Traffic sensor health monitoring using spatiotemporal graphical modeling

Linjiang Wu[1], Chao Liu[2], Tingting Huang[3], Anuj Sharma[3] and Soumik Sarkar[1]*

[1] *Department of Mechanical Engineering, Iowa State University, Ames, Iowa, 50011, USA*
lwu26,soumiks@iastate.edu

[2] *Department of Thermal Engineering, Tsinghua University, Beijing 100084, China*
cliu5@tsinghua.edu.cn

[3] *Department of Civil, Construction and Environmental Engineering, Ames, Iowa, 50011, USA*
thuang1,anujs@iastate.edu

ABSTRACT

Accurate traffic sensor data is essential for traffic operation management systems and acquisition of real-time traffic surveillance data depends heavily on the reliability of the traffic sensors (e.g., wide range detector, automatic traffic recorder). Therefore, detecting the health status of the sensors in a traffic sensor network is critical for the departments of transportation as well as other public and private entities, especially in the circumstances where real-time decision is required. With the purpose of efficiently determining the sensor health status and identifying the failed sensor(s) in a timely manner, this paper proposes a graphical modeling approach called spatiotemporal pattern network (STPN). Traffic speed and volume measurement sensors are used in this paper to formulate and analyze the proposed sensor health monitoring system and historical time-series data from a network of traffic sensors on the Interstate 35 (I-35) within the state of Iowa is used for validation. Based on the validation results, we demonstrate that the proposed approach can: (i) extract spatiotemporal dependencies among the different sensors which leads to an efficient graphical representation of the sensor network in the information space, and (ii) distinguish and quantify a sensor issue by leveraging the extracted spatiotemporal relationship of the candidate sensor(s) to the other sensors in the network.

1. INTRODUCTION

Due to the increasing dependencies on relatively cheaper sensors for condition monitoring, diagnostics, and decision making in large infrastructure systems (Wenjie, Lifeng, Zhanglong, & Shiliang, 2005) (Wang, Zhang, Sun, Gong, & Cui, 2011), the reliability of the sensors themselves is critical in terms of collecting accurate information from the system of interest. Most of the previous studies tend to use sensor redundancy approaches by considering one data source as the ground truth to validate another data source (Sallans, Bruckner, & Russ, 2005). Such systems typically have multiple collocated sensors to monitor the critical points (Jeong, Kim, Lee, & Dornfeld, 2006) (Harris et al., 1995), which may be reasonable for expensive, safety-critical systems or small systems where only limited monitoring points are needed (Bhuiyan, Wang, & Wu, 2009). However, such an approach may not be feasible in large distributed systems such as large commercial buildings (Krishnamurthy, Sarkar, & Tewari, October 2014; Bengea et al., 2015) and transportation network (Liu et al., December 2016), which may not have multiple data sources to cross-validate the data they obtained or have limited budget for sensor implementation.

In a typical road transportation network, traffic sensors are deployed on freeways primarily to collect real-time data for traffic adaptive signal control and mitigating recurring or non-recurring congestion(Klein, Mills, & Gibson, 2006). According to the U.S. Department of Transportation (the U.S. DOT), the sensors are typically installed about every 2 miles and facilitating sensor redundancy is not feasible due to the sheer length of roadways that requires monitoring in each state and the cost of sensors deployed (e.g., microwave radar sensor that covers multiple lanes costs at least $6200 without the installation fee based on the costs database of the U.S. DOT in 2002 (Klein et al., 2006)). Therefore, a robust and feasible approach to monitor the health status of the traffic sensors is required that does not rely on redundancy of collocated sensors.

Among the existing techniques that are used to monitor the health status of the traffic sensors, a method based on the traffic flow theory has been widely adopted (Wells, Smaglik, & Bullock, 2008) (Dailey, 1999) to identify the erroneous data and anomalous sensors. In this approach, the average effective vehicle length (AEVL) is computed by defining a function $F(v, c, o)$, where v, c and o denote the traffic speed, traffic volume and the sensor occupancy respectively. An assessment criteria can be formulated to report the error rate of the sensor by evaluating whether the AEVL value meets the criteria or not ("Federal size regulations for commercial motor vehicles", 2016). Although it is fast and can be used for online monitoring, the method is built upon single lane road assumption and the accuracy is seriously affected when applied to roads with multiple lanes (e.g., 2 lanes when only 1 lane has vehicle passing at the recording time).

In this paper, we consider the fact that the traffic sensors in the same freeway and direction form a sensor network in the information space. Therefore, under nominal conditions, the data collected by these sensors should follow a stable spatiotemporal relationship among themselves that can be captured by an efficient learning technique using historical data. Such a stable relationship will be affected when one or more sensors degrade in performance. Hence, discovering the relationships among the sensors during operation with respect to the (historical) nominal conditions can provide us indications whether a sensor is healthy or not. In this context, this work applies a recently proposed spatiotemporal graphical modeling approach, called the spatiotemporal pattern network (STPN, built on the concepts of symbolic dynamics filtering, SDF) (Sarkar, Sarkar, Virani, Ray, & Yasar, 2014; Liu, Ghosal, Jiang, & Sarkar, 2017; Jiang & Sarkar, 2015), to build a novel sensor health monitoring framework for traffic sensors.

Contributions: The main contributions of this paper are: (i) formulation of the traffic sensor health monitoring problem as an anomaly detection problem by modeling observations from adjacent traffic sensors using a probabilistic graphical model called STPN, the anomaly detection formulation enables us to handle the imbalance between nominal sensor data (widely abundant) and faulty sensor data (low availability), also this problem is particularly challenging as the sensors are non-collocated and hence, there is no sensor redundancy, (ii) proposing two sensor health monitoring approach, off-line (when large amount of data from the operational stage can be processed in a batch mode) and on-line (when the decision needs to be made in real-time with streaming data i.e., a large amount of operational data is not available), and (iii) validation of the proposed framework using both synthetic and real data, we demonstrate that our proposed methods are significantly more effective compared to the state-of-the-art technique based on traffic flow theory (for general sensor degradation/fault types (Najafi, Gulp, & Langari, 2004)); we also compare the pros and cons of the on-line and the off-line

techniques. Note that the on-line sensor health monitoring approach serves as a practical case study for the recently proposed STPN+RBM (RBM: Restricted Boltzmann Machine) technique (Liu, Ghosal, Jiang, & Sarkar, 2016). The traffic sensor network studied in this paper is installed on the Interstate 35 from Ankeny to Ames in the state of Iowa. The data set was collected by Wavetronix LLC. in Oct., Nov. and Dec. 2016, and the ground truth (nominal and anomalous categorization) has been established manually based on careful inspection and collected field images.

Including the introduction, this paper is organized into 5 sections. Section 2 presents the background of STPN, an information theoretic metric and an inference based metric leveraged to build the sensor health monitoring framework. Section 3 describes the three approaches for sensor health monitoring: (i) the benchmark method (AEVL), (ii) the proposed off-line method, using the information theoretic metric, and (iii) the proposed on-line approach, using the inference based metric. The results obtained with the three different methods are described in Section 4 and the paper is summarized in Section 5 along with the directions of future research.

2. BACKGROUND AND PRELIMINARIES

2.1. Spatiotemporal pattern network (STPN)

Symbolic dynamic filtering (SDF) has been recently shown to be extremely effective for extracting key textures from time-series data for anomaly detection and pattern classification (Rao, Ray, Sarkar, & Yasar, 2009; Sarkar, Sarkar, Mukherjee, Ray, & Srivastav, 2013). The core idea is that a symbol sequence (i.e., discretized time-series) emanated from a process can be approximated as a Markov chain of order D (also called depth), named as D-Markov machine (Sarkar et al., 2014) that captures key behavior of the underlying process.

