TRACE: Intra-visit Clinical Event Nowcasting via Effective Patient Trajectory Encoding

Yuyang Liang
The Chinese University of Hong
Kong, Shenzhen
Shenzhen, Guangdong, China
yuyangliang@link.cuhk.edu.cn

Yankai Chen
Cornell University
Ithaca, New York, Unite States
yankaichen@acm.org

Yixiang Fang
The Chinese University of Hong
Kong, Shenzhen
Shenzhen, Guangdong, China
fangyixiang@cuhk.edu.cn

Laks V. S. Lakshmanan
University of British Columbia
Vancouver, British Columbia, Canada
laks@cs.ubc.ca

Chenhao Ma*
The Chinese University of Hong
Kong, Shenzhen
Shenzhen, Guangdong, China
machenhao@cuhk.edu.cn

Abstract

Electronic Health Records (EHR) have become a valuable resource for a wide range of predictive tasks in healthcare. However, existing approaches have largely focused on *inter-visit* event predictions, overlooking the importance of *intra-visit* nowcasting, which provides prompt clinical insights during an ongoing patient visit. To address this gap, we introduce the task of laboratory measurement prediction within a hospital visit. We study the laboratory data that, however, remained underexplored in previous work. We propose TRACE, a Transformer-based model designed for clinical event nowcasting by encoding patient trajectories. TRACE effectively handles long sequences and captures temporal dependencies through a novel timestamp embedding that integrates decay properties and periodic patterns of data. Additionally, we introduce a smoothed mask for denoising, improving the robustness of the model. Experiments on two large-scale electronic health record datasets demonstrate that the proposed model significantly outperforms previous methods, highlighting its potential for improving patient care through more accurate laboratory measurement nowcasting. The code is available at https://github.com/Amehi/TRACE.

CCS Concepts

• **Computing methodologies** → *Knowledge representation and reasoning*.

Keywords

Electronic Health Record (EHR), Laboratory Measurement Prediction, Transformer Model, Time-aware Attention Mechanism

---

*Chenhao Ma is the corresponding author.

1 Introduction

Electronic Health Records (EHR) provide comprehensive digital records of patients' medical histories, including diagnoses, medications, procedures, and laboratory results. With the growing complexity of this data, advanced predictive models have emerged to enhance patient outcomes through accurate medical event nowcasting, early risk detection, and personalized treatment, optimizing healthcare delivery.

Traditional EHR predictive methods mainly follow conventional sequential modeling [4, 5, 30], to focus on *inter-visit* predictions, nowcasting events after the current visit, such as readmissions [22] or future diagnoses [9, 10, 25]. While valuable, these models overlook *intra-visit* predictions, which aim to provide real-time insights into a patient's condition during a single visit. Such tasks may offer more valuable insights into current patient health statuses than merely nowcasting changes between separate visits, as medical information for different diagnoses in multiple visits may not be strongly correlated. To illustrate this issue, we investigate the MIMIC-IV dataset, with our observations summarized in Figure 1. Figure 1(a) shows the Jaccard similarity of diagnosis codes for patients with multiple visits, revealing low similarity scores and relatively weak correlations between visits. Additionally, visit-based models exclude single-visit patient data, as shown in Figure 1(b), leading to significant information loss for over half of the patients in the MIMIC-IV dataset. This loss may impair model performance for newly admitted patients, limiting its real-world effectiveness.

To address these issues and provide timely healthcare, we consider the *intra-visit clinic event nowcasting*. Unlike prior works on inter-visit EHR data, we explore *laboratory measurements in EHR*, such as blood tests and metabolic panels, which provide vital insights into a patient's current clinical condition. Despite their importance, laboratory data is underutilized in existing models. In this

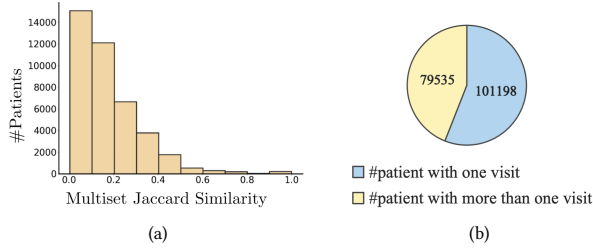Yuyang Liang, Yankai Chen, Yixiang Fang, Laks V. S. Lakshmanan & Chenhao Ma

Figure 1: Investigation on MIMIC-IV dataset. (a) Jaccard similarity of diagnosis codes between different visits of patients. (b) The proportion of patients with only one visit and patients with multiple visits in the MIMIC-IV dataset.

work, we propose the laboratory measurement prediction within a hospital visit, focusing on nowcasting a patient's clinical state in the next time window based on observed medical events. The detailed task descriptions are provided in Section 2. Specifically, We propose a patient **TR**ajectory encoding **A**pproach for **C**linal **E**vent nowcasting (TRACE), a novel Transformer-based model tailored for *intra-visit* EHR data analysis. To provide time-sensitive patient trajectory encoding, TRACE incorporates a customized timestamp embedding that captures decay and periodicity in medical events. This design allows the model to effectively account for the influence of past events while incorporating the impact of temporal cycles on laboratory test results. TRACE leverages a multi-layer Transformer architecture [29], which has been widely demonstrated to be effective across various applications [3, 6, 7, 19]. TRACE enables it to capture complex temporal dependencies and hierarchical relationships among medical events for more accurate laboratory measurement predictions. Additionally, we introduce a smoothed mask denoising module that enhances robustness by filtering out less influential information during attention calculation, enabling fine-grained denoising while maintaining differentiability for efficient backpropagation and optimization. Extensive experiments conducted showing TRACE's superior performance, achieving the highest PR-AUC and Precision@k scores on both MIMIC-III and MIMIC-IV datasets.

2 Task Descriptions

Patient Visit Trajectory. A visit trajectory in EHR for a patient can be represented as a time-ordered sequence of medical events. Formally, a trajectory for patient p during visit v is represented as:

$$\mathcal{T}_{p,v} = \{(e_1, t_1), (e_2, t_2), \ldots, (e_n, t_n)\},$$

where e_i denotes the medical code of i-th medical event, and t_i is the corresponding timestamp. Each event e_i can encompass various types of medical information such as diagnoses, procedures, medications, or laboratory results. The sequence is ordered by time, i.e., $t_1 \leq t_2 \leq \cdots \leq t_n$, reflecting the chronological order in which the events occurred during the patient's visit.

Medical Event. For medication and procedure records, e_i refers to a specific medical concept, such as a drug (e.g., aspirin) or a procedure (e.g., MRI). Laboratory measurements are more complex, with $e_i = (a_i, f_i)$, where a_i is the medical code of test subject and f_i is a flag indicating the result. In MIMIC-III, f_i is binary ($f_i = 0$ for normal, $f_i = 1$ for abnormal), while in MIMIC-IV, f_i
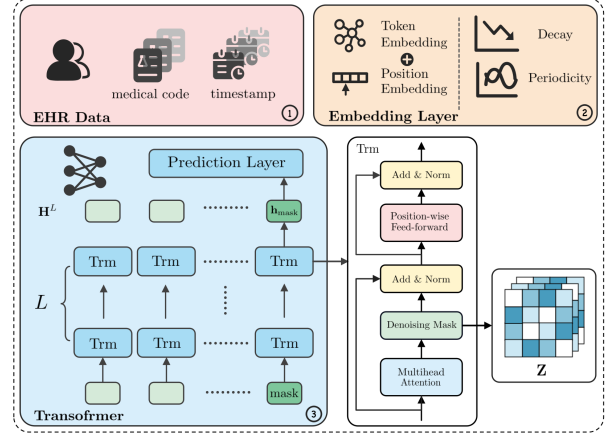

Figure 2: The TRACE model workflow.

provides more details, indicating whether the result is higher, lower, or normal relative to reference bounds. Moreover, Consecutive test records often share the same timestamp t_i, reflecting scenarios where multiple tests are conducted simultaneously, such as after a single blood draw. We use the last few records with the same timestamp as the laboratory events to predict.

Laboratory Measurement Prediction. Given a sequence of past events in a visit trajectory $\mathcal{T}_{p,v} = \{(e_1, t_1), (e_2, t_2), \ldots, (e_k, t_k)\}$ up to time t_k, the objective is to forecast a set of upcoming laboratory test events $\mathcal{T}'_{p,v} = \{(e_{k+1}, t_{k+1}), \ldots, (e_{k+m}, t_{k+m})\}$ for $t_{k+1} > t_k$ and $t_{k+1} = \cdots = t_{k+m}$.

3 TRACE

Figure 2 provides a comprehensive overview of our model design. The model begins with an embedding layer that encodes patient trajectories and timestamps into dense representations. These embeddings are processed through stacked Transformer layers to capture temporal and relational dependencies, with the final output passed to a prediction layer for nowcasting laboratory test results.

3.1 Medical Event Encoder

Following standard practices for using Transformers with sequential data, we impose a restriction N on the maximum length of the input trajectory. Given an input patient trajectory $\mathcal{T}_{p,v} = \{(e_i, t_i)\}^N$, we first calculate the hidden representation $\mathbf{h}_i$ for each medical event (e_i, t_i).

Medical Code Embedding. Each code e_i is mapped to a dense vector representation $\mathbf{s}_i$ using a learnable embedding matrix $\mathbf{E}_f \in \mathbb{R}^{|E| \times d}$, where $|E|$ is the total number of medical codes and d is the embedding dimension. In addition, in order to capture the order information of events, we inject a learnable positional encoding $\mathbf{p}_i \in \mathbb{R}^{1 \times d}$, which provides location information and helps the model understand the relative temporal order of events, to the code embedding. Finally, we obtain the medical code embedding of e_i as:

$$\mathbf{c}_i = \mathbf{s}_i + \mathbf{p}_i. \tag{1}$$

Timestamp Embedding. First, we model the decay over time based on the timestamp t_i with a square operation, as more recent

Table 1: Experiment result for laboratory measurement prediction task. The threshold is set to 0.5 for F1-score.

Datasets	Metric	Transformer	RETAIN	Dipole	HiTANET	ConCare	GRASP	PPN	TRACE
MIMIC-III	F1-score	0.5705 ± 0.0196	0.7439 ± 0.0047	0.5795 ± 0.0020	0.5669 ± 0.0049	0.5884 ± 0.0247	0.5734 ± 0.0096	0.6441 ± 0.0167	**0.7858** ± 0.0055
	PR-AUC	0.3468 ± 0.0006	0.4993 ± 0.0013	0.2745 ± 0.0081	0.5350 ± 0.0025	0.3931 ± 0.0117	0.2546 ± 0.0414	0.4478 ± 0.0044	**0.6476** ± 0.0042
	Precision@5	0.4663 ± 0.0009	0.5469 ± 0.0018	0.4374 ± 0.0136	0.5800 ± 0.0034	0.4888 ± 0.0091	0.3263 ± 0.0250	0.5215 ± 0.0011	**0.6075** ± 0.0045
	NDCG@5	0.4948 ± 0.0068	0.6086 ± 0.0021	0.4664 ± 0.0142	0.6494 ± 0.0045	0.5336 ± 0.0164	0.3426 ± 0.0346	0.5818 ± 0.0017	**0.6872** ± 0.0067
MIMIC-IV	F1-score	0.6570 ± 0.0354	0.7526 ± 0.1452	0.6276 ± 0.0020	0.5473 ± 0.0010	0.7539 ± 0.0061	0.6541 ± 0.0552	0.8000 ± 0.0071	**0.8535** ± 0.0062
	PR-AUC	0.4969 ± 0.0004	0.6893 ± 0.0027	0.3962 ± 0.0227	0.6720 ± 0.0028	0.5103 ± 0.0095	0.4400 ± 0.0772	0.6394 ± 0.0119	**0.7777** ± 0.0021
	Precision@5	0.6391 ± 0.0001	0.7252 ± 0.0021	0.5139 ± 0.0753	0.7531 ± 0.0032	0.6395 ± 0.0018	0.4872 ± 0.0303	0.7099 ± 0.0062	**0.7566** ± 0.0023
	NDCG@5	0.6413 ± 0.0001	0.7462 ± 0.0023	0.5185 ± 0.0766	0.7820 ± 0.0043	0.6446 ± 0.0074	0.4973 ± 0.0367	0.7284 ± 0.0067	**0.7960** ± 0.0031

