

[논문리뷰] An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling

<https://arxiv.org/abs/1803.01271#:~:text=We%20conduct%20a%20systematic%20evaluation%20of%20generic%20co>
250911 BOAZ WEEK9 과제

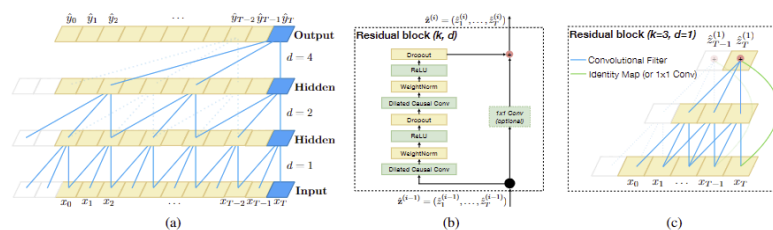
제안 배경

- 기존 순환 신경망(RNN/LSTM/GRU) 기반 sequence 모델의 한계
 - 장기 의존성 문제
 - 학습/추론 속도가 느리고, 병렬 연산 어려움
 - 최근 연구에서 CNN이 특정 작업에서 RNN보다 우수한 성능을 보일 수 있음이 보고됨.

TCN(Temporal Convolutional Networks) → CNN 기반 구조를 시퀀스 모델링에 적용

About TCN

- 1D Fully Convolutional Network 구조 기반 → 기존 RNN이 가지는 기억 유지력 문제와 병렬 연산 불가능 문제 해결



- 핵심
 - (1) Causal Convolution
 - 제안 배경 : 일반 CNN은 양방향 정보 활용, BUT sequence 에서는 미래 정보를 알고 있으면 안됨.
 - 시점 t의 출력이 과거 입력(t 이전)에만 의존하도록 제약
 - 제약 방법 = CNN의 필터가 미래 정보를 참조하지 않도록 Zero Padding 추가
 - (2) Dilated Convolution (사진 (a))
 - 제안 배경: 단기 기억 소실 문제 해결(RNN 한계)
 - 커널 사이에 공백을 두어서 깊이 쌓지 않으면서 긴 시퀀스를 포착
 - 입력 시퀀스를 일정 간격으로 건너뛰면서 적용
 - 수식

$$F(s) = (\mathbf{x} *_{\mathbf{d}} f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-di}$$

- (3) Residual Convolution (사진 (b), (c))
 - 제안 배경: 기울기 소실 문제
 - Dilated Conv → Normalization → ReLU → Dropout → Dilated Conv (잔차 연결 추가)
 - 입력과 출력을 skip-connection으로 연결
 - Gradient 소실 완화, 학습 안정성 및 깊은 네트워크 학습 가능

Experiments

: RNN과 성능 비교 & 동일 조건 하에서의 LSTM/GRU와 성능, 속도, 안정성 비교

▼ Dataset

1. Synthetic Stress Tests
 - Adding Problem (장기 기억 유지력 테스트)
 - Copy Memory Task (정보 보존력 테스트)
2. MNIST 및 변형된 P-MNIST
 - 순차적인 픽셀 데이터를 처리하는 능력 평가
3. 음악 데이터
 - JSB Chorales, Nottingham Dataset
4. 자연어 처리 (NLP) 데이터
 - PennTreebank (PTB)
 - WikiText-103
 - LAMBADA (긴 문맥 이해력 테스트)

Sequence Modeling Task	Model Size (\approx)	Models			
		LSTM	GRU	RNN	TCN
Seq. MNIST (accuracy ^b)	70K	87.2	96.2	21.5	99.0
Permuted MNIST (accuracy)	70K	85.7	87.3	25.3	97.2
Adding problem $T=600$ (loss ^d)	70K	0.164	5.3e-5	0.177	5.8e-5
Copy memory $T=1000$ (loss)	16K	0.0204	0.0197	0.0202	3.5e-5
Music JSB Chorales (loss)	300K	8.45	8.43	8.91	8.10
Music Nottingham (loss)	1M	3.29	3.46	4.05	3.07
Word-level PTB (perplexity ^e)	13M	78.93	92.48	114.50	88.68
Word-level Wiki-103 (perplexity)	-	48.4	-	-	45.19
Word-level LAMBADA (perplexity)	-	4186	-	14725	1279
Char-level PTB (bpc ^e)	3M	1.36	1.37	1.48	1.31
Char-level text8 (bpc)	5M	1.50	1.53	1.69	1.45

▼ TASK Conclusion

1 Copy Memory Task

- RNN(LSTM, GRU)은 장기 의존성 처리 어려움
- TCN은 긴 지연 시퀀스에서도 안정적으로 정보 복원

2 Sequential MNIST & Permuted MNIST

- 순차적 픽셀 입력 기반 분류
- TCN이 LSTM/GRU보다 더 나은 정확도

3 Language Modeling

- Penn Treebank, WikiText-103에서 비교
- 성능은 LSTM 수준 이상, 학습 속도는 더 빠름

4 Music Modeling (Polyphonic)

- 멀티레이어 LSTM과 유사한 성능 확보