The symbolization process (also called partitioning (Ray, 2004) (Sarkar, Mukherjee, Sarkar, & Ray, 2013)) is as follows: Let $\mathbb{X}$ represents a set of partitioning functions, $\mathbb{X} : X(t) \rightarrow S$, which can transform a general dynamic system (time-series $X(t)$) into a symbol sequence S using an alphabet set Σ. Researchers have proposed different approaches according to different objective functions, such as uniform partitioning (UP), maximum entropy partitioning (MEP), statistically similar discretization (SSD) (Sarkar & Srivastav, 2016), and maximally bijective discretization (MBD)(Sarkar, Srivastav, & Shashanka, 2013). This study uses SSD for univariate model (when only traffic speed data is used) and SSD combined with MBD for two dimensional model (when traffic speed and traffic volume data are used).

The D-Markov machine is represented by a probabilistic finite state automaton (PFSA) that is constituted by states (represents different parts of the data space) and probabilistic transitions among these states that can be captured from time-series data.

The definition and illustration of a PFSA is shown in Definition 2.1 and Fig. 1. Detailed definitions of D-Markov machine, xD-Markov machine, and the learning schemes can be found in (Sarkar et al., 2014; Mukherjee & Ray, 2014).

Definition 2.1. A probabilistic finite state automaton (PFSA) K is constructed based on 4-tuple, $K = (\Sigma, Q, \delta, \pi)$ (Adenis, Wen, & Ray, 2012):

1. Σ is a non-empty finite set, called the symbol alphabet, with cardinality $|\Sigma|$;

2. Q is a non-empty finite set, called the set of states, with cardinality $|Q|$;

3. $\delta : Q \times \Sigma \to Q$ is the state-transition map, and Σ^* is the collection of all finite-length strings with symbols from Σ including the (zero-length) empty string ε.

4. $\pi : Q \times \Sigma \to [0, 1]$ is the symbol generation function (probability morph function) that satisfies the condition $\sum_{\sigma \in \Sigma} \pi(q, \sigma) = 1, \forall q \in Q$, and π_{ij} is the probability of occurrence of a symbol $\sigma_j \in \Sigma$ at the state $q_i \in Q$.
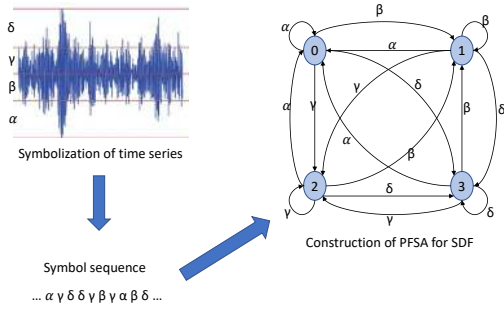

Figure 1. *The construction of a PFSA example for SDF, where {0, 1, 2, 3} are the states.*

Based on the above setup, the spatiotemporal pattern networks (STPNs) is defined as below (Liu et al., 2016).

Definition 2.2. A PFSA based STPNs is a 4-tuple as $W_D \equiv (Q^a, \Sigma^b, \Pi^{ab}, \Lambda^{ab})$, (a, b are nodes of the STPN)

1. $Q^a = \{q_1, q_2, \cdots, q_{|Q^a|}\}$ is the state set corresponding to symbol sequences S^a;

2. $\Sigma^b = \{\sigma_0, \cdots, \sigma_{|\Sigma^b|-1}\}$ is the alphabet set of symbol sequence S^b;

3. Π^{ab} is a $|Q^a| \times |\Sigma^b|$ symbol generation matrix, the ij^{th} element of Π^{ab} represents the probability of observing the symbol σ_j in the symbol list s^b while making a transition from the state q_i in the symbol sequence S^a; self-symbol generation matrices are called atomic patterns (APs) i.e., when $a = b$, cross-symbol generation matrices are called relational patterns (RPs) i.e., when $a \neq b$.

4. Λ^{ab} is a metric that can represent the importance of the learnt pattern (or degree of causality) for $a \to b$ which is a function of Π^{ab}.

An illustration of STPN is shown in Fig. 2.

2.2. Information theoretic metric for causality

Based on the above definition of STPN, we can use the atomic/relational patterns to interpret the causal dependencies among the sensors. In this context, information theoretic criteria have been widely used, e.g., transfer entropy (Wibral et al., 2011) and mutual information (Sarkar et al., 2014; Solo, 2008). In this paper, the concept of mutual information is applied for representing Λ^{ab} of the patterns (APs & RPs). The definition of Λ^{aa} and Λ^{ab} are as follow.

$$\Lambda^{aa} \triangleq I^{aa} = I(q_{k+1}^a; q_k^a) = H(q_{k+1}^a) - H(q_{k+1}^a | q_k^a) \quad (1)$$

where, I^{aa} is the mutual information of atomic pattern (a, a), H is the conditional entropy defined as follows,

$$H(q_{k+1}^a) = -\sum_{i=1}^{Q^a} P(q_{k+1}^a = i) \log_2 P(q_{k+1}^a = i)$$

$$H(q_{k+1}^a | q_k^a) = \sum_{i=1}^{Q^a} P(q_k^a = i) H(q_{k+1}^a | q_k^a = i)$$

$$H(q_{k+1}^a | q_k^a = i) = -\sum_{j=1}^{Q^a} P(q_{k+1}^a = j | q_k^a = i)$$
$$\cdot \log_2 P(q_{k+1}^a = j | q_k^a = i)$$

Here, I^{aa} essentially captures the temporal self-prediction capability of the sensor node a. Similarly, the mutual information for the relational pattern (a,b) can be expressed as:

$$\Lambda^{ab} \triangleq I^{ab} = I(q_{k+1}^b; q_k^a) = H(q_{k+1}^b) - H(q_{k+1}^b | q_k^a) \quad (2)$$

where, I^{ab} is the mutual information of pattern (a, b), H is the conditional entropy defined as follows,

$$H(q_{k+1}^b | q_k^a) = \sum_{i=1}^{Q^a} P(q_k^a = i) H(q_{k+1}^b | q_k^a = i)$$

$$H(q_{k+1}^b | q_k^a = i) = -\sum_{j=1}^{Q^b} P(q_{k+1}^b = j | q_k^a = i)$$
$$\cdot \log_2 P(q_{k+1}^b = j | q_k^a = i)$$

Detailed description of mutual information theoretic causality metric in the context of APs and RPs can be found in (Sarkar et al., 2014).

2.3. Inference based metric using STPN

The mutual information theoretic metric introduced in Section 2.2, requires significant amount of data for estimating the state transition probabilities and hence, may not be ideally suited for online decision making. Therefore, an alternative inference based metric is presented here which utilizes a short time window of data to compute the metric using a Dirchlet prior
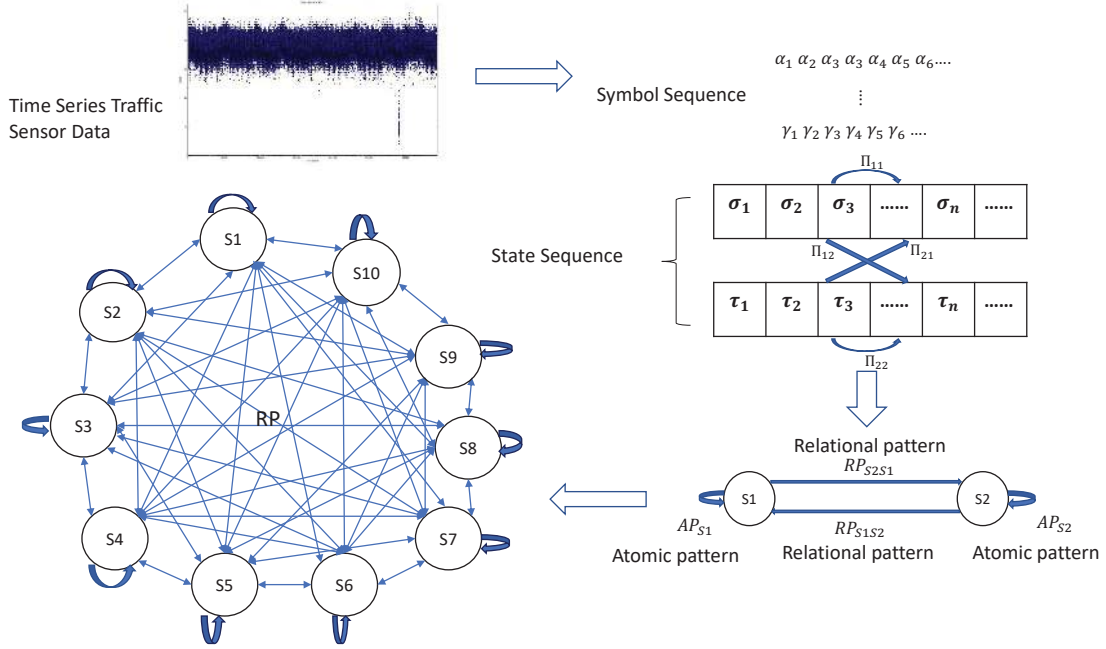
Figure 2. *Extraction of atomic and relational patterns (using D-Markov and xD-Markov machines respectively and $D = 1$, i.e., states and symbols are equivalent) in sensor network $\mathbf{S}$, where $\mathcal{N} = (1, 2, .., 10)$ sensors in the network to characterize individual sub-system behavior and interaction behavior among different sub-systems.*

on the state trasition probabilities. To compute this metric, a two-step process is needed that includes a modeling and an inference phase (Liu et al., 2016).

