DSAIL Data Science & Artificial Intelligence

KAIST

LLM기반 멀티모달 시계열 분석

박찬영

Associate Professor, KAIST

Industrial and Systems Engineering
Graduate School of Data Science
Graduate School of AI

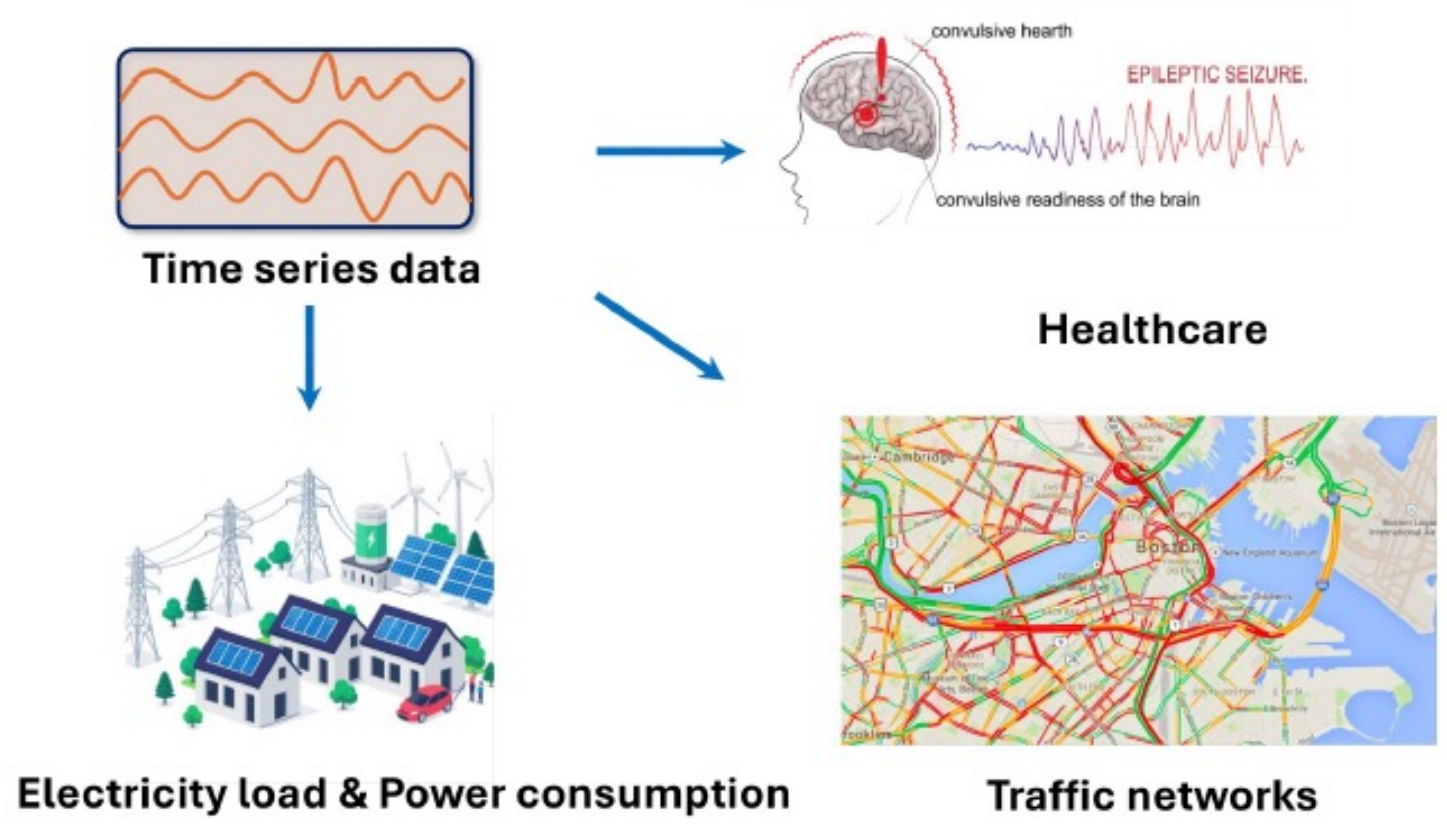
cy.park@kaist.ac.kr

Outline: Multi-modal Time Series Methods

- Background
- Multi-modal Fusion
  - Intermediate-level
  - Output-level
- Multi-modal Alignment
  - Intermediate-level: Self-attention / Cross-attention / Graph Neural Network / Contrastive Learning
  - Output-level: Retrieval / LLM reasoning

Background

- **Time Series**: 시간 순서에 따라 인덱싱된 연속적인 데이터 포인트들
  - 예: 전력 수요(Electricity Load), 뇌파(EEG), 교통량(Traffic volume)

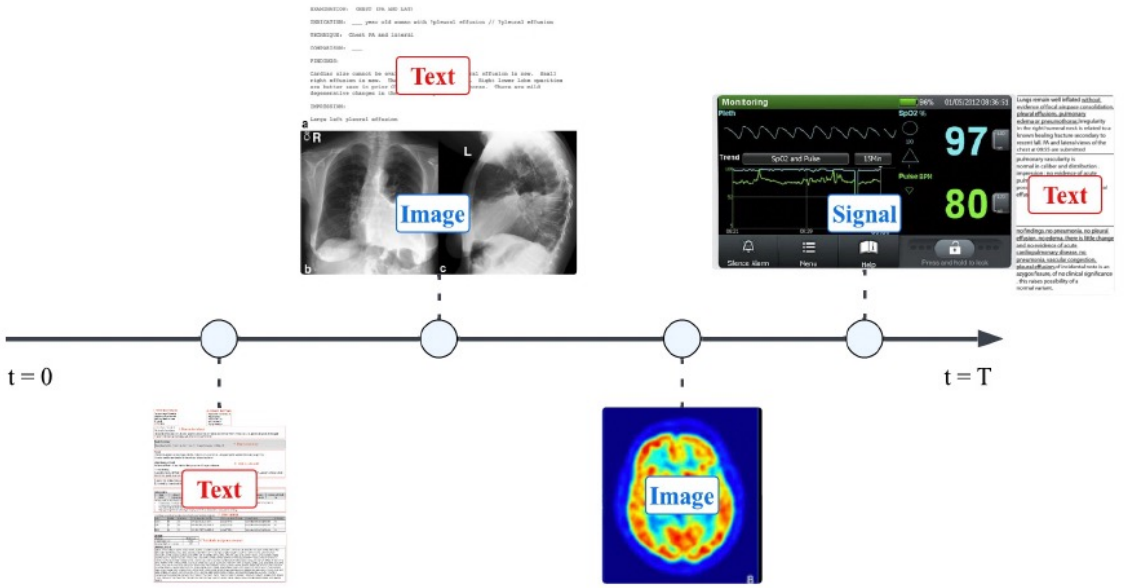


Multi-modal Time Series Analysis

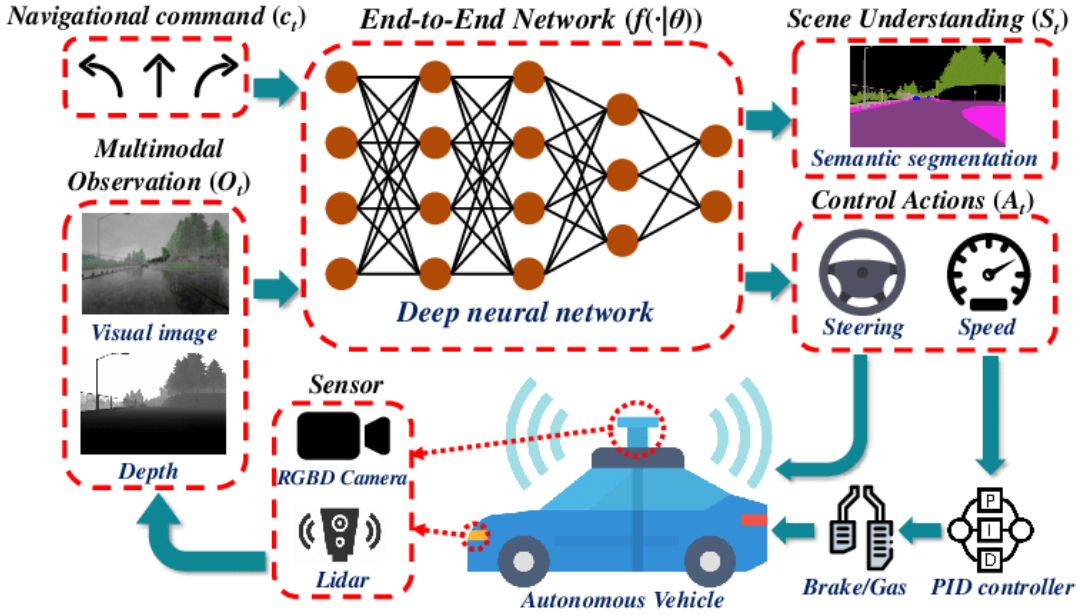
- Multi-modal: 이미지, 텍스트, 오디오, 센서 등 **여러 데이터 소스/모달리티를 함께 활용하는 것**
- Multi-modal Time Series
  - 시계열 데이터에 **외부 맥락(context) 정보가 결합된 형태**
  - 시간적 변화(temporal dynamics)와 외부 의미 정보(semantic context)를 동시에 반영
  - 보다 풍부한 표현 학습 가능
- Why is Multi-modality Important?
  - 현실 세계 시스템은 본질적으로 heterogeneous
  - 시계열 신호만으로는 상태를 완전히 설명하기 어려움
  - 다양한 모달 정보를 결합하면 **더 정확하고 안정적인 예측 가능**

Examples

- **의료 분야:** 생체 신호(시계열) + 임상 기록(텍스트)
  → 사망 예측 정확도 향상

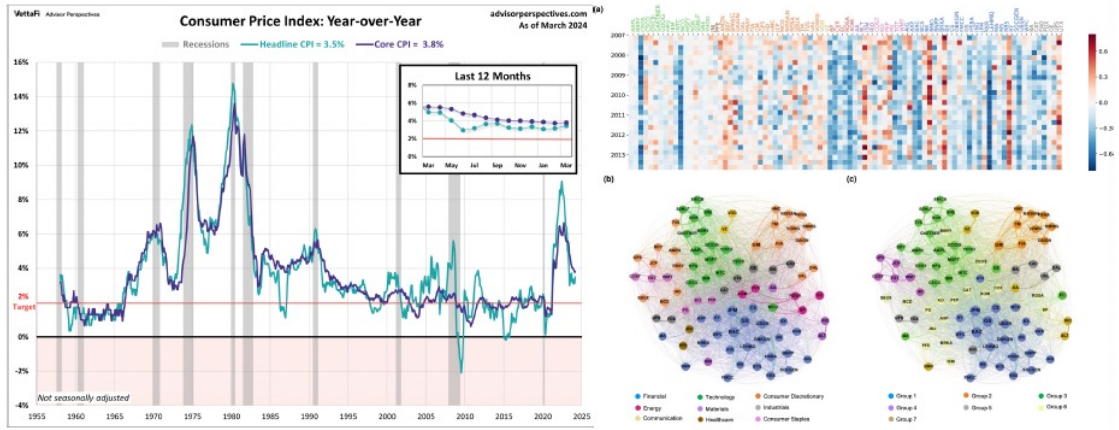


- **자율주행:** 차량 주행 궤적(시계열) + 카메라/LiDAR(영상)
  → 더 안전한 의사결정

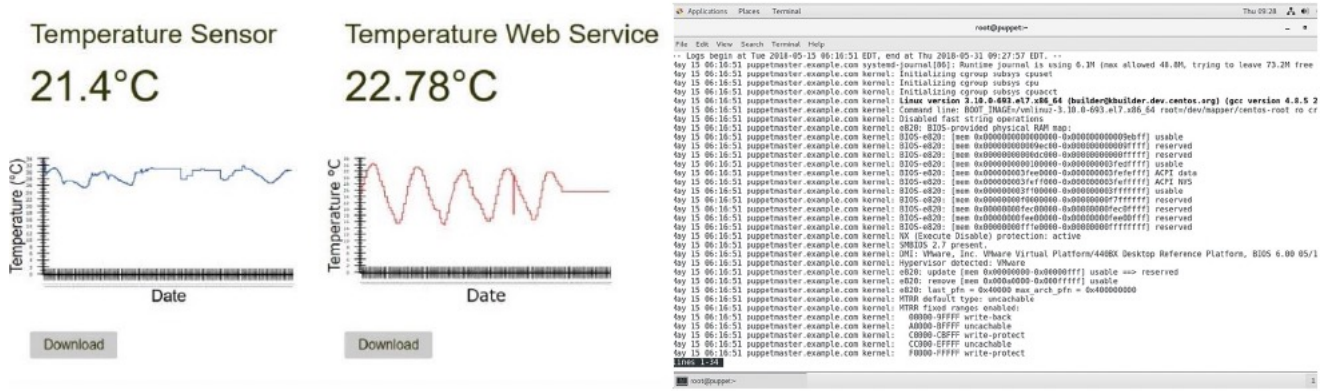


Examples

- **금융 분야**: 주가 데이터(시계열) + 뉴스/소셜 감성(텍스트)
  → 시장 급변 조기 감지



- IoT Systems (**제조 산업**): 센서 신호(시계열) + 정비 이력(텍스트)
  → 고장 진단 및 원인 분석 정확도 향상

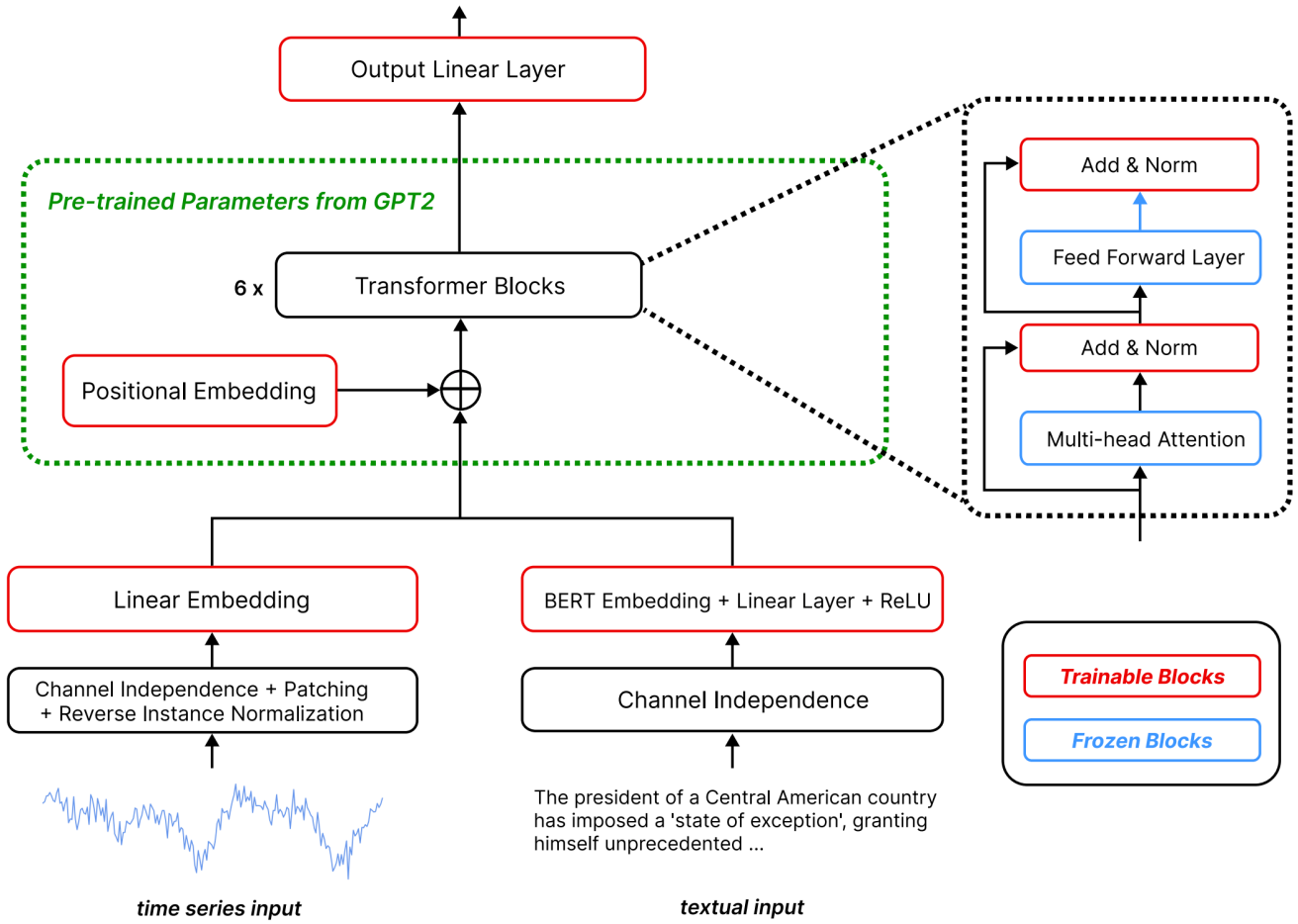


Outline: Multi-modal Time Series Methods

- Background
- Multi-modal Fusion
  - Intermediate-level
  - Output-level
- Multi-modal Alignment
  - Intermediate-level: Self-attention / Cross-attention / Graph Neural Network / Contrastive Learning
  - Output-level: Retrieval / LLM reasoning

(Fusion – Intermediate-level) GPT4MTS (AAAI 2024)

- Idea: LLM을 재학습하지 않고, prompt tuning + lightweight adapter로 해결

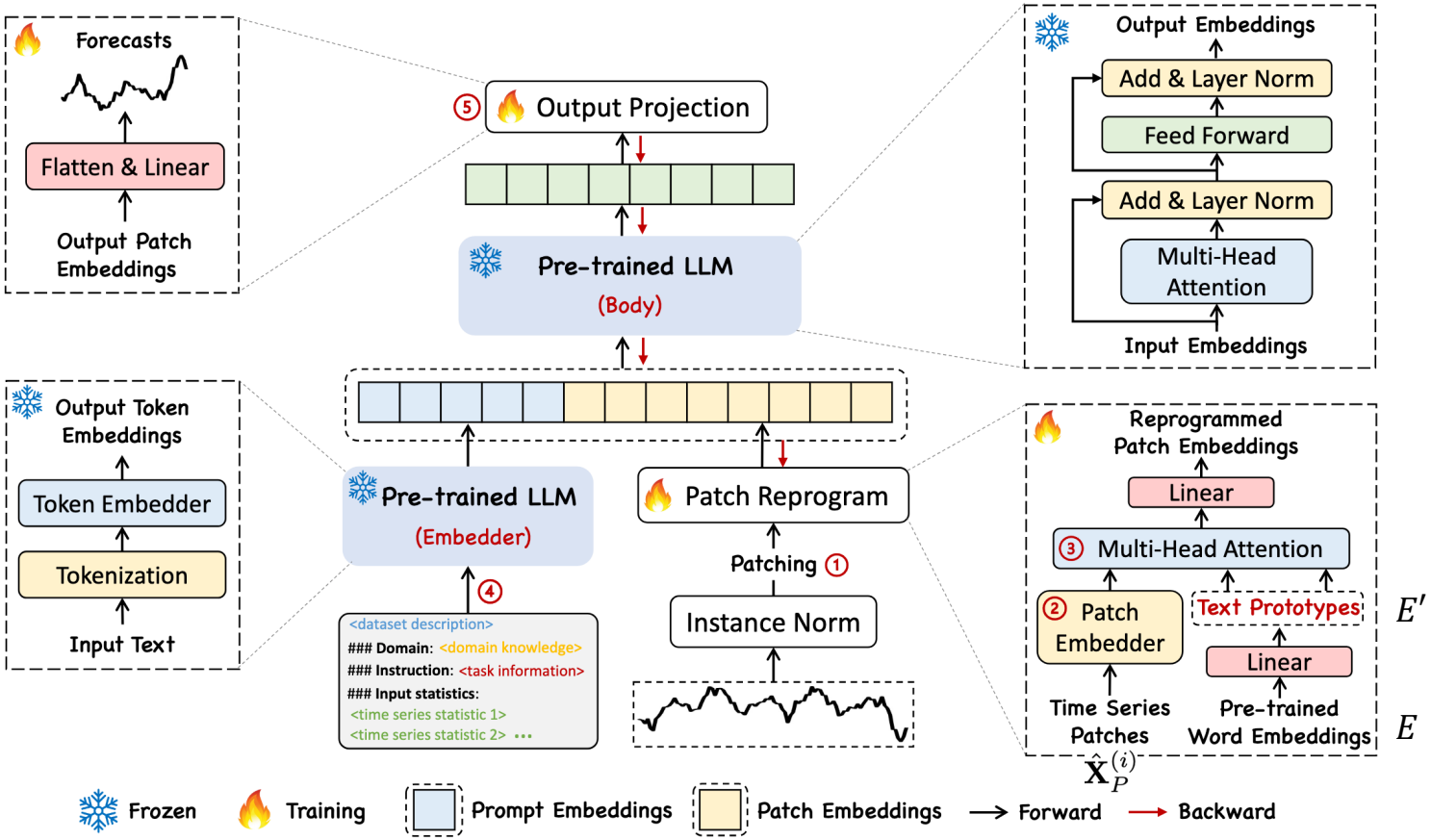


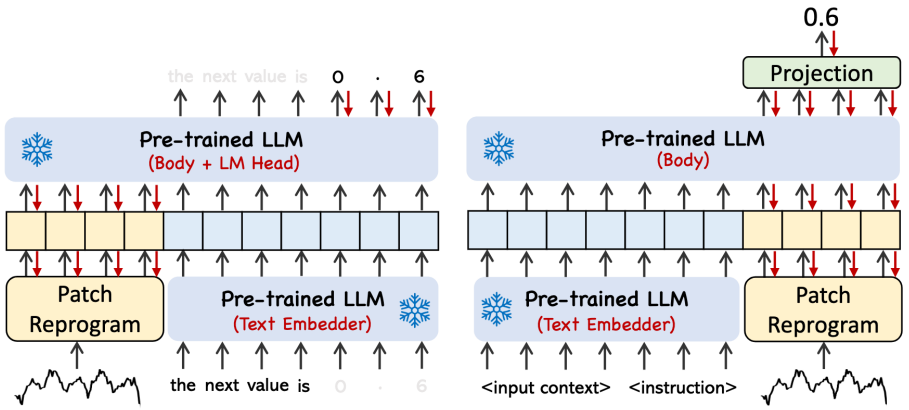
- 텍스트를 soft prompt로 사용
  - 시계열 → main signal / 텍스트 → guiding prompt
  - 즉, 텍스트는 예측보다는 LLM의 해석을 조정하는 역할
- Output Layer에서도 Time series의 hidden state만 활용해서 예측
- LLM은 minimal tuning, 나머지는 prompt tuning
- + Dataset 생성 파이프라인 자체가 contribution

(AAAI 2024) GPT4MTS: Prompt-Based Large Language Model for Multimodal Time Series Forecasting

(Fusion – Intermediate-level) Time-LLM (ICLR 2024) cited ~1,500

- Idea: LLM을 fine-tuning하지 않고, 시계열을 language space로 "재프로그래밍(reprogramming)"해서 forecasting 문제를 language task처럼 푼다



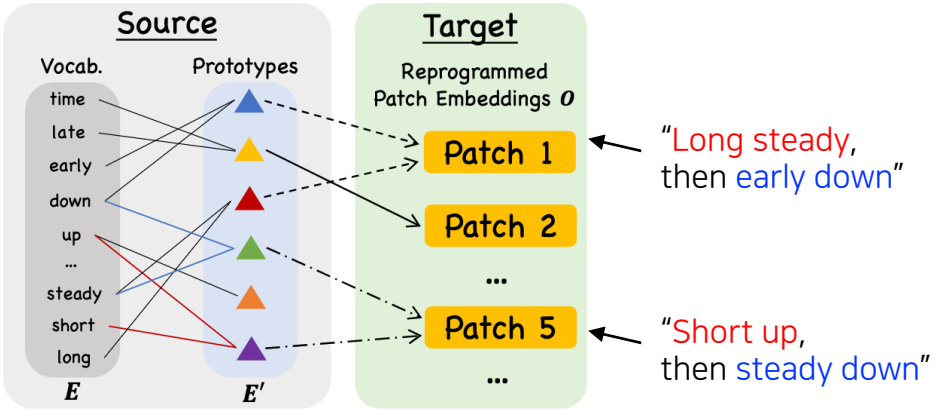


(b) Patch-as-Prefix and Prompt-as-Prefix

(ICLR 2024) Time-LLM: Time Series Forecasting by Reprogramming Large Language Models

How to fuse multi-modality?

- Patch Reprogramming



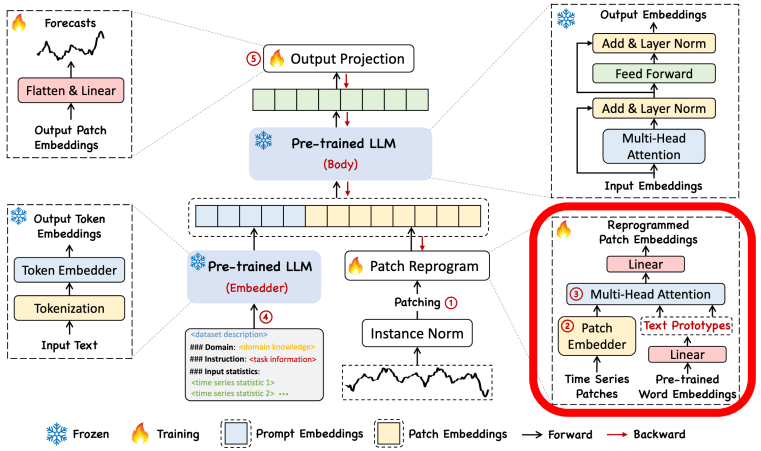
시계열 patch를 "언어적 의미 조합"으로 표현

$$\mathbf{Q}_k^{(i)} = \hat{\mathbf{X}}_P^{(i)}\mathbf{W}_k^Q \quad \mathbf{K}_k^{(i)} = \mathbf{E}'\mathbf{W}_k^K \quad \mathbf{V}_k^{(i)} = \mathbf{E}'\mathbf{W}_k^V$$

$$\mathbf{Z}_k^{(i)} = \text{ATTENTION}(\mathbf{Q}_k^{(i)}, \mathbf{K}_k^{(i)}, \mathbf{V}_k^{(i)}) = \text{SOFTMAX}\left(\frac{\mathbf{Q}_k^{(i)}\mathbf{K}_k^{(i)\top}}{\sqrt{d_k}}\right)\mathbf{V}_k^{(i)}$$

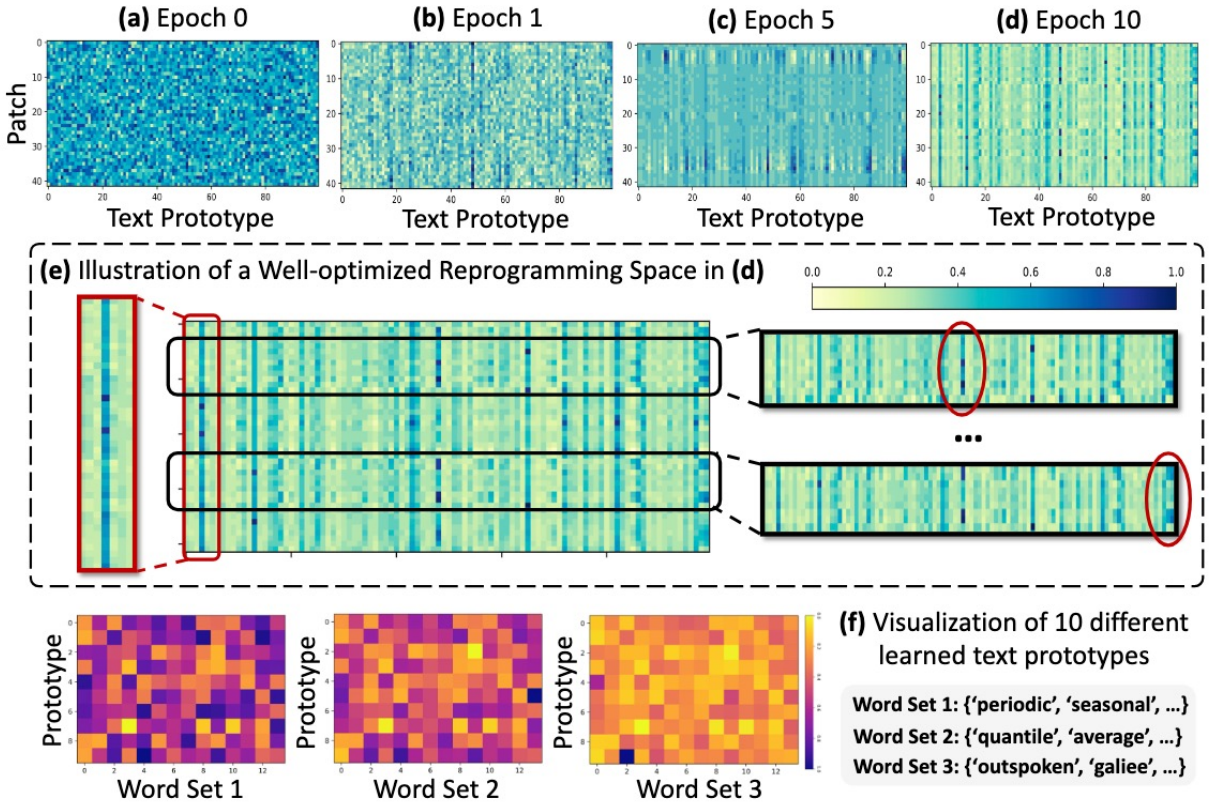
- Query: 시계열 patch embedding

- Key/Value: pretrained word embedding space의 text prototype



- 목표: 입력을 LLM이 이해할 수 있는 표현 공간으로 바꿈 (모델을 바꾸는 대신 데이터를 바꿈)
- 문제: LLM은 discrete token 기반 / 시계열은 continuous vector (Modality mismatch)
- 해결책: 시계열 patch를 LLM의 word embedding space 안에서 표현 (단, vocabulary 전체 쓰지않음)
  - 각 시계열 patch가 여러 text prototype의 weighted combination으로 표현됨

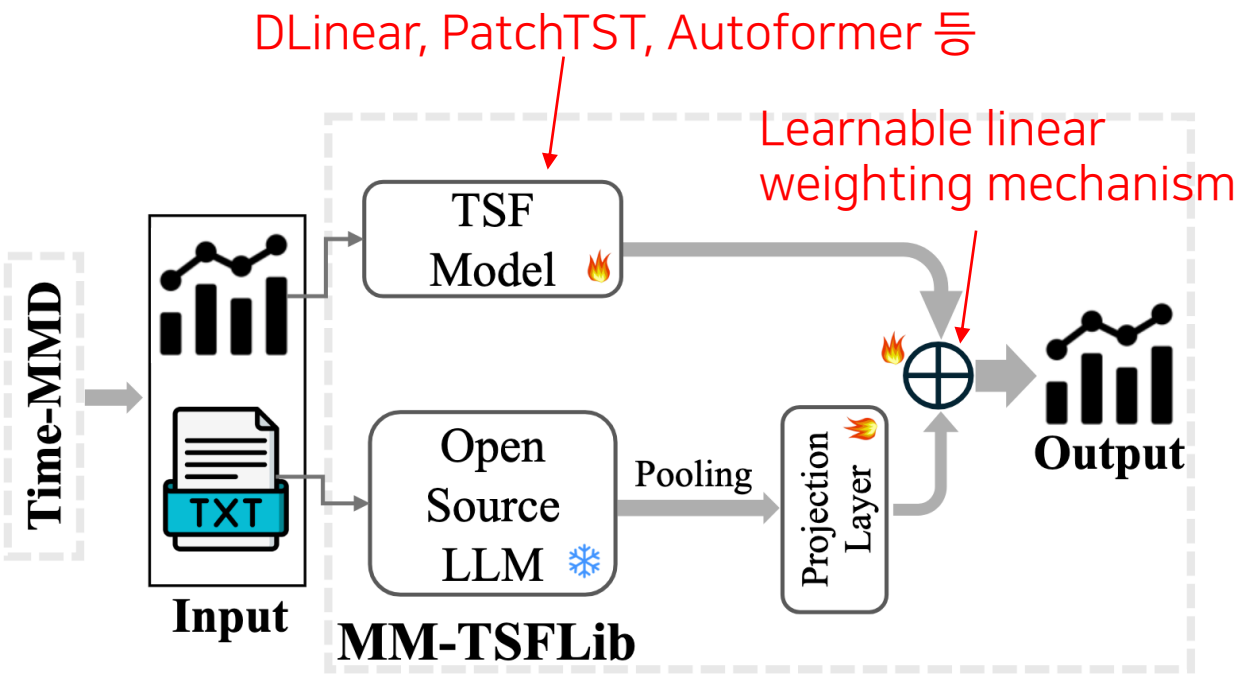
Reprogramming Interpretation



- (a-f) 소수의 prototype만이 patch의 reprogramming에 참여
- (f) Time series와 관련된 단어들 (word set 1,2)이 주로 Activate됨

(Fusion – Output-level) Time-MMD (NeurIPS D&B 2024)

- Idea: Project multiple modality outputs onto a unified space



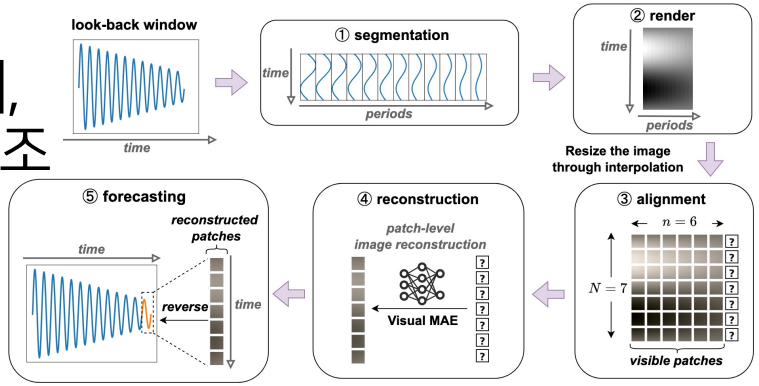
- 기존 시계열 예측 모델(TSF backbone)에 LLM 기반 텍스트 인코더를 결합하는 model-agnostic 멀티모달 확장 프레임워크
- **간단한 모델**: 모델의 복잡도 보다 데이터 alignment가 중요함
  - Temporal Alignment: 각 텍스트에 대해 [start_time, end_time] 구간을 부여하여, 숫자 시계열과 단순 날짜 매칭이 아니라 실제 영향 기간 을 반영
  - Semantic Alignment: LLM 기반 relevance filtering을 통해 해당 도메인과 인과적으로 관련 있는 텍스트만 유지하여 노이즈를 제거
  - Leakage-Free Alignment: 텍스트 내 미래 예측 문장을 제거(fact/prediction 분리)하여 forecasting task에서 데이터 누수를 방지

(NeurIPS 2024) Time-MMD: Multi-Domain Multimodal Dataset for Time Series Analysis

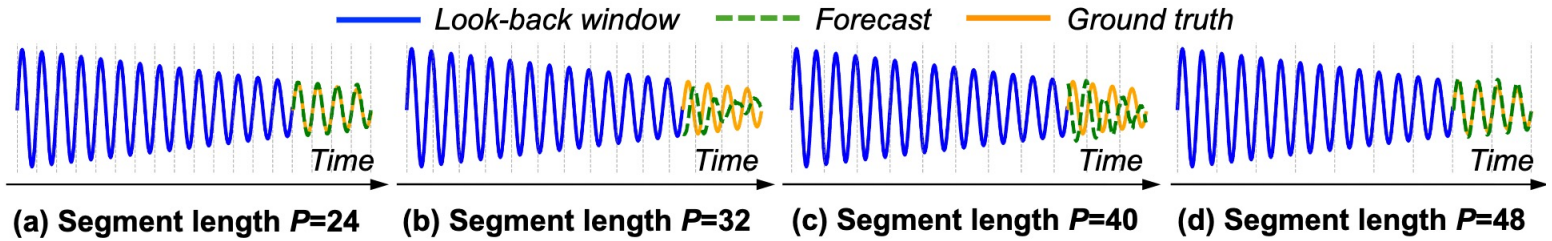
(Fusion – Output-level) DMMV (NeurIPS 2025) *image only (No text)

- Idea: Large Vision Model의 주기성(inductive bias)을 활용하기 위해, 시계열을 분해해서 numerical 모델과 vision 모델을 역할 분담시키는 구조

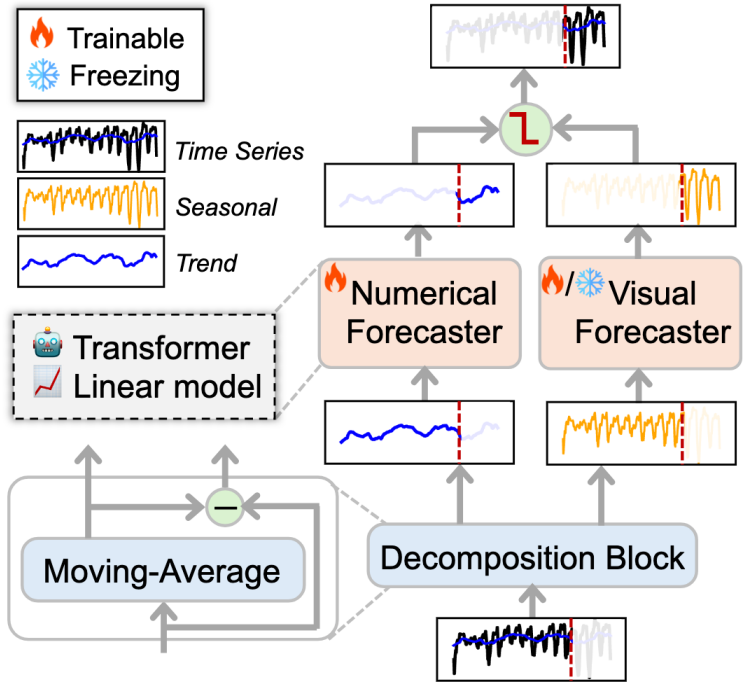
<VisionTS>



<Synthetic sinusoidal time series에 대한 VisionTS의 예측결과>



- 만약 주기가 24인데 P=24면
  → 한 segment (patch) = 정확히 한 반복 패턴 → 모델이 구조를 잘 잡음
- 하지만 P=32면
  → segment 안에 1.33개의 주기가 들어감
  → 패턴이 애매하게 잘림 → 모델이 구조를 잡기 어려움



- 주기의 배수 길이로 자른 segment를 MAE기반 모델들이 더 선호한다는 편향 존재
- 따라서, Visual Encoder의 이런 편향을 활용하자

(NeurIPS 2025) Multi-Modal View Enhanced Large Vision Models for Long-Term Time Series Forecasting

Multi-modal Fusion with Time Series

- Intermediate Level / Output Level

- Fusion은 맥락 정보의 효과적인 활용을 위해 well-aligned multi-modal data에 의존
  - 주가 시계열 + 동일 날짜 뉴스 기사 (well aligned)
  - 주가 데이터는 분 단위인데 뉴스는 하루 단위 (not aligned)
- 실제 환경에서는 이상적으로 정렬된 데이터를 사용할 수 없는 경우가 많음
- 이러한 문제를 완화하기 위해 Alignment mechanism이 활용됨

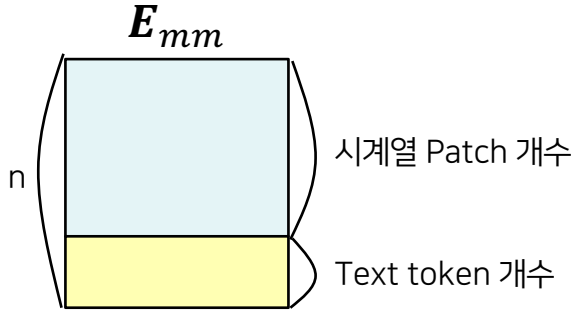
Outline: Multi-modal Time Series Methods

- Background
- Multi-modal Fusion
  - Intermediate-level
  - Output-level
- Multi-modal Alignment
  - Intermediate-level: Self-attention / Cross-attention / Graph Neural Network / Contrastive Learning
  - Output-level: Retrieval / LLM reasoning

Self-attention 기반 Alignment

- **Self-attention:** a joint and undirected alignment across all modalities by dynamically attending to important features
- Given multi-modal embeddings $\boldsymbol{E}_{mm} \in \mathbb{R}^{n \times d}$, where n is the number of modality tokens and d is the embedding dimension:

$$\text{Attention}(\boldsymbol{E}_{\text{mm}}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{\boldsymbol{d}_{\boldsymbol{k}}}}\right)\boldsymbol{V}$$

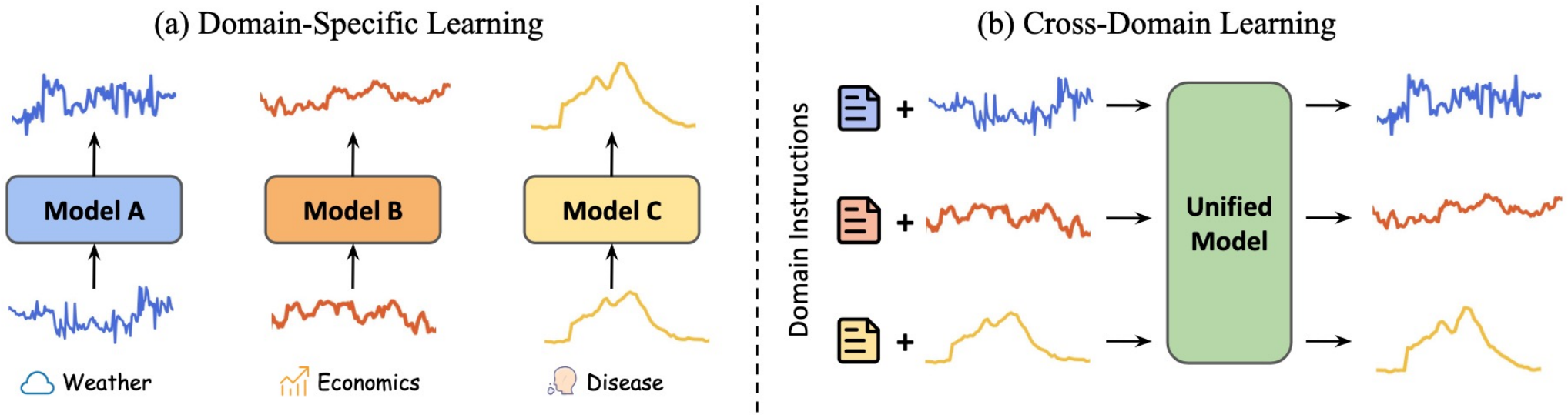


where the queries Q, keys K, and values V are linear projections of $\boldsymbol{E}_{mm}$:

$Q = \boldsymbol{E}_{mm} \times W_Q, K = \boldsymbol{E}_{mm} \times W_K, \text{V} = \boldsymbol{E}_{mm} \times W_V$ with learnable weights W_Q, W_K, W_V

(Alignment – Intermediate-level-SA) UniTime (WWW 2024)

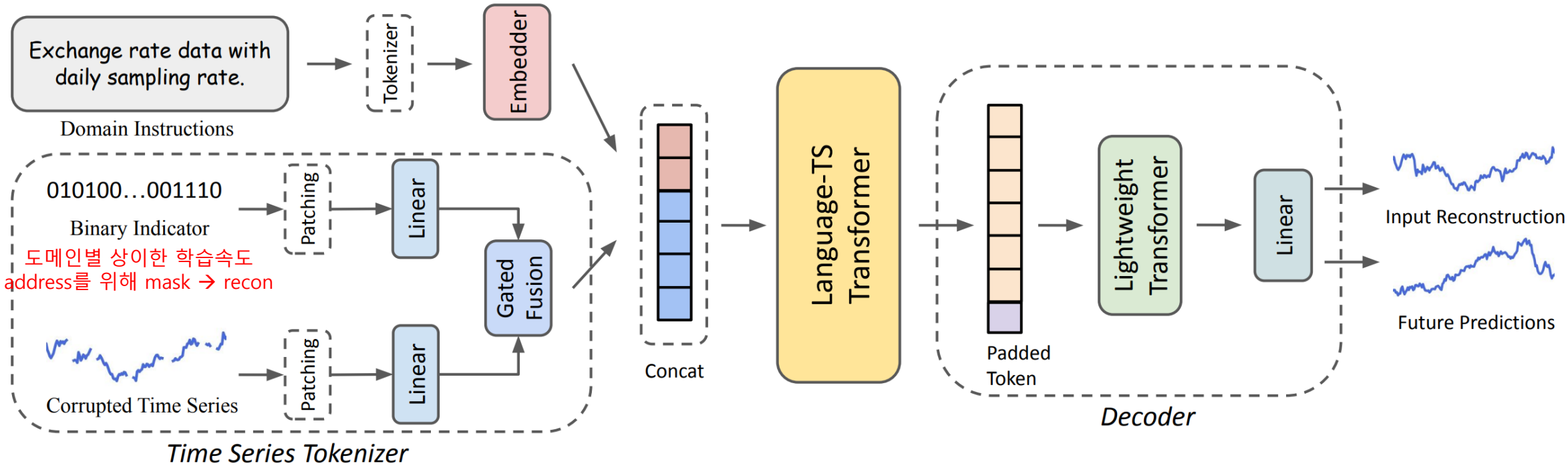
- 시계열 데이터는 도메인은 달라도 **추세(trend), 계절성(seasonality), 주기성 같은 공통 구조가 있음**
  → 하나의 통합 모델로 학습하면 더 잘 일반화할 수 있다.



