Contrast Feature Dependency Pattern Mining for Controlled Experiments with Application to Driving Behavior

Qingzhe Li, Liang Zhao, Yi-Ching Lee
George Mason University
{qli10,lzhao9,ylee65}@gmu.edu

Yanfang Ye
Case Western Reserve University
yanfang.ye@case.edu

Jessica Lin
George Mason University
jessica@gmu.edu

Lingfei Wu
IBM Research
wuli@us.ibm.com

***Abstract*—A controlled experiment is an empirical interventional study method to evaluate the causal impact of an intervention, by identifying the dynamic feature dependency patterns in the contrast multivariate time series (CMTS) collected from the control and experimental groups. Manually labeling or interpreting the effects caused by the intervention from the CMTS data has become an infeasible task even for domain experts. Thus, it is imperative to develop an integrated technique, preferably in an unsupervised manner, that can simultaneously identify and characterize feature dynamic dependencies and their contrast patterns in CMTS, which we call the contrast dynamic feature dependency (CDFD) patterns. In this paper, we propose a generative model with partial correlation-based feature dependency regularization to help analysts understand the CMTS data by jointly 1) characterizing a set of comparable multivariate Gaussian distributions from CMTS, and 2) determining whether the intervention causes the changes between two comparable distributions. Extensive experiments demonstrate the effectiveness and scalability of the proposed method. The proposed method applied to a driving behavior application demonstrates its utility and interpretability.**

I. Introduction

Controlled experiments typically contrast two multivariate time series, which are respectively from a control group and an experimental group, to identify the causal impact of an intervention. We call the control multivariate time series and the experimental multivariate time series together as a *contrast multivariate time series (CMTS)*. In this paper, we focus on quantitatively analyzing the effects of an intervention on drivers' driving behaviors. Driving behavior can be sensed by in-vehicle sensors such as brake or steering wheel positions, and jointly characterized by them via their dependency network. For example, the "steering wheel" will have lower values and the "brake" will have higher values under the driving state of deceleration. Such dependencies among the features of different sensors can form a dependency network to characterize the corresponding driving states. Therefore, the research goal of this paper is to identify whether and how much the intervention makes a difference in causing some "contrast driving behavior" under the same driving state, which raises another non-trivial problem of dynamically extracting the latent states from CMTS. The above problems amount to jointly extract and characterize various latent states, and contrast the behaviors under the intervention as exemplified in Fig. 1.

Although there exists some work for partially handling above tasks in dependency network inference [1], time series subsequence clustering [2], and contrast pattern mining [3], no work has been proposed to address the contrast pattern mining problem for controlled experiments. Several challenges prevent existing work from being directly utilized or trivially combined to address contrast pattern mining problem: **1. Difficulty in integrally modeling the contrast pattern mining problem for controlled experiments.** This problem requires not only identifying the latent states but also detecting the location of contrast patterns, as well as characterizing the contrast patterns. These subproblems tightly couple with each other, and thus should be jointly modeled. Simply using the existing models to solve the subproblems separately will fail to obtain a joint optimal solution. **2. Difficulty in differentiating the patterns featured by the latent states and those caused by the intervention.** The dependency patterns between two latent states are different from those affected by the intervention. We notice that the dependency patterns may significantly change when switching to another latent state, while the dependency patterns may slightly change with and without the intervention under the same latent state.

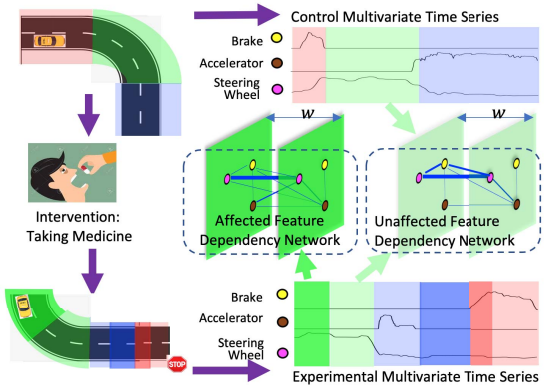


Fig. 1: A controlled driving behavior experiment: The control series without the intervention and the experimental series with the intervention are plotted at the top and bottom portions are covered in red, green and blue colors that denote the latent driving states of deceleration, turning and acceleration, respectively. Under the same latent state, the darker-colored parts denote that the dependency networks are affected by the intervention. For example, both affected and unaffected dependency networks under the "turning" latent state are plotted in the middle portion. Each node denotes the same-colored sensor in multivariate time series. There are w layers in each network to capture the dependencies across small time intervals.

To the best of our knowledge, none of the existing work can address all the above challenges and provide a concrete

2374-8486/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDM.2019.00146

IEEE computer society

TABLE I: Notations

Notation	Description
X,$\hat{X}$	contrast multivariate time series
T, $\hat{T}$	the lengths (i.e. number of rows) of X,$\hat{X}$
$Y, \hat{Y}$	latent state assignments, where $\|Y\| = T$ and $\|\hat{Y}\| = \hat{T}$
Z	contrast pattern indicator for $\hat{X}$, and $\|Z\| = \hat{T}$
$\theta_k, \hat{\theta}_k$	contrast inverse covariance matrices of the k-th latent state
K	the count of the latent states
w	sliding window size
β, γ, λ	regularization parameters

model that formulates the CDFD pattern mining problem for controlled experiments. Our main contributions are as follows:

- **Formulating the dynamic multivariate dependency pattern pattern mining problem for controlled experiments.** We formulate the contrast pattern mining problem to identify the effects of the intervention in an unsupervised and interpretable manner, which jointly optimizes the latent state assignments, contrast pattern detection, and characterization of the contrast dependency patterns in CMTS.
- **Proposing a generative model with partial correlation-based regularization.** We model the CMTS by characterizing the generative process of the control and experimental time series. Then, we propose a new partial correlation-based regularization to differentiate various patterns in CMTS.
- **Conducting experiments on both synthetic and real-world datasets.** The experiments demonstrate the effectiveness of the proposed approach on synthetic datasets. The case study shows the utility and interpretability in a real-world controlled experiment.

II. Problem Setup

We first define the relevant terminologies and then present the new research problem of contrast dynamic feature dependency pattern mining in controlled experiments.

A multivariate time series $x = [x_1, \cdots, x_m]$ is a time-ordered sequence of m vectors where $x_t \in \mathbb{R}^{n\times 1}$ is a multivariate observation that contains n variables at time t. Instead of following the independent and identically distributed (i.i.d.) assumption, the observation of x_t is also dependent on its context. To capture both dependencies among different sensors and nearby time indices, we first concatenate the observations x_t and its $w-1$ successors extracted by a sliding window of size $w \ll m$, which formulates an nw-dimensional row vector $X_t = [x_t^\intercal, \cdots, x_{t+w-1}^\intercal]$ to denote a multivariate time series *subsequence*. Then we stack all these subsequences, from X_1 to X_T, into a matrix $X \in \mathbb{R}^{T\times nw}$ where $T = m - w + 1$. By doing this, the dependencies in original time series x can be represented by the dependencies among the $n \cdot w$ **features**/columns in X. Due to the one-to-one relationship between x and X for a given w, we still call X a multivariate time series. The multivariate time series data X usually exhibits different *latent states* that may dynamically switch over time. These latent states are reflected by both the values of different features and their dependency patterns. For instance, the multivariate time series that record a driving session, can involve three latent states: "Acceleration," "Deceleration," and "Turning." For the "Acceleration" state, its dependency network should contain a strong dependency between the "Accelerator" sensor at time t and time $t+1$, and should not contain any strong dependency between other features. The other two latent states should be characterized by completely different dependency patterns. We use $Y \in \{0,1\}^{T\times K}$ to denote the assignments of the latent state for all subsequences. Specifically, $Y_{t,k} = 1$ if X_t belongs to the k-th latent state; otherwise, $Y_{t,k} = 0$. As X only contains continuous values, each latent state can be naturally characterized by a multivariate Gaussian distribution parameterized by an inverse covariance matrix $\theta_k \in \mathbb{R}^{nw\times nw}$. The inverse covariance matrix θ_k may encode the dependency network $G_k = (X_t, \theta_k)$ as a correlation network or partial correlation network [4] whose nodes denote the features and whose weighted edges denote correlations or partial correlations between connected features.

In controlled experiments, the two multivariate time series are generated from the control and the experimental sessions. They are contrasted to explore the possible differences caused by the intervention. We call the two multivariate time series in the controlled experiments as *contrast multivariate time series (CMTS)*. As other factors are strictly controlled to diminish their effects on the subject, the two multivariate time series usually share the same set of latent states. Formally, the CMTS is defined as follows:

Definition 1. ***[Contrast Multivariate Time Series]*** *The contrast multivariate time series (CMTS) contains two multivariate time series such that 1) the control multivariate time series $X \in \mathbb{R}^{T\times nw}$ are generated without the intervention and the experimental multivariate time series $\hat{X} \in \mathbb{R}^{\hat{T}\times nw}$ are generated with the intervention, 2) X and $\hat{X}$ share the same set of the latent states.*

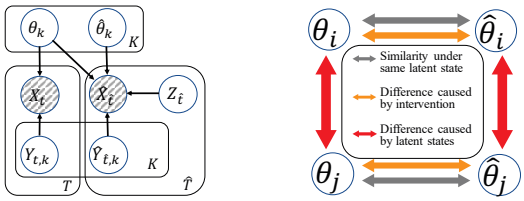
To identify the effects of the intervention, it is natural to contrast their patterns under the same latent state. Concretely, the contrast pattern in controlled experiments is formally defined as follows:

Definition 2. ***[Contrast Dynamic Feature Dependency]*** *For the subsequence $\hat{X}_{\hat{t}}$ belonging to the k-th latent state, if the intervention changes the original feature dependencies θ_k into a new one $\hat{\theta}_k$, there exists the **contrast dynamic feature dependency (CDFD)** pattern at time $\hat{t}$. Hence, we use a* contrast indicator *$Z \in \{0,1\}^{\hat{T}\times 1}$, to signify the existence of CDFD in $\hat{X}$ caused by the intervention. Specifically, $Z_t = 0$ if there exists CDFD in $\hat{X}_t$; otherwise, $Z_t = 1$.*

A driver may accelerate more quickly after she/he takes medicine that stimulates adrenaline in some road segments, which demonstrates the contrast pattern in the controlled experiments. In this paper, our goal is to identify and characterize the CDFD patterns for CMTS in controlled experiments. The problem is formally defined as follows:

Problem Formulation: Given the CMTS X and $\hat{X}$, our goal is to simultaneously discover the interpretable CDFD patterns, including 1) to determine the latent state assignments Y and $\hat{Y}$ for X and $\hat{X}$, respectively, 2) to characterize the K latent states by learning their CDFD patterns $\theta = \{\theta_k\}_k^K$ and $\hat{\theta} = \{\hat{\theta}_k\}_k^K$, and 3) to decide the Z assignments by detecting the CDFD.

For example, for the problem of mining the CDFD patterns in the controlled experiment on driving behavior, in order to test the effectiveness of taking some medicine, the research goals are: 1) to determine the driving state assignments Y and



(a) Plate notations of the generative model

(b) Relationships among the CDFD networks

Fig. 2: The Contrast Pattern Mining Model

$\hat{Y}$, 2) to characterize the K latent driving states encoded by θ and $\hat{\theta}$, and 3) to decide the Z assignments based on whether driving behaviors have been changed after medication.

III. The Methodologies

This section proposes the model of *Contrast dynamic feature dependency Pattern Mining with Partial correlation-based regularization (CPM-P)*. We first propose the generative model, the partial correlation-based regularization and the temporal regularization in turn. Then the overall objective function along with its optimization algorithms are presented.

A. Generative Model of CMTS

The time series subsequences X_t and $\hat{X}_{\hat{t}}$ in CMTS are continuous variables in controlled driving behavior experiments, so they are modeled to be sampled from a set of multivariate Gaussian distributions. The generative process is depicted in Figure 2a. Specifically, by Definition 1 for any $(t = 1, \cdots, T)$, X_t belongs to one of the K latent states, and hence for each subsequence X_t at time index t, we draw $X_t \sim \mathcal{N}(X_t|\theta_k, \mu_k)$ in Figure 2a, where θ_k and μ_k are respectively the inverse covariance matrix and the mean vector to be estimated by the X_t data assigned to the k-th latent state. The joint conditional distribution of X is: $p(X|Y,\theta) = \prod_{k,t}^{K,T} \mathcal{N}(X_t|\theta_k, \mu_k)^{Y_{t,k}}$, where $\theta = \{\theta_k\}_k^K$. Similarly, for any $(\hat{t} = 1, \cdots, \hat{T})$, $\hat{X}_{\hat{t}}$ also belongs to one of the K latent states. However, the vectors $\hat{X}_{\hat{t}}$ that belong to the k-th latent state are possibly generated either from a new distribution $\mathcal{N}(\hat{X}_{\hat{t}}|\hat{\theta}_k, \hat{\mu}_k)$ or from $\mathcal{N}(\hat{X}_{\hat{t}}|\theta_k, \hat{\mu}_k)$. Here, $\hat{\theta}_k$ and $\hat{\mu}_k$ are respectively the inverse covariance matrix and the mean vector to be estimated by the $\hat{X}_{\hat{t}}$ data assigned to the k-th latent state. Specifically, when $Z_{\hat{t}} = 0$ (i.e., the CDFD pattern exists), we draw $\hat{X}_{\hat{t}} \sim \mathcal{N}(\hat{X}_t|\hat{\theta}_k, \hat{\mu}_k)$ in Figure 2a. Otherwise, when $Z_{\hat{t}} = 1$ (i.e., the CDFD pattern does not exist), we draw $\hat{X}_{\hat{t}} \sim \mathcal{N}(\hat{X}_{\hat{t}}|\theta_k, \hat{\mu}_k)$ in Figure 2a. Therefore the conditional joint distribution of $\hat{X}_{\hat{t}}$ is: $p(\hat{X}|\hat{Y}, Z, \theta, \hat{\theta}) = \prod_{k,\hat{t}}^{K,\hat{T}} [\mathcal{N}(\hat{X}_{\hat{t}}|\theta_k, \hat{\mu}_k)^{Y_{t,k}}]^{Z_{\hat{t}}} [\mathcal{N}(\hat{X}_{\hat{t}}|\hat{\theta}_k, \hat{\mu}_k)^{\hat{Y}_{t,k}}]^{(1-Z_{\hat{t}})}$. Based on the equations above, the joint likelihood of $(X, \hat{X})$ conditioned on the parameters $Y, \hat{Y}, Z, \theta$, and $\hat{\theta}$ is:$p(X, \hat{X}|Y, Z, \theta, \hat{\theta}) = p(X|Y,\theta) \cdot p(\hat{X}|\hat{Y}, \theta, \hat{\theta})$. Therefore, given the CMTS $(X, \hat{X})$ data, maximizing the likelihood is equivalent to minimizing the negative log likelihood, which leads to our loss function:

$$\mathcal{L}(Y, \hat{Y}, Z, \theta, \hat{\theta}) = -\sum\nolimits_{t,k}^{T,K} Y_{t,k} \ell\ell(X_t, \theta_k) - \sum\nolimits_{\hat{t},k}^{\hat{T},K} \hat{Y}_{t,k} [Z_{\hat{t}} \ell\ell(\hat{X}_{\hat{t}}, \theta_k) + (1 - Z_{\hat{t}}) \ell\ell(\hat{X}_{\hat{t}}, \hat{\theta}_k)], \quad (1)$$

where $\ell\ell(a, \Gamma) = -\frac{1}{2}(a^\intercal - \mu)^\intercal \Gamma (a^\intercal - \mu) + \frac{1}{2} \log\det\Gamma - \frac{n}{2}\log(2\pi))$ denotes the log likelihood that vector a comes from the Gaussian distribution with the inverse covariance matrix Γ.

B. Partial Correlation-based Regularization

As discussed in the previous section, it is only meaningful to contrast two feature dependency patterns for the same latent state. However, for existing models, it is very difficult to characterize the complicated relationships among feature dependency networks belonging to different latent states with or without the contrast patterns. For example, consider the four feature dependency patterns encoded by $\theta_i, \hat{\theta}_i, \theta_j, \hat{\theta}_j$, as shown in Figure 2b. θ_i and $\hat{\theta}_i$ should be similar (i.e., the grey arrows) such that both of them characterize the i-th latent state, but on the other hand, they should also be different (i.e., the orange arrows) to characterize the effect caused by the intervention. In addition, θ_i and θ_j (or $\hat{\theta}_i$ and $\hat{\theta}_i$) should characterize the differences between the i-th and j-th latent states in a different way (i.e., the red arrows). To ensure θ_k and $\hat{\theta}_k$ are characterizing the same latent state, the difference between θ_k and $\hat{\theta}_k$ should be regularized under appropriate metrics. Traditionally, the inverse covariance matrices are mostly regularized by penalizing the element-wise differences with an L1-norm or L2-norm (e.g., [1]). However, these regularizations are flawed because a single element in the inverse covariance matrix does not have any mathematical or statistical meaning. Moreover, the scales and values of the non-diagonal elements depend on the diagonal elements. Both flaws prohibit directly regularizing the inverse covariance matrices. To address the above flaws, we propose to regularizing the element-wise distance between the partial correlation coefficient matrices. Doing so has the following advantages: 1) The partial correlation captures the feature dependency better than other correlations. 2) The partial correlation coefficients share the same scale and range. So it is more reasonable to apply element-wise distance on partial correlation matrices ρ_k and $\hat{\rho}_k$ rather than inverse covariance matrices θ_k and $\hat{\theta}_k$. Therefore, we propose a new partial correlation based regularization as follows:

$$\mathcal{R}_{\mathcal{C}}(\theta, \hat{\theta}) = \lambda \cdot \sum\nolimits_k^K \|\rho_k - \hat{\rho}_k\|_F^2,$$

where the elements of the partial correlation matrices are $\rho_{k,i,j} = -\theta_{k,i,j}/(\theta_{k,i,i}\theta_{k,j,j})^{\frac{1}{2}}$ and $\hat{\rho}_{k,i,j} = -\hat{\theta}_{k,i,j}/(\hat{\theta}_{k,i,i}\hat{\theta}_{k,j,j})^{\frac{1}{2}}$.

C. Temporal Regularization

Due to the nature of temporal continuity in time series, neighboring points tend to have consistent latent state assignments and contrast indicator values. We thus penalize the divergence of the assignments between the neighboring time indices by proposing the following smoothing term:

$$\mathcal{R}_{\mathcal{T}}(Y, \hat{Y}, Z) = \sum\nolimits_{t=2}^{T} \gamma \mathbb{1}(Y_t \neq Y_{t-1}) + \sum\nolimits_{\hat{t}=2}^{\hat{T}} \beta \mathbb{1}(Z_{\hat{t}} \neq Z_{\hat{t}-1}) + \gamma \mathbb{1}(\hat{Y}_{\hat{t}} \neq \hat{Y}_{\hat{t}-1})$$

where $\mathbb{1}(\cdot)$ is an indicator function that maps "True" values to 1 and "False" values to 0, β is the penalty if $Z_t \neq Z_{t-1}$, and γ is the penalty of switching among the K latent states.

D. The Overall Objective Function:

Based on above components, the overall objective function of the proposed CPM-P model is as follows:

$$\arg\min_{\theta, \hat{\theta}, Y, \hat{Y}, Z} \quad \mathcal{L}(Y, \hat{Y}, Z, \theta, \hat{\theta}) + \mathcal{R}_{\mathcal{C}}(\theta, \hat{\theta}) + \mathcal{R}_{\mathcal{T}}(Y, \hat{Y}, Z), \quad (2)$$

where $\{\theta_k, \hat{\theta}_k\} \succ 0$ are positive definite matrices such that $\log\det(\cdot)$ is defined in a valid domain. The hyper-parameters K and w, can be chosen based on prior knowledge, through cross-validation, or by a principled method such as the Bayesian information criterion [5]. If the number of subsequences assigned to any latent state is too small (e.g. <30) to learn a good θ_k and $\hat{\theta}_k$, this indicates that the value of K should be decreased. Since the short term temporal dependency is much

