

LSTMを用いた任意パラメータ指定可能な ネットワークトラヒック生成

栗山 海渡 渡部 康平

長岡技術科学大学 大学院工学研究科

目次

- ◆ 背景・目的
- ◆ 関連研究
- ◆ 提案モデル
- ◆ 実験
- ◆ まとめ

目次

◆背景・目的

◆関連研究

◆提案モデル

◆実験

◆まとめ

背景・目的

- ◆近年、ユーザの利用する端末が多様化し、トラフィックが増加
- ◆そのため、ネットワークの評価が重要になってきている
 - ◆ex) 負荷テスト, シミュレーション etc...
- ◆ネットワークの評価には、**大量のトラフィックデータ**が必要となる
 - ◆これを用意するのは、コストやプライバシーの観点から難しい

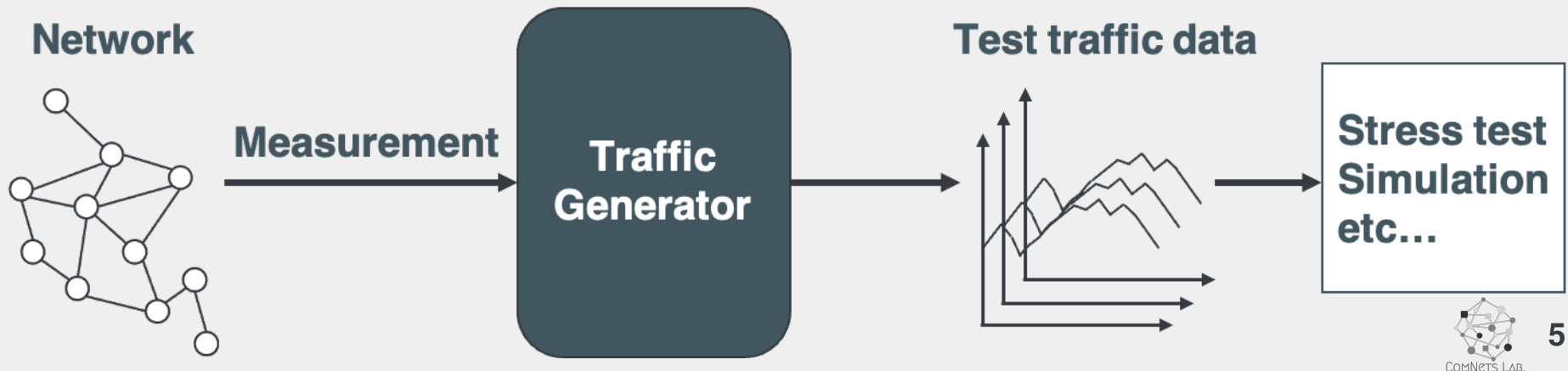


背景・目的

- ◆一般的に**トラフィックジェネレータ**を用いて試験用トラフィックを生成
- ◆既存のトラフィックジェネレータにおける問題
 - ◆リアルなトラフィックを作るのが困難
 - ◆任意の特性を持ったトラフィックが生成できない

◆目的

**トラフィックの特性を多面的に再現し、任意パラメータ指定可能な
新たなトラフィックジェネレータを提案**



目次

◆背景・目的

◆関連研究

◆提案モデル

◆実験

◆まとめ

一般的な生成モデル

- ◆ 確率分布を推定する生成モデルは、以下の2つに分類される
 - ◆ 統計学的手法に基づくモデル
 - ◆ Markov Modulated Poisson Process (MMPP)
 - ◆ Auto Regressive Integrated Moving Average (ARIMA)
 - ◆ Covariation Orthogonal (CO) 線形予測
 - ◆ 機械学習に基づくモデル
 - ◆ Convolutional Neural Network (CNN)
 - ◆ Long Short Term Memory (LSTM) [2]
 - ◆ Generative Adversarial Networks (GAN) [1]
- ◆ 通信トラフィックを対象とした生成モデルは限定的であり、十分に研究がなされていない

