Difference between Similars: a Novel Method to Use Topic Models for Sensor Data Analysis

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

***Abstract*—We propose a novel method to use the topics obtained by topic modeling for sensor data analysis. This paper describes a case study where we perform an exploratory data analysis of manufacturing sensor data by using latent Dirichlet allocation (LDA) as a tool to discover remarkable change patterns. Our target is a set of time-series data originating from the sensors installed in a closed factory environment. Each sensor gives a different type of measurement of the same manufacturing process, which is operated repeatedly in a lot-by-lot manner. We first discretize the data based on the histogram of sensor measurements and construct a bag-of-words representation. We then apply LDA to discover change patterns across tens of thousands of lots. When we apply LDA to natural language documents, the resulting topics are widely different from each other because the documents intrinsically show considerable diversity. In contrast, our data, which come from the repeatedly operated manufacturing process, only show limited diversity. As a result, LDA provides topics closely similar to each other. Our main and unexpected finding is that the difference between similar topics is useful in discovering remarkable change patterns. We performed an experiment over the data sets containing sensor measurements collected in the factory. The results have revealed that subtle difference between very similar topics often corresponds to an interesting change pattern of sensor measurements.**

***Index Terms*—Sensors, Data analysis, Text mining**

I. Introduction

A. LDA-Based Analysis of Bag-of-Words Sensor Data

Sensor data analysis is one among the most important applications of machine learning in IoT [12], [13]. This paper considers large data sets provided by the sensors installed in a closed manufacturing environment. A distinctive feature of such sensor data is that they show apparent regularities because the data originate from a production process repeatedly operated in a lot-by-lot manner.[1] Therefore, the main issue in the analysis of such sensor data is to detect deviations from regularly observed behaviors of the sensors. Such deviations are critical because they can lead to severe production defects.

Our primary research question is thus how to discover such deviations in a large amount of sensor measurements generated by the daily processing of tens of thousands of product lots. We address this question with topic models [1]. We only consider latent Dirichlet allocation (LDA) [2] in this paper. However, other variants of LDA, e.g. Author Topic model [5], [11], Structural Topic Model [10], etc., may also be adopted.

LDA extracts a specified number of *topics* from a document set. Each topic is represented as a probability distribution defined over vocabulary. When the words like “election”, “Democrats”, “Republican”, etc. have high probabilities in one among the extracted topics and the words like “curveball”, “strikeout”, “inning”, etc. have high probabilities in another, we can know that some documents discuss politics and others baseball. LDA further provides the probabilities of topics in each document. When in some document the probability of a topic in which the words like “singer”, “stage”, “Billboard”, etc. have high probabilities is larger than those of all other topics, we can know that the document mainly talks about music. In short, LDA represents each topic as a word probability distribution and each document as a topic probability distribution, and we can get important information about the data set by combining the distributions. Since LDA is an unsupervised method, we can adopt it even when we have little clear prior knowledge about the data set. The unsupervised nature of LDA is an advantage for sensor data analysis because sensor data often have no intuitive meaning for humans.

The target of our LDA-based exploratory data analysis is sensor data. Before applying LDA, we need to convert the data set into a document set. Therefore, we first construct a vocabulary by discretizing sensor outputs. Our words are a 3-tuple consisting of a sensor ID, a discretized output value, and a symbol telling whether the output value at the next time point is smaller than (’$>$’), equal to (’$=$’), or larger than (’$<$’) the current one, e.g. “(sensor Q, -30, $<$)”.[2] We next obtain bag-of-words documents by regarding the sensor outputs coming from every single lot as forming a single document. Therefore, we have as many documents as processed lots. In this paper, we analyze three document sets obtained at different sites in the factory. Each set contains around 40,000 documents composed from the outputs of about 30 sensors.[3]

[1] In this paper, the word *lot* means the smallest unit of products manufactured in the factory. A single lot may consist of multiple items that are processed simultaneously in the same run.

[2] The comparison is done by using the raw outputs before discretization.

[3] The details of the sensors cannot be disclosed. We thus have shown the data specification in round numbers. The length of the time period we collected the data is also not disclosed. We get rid of the unit of the sensor measurements and further apply a random 1D affine transformation.

B. Example of Exploratory Data Analysis with LDA

We here show an example of our analysis to explain the outline of the proposed method. Fig. 1 presents a remarkable change pattern we have discovered for the sensor Q[4] in one of the three data sets we consider. The horizontal axis is the time axis. The vertical axis gives the sensor output. Fig. 1 shows a change pattern, where the sensor outputs fall

- in the interval $[-30, 30]$ before Day A,
- in the interval $[-30, 10]$ from Day A to Day B, and
- in the interval $[20, 60]$ after Day B.[5]

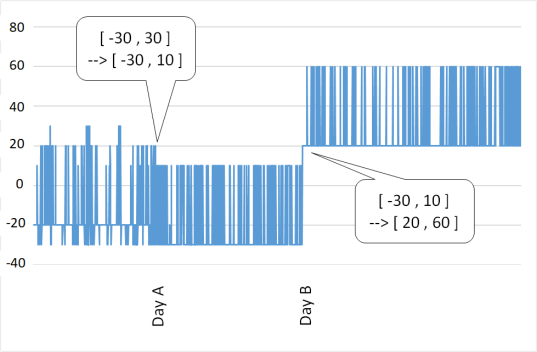

Fig. 1. The first example of a discovered change pattern.

