Research on Multimodal Fusion of Temporal Electronic Medical Records

Zhaokun Wang

Uni Heidelberg

November 10, 2024

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Introduction to Electronic Medical Records (EMRs)

- **What is EMR:**
  - Digital versions of patients' paper charts.
- **Why EMR Data Matters:**
  - Supports better clinical decision-making
- **Data types:**
  - Structured, unstructured, and temporal.

The Need for Multimodal EMR Fusion

- **Background**:
  - **EMR** systems are rapidly growing, providing rich patient data.
- **Problem**:
  - Previous research focused on **single-modality** data, underutilizing multimodal EMR data.

Challenges in Multimodal EMR Fusion

- **Data Irregularity**:
  - Irregular time intervals and missing entries complicate analysis.
- **Data Variety**:
  - Integrating **structured (tabular)** and **unstructured (textual)** data types.
- **Long-Term Dependencies**:
  - Capturing patient health trends over extended periods.

Study Objectives and Contributions

- **Goal**: Develop a robust multimodal fusion model, termed **T-MAG** (Time-series Multimodal Adaptation Gate), to combine diverse EMR data effectively.
- **Research Focus**:
  - Integrate four distinct EMR data types: **static tabular**, **static notes**, **temporal tabular**, and **temporal notes**.
  - Implement mechanisms like **attention-backtracking** to handle long-term dependencies and improve predictive capabilities.

Understanding MAG and LSTM:

- **MAG:**
  Imagine a traffic controller that directs information flow, ensuring the most important data gets through.
- **LSTM:**
  Picture a note-taker that remembers key points over time, even when there's a lot of information.

Methods Overview

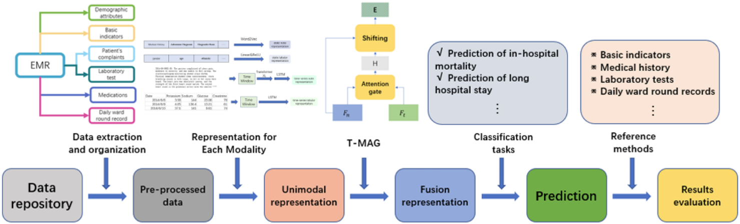

Figure 1: An overview of the study.

- **Objective**: Develop a multimodal fusion model to integrate temporal and static EMR data, enhancing prediction accuracy for medical tasks like in-hospital mortality and long stay.

Data Sources

Table 1. Details and examples of each modal data.

Modal	Feature Category	AMI Dataset	Stroke Dataset	Example
		Number of Features		
Static tabular	Gender	2	2	male, female
	Age	2	2	>60 years, ≤60 years
	Ethnicity	12	16	Han ethnicity, Hui ethnicity
	Marital status	3	3	married, divorced, unmarried
	Department	7	18	neurosurgery, vascular surgery
	Admission method	2	2	emergency department
	Basic indicators	7	7	height, weight, temperature
Static note	Chief Complaint	1	1	chief complaint
	Medical History	5	5	current medical history, past medical history
	Specialized Examination	2	2	specialist examination, auxiliary examination
	Admission Diagnosis	3	3	confirm diagnosis, supplementary diagnosis
	Characteristics	1	1	patient characteristics
	Diagnostic Basis	2	2	diagnostic basis, differential diagnosis
	Treatment Plan	1	1	treatment plan
Temporal tabular	Laboratory tests	73	71	serum triglyceride, serum creatinine
	Medications	17	21	angiotensin-converting enzyme inhibitor, heparin
	Vital signs measured	5	5	respiratory rate, pulse rate
Temporal note	Daily ward round records	1	1	daily ward round records

- **Data Sources**: EMR data from Beijing hospital (2014-2016), including 1271 AMI and 6450 stroke records.
- **Static Data**: Patient demographics, admission data.
- **Temporal Data**: Laboratory results, progress notes.
- **Data Preprocessing:** Interpolate missing values and segment data into manageable parts.

Model Training Process

- **Data Splitting**: Randomly split into training (80%), validation (10%), and test (10%) sets.
- **Optimizer and Hyperparameters**: Adam optimizer with a learning rate of 0.0001, batch size of 64, and a maximum of 50 epochs.
- **Loss Function**: Binary cross-entropy loss for calculating the difference between true and predicted values.
- **Evaluation Metrics**: AUROC, AUPRC, and F1 Score.
- **Model Selection**: Choose the parameter set with the minimum validation loss.
- **Implementation**: All deep learning models are implemented using PyTorch 1.10.

