is introduced in Section 2. Section 3 provides details of the proposed TLDA model. The criteria presented in Section 4 are not only for performance evaluation but also show the potential applications of the framework. Section 5 describes the experimental results in the real failure histories of the HVCBs. Finally, Section 6 concludes the paper.

## 2. Data Preprocessing

### 2.1. Description of the HVCB Event Logs

To provide necessary context, the format of the collected data is described below. The HVCBs' failure dataset was derived from 149 different types of HVCBs from 219 transformer substations in a regional power grid in South China. The voltage grades of the HVCBs were 110 kV, 220 kV, and 500 kV, and the operation time ranged from 1 to 30 years. Most of the logs were retained for 10 years, aligned with the use of the computer management system. Detailed attributes of each entry are listed in Table 1. In addition to the device identity information, the failure description, failure reason, and processing measures fields contain key information about the failure.

**Table 1.** Attributes of the failure logs.

Attribute	Content
ID	Numerical order of a failure entry
Voltage grade	110 kV, 220 kV, or 550 kV
Substation	Location of the equipment failure, e.g., ShenZhen station
Product model	Specified model number, e.g., JWG1-126
Equipment type	A board taxonomy, e.g., high-voltage isolator, gas insulated switchgear (GIS)
Failure description	Detailed description of the phenomena observed
Failure reason	Cause of the failure
Failure time	Time when a failure was recorded
Processing measures	Performed operation to repair the high voltage circuit breaker (HVCB)
Processing result	Performance report after repair
Repair time	Time when a failure was removed
Installing time	Time when a HVCB was first put into production
Others	Including the person responsible, mechanism type, a rough classification, manufacturers, etc.

### 2.2. Failure Classification

One primary task of pre-processing is to reduce the unavoidable subjectivity and compress the redundant information. Compared to the automatically generated logs, the failure descriptions of HVCBs are created manually by administrators containing an enormous amount of text information. Therefore, the skill of the administrators significantly influences the results. Underreporting or misreporting the root cause may reduce the credibility of the logs. Only by consolidating multiple information sources can a convincing failure classification be generated. An illustrating example is presented in Table 2. The useful information is hidden in the last part and can be classified as an electromotor failure. Completing this task manually is time-consuming and is highly prone to error. Automatic text classification has traditionally been a challenging task. Straightforward procedures, including keyword searches or regular expression, cannot meet the requirements.

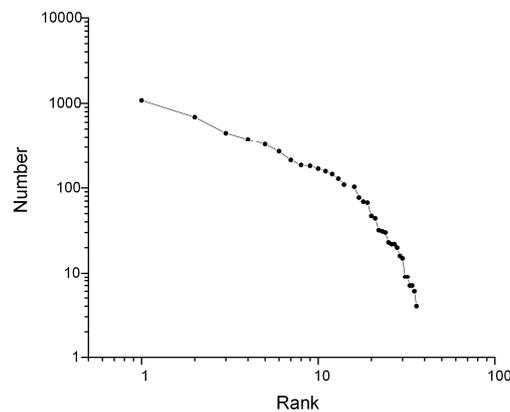
**Table 2.** A typical manual log entry sample.

Failure Description	Failure Reason	Processing Measures
Circuit breakers connected with the high voltage side of the main transformer cannot close or open. The power supply runs faultlessly during inspection	Bad manufacturing quality	Replace the electromotor

Due to progress in deep learning technology, establishing an end-to-end text categorization model has become easier. In this study, the Google seq2seq [35] neural network was adopted by replacing the decoder part with a SVM. The text processing procedure was as follows: (1) An expert administrator manually annotated a small set of the examples with concatenated texts, including the failure description, failure reason, and processing measures; (2) After tokenization and conjunction removal, the labeled texts were used to train a neural network; (3) Another small set of failure texts were predicted by the neural network. Wrong labels were corrected and added to the training set;

(4) Steps (2) and (3) were repeated until the classification accuracy reached 90%; (5) The trained network was used to replace manual work. The preferential classification taxonomy was the accurate component location that broke the operation. The failure phenomenon was recorded when no failure location was available. Finally, 36 kinds of failures were extracted from the man-machine interaction.

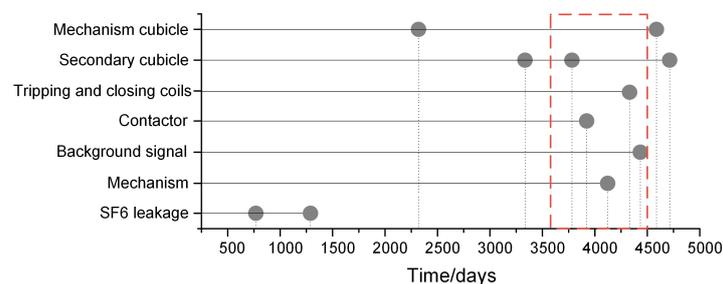
The numbers of different failures were ranked in descending order and plotted in a log-log axis shown in Figure 1. The failure numbers satisfy a long-tail distribution [36], making it hard to recall the failures with a lower occurrence frequency.



**Figure 1.** Long tail distribution of the failure numbers.

### 2.3. Sequence Aligning and Spatial Compression

The target outputs of the sequence pre-processing are event chains in chronological order. As mentioned earlier, the accessibility to the failure data was limited to the last 10 years. Therefore, the visible sequences were bilaterally truncated, creating new difficulties for comparing different sequences. Instead of using the actual failure times, the times of origin of the HVCBs were changed to their installation time to align different sequences. To mitigate the sparsity problem, spatial compression was used by clustering failure events from the same substation of the same machine type, as they often occur in bursts. Finally, of the 43,738 raw logs, 7637 items were HVCB-related. After sequence aligning and spatial compression, 844 independent failure sequences were extracted, with an average length of nine. A sequence example can be found in Figure 2. Different failures that break the device operation continually occurred along the time axis.



**Figure 2.** A graphical illustration of a failure sequence.

## 3. Proposed Method

The key idea behind all failure tracking predictions is to obtain the probability estimations using the occurrence of previous failures. The problem is unique because both the training sets and the test sets are categorical failure data. A detailed expression of the sequential mining problem studied in this paper can be summarized as follows: the HVCB failure prognosis problem is a topic of sequential mining concerned with estimating the future failure distribution of a HVCB, based on the failure

history of itself, and the failure sequences of all the other HVCBs, under the limitations of short sequences and multiple categories.

This section will present how the TLDA provides a possible solution to this problem by embedding the temporal association into the LDA model.

### 3.1. Latent Dirichlet Allocation Model

LDA is a three-level hierarchical Bayesian model originally used in natural language process. It posits that each document is modeled as a mixture of several topics, and each topic is characterized by an infinite mixture of words with certain probabilities. A LDA example is shown in Figure 3. A document consists not only of words but also the topics assigned to the words, and the topic distribution provides a sketch of the document subject. LDA introduces topics as a fuzzy skeleton to combine the discrete words into a document. Meanwhile, the shared topics provide a convenient indicator to compare the similarity between different documents. LDA has had success in a variety of areas by extending the concepts of document, topic, and word. For example, a document can be a gene [37], an image [38], or a piece of code [39], with a word being a feature term, a patch, or a programming word. Likewise, a failure sequence can be treated as a document, and a failure can be recognized as a word. The topics in LDA can be analogous to failure patterns that represent the kinds of failures that cluster together and how they develop with equipment aging. Two foundations of LDA are the Dirichlet distribution and the idea of latent layer.

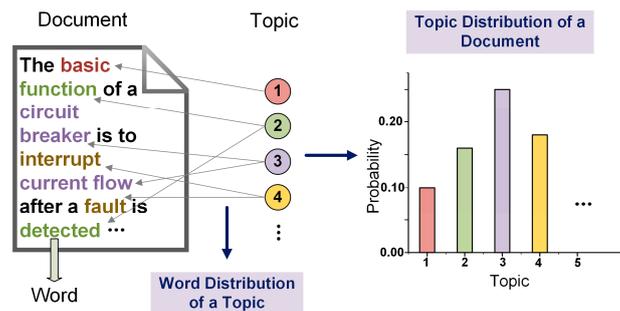


Figure 3. An illustrating example of Latent Dirichlet Allocation (LDA).

#### 3.1.1. Dirichlet Distribution

Among the distribution families, the multinomial distribution is the most intuitive for modeling a discrete probability estimation problem. The formulation of the multinomial distribution is described as:

$$f(x_1, \dots, x_k; n; p_1, \dots, p_k) = \frac{\Gamma(\sum_i x_i + 1)}{\prod_i \Gamma(x_i + 1)} \prod_{i=1}^k p_i^{x_i} \quad (1)$$

which satisfies  $\sum_i x_i = n$  and  $\sum_i p_i = 1$ . Multinomial distribution represents the probability of  $k$  different events for  $n$  experiments, with each category having a fixed probability  $p_i$  happening  $x_i$  times.  $\Gamma$  is the gamma function. Furthermore, the Maximum Likelihood Estimation (MLE) of  $p_i$  is:

$$\hat{p} = \frac{x_i}{\sum_i x_i} \quad (2)$$

which implies that the theoretical basis of the statistic method is MLE estimation of a multinomial distribution. Effective failure prognosis methods must balance the accuracy and details of the adequate grain information. However, we supposed that the dataset has  $M$  sequences and  $N$  kinds of failures. Modeling a multinomial distribution for each HVCB will result in a parameter matrix with the shape of  $M \times N$ . These statistics for individuals will cause most elements to be zero. Taking the failure sequence in Figure 1 as an example, among the 36 kinds of failure, only 7 have been seen, making providing a reasonable probability estimation for the other failures impossible. This is why much of the statistical analysis relies on a special classifying standard to reduce types of failure, or ignores the independence

of the HVCBs. Two solutions are feasible for alleviating the disparities: introduce a priori knowledge or mine associations among different failures and different HVCBs.

One possible way to introduce a priori knowledge is based on Bayes' theorem. Bayesian inference is a widely used method of statistical inference to estimate the probability of a hypothesis when insufficient information is available. By introducing a prior probability on the parameters, Bayesian inference acts as a smoothing filter. Conjugate prior is a special case where the prior and posterior distribution have the same formulation. The conjugate prior distribution of multinomial distribution is Dirichlet distribution, which is:

$$Dir(\vec{p} | \vec{\alpha}) = f(p_1, \dots, p_k; \alpha_1, \dots, \alpha_k) = \frac{1}{\Delta \vec{\alpha}} \prod_{i=1}^k p_i^{\alpha_i - 1} \tag{3}$$

with the normalization coefficient being:

$$\Delta \vec{\alpha} = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^k \alpha_i)} \tag{4}$$

similar to the multinomial distribution. Due to the Bayesian rule, the posterior distribution of  $\vec{p}$  with new observations  $\vec{x}$  can be proven as:

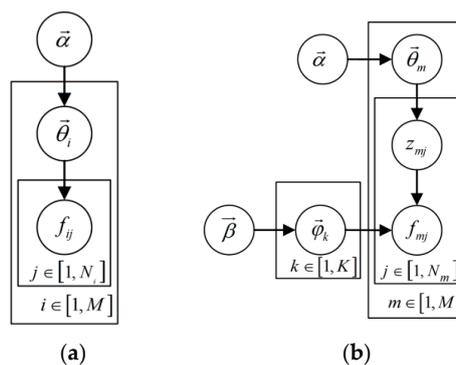
$$p(\vec{p} | \vec{\alpha}, \vec{x}) = Dir(\vec{p} | \vec{\alpha} + \vec{x}) \tag{5}$$

with the mean being:

$$\vec{p}_i = \frac{x_i + \alpha_i}{\sum_i (x_i + \alpha_i)} \tag{6}$$

From Equation (6), even the failures with no observations are assigned to a prior probability associated with  $\alpha_i$ . The conjugate relation can be described as a generative process shown in Figure 4a:

- (1) Choose  $\vec{\theta}_i \sim Dir(\vec{\alpha})$ , where  $i \in \{1, 2, 3, \dots, M\}$ ;
- (2) Choose a failure  $f_{ij} \sim Multinomial(\vec{\theta}_i)$ , where  $j \in \{1, 2, 3, \dots, N_i\}$ .



**Figure 4.** Graphical representations comparison of the Dirichlet distribution and LDA: (a) graphical representation of the Dirichlet distribution; (b) graphical representation of LDA.

### 3.1.2. Latent Layer

Matrix completion is another option for solving the sparsity problem that establishes global correlation among units. The basic task of matrix completion is to fill the missing entries of a partially observed matrix. In sequential prediction with limited observations, predicting the probabilities of failures that have never appeared is a problem. Using the recommend system as an example, for a sparse user-item rating matrix  $R$  with  $m$  users and  $f$  items, each user had only rated several items.

To fill the unknown space,  $R$  is first decomposed as two low dimensional matrices  $P \in R^{m \times f}$  and  $Q \in R^{n \times f}$  satisfying:

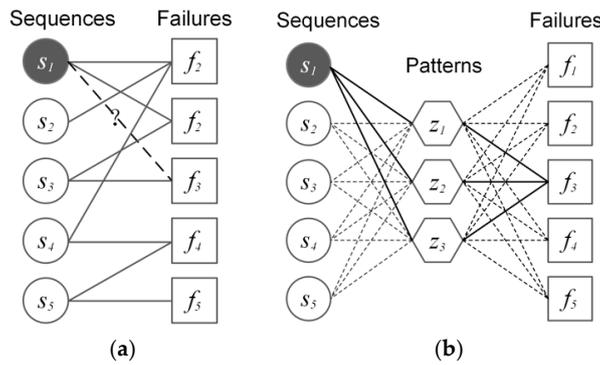
$$R \approx PQ^T = \hat{R} \tag{7}$$

with the aim of making  $\hat{R}$  as close to  $R$  as possible. Then, the rating of user  $u$  to item  $\hat{R}(u, i) = \hat{r}_{ui}$ , can be inferred as:

$$\hat{r}_{ui} = \sum_f p_{uf}q_{if} \tag{8}$$

Many different realizations of Equation (7) can be created by adopting different criteria to determine whether the given matrices are similar. The spectral norm or the Frobenius norm creates the classical singular value decomposition (SVD) [40], and the root-mean-square error (RMSE) creates the latent factor model (LFM) [41] model. In addition, regularization terms are useful options to increase the generalization of the model.

Analogously, a latent layer with  $L$  elements can be introduced between the HVCB sequences and the failures. For  $M$  sequences with  $N$  kinds of failures, instead of  $M N$ -parameter multinomial distributions described above,  $M L$ -parameter multinomial models, and  $L N$ -parameter multinomial models are preferred, where  $L$  failure patterns are extracted. A schematic diagram of the comparison is shown in Figure 5. No direct observations exist to fill the gap between  $s_1$  and  $f_3$ ; the connection of  $s_1-z_1-f_3, s_1-z_2-f_3, s_1-z_3-f_3$  will provides a reasonable suggestion.



**Figure 5.** Schematic diagram of the matrix completion: (a) the graphical representation of the failure probability estimation task. The solid lines represent the existing observations, and the dotted line represents the probability of the estimate. (b) The model makes an estimation by the solid lines after matrix decomposition.