We have found this change pattern with the topics given by LDA in a manner explained in the following paragraphs.

Fig. 2 presents the change of the probabilities of the three topics, whose IDs are #04, #24, and #18 from top to bottom, among the 30 topics extracted by LDA. We will explain in Section IV how to set the number of topics. The horizontal axis in Fig. 2 is the time axis. The vertical axis gives the topic probability in each document. The three charts in Fig. 2 are obtained for the same time period as that of Fig. 1 after sorting the documents in chronological order along the time axis. Days A and B in Fig. 2 are the same with Days A and B in Fig. 1 respectively. We have picked up these three topics #04, #24, and #18 because only the probabilities of them change remarkably in this period. We can observe the followings:

- The probability of the topic #04 is nearly equal to 0.1 after Day A while that of the topic #24 is almost equal to 0.0 after Day A.
- The probability of the topic #04 is almost negligible after Day B while that of the topic #18 takes off from zero after Day B.

The above two observations have led to the discovery of the change pattern depicted in Fig. 1 as follows. We have computed the word probability differences between the topics #04 and #24 and then have found that the word “(sensor Q, -30, $<$)” gives the largest absolute difference. We also have computed the word probability differences between the topics #04 and #18 and then have found that the word “(sensor Q, -40, $<$)” gives the largest absolute difference. Consequently, we have guessed that the output values around -40 or -30 of the sensor Q are important. Moreover, it has turned out that the topics #24 and #18 are the top closest and the second closest to the topic #04 respectively in L_1-distance.

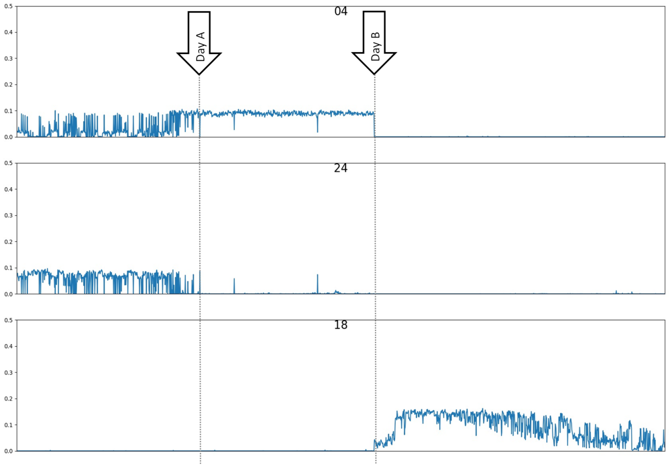

Fig. 2. Topic probability change revealing the change pattern in Fig. 1.

We next go back to the outputs of the sensor Q. The processing of each single lot produces a series of outputs at the sensor Q as depicted in Fig. 3. The horizontal axis gives a time interval of a few minutes because every lot is processed in a few minutes. The vertical axis gives the sensor Q’s output. We obtain similar waveforms from the sensor Q also for other lots. Since the output values around -40 or -30 appear only in the interval enclosed by the red oval, we have guessed that the lot-by-lot changes of the sensor outputs in this interval correspond to the document-by-document changes of the topic probabilities in Fig. 2. Therefore, we get the most frequent value from the interval enclosed by the red oval as a representative value and plot it for the same time period as that of Fig. 2. As a result, we have obtained Fig. 1. That is, the most frequent value in the enclosed area of Fig. 3 changes as presented in Fig. 1 over thousands of lots. In this manner, we have come across the change pattern in Fig. 1.

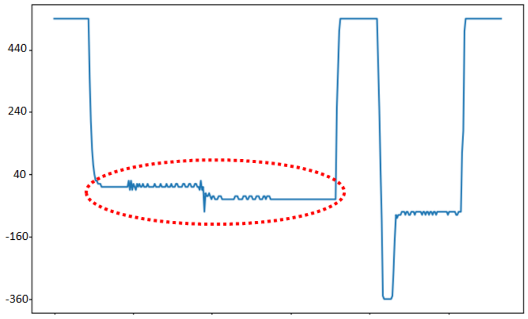

Fig. 3. Example of a per-lot sequence of the sensor Q’s outputs.

In our experiment we have often encountered remarkable

[4] In this paper, we anonymize the sensors by using alphabet letters.

[5] We also anonymize the actual dates.

change patterns similar to that in Fig. 1 by computing the *difference* of word probabilities between *very similar* topics. In the above example the difference between the topics #04 and #24 and that between #04 and #18 have told us which output values of which sensor we should focus on. Considering the difference between similar topics is a novel way to use the topics extracted by LDA and is not a standard way in case of natural language documents because the extracted topics show a considerable diversity. If we obtain closely similar topics when analyzing natural language documents with LDA, we may conclude that the number of topics is too large and reduce the number of topics to remove redundant topics. In contrast, the periodical regularity intrinsic to factory sensor data makes us hesitate to adopt this standard way because the extracted topics often show a non-negligible similarity. Our important and unexpected finding is that the topic similarity often conveys important information. Therefore, we have proposed a novel method using the difference between similar topics. We call our approach *difference-between-similars* approach.

