

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

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

1. Overview

BasisFormer는 시계열 예측용으로 제안된 새로운 프레임워크. 기존 Transformer 기반 모델들이 가진 복잡성, 일반화 부족 문제 해결하고자 등장함.

핵심은 학습 가능한 basis를 사용해서 시계열의 주요 패턴을 압축해 표현함.

예측 시에는 양방향 크로스 어텐션으로 과거와 미래 정보 연결함.

고차원에서 직접 예측하지 않고, 더 공간 효율적이고 해석 가능한 basis 공간에서 예측 수행.

이를 통해 예측 성능 향상 + 모델 해석력 강화.

2. Challenges

high-dimensional forecasting: 원래 시계열 데이터가 고차원이라 모델이 복잡해지고 오버피팅 위험 높음

limited interpretability: Transformer 계열 모델은 성능은 좋지만 예측의 근거를 파악하기 어려워 해석성 떨어짐

inefficient modeling of temporal patterns: 반복적이거나 계절적인 시계열 특성을 잘 포착하지 못하는 경우 많음

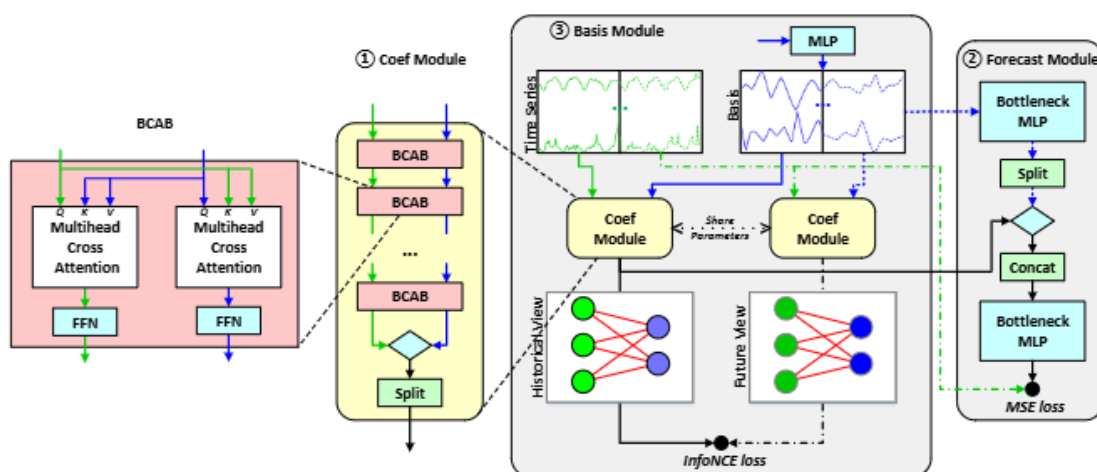
3. Method

basis 학습: 데이터에서 반복적인 패턴을 담은 basis를 자동으로 학습해 시계열을 압축된 형태로 표현함

디코딩 예측: 미래 예측을 원래 시계열 공간이 아니라 basis space에서 수행해 효율성과 성능을 높임

양방향 어텐션: 과거 시계열, basis, 미래 basis 간 관계를 양방향 cross-attention으로 계산해 정보 손실 최소화함

BasisFormer는 시계열 예측을 위한 새로운 구조로, 다음 세 가지 주요 구성 요소로 구성됨



1. coefficient module (coef module)

$$\text{Attn}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

과거 시계열과 학습된 basis 간 유사성 계수 계산함. 이때 Bidirectional Cross-Attention Block(BCAB) 사용해 시간 정보 고려한 정밀한 매칭 수행. 시계열의 장기 의존성 효과적으로 학습 가능하며, Learnable Positional Encoding 통해 위치 정보 일반화 문제 해결함. 결과적으로 위치 정보와 basis 간 유사도 정교하게 반영 가능하며, 계산된 계수는 예측 시 어떤 basis를 얼마나 반영할지 결정하는 핵심 요소로 작용함.

2. basis learner

입력 시계열 기반으로 학습 가능한 basis 집합 생성함. 이 basis는 주기적·반복적 패턴을 포착하고, 모델이 반복적으로 활용 가능한 기초 시퀀스 역할을 수행함. 자기 지도 학습 방식으로 지속적으로 업데이트되며, 데이터 특성에 맞게 최적화됨. 예측은 Bottleneck MLP 구조(입력 → 저차원 병목층 → 다시 확장 → 출력)를 통해 중요한 특징만 필터링한 뒤, Coef 모듈에서 얻은 유사도를 활용해 수행됨. 이로 인해 기존 모델 대비 더 높은 예측 성능과 해석 가능성 확보 가능함.