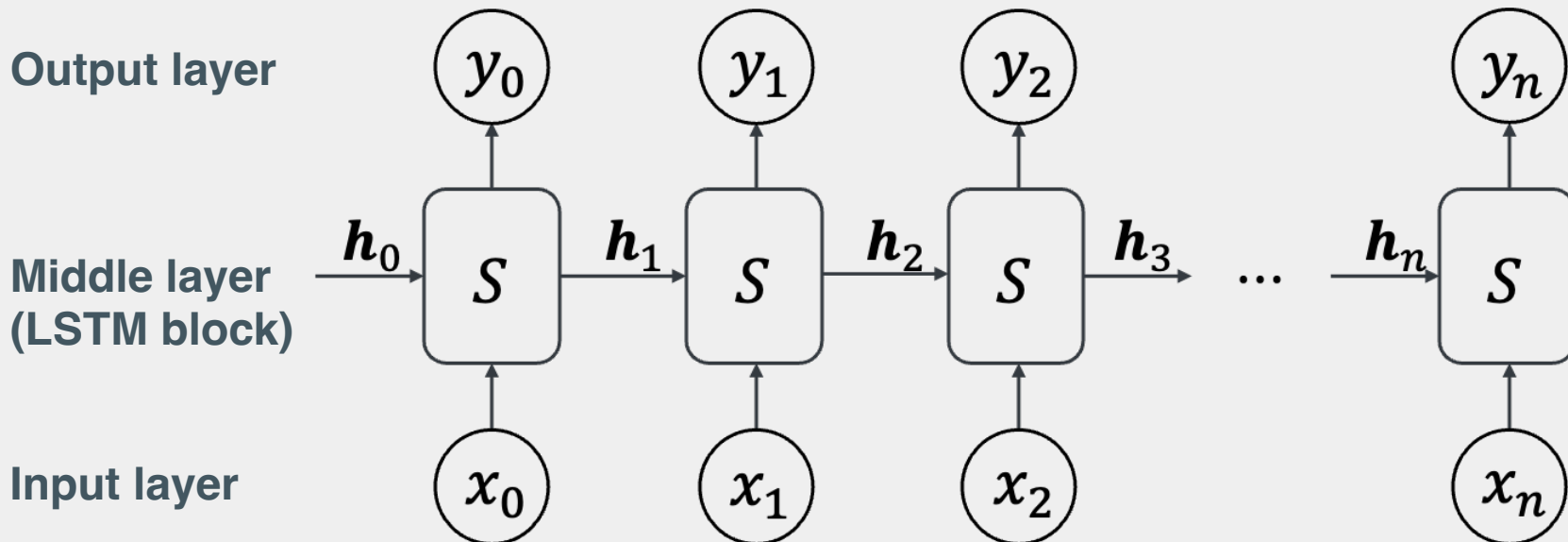
LSTM

◆長期的な依存関係を学習できるため、時系列データに対してよく用いられる

◆LSTMの構成

◆セル (h) : セル状態に情報を追加または削除していく

◆ゲート: シグモイド層 (σ) によって、0~1の数値を出力することで前の情報 (h) を引き継ぐ



GAN

◆本物のデータに類似したデータを生成する教師なし学習モデル

◆GANの構成

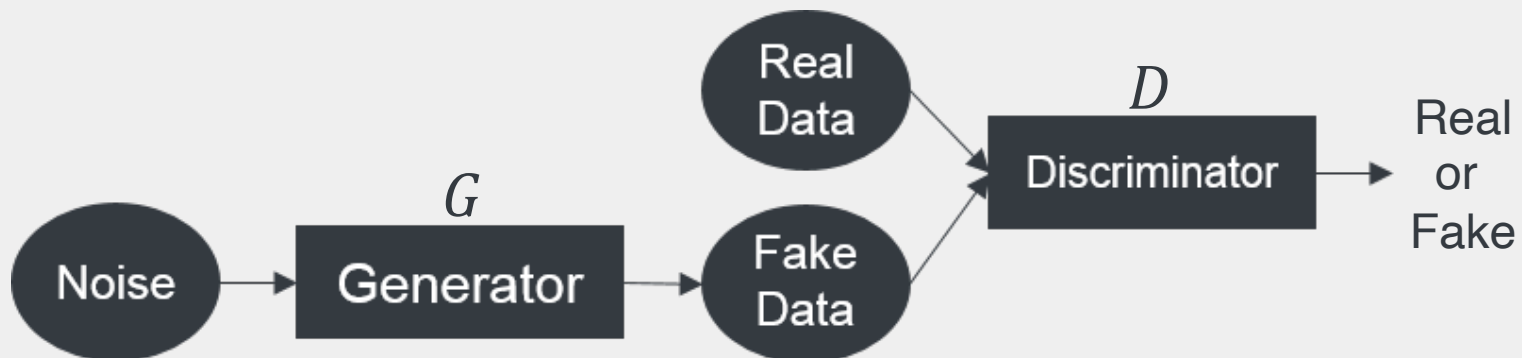
◆識別器(Discriminator)

◆学習データなら1, 生成データなら0となるように学習

◆生成器(Generator)