### 3.1.3. Latent Dirichlet Allocation

The combination of Bayesian inference and matrix completion creates the LDA. Two Dirichlet priors are assigned to the two-layer multinomial distributions. A similar idea is shared by LFM, where the regularization items can be theoretically deduced from the assumption of Gaussian priors. The major difficulties in realizing LDA lie in the model inference. In LDA, it is assumed that the  $j$ th failure in sequence  $m$   $f_{mj}$  comes from a failure pattern  $z_{mj}$ , making  $f_{mj}$  satisfying a multinomial distribution parameterized with  $\vec{\varphi}_{z_{mj}}$ . In addition, the failure pattern  $z_{mj}$  also originates from a multinomial distribution whose parameters are  $\vec{\theta}_m$ . Finally, from the perspective of Bayesian statistics, both  $\vec{\varphi}_{z_{mj}}$  and  $\vec{\theta}_m$  are sampled from two Dirichlet priors with parameters  $\vec{\alpha}$  and  $\vec{\beta}$ . The original Dirichlet-multinomial process can evolve to a three-layer sampling process as follows:

- (1) Choose  $\vec{\theta}_m \sim Dir(\vec{\alpha})$ , where  $m \in \{1, 2, 3, \dots, M\}$ ;
- (2) Choose  $\vec{\varphi}_k \sim Dir(\vec{\beta})$ , where  $k \in \{1, 2, 3, \dots, K\}$ ;

For each failure  $f_{mj}$ ,

- (3) Choose a latent value  $z_{mj} \sim \text{Multinomial}(\vec{\theta}_m)$ ;
- (4) Choose a failure  $f_{mj} \sim \text{Multinomial}(\vec{\varphi}_{z_{mj}})$ .

where  $m \in \{1, 2, 3, \dots, M\}$ , and  $j \in \{1, 2, 3, \dots, N_m\}$ .  $N_m$  is the failure number in sequence  $m$ , and  $M$  is the total sequence number.

The probabilistic graphic of LDA is shown in Figure 4b, and the joint probability distribution of all the failures under this model is given by:

$$p(\vec{f}, \vec{z}) = \prod_{m=1}^M \prod_{j=1}^{N_m} \sum_{k=1}^K p(f_{mj} | z_{mj} = k) p(z_{mj} = k) \tag{9}$$

The learning targets of LDA include  $\vec{\theta}_m$  and  $\vec{\varphi}_k$ . They can both be inferred from the topic assigning  $\vec{z}$ . The posterior distribution of  $\vec{z}$  cannot be directly solved. Gibbs sampling is one possible solution. First, the joint probability distribution can be reformulated as:

$$p(\vec{f}, \vec{z} | \vec{\alpha}, \vec{\beta}) = \prod_{k=1}^K \frac{\Delta(\vec{n}_k + \vec{\beta})}{\Delta(\vec{\beta})} \prod_{m=1}^M \frac{\Delta(\vec{n}_m + \vec{\alpha})}{\Delta(\vec{\alpha})} \tag{10}$$

where  $\vec{n}_k = \{n_k^w\}_{w=1:V}$ , and  $\vec{n}_m = \{n_m^k\}_{k=1:K}$  are the statistics of the failures count under topic  $k$ , and topic count under failure sequence  $m$ .  $V$  is the number of failure types. The conditional distribution of the Gibbs sampling can be obtained as:

$$p(z_{mj} = k | \vec{z}_{-mj}, \vec{\alpha}, \vec{\beta}) = \frac{n_k^{f_{mj}} - 1 + \beta_{f_{mj}}}{\sum_{i=1}^V (n_{k,-mj}^i - 1 + \beta_i)} \cdot \frac{n_m^k - 1 + \alpha_k}{\sum_{i=1}^K (n_{m,-mj}^i + \alpha_i)} \tag{11}$$

where  $n_{k,-mj}^i$  is the number of failures with the index  $i$  assigned to topic  $k$ , excluding the failure  $f_{mj}$ , and  $n_{m,-mj}^i$  is the number of failures in sequence  $m$  with topic  $i$ , excluding the failure  $f_{mj}$ . After certain iterations, the posterior estimation of  $\vec{\theta}_m$  and  $\vec{\varphi}_k$  can be inferred with:

$$\theta_{mk} = \frac{n_m^k + \alpha_k}{\sum_{i=1}^K (n_m^i + \alpha_i)} \tag{12}$$

$$\varphi_{kw} = \frac{n_k^w + \beta_w}{\sum_{i=1}^V (n_k^i + \beta_i)} \tag{13}$$

Finally, the posterior failure distribution of the  $i$ th HVCB can be predicted with:

$$\vec{p}_m = \sum_{k=1}^K p(z_k) p(\vec{w} | z_k) = \sum_{k=1}^K \theta_{mk} \vec{\varphi}_k \tag{14}$$

### 3.2. Introducing the Temporal Association into LDA

Even with the promising advantage of finding patterns in the categorical data, directly borrowing LDA to solve the sequence prediction problem has some difficulties. LDA assumes that data samples are fully exchangeable. The failures are assumed to be independently drawn from a mixture of multinomial distributions independently, which is not true. In the real world, failure data are naturally collected in time order, and different failure patterns evolve. So, it is important to exploit the temporal characteristics of the failure sequences.

To introduce the time attributes in LDA, we first assumed that the future failure was most related to the other failures within a time slice. Instead of using the failure sequences of the full life circle, the long sequences were divided into several sub-sequences by a sliding time window with width  $W$ . The sub-sequences may overlap with each other. Under this assumption, a simple way to use LDA in a time series is to directly exploit the pattern distributions  $\vec{\theta}_m$  in different time-slices. However, this

approach does not consider the dependence among different slices. In the LDA model, the dependence among different sub-sequences can be represented by the dependency among the pattern distributions. A modified probabilistic graph is shown in Figure 6, where  $u_{ms}$  represents the topic distribution of a specified sub-sequence, and  $\vec{w}$  are the prior parameters, with the joint distribution being:

$$p(\vec{f}, \vec{z} | \vec{w}) = \int \int p(\vec{u}_{m0} | \vec{w}) \prod_{s=1}^{J_m} p(\vec{u}_{ms} | \vec{u}_{m0}, \vec{u}_{m1}, \dots, \vec{u}_{m,s-1}, \vec{w}) \prod_{j=1}^{N_{ms}} p(z_{msj} | \vec{u}_{ms}) d\vec{u}_{ms} d\vec{u}_{m0} \quad (15)$$

where  $J_m$  is the number of sub-sequences in sequence  $m$ ,  $N_{ms}$  is the number of failures in the sub-sequence  $s$ , and  $\vec{u}_{ms}$  is the topic distribution of a specified sub-sequence.

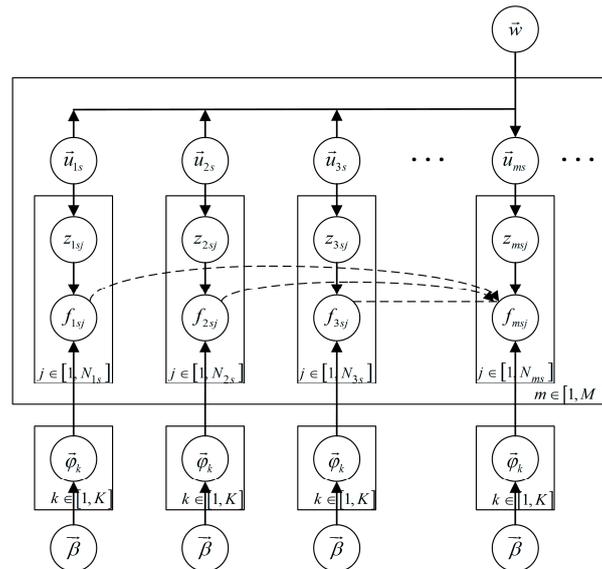


Figure 6. Graphical representation for a general sequential extension of LDA.