events are more influential for future predictions. The decay is represented by:

$$\mathbf{f}_i^{decay} = \mathbf{W}_d \left(1 - \tanh\left((\mathbf{W}_t t_i - \mathbf{b}_t)^2 \right) \right) - \mathbf{b}_d, \quad (2)$$

where $\mathbf{W}_d \in \mathbb{R}^{d \times h}$ and $\mathbf{W}_t \in \mathbb{R}^{h \times 1}$ are learnable weighted matrix, $\mathbf{b}_d \in \mathbb{R}^d$ and $\mathbf{b}_t \in \mathbb{R}^h$ are bias vectors. Next, to capture the periodic patterns in medical treatments, such as regular laboratory tests or medications with daily cycles, we apply sine and cosine transformations to t_i, leveraging their cyclic nature. This periodicity is represented by:

$$\mathbf{f}_i^{periodic} = \mathbf{W}_p \cdot \begin{bmatrix} \sin\left(2\pi \frac{t_i}{\omega}\right) \\ \cos\left(2\pi \frac{t_i}{\omega}\right) \end{bmatrix} + \mathbf{b}_p, \quad (3)$$

where $\mathbf{W}_p \in \mathbb{R}^{d \times 2}$ is a learnable weight matrix, $\mathbf{b}_p \in \mathbb{R}^d$ is the corresponding bias term, and ω represents the periodic time (e.g., for daily periodicity, $\omega = 24$). The decay and periodic components are combined via element-wise addition to form the final timestamp embedding:

$$\mathbf{f}_i = \mathbf{f}_i^{decay} + \mathbf{f}_i^{periodic}. \quad (4)$$

By adding the medical code embedding and the timestamp embedding, we obtain the hidden representation of the input trajectory:

$$\mathbf{h}_i = \mathbf{c}_i + \mathbf{f}_i. \quad (5)$$

3.2 Time-aware Transformer

Transformer layers iteratively calculate the hidden representation $\mathbf{h}_i^l$ for the i-th event at layer l, effectively encoding the patient's trajectory to capture temporal and contextual dependencies. We first employ a multi-head self-attention mechanism to capture the complex relationships between medical events. Let h denote the number of heads and $\mathbf{H}^l \in \mathbb{R}^{n \times d}$ denote the matrix of input embeddings obtained by stacking the input sequence of $\mathbf{h}_i^l$. The multi-head attention is computed as:

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h \right) \mathbf{W}^O, \quad (6)$$

$$\text{head}_i = \text{softmax}\left(\frac{Q^l K^{l^T} \odot Z^l}{\sqrt{d/h}} \right) V^l, \quad (7)$$

$$Q^l = \mathbf{H}^l \mathbf{W}_i^Q, K^l = \mathbf{H}^l \mathbf{W}_i^K, V^l = \mathbf{H}^l \mathbf{W}_i^V. \quad (8)$$

Here, $\mathbf{W}_i^Q \in \mathbb{R}^{d \times d/h}, \mathbf{W}_i^K \in \mathbb{R}^{d \times d/h}$, and $\mathbf{W}_i^V \in \mathbb{R}^{d \times d/h}$ are projection matrices for each head. $\mathbf{W}^O \in \mathbb{R}^{d \times d}$ is the aggregation matrix. Z^l is the denoising matrix. Sequential medical data has temporal irregularities and varying event importance. Additionally, noise like

inconsistent time gaps or redundant records can affect performance. To address this, we apply a smoothed mask to denoise time-aware attention. For each layer l, we define a mask matrix $\mathbf{Z}^l \in \mathbb{R}^{N \times N}$, where N is the length of input trajectory. $\mathbf{Z}^l$ contains logits that represent the relative importance between each query and key. Unlike binary masks [1, 2], we use a smoothed mask matrix for more gradual modifications and easier optimization via gradient descent to improve denoising. $\mathbf{Z}^l$ is randomly initiated with high positive values, and a loss function penalizes excessive logits in $\mathbf{Z}^l$ to ensure effective denoising. Formally, the denoising loss is defined by:

$$\mathcal{L}_{\text{denoise}} = \sum_{l=1}^{L} \|\mathbf{Z}^l\|_2. \quad (9)$$

Following the attention mechanism, a position-wise feed-forward network (FFN) introduces non-linearity and transforms input features to capture dependencies beyond attention. Each sublayer (Figure 2) includes residual connections and layer normalization to enhance performance. We stack such Transformer layers to capture complex patterns and hierarchical representations.