◆Discriminatorを騙せるほど類似したデータを生成

◆Discriminatorの出力を1に近づけるように学習



目次

◆背景・目的

◆関連研究

◆**提案モデル**

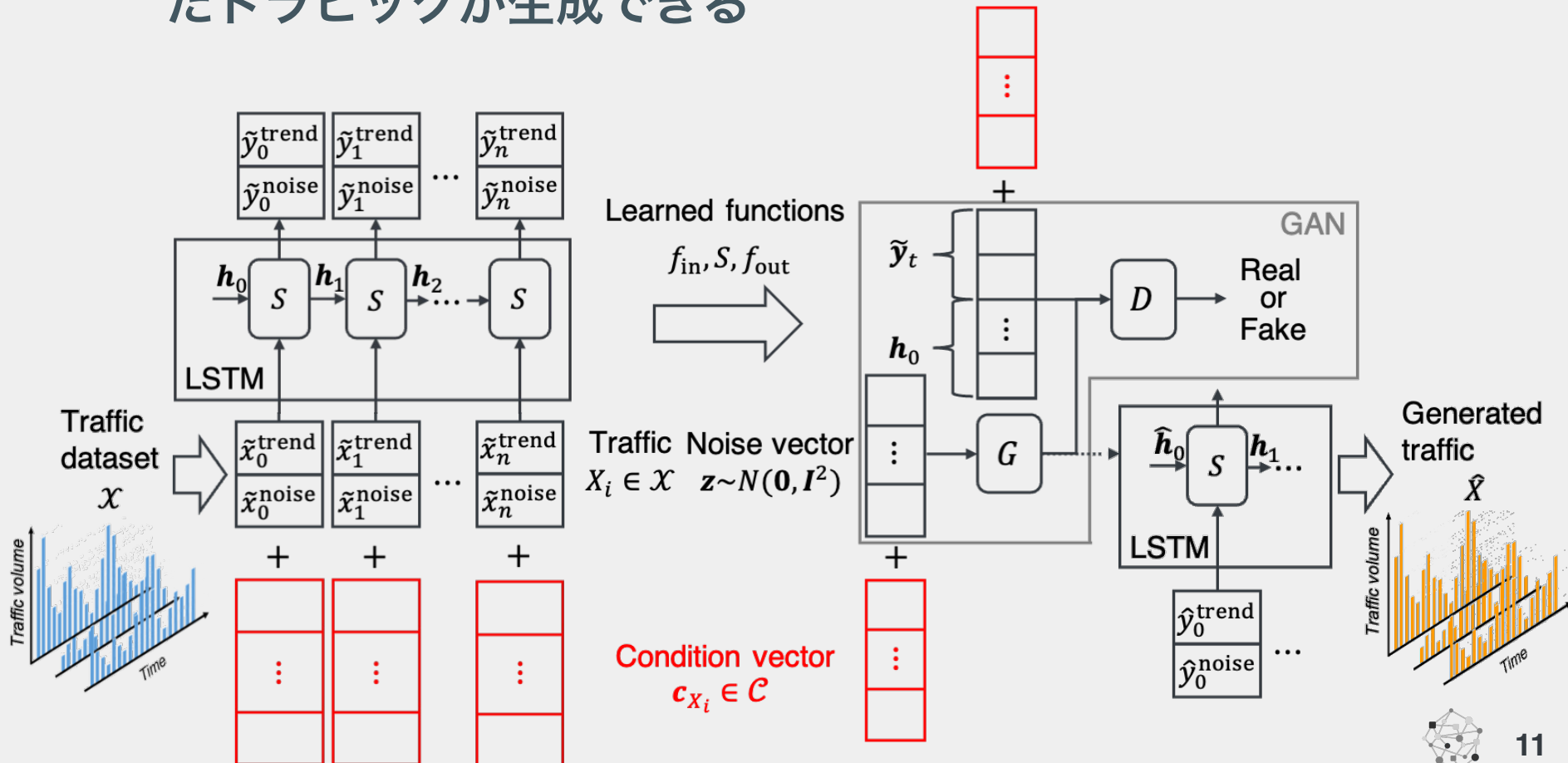
◆実験

◆まとめ

提案モデルの構造

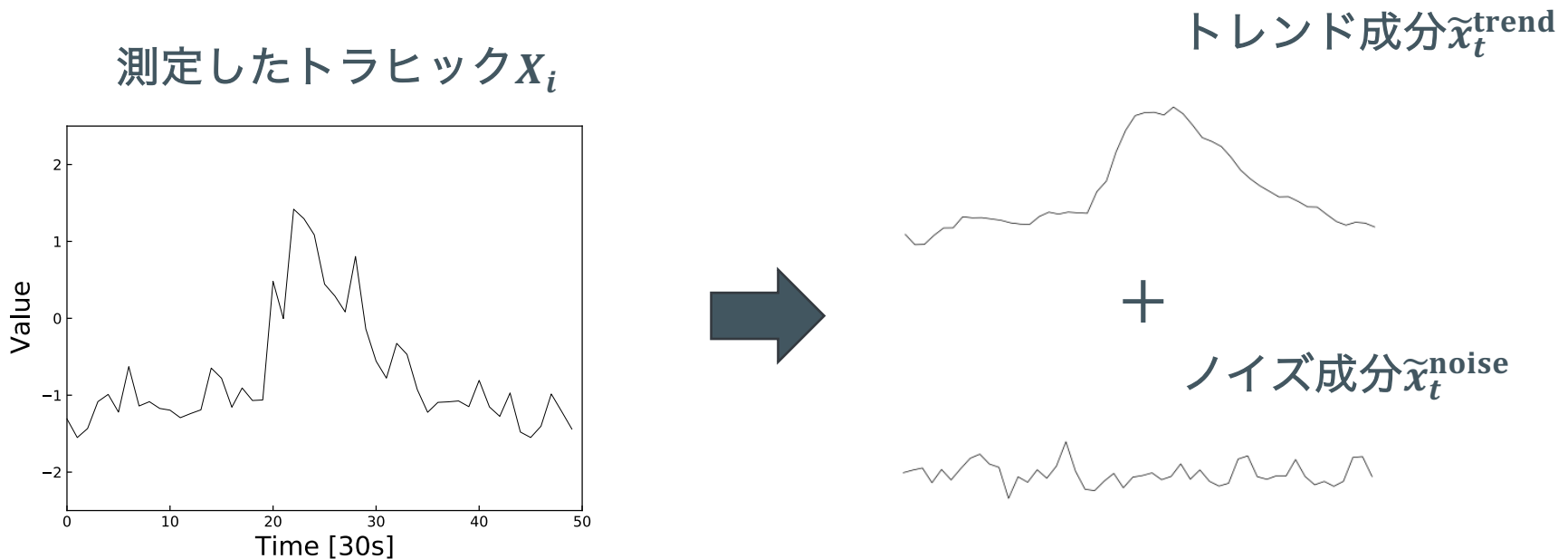
◆ LSTMとGANを組み合わせたトラフィック生成モデル

- ◆ **特徴量 (condition)** を入力に付与することで、特徴量を反映したトラフィックが生成できる



提案モデルの構造: 前処理

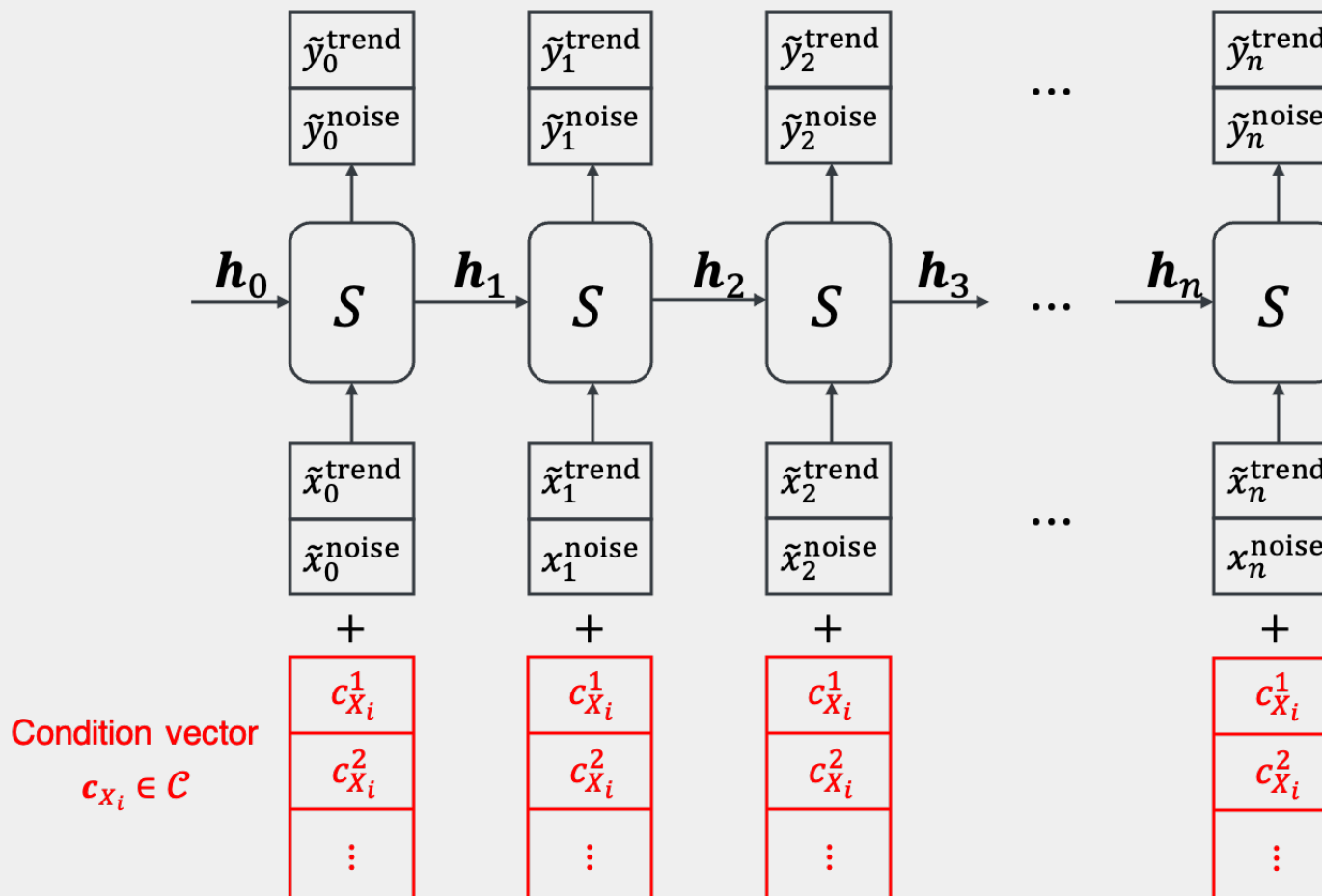
- ◆トレンド成分 $\tilde{x}_t^{\text{trend}}$ とノイズ成分 $\tilde{x}_t^{\text{noise}}$ に分割する
 - ◆トレンド成分 $\tilde{x}_t^{\text{trend}}$: Exponential Moving Average (EMA)
 - ◆ノイズ成分 $\tilde{x}_t^{\text{noise}}$: 測定したトラフィック – EMA



提案モデルの構造: LSTM

◆ LSTMに分割されたトラフィック $[\tilde{x}_t^{\text{trend}}, \tilde{x}_t^{\text{noise}}]^T$ を入力する