- 하지만 cross-domain 학습에는 3가지 문제가 존재
  1. **도메인마다 변수 개수, 길이 등이 다름**
  2. 모델이 **어느 도메인 데이터인지 헷갈리는 문제** (domain confusion)
  3. 어떤 도메인은 빨리 overfit, 어떤 건 느리게 학습되는 **수렴 속도 불균형 문제**

UniTime

- Idea
  - 1) Channel-independent 설계를 통해 변수 개수 차이 문제 해결
  - 2) 명확한 도메인 식별 정보를 제공하기 위해 사람이 직접 작성한 지침을 추가함으로써, 도메인 혼동 문제를 완화
  - 3) Masking을 통해 수렴 불균형 해결

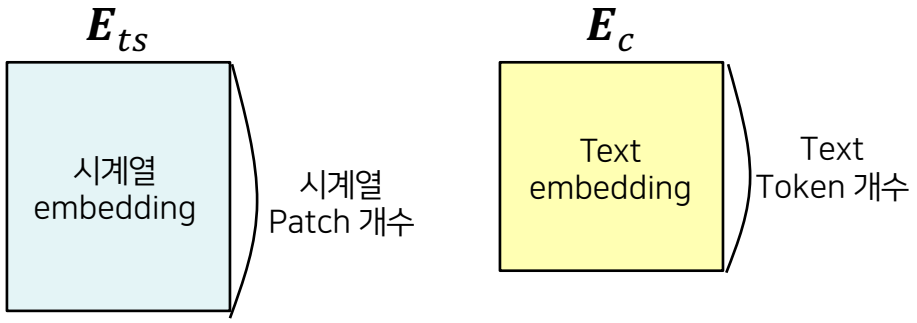


Cross-attention 기반 Alignment

- **Cross-attention**: time series serves as the query modality to get contextualized by other modalities, providing a directed alignment that ensure auxiliary modalities contribute relevant contexts while preserving the temporal structure of time series.
- Given multi-modal embeddings $\boldsymbol{E}_{ts} \in \mathbb{R}^{n \times d}$, where n is the number of time series patches and d is the embedding dimension

$$\text{CrossAttention}(\boldsymbol{E}_{\text{ts}}, \boldsymbol{E}_{\text{C}}) = \text{softmax}\left(\frac{\boldsymbol{Q}_{\text{ts}}\boldsymbol{K}_{\text{C}}^{\top}}{\sqrt{\boldsymbol{d}_{\boldsymbol{k}}}}\right)\boldsymbol{V}_{\text{C}}$$

Q K,V

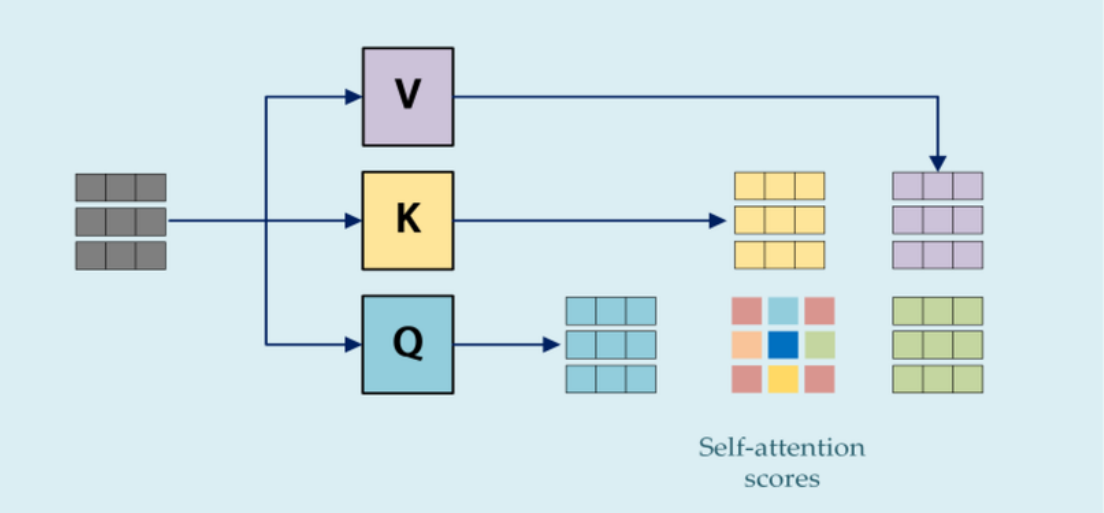


where the queries Q_{ts}, keys K_C, and values V_C are linear projections of $\boldsymbol{E}_{ts}$ and $\boldsymbol{E}_C$:

$Q_{ts} = \boldsymbol{E}_{ts} \times W_Q, K_C = \boldsymbol{E}_C \times W_K, \mathrm{V}_C = \boldsymbol{E}_C \times W_V$ with learnable weights W_Q, W_K, W_V

Self-attention vs Cross-attention

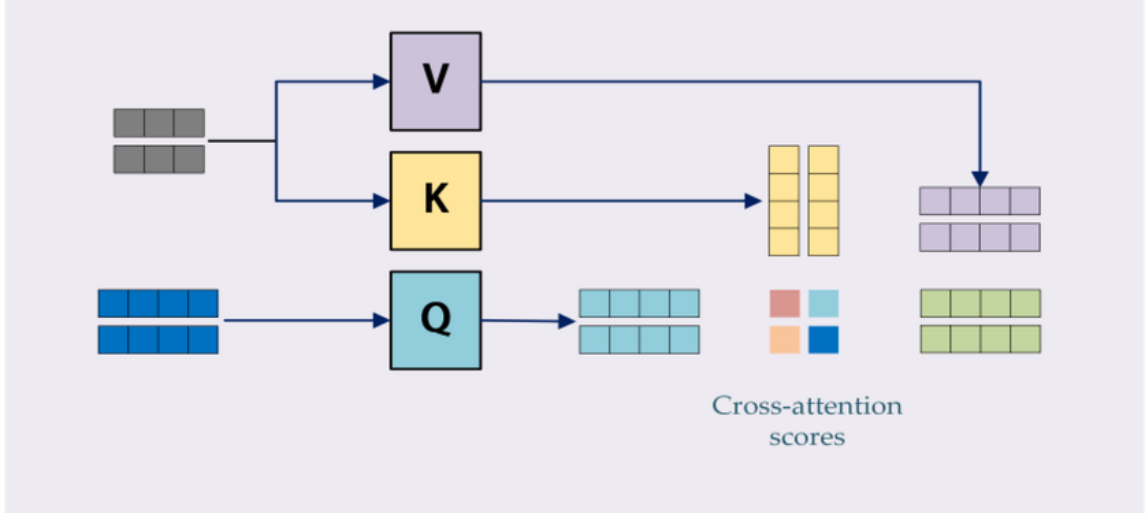
Self-attention



- Query, Key, Value가 **같은 입력**에서 생성
- 동일 시퀀스 내 **내부 관계(intra-dependency)** 학습
- 문맥 이해, 시계열 내 패턴 학습에 사용
- 예시: 문장 내 단어 간 관계 / 시계열 내부 시점 간 상관성

👉 **두 모달을 이미 같은 표현 공간으로 투영했다**는 가정

Cross-attention

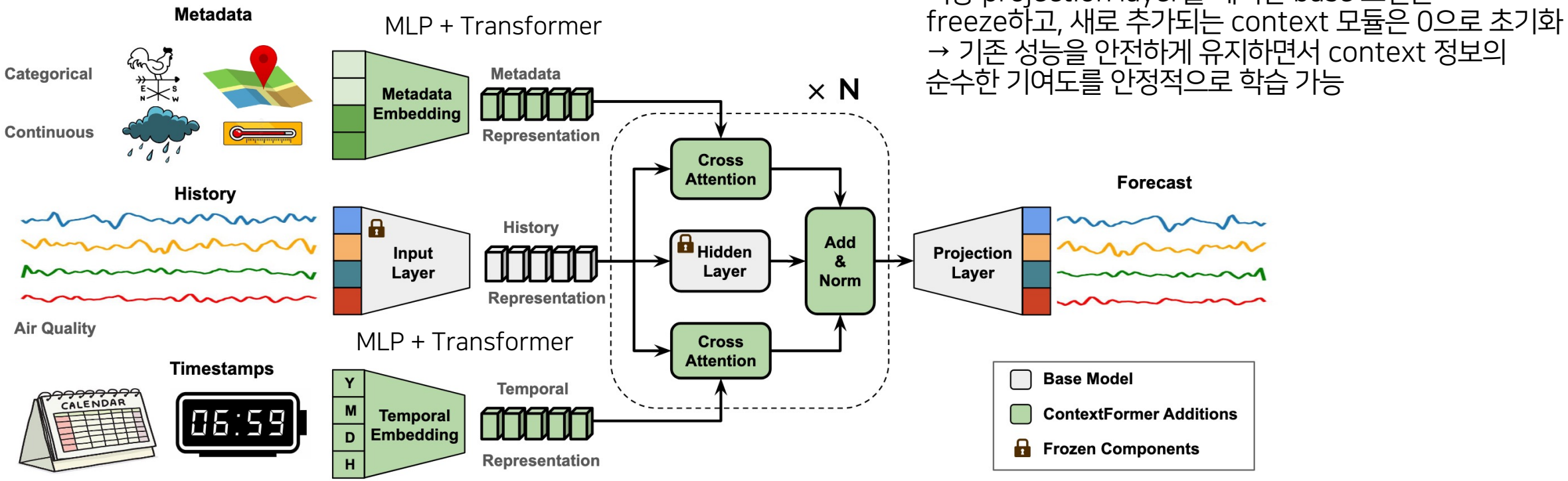


- Query와 Key/Value가 **다른 입력**에서 생성
- 서로 다른 모달리티/시퀀스 간 **정렬 및 융합(inter-dependency)** 학습
- 멀티모달 fusion의 핵심 메커니즘
- 예시: 시계열 + 텍스트 융합 / 이미지 + 텍스트 정렬

👉 **아직 두 모달이 같은 공간은 아니라는** 가정

(Alignment – Intermediate-level-CA) ContextFormer (arxiv 2025)

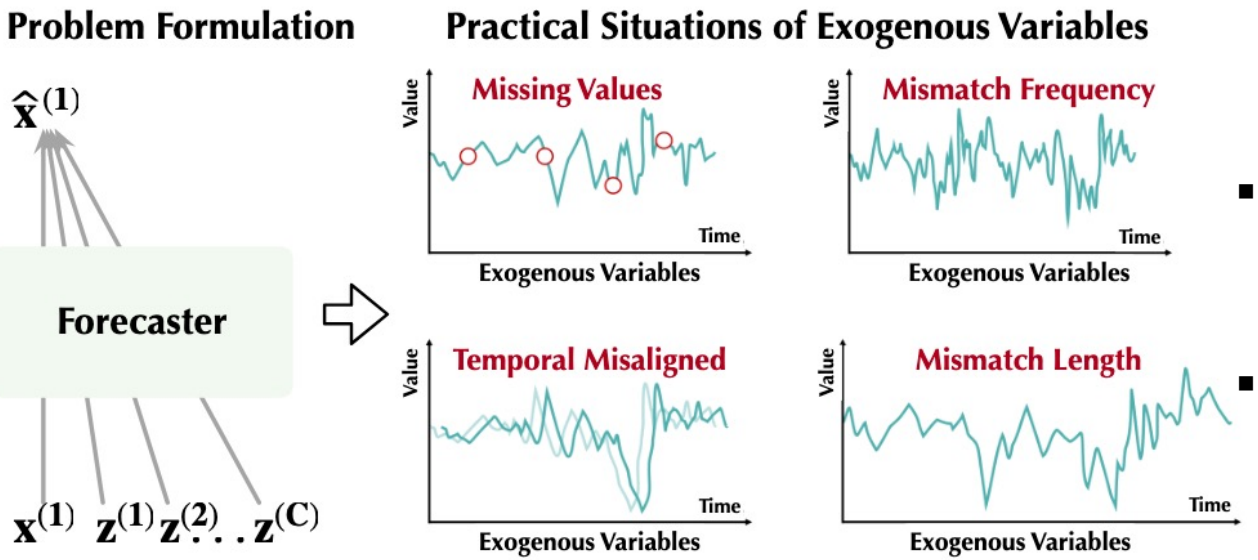
- Motivation: 현실 시계열 예측에는 맥락 정보가 필수지만, 기존 SOTA 모델들은 이를 효과적으로 활용하지 못함
- Key idea: 기존 시계열 모델 위에 cross-attention 기반 plug-and-play 모듈을 얹어, 성능 저하 없이 예측에 유효한 맥락만 선택적으로 통합
- 학습전략: Base Model Freeze + Zero init.
  - 최종 projection layer를 제외한 base 모델은 freeze하고, 새로 추가되는 context 모듈은 0으로 초기화 → 기존 성능을 안전하게 유지하면서 context 정보의 순수한 기여도를 안정적으로 학습 가능



(arxiv 2025) Context Matters: Leveraging Contextual Features for Time Series Forecasting

(Alignment – Intermediate-level-CA) TimeXer (NeurIPS 2024)

- Motivation: 단순 concat을 넘어, Exogenous 변수의 시간 지연과 변수 간 관계를 구조적으로 모델링할 필요성



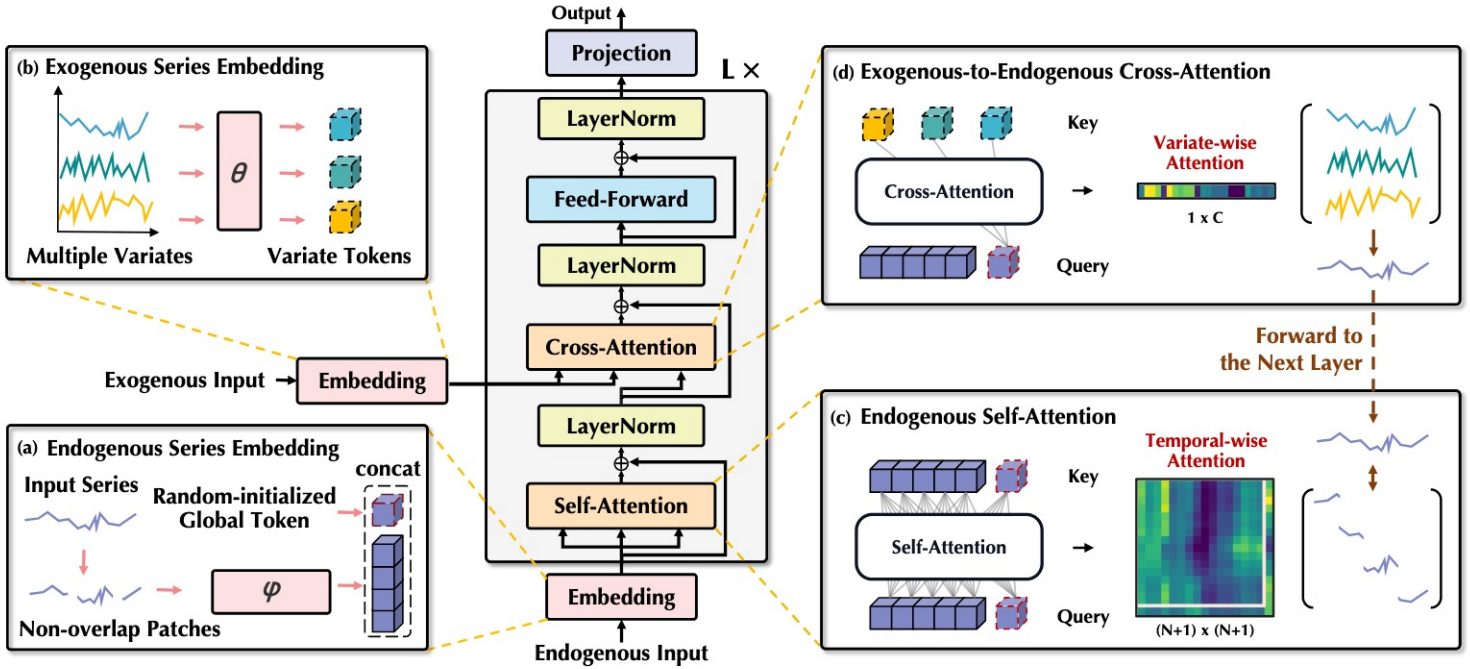
- 현실 세계 예측은 외부 요인의 영향을 받음
  - 실제 시계열 예측에서는 외부 변수들이 중요한 보조 정보로 작용
  - 예시 – 전력가격) 전력 가격은 수요/공급 같은 외부 요인의 영향을 크게 받아서, 과거 가격만으로는 본질적으로 한계가 있음
- 기존 방법은 concat에 의존
  - Endogenous와 exogenous를 시간축에 맞춰 단순 concat하는 방식은 시간 불일치, 결측치, 길이 차이 문제를 유발
- Exogenous 변수는 시간 지연(lag) 및 변수간 인과 관계 고려 필수
  - 외부 변수의 영향은 시간이 지나 나타날 수 있기 때문에, 모델은 "시간 안에서의 변화"와 "변수 간 영향"을 동시에 학습해야 함
  - 예시
    - 전력가격) 기온 상승 → 냉방 수요 증가 → 몇 시간 뒤 가격 상승
    - 기업매출) 금리 인상 → 몇 달 뒤 소비 감소 → 기업 매출 하락