3.3 Masked Representation Prediction

We adopt a masked representation prediction mechanism by appending a mask token $(e_{\text{mask}}, t_{\text{mask}})$, at the end of the sequence as a placeholder for the patient's future state (Figure 2). The model predicts the next clinical representation using the hidden representation of the mask event $\mathbf{h}_{\text{mask}}^L$, processed through a linear layer with *sigmoid* activation. The predicted probabilities $\hat{\mathbf{y}}$ are compared with the ground truth $\mathbf{y}$ using cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \left(\mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1 - \mathbf{y}_i) \log(1 - \hat{\mathbf{y}}_i) \right). \quad (10)$$

Combined with the denoising module's loss, the final objective is:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{denoise}}. \quad (11)$$

4 Experiment

4.1 Experiment Setup

Datasets. We evaluate our model on two real-world EHR datasets MIMIC-III [15] and MIMIC-IV [14]. They are randomly split into training, validation, and test sets, with a ratio of 0.75: 0.10: 0.15.

Baselines. We compare our model with the listed methods **(1)** Vanilla Transformer [29], **(2)** RETAIN [8], **(3)** Dipole [24], **(4)** HiTANET [23], **(5)** ConCare [26], **(6)** GRASP [33], and **(7)** PPN [32].

Table 2: Average count of model predictions. For simplicity, we selected a subset of baselines with relatively better performance. The threshold is set to 0.5.

Model	MIMIC-III	MIMIC-IV
HiTANET	11670.86 ± 483.92	2118.32 ± 1223.93
RETAIN	3098.37 ± 179.53	2718.27 ± 601.42
PPN	3.95 ± 0.68	14.49 ± 0.52
TRACE	25.44 ± 1.29	20.49 ± 1.06
Ground Truth	12.17 ± 8.71	18.21 ± 9.97

Table 3: Ablation study results on MIMIC-III and MIMIC-IV.

Model	MIMIC-III		MIMIC-IV	
	PR-AUC	Precision@5	PR-AUC	Precision@5
w/o D	0.54(↓ 17.02%)	0.55(↓ 08.67%)	0.76(↓ 01.81%)	0.75(↓ 01.47%)
w/o P	0.63(↓ 03.23%)	0.59(↓ 02.37%)	0.77(↓ 00.84%)	0.75(↓ 00.38%)
w/o DP	0.48(↓ 25.98%)	0.53(↓ 12.61%)	0.73(↓ 06.26%)	0.74(↓ 02.58%)
w/o DPM	0.47(↓ 26.81%)	0.53 (↓ 13.31%)	0.72(↓ 06.69%)	0.73(↓ 02.96%)
TRACE	**0.65**	**0.61**	**0.78**	**0.76**

Evaluation Protocols and Metrics. We evaluate our model using two protocols with distinct metrics. For multi-label classification, we use *PR-AUC* and *F1-score* to measure the model's performance on imbalanced datasets. For ranking-based evaluation, we apply *Precision@k* and *NDCG@k* to assess its ability to rank relevant items at top positions. In *NDCG@k*, ground truth events with identical timestamps are assigned the same rank.

4.2 Experimental Results

Table 1 presents the experimental results, comparing TRACE with baseline models across two datasets and evaluating performance on various metrics to highlight strengths and limitations. Our analysis reveals key patterns and insights into how different models handle the newly defined laboratory measurement prediction task. Our analyses are threefold:

(a) The results on MIMIC-IV are generally better than on MIMIC-III, due to its richer, more comprehensive data, including medications and procedures. Unlike MIMIC-III, which focuses mainly on laboratory measurements, MIMIC-IV offers more detailed information including flags for test results, helping the model better understand the context and improve prediction accuracy.