Algorithm 1 Overall Algorithm for optimizing CPM-P model

Require: $X, \hat{X}, K, w, n, \lambda, \beta, \gamma$
Ensure: solution $Y, \hat{Y}, Z, \theta, \hat{\theta}$
1: $\{Y, \hat{Y}\} \leftarrow$ random initialization
2: $Z \leftarrow \mathbf{0}$
3: **repeat**
4: **for** $k = 1, \cdots, K$ **do**
5: $\Phi_k \leftarrow \{X_t | Y_{t,k} = 1\} \bigcup \{\hat{X}_{\hat{t}} | \hat{Y}_{\hat{t},k} = 1 \text{ AND } Z_{\hat{t}} = 1\}$
6: $\Psi_k \leftarrow \{\hat{X}_{\hat{t}} | \hat{Y}_{\hat{t},k} = 1 \text{ AND } Z_{\hat{t}} = 0\}$
7: $[\theta_k, \hat{\theta}_k] \leftarrow$ ADMM_solver$(\lambda, \Psi_k, \Phi_k)$
8: **end for**
9: $Y \leftarrow$ Updating Y by fixing θ
10: $(\hat{Y}, Z) \leftarrow$ Updating $\hat{Y}$ and Z by fixing θ and $\hat{\theta}$
11: **until** $Y, \hat{Y}$ and Z assignments are stationary
12: **return** $Y, \hat{Y}, Z, \theta, \hat{\theta}$

stronger than the long term one in real-world applications, the window size w should be small (e.g. $w < 10$).

E. Model Optimization

In this section, we briefly introduce our parameter optimization algorithm for the proposed CPM-P model. The details and implementation of the algorithm are provided in [6].

The overall objective function defined in Equation (2) is a mixture of combinational optimization of discrete variables (i.e., $Y, \hat{Y}, Z$) and continuous variables (i.e., $\theta, \hat{\theta}$) with the non-convex term (i.e. the partial correlation-based regularization term). Jointly optimizing these variables is prohibitively difficult to be solved by the existing algorithms. To address this challenge and optimize the proposed model, we develop an Expectation Maximization (EM)-like optimization algorithm outlined in Algorithm 1. After a random initialization of the discrete variables' assignments, Lines 3-12 alternatively optimize the continuous variables and discrete variables until the discrete assignments are stationary. Specifically, the maximization step (M-step) optimizes θ and $\hat{\theta}$ in Lines 4-8, and then the expectation step (E-step) optimizes the $Y, \hat{Y}$ and Z assignments in Lines 9-10.

IV. Experiments

The performance of the proposed CPM-P method is evaluated on synthetic and real-world datasets with different settings.

A. Evaluation on Synthetic Datasets

Because of the unsupervised nature and the lack of publicly available real-world datasets with labels for the contrast pattern mining problem, we first evaluate the effectiveness of the proposed CPM-P model on 3 synthetic datasets generated with random latent state assignments.

Generating the Synthetic Datasets: Due to the space limitation, please refer to [6] for this part.

Evaluation metrics: To compare the effectiveness of the proposed method and other methods, the predicted latent state and CDFD pattern assignments are evaluated with the ground truth labels described above. To ensure a fair comparison of effectiveness among all methods, the number of latent states K in all the methods is fixed to the corresponding K used to generate the datasets. All methods are evaluated as a clustering problem with K clusters for the latent state assignments by using macro F_1 score defined as the mean F_1 scores of all clusters. The predicted CDFD pattern assignments are evaluated as an anomaly detection problem by using F_1 score defined as the harmonic mean of the precision and recall. The closer the (macro) F_1 score to 1, the better the result.

TABLE II: The (macro) F_1 scores of predicted $Y, \hat{Y}$, and Z assignments

Method	Dataset 1			Dataset 2			Dataset 3		
	Y	$\hat{Y}$	Z	Y	$\hat{Y}$	Z	Y	$\hat{Y}$	Z
K-means+1SVM	0.50	0.51	0.58	0.33	0.34	0.61	0.28	0.27	0.60
K-means+EE	0.50	0.51	0.23	0.33	0.34	0.25	0.28	0.27	0.25
K-means+IF	0.50	0.51	0.23	0.33	0.34	0.26	0.28	0.27	0.26
K-means+LOF	0.50	0.51	0.15	0.33	0.34	0.18	0.28	0.27	0.21
K-shape+1SVM	0.51	0.51	0.54	0.34	0.33	0.56	0.26	0.24	0.55
K-shape+EE	0.51	0.51	0.23	0.34	0.33	0.25	0.26	0.24	0.25
K-shape+IF	0.51	0.51	0.24	0.34	0.33	0.26	0.26	0.24	0.25
K-shape+LOF	0.51	0.51	0.14	0.34	0.33	0.19	0.26	0.24	0.21
TICC+1SVM	**0.99**	0.72	0.47	0.29	0.24	0.48	0.25	0.23	0.51
TICC+EE	0.99	0.72	0.35	0.29	0.24	0.25	0.25	0.23	0.25
TICC+IF	0.99	0.72	0.29	0.29	0.24	0.27	0.25	0.23	0.25
TICC+LOF	0.99	0.72	0.30	0.29	0.24	0.20	0.25	0.23	0.25
GMM+1SVM	0.95	0.87	0.49	0.85	**0.80**	0.50	0.83	**0.78**	0.52
GMM+EE	0.95	0.87	0.22	0.85	0.80	0.22	0.83	0.78	0.24
GMM+IF	0.95	0.87	0.23	0.85	0.80	0.24	0.83	0.78	0.25
GMM+LOF	0.95	0.87	0.16	0.85	0.80	0.18	0.83	0.78	0.21
Baseline ($\lambda = 0$)	0.94	**0.92**	**0.89**	**0.86**	0.63	**0.88**	**0.83**	0.59	**0.76**
CPM-P (ours)	**0.99**	**0.99**	**0.98**	**0.99**	**0.75**	**0.89**	**0.99**	**0.99**	**0.98**