(NeurIPS 2024) TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables

TimeXer

- Key Idea: endogenous는 시간 패턴이, exogenous는 변수 관계가 중요
  - endogenous는 시간적 변화(temporal variation)가 핵심 → patch로 잘라야 잘 잡힘
  - exogenous는 수십~수백 개일 수 있고, 각 변수는 "외부 원인"이라서 변수 간 서로 복잡하게 섞기보다 target과의 관계에 중점

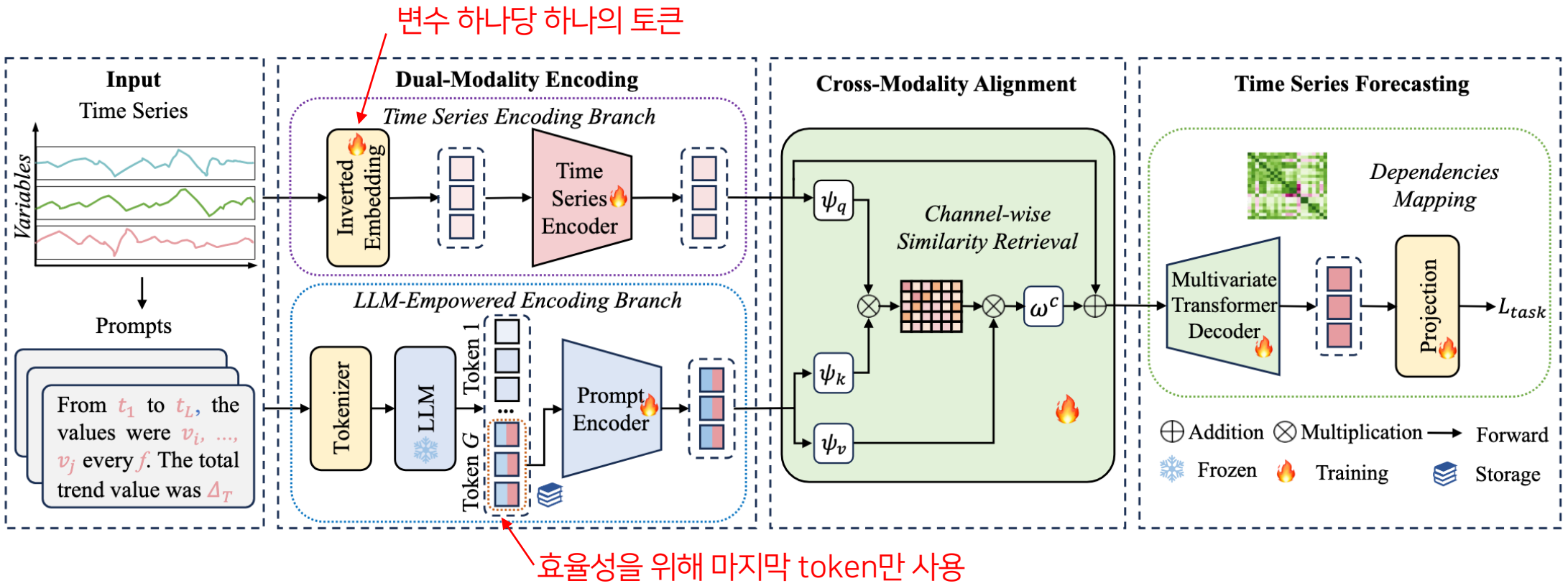
  → 두 모달을 **같은 수준의 토큰으로 섞지 않고**, 역할에 맞게 **다른 granularity로 임베딩**함



- Patch-level Temporal Modeling
  - Endogenous 시계열을 patch 단위 토큰으로 분할하여, self-attention을 통해 장기 시간 의존성(temporal dependency)을 학습
- Variate-level Exogenous Tokenization
  - 각 외생 변수를 하나의 variate-level 토큰으로 임베딩하여, 변수 간 불필요한 상호작용을 줄이고 효율적으로 표현
- Global Token 기반 Cross-Attention Alignment
  - Learnable global token을 매개로 endogenous–exogenous 간 cross-attention을 수행하여, 단순 concat 대신 구조적 정렬(alignment)을 구현.

(Alignment – Intermediate-level-CA) TimeCMA (AAAI 2025)

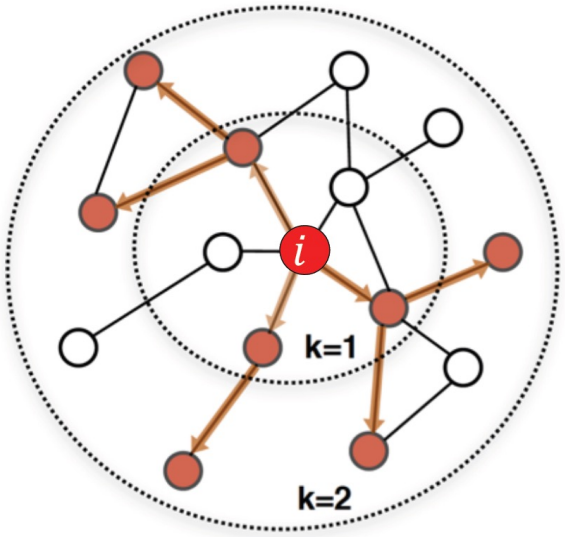
- LLM은 대규모 언어 데이터로부터 학습된 **강건한 표현력** 덕분에 시계열 예측에서도 성능 향상을 보이나,
  - 시계열+텍스트 결합 시 **임베딩 entanglement** 발생 → 텍스트가 노이즈로 작용
- Key idea: 시계열 임베딩을 query로 하는 cross-modality alignment를 통해 LLM prompt에서 강건하면서도 disentangled된 시계열 표현만을 선택적으로 추출



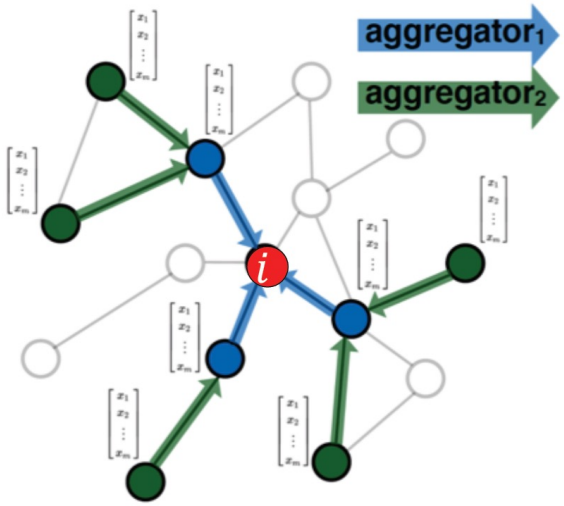
(AAAI 2025) TimeCMA: Towards LLM-Empowered Multivariate Time Series Forecasting via Cross-Modality Alignment

GNN 기반 Alignment

- The topological structure from external contexts can be used for alignment. It explicitly aligns representations with relational structures, enabling context-aware feature propagation across modalities.



Determine node computation graph



Propagate messages and transform information

$$\boldsymbol{h}_v^{(l+1)} = \sigma\left(\boldsymbol{W}_l \sum_{u \in N(v)} \frac{\boldsymbol{h}_u^{(l)}}{|N(v)|} + \boldsymbol{B}_l \boldsymbol{h}_v^{(l)}\right),$$

Weight matrix

Embedding of v at layer l

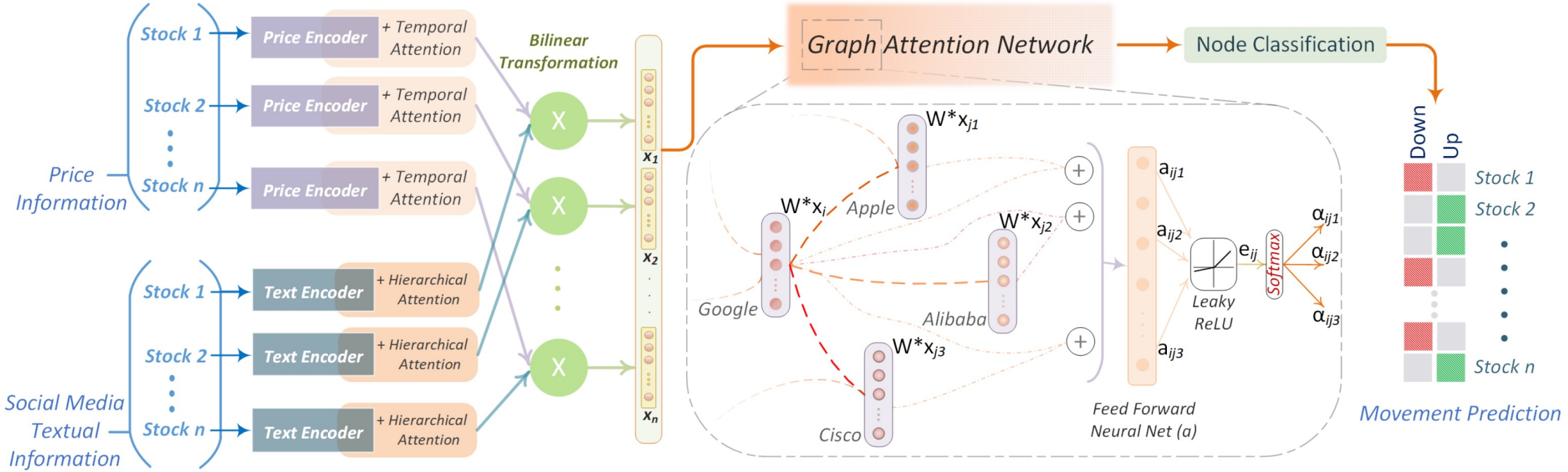
Average of neighboring nodes' previous layer embeddings

$$\forall l \in \{0,1,\ldots,L-1\}$$

Total number of layers

(Alignment – Intermediate-level - GNN) MAN-SF (EMNLP 2020)

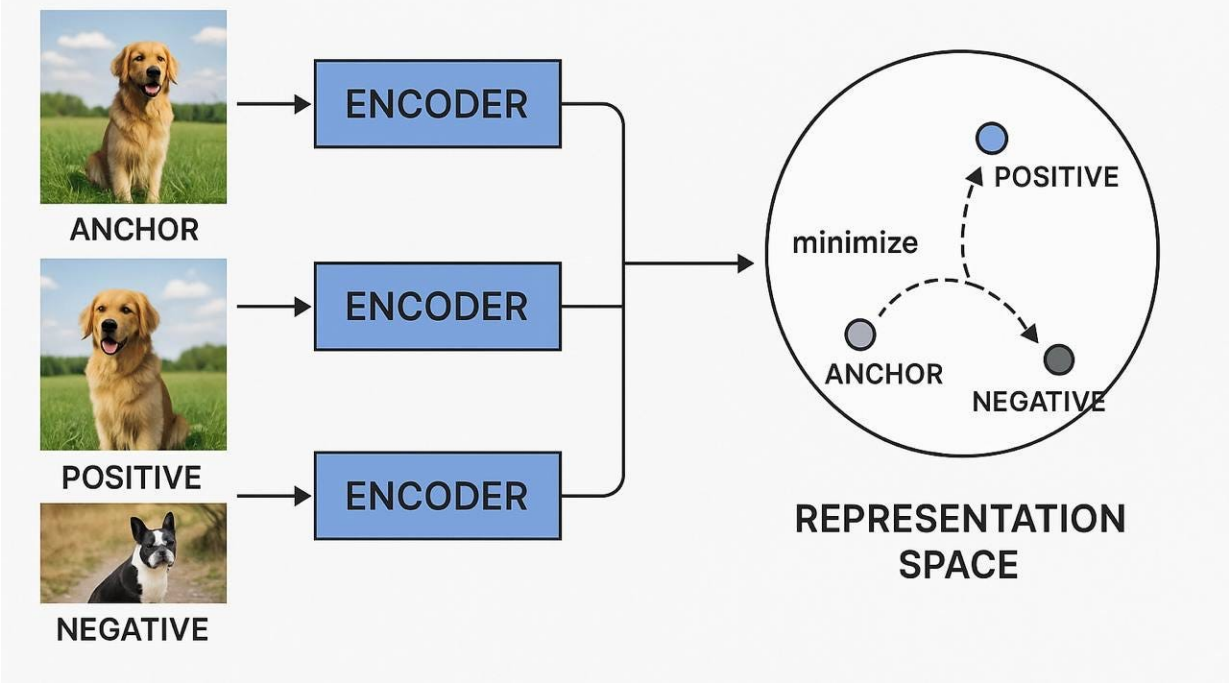
- 기업 간 관계를 Wikidata 기반 그래프로 구성하고, Graph Attention Network(GAT)를 통해 주식 간 영향력을 가중 학습하여 개별 종목 예측에 반영
- Graph 생성 방법
  - 1차 관계: Wells Fargo → owned by → Berkshire Hathaway
  - 2차 관계: Microsoft → owned by → Bill Gates / Bill Gates → board member of → Berkshire Hathaway



(EMNLP 2020) Deep Attentive Learning for Stock Movement Prediction from Social Media Text and Company Correlations

Contrastive Learning 기반 Alignment

- **Contrastive learning**: a self-supervised representation learning approach that pulls semantically similar (positive) pairs closer together while pushing dissimilar (negative) pairs farther apart in the embedding space.