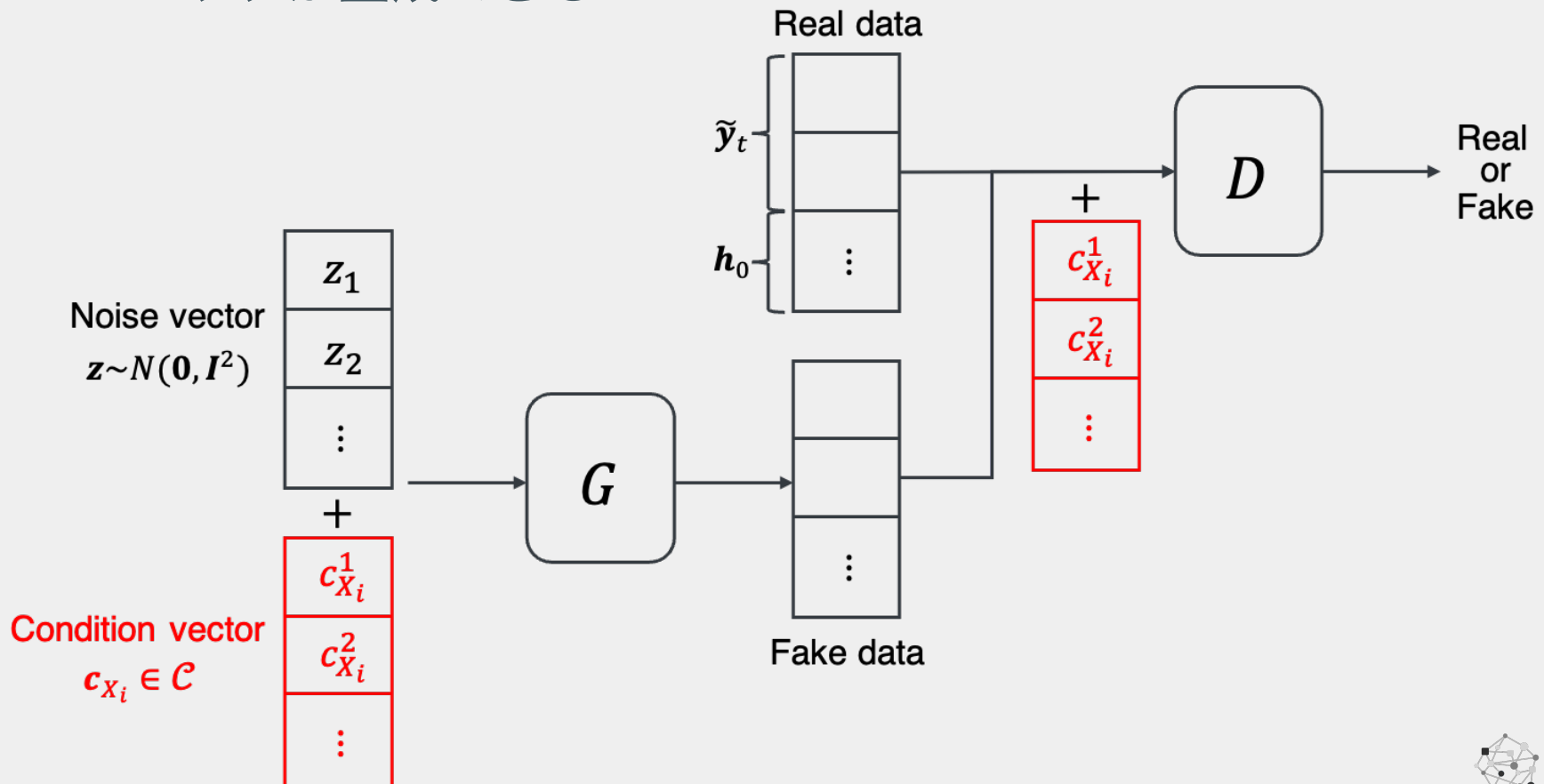
◆ 特徴量 (condition) を付与することで、特徴量を学習する



提案モデルの構造: GAN

◆GANでLSTMが学習した時系列的な特性を学習する

◆特徴量 (condition)を付与することで、特徴量を反映したトラフィックが生成できる



目次

◆背景・目的

◆関連研究

◆提案モデル

◆**実験**

◆まとめ

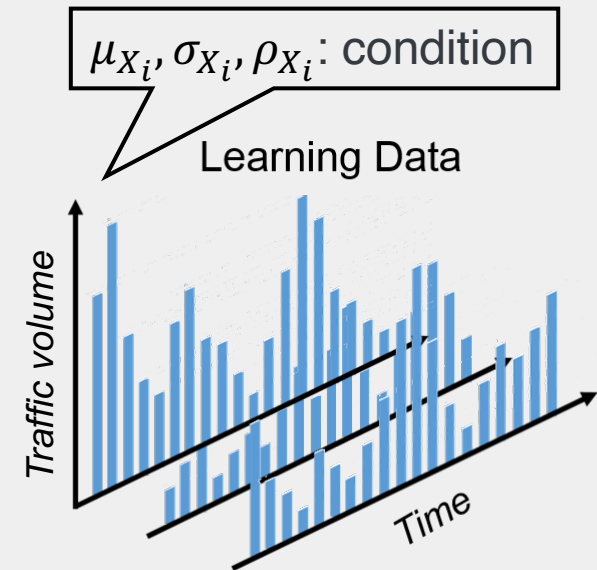
特徴量

◆特徴量とは

- ◆トラフィックが持つ固有の値のこと
- ◆特徴量は各トラフィック毎に異なる値を持つ

◆今回、学習に用いる特徴量は以下の通り

- ◆平均: μ_{X_i}
- ◆標準偏差: σ_{X_i}
- ◆トレンド (線形近似の傾き): ρ_{X_i}



◆特徴量の値を特徴量ベクトル $[\mu_{X_i}, \sigma_{X_i}, \rho_{X_i}]^T$ に調整した生成を目指す

学習データの作成

◆データセット [5]

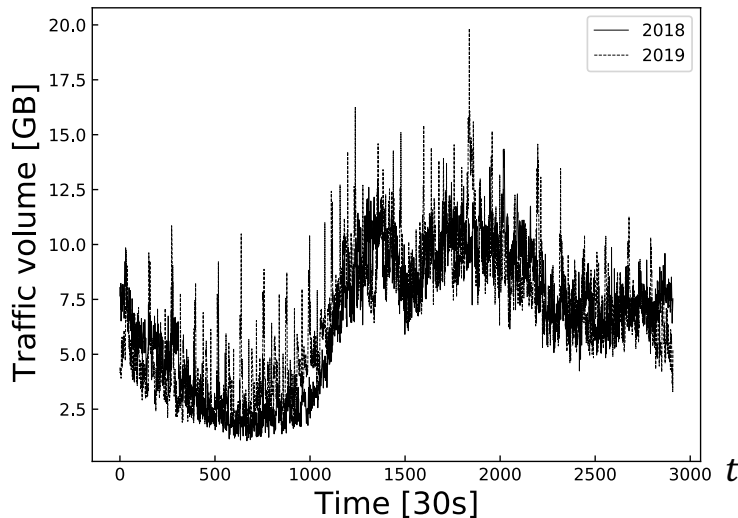
◆ある1日の30秒毎に何バイト通信されたかを表したデータ

◆学習データ

◆前処理として、標準化を行う (平均値 : 0, 標準偏差 : 1)