Comparison methods and our methods: To the best of our knowledge, there is no integrated method capable of mining CDFD pattern for CMTS generated from controlled experiments. Therefore, all the comparison methods need to predict the latent state and the CDFD pattern assignments in two steps. In Step 1, the subproblem of determining latent state assignments are equivalent to the time series subsequence clustering problem. We compared with two traditional distance-based clustering methods, including the classic K-means and the state-of-the-art K-shape [7] methods, and two model-based methods, including the classic Gaussian Mixture Models (GMM) [8] and the state-of-the-art TICC [2] methods. For Step 2, the contrast patterns detection problem can be evaluated as an anomaly detection problem. The control time series, which only contain the "normal" data, are used to train the anomaly detection model, and the experimental time series, which contain both "normal" and "abnormal" data, are used to detect the anomalies. Therefore, we consider four anomaly detection methods for the contrast pattern detection problem: one-class support vector machine (1SVM) [9] (with linear kernel), Elliptic Envelope (EE) [10], Isolation Forest (IF) [11], and Local Outlier Factor (LOF) [12]. For all these anomaly detection methods, the default values are used for all parameters except for the "outlier ratio," which is set to 50%. That is the same as the "true" outlier ratio in the synthetic datasets to get relatively good results for the comparison methods. Notice that our method does not need prior knowledge of the outlier ratio. To validate the proposed regularization terms, the baseline method, which only contains our loss function and the temporal regularization term by setting $\lambda = 0$ in Equation (2), is also considered. For our CPM-P method, we set the hyper-parameters in our methods by $\lambda = 1000, \beta = 2, \gamma = 10$.

Performance: In this section, the effectiveness of the comparison methods and the proposed models are evaluated on Dataset 1, Dataset 2, and Dataset 3, which respectively include two, three, and four latent states. The (macro) F_1 scores of the predicted latent state (i.e.,$Y, \hat{Y}$) and CDFD pattern (i.e.,Z) assignments are shown in Table II. The method named with the "+" sign in its name is a two-step method. For each column, the two best performers are written in bold and the best performer is underlined. As the results show, our method is always one of the top-two performers in all (macro) F_1 scores. Our method is 22% better on average than most the best comparison methods

except for the baseline method using our model without the partial correlation-based regularization term. The distance-based methods perform worse than the model-based methods, as opposed to the dependency-based patterns in these datasets. After intensively tuning the hyper-parameters, TICC achieves the macro F_1 score of 0.99 on Y assignments in Dataset 1, but still fails in other datasets because Dataset 1 only contains two latent states, which is a relatively simple dataset. For the datasets with more than two latent states, the TICC method with the $L1$ regularization term still suffers from differentiating the latent states and the contrast patterns caused by the intervention. GMM generally performs better than other comparison methods on Y and $\hat{Y}$ assignments. For Z assignments, all comparison methods except for our methods and the baseline method do not perform well with the highest score at 0.60, which is still close to random guessing. Because the contrast patterns highly depend on the latent state assignments, the imperfect results on the latent state assignments lead to worse results for the CDFD pattern assignments. As none of the comparison methods can solve the CDFD pattern mining problem, we will focus on analyzing the results of our method in the rest of this section. The performance of the baseline is better than other comparison methods, which validates the effectiveness of our generative model proposed in Section III-A, but still worse than our CPM-P method, which validates the usefulness of our partial correlation-based regularization.

B. Evaluations on Driving Behavior Experiments

In this section, we apply our CPM-P method to a controlled driving behavior experiment, which evaluates the influence of an attention deficit hyperactivity disorder (ADHD) medicine by contrasting driving behaviors [13] [14].

In this controlled experiment, the same driver is asked to drive twice on the same high-fidelity driving simulator, once before and once after taking the ADHD medicine. Specifically, the control multivariate time series is recorded from the driving session before taking the ADHD medicine, and the experimental multivariate time series is recorded from the driving session after taking the medicine. The driving simulator can record the multivariate observation in six dimensions, which are Brake (B), steering Wheel (W), Accelerator (A), Velocity (V), latitude, and longitude. We use the first three dimensions' data (i.e., 3-D time series) to predict the latent states and the contrast patterns, then use the velocity time series and the latitude-longitude time series (i.e.,trajectory) to validate the predictions by our method. We choose the number of latent states $K = 4$ for this dataset and for any value of $K \geq 4$, the model still assigns most of the points to four latent states. The other hyper-parameters are set as follows: $w=5, \lambda=1000, \beta=5, \gamma=10$.