In the modeling phase, the entire time-series in the nominal condition is considered, where the multivariate time-series is denoted by $X = \{X^{\mathcal{N}}(t), t \in \mathbb{N}, \mathcal{N} = 1, 2, \cdots, 10\}$, where $\mathcal{N}$ is the number of traffic sensors in the network. The multivariate time-series is symbolized into $S = \{S^{\mathcal{N}}\}$ and then state sequences are generated with the STPN formulation, noted by $Q = \{Q^a, a = 1, 2, \cdots, 10\}$.

In the inference phase, a short time-series is analyzed, $\tilde{X} = \{\tilde{X}^{\mathcal{N}}(t), t \in \mathbb{N}^*, \mathcal{N} = 1, 2, \cdots, 10\}$, where $\mathbb{N}^*$ is a subset of $\mathbb{N}$. The length of the short time-series depends on the selection of a window size, which is flexible and can be overlapping. The symbolic subsequences for the short time-series is noted as $\tilde{S} = \{\tilde{S}^{\mathcal{N}}\}$, and the state sequences is noted as $\tilde{Q}$. An importance metric Λ^{ab} is defined for a given short subsequence (described by short state subsequence $\tilde{Q}$ and short symbol subsequence $\tilde{S}$). The value of this metric suggests the importance of the pattern Π^{ab} or the degree of causality in $a \rightarrow b$ as evidenced by the short subsequence. In this context, we consider

$$\Lambda^{ab}(\tilde{Q}, \tilde{S}) \propto Pr(\{\tilde{Q}^a, \tilde{S}^b\}|\Pi^{ab}) \qquad (3)$$

where $Pr(\{\tilde{Q}^a, \tilde{S}^b\}|\Pi^{ab})$ is the conditional probability of the joint state-symbol subsequence given the pattern Π^{ab}.

With this definition of Λ^{ab} and with proper normalization, the

inference based metric $\Lambda^{ab}(\tilde{Q}, \tilde{S})$ can be obtained as follows,

$$\Lambda^{ab}(\tilde{Q}, \tilde{S}) =$$

$$K \prod_{m=1}^{|Q^a|} \frac{(\tilde{N}_m^a)!(N_m^a + |\Sigma^b| - 1)!}{(\tilde{N}_m^a + N_m^a + |\Sigma^b| - 1)!} \prod_{n=1}^{|\Sigma^b|} \frac{(\tilde{N}_{mn}^{ab} + N_{mn}^{ab})!}{(\tilde{N}_{mn}^{ab})!(N_{mn}^{ab})!} \quad (4)$$

where, K is a proportional constant, $N_{mn}^{ab} \triangleq |\{(Q^a(k), S^b(k+1)) : S^b(k+1) = \sigma_n^b \mid \mathcal{Q}^a(k) = q_m^a\}|$, $N_m^{ab} = \sum_{n=1}^{|\Sigma^b|}(N_{mn}^{ab})$, $\tilde{N}_{mn}^{ab}$ and $\tilde{N}_m^{ab}$ are similar to N_{mn}^{ab} and N_m^{ab}, $|Q^a|$ is number of states in state sequence $\tilde{Q}$, and $|\Sigma^b|$ is number of symbols in symbol sequence $\tilde{S}$.

A detail derivation can be found in (Liu et al., 2016).

Thus, with Eq. 4, inference metrics Λ^{ab} of APs (i.e., when $a = b$) and RPs (i.e., when $a \neq b$) are obtained with respect to the *short subsequences*.

Remark. In above definition and preliminaries, a,b nodes are the representation of sensors in the traffic sensor network, such as $\mathcal{S}1, \mathcal{S}2$ in Fig. 2.

2.4. Online anomaly detection with STPN+RBM framework

The inference metrics (Λ^{ab}) shown in the above section can be further normalized and converted into binary states (0 for low values and 1 for high values) for APs and RPs, and then are the inputs of Restricted Boltzmann Machine (RBM). With multiple short subsequences, a large number of examples can be formed, whose characteristics represent the systematic behav-

4

ior. Then, RBM is trained to capture the most likely system-wide behavior, and detect any anomaly via identifying a low probability event.

For RBM, weights and biases are learnt so that the feature configurations observed during nominal operation of the system obtain low energy (or high probability). Consider a system state that is described by a set of visible variables $\mathbf{v} = (v_1, v_2, \cdots, v_D)$ and a set of hidden (latent) variables $\mathbf{h} = (h_1, h_2, \cdots, h_F)$. Here, the normalized inference metrics (Λ^{ab} are used as the inputs for the RBM, thus, $|\mathbf{v}| = |\Lambda^{ab}|$, i.e., the number of visible units equal to the number of patterns learned by the STPN. The variables can be binary or real-valued depending on the need. Now, each joint configuration of these variables determines a particular state of the system and an energy value $E(\mathbf{v}, \mathbf{h})$ is associated with it. The energy values are functions of the weights of the links between the variables (for RBM, internal links within the visible variables and the hidden variables are not considered) and bias terms related to the variables.

With this setup, the probability of a state $P(\mathbf{v}, \mathbf{h})$ depends only on the energy of the configuration $(\mathbf{v}, \mathbf{h})$ and follows the Boltzmann distribution

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}} \qquad (5)$$

Anomaly detection process. During training, weights and biases are obtained via maximizing likelihood of the training data. During testing, short testing subsequences are converted into an $\mathbb{N}^2$-dimensional binary vectors using the same *inference phase* of the training process. Multiple testing (possibly overlapping) subsequences are applied to compute a distribution of free energy. For the nominal condition, the distribution of free energy should be close to that of the training data, while the anomalous data should differ from the nominal condition. Further details of the STPN+RBM framework can be found in (Liu et al., 2016).

3. PROBLEM SETUP AND METHODOLOGY

3.1. Problem setup

Consider a sensor network $\mathbf{S} = \{\mathcal{S}1, \mathcal{S}2, ..., \mathcal{S}\mathcal{N}\}$ with $\mathcal{N}$ sensors (as illustrated in Fig. 2). Each sensor is represented by the measurements in our case, which can be univariate time-series (speed) or multivariate time-series (speed, volume, and occupancy). The sensor health monitoring task is to find out the anomalous sensor(s) based on a certain performance metric $\mathcal{M}$. Therefore, the sensor health monitoring problem can be formulated as:

$$\text{Finding } \mathbf{S}^{ano} \subset \mathbf{S} \qquad (6)$$

where $\mathbf{S}^{ano}$ is the subset of the sensor network that are anomalous.