Due to the lack of conjugacy between Dirichlet distributions, the posterior inference of Equation (15) can be intractable. Simplifications, such as the Markov assumption and specified conditional distributions, can elucidate the posterior distribution out [42,43]. However, the formulation does not need to be Markovian, and the time dependency can still be complicated. To overcome this problem, an alternative method of creating a new co-occurrence mode is proposed to establish the long-term dependency among different sub-sequences. Specifically, from Equations (12) and (13), the failures that occur together are likely to have the same failure pattern. In other words, co-occurrence is still the foundation for deeper pattern mining in LDA. Therefore, instead of specifying the dependency among the topic distributions, as shown by the dotted line in Figure 6, a direct link was constructed between the current and earlier failures by adding the past failures into current sub-sequence with certain probabilities. Additionally, the adding operation should embed the temporal information by assigning a higher probability to the closer ones. Based on the requirements, a sampling rate comforting exponential decay is implemented as follows:

$$p(x) = \begin{cases} \exp(-\frac{x}{\Delta}), & 0 \leq x < T \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where the attenuation coefficient  $\Delta$  controls the decreasing speed of  $p(x)$  along the time interval  $x$ .  $T$  is the time at the left edge of the current time window. Figure 7 shows the schematic diagram of the process for constructing new co-occurrence patterns. To predict the future failure distribution, the failures ahead of the current time window are also included. Each iteration generates new data combinations to argument the data. An outline of the Gibbs sampling procedure with the new data generation method is shown in Algorithm 1.

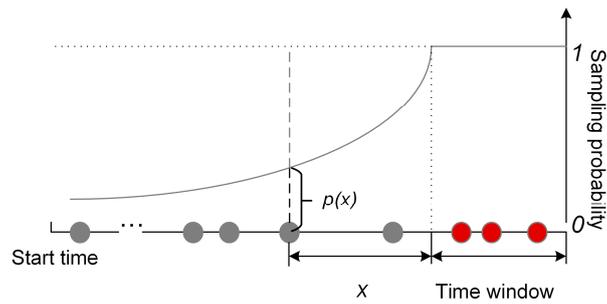


Figure 7. The sampling probability within and prior to the time window.

---

**Algorithm 1** Gibbs sampling with the new co-occurrence patterns

---

Input: Sequences, MaxIteration,  $\vec{\alpha}$ ,  $\vec{\beta}$ ,  $\Delta$ ,  $W$

Output: posterior inference of  $\vec{\theta}$  and  $\vec{\varphi}$

- 1: Initialization: randomly assign failure patterns and make sub-sequences by  $W$ ;
  - 2: Compute the statistics  $n_k^{f_{mj}}$ ,  $n_m^k$ ,  $n_{k,-mj}^i$ ,  $n_{i,-mj}^k$  in Equation (11) for each sub-sequence;
  - 3: **for** iter in 1 to MaxIteration **do**
  - 4:     **foreach** sequence in Sequences **do**
  - 5:         **foreach** sub-sequence in sequence **do**
  - 6:             Add new failures in the current sub-sequence based on Equation (16);
  - 7:             **foreach** failure in the new sub-sequence **do**
  - 8:                 Draw new  $z_{mj}$  from Equation (11);
  - 9:                 Update the statistics in Equation (11);
  - 10:             **end for**
  - 11:         **end for**
  - 12:     **end for**
  - 13:     Compute the posterior mean of  $\vec{\theta}$  and  $\vec{\varphi}$  based on Equations (12) and (13)
  - 14: **end for**
  - 15: Compute the mean of  $\vec{\theta}$  and  $\vec{\varphi}$  of last several iterations
- 

Based on the above premise, the TLDA framework for extracting the semantic characteristics and predicting the failure distribution is shown in Figure 8. After preprocessing and generating the sub-sequences, an alternating renewal process was implemented between the new co-occurrence pattern construction and the Gibbs sampling. The final average output reflects the time decrease presented in Equation (16) due to the multi-sampling process. Finally, Equation (14) provides the future distribution prognosis using the learned parameters of the last sub-sequence of each HVCB.

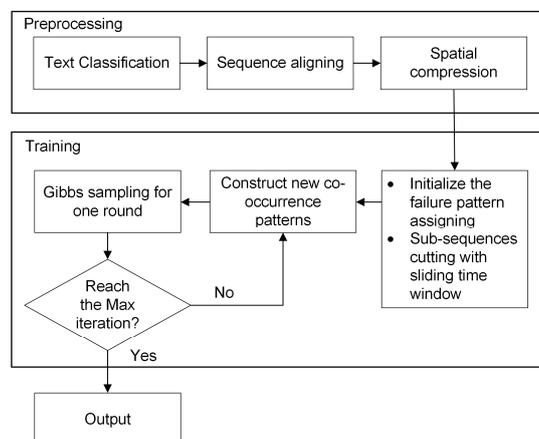


Figure 8. Log analysis framework by the temporal Latent Dirichlet Allocation (TLDA).

## 4. Evaluation Criteria

The output of the proposed system is the personalized failure distribution for each HVCB. However, directly verifying the prediction result is impossible due to the sparsity of the failure sequences. Therefore, several indirectly quantitative and qualitative criteria are proposed as follows.

### 4.1. Quantitative Criteria

Instead of verifying the entire distribution, the prognosis ability of the model was testified by predicting the upcoming one failure. Several evaluation criteria are developed as follows.

#### 4.1.1. Top-N Recall

The Top-N prediction is originally used in the recommend system to check if the recommended  $N$  items satisfy the customers. Precision and recall are the most popular metrics for evaluating the Top-N performance [44]. With only one target behavior, the recall becomes proportional to the precision, which can be simplified as:

$$Recall = \frac{\sum_{h \in H} |R_N(h) \cap T(h)|}{|H|} \quad (17)$$

where  $R_N(h)$  is the failure set with the Top-N highest prediction probabilities for HVCB  $h$ ,  $T(h)$  is the failure that subsequently occurred, and  $|H|$  is the sum of HVCBs to be predicted. The recall indicates whether the failure that subsequently occurred is included in the Top-N predictions. Considering the diversity of the different failure categories, Top-1, Top-5 and Top-10 recalls were used.

#### 4.1.2. Overlapping Probability

The overlapping probability  $P_o$  is proposed as an aided index to the Top-1 recall, which is defined as the probability the model assigns for  $T(h)$ . For instance, assuming a model concludes that the next failure probabilities for  $a$ ,  $b$ ,  $c$  are 50%, 40%, and 10%, respectively, after a while, failure  $b$  actually occurs. Then, the overlapping probability is 40%. This index provides an outline of how much the probability distribution overlaps with the real one-hot distribution, which can also be understood as the confidence. With similar Top-1 recall, higher mean overlapping probability represents a more reliable result.

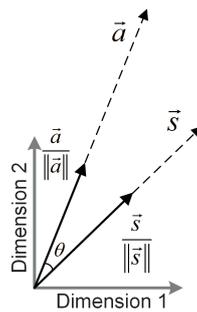
These two kinds of quantitative criteria are suitable for different maintenance strategies, considering the limitation of the maintainers' rigor. The Top-N recall corresponds to the strategy of focusing on the Top-N rank failure types, whereas the overlapping probability is another possible strategy of monitoring on the probabilities of the failure types that exceed the threshold.

### 4.2. Qualitative Criteria

The TLDA can provide explicit semantic characteristics. The results of our algorithm offer a deep new perspective for understanding the failure modes and their variation trends. For example, different failure patterns can be extracted by examining the failures with high proportions. By considering  $\vec{\theta}$  as a function of time, investigating rise and fall of different failures and how they interact is easy, either from a global perspective or when focusing on one sample. In addition, by introducing the angle cosine distance as a measurement, the similarity between failure  $p$  and failure  $q$  can be calculated as:

$$I_{pq} = \frac{\sum_{i=1}^K \varphi_{ip} \varphi_{iq}}{\sqrt{\sum_{i=1}^K \varphi_{ip}^2} \sqrt{\sum_{i=1}^K \varphi_{iq}^2}} \quad (18)$$

Figure 9 depicts the cosine distance computing method. Only the angle between the two vectors affects this indicator. A higher cosine distance often indicates more similar failure reasons.



