

BasisFormer: Attention-based Time Series Forecasting with Learnable and Interpretable Basis

<https://arxiv.org/abs/2310.20496>

0. Overview

- BasisFormer는 시계열 예측을 위해 제안된 새로운 Transformer 계열 프레임워크임.
- 기존 Transformer 기반 예측 모델의 복잡성, 일반화 한계, 해석 불가능성을 개선하고자 등장함.
- 핵심 아이디어는 학습 가능한 basis를 통해 반복적/주기적 패턴을 추출하고, 원래 시계열 공간이 아닌 basis 공간에서 예측을 수행하는 것임.
- 예측 과정에서 양방향 크로스 어텐션을 통해 과거-미래 정보를 연결하고, 효율적이고 해석 가능한 방식으로 예측을 수행함.

1. Introduction

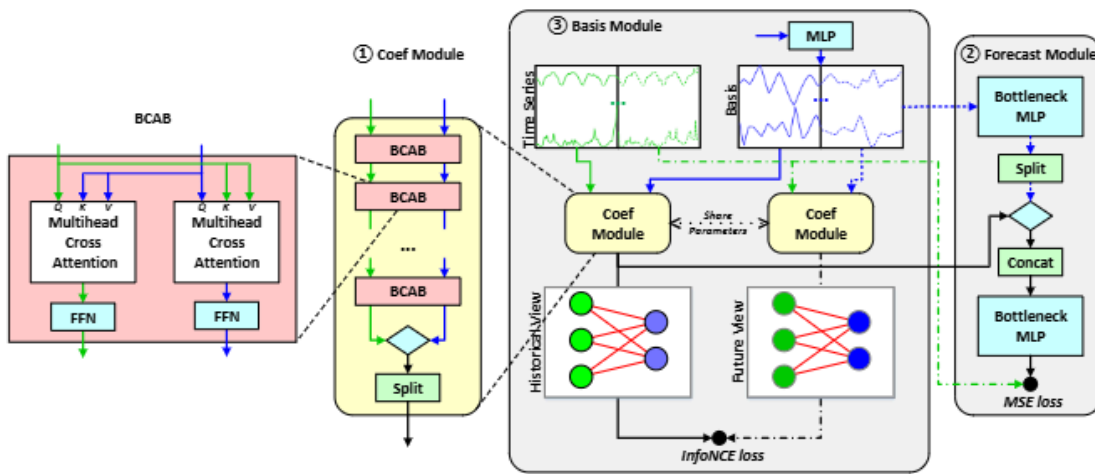
- 시계열 예측은 금융, 에너지, 교통, 헬스케어 등 다양한 분야에서 중요한 과제임.
- 기존 Transformer 모델들은 성능이 뛰어나지만 데이터 고차원성, 해석력 부족, 연산 효율성 문제로 한계가 존재함.
- BasisFormer는 시계열 데이터를 보다 압축적이고 의미 있는 basis 표현으로 변환하여 학습과 예측을 수행함으로써, 성능과 해석 가능성을 동시에 확보하는 것을 목표로 함.

2. Challenges

- high-dimensional forecasting → 시계열이 고차원일수록 모델 복잡도와 오버피팅 위험이 증가함

- limited interpretability → 기존 Transformer 예측 근거 파악이 어려워 신뢰성이 떨어짐
- inefficient temporal modeling → 반복적/계절적 패턴을 충분히 학습하지 못하는 경우 존재

3. Method



- basis learner: 데이터에서 반복적·주기적 패턴을 추출하여 학습 가능한 basis 집합 생성
- coefficient module: 과거 시계열과 basis 간 유사도를 Bidirectional Cross-Attention Block(BCAB)을 통해 계산, positional encoding으로 위치 정보 보정
- forecast module: coefficient와 basis를 조합해 미래를 basis 공간에서 예측한 뒤 시계열 공간으로 복원
- 학습 과정에서 self-supervised + contrastive learning(InfoNCE loss)을 활용해 label 없이 의미 있는 basis를 자동 학습

4. Experiments

- 논문에서는 다양한 도메인의 시계열 데이터셋을 활용하여 BasisFormer를 평가함:
 - 전력 부하 (Electricity Load Forecasting): 수천 개의 고객 전력 소비 시계열을 활용해 단기/중기 예측 성능 평가

- 교통 데이터 (Traffic Forecasting): 도시 도로 네트워크의 교통량 데이터셋으로 주기성과 변동성 평가
- 기상 데이터 (Weather / ETT): 온도, 압력 등 기상 시계열 데이터를 통해 계절성과 외부 변동성 처리 능력 검증
- 금융 시계열 (Exchange-Rate, Stock-like): 환율·시장 데이터에서 노이즈가 많은 시계열 예측 성능 확인
- 모든 실험은 MSE / MAE 지표를 사용해 평가되며, Fourier / Sine-Cosine 기반 고정 basis 모델 및 Transformer 계열 최신 모델과 비교 수행