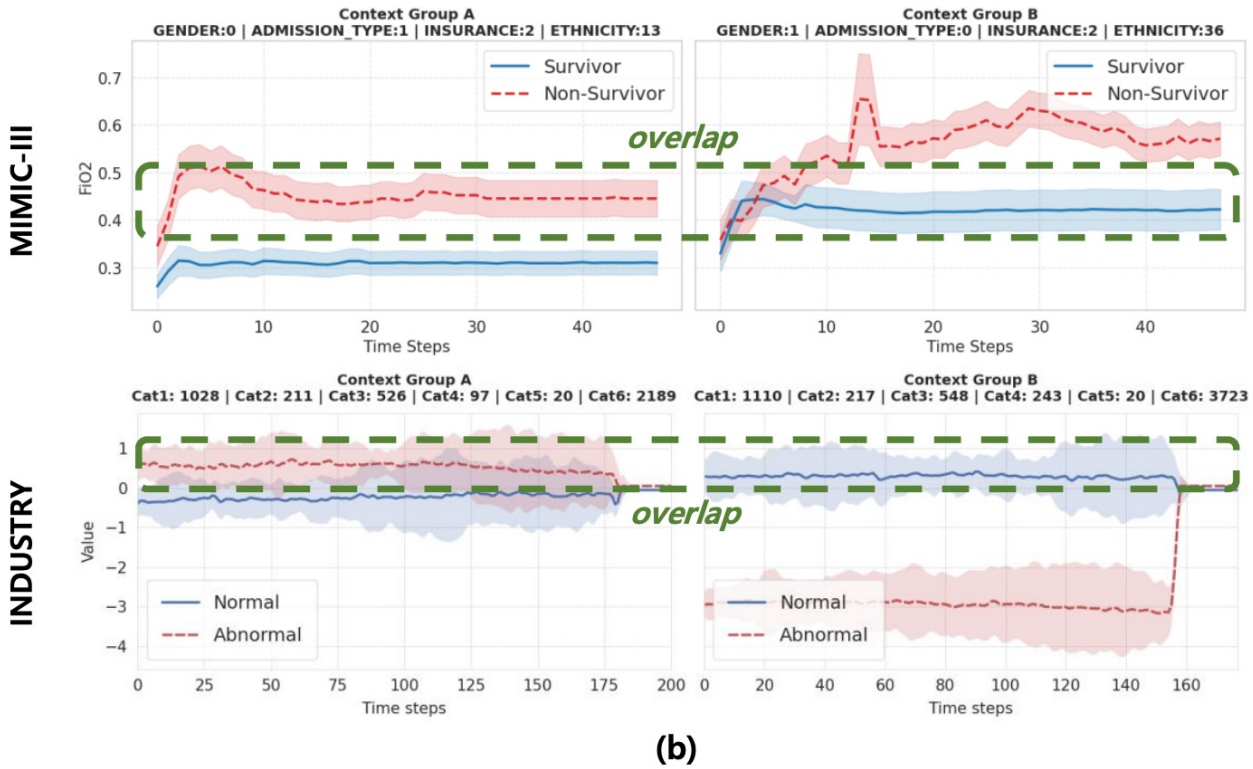


$$L_i = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)}$$

- z_i: Anchor representation
- z_j: Positive sample
- z_k: batch내 모든 샘플
- τ: Temperature

(Alignment – Intermediate-level – CL) TS-C3 (Under Review)

- Motivation: 다른 context group에 속한 서로 다른 클래스가 매우 유사한 시계열 분포를 가질 수 있음
- Label은 단순히 x로 결정되지 않고 $(x, context)$로 결정됨



(b)

예시) 의료 도메인

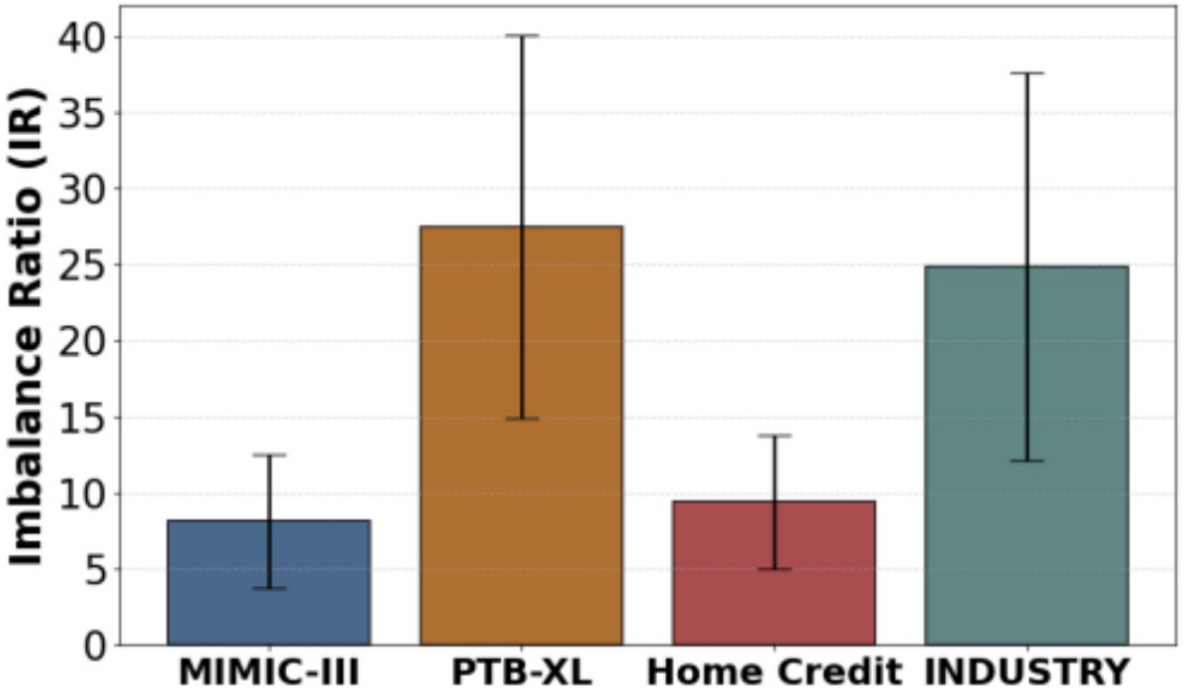
- 동일한 혈압·심박수 시계열 패턴
- 그러나
  - 젊은 환자 → 정상
  - 고령 + 기저질환 환자 → 중증

→ 같은 패턴이지만 나이에 따라 진단(label)이 달라질 수 있음

이를 무시하면 Global한 Decision boundary를 학습. 즉, 같은 패턴인데 context에 따라 label이 달라지는 상황을 처리하지 못함 (Decision boundary heterogeneity)

TS-C3

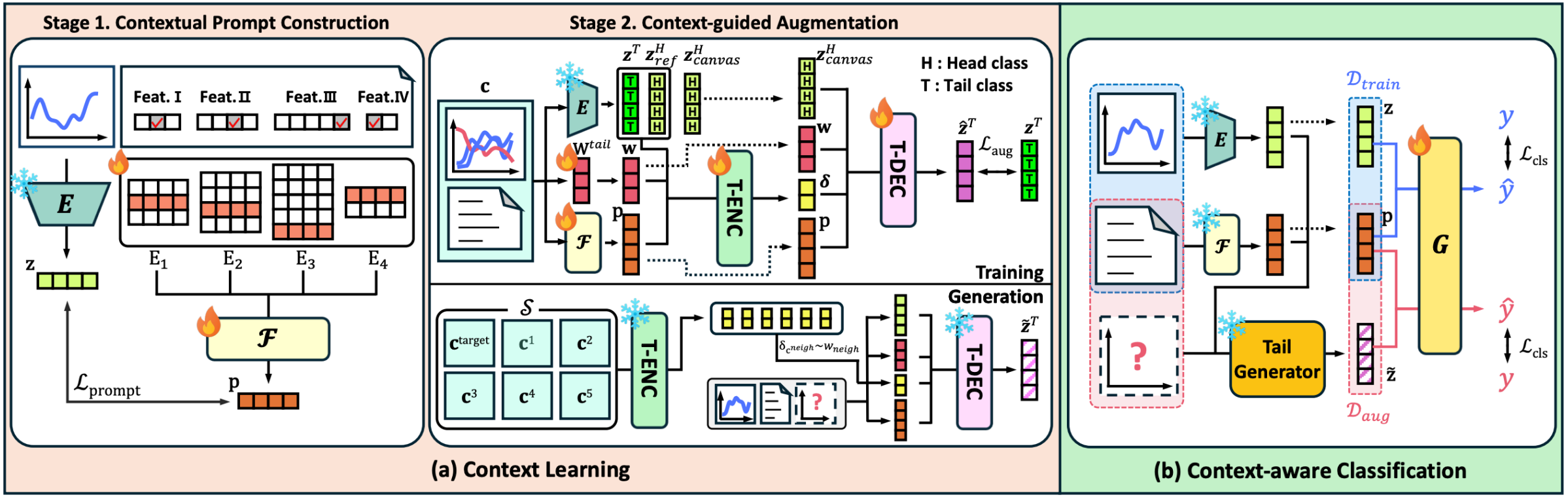
- 클래스 불균형 정도가 context group마다 크게 다름



- 각 데이터셋에 대해 **context group별 imbalance ratio (IR)** 를 계산한 뒤 그 **평균/표준편차** 를 시각화
  - IR: # head samples / # tail samples
  - 평균보다 표준편차에 주목
- 해석: 어떤 context group에서는 tail class가 매우 희소할 수 있고 다른 context group에서는 비교적 균형적일 수 있음
  - Global imbalance보다 Context별 imbalance 차이가 더 문제다

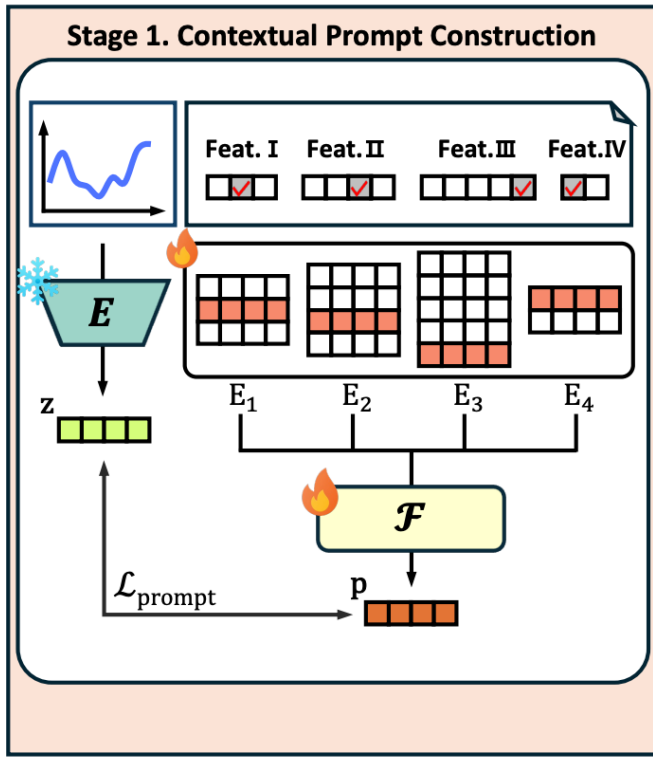
→ 이를 무시한채 **global하게 하나의 imbalance strategy를 적용하면 잘못된 보정이 될 수 있음** (class distribution heterogeneity)

TS-C3


Stage 1. Contextual Prompt Construction
Feat. I
Feat. II
Feat. III
Feat. IV
E
z
E1
E2
E3
E4
F
Lprompt
p
Stage 2. Context-guided Augmentation
c
E
z^T z^H_ref z^H_canvas
z^H_canvas
H : Head class
T : Tail class
w^tail
w
F
p
T-ENC
w
δ
p
T-DEC
ẑ^T
L_aug
z^T
Training
Generation
S
c^target
c^1
c^2
c^3
c^4
c^5
T-ENC
δ_c^neigh ~ w_neigh
?
T-DEC
z̃^T
(a) Context Learning
D_train
E
z
F
p
?
Tail Generator
G
z̃
D_aug
y
L_cls
ŷ
ŷ
L_cls
y
(b) Context-aware Classification

TS-C3

- Idea: Context 자체를 distribution-level에서 학습하자 (Decision boundary heterogeneity 해결)



- 1. Prompt 생성
  - 해당 context group의 distribution blueprint
- 2. Temporal Representation과 정렬
  - 동일 context 내 여러 시계열 표현 z_i들의 **분포 중심 방향으로 정렬되도록 학습**
  - Context간 Representation 분리

$$p_c \approx E[z|c]$$

- 3. Decoupled Learning for Classification
  - Classification과 분리된 decoupled 학습을 통해 context-dependent decision boundary를 가능하게 함

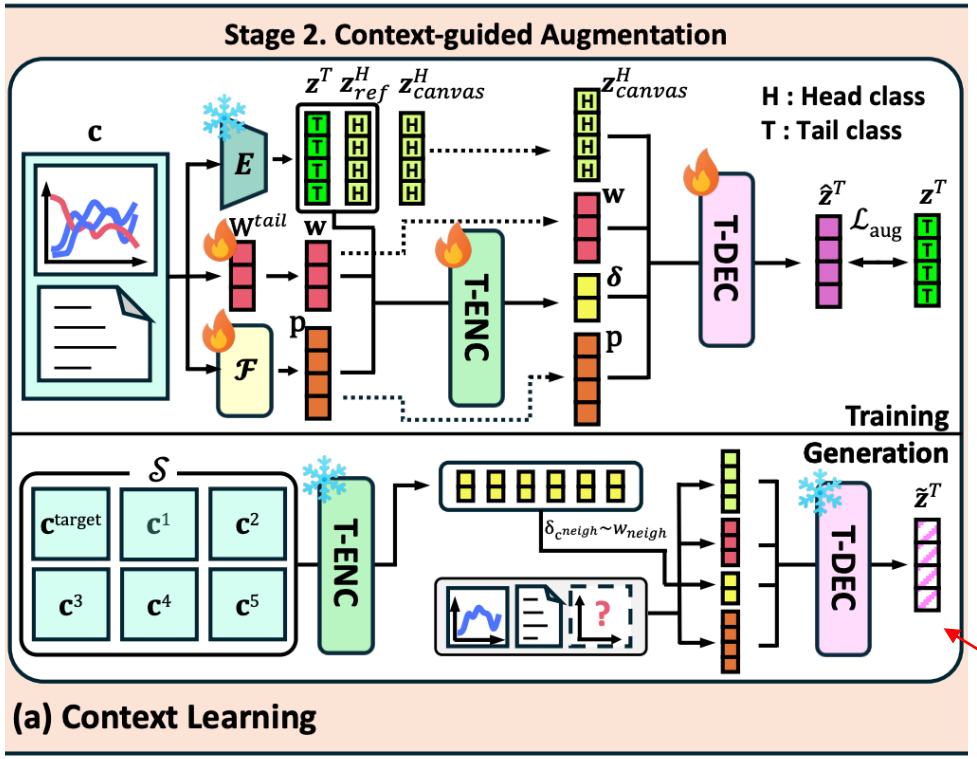
$$\hat{y} = G([z||p_c])$$

  - 같은 z라도 p_c에 따라 다른 예측

$$\mathcal{L}_{\text{prompt}} = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{x}_i \in \mathcal{B}} \sum_{r=1}^{R} \max\left(0, \alpha + \|\mathbf{z}_i - \mathbf{p}_i\|_2 - \|\mathbf{z}_i - \bar{\mathbf{p}}_{i,r}\|_2\right),$$

TS-C3

- 목표: minority class를 context-aware하게 보강(augmentation) 하자
- Idea: Head와 Tail의 의미적 차이를 context 조건 하에 학습해, 각 context의 분포를 유지하면서 tail을 생성하고 불균형을 보정 (Class distribution heterogeneity 해결)

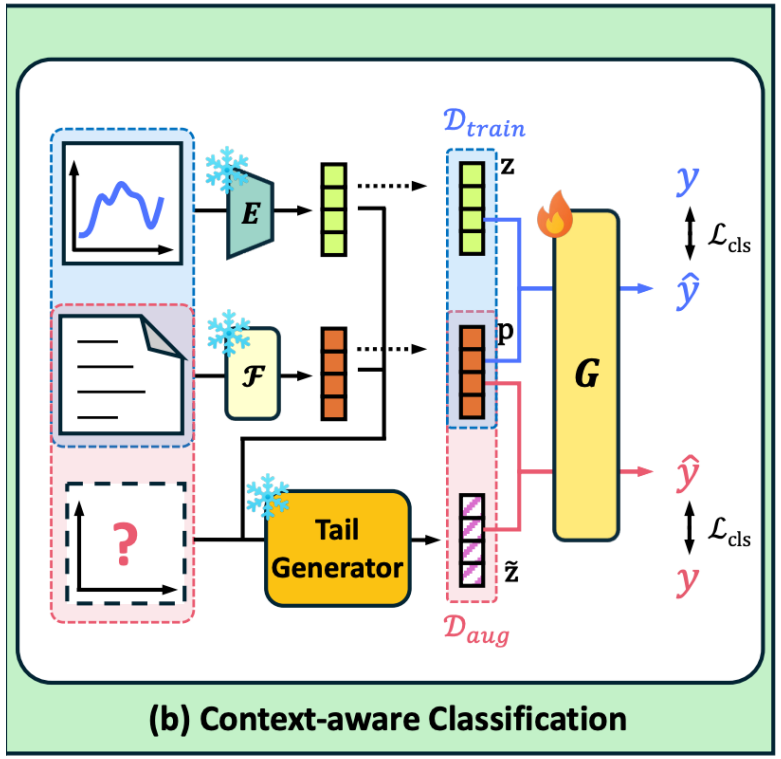


- Head→Tail latent transition을 학습하여, 풍부한 head representation을 기반으로 tail representation을 생성함
  - 의료 예시
    - 고령 환자 group에서 생존 케이스와 사망 케이스의 차이 학습
    - 사망 환자는 더 불안정한 혈압 / 점진적 산소 요구량 증가 등