Three performance metrics are illustrated in the following sections where the first one is based on the traffic flow theory and used as the benchmark, and the later two are proposed in this paper for the purposes of off-line and on-line detection. Note that in this work, sensor health monitoring with univariate time-series (speed) is noted as the 1D model, and the one with two time-series (speed, volume) is noted as the 2D model.
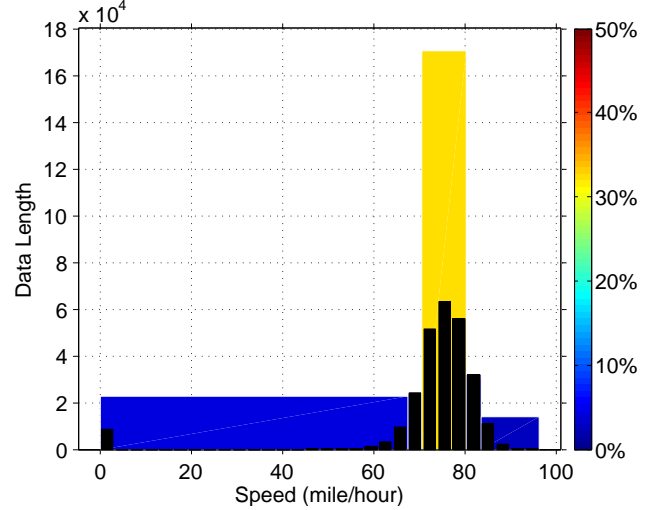

Figure 3. *Speed distribution in histogram which shows that the speed of 10 sensors is mainly in range from 65 mile/hour to 83 mile/hour, traffic congestion or no vehicle passing leads to 0 mile/hour.*

3.2. Benchmark method based on traffic flow theory

According to the traffic flow theory, there is an inherent relationship between speed, volume and occupancy (Dailey, 1999) and hence, such relationship can be applied to assess sensor data quality. Authors in (Wells et al., 2008) proposed a method to identify sensor errors via evaluating the relationship among speed, volume and sensor occupancy. In this framework, the sum of average effective vehicle length (AEVL) and the detection range (DTR) is estimated by the following empirical rule:

$$AEVL + DTR = \frac{5280 * Speed * Occupancy}{Volume}, \qquad (7)$$

where AEVL is in feet, DTR of the sensor is in feet, speed is in miles per hour, occupancy is a fractional number between 0 and 1 representing percentage of time when the sensor is occupied, volume is in vehicles per hour and the scalar 5280 is used for unit standardization.

We use this relationship based on our radar sensor as the benchmark method for sensor health monitoring. Note, DTR is the length of a loop detector (Chen, Petty, Skabardonis, Varaiya, & Jia, 2001), a typical sensor used in this type of applications. As the data from Wavetronix HD sensors uses a virtual line

to represent the detector("Wavetronix Smartsensor HD user guide", 2016), DTR equals to 0. According to ("Federal size regulations for commercial motor vehicles", 2016; Minge, Peterson, Weinblatt, Coifman, & Hoekman, 2012), the possible distance between vehicles should fall within the range of 10 ft to 75 ft, which provides a method to monitor the health of the sensor by using the AEVL equation. A sensor reading is identified as erroneous when the output falls outside this range and the ratio of error counts to the total number of data points is called the error rate E_r.

3.3. Off-line sensor fault detection using STPN

Among the proposed STPN-based sensor monitoring solutions, we first present the off-line method that uses the information theoretic metric defined in Section 2.2. We begin the discussion with the symbolization procedures which is critical for the success of the proposed schemes.

3.3.1. 1D data and 2D data symbolization

In 1D STPN model, we use traffic speed as the univariate input. The speed distribution from historical data shown in Fig. 3 demonstrates that speed mainly falls in the range of 65 miles per hour to 83 miles per hour and is normally distributed. In this case, we found that compared with UP and MEP discretization, SSD discretization that aims to preserve the nature of the continuous data distribution in the discrete domain, is much more effective.

While detail formulation of the SSD scheme can be found in (Sarkar & Srivastav, 2016), we provide a brief description here in **Algorithm 1** for completeness. Let $X(t) \in \Omega^1$ be a

---

Algorithm 1 Statistical Similarity-based Discretization (SSD)

1: Input variable X(t)
2: Input tolerance ξ on d for univariate discretization of X(t)
3: Compute empirical $F_n(x)$ of x as the set $\{(x^k, p^k) | Prob(x < x^k) = p^k, k = 0, 1, 2, 3, ..., K\}$ and $p^K = 1$
4: fit a line segment L_1 through the endpoints (x^0, p^0) and (x^K, p^K)
5: Initialization of the line segments $\mathbb{L} = \{L_1\}$
6: Find a split point k_s to maximize the distance between L_1 and cumulative density function $F_n(x)$
7: **while** d$\geq \xi$ **do**
8:     **for** i in k **do**
9:         Generate two new line segments L_l and L_r corresponding to the split point k_s for L_i
10:        Update $\mathbb{L} = \mathbb{L} \cup \{L_l, L_r\} \backslash L_i$
11:    **end for**
12:    Find the k_s for segments $\mathbb{L}$ and $F_n(x)$ have the maximum distance, where the bin boundary a is determined.
13: **end while**

---

one dimensional time-series data, speed, where interval Ω^1 is a compact subset of $\mathbb{R}$. Let $a = \{a_1, a_2, ..., a_k\}$, $a_i \in \Omega^1$, $a_i < a_j$ for $i < j$ be an ordered set with discrete levels that have to be

determined to discretize Ω^1. There exists a function $F(X(t))$ which represents the true underlying cumulative density function (CDF) of X(t). In the discrete domain, we define the density $\tilde{F}_a(X(t))$ that aims to preserve the statistical properties of X(t) in the sense of minimizing the distance d with respect to F, where $\tilde{F}_a(X(t))$ is a piece-wise linear and continuous function. In this paper, we use the Kolmogorov-Smirnov statistic d to compute the distance, where $d = \sup_{x(t)} |\tilde{F}_a - F|$. Note, generally, the true density $F(X(t))$ is not available, here we propose to use the empirical density $F_n(X(t))$(observed), where n is the number of data points. With the SSD technique, partitioning of the speed data is implemented and the symbol sequences for the speed sensors in the network are generated.

In this paper, we also construct STPN model for 2D data, where traffic speed and volume are used (data are in 20s and from the same data source as the 1 dimensional model), the correlation between speed and volume can be viewed in the plot presented in Fig. 4, The distribution of volume is shown in Fig. 5. To implement the joint (2D) symbolization of the correlated variables speed and volume, we adopt a two-step technique, SSD followed by MBD.

The main idea of MBD is as follows. Let $u(t)$ denote the volume and $y(t)$ denote the speed and let $\mathbb{L}=(L_1,...,L_g)$ denote the set of discrete bins for volume u after implementing the SSD discretization. The main objective of the MBD scheme is to find the bins for the speed variable $y(t)$ which is denoted by $a=(a_1,...,a_k)$, k is the number of bins, such that there is a maximum possible one-to-one correspondence (hence, maximally bijective) between the bins for y and those for u. In this

---

Algorithm 2 Maximally Bijective Discretization(MBD)

1: $c_1 = \min(y)$
2: r_j is a dummy variable that scans through the range of the variable y
3: $k=1$
4: **while** $r_j < \max y(t)$ **do**
5:     Select L_i such that $P(L_i|a_j) \Rightarrow P(a_j)$
6:     Select L_m such that $P(L_m|a_j + dy) \Rightarrow P(a_j + dy)$
7:     L_i, L_m represent the existing i^{th}, m^{th} bins for $u(t)$
8:     **if** $i \neq m$ **then**
9:         $c_k = r_j$
10:        k$\leftarrow$ k+1
11:    **end if**
12:    $r_j \leftarrow r_j + dy$
13: **end while**
14: bin boundaries for MBD, $c=(c_1,...,c_{k-1})$ is obtained

---

context, we call bin L_i corresponds to bin a_j, i.e., $L_i \Rightarrow a_j$ if $i = arg \max_h P(L_h|a_j \in a)$, $h \in (1,..,g)$, where $P(\cdot)$ is a probability function, L_i is the ith bin for volume and a_j is the jth bin for speed. Then, a reward function is defined for the discretization as follows: $R(a_j) = P(L_i|a_j \in a)$ s.t. $L_i \Rightarrow a_j$. A higher reward value means that an existing (SSD) bin better corresponds to a MBD bin. The total expected reward value

can be calculated by: $TR(a) = \int_a R(a_j)P(a_j)da_j$. In the MBD scheme, the goal is to maximize the total reward TR.

