

REAL-VALUED (MEDICAL) TIME SERIES GENERATION WITH RECURRENT CONDITIONAL GANs

1. 소개

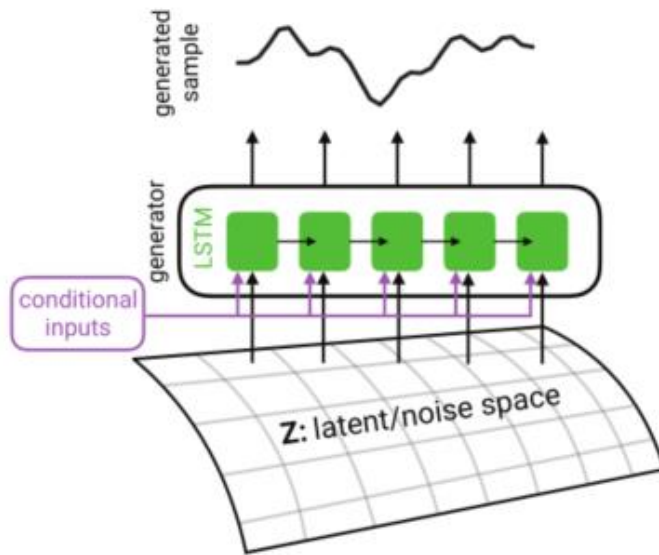
- 1. 시계열 데이터를 생산하기 위해 반복 신경망을 가진 GAN을 사용
- 2. 이런 GAN으로 생성된 데이터 평가
 - 실제 데이터로 쓰일 만큼 사실적인지 판단(maximum mean discrepancy[MMD]을 이용)
 - 데이터가 유용한지 판단
 - 훈련 데이터가 아닌지 판단

2. 데이터를 생산하는 이유

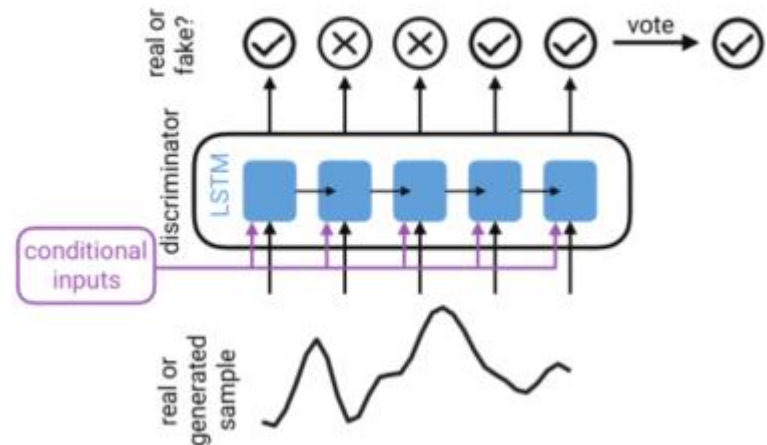
- 1. 데이터 공유 문제 해결 - 환자의 개인정보
- 2. 데이터 수 증가 - 적은 환자 수, 개인정보 등
- 3. 실험 - 적은 데이터 수

3. Recurrent GAN and Recurrent Conditional GAN

- 판별자와 생성자 둘다 반복 신경망(예: LSTM)을 사용



generator



discriminator

3. Recurrent GAN and Recurrent Conditional GAN

- Discriminator의 loss 함수

$$D_{\text{loss}}(X_n, \mathbf{y}_n) = -\text{CE}(\text{RNN}_D(X_n), \mathbf{y}_n)$$

- Generator의 loss 함수

$$G_{\text{loss}}(Z_n) = D_{\text{loss}}(\text{RNN}_G(Z_n), \mathbf{1}) = -\text{CE}(\text{RNN}_D(\text{RNN}_G(Z_n)), \mathbf{1})$$

4. Maximum Mean Discrepancy (MMD)

$$\widehat{\text{MMD}}_u^2 = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K(x_i, x_j) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m K(x_i, y_j) + \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j \neq i}^m K(y_i, y_j)$$

- Kernel 함수

$$K(x, y) = \exp(-\|x - y\|^2 / (2\sigma^2)).$$



MMD 설명 pptp

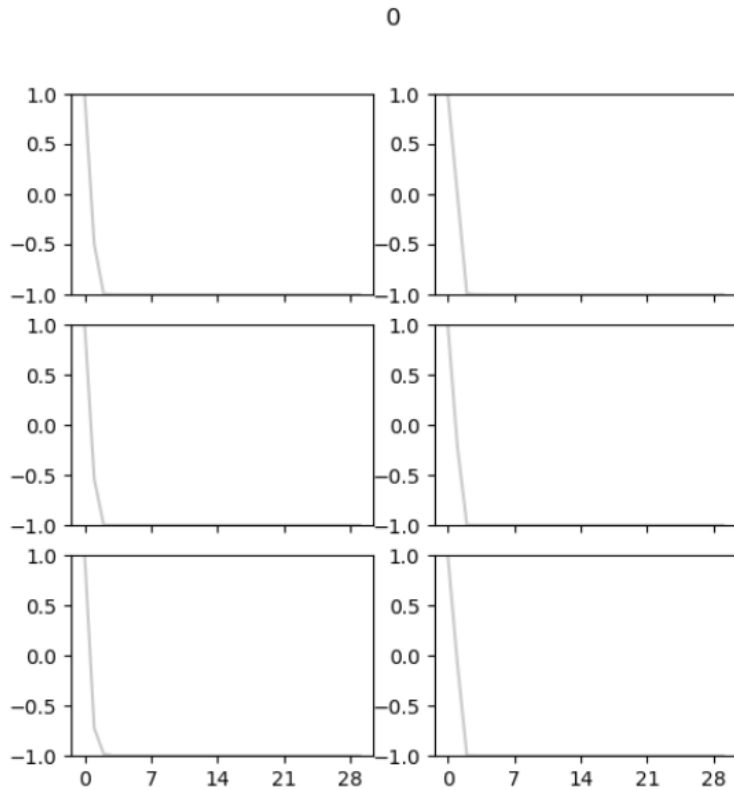
5. Train on Synthetic, Test on Real (TSTR)

- 우리는 GAN에 의해 생성된 데이터 세트를 사용하여 모델을 훈련시키고, 그런 다음 실제 예들을 가지고 시험한다.