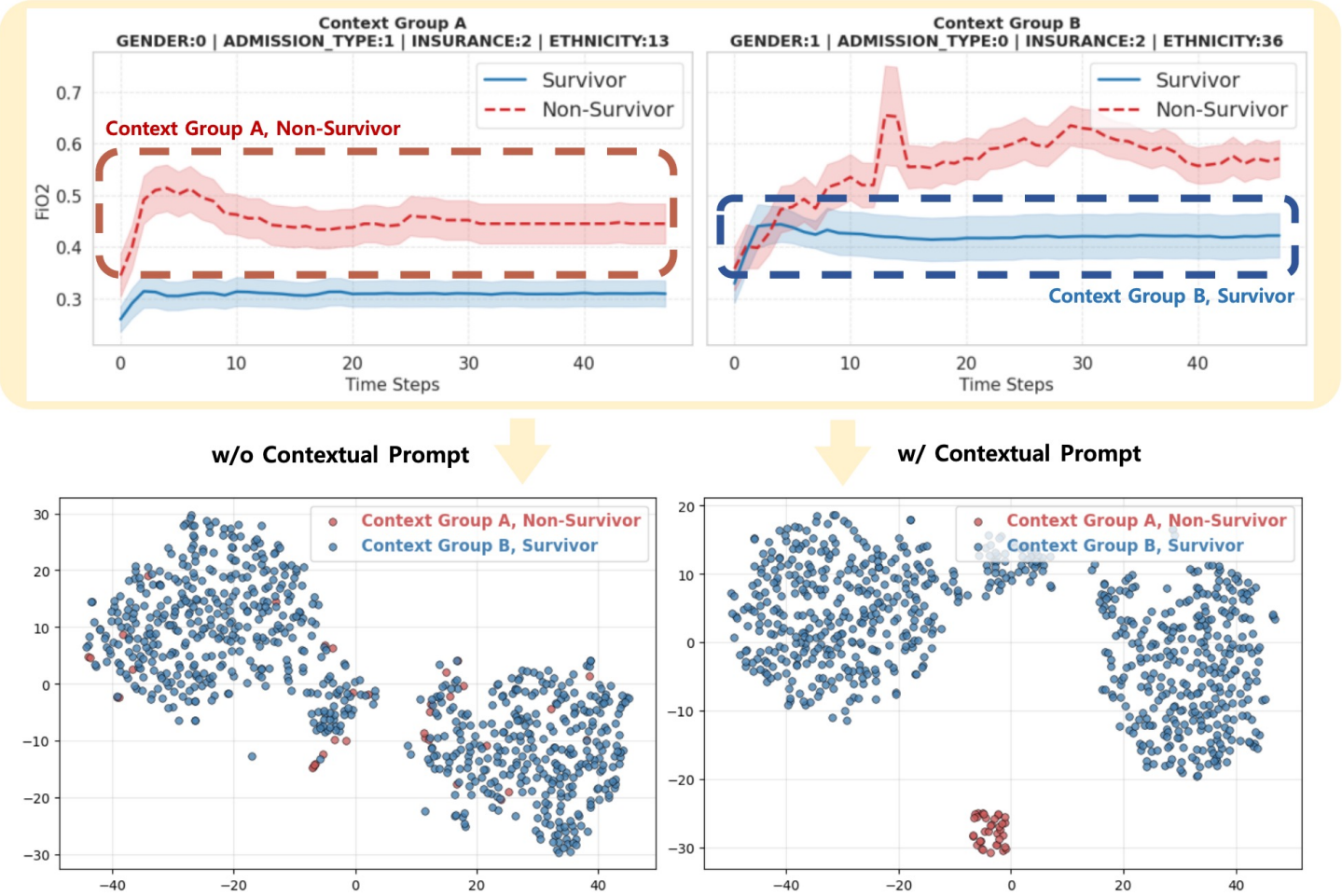
    → 이 패턴 차이가 latent space에 반영됨.
- Contextual prompt를 조건으로 사용해 생성된 샘플이 해당 context의 분포 특성을 따르도록 함
  - 의료 예시
    - 고령 환자 group에서 사망 케이스가 부족할 때, 젊은 환자 패턴을 복사하면 안 됨
    - 고령 context manifold 위에서 synthetic sample 생성해야 함
- 유사 context 간 knowledge transfer를 활용해 tail이 극도로 희소한 그룹에서도 안정적으로 균형을 맞춤

TS-C3

- Contextual prompt와 temporal representation을 결합하여, context-dependent decision boundary를 학습하는 최종 분류



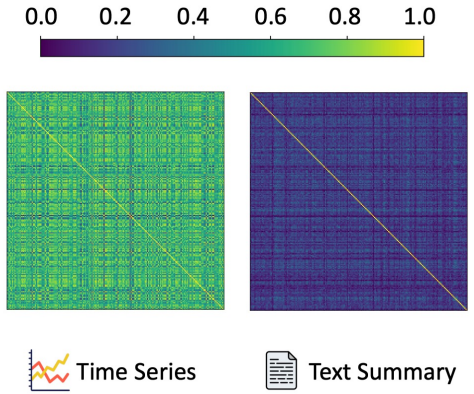
Result



- Without contextual prompt (좌측)
  - 서로 다른 context에 속하지만 시계열 패턴이 유사한 샘플들이 latent 공간에서 겹쳐 나타남 → global boundary로는 구분이 어려움
- With contextual prompt (우측)
  - z에 context prompt p를 결합하자, 같은 시계열이라도 context에 따라 representation이 분리됨 → context별 cluster 형성
- 결론: Contextual prompt는 단순 feature 추가가 아니라, context-dependent decision boundary를 가능하게 하는 분포 분리 역할을 수행함

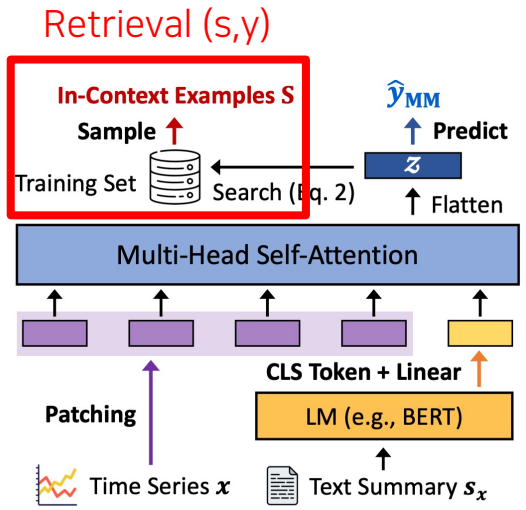
(Alignment – Output-level – Retrieval) TimeCAP (AAAI 2025)

- LLM을 시계열의 Predictor뿐 아니라 contextualizer로도 활용
- 멀티모달 인코더와 결합해 contextualize-augment-predict 구조로 시계열 이벤트 예측 성능을 크게 향상

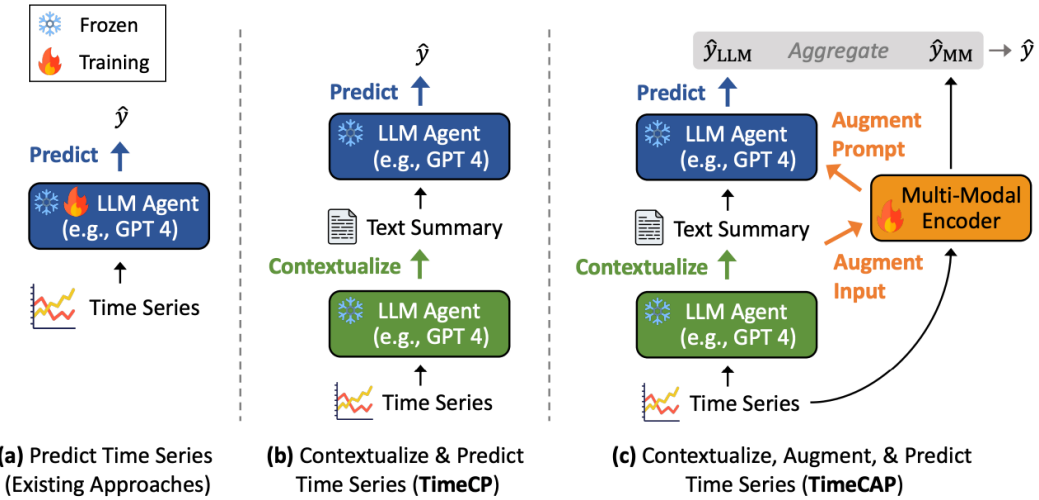


(b) Sample-wise Similarity

시계열 자체만 보면 샘플들이 서로 크게 구분되지 않지만, LLM이 생성한 텍스트 요약은 서로 다른 context를 더 잘 드러냄



(a) Multi-Modal Encoder $\mathcal{E}_\phi$



(a) Predict Time Series (Existing Approaches)

(b) Contextualize & Predict Time Series (**TimeCP**)

(c) Contextualize, Augment, & Predict Time Series (**TimeCAP**)

- Dual LLM Agent Architecture (Contextualize → Predict)
  - 첫 번째 LLM agent가 시계열을 도메인 지식 기반 텍스트 요약 s_x 으로 변환하고, 두 번째 agent (AP)가 s_x를 입력으로 이벤트를 예측하는 2-stage 구조 (zero-shot 기반)
- Multi-Modal Encoder 기반 Dual Augmentation
  - Trainable encoder가 시계열 + 텍스트를 공동 임베딩하여
    (1) 입력 증강(input augmentation) 수행 및
    (2) embedding 기반 k-NN으로 in-context example을 선택해 prompt augmentation 수행
- LLM Prediction + Encoder Prediction Fusion

Results: Interpretability

Text Summary (June 26th, 2017) | **Answer: Rain**

Over the past 24 hours, New York City has experienced a gradual decrease in temperature followed by a moderate increase, indicating a typical diurnal variation with cooler conditions overnight and warmer conditions during the day. Humidity levels showed a general upward trend throughout the night, reaching a peak before starting to decrease, which could be indicative of a passing weather front. Air pressure remained relatively stable with only slight fluctuations, suggesting a steady state of the atmosphere without any significant weather systems moving in or out of the area. Wind speeds varied, with some calmer periods and occasional gusts, but no sustained high winds were observed, which would have been noteworthy for weather impacts. Wind direction shifts were not consistent, indicating the presence of variable winds rather than a dominant weather pattern influencing the direction over the period.

Implicit Interpretation (w/o in-context examples) | **Prediction: Not Rain**

The summary indicates stable air pressure and variable winds, without a dominant weather pattern that would typically precede a significant rain event. The increase in humidity could suggest the possibility of rain, but without the presence of a significant weather system or sustained high winds to bring in moisture, the stable conditions suggest that significant precipitation is unlikely in the next 24 hours.

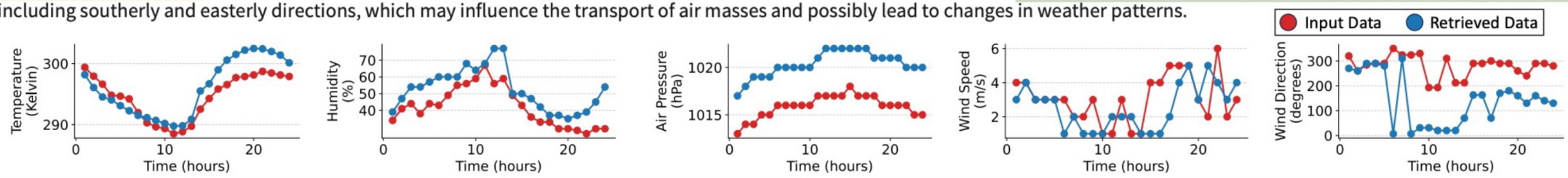
Implicit Interpretation (w/ in-context examples) | **Prediction: Rain**

The rationale for predicting rain is based on the pattern of a general upward trend in humidity, which could suggest an approaching front, and the stable air pressure with slight fluctuations, which is similar to the scenarios in Summary #1 and Summary #2, both of which resulted in rain. Although wind speeds and directions were variable and not indicative of a strong weather system, these factors combined with the humidity trend lean towards the possibility of rain.