**Figure 9.** Schematic diagram of the cosine distance with two dimensions.

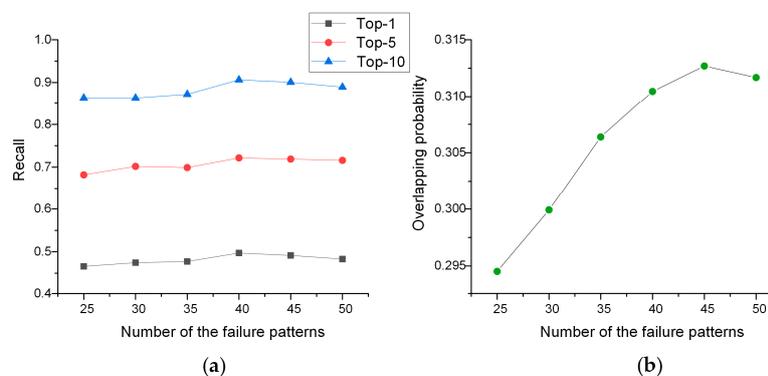
## 5. Case Study

The experimental dataset was based on the real-world failure records described in Section 2. After data processing, the failure history of each HVCB was listed as a failure sequence in chronological order. A cross-validation test was used to assess the performance with the following process. Firstly, the last failure of each sequence was separated as the test set. Then, the remaining instances were used to train the TLDA model based on Algorithm 1. For each validation round, the tail part of each failure sequences was randomly abandoned to obtain the new test sets.

### 5.1. Quantitative Analysis

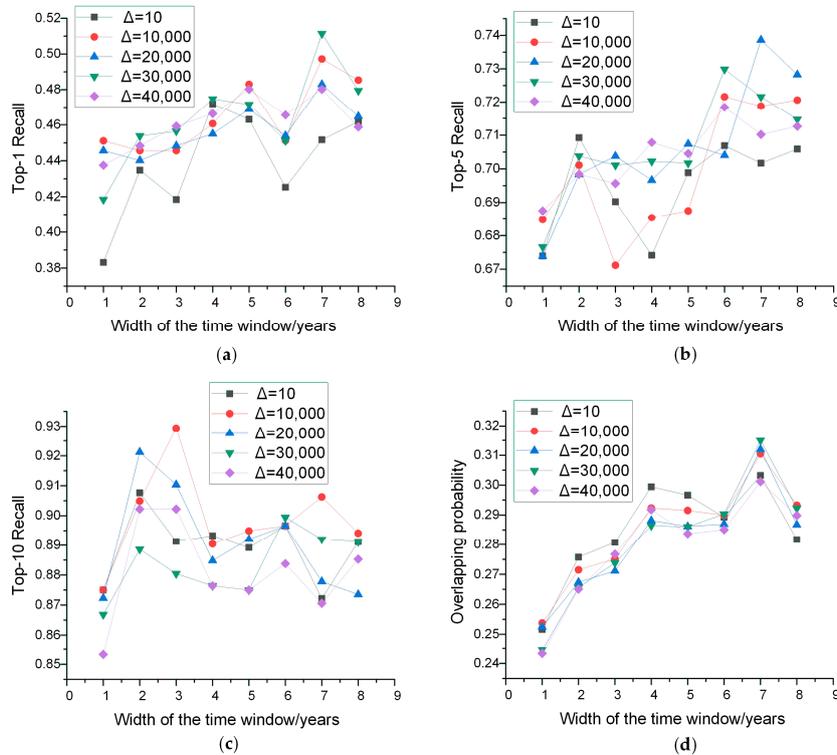
#### 5.1.1. Parameter Analysis

Hyper-parameters of the proposed method include the number of the failure patterns  $K$ , the width of the time window  $W$ , and the attenuation coefficient  $\Delta$ . For all runs of the algorithm, the Dirichlet parameters  $\vec{\alpha}$  and  $\vec{\beta}$  were assigned with symmetric priors of  $1/K$  and  $0.01$ , respectively, which were slightly different from the common setting [45]. Gibbs sampling of 300 iterations was sufficient for the algorithm to converge. For each Gibbs sampling chain, the first 200 iterations were discarded, and the average results of the last 100 iterations were taken as the final output. The first set of experiments was conducted to analyze the model performance with respect to  $K$  among  $\{25, 30, 35, 40, 45, 50\}$ . Figure 10 shows the results of Top-1, Top-5, Top-10 recalls, and the overlapping probability under fixed  $W$  and  $\Delta$  of six years and 10,000 days, respectively. These evaluation indexes do not appear to be much affected by the number of failure patterns. The failure pattern of 40 surpasses the others slightly for the Top-N recalls. The overlapping probability increased to relatively stable numerical values after 40. The overfitting phenomenon, which perplexes many machine learning methods, was not serious with high numbers of failure patterns.



**Figure 10.** Performance comparison versus the number of failure patterns: (a) the Top-1, Top-5, and Top10 recalls with respect to the number of failure patterns. (b) the overlapping probability with respect to the number of failure patterns.

In the next experiment, the qualitative criteria were examined as a function of the time window  $W$  and the attenuation coefficient  $\Delta$ , with the number of the failure patterns  $K$  fixed at 40. The results are shown in Figure 11. The peak values of different criteria were achieved with different parameters. The optimal parameters with respect to the performance metrics are summarized in Table 3.



**Figure 11.** Performance comparison versus time window length and the attenuation coefficient: (a) the Top-1 recall versus the model parameters; (b) the Top-5 recall versus the model parameters; (c) the Top-10 recall versus the model parameters; and (d) the overlapping probability versus the model parameters.

**Table 3.** Optimal parameters for different prediction tasks.

Performance Criteria	$W$ (Years)	$\Delta$ (Days)
Top-1	7	30,000
Top-5	7	20,000
Top-10	3	10,000
Overlapping Probability	7	30,000

From Table 3, the high Top-1 recall calls for a relatively large window size of seven years and a large decay parameter of 30,000 days, while the best Top-10 recall was obtained with smaller parameters of three years and 10,000 days. The Top-5 recall also requires a large  $W$  of seven years but a smaller  $\Delta$  of 20,000 days when compared to the Top-1 recall. The overlapping probability also shares similar optimal parameters with Top-10. The difference among the parameter selection for different evaluation parameters may be explained as follows. With wider  $W$  and larger  $\Delta$ , the sub-sequence tends to include more failure data. A duality exists where more data may help the model discover the failure pattern more easily or limit its extension ability. With more data, the model tends to converge on several certain failure patterns and provides more confidence in the failures. This explains why the Top-1 recall and the overlapping probabilities share the same optimal parameters. However, this kind of converge may neglect the other related failures. For the Top-10 recall, the most important criterion is the fraction of coverage, rather than one accurate hit. Training and predicting with relatively less data

focuses more on the mutual associations, which provides more insight into the hidden risk. Generally, the difference between the optimal parameters of Top-1 and Top-10 recalls reflects a dilemma between higher confidence and wider coverage in machine learning methods.

### 5.1.2. Comparison with Baselines

The best results were also compared with several baseline algorithms, including the statistical approach, Bayesian method, and the newly developed Long Short-Term Memory (LSTM) neural network. The statistical approach is the most common method for log analysis in power grids, which accounts for a large proportion in the annual report of power enterprises. A global average result that mainly focuses on the proportion of different failures is used to guide the production of the next year. The Bayesian method is one of the main approaches for distribution estimation. A sequential Dirichlet update initialized with the statistical average was conducted to provide a personalized distribution estimates for each HVCB. In past years, deep learning has exceeded the traditional methods in many areas. As one branch of deep learning for handling sequential data, LSTM has been applied to HVCB log processing. The key parameters of the LSTM include the embedding dimension of eight and the fully connected layer, having 100 units. Additionally, sequences shorter than 10 are padded to ensure a constant input dimension.