3. forecast module

$$L = -\log \frac{\exp(\text{sim}(z, z^+)/\tau)}{\sum_{i=0}^N \exp(\text{sim}(z, z_i)/\tau)}$$

예측을 위해 basis와 coefficient를 조합해 basis space에서 미래 값을 구성하고, 이를 시계열 공간으로 복원해 최종 예측값 생성함. 이 방식은 계산 효율성과 해석 가능성을 동시에 확보함. 기저 학습은 Self-Supervised (입력(X)만 필요, 입력데이터에서 라벨 생성 후 학습)+ Contrastive Learning(InfoNCE)(Softmax 기반 확률 모델링) 기반으로 진행됨. Anchor-Positive는 가깝게, Anchor-Negative는 멀게 하여 시계열 내 유의미한 기저를 데이터 기반으로 자동 학습함. 사전 정의된 기저가 아닌, 데이터에 최적화된 기저를 생성함.

4. Experiments

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.

(MSE 기준 작을수록 좋음)

1. 학습 가능한 기저 vs 고정된 기저 → 기존 Fourier / Sine-Cosine 대비 5~10% 성능 향상

2. 다중 헤드 개수 분석 → 16개 헤드에서 최고 성능 달성
3. 대조 학습(Contrastive Learning) 효과 → InfoNCE Loss 적용 시 7.2% 성능 향상
4. 기저 개수 영향 → 너무 많으면 중복된 기저가 학습됨 → 성능 저하 가능

6. Insight

1. basis로의 차원 축소는 매우 직관적이면서도 효과적임

시계열 데이터는 계절성, 주기성 등 반복적인 패턴이 많은데, 이런 특성을 basis로 추출해서 예측을 시계열 원공간이 아닌 basis space에서 수행하는 방식은 예측 정확도와 계산 효율을 동시에 잡은 접근임.

2. 기계 학습에 해석 가능성을 도입하려는 시도

Transformer 계열 모델이 보통 블랙박스처럼 작동하는 반면, BasisFormer는 해석 가능한 basis를 사용해서 모델 작동 원리에 대해 일정 부분 설명 가능하게 만듦. 실제 산업적 활용이나 의사결정에서 신뢰성을 높이는 요소로 작용할 수 있음.

3. self-supervised learning의 실용적인 적용 사례

레이블 없이도 basis를 학습할 수 있게 한 자기 지도 학습이 잘 통합되어 있음. 이로 인해 label scarcity 문제를 겪는 시계열 예측 분야에서 유용한 솔루션이 될 수 있음.

4. basis의 해석 가능성은 여전히 주관적임

논문에서는 “해석 가능한 basis”를 제시하지만, 실제로는 해당 basis가 무엇을 의미하는지 정량적으로 설명하는 체계가 부족함. 시각화를 통해 해석을 시도하긴 했지만, 특정 basis가 예를 들어 ‘평일 트렌드’, ‘계절 효과’ 등을 포착하는지를 명확히 판별할 수 있는 라벨링 도구나 기준은 부재함.

→ 특정 basis가 어떤 패턴(예: 평일/주말, 계절성 등)을 포착하고 있는지 어떻게 확인할 수 있는가?

5. 복잡한 구조 대비 구현 난이도 및 효율 이슈 존재

양방향 cross-attention, basis space 활용 등으로 인해 구조가 복잡해짐. 실제 산업 환경에서 적용하려면 구현 난이도와 서버 비용이 높을 수 있음. baseline 모델 대비 효율성 우위가 항상 보장되는 건 아님.

→ 논문에서는 training 속도는 기존 Transformer보다 빠르다고 주장함. 하지만 이 역시 계산 비용에 대한 더 정량적이고 상세한 비교 실험이 필요함.

6. 기존 SOTA 모델 대비 명확한 우위가 모든 데이터셋에서 보이진 않음

논문에서는 성능 향상을 보고하고 있지만, 일부 데이터셋에서는 개선 폭이 작거나, 특정 지표에서만 개선된 사례도 있음. 전반적인 generalization 성능에 대해선 좀 더 다양한 도메인에서의 검증이 필요함.

→ 실질적으로 전천후 모델인지, 아니면 특정 유형의 시계열에만 효과적인 모델인지는 추가 분석 필요.

7. 외부 요인을 고려하지 않는 한계

BasisFormer는 시계열 내부의 패턴 학습에 초점을 맞추고 있어, 날씨, 이벤트, 휴일 등 외부 변수는 반영하지 않음. 이는 설계상의 의도일 수 있으나, 다양한 소스를 사용하는 실제 시계열 예측 작업에서는 적용에 제한이 생길 수 있음.

→ 멀티모달 시계열 예측이나 외부 요인이 중요한 도메인(예: 에너지 수요, 주식시장 등)에서는 한계로 작용할 수 있음. ↔ 반대로 도메인 특화 모델 발전 가능성은 있음