5. Results

Models		Fedformer		Autoformer		N-HiTS		Film		Dlinear		TCN		Basisformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT	96	0.203	0.287	0.255	0.339	0.192	0.265	0.183	0.266	0.193	0.292	3.041	1.330	0.184	0.266
	192	0.269	0.328	0.281	0.340	0.287	0.329	0.247	0.305	0.284	0.362	3.072	1.339	0.248	0.307
	336	0.325	0.366	0.339	0.372	0.389	0.389	0.309	0.343	0.369	0.554	3.105	1.348	0.321	0.355
	720	0.421	0.415	0.422	0.419	0.591	0.491	0.407	0.399	0.554	0.522	3.135	1.354	0.410	0.404
electricity	96	0.193	0.308	0.201	0.317	1.748	1.020	0.199	0.276	0.199	0.284	0.985	0.813	0.165	0.259
	192	0.201	0.315	0.222	0.334	1.743	1.018	0.198	0.279	0.198	0.287	0.996	0.821	0.178	0.272
	336	0.214	0.329	0.231	0.338	-	-	0.217	0.301	0.210	0.302	1.000	0.824	0.189	0.282
	720	0.246	0.355	0.254	0.361	-	-	0.280	0.358	0.245	0.335	1.438	0.784	0.223	0.311
exchange	96	0.148	0.278	0.197	0.323	1.685	1.049	0.083	0.201	0.088	0.218	3.004	1.432	0.085	0.205
	192	0.271	0.380	0.300	0.369	1.658	1.023	0.179	0.300	0.176	0.315	3.048	1.444	0.177	0.299
	336	0.460	0.500	0.509	0.524	1.566	0.988	0.337	0.416	0.313	0.427	3.113	1.459	0.336	0.421
	720	1.195	0.841	1.447	0.941	1.809	1.055	0.642	0.610	0.839	0.695	3.150	1.458	0.854	0.670
traffic	96	0.587	0.366	0.613	0.388	2.138	1.026	0.652	0.395	0.650	0.396	1.438	0.784	0.444	0.315
	192	0.604	0.373	0.616	0.382	-	-	0.605	0.371	0.605	0.378	1.463	0.794	0.460	0.316
	336	0.621	0.383	0.622	0.387	-	-	0.615	0.372	0.612	0.382	1.479	0.799	0.471	0.317
	720	0.626	0.382	0.660	0.408	-	-	0.692	0.428	0.645	0.394	1.499	0.804	0.486	0.318
weather	96	0.217	0.296	0.266	0.336	0.648	0.492	0.193	0.234	0.196	0.255	0.615	0.589	0.173	0.214
	192	0.276	0.336	0.307	0.367	0.616	0.479	0.238	0.270	0.237	0.296	0.629	0.600	0.223	0.257
	336	0.339	0.380	0.359	0.395	0.579	0.462	0.288	0.304	0.283	0.335	0.639	0.608	0.278	0.298
	720	0.403	0.428	0.419	0.428	0.541	0.447	0.358	0.350	0.343	0.383	0.639	0.610	0.355	0.347
illness	24	3.228	1.260	3.486	1.287	3.297	1.679	2.198	0.911	2.398	1.040	6.624	1.830	1.550	0.814
	36	2.679	1.080	3.103	1.148	2.379	1.441	2.267	0.926	2.646	1.088	6.858	1.879	1.516	0.819
	48	2.622	1.078	2.669	1.085	3.341	1.751	2.348	0.989	2.614	1.086	6.968	1.892	1.877	0.907
	60	2.857	1.157	2.770	1.125	2.278	1.493	2.508	1.038	2.804	1.146	7.127	1.918	1.878	0.902

Experiment with '-' means it reported an out-of-memory error on a computer with 80G memory.

- 학습 가능한 basis는 고정된 Fourier/Sine 기저보다 5~10% 성능 향상을 보임
- 다중 헤드 실험에서 16개 헤드가 최적 성능을 기록함
- InfoNCE 대조 학습 적용 시 약 7.2% 성능 향상 달성
- basis 개수는 너무 많으면 중복 학습으로 성능 저하 발생

6. Insight

- basis 공간으로의 차원 축소는 직관적이며 계산 효율성과 정확도를 동시에 높임
- 해석 가능한 basis를 통해 Transformer의 블랙박스 문제 일부 해소
- self-supervised 학습으로 레이블 부족 문제에 효과적 대응
- 그러나 basis의 해석 가능성은 여전히 주관적이며, 특정 basis가 의미하는 패턴을 정량적으로 설명하는 체계 부족
- 양방향 cross-attention 등 복잡한 구조로 인해 산업 적용 시 효율성/구현 난이도가 이슈가 될 수 있음
- SOTA 대비 모든 데이터셋에서 압도적 우위는 아니며, 특정 유형 시계열에서만 큰 개선이 나타날 가능성 존재
- 외부 요인(날씨, 이벤트, 정책 등)을 반영하지 않아 실제 도메인에서는 한계가 있을 수 있음