◆ t を50ずつずらして学習データを作成 ($t = 0 \sim 50, 51 \sim 100, \dots$)

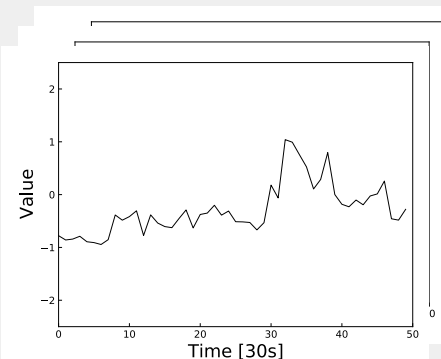
◆**ランダム**に訓練データとテストデータを分割



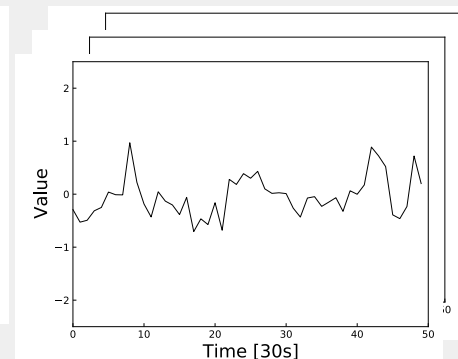
データセット(24, 24時間)



訓練データ
(80個)



テストデータ
(20個)



学習データ(100個)

パラメータ設定

◆LSTMのハイパーパラメータ

- ◆層数: 4
- ◆隠れ層の次元: 256
- ◆最適化関数: SGD
- ◆学習回数: 50000
- ◆学習率: 0.01

◆GANのハイパーパラメータ

- ◆正規乱数ベクトルの次元: 100
- ◆最適化関数: Adam
- ◆バッチサイズ: 2
- ◆学習回数: 5000
- ◆学習率: 0.01

評価指標

- ◆ Kolmogorov-Smirnov (KS)検定
 - ◆ グラフ分布における空間的な特性の類似度を評価
- ◆ フーリエ変換, 自己相関
 - ◆ 時系列的な特性の類似度を評価
- ◆ 平均, 標準偏差: μ_{X_i}, σ_{X_i}
 - ◆ 平均, 標準偏差の分布の類似度を評価
- ◆ トレンド: ρ_{X_i}
 - ◆ 時系列の長期的傾向の類似度を評価

訓練データの特徴を統計的に反映しつつ, 訓練データには存在しない, 新たなトラヒックを生成したい

- 訓練データとテストデータの特徴 (統計量) と **同程度の類似度を持つデータ**が望ましい

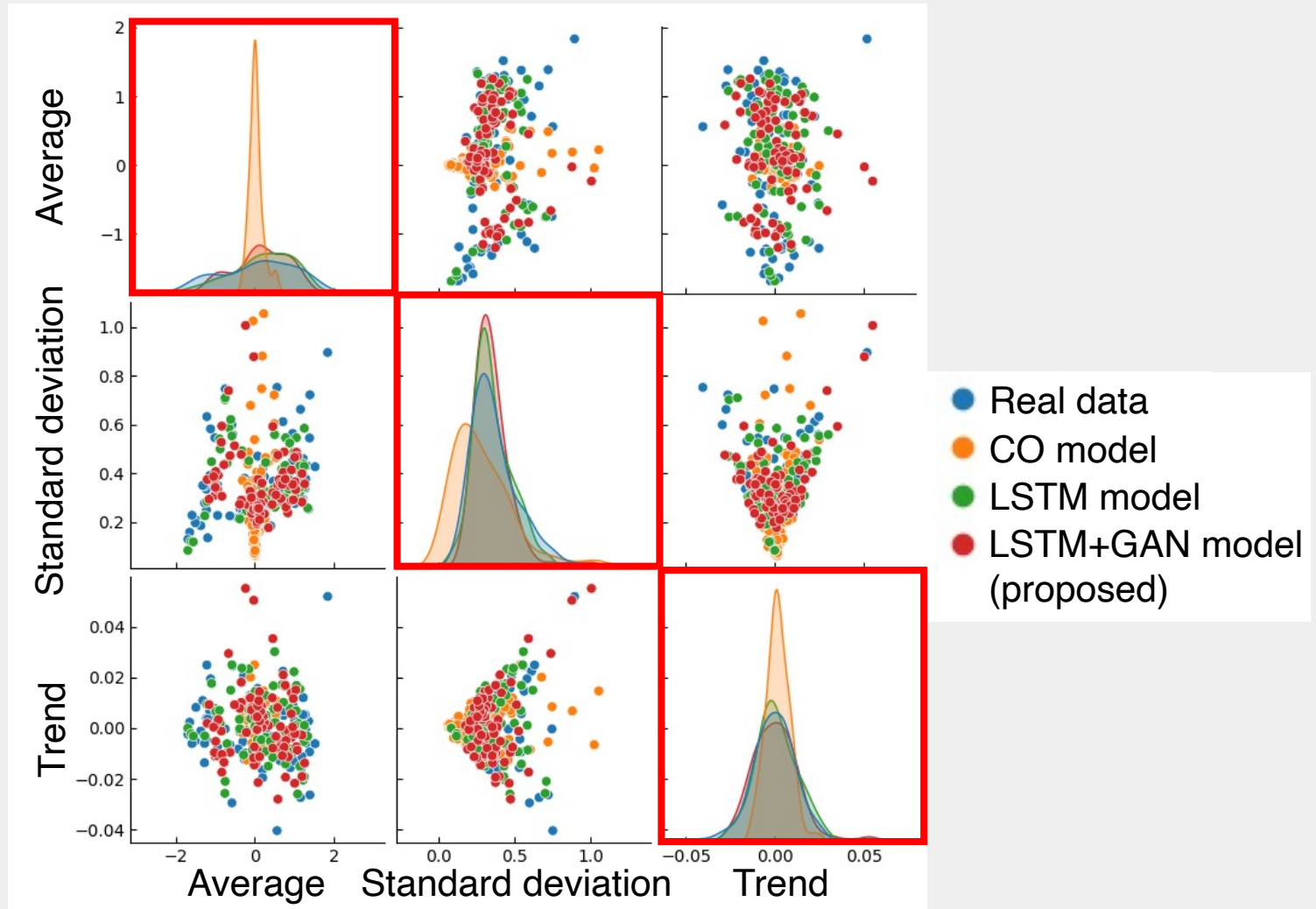
特徴量を指定しない生成結果: 評価指標

◆特徴量を指定しない生成結果の評価

Evaluation index	CO model	LSTM model	Proposed model	Real data
Average of KS statistics	0.347	0.223	0.229	0.259
RMSE of frequency spectrum	0.00163	0.00141	0.00124	0.00145
RMSE of autocorrelation	0.0899	0.0489	0.0690	0.0313
Average	0.0163	0.168	0.154	0.0719
Standard deviation	0.297	0.348	0.347	0.368
Trend	0.00163	-0.00197	0.00115	-0.000277

- ◆ LSTM及び提案モデルがほぼ全ての評価指標において同程度の精度
 - 多面的な観点から見て、再現精度が優れているといえる

特徴量を指定しない生成結果: 分布



◆ LSTM及び提案モデルがReal dataの分布とほぼ一致している

➤ 多面的な観点から見て、再現精度が優れているといえる

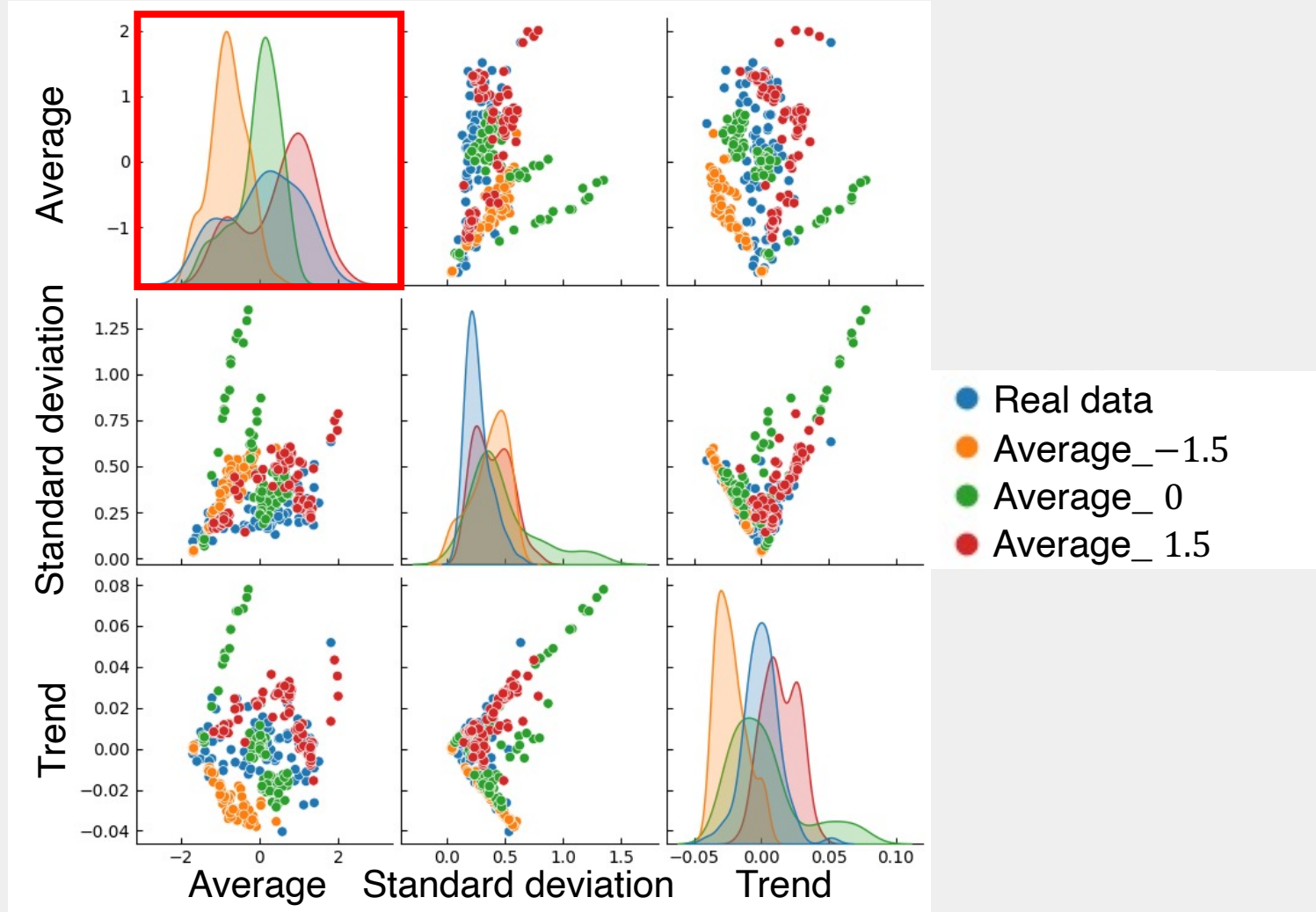
特徴量を指定した生成結果: 評価指標

◆特徴量を指定した生成結果の評価

Evaluation index	condition	Proposed model
Average	-1.5	-0.529
	0	-0.00785
	1.5	0.698
Standard deviation	0.2	0.428
	0.4	0.460
	0.6	0.468
Trend	-0.02	-0.0147
	0	-0.000281
	0.02	0.0123