Model Evaluation

- **Tasks**:
  - In-hospital mortality prediction.
  - Length of stay prediction (threshold: 14 days).
- **Metrics**:
  - AUROC (Area Under the Receiver Operating Characteristic Curve)
  - AUPRC (Area Under Precision-Recall Curve)
  - F1 Score

Evaluation of the T-MAG-Based Multimodal Fusion Model

These metrics provide insights into the performance of the model, especially on imbalanced datasets.

- **Prediction Formula**

$$\hat{Y} = \text{sigmoid}(\text{Linear}(Z)) \tag{1}$$

where $\hat{Y}$ is the predicted value and Z is the final fused representation.

- **Loss Function**

$$L = -\frac{1}{B}\sum_{i=1}^{B} Y_i \log(\hat{Y}_i) + (1 - Y_i)\log(1 - \hat{Y}_i) \tag{2}$$

where Y is the true value and $\hat{Y}$ is the predicted value. This cross-entropy loss function measures the difference between the true values and the predictions.

T-MAG Model Structure

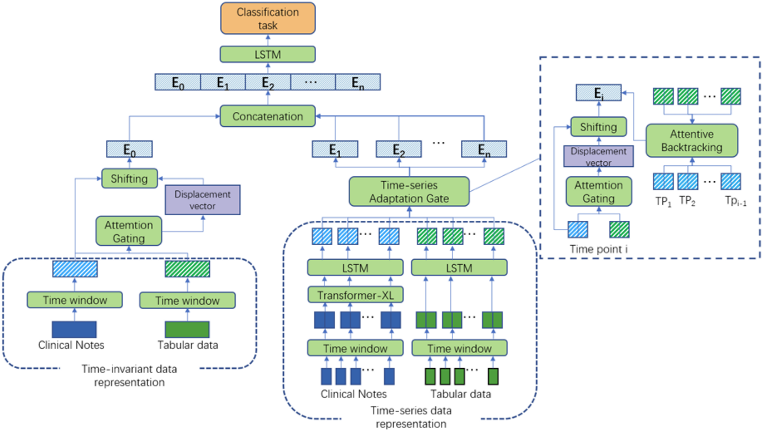

Figure 2: T-MAG Model Structure

T-MAG Model Structure

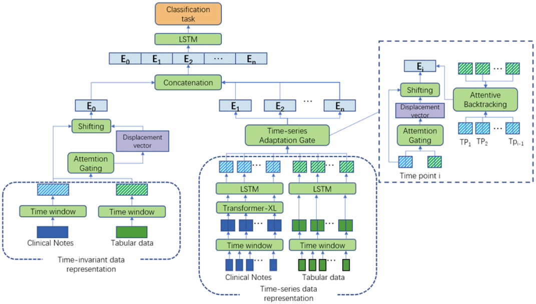

Figure 3: T-MAG Model Structure

- There are specialized encoders for different data types.
- **Fusion:** MAG control the influence of auxiliary modalities.
- **Attention-Backtracking Module:** Captures long-term dependencies in temporal data.

T-MAG Model Structure

- **T-MAG Model**: Integrates four EMR data modalities.
  - **Static Data**: Encoded using one-hot for demographic and doc2vec for note data.
  - **Temporal Data**: Encoded with LSTM and Transformer-XL for tabular and note data.

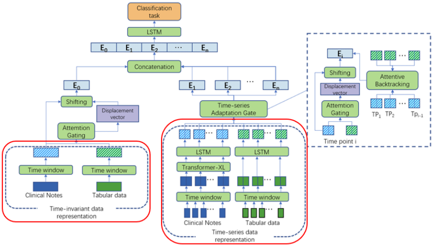

Figure 4: T-MAG Model Structure

Static Data Processing and Flow

- **Feature Embedding for Static Tabular Data**

$$F_{st} = \mathrm{Onehot}(I_{st})$$

Static tabular data is divided into continuous and discrete variables, which are then discretized and one-hot encoded to generate the static tabular representation.

- **Feature Embedding for Static Note Data**

$$F_{sn} = \mathrm{doc2vec}(I_{sn})$$

Static note data is processed using doc2vec to obtain the representation of the patient's static notes.

Temporal Data Processing and Flow

- **Feature Embedding for Temporal Tabular Data**

$$F_{tt}^{i\,\prime} = \mathrm{ReLU}(\mathrm{Linear}(I_{tt}^{i}))$$

$\mathrm{F}_{tt}^{i} = \mathrm{LSTM}(F_{tt}^{i\,\prime})$ Time series tabular data is encoded using a fully connected neural network with ReLU activation, followed by an LSTM network to obtain the time-structured representation for each subsequence.

- **Feature Embedding for Temporal Note Data**

$$F_{tn}^{i\,\prime} = \mathrm{TransformerXL}(I_{tn}^{i})$$

$\mathrm{F}_{tn}^{i} = \mathrm{LSTM}(F_{tn}^{i\,\prime})$ Time series clinical note data is encoded using Transformer-XL, followed by an LSTM network to obtain the temporal structure information for each subsequence.

