

2025 11 13
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

- Crossformer
 - 논문 리뷰
 - 실험 세팅 및 결과
 - 결과 정리

CROSSFORMER: TRANSFORMER UTILIZING CROSS-DIMENSION DEPENDENCY FOR MULTIVARIATE TIME SERIES FORECASTING

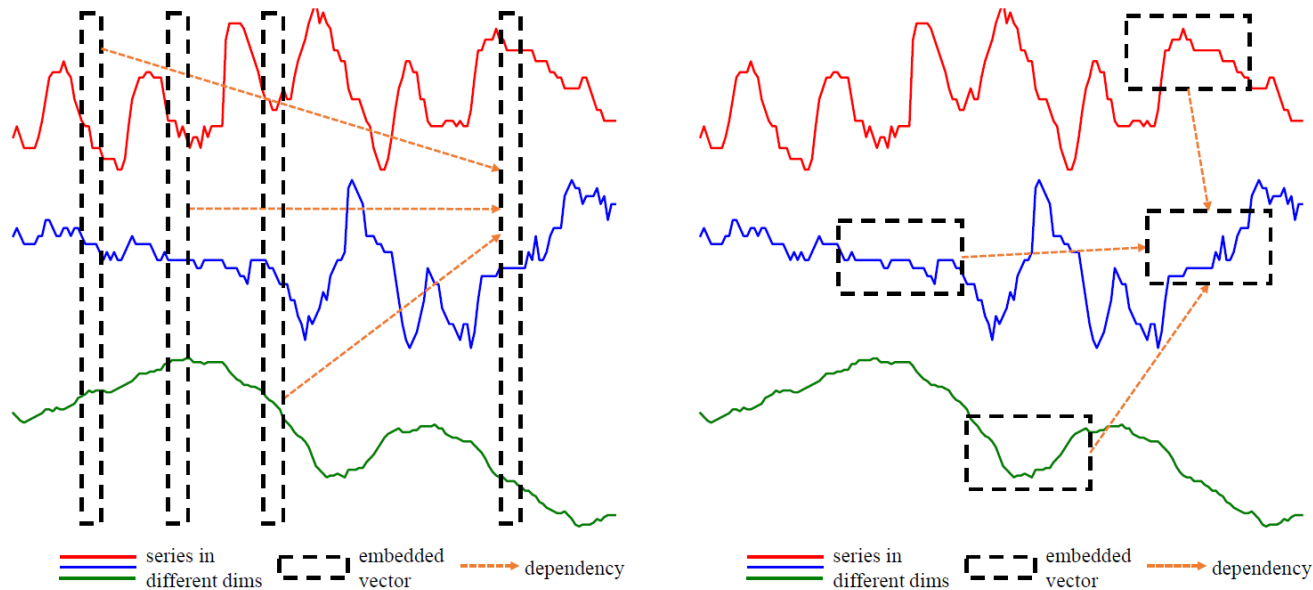
Yunhao Zhang & Junchi Yan*

MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University and Shanghai AI Lab
{zhangyunhao, yanjunchi}@sjtu.edu.cn

- arXiv 등록일: 2023-01-23
- 인용 수: 1559회(Google Scholar, 2025-11-07)
- Time Series Library 등록
- Published at ICLR 2023
- 링크: <https://openreview.net/pdf?id=vSVLM2j9eie>

Crossformer

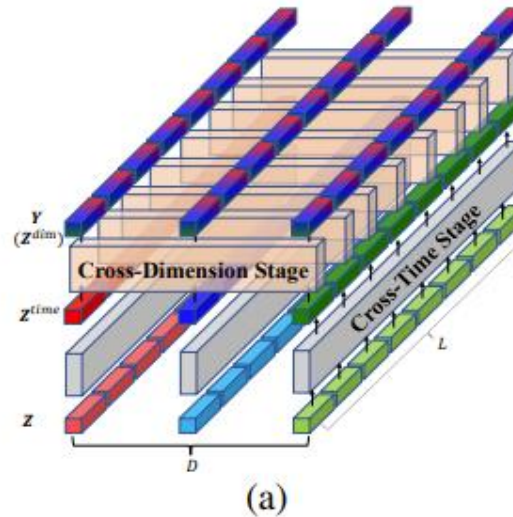
데이터 임베딩 및 처리 방식 비교 (Transformer vs Crossformer)