Table 4 reports the experimental result, where the model with the best performance is marked in bold font. The TLDA had the best performance for the Top-1, Top-5, and Top-10 tasks with 51.13%, 73.86%, and 92.93%, respectively, whereas the best overlapping probability was obtained by the Bayesian method. Although the Bayesian method obtained good overlapping probability and Top-1 recall, its Top-5 and Top-10 performances were the worst among the tested methods because the Bayesian method places too much weight on individual information and ignores the global correlations. On the contrary, the statistical approach obtained a slightly better result in Top-5 and Top-10 recall owing to the long tail distribution. However, its Top-1 recall was the lowest. The unbalanced datasets create a problem for the LSTM for obtaining high Top-1 recall. However, the LSTM still demonstrated its learning ability as reflected in its Top-5 and Top-10 recalls.

**Table 4.** Performance comparison with different methods.

Method	Top-1 (%)	Top-5 (%)	Top-10 (%)	$P_o$ (%)
TLDA	51.13	73.86	92.93	31.50
Statistical Approach	19.79	58.33	78.12	5.87
Bayesian Sequential Method	41.67	55.21	65.63	33.96
Neural Network(LSTM)	32.29	67.70	81.25	15.52

## 5.2. Qualitative Analysis

### 5.2.1. Failure Patterns Extraction

As mentioned before, the LDA method treats each failure sequence as a mixture of several failure patterns. Some interesting failure modes and failure associations can be mined by visualizing the failures. For simplicity, failure patterns of 10 were adopted to train a new TLDA model. Table 5 lists the failures that account for more than 1% in each failure pattern. All the failure patterns were extracted automatically and the titles were summarized afterward. Notably, error records may exist, the most common of which was the confusion between causes and phenomena. For example, various failure categories can be mistaken for operating mechanism failure as that the mechanism is the last step of a complete HVCB operation. A summary of the extracted failure patterns is as follows.

Table 5. Top failures in each failure pattern.

1. Operation Error by Machinery Parts	2. Operation Error by Driving System	3. Operation Error by Tripping Coils	4. Cubicles and Its Auxiliary System
Operating mechanism Assistive component damage High voltage indicating device SF6 leakage Electromotor stalling Travel switch	Electromotor stalling Travel switch Relay Reason Unidentified Electromotor on file Safe-blocked circuit Closing instruction Operating mechanism	Tripping and closing coil Secondary cubicle Operating mechanism Humidity ovenproof Mechanism cubicle Safe-blocked circuit	Mechanism cubicle High temperature Secondary cubicle Insulator Closing instructions Auxiliary switch Incomplete installation High voltage indicating device Operating mechanism Travel switch
5. SF6 Leakage	6. Operation Error by Secondary System	7. Pneumatic Mechanism	8. Measuring System
SF6 leakage	Remote control signal Auxiliary switch Rejecting action Operating mechanism Tripping and closing coil Gas pressure meter	Pneumatic mechanism leakage Poor contact Misjudgment Air compressor stalling Air compressor leakage Mechanism cubicle SF6 leakage Remote control signal Relay Electromotor stalling	Closing instructions High voltage indicating device Operation counter False wiring Transmission bar Operating mechanism SF6 leakage Gas pressure meter SF6 constituents Tripping and closing coil
9. Secondary System		10. Hydraulic Mechanism	
Contactator Safe-blocked circuit Air compressor stalling Main circuit High voltage indicating device SF6 constituents False wiring Contactator		Hydraulic mechanism leakage Closing instructions SF6 leakage Operating mechanism	

Failure pattern 1 mainly contains the operating mechanism's own failures, while pattern 2 reveals the co-occurrence of the operating mechanism within the driving system. Analogously, pattern 3 and pattern 6 mainly focus on how the operation may be broken by the tripping coils and secondary parts such as remote control signal. Pattern 7 and pattern 10 cluster the failures of pneumatic and hydraulic mechanism together. The other patterns also show different features. Different failure patterns have special emphasis and overlap. For example, though both contain secondary components, pattern 9 only considers their manufacturing quality, while pattern 6 emphasizes the interaction between the secondary components and the final operation.

### 5.2.2. Temporal Features of the Failure Patterns

The average value of  $\theta$  in different time slices can be calculated as a function of time to show the average variation tendency of different failure patterns. As shown in Figure 12, the failure modes of hydraulic mechanism, pneumatic mechanism and cubicles increase along with operation years, while the percentage of the measuring system, tripping and closing coils decrease. The SF6 leakage and machinery failures always share a large portion. The rise and fall of different failure patterns reflect the dynamic change of the device state, which is useful for targeted action scheduling.

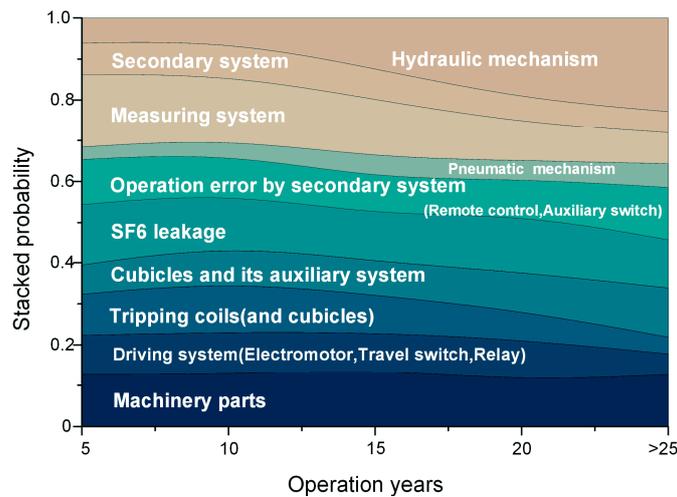


Figure 12. Average time-varying dynamics of the extracted 10 failure patterns.

Additionally, the concentration can be placed on one sequence to determine how each event change the mixture of the failure modes. Figure 13 shows the failure mode variation of the sample. At first, the SF6 leakage and the cubicle failures allocates a large portion to the corresponding modes. Then, the contactor failure improves the failure pattern of the secondary system. Afterward, the operation mechanism creates a peak in the pattern of machinery parts. However, its shares are quickly replaced by the failure mode of the tripping coils. This can be considered as the model’s self-correction to distinguish failures caused by the operating mechanism itself or its preorder system. At last, the remote control failure causes a portion shift from the failure mode of the secondary system to the operation error by secondary system.

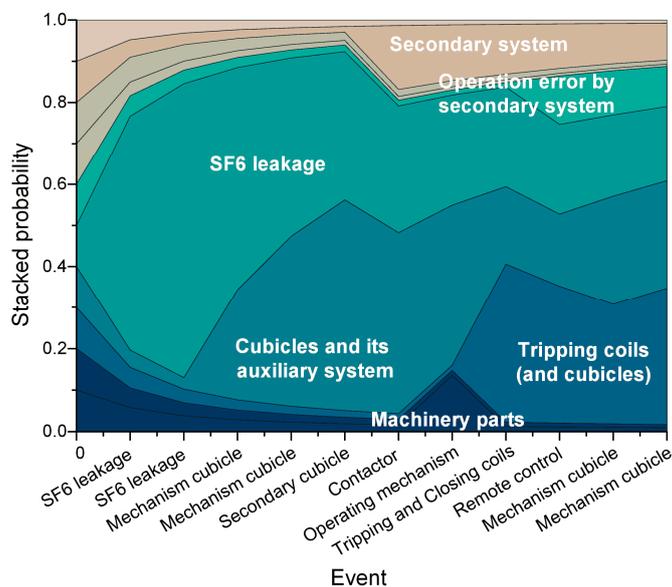


Figure 13. Time-varying dynamics of the failure patterns for an individual HVCB.