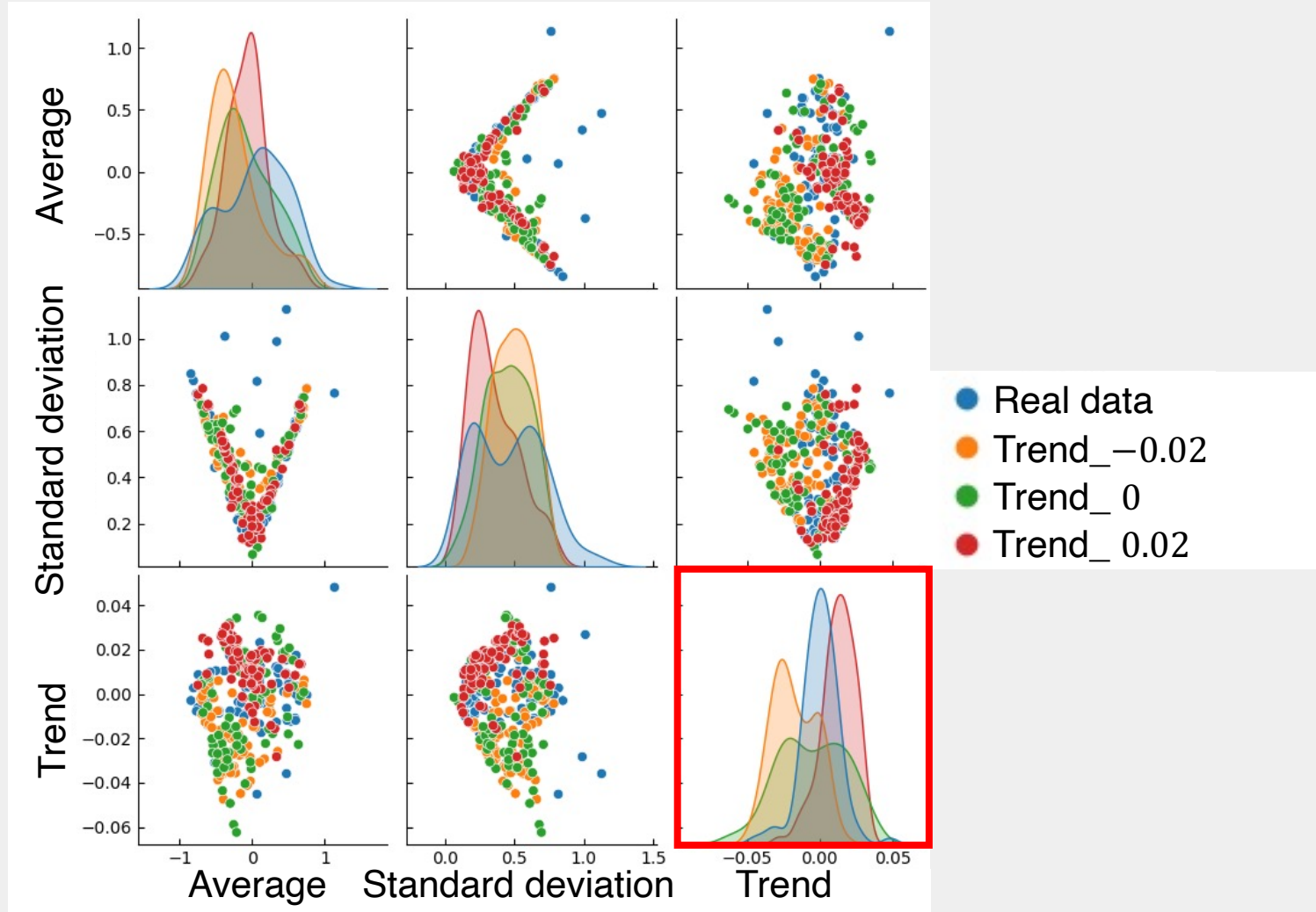
- ◆特徴量 (condition)を指定することで生成結果が変化している
 - 任意の特徴量を調整したトラヒックが生成できたといえる

特徴量を指定した生成結果: 分布 (平均を指定)



- ◆ 特徴量を指定することによって、分布が変化していることがわかる
 - 任意の特徴量を調整したトラフィックが生成できた

特徴量を指定した生成結果: 分布 (トレンドを指定)



- ◆ 特徴量を指定することによって、分布が変化していることがわかる
 - 任意の特徴量を調整したトラフィックが生成できた

目次

◆背景・目的

◆関連研究

◆提案モデル

◆実験

◆まとめ

まとめ・今後の予定

◆まとめ

- ◆LSTMとGANを組み合わせたトラフィック生成モデルの提案
- ◆実際のトラフィックトレースを使用して比較評価を実施
 - **トラフィック特性を多面的に再現**したトラフィックが生成できた
 - **任意の特徴量を調整したトラフィック**が生成できた
 - しかし、指定した特徴量と値が離れていた

◆今後の予定

- ◆GANによるデータセットの拡張
- ◆他の特徴量の検討

参考文献

- ◆ [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), 2014.

- ◆ [2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol.9, no.8, pp.1735–1780, 1997.

- ◆ [5] C. Zhang, X. Ouyang, and P. Patras, “Traffic Data Repository at the WIDE Project,” Proceedings of 2000 USENIX Annual Technical Conference (USENIX ATC 2000) FREENIX Track, 2000.