- Transformer
 - 동일 시점의 모든 변수를 하나의 벡터로 묶어 임베딩
 - 시간 순서에 따른 시점 간 패턴은 학습하지만, 변수 간 상호 관계는 직접적으로 고려하지 않음
- Crossformer
 - Transformer 기반 모델과는 달리, 각 변수를 개별적으로 Segment 단위로 나누어 임베딩
 - 시간 축과 변수 축을 분리하여, 시간 축 내 패턴과 변수 간 상호관계를 동시에 학습 가능

Crossformer

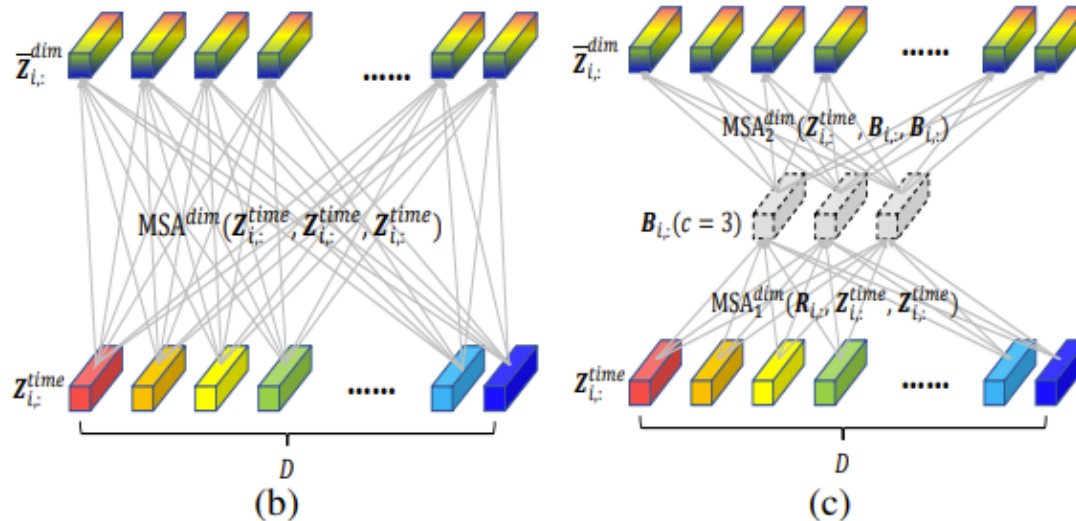
Two-Stage Attention 구조



- DSW 임베딩 (전처리)
 - 입력 시계열을 2D 형태인 변수 \times 시간으로 구성
 - 각 변수를 개별적으로 segment 단위로 나누어 임베딩
- TSA Cross-Dimension Stage
 - 변수 간 상호 의존성을 학습
 - 시간 축에서 얻은 정보를 통해 변수 간 의존성을 학습
- TSA Cross-Time Stage
 - 시간 축 내 패턴을 학습
 - 동일 변수 내에서 시간 순서 정보를 포착