Explicit Interpretation (September 4th, 2014) | **Prediction: Rain**

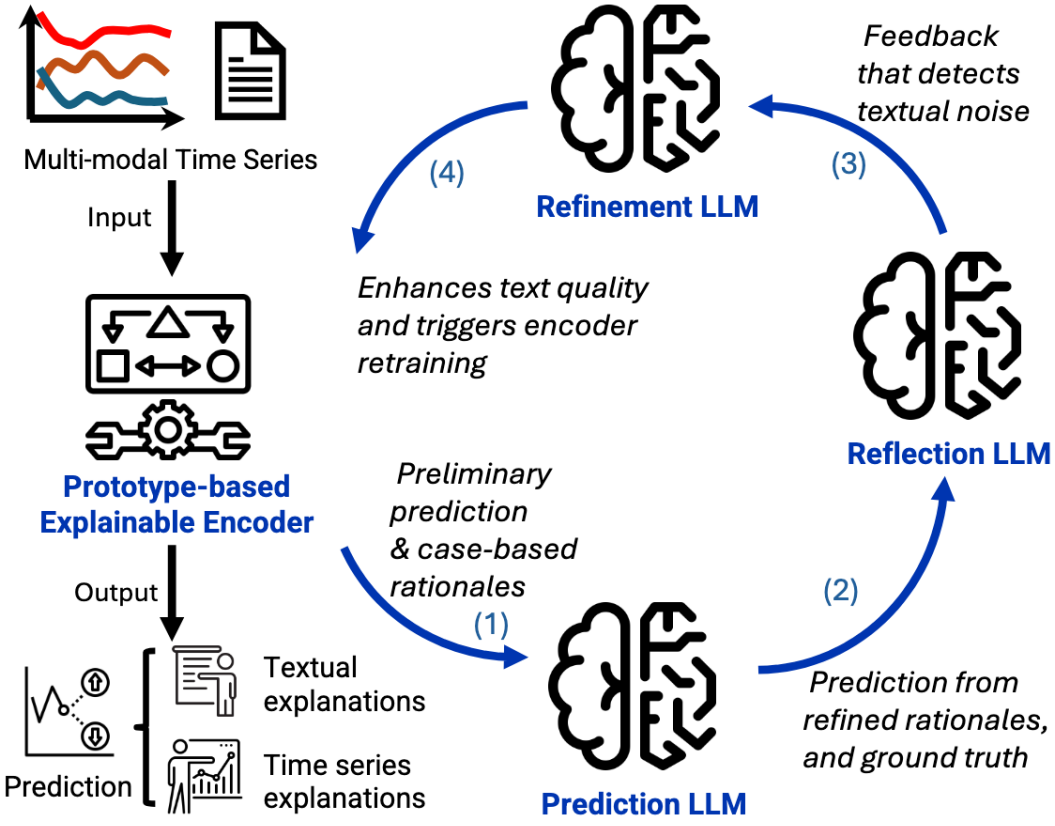
가장 관련있는 in-context sample 선택하도록 시킴

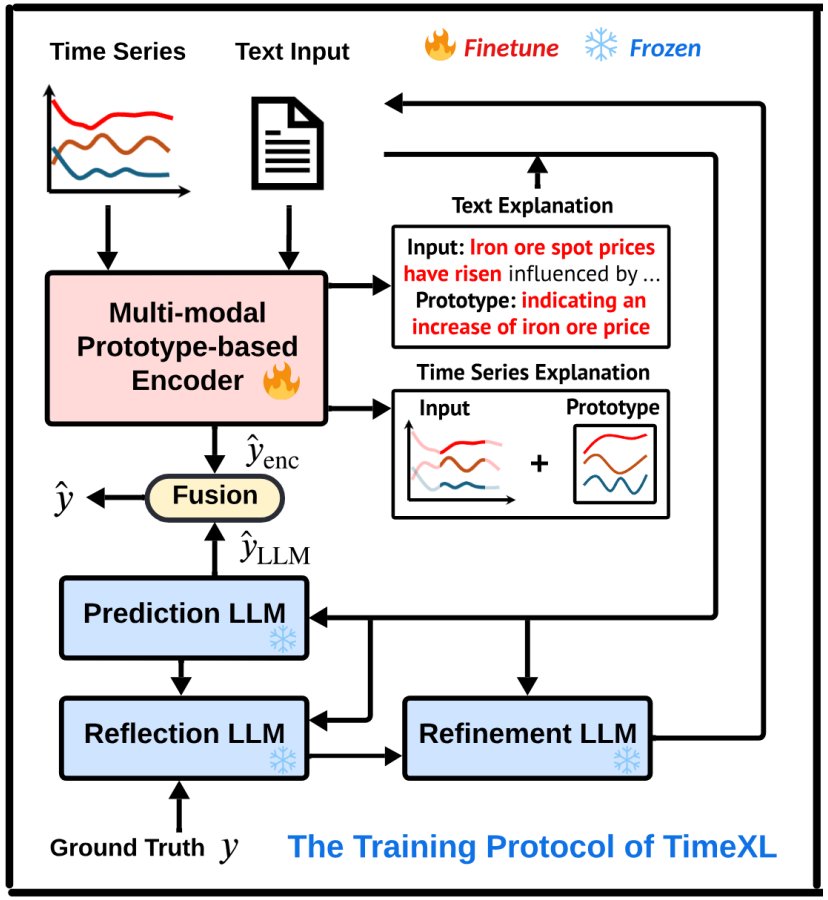
Over the past 24 hours in New York City, temperatures have experienced a notable increase, indicative of a warming trend that peaked in the late afternoon before gradually declining toward the evening. Humidity levels have shown fluctuations, starting relatively low, increasing significantly during the early hours, and then decreasing during the day, which could be associated with a front passing through the area. Air pressure remained relatively stable throughout the period, with only slight variations, suggesting a period of settled weather. Wind speeds varied modestly, with calmer conditions prevailing for most of the day before a slight increase later on, while wind direction shifted from primarily westerly to more variable, including southerly and easterly directions, which may influence the transport of air masses and possibly lead to changes in weather patterns.



(Alignment – Output-level – LLM Reasoning) TimeXL (NeurIPS 2025)

- Prototype 기반 멀티모달 시계열 인코더와 LLM의 prediction-reflection-refinement 루프를 결합해 성능과 설명 가능성을 동시에 향상시키는 프레임워크

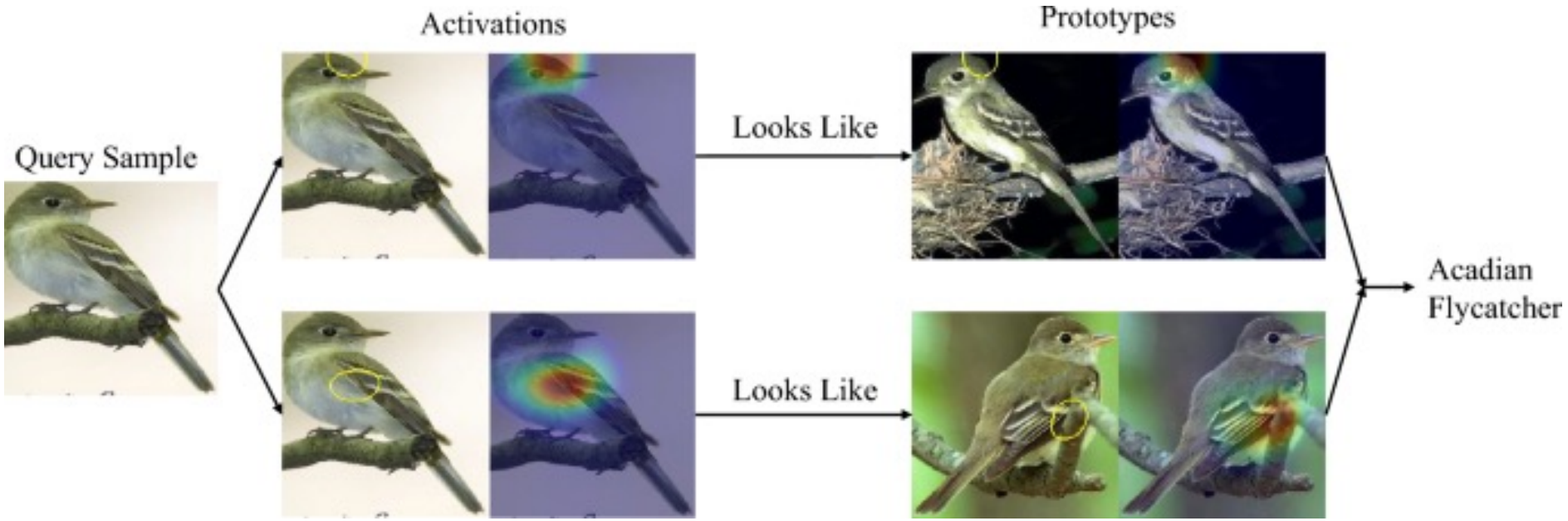




(NeurIPS 2025) TimeXL: Explainable Multi-modal Time Series Prediction with LLM-in-the-Loop

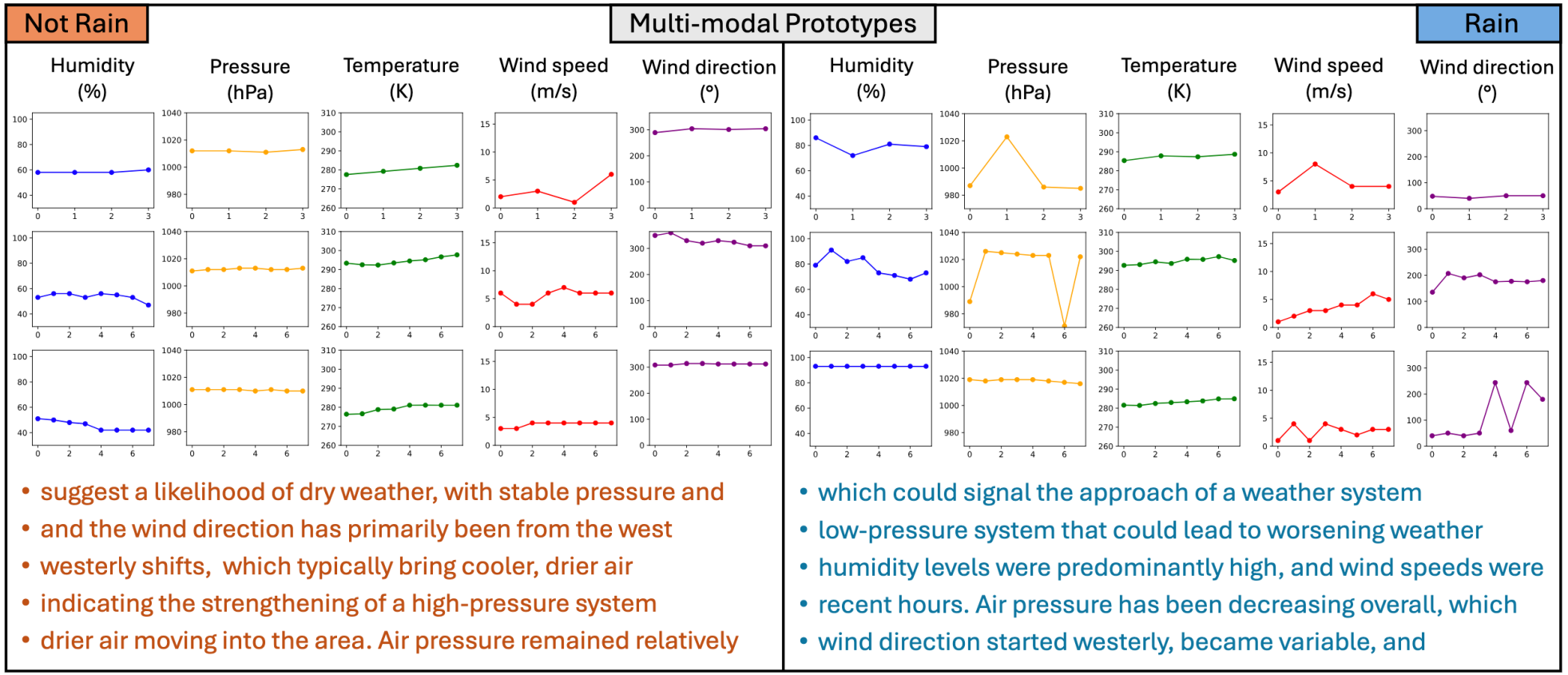
Prototype-based Explanations

- 일반 CNN: feature → fully connected → softmax (내부 reasoning이 보이지 않음)
- Prototype 모델: Similarity(query patch, prototype) 를 직접 계산해서 "이 부분이 저 클래스의 전형적인 모습과 닮았다" 라고 설명 가능



"이 새는 ○○ 클래스다. 왜냐하면 이 이미지의 특정 부분이 해당 클래스의 프로토타입과 비슷하기 때문이다."

Results: Prototypes



Rain/Not Rain 클래스별로 전형적인 시계열 패턴과 텍스트 표현을 프로토타입 형태로 학습하여, 예측 근거를 인간이 이해 가능한 multi-modal case-based explanation으로 제시함

Results: Refinement

Original Text Reasoning | Truth: Rain | Prediction: Not rain

Original Text: Over the past 24 hours, New York City experienced a gradual increase in temperature throughout the day, peaking in the late afternoon before a modest decline towards the evening. Humidity levels started high in the early hours, dropped significantly during the day, and then showed a rising trend at the end of the period, indicating a possible increase in moisture content in the air. Air pressure remained relatively stable with only minor fluctuations, suggesting a period of stable weather conditions. Wind speeds varied, with a noticeable increase during the late afternoon, which could have contributed to a brief period of cooler and more turbulent conditions. Wind direction was predominantly from the south-southwest, shifting slightly to a more south-southeast orientation later in the day, which is typical for the region's weather patterns during this time.

Prototypes:

Not rain: City has experienced relatively stable temperatures with a slight warming trend observed (0.78)

Not rain: peak during the late afternoon before beginning to decrease slightly into the (0.64)

Not rain: dropping again, indicative of typical diurnal variation (0.51)

TimeXL: Text Reasoning | Prediction: Rain

Refined Text: Over the past 24 hours, New York City experienced a stable air pressure pattern with minor fluctuations, indicating stable weather conditions. The day saw a gradual increase in temperature, peaking in the late afternoon before declining in the evening. Humidity levels were high early on, dropped significantly during the day, and rose again later, suggesting increased moisture content. Wind direction shifted from south - southwest to south - southeast, bringing moisture-laden air, which could increase the likelihood of rain.

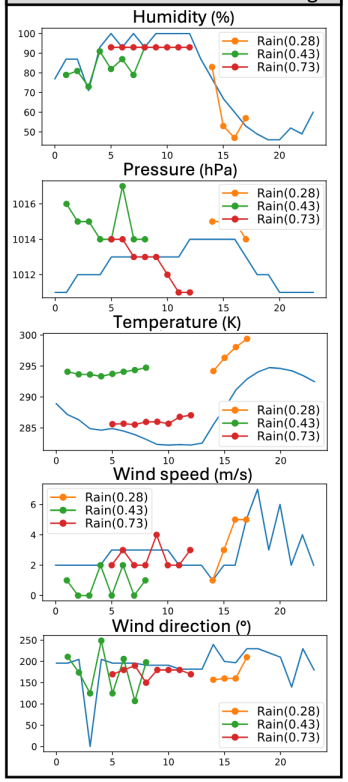
Prototypes:

Rain: direction was variable without a consistent pattern. These indicators suggest (0.47)

Rain: wind direction started westerly, became variable, and (0.64)

Rain: which could signal the approach of a weather system (0.53)

TimeXL: Time Series Reasoning



- Text refinement를 통해 예측이 Not Rain → Rain으로 교정되며, 프로토타입 매칭이 온도 안정성 중심(비강수)에서 습도·바람 변화·기압 하강(강수 신호) 중심으로 전환
- 시계열 프로토타입과 텍스트 프로토타입이 일관되게 강수 패턴을 support하며, Multi-modal case-based reasoning이 예측 정확도와 설명 신뢰도를 동시에 강화

Summary

- 1. Big Picture
  - 시계열은 더 이상 단일 모달 문제가 아니다. (multi-modal view vs. multi-modal data)
  - 현실 예측 문제는 항상 Context + Multi-modality + Heterogeneity를 포함한다.
  - 핵심은 "더 큰 모델"이 아니라 어떻게 정렬(Alignment)하느냐이다.
    - Data alignment
    - Multi-modal alignment

- 2. Two Design Axes
  - (1) Fusion Level
    - Intermediate-level: Patch Reprogramming / Prompt tuning
    - Output-level: Gating
  - (2) Alignment Strategy
    - Intermediate-level: Self / Cross-attention / GNN (Relational alignment) / Contrastive learning
    - Output-level: Retrieval / LLM reasoning

Reference

- Papers
  - (AAAI 2024) GPT4MTS: Prompt-Based Large Language Model for Multimodal Time Series Forecasting
  - (ICLR 2024) Time-LLM: Time Series Forecasting by Reprogramming Large Language Models
  - (NeurIPS 2024) Time-MMD: Multi-Domain Multimodal Dataset for Time Series Analysis
  - (NeurIPS 2025) Multi-Modal View Enhanced Large Vision Models for Long-Term Time Series Forecasting
  - (NeurIPS 2024) Are Language Models Actually Useful for Time Series Forecasting?
  - (WWW 2024) UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting
  - (arxiv 2025) Context Matters: Leveraging Contextual Features for Time Series Forecasting
  - (NeurIPS 2024) TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables
  - (AAAI 2025) TimeCMA: Towards LLM-Empowered Multivariate Time Series Forecasting via Cross-Modality Alignment
  - (EMNLP 2020) Deep Attentive Learning for Stock Movement Prediction from Social Media Text and Company Correlations
  - (Under Review) Beyond Time Series: Integrating Categorical Context for Robust Time Series Classification
  - (AAAI 2025) TimeCAP: Learning to Contextualize, Augment, and Predict Time Series Events with Large Language Model Agents
  - (NeurIPS 2025) TimeXL: Explainable Multi-modal Time Series Prediction with LLM-in-the-Loop

- (KDD 2025 Tutorial) Multi-modal Time Series Analysis — Methods, Datasets, and Applications
  - Link: https://arxiv.org/pdf/2503.13709