The remainder of the paper is structured as follows. Section II discusses related work. Section III presents the data discretization method, the details of LDA data modeling, and our novel way to use the topics. Section IV provides the data specifications and two more examples of remarkable change pattern, similar to that illustrated in Subsection I-B. The conclusions of this work are given in Section V.

II. Related Work

Due to the growing interest in IoT, we can find many prominent proposals of the applications of topic modeling to sensor data analysis. In [3], the authors apply LDA to room occupancy logs generated by the sensor network installed in an office environment to find interesting occupancy patterns. While the occupancy states are represented in binary, the timestamps of the occupancy states are continuous data. Therefore, the authors discretize them by dividing the time period under consideration into several segments. The words in the vocabulary for LDA are then constructed by combining the binary occupancy states and the discretized time intervals. Another proposal [4] adopts LDA to learn activity routines from the sensor network data similarly gathered in an office environment. Since the timestamps are continuous also in this work, the authors obtain a bag-of-words representation by discretizing them. As shown in these proposals, when we have continuous numerical data, we need to discretize them in a way depending on the characteristics of the data. Such preprocessing aims to obtain a bag-of-words representation of the data to which we apply LDA.

In the proposals given above, only the timestamp data are continuous. However, in [9], the outputs of the sensor are also continuous. Therefore, the authors discretize them by using nonoverlapping bins. Interestingly, this work also performs a fault detection over a test set by computing KL-divergence between the test data word distributions and the per-topic word distributions. This is a fascinating way to use the analysis results achieved with LDA. However, it is beyond the scope of our current work to devise a defect detection utilizing the discoveries our difference-between-similars method provides. In [8], the data collected from the wearable sensors are analyzed with LDA to discover remarkable human activity patterns. Since the sensor outputs continuous data, the authors give annotations with tens of labels to the data. In our case, the sensor data are discretized not by assigning annotation but by using nonoverlapping bins as in [9].

Depending on the nature of the data, we may adopt an extended version of LDA as in [5], where the model called Author-Topic Model [11] is adopted. The model proposed in [6] extends LDA to represent word sequences, where each word is a pair of a place visited by mobile phone users and a discretized time interval. This model, called DNTM, represents word sequences by using multiple word probability distributions. Although we only consider the vanilla LDA in this paper, it is a promising future research direction to adopt an existing extension of LDA or to propose a new topic model by extending LDA for explicitly representing important aspects of the sensor data that the vanilla LDA cannot address.

All the above proposals interpret the high probability words in each topic as corresponding to some remarkable patterns. The data considered in these proposals show no repetitive regularity similar to that observable in our factory sensor data. When the data shows a great diversity as in the case of natural language documents, the topics LDA provides also show considerable variety. That is, different topics have widely different high probability words. Therefore, we can use such words to focus on a specific part of the data set. For example, the sensor data analyzed in [9] show no great regularity like that we can observe in our factory data because the data are obtained in an open environment from an autonomous underwater vehicle moving. Therefore, the high probability words in each topic can be regarded as corresponding to noticeable behaviors of the underwater vehicle.

However, when the data shows considerable regularity, the extracted topics also show regularity. That is, we obtain multiple topics whose word probability distributions are closely similar to each other. If we obtain closely similar topics for natural language documents or for the sensor data collected in an open environment, the similarity is interpreted as redundancy. When we have redundant topics, we reduce the number of topics and rerun LDA over the same document set. However, in case of the data showing a considerable regularity, we can *utilize* the similarity between topics. Our contribution is to provide a procedure for utilizing the topic similarity to discover remarkable change patterns for the sensor data collected in a closed factory environment.

III. Method

A. Data Discretization

LDA is a method originally proposed for a set of documents each represented as a bag-of-words, i.e., a multiset of words. When we apply LDA to other types of data, we need to obtain a bag-of-words representation by defining a vocabulary. For continuous value data, we construct the vocabulary by using symbolic representations each corresponding to a bin, i.e., a specific range of values. The width of the bins should be determined carefully. Too narrow a bin width results in too many words in the vocabulary, and too broad a bin width results in an oversimplification of sensor outputs. Therefore, we visualize the frequencies of all possible output values of each sensor with histogram and find the best bin width by modifying the bin width manually. Since the number of sensors we consider was around 30, the manual decision of the bin width was not a burden. Fig. 4 presents the histogram obtained for the sensor Q considered in Subsection I-B.[6]

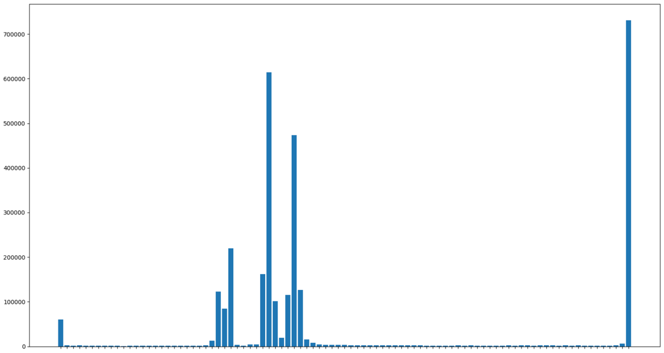

Fig. 4. Histogram of the sensor Q's outputs.