### 5.2.3. Similarities between Failures

The similarities between different failures based on Equation (18) are shown in Figure 14. A wealth of associations can be extracted combined with the equipment structure knowledge. In general, the failures with high similarities can be classified into four types.

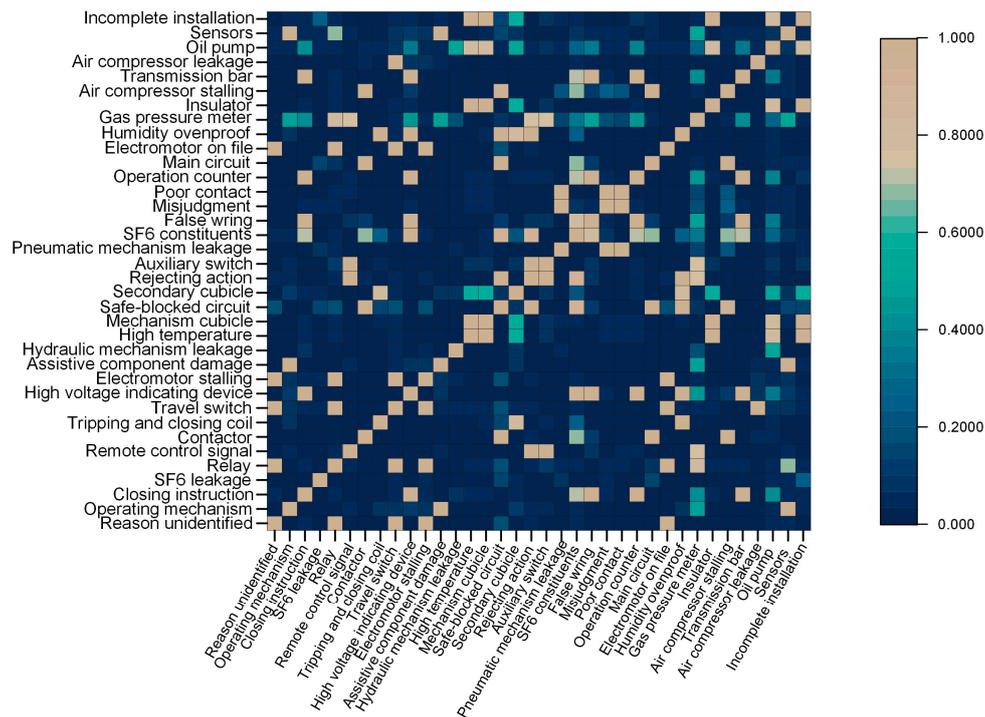


Figure 14. Similarity map for all the failures in the real-word dataset.

The first type is causal relationship, where the occurrence of one failure is caused by another. For example, the failure of a rejecting action may be caused by the remote control signal, safe-blocked circuit, auxiliary switch, SF6 constituents, and humidity ovenproof which may cause blocking according to the similarity map. The second type is wrong logging. Failures with wrong logging relationships often occur in a functional chain, facilitating wrong error location. The similarity between electromotor stalling and relay or travel switch failures, and the similarity between secondary cubicle and tripping coil may belong to this type. The third type is common cause failures. The failures are caused by similar reasons, such as the similarities among the measurement instruments, including the closing instructions, the high voltage indicating device, the operation counters, and the gas pressure meter. The strong association between the secondary cubicle and the mechanism cubicle may be caused by the deficient sealing, and a bad choice of motors assigns high similarity between the electromotor and oil pump. The fourth type is relation transmission. Similarities are built on indirect association. For example, the transmission bar has a direct connect to the operation counter, and the counter shares a similar aging reason with the other measurement instrument, making the transmission bar similar in number to the high voltage indicating device and the gas pressure meter. Likewise, the safe-blocked circuit may act as the medium between the air compressor stalling and SF6 constituents.

This similarity map may help establish a failure look-up table for fast failure reason analysis and location.

## 6. Conclusions and Future Work

In this paper, the event logs in a power grid were considered a promising data source for the goal of predicting future critical events and extracting the latent failure patterns. A TLDA framework is presented as an extension of the topic model, introducing a failure pattern layer as the medium between the failure sequences and the failures. The conjunction relation between the multinomial distribution and the Dirichlet distribution is embedded into the framework for better generalizations. Using a mixture of hidden variables for a failure representation not only enables pattern mining from the sparse data but also enables the establishment of quantitative relationships among failures. Furthermore, a simple but effective

temporal new co-occurrence pattern was established to introduce strict chronological order of events into the originally exchangeable Bayesian framework. The effectiveness of the proposed method was verified by thousands of real-word failure records of the HVCBs from both quantitative and qualitative perspectives. The Top-1, Top-5, and Top-10 results revealed that the proposed method outperformed the existing methods in predicting potential failures before they occurred. The parameter analysis showed a different parameter preference for higher confidence or a wider coverage. By visualizing the temporal structures of the failure patterns, the TLDA showed its ability to extract meaningful semantic characteristics, providing insight into the time variation and interaction of failures.

As future work, experiments can be conducted in other application areas. Furthermore, as a branch of the state space model, the attempt to use the trained TLDA embedding in the Recurrent Neural Network may provide better results.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (No. 51521065).

**Author Contributions:** Gaoyang Li and Xiaohua Wang conceived and designed the experiments; Mingzhe Rong provided theoretical guidance and supported the study; Kang Yang contributed analysis tools; Gaoyang Li wrote the paper; Aijun Yang revised the contents and reviewed the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hale, P.S.; Arno, R.G. Survey of reliability and availability information for power distribution, power generation, and HVAC components for commercial, industrial, and utility installations. *IEEE Trans. Ind. Appl.* **2001**, *37*, 191–196. [[CrossRef](#)]
- Lindquist, T.M.; Bertling, L.; Eriksson, R. Circuit breaker failure data and reliability modelling. *IET Gener. Transm. Distrib.* **2008**, *2*, 813–820. [[CrossRef](#)]
- Janssen, A.; Makareinis, D.; Solver, C.E. International Surveys on Circuit-Breaker Reliability Data for Substation and System Studies. *IEEE Trans. Power Deliv.* **2014**, *29*, 808–814. [[CrossRef](#)]
- Pitz, V.; Weber, T. Forecasting of circuit-breaker behaviour in high-voltage electrical power systems: Necessity for future maintenance management. *J. Intell. Robot. Syst.* **2001**, *31*, 223–228. [[CrossRef](#)]
- Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [[CrossRef](#)]
- Liu, H.; Wang, Y.; Yang, Y.; Liao, R.; Geng, Y.; Zhou, L. A Failure Probability Calculation Method for Power Equipment Based on Multi-Characteristic Parameters. *Energies* **2017**, *10*, 704. [[CrossRef](#)]
- Peng, Y.; Dong, M.; Zuo, M.J. Current status of machine prognostics in condition-based maintenance: A review. *Int. J. Adv. Manuf. Technol.* **2010**, *50*, 297–313. [[CrossRef](#)]
- Rong, M.; Wang, X.; Yang, W.; Jia, S. Mechanical condition recognition of medium-voltage vacuum circuit breaker based on mechanism dynamic features simulation and ANN. *IEEE Trans. Power Deliv.* **2005**, *20*, 1904–1909. [[CrossRef](#)]
- Rusek, B.; Balzer, G.; Holstein, M.; Claessens, M.S. Timings of high voltage circuit-breaker. *Electr. Power Syst. Res.* **2008**, *78*, 2011–2016. [[CrossRef](#)]
- Natti, S.; Kezunovic, M. Assessing circuit breaker performance using condition-based data and Bayesian approach. *Electr. Power Syst. Res.* **2011**, *81*, 1796–1804. [[CrossRef](#)]
- Cheng, T.; Gao, W.; Liu, W.; Li, R. Evaluation method of contact erosion for high voltage SF6 circuit breakers using dynamic contact resistance measurement. *Electr. Power Syst. Res.* **2017**. [[CrossRef](#)]
- Tang, J.; Jin, M.; Zeng, F.; Zhou, S.; Zhang, X.; Yang, Y.; Ma, Y. Feature Selection for Partial Discharge Severity Assessment in Gas-Insulated Switchgear Based on Minimum Redundancy and Maximum Relevance. *Energies* **2017**, *10*, 1516. [[CrossRef](#)]
- Gao, W.; Zhao, D.; Ding, D.; Yao, S.; Zhao, Y.; Liu, W. Investigation of frequency characteristics of typical pd and the propagation properties in gis. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 1654–1662. [[CrossRef](#)]
- Yang, D.; Tang, J.; Yang, X.; Li, K.; Zeng, F.; Yao, Q.; Miao, Y.; Chen, L. Correlation Characteristics Comparison of SF6 Decomposition versus Gas Pressure under Negative DC Partial Discharge Initiated by Two Typical Defects. *Energies* **2017**, *10*, 1085. [[CrossRef](#)]