Crossformer

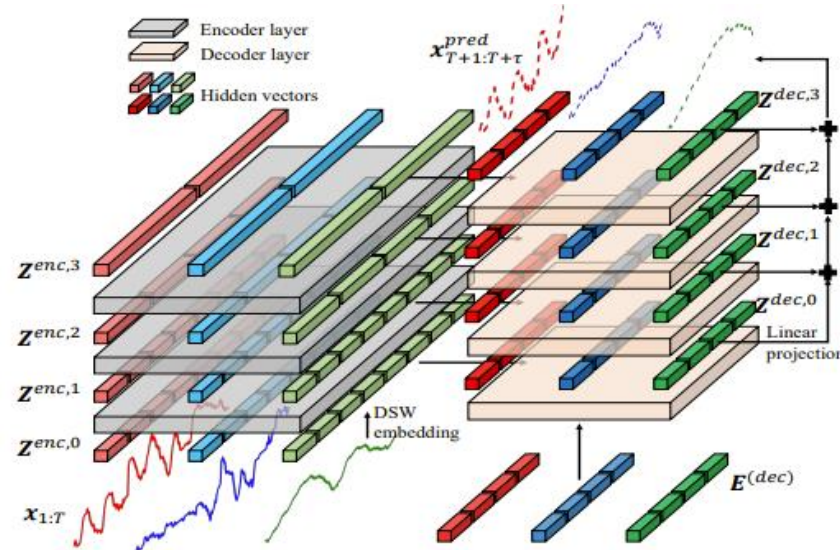
Two-Stage Attention 상세 설명



- Cross-Dimension Stage
 - D개의 임베딩 벡터들이 한꺼번에 입력되고, 모든 변수 벡터들이 서로에게 Multi-Head Self-Attention 수행
 - 이후 각 벡터는 다른 모든 변수와의 관계 정보가 반영된 새로운 벡터가 됨
 - 라우터 벡터를 통해 계산량 문제 해결 및 변수 간의 상호 의존성 학습
- Cross-Time Stage
 - 각 변수별로 독립적으로 Multi-Head Self-Attention을 적용
 - 동일 변수 내의 서로 다른 segment 간의 시간적 의존성 학습

Crossformer

Hierarchical Encoder-Decoder



- Crossformer는 여러 개의 Encoder-Decoder를 쌓은 계층적 구조를 가짐
- Encoder
 - Cross-Dimension Stage와 Cross-Time Stage로 구성됨
 - 층이 올라갈 때마다 segment 병합을 통해 인접 segment를 합쳐 더 큰 스케일의 정보 학습
- Decoder
 - 기존 Transformer의 Decoder와 유사하며, Self-Attention과 Cross-Attention 수행
 - 인코더의 각 층에 1대1로 대응, 모든 층에서 개별 예측 생성
 - 이후 모든 층의 예측을 합산해 최종 예측값 도출