(b) Transformer-based models outperform RNN-based models in handling long sequences with new input structures. HiTANET, leveraging the Transformer architecture, achieves the highest Precision, NDCG, and AUC-PR on MIMIC-III by incorporating temporal information. Baseline models originally designed for *inter-visit* predictions struggle with handling event-level data. RETAIN's linear embedding layer enables it to effectively model event sequences, outperforming those models.

(c) TRACE excelled across all evaluation metrics, outperforming other models on both MIMIC-III and MIMIC-IV. With Transformer architecture and a strong embedding strategy, TRACE adapts effectively to intra-visit prediction tasks and achieves outstanding predictive accuracy. For the multi-label classification task, TRACE also outperformed baseline models in predicting the correct number
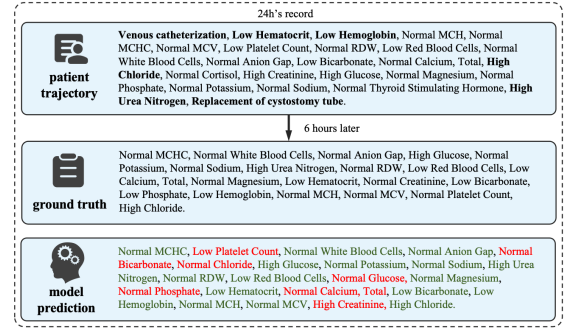

Figure 3: A sample case of a patient with kidney dysfunction. For simplicity, the ordered patient trajectory only contains several important medical events. The wrong predictions are colored with red and the correct ones are colored with green.

of laboratory measurements. As shown in Table 2, while baselines often over-or underestimated the number of measurements, TRACE aligned more precisely with the actual number of events.

4.3 Abalation Study

To evaluate key components of TRACE, we conducted an ablation study focusing on the decay feature (D), periodicity feature (P) of the timestamp embedding, and the denoising module (M). Experiment results are shown in Table 3.

Without D/P. The results show a significant performance decline when temporal features are excluded. Removing the decay feature led to a 17.02% drop in PR-AUC and an 8.67% drop in Precision@5 on MIMIC-III. Excluding the periodicity feature (w/o P) caused a 3.23% decrease in PR-AUC on MIMIC-III and 0.84% on MIMIC-IV, highlighting both features' importance.

Without DP. When both decay and periodicity features were removed, PR-AUC dropped by 25.98% and Precision@5 by 12.61% on MIMIC-III, confirming their complementary roles in capturing temporal dynamics. Smaller but notable declines on MIMIC-IV further underscore their importance across datasets.

Without DPM. The masking mechanism showed a notable impact. TRACE w/o DPM exhibited the worst performance across all metrics on both datasets. This highlights the essential role of the denoising module in filtering out irrelevant or noisy information, thereby enhancing the model's robustness and predictive accuracy.

4.4 Case Study

In this case study, we analyze a hospitalized patient's trajectory and demonstrate how TRACE predicts critical laboratory measurements, aiding clinical decision-making. Figure 3 depicts a sample trajectory and TRACE's predictions.

Patient Background. The hospitalized patient underwent procedures like *venous catheterization* and *cystostomy tube replacement*, indicating fluid management and urinary tract care. Key laboratory tests showed elevated *urea nitrogen* and *chloride*, suggesting renal impairment and metabolic acidosis, along with low *hematocrit*, *hemoglobin*, and *platelets*, indicating anemia.

Model Prediction. TRACE accurately predicts a range of laboratory values from the patient's prior trajectory, aligning well with

the ground truth. It correctly identified a normal MCHC, low RBC, and high urea nitrogen, but struggles with values like glucose and creatinine, likely due to limited 24-hour historical data. Despite discrepancies, TRACE helps clinicians prioritize management and guide treatment decisions.