In this paper, we make all bins have the same width for the same sensor. One possible future research direction is to determine the width of each bin adaptively based on the frequencies of the output values. After discretizing the sensor outputs, we compose a vocabulary for LDA. In this paper, we propose to compose each word as a 3-tuple consisting of a sensor ID, a discretized output value, and a symbol telling whether the output value at the next time point is smaller than ('$>$'), equal to ('$=$'), or larger than ('$<$') the current one, e.g. "(sensor X, -8.4, $>$)". The comparison is made for the output values before discretization. The third entry of each word thus represents a local change of the raw sensor readings.

B. Latent Dirichlet Allocation

We use latent Dirichlet allocation (LDA) to analyze our sensor data sets. Its precise details can be found in the original paper [2]. We here explain the input and the output of LDA.

LDA assumes that each document is represented as a bag-of-words. Therefore, the ordering of words is not relevant to LDA. Only the frequencies of words in each document are relevant. In this paper, the 3-tuples defined in Section III-A are used as words. Let $n_{i,j}$ denote the frequency of the word w_j in the document d_i. Bag-of-words documents are given as a vector whose entries are word frequencies, i.e., $(n_{i,1}, \dots, n_{i,J})$, where J is the vocabulary size. In our experiment, the sensor outputs collected for a single lot are regarded as composing a single document. Therefore, we have as many documents as lots processed in the factory for the time period we consider.

By applying LDA to a given document set, we can obtain per-topic word probabilities and per-document topic probabilities. We denote the probability of the word w_j in the topic t_k by $\phi_{k,j}$. For each k, $\sum_{j=1}^{J} \phi_{k,j} = 1$ holds. We denote the probability of the topic t_k in the document d_i by $\theta_{i,k}$. For each i, $\sum_{k=1}^{K} \theta_{i,k} = 1$ holds, where K is the number of topics. K needs to be specified manually before we use LDA.

Applying LDA means performing an inference to estimate the $\phi_{k,j}$'s and the $\theta_{i,k}$'s. In this work we perform the mini-batch variational inference proposed in [7] by implementing it with PyTorch.[7] The details are given in Algorithm 1, where $\Psi(\cdot)$ is the digamma function. The variational inference estimates model parameters by maximizing the evidence lower bound (ELBO) [2]. In this paper, we assign a symmetric Dirichlet prior distribution Dirichlet(α) to the per-document topic probabilities $\theta_{i,k}$ and estimate the parameters of the corresponding Dirichlet posterior Dirichlet$(\boldsymbol{\lambda}_i)$ for each document d_i. On the other hand, we estimate the per-topic word probabilities $\phi_{k,j}$ directly via the maximization of ELBO without assigning any prior to them. Further, additional posterior parameters $\gamma_{i,j,k}$ are introduced for representing the probability that an occurrence of the word w_j in the document d_i is assigned to the topic t_k. The detailed discussion of the variational inference is referred to [2] and [7]. The ELBO to be maximized in our mini-batch variational inference is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\gamma}; \alpha) \equiv & \sum_{i,j,k} n_{i,j} \gamma_{i,j,k} \log \phi_{k,j} \\
& + \sum_{i,k} \Big(\alpha + \sum_j n_{i,j} \gamma_{i,j,k} - \lambda_{i,k} \Big) \Big\{ \Psi(\lambda_{i,k}) - \Psi\Big(\sum_{k'} \lambda_{i,k'} \Big) \Big\} \\
& - \sum_{i,j,k} n_{i,j} \gamma_{i,j,k} \log \gamma_{i,j,k} + \log \Gamma(K\alpha) - K \log \Gamma(\alpha) \qquad (1)
\end{aligned}
$$

where $\Gamma(\cdot)$ is the gamma function.

Algorithm 1 Mini-batch variational inference for LDA

```
Input: $(n_{i,1}, \dots, n_{i,J})$ for all $i$ and the hyperparameter $\alpha$
Initialize a $J \times K$ matrix $\boldsymbol{W}$ and a $J$-dimensional vector $\boldsymbol{b}$ randomly
repeat
    $\boldsymbol{\phi}_k \leftarrow$ Softmax$(\boldsymbol{W}\boldsymbol{e}_k + \boldsymbol{b})$ for all $k$
    Get next mini-batch
    for each document $d_i$ in mini-batch do
        Initialize $\gamma_{i,j,k}$ randomly
        repeat
            $\lambda_{i,k} \leftarrow \alpha + \sum_j n_{i,j}\gamma_{i,j,k}$
            $\gamma_{i,j,k} \leftarrow\propto \exp\big(\Psi(\lambda_{i,k})\big) \times \phi_{k,v}$
        until change in $\lambda_{i,k}$ is negligible
    end for
    Make computational graph of the negative of ELBO
    Backpropagate
    Update $\boldsymbol{W}$ and $\boldsymbol{b}$
until change in ELBO is negligible
```

[6] We have erased the horizontal axis values to hide the exact value range.