Crossformer

주요 모델 성능 비교

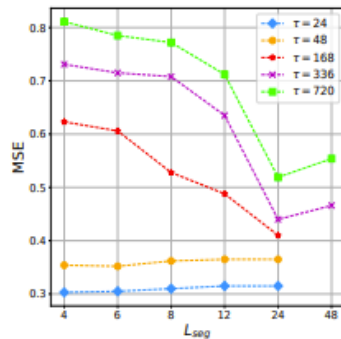
Models		LSTMa		LSTnet		MTGNN		Transformer		Informer		Autoformer		Pyraformer		FEDformer		Crossformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	0.650	0.624	1.293	0.901	0.336	0.393	0.620	0.577	0.577	0.549	0.439	0.440	0.493	0.507	<u>0.318</u>	<u>0.384</u>	0.305	0.367
	48	0.720	0.675	1.456	0.960	0.386	0.429	0.692	0.671	0.685	0.625	0.429	0.442	0.554	0.544	0.342	<u>0.396</u>	<u>0.352</u>	0.394
	168	1.212	0.867	1.997	1.214	0.466	0.474	0.947	0.797	0.931	0.752	0.493	0.479	0.781	0.675	<u>0.412</u>	<u>0.449</u>	0.410	0.441
	336	1.424	0.994	2.655	1.369	0.736	0.643	1.094	0.813	1.128	0.873	0.509	0.492	0.912	0.747	<u>0.456</u>	<u>0.474</u>	0.440	0.461
	720	1.960	1.322	2.143	1.380	0.916	0.750	1.241	0.917	1.215	0.896	0.539	0.537	0.993	0.792	<u>0.521</u>	0.515	0.519	<u>0.524</u>
ETTm1	24	0.621	0.629	1.968	1.170	<u>0.260</u>	<u>0.324</u>	0.306	0.371	0.323	0.369	0.410	0.428	0.310	0.371	0.290	0.364	0.211	0.293
	48	1.392	0.939	1.999	1.215	0.386	0.408	0.465	0.470	0.494	0.503	0.485	0.464	0.465	0.464	<u>0.342</u>	<u>0.396</u>	0.300	0.352
	96	1.339	0.913	2.762	1.542	0.428	0.446	0.681	0.612	0.678	0.614	0.502	0.476	0.520	0.504	<u>0.366</u>	<u>0.412</u>	0.320	0.373
	288	1.740	1.124	1.257	2.076	0.469	0.488	1.162	0.879	1.056	0.786	0.604	0.522	0.729	0.657	0.398	<u>0.433</u>	<u>0.404</u>	0.427
	672	2.736	1.555	1.917	2.941	0.620	0.571	1.231	1.103	1.192	0.926	0.607	0.530	0.980	0.678	0.455	0.464	<u>0.569</u>	<u>0.528</u>
WTH	24	0.546	0.570	0.615	0.545	0.307	<u>0.356</u>	0.349	0.397	0.335	0.381	0.363	0.396	<u>0.301</u>	0.359	0.357	0.412	0.294	0.343
	48	0.829	0.677	0.660	0.589	0.388	0.422	0.386	0.433	0.395	0.459	0.456	0.462	<u>0.376</u>	<u>0.421</u>	0.428	0.458	0.370	0.411
	168	1.038	0.835	0.748	0.647	<u>0.498</u>	<u>0.512</u>	0.613	0.582	0.608	0.567	0.574	0.548	0.519	0.521	0.564	0.541	0.473	0.494
	336	1.657	1.059	0.782	0.683	<u>0.506</u>	<u>0.523</u>	0.707	0.634	0.702	0.620	0.600	0.571	0.539	0.543	0.533	0.536	0.495	0.515
	720	1.536	1.109	0.851	0.757	0.510	0.527	0.834	0.741	0.831	0.731	0.587	0.570	0.547	0.553	0.562	0.557	<u>0.526</u>	<u>0.542</u>
ECL	48	0.486	0.572	0.369	0.445	<u>0.173</u>	<u>0.280</u>	0.334	0.399	0.344	0.393	0.241	0.351	0.478	0.471	0.229	0.338	0.156	0.255
	168	0.574	0.602	0.394	0.476	<u>0.236</u>	<u>0.320</u>	0.353	0.420	0.368	0.424	0.299	0.387	0.452	0.455	0.263	0.361	0.231	0.309
	336	0.886	0.795	0.419	0.477	0.328	<u>0.373</u>	0.381	0.439	0.381	0.431	0.375	0.428	0.463	0.456	0.305	0.386	<u>0.323</u>	0.369
	720	1.676	1.095	0.556	0.565	0.422	0.410	0.391	0.438	0.406	0.443	<u>0.377</u>	0.434	0.480	0.461	0.372	0.434	0.404	<u>0.423</u>
	960	1.591	1.128	0.605	0.599	0.471	0.451	0.492	0.550	0.460	0.548	0.366	0.426	0.550	0.489	<u>0.393</u>	0.449	0.433	<u>0.438</u>
ILI	24	4.220	1.335	4.975	1.660	4.265	1.387	3.954	1.323	4.588	1.462	3.101	1.238	3.970	1.338	2.687	1.147	<u>3.041</u>	<u>1.186</u>
	36	4.771	1.427	5.322	1.659	4.777	1.496	4.167	1.360	4.845	1.496	<u>3.397</u>	1.270	4.377	1.410	2.887	1.160	3.406	<u>1.232</u>
	48	4.945	1.462	5.425	1.632	5.333	1.592	4.746	1.463	4.865	1.516	<u>2.947</u>	<u>1.203</u>	4.811	1.503	2.797	1.155	3.459	1.221
	60	5.176	1.504	5.477	1.675	5.070	1.552	5.219	1.553	5.212	1.576	<u>3.019</u>	<u>1.202</u>	5.204	1.588	2.809	1.163	3.640	1.305
Traffic	24	0.668	0.378	0.648	0.403	<u>0.506</u>	<u>0.278</u>	0.597	0.332	0.608	0.334	0.550	0.363	0.606	0.338	0.562	0.375	0.491	0.274
	48	0.709	0.400	0.709	0.425	0.512	<u>0.298</u>	0.658	0.369	0.644	0.359	0.595	0.376	0.619	0.346	0.567	0.374	<u>0.519</u>	0.295
	168	0.900	0.523	0.713	0.435	<u>0.521</u>	<u>0.319</u>	0.664	0.363	0.660	0.391	0.649	0.407	0.635	0.347	0.607	0.385	0.513	0.289
	336	1.067	0.599	0.741	0.451	<u>0.540</u>	<u>0.335</u>	0.654	0.358	0.747	0.405	0.624	0.388	0.641	0.347	0.624	0.389	0.530	0.300
	720	1.461	0.787	0.768	0.474	0.557	<u>0.343</u>	0.685	0.370	0.792	0.430	0.674	0.417	0.670	0.364	0.623	0.378	<u>0.573</u>	0.313