We first examine the latent state assignments (i.e., the Y and $\hat{Y}$ assignments) predicted by our model. The results are plotted in Figure 3. Each of the predicted latent states can be visually validated by the velocity time series and the trajectory plotted in the bottom portion. For example, for the velocity time series segments corresponding to the blue latent state, the velocity first decreases to 0 and then increases, which can be interpreted as "stopping at stop sign." The green segments in the velocity time series mostly increase or remain stable, which can be interpreted as driving straight. All the corners in both trajectories belong to the orange latent state, which corresponds the "turning" latent state. The red latent state is interpreted as deceleration, since red segments in the velocity time series always decrease.

To examine the discovered contrast patterns, we use the closeness centrality score [15] to determine the "significance" of each feature F_i in the partial correlation network, where the nodes are the features, the weight of the edge between two features is the absolute value of the partial correlation coefficient, $F \in \{W, A, V, B\}$ denotes the variables, and i denotes the relative index within the subsequences. For example, the feature B_0 denotes the feature corresponding to the brake variable at the first observation of the subsequence. Then the closeness centrality scores are computed for each node in the partial correlation network. The higher the closeness centrality score, the more significant the node is, and the more important this feature is. Finally, we plot the zero-normalized closeness centrality scores for each latent state in Figure 4. It is easily seen that, first, each latent state has a unique "shape" in terms of the relative importance of the features; second, the areas under the same latent state mostly overlap, which validates the partial correlation-based regularization for pairing the latent states; and third, the differences of the centrality scores under the same latent state is one way to visualize the contrast pattern. In addition, the contrast patterns mined by our model are highly interpretable. For example, the plots of the "Turning" latent state suggest the steering wheel plays a relatively more important role after taking the medication. It can be interpreted as the driver being more likely to turn the vehicle by proactively adjusting the steering wheel rather than adjusting the brake and the accelerator after taking the ADHD medicine. This contrast pattern indicates the driver turns the vehicle with less variation of speed, which is a safer driving behavior [16]. For another example, the plots of the "Deceleration" latent state show that the centrality score of the feature B_0 is higher before taking the medicine, and the centrality scores of features B_1 to B_4 are higher after taking the medicine. This observation suggests that 1) before taking the ADHD medicine, the driver is likely to use the brake in the early stage of the deceleration period but is unlikely to continue braking through the entire deceleration period, which is a recognized as a typical ADHD driving behavior [16]; 2) after taking the ADHD medicine, the driver reduces the speed by using the brake in a more consistent way during the deceleration, which is closer to a normal driver's driving behavior.

V. Related Work

The work related to this research is summarized as follows: **Dependency-based multivariate time series clustering:** Most dependency-based time series clustering approaches such as those based on an autoregressive moving average model [17], Gaussian mixture model [8], [18], or hidden Markov model [19], typically consider the whole sequence rather than subsequence. Recently, Hallac et al. proposed the Toeplitz

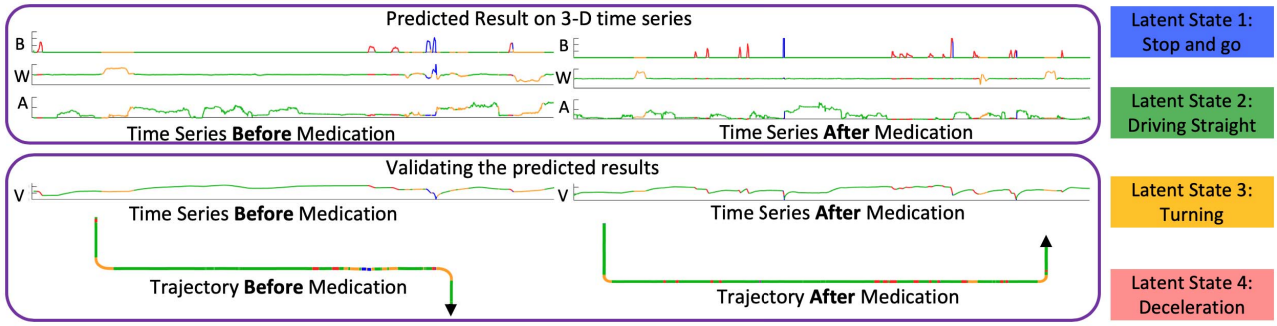


Fig. 3: ADHD medicine controlled experiment case study: The predicted latent states in the multivariate time series are plotted in the top portion. Each predicted latent state plotted can be interpreted as a driving state, which can be validated in the lower portion.

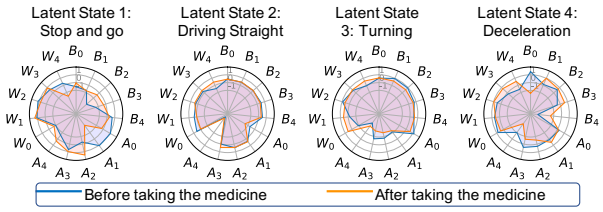


Fig. 4: The contrast patterns are analyzed by the (zero-normalized) closeness centrality scores of all features (e.g., B_0). Under the same latent state, the areas of before and after taking the medication mostly overlap with each other. The non-overlapped areas indicate the driving behaviors have been changed by the medicine.

Inverse Covariance-based Clustering (TICC) [2] method to cluster the subsequences in a single multivariate time series according to connectivity patterns estimated by a graphical lasso. However, TICC only focuses on single time series, which neither considers the relationship between the control and experimental time series nor mines their contrast patterns.

Contrast pattern mining for time series: Research on contrast pattern mining between two multivariate time series has recently emerged. Researchers have explored multivariate time series generated in functional MRI to mine the contrast patterns by proposing various network inference models [3], [20], [21]. For instance, Lee et al. proposed a CNN based deep neural network [20] to identify contrasting dependency networks inferred from the entire time series. Similarly, Liu et al. proposed a contrast graphical lasso model [3] for whole time series. The model derives a single contrast dependency network that corresponds to two multivariate time series. However, neither of these methods considers the fact that the contrast patterns are only meaningful while they are compared under the same latent state in the subsequence level.

VI. Conclusion

In this paper, we define a new contrast pattern mining problem for controlled experiments with CMTS data. We propose a novel CPM-P model to formulate the contrast pattern mining problem as an optimization problem, which integrates latent state identification, dynamic feature dependency inference, and contrast pattern detection. Experiments on both synthetic and real-world dataset validate its effectiveness and interpretability.

Acknowledgement

This work was supported by the National Science Foundation grant: #1841520, #1755850, and NVIDIA GPU Grant. Y. Ye's work was partially supported by the NSF under grants CNS-1618629, CNS-1814825, CNS-1845138, OAC-1839909 and III-1908215, the NIJ 2018-75-CX-0032.

References

[1] D. Hallac, Y. Park, S. Boyd, and J. Leskovec, "Network inference via the time-varying graphical lasso," ser. KDD, 2017, pp. 205–213.
[2] D. Hallac, S. Vare, S. Boyd, and J. Leskovec, "Toeplitz inverse covariance-based clustering of multivariate time series data," ser. KDD, 2017, pp. 215–223.
[3] X. Liu, X. Kong, and A. B. Ragin, "Unified and contrasting graphical lasso for brain network discovery," ser. SIAM '17, 2017, pp. 180–188.
[4] S. L. Lauritzen, *Graphical models*. Clarendon Press, 1996, vol. 17.
[5] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, pp. 461–464, 1978.
[6] "Code and supplemental material," https://github.com/qingzheli/partial-correlation-based-contrast-pattern-mining.
[7] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD*, 2015, pp. 1855–1870.
[8] J. D.Banfield and A. E.Raftery, "Model-based gaussian and non-gaussian clustering," *Biometrics*, pp. 803–821, 1993.
[9] B. Schölkopf, J. C. Platt, and et.al., "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
[10] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
[11] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *TKDD*, vol. 6, no. 1, p. 3, 2012.
[12] H.-P. K. Markus M Breunig and et.al., "Lof: identifying density-based local outliers," in *ACM sigmod record*, vol. 29, no. 2, 2000, pp. 93–104.
[13] D. Barragan and Y.-C. Lee, "Pre-crash driving behaviour of individuals with and without adhd," in *6th International Conference on Driver Distraction and Inattention, Gothenburg, Sweden*, 2018.
[14] Y.-C. Lee, C. W. McIntosh, and et al., "Design of an experimental protocol to examine medication non-adherence among young drivers diagnosed with adhd: A driving simulator study," *Contemporary clinical trials communications*, vol. 11, pp. 149–155, 2018.
[15] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.
[16] M. J. Groom, E. Van Loon, D. Daley, and et al., "Driving behaviour in adults with attention deficit/hyperactivity disorder," *BMC psychiatry*, vol. 15, no. 1, p. 175, 2015.
[17] Y. Xiong and D.-Y. Yeung, "Time series clustering with arma mixtures," *Pattern Recognition*, pp. 1675–1689, 2004.
[18] J. Luo, A. Brodsky, and Y. Li, "An em-based ensemble learning algorithm on piecewise surface regression problem," *International Journal of Applied Mathematics and Statistics*, vol. 28, no. 4, pp. 59–74, 2012.
[19] P. Smyth, "Clustering sequences with hidden markov models," in *NIPS'97*, M.C.Mozer, M.I.Jordan, and T.Petsche, Eds., 1997, pp. 648–654.
[20] J. B. Lee, X. Kong, Y. Bao, and C. Moore, "Identifying deep contrasting networks from time series data: Application to brain network analysis," ser. SIAM '17, 2017, pp. 543–551.
[21] A. Brodsky and J. Luo, "Decision guidance analytics language (dgal) - toward reusable knowledge base centric modeling," in *17th ICEIS*, vol. 1, 2015, pp. 67–78.