[7] https://pytorch.org/

In Algorithm 1, we parameterize the per-topic word probabilities $\phi_{k,j}$ as $\boldsymbol{\phi}_k \equiv \text{Softmax}(\boldsymbol{W}\boldsymbol{e}_k + \boldsymbol{b})$ for each k, where $\boldsymbol{e}_k$ is the k-th standard basis. When compared to the case where we parameterize $\boldsymbol{\phi}_k$ simply as $\boldsymbol{\phi}_k \equiv \text{Softmax}(\boldsymbol{W}\boldsymbol{e}_k)$, the bias vector $\boldsymbol{b}$ provides an effect similar to that achievable by smoothing in language modeling. By using the per-document Dirichlet posterior parameters $\lambda_{i,k}$ given by Algorithm 1, we can obtain an estimation of the per-document topic probabilities $\theta_{i,k}$ as $\theta_{i,k} \equiv \lambda_{i,k} / \sum_{k'} \lambda_{i,k'}$, which is the mean of the Dirichlet posterior distribution.

C. Difference-between-Similars Approach

We use the $\theta_{i,k}$ and the $\phi_{k,j}$ given by Algorithm 1 in our difference-between-similars approach as follows.

First, we sort all documents in chronological order and plot the $\theta_{i,k}$ along the time axis for each topic t_k, $k = 1, \ldots, K$. We can visualize the change pattern of topic probabilities as in Fig. 5, which presents an example of a change pattern of the 30 topics for a small part of the time period we consider in our experiment. We often observe that two topics change their probabilities in a synchronized manner as enclosed by the red oval in Fig. 5. In most cases, such two topics are closely similar to each other in terms of L_1-distance. That is, $\sum_{j=1}^{J} |\phi_{k,j} - \phi_{k',j}|$ is often small for the topics t_k and $t_{k'}$ when they change their probabilities in a synchronized manner. Moreover, we sometimes observe that a single topic shows an interesting trend in its probability (cf. Fig. 10). For such a topic, we pick up the one or two topics closest to it in terms of L_1-distance.

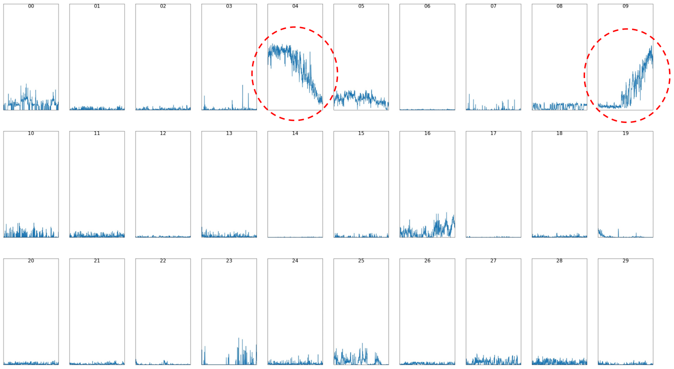

Fig. 5. Example of a change pattern of all topics ($K = 30$).

Second, for a pair of two closely similar topics t_k and $t_{k'}$, we compute the absolute probability difference $|\phi_{k,j} - \phi_{k',j}|$ for all words w_j, $j = 1, \ldots, J$ and sort them in the decreasing order to select a few top-ranked words. Recall that our words are a 3-tuple consisting of a sensor ID, a discretized output value, and a symbol telling whether the output value at the next time point is smaller than ('$>$'), equal to ('$=$'), or larger than ('$<$') the current one (cf. Section III-A). The three entries of the words giving a large absolute difference provide a clue to discover remarkable change patterns. That is, the three entries tell us what output value of what sensor we should focus on.

The practitioners using LDA have so far focused only on the words having high probabilities in each topic separately. The word probabilities of closely similar topics have not been compared as far as we know. Therefore, we can say that our difference-between-similars approach has discovered a new way to use LDA. In Section IV-B we will illustrate the difference-between-similars approach with actual examples.

IV. Experimental Results

A. Experimental Settings

Our experiment used three data sets I, II, and III, which we collected at the three different sites of the factory. The result presented in Subsection I-B was obtained for the data set III. The data came from around 30 different sensors, each of which provided a sequence of approximately 350 outputs for every single lot. Therefore, we had around $30 \times 350 \approx 10,000$ word tokens in each document because we composed one document for each lot by discretizing the sensor outputs as explained in Section III-A. Our factory processed around 40,000 lots during the period we collected data. Therefore, each of the three data sets contains around 40,000 documents. The vocabulary size is around 2,500 for all data sets because the same types of sensors are used at all of the three sites of the factory.

We first applied LDA to each data set by setting the number of topics K as 50. However, the probabilities of many topics were almost zero for all documents. Therefore, we reduced the topic number K to 30. This number might be considered to be reasonable because the number of the data sources, i.e., the number of the sensors, is around 30. The results presented in this paper were thus obtained for $K = 30$. We ran the mini-batch variational inference in Algorithm 1 by setting the mini-batch size as 50. Every epoch took around five minutes on NVIDIA GeForce GTX 1060. We ran the variational inference for 200 epochs. The symmetric Dirichlet hyperparameter α was set as 0.01. The mini-batch size, the epoch number, and the Dirichlet hyperparameter were determined by the evaluation in perplexity [4]–[6], [9].

B. Example of a Discovered Change Pattern

This subsection presents two more remarkable change patterns discovered by the difference-between-similars approach.

Fig. 6 presents a change pattern discovered for the two sensors L and X in the data set I. The horizontal axis is the time axis. The vertical axis gives the sensor output. The top panel plots the outputs of the sensor L. The output value increases gradually from Day A to Day C and keep staying at almost the same level after Day C. The bottom panel plots the outputs of the sensor X. The value increases gradually from Day A to Day B and keeps staying at almost the same level after Day B. This change pattern has been revealed by the probabilities of the two topics #04 and #09 depicted in Fig. 7. The horizontal axis represents the same time period with Fig. 6. The vertical axis gives the topic probability. We can observe the followings:

- The probability of the topic #04 decreases after Day B, and that of the topic #09 increases after this time point.

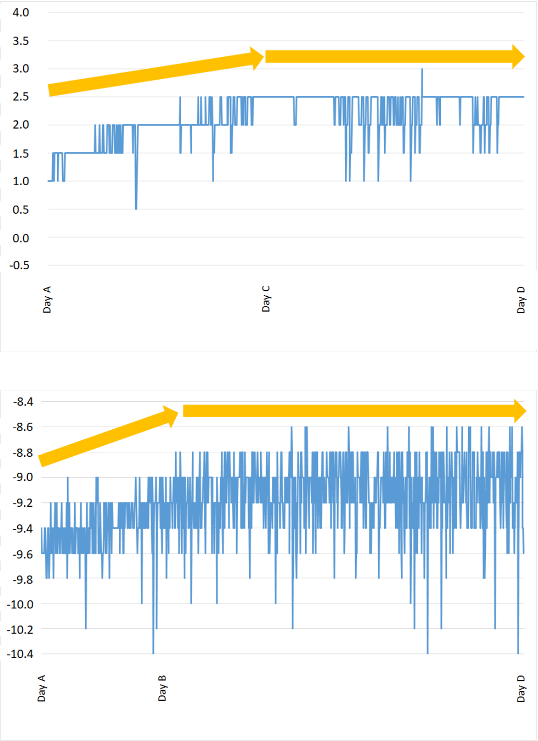

Fig. 6. The second example of a discovered change pattern.

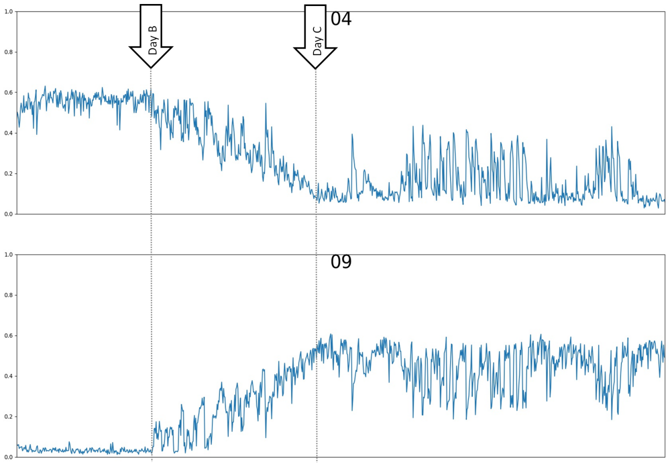

Fig. 7. Topic probability change revealing the change pattern in Fig. 6.

- Both the decrease of the topic #04's probability and the increase of the topic #09's probability stop on Day C.

Once we have found the topic probability change like that given in Fig. 7, we can use the difference-between-similars approach. By comparing the word probabilities in the topics #04 and #09, we have found that the two topics show a tiny difference. Table I gives the top five high probability words in each topic. The words are the same, and the probabilities are also almost the same. If this result were obtained for natural language documents, we would conclude that we got redundant topics due to too large a number of topics. However, mechanical regularity prevails in the factory sensor data. Therefore, LDA can give closely similar topics as in Table I. Even if two topics are similar to each other to this extent, we should not discard them. We should rather focus on their difference.

TABLE I
TOP FIVE HIGH PROBABILITY WORDS IN THE TOPICS #04 AND #09

topic #04		topic #09	
(sensor E, 28.2, >)	0.0514	(sensor E, 28.2, >)	0.0508
(sensor P, 0.3, >)	0.0513	(sensor P, 0.3, >)	0.0508
(sensor F, 57.0, >)	0.0477	(sensor F, 57.0, >)	0.0493
(sensor y, 8.0, >)	0.0418	(sensor y, 8.0, >)	0.0413
(sensor Z, -11.0, >)	0.0380	(sensor Z, -11.0, >)	0.0375

By computing the word probability differences, we have found that the following five words give the largest absolute differences between the topics #04 and #09: "(sensor X, -8.6, >)", "(sensor X, -8.4, >)", "(sensor X, -8.2, >)", "(sensor L, 2.0, >)", and "(sensor X, -8.8, >)". Note that these five words do not appear in Table I. If we only focused on the high probability words in each topic, we could not find these five words. The five words tell us that the sensors L and X are important. The sensor L gives an output sequence as shown in Fig. 8 for each lot. Since the word "(sensor L, 2.0, >)" gives a large probability difference, we search the values around 2.0 in Fig. 8. It turns out that the sensor L outputs values around 2.0 in the interval enclosed by the red oval. Therefore, we get the minimum value from this interval as a representative and plot it from Day A to Day D. The result is the top panel in Fig. 6. We perform a similar analysis also for the sensor X.