- 기존 Transformer 기반 모델에 비해 시간 축과 변수 축을 분리하여 학습하는 구조의 Crossformer가 대부분의 예측 길이에서 좋은 성능을 보여줌

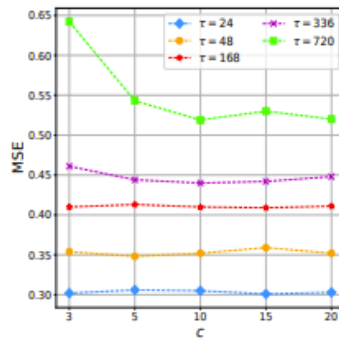
Crossformer

구성 요소별 성능 분석

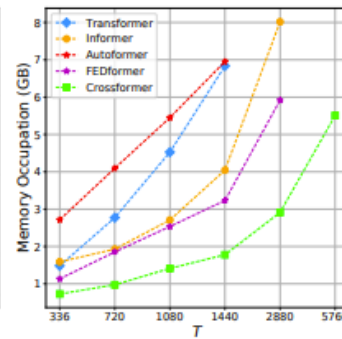
Models	Transformer		DSW		DSW+TSA		DSW+HED		DSW+TSA+HED	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
24	0.620	0.577	0.373	0.418	0.322	0.373	0.406	0.454	0.305	0.367
48	0.692	0.671	0.456	0.479	0.365	0.403	0.493	0.512	0.352	0.394
168	0.947	0.797	0.947	0.731	0.473	0.479	0.614	0.583	0.410	0.441
336	1.094	0.813	0.969	0.752	0.553	0.534	0.788	0.676	0.440	0.461
720	1.241	0.971	1.086	0.814	0.636	0.599	0.841	0.717	0.519	0.524



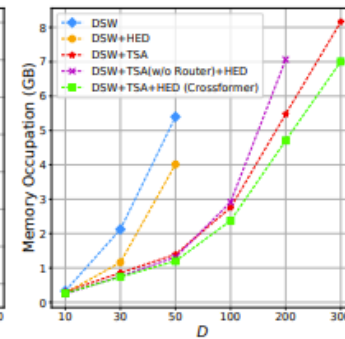
(a)



(b)



(c)



(d)

- Crossformer의 핵심 구조인 Two-Stage Attention을 추가했을 때가 성능 향상 폭이 가장 크며, 여기에 계층 구조(HED)를 추가했을 때의 성능이 모든 예측 길이에서 가장 좋은 성능을 보여줌
- Segment 길이 변화에도 안정적인 성능을 유지하며, 메모리 사용량이 낮아 입력 길이나 변수 개수가 증가하더라도 효율성이 뛰어남

실험 세팅

- 사용한 모델: Crossformer
- 재현 실험 데이터셋: ETTh1

Experiment	ETTh1
Learning rate	10^{-4}
Epoch	20
Batch size	32
Loss function	MSE Loss
Segment Length	6/24
input_feature	7
Pred_len	24/48/168/336/720
d_model	256
d_ff	512

Crossformer 재현 실험 (ETTh1)

- 단기 예측에서는 논문과 거의 동일한 성능을 보였고, 장기 예측으로 갈수록 약간의 오차가 발생하였지만, 전반적으로 논문과 비슷한 성능을 보였음

	ETTh1 Paper		ETTh1 Reproduction	
Pred len	MSE	MAE	MSE	MAE
24	0.305	0.367	0.302	0.359
48	0.352	0.394	0.346	0.387
168	0.410	0.441	0.406	0.441
336	0.440	0.461	0.449	0.471
720	0.519	0.524	0.556	0.542

실험 결과 정리

- 재현 실험 결과, 오차가 조금 있긴 하지만 논문과 거의 동일한 성능이 나온 것을 확인