5 Conclusion and Future Work

In conclusion, this work defines the novel task of laboratory measurement prediction within a hospital visit. We propose TRACE, which successfully captures intricate temporal dependencies, enabling fine-grained predictions that reflect the dynamics of patient data. Our findings emphasize the value of *intra-visit* information and demonstrate how processing long sequences enhances clinical decision-making accuracy and timeliness. Future work will focus on two directions: (1) Leveraging the external knowledge [16, 17] and/or Large Language Models with particular safety and privacy considerations [18, 20, 21] for knowledge adaptation and EHR data parsing [27]. (2) Enhancing the model by integrating advanced temporal learning and trajectory analysis techniques to further improve its accuracy, robustness, and continual learning capability [11–13, 28, 31, 34].

Acknowledgments

References

[1] Jasmijn Bastings, Wilker Aziz, and Ivan Titov. 2019. Interpretable neural predictions with differentiable binary variables. *arXiv preprint arXiv:1905.08160.*

[2] Huiyuan Chen, Yusan Lin, Menghai Pan, Lan Wang, Chin-Chia Michael Yeh, Xiaoting Li, Yan Zheng, Fei Wang, and Hao Yang. 2022. Denoising Self-Attentive Sequential Recommendation (*RecSys '22*).

[3] Yankai Chen, Yixiang Fang, Qiongyan Wang, Xin Cao, and Irwin King. 2024. Deep Structural Knowledge Exploitation and Synergy for Estimating Node Importance Value on Heterogeneous Information Networks. In *AAAI*, Vol. 38. 8302–8310.

[4] Yankai Chen, Quoc-Tuan Truong, Xin Shen, Jin Li, and Irwin King. 2024. Shopping Trajectory Representation Learning with Pre-training for E-commerce Customer Understanding and Recommendation. *SIGKDD* (2024).

[5] Yankai Chen, Quoc-Tuan Truong, Xin Shen, Ming Wang, Jin Li, Jim Chan, and Irwin King. 2023. Topological Representation Learning for E-commerce Shopping Behaviors. *MLG-KDD*.

[6] Yankai Chen, Yaozu Wu, Shicheng Ma, and Irwin King. 2020. A literature review of recent graph embedding techniques for biomedical data. In *ICONIP*. Springer, 21–29.

[7] Yankai Chen, Yifei Zhang, Huifeng Guo, Ruiming Tang, and Irwin King. 2022. An Effective Post-training Embedding Binarization Approach for Fast Online Top-K Passage Matching. In *AACL*. 102–108.

[8] Edward Choi, Mohammad Taha Bahadori, Joshua A. Kulas, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. [n. d.]. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism (*NIPS'16*).

[9] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-based Attention Model for Healthcare Representation Learning. In *KDD* (*KDD '17*).

[10] Jingyue Gao, Xiting Wang, Yasha Wang, Zhao Yang, Junyi Gao, Jiangtao Wang, Wen Tang, and Xing Xie. 2019. CAMP: Co-Attention Memory Networks for Diagnosis Prediction in Healthcare. In *ICDM*.

[11] Xiaolin Han, Reynold Cheng, Chenhao Ma, and Tobias Grubenmann. 2022. DeepTEA: Effective and efficient online time-dependent trajectory outlier detection. *Proceedings of the VLDB Endowment* 15, 7 (2022), 1493–1505.

[12] Xiaolin Han, Tobias Grubenmann, Reynold Cheng, Sze Chun Wong, Xiaodong Li, and Wenya Sun. 2020. Traffic incident detection: A trajectory-based approach. In *ICDE*. IEEE, 1866–1869.

[13] Xiaolin Han, Tobias Grubenmann, Chenhao Ma, Xiaodong Li, Wenya Sun, Sze Chun Wong, Xuequn Shang, and Reynold Cheng. 2024. FDM: Effective and efficient incident detection on sparse trajectory data. *Information Systems* (2024), 102418.

[14] Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Benjamin Moody, Brian Gow, Li wei H. Lehman, Leo Anthony Celi, and Roger G. Mark. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* (2023).

[15] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad Mahdi Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Sci Data* (2016).