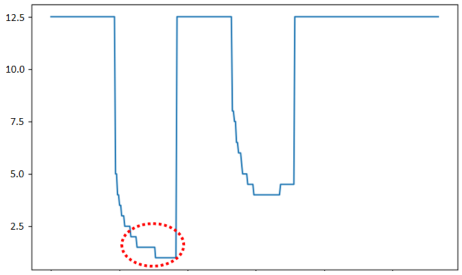

Fig. 8. Example of a per-lot sequence of the sensor L's outputs.

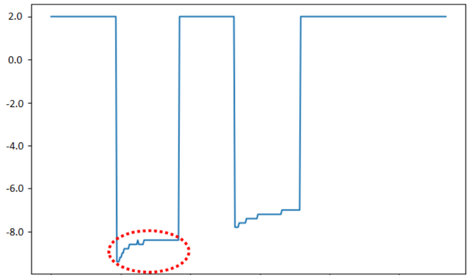

Fig. 9. Example of a per-lot sequence of the sensor X's outputs.

The sensor X gives an output sequence as shown in Fig. 9

for each lot. Since the words “(sensor X, -8.6, $>$)”, “(sensor X, -8.4, $>$)”, “(sensor X, -8.2, $>$)”, and “(sensor X, -8.8, $>$)” give a large probability difference, we search the values around -8.5 in Fig. 9. It turns out that the sensor X outputs values around -8.5 in the interval enclosed by the red oval. Therefore, we get the minimum value from this interval as a representative and plot it from Day A to Day D. The result is the bottom panel in Fig. 6. Our difference-between-similars approach has worked in this manner to reveal a remarkable change pattern given in Fig. 6.

The next example may be more interesting. After applying LDA to the data set II, we have found that the probability of the topic #03 shows a monotonically increasing trend over a relatively long period, as depicted in Fig. 10. We also have found that the topics #06 and #18 are the two most similar topics to the topic #03 in terms of L_1-distance. Both for the pair of the topics #03 and #06 and the pair of the topics #03 and #18, the two words “(sensor T, -6.0, $<$)” and “(sensor T, -6.0, $>$)” have shown a probability difference far larger than the other words. We thus have focused on the sensor T.

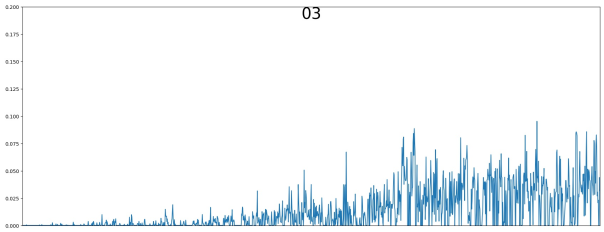

Fig. 10. Topic probability change revealing the change pattern in Fig. 12.

Fig. 11 presents an example of the output sequence of the sensor T for a single product lot. It can be observed that the sensor T only outputs three different values: -8.0, -6.0, and -4.0. By inspecting the sensor T’s outputs for all lots, it has been confirmed that this sensor only outputs these three values. It has also been confirmed that the frequency of the value -8.0 is the smallest among the three values. Therefore, we have guessed that the monotonically increasing trend in Fig. 10 corresponds to the frequency of the value -8.0. Fig. 12 presents per-lot frequencies of the value -8.0 of the sensor T. The horizontal axis represents the same time period with Fig. 10. The vertical axis gives how many times the sensor T outputs the value -8.0 during the processing of every single lot. As Fig. 12 shows, the sensor T rarely outputs -8.0 from Day A to Day B. However, the sensor T outputs -8.0 more than 50 times for many of the lots processed from Day C to Day D.

We could find many other remarkable change patterns by using the difference-between-similars approach, though we only provide three examples in this paper.

V. Conclusions

This paper provides a case study where we utilized LDA to analyze the factory sensor data in a special manner. Our contribution is to propose a novel approach, called *difference-between-similars* approach, to use the per-document topic probabilities and the per-topic word probabilities given by LDA. In this approach, we

- Find a pair of topics whose word probabilities are very similar to each other, and then
- Find the words whose probability differences between the two topics are larger than other words.

As depicted in Fig. 2 and in Fig. 7, a synchronized change of the per-document topic probabilities often suggest the existence of a pair of very similar topics. We then compute the difference of their word probabilities to find the dominant factors bringing about the difference between those two very similar topics. It can be said that the difference between closely similar things is sometimes more informative than the difference between widely different things. The practitioners using LDA have so far focused only on the words having high probabilities in each topic separately. That is, they have focused only on the rationale of the difference between widely different topics. However, when analyzing the data showing apparent regularities, we can utilize the difference between closely similar topics to reveal some important aspects of the data.