Algorithm 2 aims to achieve the maximally bijective discretization of $y(t)$, where $\{c_1, c_2, \cdots, c_{k-1}\}$ denotes the set of MBD bin boundaries and a_j is the bin with boundaries c_j and c_{j+1}:
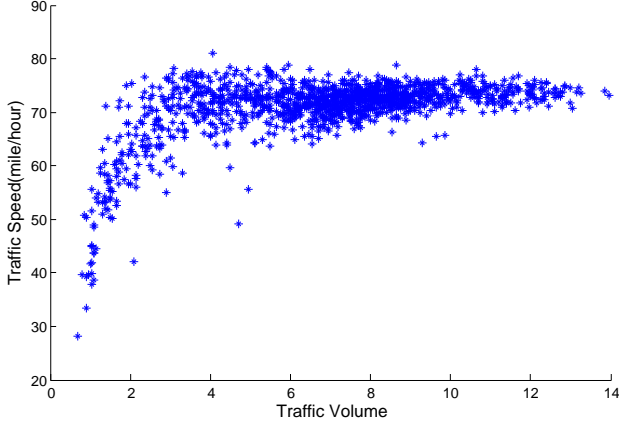

Figure 4. *Relationship between speed and volume, it reasonably explains the high correlations between speed and volume.*

3.3.2. Fault detection

After symbolization, a set of symbol sequences $S^1, S^2, ..., S^{10}$ (representing training data and testing data respectively, and corresponding to sensor network shown in Fig. 2) are generated in time-series based on the training symbols Σ. Thus, based on section 2.2, the mutual information matrices (10 by 10 matrix) of the training data and testing data in the network can be obtained as:

$\underline{\Lambda}_{trg} = (I^{S1S1}, I^{S1S2}, ..., I^{S10S10})$
$\underline{\Lambda}_{tst} = (I^{S1S1}, I^{S1S2}, ..., I^{S10S10})$

Where $\underline{\Lambda}_{trg}$ and $\underline{\Lambda}_{tst}$ are the training and testing mutual information matrices of the sensor network respectively.

The difference $\Delta\underline{\Lambda}$ between $\underline{\Lambda}_{trg}$ and $\underline{\Lambda}_{tst}$, can be used to detect and isolate the anomalous sensor(s) in the sensor network.

3.4. Online detection with inference based on spatiotemporal graphical modeling

With STPN+RBM framework presented in Section 2.4, an anomaly is detected as a high energy (low probability) event, and the distribution of free energy in anomalous condition differs from that in the nominal condition. Then a sequential state switching method i.e. root cause analysis method (RCA), can be used to further localize the fault in the sensor network. The idea for sequential state switching is to find potential pattern(s) that, if changed, can transition the system from a high to a low energy state. The probabilities of AP and RP's existence are
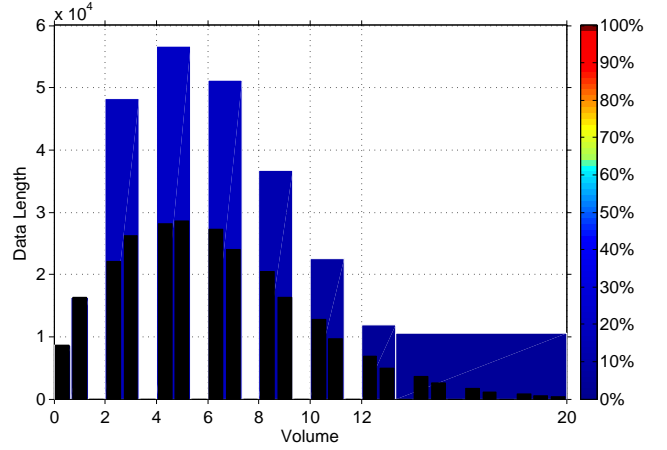

Figure 5. *Volume distribution in histogram. In 20s timestamp data, the volume is mainly distributed in range from 0 to 12.*

discovered by the STPN, and an anomaly will influence the causality of specific patterns. Hence, by switching/flipping a pattern, its contribution on the energy status of the system can be estimated and attributed to a possible fault.

Based on sections 2.3 & 2.4, with the weights and biases of RBM using training data, free energy can be computed (Hinton, 2012):

$$F(v) = -\sum_i v_i a_i - \sum_j \ln(1 + e^{b_j + \sum_i v_i w_{ij}})$$

The free energy in nominal conditions is noted as $\tilde{F}$. In fault conditions, a failed pattern will shift the energy from a lower state to a higher state. Assume that the patterns can be categorized into two sets, $\mathbf{v}^{nom}$ and $\mathbf{v}^{ano}$. By flipping the set of anomalous patterns $\mathbf{v}^{ano}$, a new expression for free energy is obtained:

$$F^s(v) = -\sum_g v_g a_g - \sum_j \ln(1 + e^{b_j + \sum_g v_g w_{gj}})$$
$$- \sum_h v_h^\star a_h - \sum_j \ln(1 + e^{b_j + \sum_h v_h^\star w_{hj}}), \quad (8)$$
$$\{v_g\} \in v^{nom}, \{v_h^\star\} \in v^{\star, ano}$$

Here, $v^\star$ has the opposite state to v and represents that the probability of the pattern has been significantly changed. In this work, the probabilities of the patterns are binary (i.e. 0 or 1). Hence, we have that $v^\star = 1 - v$. The sequential state switching is formulated by finding a set of patterns v^{ano} via $\min(F^s(v^{ano}, v^{nom}) - \tilde{F})$. Flowchart of sequential state switching method is shown in Fig. 6.

Note that the sequential state switching method is pattern based, and the anomalous patterns are associated to the candidate nodes (sensors in this case) using a sequential search
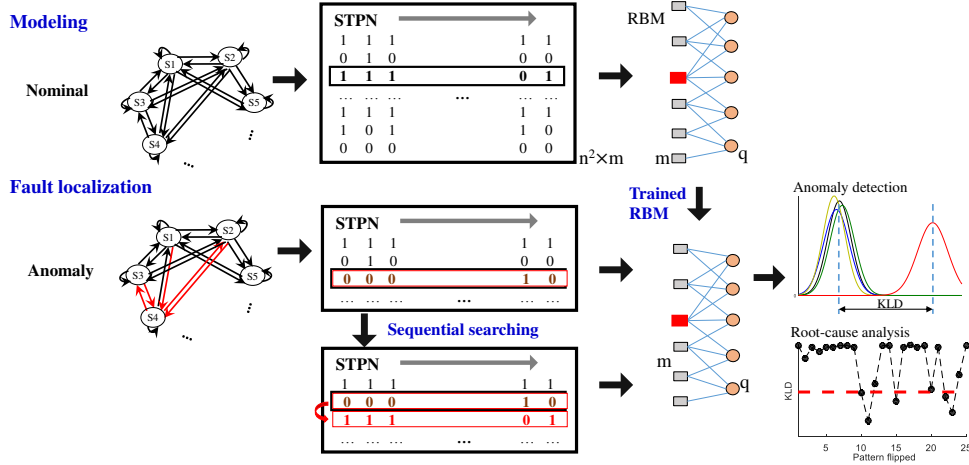
Figure 6. *Online detection via sequential state switching approach based on STPN+RBM framework. Spatiotemporal features are extracted from both nominal and anomalous data, with online STPN inference. Multiple sub-sequences of APs and RPs form input vectors to the RBM. Here, the RBM is only trained with nominal data, and the anomalous data is used as input to compute free energy. Anomaly is detected by identifying its high energy state. In the root-cause analysis phase (fault detection/isolation), the potentially failed patterns are obtained via evaluating free energy of the system with the perturbations.*

method, to find a subset $\hat{\mathbf{X}}$ of $X = \{X^{\mathcal{N}}(t), t \in \mathbb{N}, \mathcal{N} = 1, 2, \cdots, 10\}$ that can interpret all of the anomalous patterns $\mathbf{\Lambda}^{ano}$. For example, a pattern $N_i \to N_j$ is identified as failed, and it indicates that the two nodes (N_i, N_j) are potentially failed. If multiple patterns from or to a node are detected as anomalous, the node is more reliable to be classified as anomalous. Thus, the node inference can be carried out via computing the anomaly score of each node. Also, as the failed patterns contribute differently to the system (in terms of energy increase in RBM, weights of failed patterns are defined. The weights of failed patterns can be formed by the difference of free energy with and without the failed pattern and then can be associated to the anomaly scores for the patterns.