[16] Muzhi Li, Cehao Yang, Chengjin Xu, Xuhui Jiang, Yiyan Qi, Jian Guo, Ho fung Leung, and Irwin King. 2024. Retrieval, Reasoning, Re-ranking: A Context-Enriched Framework for Knowledge Graph Completion. arXiv:2411.08165

[17] Muzhi Li, Cehao Yang, Chengjin Xu, Zixing Song, Xuhui Jiang, Jian Guo, Ho fung Leung, and Irwin King. 2024. Context-aware Inductive Knowledge Graph Completion with Latent Type Constraints and Subgraph Reasoning. arXiv:2410.16803

[18] Yangning Li, Qingsong Lv, Tianyu Yu, Yinghui Li, Shulin Huang, Tingwei Lu, Xuming Hu, Wenhao JIang, Hai-Tao Zheng, and Hui Wang. 2024. UltraWiki: Ultra-fine-grained Entity Set Expansion with Negative Seed Entities. *arXiv preprint arXiv:2403.04247* (2024).

[19] Sida Lin, Zhouyi Zhang, Yankai Chen, Chenhao Ma, Yixiang Fang, Shan Dai, and Guangli Lu. 2024. Effective Job-market Mobility Prediction with Attentive Heterogeneous Knowledge Learning and Synergy. In *CIKM*. 3897–3901.

[20] Aiwei Liu, Leyi Pan, Xuming Hu, Shuang Li, Lijie Wen, Irwin King, and Philip S. Yu. 2024. An Unforgeable Publicly Verifiable Watermark for Large Language Models. In *ICLR*.

[21] Aiwei Liu, Leyi Pan, Xuming Hu, Shiao Meng, and Lijie Wen. 2024. A Semantic Invariant Robust Watermark for Large Language Models. In *ICLR*.

[22] Changshuo Liu, Wenqiao Zhang, Beng Chin Ooi, James Wei Luen Yip, Lingze Zeng, and Kaiping Zheng. 2023. Toward Cohort Intelligence: A Universal Cohort Representation Learning Framework for Electronic Health Record Analysis.

[23] Junyu Luo, Muchao Ye, Cao Xiao, and Fenglong Ma. 2020. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *SIGKDD*.

[24] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *SIGKDD'23*.

[25] Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. 2018. KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare. In *KAME (CIKM '18)*.

[26] Liantao Ma, Chaohe Zhang, Yasha Wang, Wenjie Ruan, Jiangtao Wang, Wen Tang, Xinyu Ma, Xin Gao, and Junyi Gao. 2020. Concare: Personalized clinical feature embedding via capturing the healthcare context. In *AAAI*, Vol. 34. 833–840.

[27] Zexuan Qiu, Jieming Zhu, Yankai Chen, Guohao Cai, Weiwen Liu, Zhenhua Dong, and Irwin King. 2024. EASE: Learning Lightweight Semantic Feature Adapters from Large Language Models for CTR Prediction. In *CIKM*. 4819–4827.

[28] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. 2020. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637* (2020).

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS (NIPS'17)*.

[30] Yaozu Wu, Yankai Chen, Zhishuai Yin, Weiping Ding, and Irwin King. 2023. A survey on graph embedding techniques for biomedical data: Methods and applications. *Information Fusion* 100 (2023), 101909.

[31] Dianzhi Yu, Xinni Zhang, Yankai Chen, Aiwei Liu, Yifei Zhang, Philip S Yu, and Irwin King. 2024. Recent Advances of Multimodal Continual Learning: A Comprehensive Survey. *arXiv preprint arXiv:2410.05352* (2024).

[32] Zhihao Yu, Chaohe Zhang, Yasha Wang, Wen Tang, Jiangtao Wang, and Liantao Ma. 2024. Predict and Interpret Health Risk Using Ehr Through Typical Patients. In *ICASSP*.

[33] Chaohe Zhang, Xin Gao, Liantao Ma, Yasha Wang, Jiangtao Wang, and Wen Tang. 2021. GRASP: Generic Framework for Health Status Representation Learning Based on Incorporating Knowledge from Similar Patients. *AAAI* (2021).

[34] Xinni Zhang, Yankai Chen, Chenhao Ma, Yixiang Fang, and Irwin King. 2024. Influential Exemplar Replay for Incremental Learning in Recommender Systems. In *AAAI*, Vol. 38. 9368–9376.