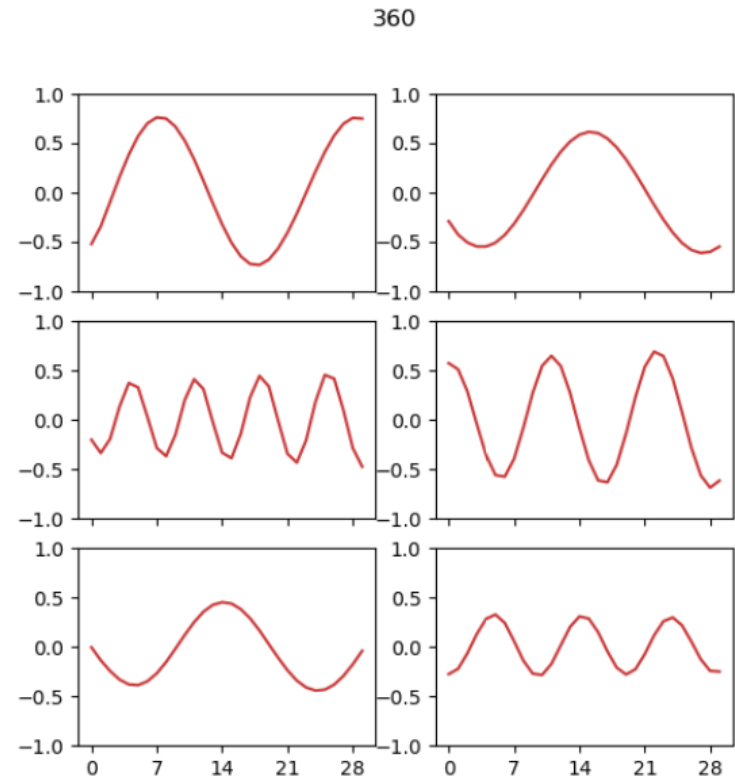
Algorithm 1 (TSTR) Train on Synthetic, Test on Real

```
1: train, test = split(data)
2: discriminator, generator = train_GAN(train)
3: with labels from train:
4:   synthetic = generator.generate_synthetic(labels)
5:   classifier = train_classifier(synthetic, labels)
6:   If validation set available, optionally optimise GAN over classifier performance.
7: with labels and features from test:
8:   predictions = classifier.predict(features)
9:   TSTR_score = score(predictions, labels)
```

6. Sine waves



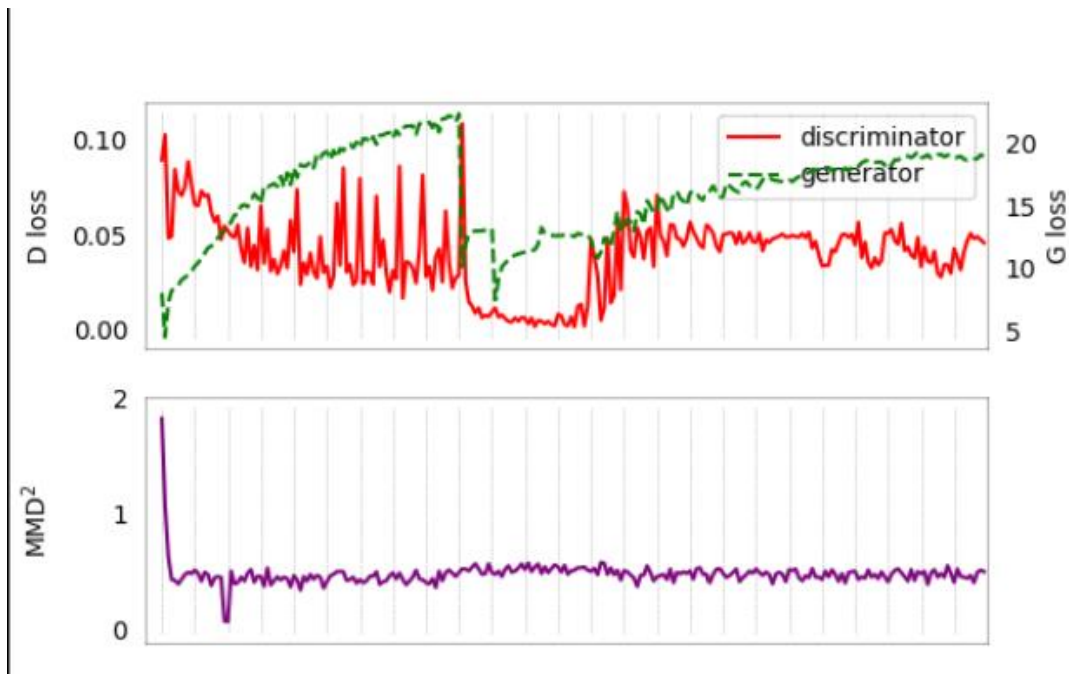
epoch 0



epoch 250

6. Sine waves

GAN sample 평가



REAL-VALUED (MEDICAL) TIME SERIES GENERATION WITH RECURRENT CONDITIONAL GANs

감사합니다