15. Huang, N.; Fang, L.; Cai, G.; Xu, D.; Chen, H.; Nie, Y. Mechanical Fault Diagnosis of High Voltage Circuit Breakers with Unknown Fault Type Using Hybrid Classifier Based on LMD and Time Segmentation Energy Entropy. *Entropy* **2016**, *18*, 322. [[CrossRef](#)]
16. Wang, Z.; Jones, G.R.; Spencer, J.W.; Wang, X.; Rong, M. Spectroscopic On-Line Monitoring of Cu/W Contacts Erosion in HVCBs Using Optical-Fibre Based Sensor and Chromatic Methodology. *Sensors* **2017**, *17*, 519. [[CrossRef](#)] [[PubMed](#)]
17. Tang, J.; Zhuo, R.; Wang, D.; Wu, J.; Zhang, X. Application of SA-SVM incremental algorithm in GIS PD pattern recognition. *J. Electr. Eng. Technol.* **2016**, *11*, 192–199. [[CrossRef](#)]
18. Liao, R.; Zheng, H.; Grzybowski, S.; Yang, L.; Zhang, Y.; Liao, Y. An integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning. *IEEE Trans. Power Deliv.* **2011**, *26*, 1111–1118. [[CrossRef](#)]
19. Jiang, T.; Li, J.; Zheng, Y.; Sun, C. Improved bagging algorithm for pattern recognition in UHF signals of partial discharges. *Energies* **2011**, *4*, 1087–1101. [[CrossRef](#)]
20. Mazza, G.; Michaca, R. The first international enquiry on circuit-breaker failures and defects in service. *Electra* **1981**, *79*, 21–91.
21. International Conference on Large High Voltage Electric Systems; Study Committee 13 (Switching Equipment); Working Group 06 (Reliability of HV circuit breakers). *Final Report of the Second International Enquiry on High Voltage Circuit-Breaker Failures and Defects in Service*; CIGRE: Paris, France, 1994.
22. Ejnár, S.C.; Antonio, C.; Manuel, C.; Hiroshi, F.; Wolfgang, G.; Antoni, H.; Dagmar, K.; Johan, K.; Mathias, K.; Dirk, M. Final Report of the 2004–2007 International Enquiry on Reliability of High Voltage Equipment; Part 2—Reliability of High Voltage SF Circuit Breakers. *Electra* **2012**, *16*, 49–53.
23. Boudreau, J.F.; Poirier, S. End-of-life assessment of electric power equipment allowing for non-constant hazard rate—Application to circuit breakers. *Int. J. Electr. Power Energy Syst.* **2014**, *62*, 556–561. [[CrossRef](#)]
24. Salfner, F.; Lenk, M.; Malek, M. A survey of online failure prediction methods. *ACM Comput. Surv.* **2010**, *42*. [[CrossRef](#)]
25. Fu, X.; Ren, R.; Zhan, J.; Zhou, W.; Jia, Z.; Lu, G. LogMaster: Mining Event Correlations in Logs of Large-Scale Cluster Systems. In Proceedings of the 2012 IEEE 31st Symposium on Reliable Distributed Systems (SRDS), Irvine, CA, USA, 8–11 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 71–80.
26. Gainaru, A.; Cappello, F.; Fullop, J.; Trausan-Matu, S.; Kramer, W. Adaptive event prediction strategy with dynamic time window for large-scale hpc systems. In Proceedings of the Managing Large-Scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques, Cascais, Portugal, 23–26 October 2011; ACM: New York, NY, USA, 2011; p. 4.
27. Wang, F.; Lee, N.; Hu, J.; Sun, J.; Ebadollahi, S.; Laine, A.F. A framework for mining signatures from event sequences and its applications in healthcare data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 272–285. [[CrossRef](#)] [[PubMed](#)]
28. Macfadyen, L.P.; Dawson, S. Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Comput. Educ.* **2010**, *54*, 588–599. [[CrossRef](#)]
29. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 25–28 May 1993; ACM: New York, NY, USA, 1993; pp. 207–216.
30. Li, Z.; Zhou, S.; Choubey, S.; Sevenpiper, C. Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE Trans.* **2007**, *39*, 303–315. [[CrossRef](#)]
31. Fronza, I.; Sillitti, A.; Succi, G.; Terho, M.; Vlasenko, J. Failure prediction based on log files using random indexing and support vector machines. *J. Syst. Softw.* **2013**, *86*, 2–11. [[CrossRef](#)]
32. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *Comput. Sci.* **2013**.
33. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
34. Guo, C.; Li, G.; Zhang, H.; Ju, X.; Zhang, Y.; Wang, X. Defect distribution prognosis of high voltage circuit breakers with enhanced latent Dirichlet allocation. In Proceedings of the Prognostics and System Health Management Conference (PHM-Harbin), Harbin, China, 9–12 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7.
35. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *Adv. Neural Inf. Process. Syst.* **2014**, *4*, 3104–3112.

36. Anderson, C. *The Long Tail: Why the Future of Business Is Selling Less of More*; Hachette Books: New York, NY, USA, 2006.
37. Pinoli, P.; Chicco, D.; Masseroli, M. Latent Dirichlet allocation based on Gibbs sampling for gene function prediction. In Proceedings of the 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, Honolulu, HI, USA, 21–24 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–8.
38. Wang, X.; Grimson, E. Spatial Latent Dirichlet Allocation. In Proceedings of the Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1577–1584.
39. Maskeri, G.; Sarkar, S.; Heafield, K. Mining business topics in source code using latent dirichlet allocation. In Proceedings of the India Software Engineering Conference, Hyderabad, India, 9–12 February 2008; pp. 113–120.
40. Golub, G.H.; Van Loan, C.F. *Matrix Computations*; Johns Hopkins University Press: Baltimore, MD, USA, 1983; pp. 392–396.
41. Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37. [[CrossRef](#)]
42. Blei, D.M.; Lafferty, J.D. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; ACM: New York, NY, USA, 2006; pp. 113–120.
43. Du, L.; Buntine, W.; Jin, H.; Chen, C. Sequential latent Dirichlet allocation. *Knowl. Inf. Syst.* **2012**, *31*, 475–503. [[CrossRef](#)]
44. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. [[CrossRef](#)]
45. Wei, X.; Croft, W.B. LDA-based document models for ad-hoc retrieval. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 6–10 August 2006; ACM: New York, NY, USA, 2006; pp. 178–185.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).