The data exploration required for our method is now performed manually except the variational inference for LDA. We think that this is not necessarily a burden. The detection of topic probability change (e.g. Fig. 2) and the inspection of the sequences of the sensor outputs (e.g. Fig. 3) can both be performed only by using a banal data visualization. However,

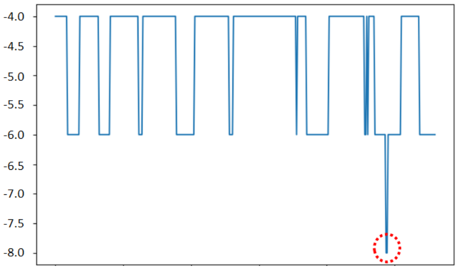

Fig. 11. Example of a per-lot sequence of the sensor T’s outputs.

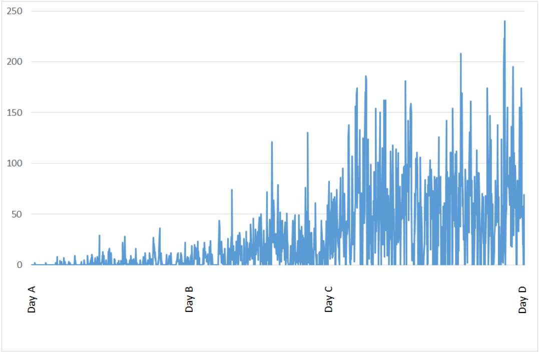

Fig. 12. Frequency of the value -8.0 of the sensor T in the data set II.

future work may help to automatize such data exploration. It is also an important future research direction to provide a defect detection system based on the findings presented in this paper.

References

[1] D. M. Blei, “Probabilistic topic models,” *Commun. ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012. [Online]. Available: http://doi.acm.org/10.1145/2133806.2133826

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[3] F. Castanedo, H. Aghajan, and R. Kleihorst, “Modeling and discovering occupancy patterns in sensor networks using latent Dirichlet allocation,” in *Proceedings of the 4th International Conference on Interplay Between Natural and Artificial Computation - Volume Part I*, ser. IWINAC’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 481–490. [Online]. Available: http://dl.acm.org/citation.cfm?id=2009405.2009455

[4] F. Castanedo, D. L. de Ipiña, H. K. Aghajan, and R. Kleihorst, “Learning routines over long-term sensor data using topic models,” *Expert Sys: J. Knowl. Eng.*, vol. 31, no. 4, pp. 365–377, Sep. 2014. [Online]. Available: http://dx.doi.org/10.1111/exsy.12033

[5] K. Farrahi and D. Gatica-Perez, “Discovering routines from large-scale human locations using probabilistic topic models,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, pp. 3:1–3:27, Jan. 2011. [Online]. Available: http://doi.acm.org/10.1145/1889681.1889684

[6] ——, “Extracting mobile behavioral patterns with the distant n-gram topic model,” in *Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC)*, ser. ISWC ’12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ISWC.2012.20

[7] M. D. Hoffman, D. M. Blei, and F. Bach, “Online learning for latent Dirichlet allocation,” in *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’10. USA: Curran Associates Inc., 2010, pp. 856–864. [Online]. Available: http://dl.acm.org/citation.cfm?id=2997189.2997285

[8] T. Huynh, M. Fritz, and B. Schiele, “Discovery of activity patterns using topic models,” in *Proceedings of the 10th International Conference on Ubiquitous Computing*, ser. UbiComp ’08. New York, NY, USA: ACM, 2008, pp. 10–19. [Online]. Available: http://doi.acm.org/10.1145/1409635.1409638

[9] B.-Y. Raanan, J. Bellingham, Y. Zhang, M. Kemp, B. Kieft, H. Singh, and Y. Girdhar, “Detection of unanticipated faults for autonomous underwater vehicles using online topic models,” *Journal of Field Robotics*, vol. 35, no. 5, pp. 705–716, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21771

[10] M. E. Roberts, B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand, “Structural topic models for open-ended survey responses,” *American Journal of Political Science*, vol. 58, no. 4, pp. 1064–1082, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12103

[11] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, “Probabilistic author-topic models for information discovery,” in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’04. New York, NY, USA: ACM, 2004, pp. 306–315. [Online]. Available: http://doi.acm.org/10.1145/1014052.1014087

[12] M. Strohbach, H. Ziekow, V. Gazis, and N. Akiva, *Towards a Big Data Analytics Framework for IoT and Smart City Applications*. Cham: Springer International Publishing, 2015, pp. 257–282. [Online]. Available: https://doi.org/10.1007/978-3-319-09177-8_11

[13] Y. Zhang and H. Hung, “Using topic models to mine everyday object usage routines through connected IoT sensors,” in *Proceedings of the 8th International Conference on the Internet of Things*, ser. IOT ’18. New York, NY, USA: ACM, 2018, pp. 27:1–27:4. [Online]. Available: http://doi.acm.org/10.1145/3277593.3277634