With sequential state switching method, the fault sensor(s) $\mathbf{S}^{ano}$ in the sensor network can be identified and the performance metric $\mathcal{M}$ is formed using the anomaly score(s) AS, which is a measure of the energy increase (of the candidate sensor) in the RBM. By defining a threshold AS^{thres}, the online detection approach for the sensor network is formulated as:

$$\text{Finding a subset } \mathbf{S}^{ano} \subset \mathbf{S}, \text{ where } AS > AS^{thres} \quad (9)$$

Note that, as the sequential state switching method is built upon short sequences, it only needs short time-series and is suitable for online detection.

4. RESULTS AND DISCUSSIONS

While we collected real data from Iowa interstate traffic scenarios (a network of 10 sensors on Interstate 35 from Ankeny to Ames in the state of Iowa, the data set was collected by Wavetronix LLC. in Oct., Nov. and Dec. 2016) to validate

our proposed technique, actual sensor faults are somewhat rare. Therefore, it becomes difficult to use only that data for comprehensive validation. However, representative sensor degradations of different types and severity levels can be artificially injected in real nominal data and that is how we begin presenting our results.

4.1. Simulation results

Based on the collected real data, the sensor faults are simulated in two ways: (i) drift–the measured speeds of the sensor(s) are artificially modified by adding different levels of bias and (ii) noise–the measurements of the sensor(s) are contaminated by a predefined level (variance) of Gaussian noise.

The original data is initially divided into a training, $X_{trg}(t)$ and testing, $X_{tst}(t)$ data for validating with real data. However, initially the training set $X_{trg}(t)$ is further divided into $\hat{X}_{trg}(t)$ (80% with 32000 data points) and $\hat{X}_{tst}(t)$ (20% with 8000 data points) sets that are treated as the training and testing sets for a simulation based validation. Then we artificially inject sensor faults into the testing set as described above. Simulation cases include: (1) adding drift from 1 mile per hour to 10 miles per hour to one sensor (sensor 3, $\mathcal{S}3$), two sensors ($\mathcal{S}3$ & $\mathcal{S}7$), and five sensors ($\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$, $\mathcal{S}6$, and $\mathcal{S}7$) respectively; (2) adding Gaussian noise with standard deviation from 1 to 10 to one sensor ($\mathcal{S}3$), two sensors ($\mathcal{S}3$ & $\mathcal{S}7$), and five sensors ($\mathcal{S}1$, $\mathcal{S}2$, $\mathcal{S}3$, $\mathcal{S}6$, and $\mathcal{S}7$) respectively. The cases are tested with AEVL, off-line (STPN) and online (RCA) approaches as presented in Section 3. Note that we call the STPN+RBM based online method as RCA as it essentially uses a root-cause analysis (RCA) approach to solve the sensor fault detection and isolation problem.

Table 1. Anomaly detection results with one-sensor-fault simulations

Noise Type	Data Type	Method	Severity									
			1	2	3	4	5	6	7	8	9	10
Drift(mile/hour)	S	STPN	0/1	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1
		RCA	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	S+V	STPN	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
		RCA	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	S+V+O	AEVL	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1
Gaussian Noise(mile/hour)	S	STPN	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
		RCA	0/1	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	S+V	STPN	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
		RCA	0/1	0/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
	S+V+O	AEVL	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	1/1

Table 2. Anomaly detection results with two-sensors-fault simulations

Noise Type	Data Type	Method	Severity									
			1	2	3	4	5	6	7	8	9	10
Drift(mile/hour)	S	STPN	0/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
		RCA	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
	S+V	STPN	0/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
		RCA	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
	S+V+O	AEVL	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	2/2
Gaussian Noise(mile/hour)	S	STPN	0/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
		RCA	0/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
	S+V	STPN	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
		RCA	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
	S+V+O	AEVL	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2	2/2

The results are summarized in Table 1–3 where the drift and noise variance levels are treated as severity levels. The simulation results show that 2 dimensional models have higher sensitivity than 1 dimensional models when using STPN and RCA and both can isolate the faulty sensors with a higher sensitivity. The benchmark method AEVL does not perform as well despite using 3 sensor modalities.

4.2. Sensor fault detection with real data

The real use case involves the original training data and testing data ($X_{trg}(t)$ and $X_{tst}(t)$) that are collected for the same sensor network every 20s by Wavetronix LLC. in 2016. Using detail manual investigation, we find that sensor 6 is anomalous during the testing period and hence used as the ground truth. In the benchmark method, we can compute the AEVL for each sensor and use this value to compare with the length range (10 feet to 75 feet as shown in section 3.2) and compute the number of error occurrences (denoted by e). Then, the AEVL differential error will be: $\delta E_r^i = \frac{e_{trg}^i}{\#X_{trg}^i} - \frac{e_{tst}^i}{\#X_{tst}^i}$ for the ith sensor, and $\#X_{trg}^i$ represents the total number of data points in the training data. Then, δE_r^i is the error rate or anomaly score in this case for each sensor in the network.

For the off-line method (i.e., STPN), the symbol sequences, S^1, S^2, ..., S^{10}, are generated by symbolizing $X(t)$ (both $X_{trg}(t)$ and $X_{tst}(t)$) via SSD and MBD partitioning techniques. As the depth $D = 1$, the state sequences Q^1, Q^2, ...,

Q^{10} are equivalent to the corresponding symbol sequences that are the inputs to the STPN model for extracting the features APs and RPs. Based on these features, the training and testing mutual information matrices $\underline{\Lambda}_{trg}$ and $\underline{\Lambda}_{tst}$ can be obtained. To identify the anomalous sensor, deviation in the mutual information matrix, $\Delta\underline{\Lambda} = \underline{\Lambda}_{trg} - \underline{\Lambda}_{tst}$ can be computed as visualized in Figs. 7-8. As shown in Fig. 7, we can conclude that the faulty sensor 6 can be identified by STPN only using the speed data. However, there may be ambiguities and false alarms as evidenced visually by the deviation matrix. On the other hand, when using both speed and volume, the off-line STPN method can detect the fault more accurately without any significant ambiguity as seen in Fig. 8.

Remark. In order to get the testing mutual information matrix $\underline{\Lambda}_{tst}$, a long historic data set is required for computing the information theoretic metrics, as shows in section 2.2. Thus, we call it an off-line traffic sensor health monitoring method.

The on-line method adopts short sequences U, where $U \subset (S^1, S^2, ..., S^{10})$ to form the inference metric $\Lambda^{ab}(\tilde{Q}, \tilde{S})$ which is then used to construct the input to the RBM model. Based on the RBM training and testing procedures, the energy values for the training and testing time-series are obtained as $E_{trg}(\mathbf{v}, \mathbf{h})$, $E_{tst}(\mathbf{v}, \mathbf{h})$ respectively. Using the free energy method, we can detect the presence of an anomaly and subsequently implement root-cause analysis to identify the anomalous sensor(s). Table 4 presents the ranking of the possibly faulty sensors along with their respective anomaly scores. The on-

Table 3. Anomaly detection results with five-sensors-fault simulations

Noise Type	Data Type	Method	Severity									
			1	2	3	4	5	6	7	8	9	10
Drift(mile/hour)	S	STPN	0/5	2/5	2/5	2/5	3/5	3/5	3/5	3/5	5/5	5/5
		RCA	0/5	2/5	2/5	2/5	3/5	3/5	3/5	3/5	4/5	4/5
	S+V	STPN	0/5	2/5	2/5	3/5	3/5	3/5	5/5	5/5	5/5	5/5
		RCA	2/5	2/5	3/5	3/5	3/5	3/5	4/5	4/5	4/5	5/5
	S+V+O	AEVL	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	5/5
Gaussian Noise(mile/hour)	S	STPN	0/5	0/5	0/5	3/5	3/5	4/5	4/5	4/5	5/5	5/5
		RCA	2/5	2/5	2/5	2/5	3/5	3/5	3/5	4/5	4/5	4/5
	S+V	STPN	1/5	4/5	4/5	4/5	4/5	5/5	5/5	5/5	5/5	5/5
		RCA	2/5	3/5	3/5	3/5	3/5	4/5	4/5	4/5	4/5	5/5
	S+V+O	AEVL	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5	5/5

