
Clase 138 — 1D CNNs y WaveNet

Parte: 2 — Deep Learning · Fuente: Géron, cap. 15 § Handling Long Sequences + WaveNet paper (van den Oord et al. 2016). Duración estimada: 60 min.

Clase 138 — 1D CNNs y WaveNet

Parte: 2 — Deep Learning · Fuente: Géron, cap. 15 § Handling Long Sequences + WaveNet paper (van den Oord et al. 2016). Duración estimada: 60 min.

Objetivo

Conocer la alternativa a RNN para secuencias: Conv1D y WaveNet (dilated causal convolutions). Más rápido que LSTM (paralelizable), receptive field amplio con pocas capas (dilated convolutions). Útil para audio, series temporales y como capa de preprocesamiento.

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Aplicar Conv1D(filters, kernel_size, padding='causal') sobre secuencias.
- Implementar dilated convolutions (kernel salta posiciones).
- Reconocer causal convolution: el output en t depende solo de inputs $\leq t$.
- Construir un mini-WaveNet con dilation rates exponenciales (1, 2, 4, 8, ...).
- Comparar Conv1D vs LSTM en speed/accuracy.

Temas

- Conv1D fundamentos: stride, padding, dilation.
- Causal padding: para no ver el futuro.
- Dilated convolutions: receptive field grande sin más layers.
- WaveNet: stack de dilated causal convolutions (1, 2, 4, ..., 512).
- Conv1D vs LSTM: paralelización, receptive field, parameter count.

Definiciones y características

- Conv1D: convolución sobre una sola dimensión espacial/temporal.
- padding='causal': padding asimétrico de modo que output[t] depende solo de input[0..t].
- dilation_rate: gap entre los elementos del kernel. dilation=2 con kernel=3 \rightarrow ve t, t-2, t-4.
- Receptive field: rango temporal que afecta el output. Con dilated stack 1,2,4,8,16: $2^N + 1$.
- WaveNet: red de Conv1D dilated causal stackeadas, originalmente para generación de audio.

Dataset / recursos

- Serie temporal del ejercicio 119.
- Audio simple (sintético: superposición de senos).
- Librerías: tensorflow, keras.

Ejercicios

1. Conv1D vs LSTM: forecasting de serie. Conv1D(32, 5, padding='causal') \rightarrow Conv1D(32, 5,

- padding='causal') → Flatten → Dense(1). Comparar con LSTM equivalente.
- 2. Dilation rates: stack de 4 Conv1D con dilation_rate {1, 2, 4, 8}. Calcular receptive field.
- 3. Speed test: medir tiempo de training Conv1D vs LSTM para misma data. Conv1D suele ser 5-20× más rápido en GPU.
- 4. WaveNet mini: implementar stack de 10 Conv1D causal con dilations {1, 2, 4, ..., 512} para una serie larga.
- 5. Visualización del receptive field: para un output [t], marcar qué inputs lo afectan.

Homework verificable

Forecasting del dataset eléctrico (clase 119) con mini-WaveNet:

1. Stack de 6 Conv1D causal dilated con rates 1, 2, 4, 8, 16, 32.
2. Cada capa con 32 filtros, kernel 2.
3. Comparar MAE en test vs LSTM del ejercicio 119.

Criterio de aceptación: el WaveNet debe ser competitivo o mejor que LSTM, y notablemente más rápido en GPU.

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
padding='same' en forecasting causal	Ve el futuro. Fix: padding='causal'.
Stack sin dilation → receptive field chico	Necesitás muchas capas. Fix: dilations exp
Conv1D + Flatten → Dense con muchos params	Modelo grande. Fix: GlobalAvgPool1D antes
Mezclar Conv1D no causal y causal	Inconsistente. Fix: si forecasting, todo c
Audio en [-1, 1] con Conv1D sin normalizar	Funcionar funciona, pero converge mejor co

Preguntas frecuentes

¿Conv1D supera a Transformer en algún caso?

Sí, en audio raw y series largas (>1000 pasos) con patrones locales. Transformer atención $O(N^2)$ en memoria.

¿WaveNet sigue siendo state-of-the-art?

En generación de audio, fue reemplazada por modelos de difusión (clase 133) y autoregresivos sobre tokens de audio (EnCodec + Llama-style).

¿Combinar Conv1D + LSTM/Transformer?

Sí. Patrón "convs como tokenizer" + Transformer arriba es estándar en audio (Wav2Vec, Whisper).

¿dilation_rate y kernel_size cómo se combinan?

$\text{receptive_field} = (\text{kernel_size} - 1) * \text{dilation_rate} + 1$ para una sola capa. Stack: suma de eso por capa.

¿Por qué Conv1D paralelizable?

Todas las posiciones se computan en paralelo (no hay dependencia temporal explícita como RNN). Ideal para GPU.

Referencias

- Géron, cap. 15 — Using 1D Convolutional Layers + WaveNet.
- van den Oord et al. (2016), WaveNet.
- Bai et al. (2018), An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.

Siguiente clase

Clase 139 — Generación de texto char-RNN

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb