

# FITS: Modeling Time Series with 10k Parameters

<https://arxiv.org/abs/2307.03756>

## 0. Introduction

- 기존의 시계열 예측 모델은 수십만 개 이상의 파라미터를 사용하여, 작은 디바이스나 제한된 리소스 환경에서 사용이 어려움.
- 특히 Transformer 기반 모델은 계산량이 크고 복잡함.

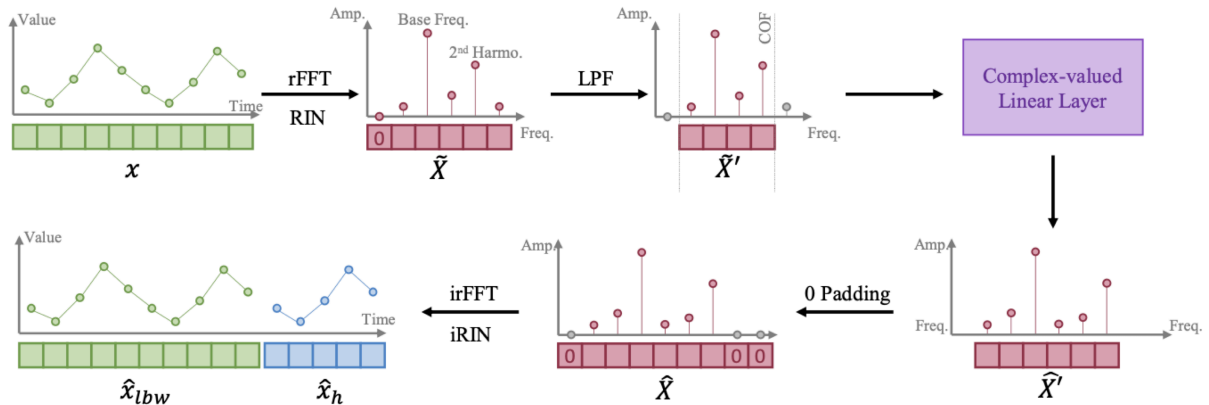
## 1. Overview

- FITS(Frequency Interpolation for Time Series)는 약 10,000개 미만의 파라미터로 동작하는 초경량 시계열 예측 모델.
- 기존의 시간 영역에서 직접 연산하는 방식과 달리, FITS는 rFFT(Real Fast Fourier Transform)를 활용해 주파수 영역에서 시계열을 모델링함으로써 연산 효율과 구조 단순화를 동시에 달성.

## 2. Challenges

- 기존 Transformer 기반 시계열 예측 모델은 수십만 개 이상의 파라미터를 갖고 있어, 학습과 추론 시 메모리와 연산량이 매우 큼
- 특히 self-attention 연산은 시계열 길이에 따라 연산량이  $O(L^2)$ 로 증가해, 긴 시계열을 다룰수록 비효율적
- 경량화 시 기존 성능이 급격히 저하되는 문제가 있어, 정확도와 모델 크기의 균형을 맞추는 것이 핵심 과제였음
- 비주기적 시계열이나 다양한 주기를 가지는 복합 시계열의 경우, 단순한 구조로는 반복성 탐지가 어렵고 오버피팅 가능성도 존재
- 실시간 환경이나 엣지 디바이스와 같이 리소스 제약이 있는 시스템에서는 기존 모델을 그대로 적용하기 어려움

### 3. Method



- 주파수 보간(Frequency Interpolation)
  - 과거 시계열 데이터를 rFFT를 통해 주파수 영역으로 변환한 뒤, 핵심 성분인 진폭(amplitude)과 위상(phase) 정보를 추출.
  - 이 주파수 정보를 보간(interpolation)하여 미래 구간의 주파수를 예측하고, 이를 역 FFT(irFFT)로 변환해 시계열을 복원.
- 저역 통과 필터링(Low-pass Filtering)
  - 예측에 불필요한 고주파 성분을 제거하고, 의미 있는 저주파 성분만을 남겨 데이터의 주요 패턴을 보존.
- 복소수 선형층(Complex Linear Layer)
  - 주파수 도메인에서의 진폭과 위상을 복소수 형태로 표현하고, 이를 학습 가능한 선형층으로 처리하여 더 정밀한 주파수 기반 예측을 가능하게 .

### 4. Experiments

- 사용된 데이터셋:
  - ETTh1 / ETTm1: 에너지 변동 시계열 (시간별/분별)
  - Weather: 기상 관측 데이터
  - Traffic: 도로 교통량 데이터
  - 이들은 장기 예측 성능을 평가하기 위해 자주 사용되는 공개 벤치마크임
- 비교 대상 모델:

- Informer, Autoformer, FEDformer, PatchTST 등 Transformer 계열 SOTA 모델
- FITS는 이들보다 모델 크기가 수백 배 작지만 유사하거나 더 나은 성능 달성
- 평가 지표:
  - 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE)
  - 각 시계열 길이에 따라 다양한 horizon에 대한 성능 비교 수행
- 실험 설정:
  - 입력 시계열 길이와 예측 horizon은 논문 내 실험 조건에 맞게 고정
  - 학습은 Adam 옵티마이저 기반으로 진행
  - 모델 파라미터 수는 약 10K로 제한하여 실험

## 5. Results

- 초소형 모델
  - 전체 파라미터 수가 약 10K에 불과하며, 이는 대부분의 시계열 예측 모델 대비 수백 배 가벼운 구조
- 경쟁력 있는 성능
  - ETTh1, ETTm1, Traffic 등 다양한 벤치마크에서 기존 Transformer 기반 모델과 동등하거나 더 우수한 예측 성능을 보임
- 엣지 환경 적합성
  - 가벼운 구조와 낮은 연산 복잡도로 인해, 모바일·IoT·엣지 디바이스와 같은 리소스 제약 환경에서도 효과적으로 동작할 수 있음

## 6. Insight

- 비주기적 시계열에 대한 제한
  - FITS는 주로 주기적 구조를 갖는 시계열 데이터에 최적화되어 있어, 비주기적이거나 불규칙한 패턴이 강한 시계열에서는 성능이 저하될 수 있음
- 입력 길이에 대한 민감성

- 모델의 성능은 입력 시계열의 길이 및 변환 파라미터에 민감하게 반응할 수 있으며, 이로 인해 일부 환경에서는 튜닝이 요구될 수 있음