S:speed, V: volume, O: occupancy. The results in above tables (table 1–3) are represented in m/n, where m represents the detected anomaly sensor(s) and n denotes the sensor(s) are labeled anomaly in the sensor network. The severity levels correspond to the different levels of bias and noise variance synthetically added to the test data.
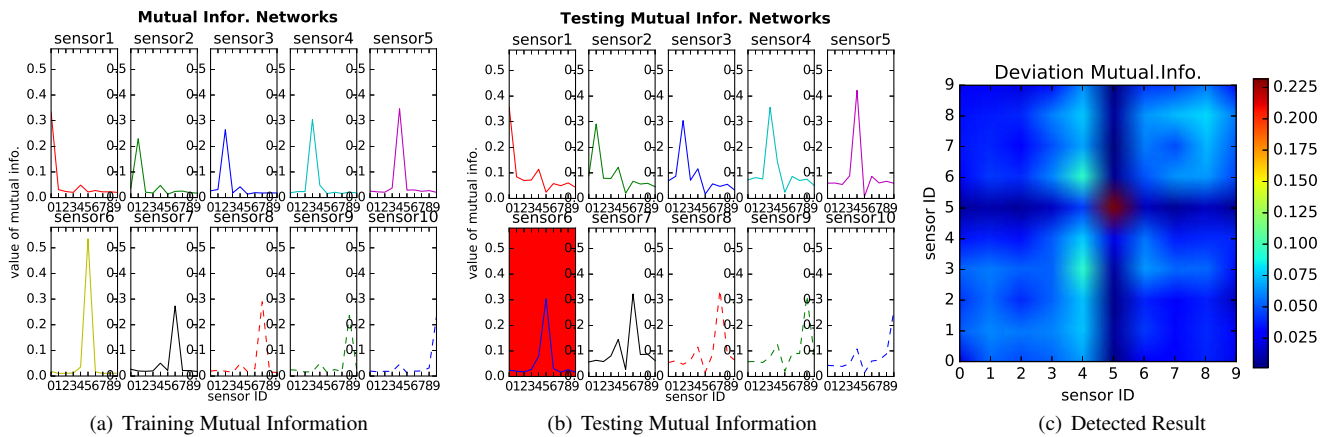

(a) Training Mutual Information (b) Testing Mutual Information (c) Detected Result

Figure 7. *Labeled sensor health monitoring using STPN with vehicle speed data, where sensor 6 has been detected as an anomaly sensor (matched with the labeled data) and the result is directly reported in image c with some ambiguity.*

line approach can correctly isolate the faulty sensor (which shows the highest anomaly score), while the AEVL method can not identify the faulty sensor despite using 3 different sensing modalities.

Remark. Although the proposed online detection technique is susceptible to false alarms, it plays a critical role in the entire traffic sensor health monitoring framework. This can be used for getting early indication of possible sensor anomalies and failures in traffic systems. They can be manually verified only if a potentially anomalous sensor is crucial such as being safety-critical. In this case, a certain level of false alarm can be acceptable. The online detection technique is complemented by the off-line method which can verify the alarms from the online technique using batch processing.

5. CONCLUSION AND FUTURE WORK

By applying the concept of spatiotemporal pattern network, this work proposes two ways (online and off-line) to monitor sensor health via graphical modeling of sensor network data. Both approaches are designed to process large-scale time-series data in sensor networks with advantages in: (1)

extracting spatiotemporal features to discover relationships among sensors, (2) detecting anomaly in an off-line manner by computing an information theoretic metric, (3) monitoring and localizing anomalous sensor in real-time by computing an inference based metric.

Based on the results, it can be concluded that: (i) compared with the benchmark AEVL method, off-line and on-line methods can isolate the anomalous sensor more accurately and can be very effective for different types of sensor anomalies such as bias/drift and increased noise variance, (ii) 2D model using both speed and volume data can distinguish the anomalous sensor more clearly than the 1D model. The possible reason is that, regional congestion may affect the 1D result since the speed is very low at that point, while the relationship among the sensors remain preserved while using both the speed and volume information. Also, while the proposed off-line method is more stable and sensitive, the on-line approach is fast, i.e., low time-to-detect (suitable for real-time application) but may have more false alarms.

This paper validates the two proposed approaches by applying them in an Iowa interstate segment sensor network with 10
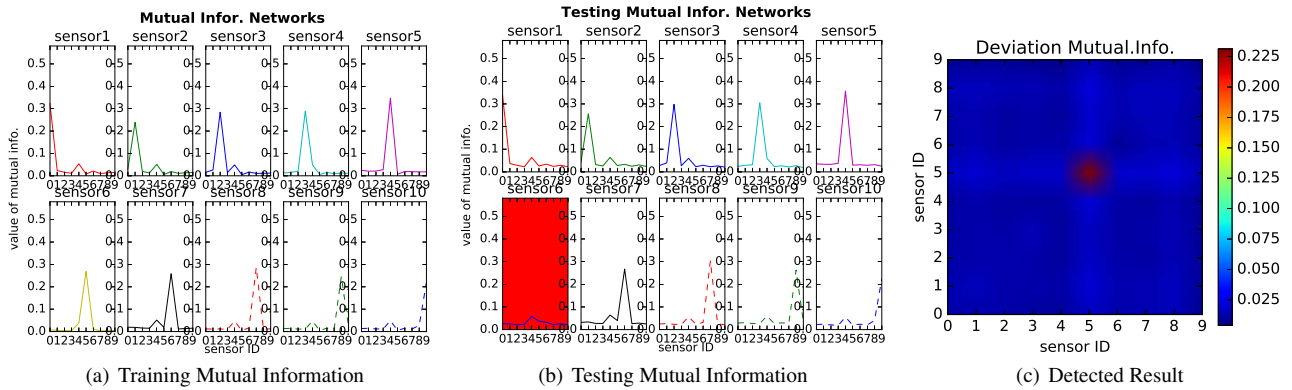
10

(a) Training Mutual Information   (b) Testing Mutual Information   (c) Detected Result

Figure 8. *Labeled sensor health monitoring using STPN with vehicle speed and volume data, where the labeled fault sensor 6 has been detected with less ambiguity as in Fig.7.*

Table 4. Results of RCA and AEVL with real data

Methodology	Data Type	Probability of detected faults (ranked)			
RCA	Speed	**S6=0.512**	S1=0.237	S2=0.151	
	Speed+Volume	**S6=0.725**	S10=0.210		
AEVL	Speed+Volume+Occupancy	S10=0.061	S8=0.051	S7=0.031	S5=0.029

S1 to S10 represents the sensor ID

traffic sensors. The future work will pursue: (1) larger sensor network monitoring in urban roads and (2) distinguishing anomalous traffic events (e.g., accidents, congestion) with sensor anomalies.

REFERENCES

Adenis, P., Wen, Y., & Ray, A. (2012). An inner product space on irreducible and synchronizable probabilistic finite state automata. *Mathematics of Control, Signals, and Systems*, *23*(4), 281–310.

Bengea, S., Li, P., Sarkar, S., Vichik, S., Adetola, V., Kang, K., . . . Kelman, A. (2015). Fault-tolerant optimal control of a building heating, ventilation and air conditioning system. *Science and Technology for the Built Environment*, *21*, 734-751.

Bhuiyan, M. Z. A., Wang, G., & Wu, J. (2009). Target tracking with monitor and backup sensors in wireless sensor networks. In *Computer communications and networks, 2009. iccen 2009. proceedings of 18th internatonal conference on* (pp. 1–6).

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., & Jia, Z. (2001). Freeway performance measurement system: mining loop detector data. *Transportation Research Record: Journal of the Transportation Research Board*(1748), 96–102.

Dailey, D. J. (1999). A statistical algorithm for estimating speed from single loop volume and occupancy measurements. *Transportation Research Part B: Methodological*, *33*(5), 313–322.

Federal size regulations for commercial motor vehicles. (2016). *U.S. Department of Transportation Federal Highway Administration*.

Harris, T., Gamlyn, L., Smith, P., MacIntyre, J., Brason, A., Palmer, R., . . . Slater, A. (1995). 'neural-maine': intelligent on-line multiple sensor diagnostics for steam turbines in power generation. In *Neural networks, 1995. proceedings., ieee international conference on* (Vol. 2, pp. 686–691).

Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade* (pp. 599–619). Springer.

Jeong, H., Kim, H., Lee, S., & Dornfeld, D. (2006). Multisensor monitoring system in chemical mechanical planarization (cmp) for correlations with process issues. *CIRP Annals-Manufacturing Technology*, *55*(1), 325–328.

Jiang, Z., & Sarkar, S. (2015). Understanding wind turbine turbine interactions using spatiotemporal pattern network. In *Proceedings of asme dynamics systems and control*

conference.

Klein, L. A., Mills, M. K., & Gibson, D. R. (2006). *Traffic detector handbook: -volume ii* (Tech. Rep.).

Krishnamurthy, S., Sarkar, S., & Tewari, A. (October 2014). Scalable anomaly detection and isolation in cyber-physical systems using bayesian networks. In *Proceedings of asme dynamical systems and control conference, san antonio, tx, usa.*

Liu, C., Ghosal, S., Jiang, Z., & Sarkar, S. (2016). An unsupervised spatiotemporal graphical modeling approach to anomaly detection in distributed cps. In *Cyber-physical systems (iccps), 2016 acm/ieee 7th international conference on* (pp. 1–10).

Liu, C., Ghosal, S., Jiang, Z., & Sarkar, S. (2017). An unsupervised anomaly detection approach using energy-based spatiotemporal graphical modeling. *Cyber-Physical Systems*. doi: 10.1080/23335777.2017.1386717

Liu, C., Huang, B., Zhao, M., Sarkar, S., Vaidya, U., & Sharma, A. (December 2016). Data driven exploration of traffic network system dynamics using high resolution probe data. In *Proceedings of ieee conference on decision and control, las vegas, nv.*

Minge, E. D., Peterson, S., Weinblatt, H., Coifman, B., & Hoekman, E. (2012). *Loop-and length-based vehicle classification: Federal highway administration-pooled fund program [tpf-5 (192)]* (Tech. Rep.). Minnesota Department of Transportation, Research Services.

Mukherjee, K., & Ray, A. (2014). State splitting and merging in probabilistic finite state automata for signal representation and analysis. *Signal processing*, *104*, 105–119.

Najafi, M., Gulp, C., & Langari, R. (2004). Enhanced auto-associative neural networks for sensor diagnostics (e-aann). In *Fuzzy systems, 2004. proceedings. 2004 ieee international conference on* (Vol. 1, pp. 453–456).

Rao, C., Ray, A., Sarkar, S., & Yasar, M. (2009). Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing*, *3*(2), 101-114.

Ray, A. (2004). Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Processing*, *84*(7), 1115–1130.

Sallans, B., Bruckner, D., & Russ, G. (2005). Statistical model-based sensor diagnostics for automation systems. *IFAC Proceedings Volumes*, *38*(2), 239–246.

Sarkar, S., Mukherjee, K., Sarkar, S., & Ray, A. (2013). Symbolic dynamic analysis of transient time series for fault detection in gas turbine engines. *Journal of Dynamic Systems, Measurement, and Control*, *135*(1), 014506.

Sarkar, S., Sarkar, S., Mukherjee, K., Ray, A., & Srivastav, A. (2013). Multi-sensor information fusion for fault detection in aircraft gas turbine engines. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *227*(12), 1988–2001.

Sarkar, S., Sarkar, S., Virani, N., Ray, A., & Yasar, M. (2014). Sensor fusion for fault detection and classification in distributed physical processes. *Frontiers in Robotics and AI*, *1*, 16.

Sarkar, S., & Srivastav, A. (2016). A composite discretization scheme for symbolic identification of complex systems. *Signal Processing*, *125*, 156–170.

Sarkar, S., Srivastav, A., & Shashanka, M. (2013). Maximally bijective discretization for data-driven modeling of complex systems. In *American control conference (acc), 2013* (pp. 2674–2679).

Solo, V. (2008). On causality and mutual information. In *Decision and control, 2008. cdc 2008. 47th ieee conference on* (pp. 4939–4944).

Wang, R., Zhang, L., Sun, R., Gong, J., & Cui, L. (2011). Easitia: A pervasive traffic information acquisition system based on wireless sensor networks. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 615–621.

Wavetronix smartsensor hd user guide. (2016). *Wavetronix*.

Wells, T. J., Smaglik, E. J., & Bullock, D. M. (2008). Health monitoring procedures for freeway traffic sensors, volume 1: Research report. *Joint Transportation Research Program*, 318.

Wenjie, C., Lifeng, C., Zhanglong, C., & Shiliang, T. (2005). A realtime dynamic traffic control system based on wireless sensor network. In *Parallel processing, 2005. icpp 2005 workshops. international conference workshops on* (pp. 258–264).

Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., & Kaiser, J. (2011). Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Progress in biophysics and molecular biology*, *105*(1), 80–97.

BIOGRAPHIES

Linjiang Wu is a graduate student at Iowa State University, pursuing his M.Sc. degree in Mechanical Engineering. He has received his B.Eng degree in Automotive Engineering in 2012 from Wuhan University of Technology, Wuhan, China and then worked at Ford Motor Company (China) as a product development engineer for over 3 years. Linjiang's research interests are System Dynamics, Transportation and Sensor Health Monitoring.

Chao Liu is a research assistant professor at Department of Thermal Engineering, Tsinghua University, Beijing, China. Previously, he was with the Iowa State University as a Postdoctoral Fellow. He received the B.Sc. degree from Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2013. Dr. Lius research inter-

ests include structure dynamics, machine learning, diagnosis, prognosis, and health monitoring.

Tingting Huang is a Ph.D. Candidate in Transportation Engineering at Iowa State University. She is working at Institute for Transportation as a graduate research assistant since 2014. She focuses on big data analytics in transportation operation and safety. Huangs research interests include data quality control, traffic state estimation, signalized intersection analysis, highway safety and work zone operation analysis. She is served as a reviewer for several committees in Transportation Research Board.

Anuj Sharma is an associate professor in the Civil Construction and Environmental Engineering Department at Iowa State University. He also holds a joint appointment as a Research Scientist with the Institute of Transportation. In these positions, he teaches transportation engineering courses to undergraduate and graduate civil engineering students, conducts research in the transportation operations area, and participates in numerous professional organizations. Dr. Sharma is currently leading the REACTOR (REaltime AnalytiCs of TranspORtation data) laboratory. The lab can ingest multiple streams of real-time data to assist in driving transportation policy decisions. The efforts are focused on ingestion, real-time analytics, batch processing, visualization/front end development, and archiving of numerous data streams.

He coauthored more than 94 peer-reviewed publications including 31 journal papers, 1 book chapters and one patent. He has also served as a reviewer and session chair for several technical journals and conferences.

Soumik Sarkar is an assistant professor of Mechanical Engineering at Iowa State University. Previously, he was with the United Technologies Research Center for 3 years as a Senior Scientist. Dr. Sarkars research interests include Statistical Signal Processing, Machine Learning, Sensor Fusion, Fault Diagnostics and Prognostics, Distributed Control and Complexity Analysis with applications to complex Cyber-Physical Systems such as aerospace, energy and smart building systems, transportation, manufacturing and agriculture systems. He coauthored more than 80 peer-reviewed publications including 30 journal papers, 4 book chapters and one magazine article. He has also served as a reviewer and session chair for several technical journals and conferences. Dr. Sarkar is currently serving as an Associate Editor of Frontiers in Robotics and AI: Sensor Fusion and Machine Perception journal.