Fusion Process

- **Fusion Process**:
  - Separate fusion for static and temporal representations using Multimodal Adaptation Gate (MAG).
  - Use of an **attention-backtracking** mechanism to capture long-term dependencies in patient data.
  - **Output**: Fused representation passed through LSTM for prediction tasks.

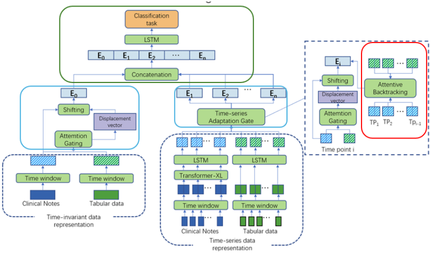

Figure 5: Fusion Process

Multimodal Adaptation Gate (MAG)

- **MAG Function**: Dynamically adjusts the influence of auxiliary data.
  - **Gating Mechanism**: Controls importance of auxiliary embeddings to minimize redundant information.
  - **Process**:
    - Treats one modality as primary and others as auxiliary.
    - Influence of auxiliary modalities on the primary modality is represented as a displacement vector H.
    - Fusion is interpreted as the addition of the primary modality and the displacement vector.

T-MAG Architecture: Data Fusion (Part 1)

- **Static MAG Calculation**

$$g_S = \sigma(W_{gS}[F_{Sn}; F_{St}] + b_{gS})$$

where g_S is the gating value, W_{gS} is the weight matrix, and b_{gS} is the scalar base, and $\sigma(x)$ is the sigmoid function.

T-MAG Architecture: Data Fusion (Part 2)

- **Displacement Vector Calculation**

$$H_S = g_S \cdot (W_{hS} F_{St}) + b_{hS}$$

where W_{hS} is the weight matrix for static tabular modality and b_{hS} is the bias vector.

- **Multimodal Representation Creation**

$$E_S = F_{Sn} + \alpha_S H_S \quad \text{where} \quad \alpha_S = \min\left(\frac{\|F_{Sn}\|^2}{\|H_S\|^2 \beta_S}, 1\right)$$

where β_S is a randomly initialized hyper-parameter, and α_S is the scaling factor.

- **Attention-Backtracking Module**
- **Final Patient Fusion Representation**
  - Concatenates static fusion representation with each time series subsequence fusion representation.
  - Inputs into a 2-layer LSTM to obtain the final patient fusion representation, denoted as Z.

Attention-Backtracking Mechanism

- **Purpose**: Enhances model's ability to capture long-term dependencies.
- **Mechanism**:
  - Uses self-attention on prior subsequences to inform the current sequence representation.
  - Outputs a weighted sum that includes context from previous records, improving time-series data interpretation.

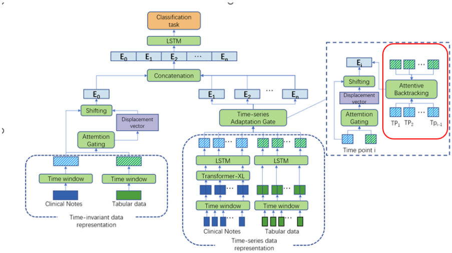

Figure 6: Attention-Backtracking Mechanism

Results Overview

- **Evaluation Structure**:
  - Impact of Main Modality Selection
  - Performance of Different Data Subsets
  - Comparative Experiments with Baseline Models
  - Ablation Study of Key Components

Impact of the Main Modality

Table 2. Predictive performance of models with different main modality combinations.

Data Set	Selection of Main Modality	Prediction of In-Hospital Mortality			Prediction of Long Hospital Stay		
		AUROC	AUPRC	F1	AUROC	AUPRC	F1
AMI	Static notes and temporal notes	**0.928**	**0.363**	**0.535**	**0.881**	**0.632**	**0.478**
	Static tabular data and temporal notes	0.923	0.351	0.520	0.879	0.630	0.473
	Static notes and temporal tabular data	0.925	0.359	0.528	0.877	0.626	0.466
	Static tabular data and temporal tabular data	0.919	0.346	0.516	0.874	0.621	0.454
Stroke	Static notes and temporal notes	**0.954**	**0.455**	**0.671**	**0.847**	**0.508**	**0.376**
	Static tabular data and temporal notes	0.951	0.447	0.665	0.836	0.486	0.359
	Static notes and temporal tabular data	0.945	0.438	0.651	0.834	0.479	0.352
	Static tabular data and temporal tabular data	0.933	0.425	0.644	0.818	0.443	0.334

The bold in the table represents the optimal result.

- **Objective**:Find which modality selection yields the best results.
- **Key Findings**: Using static notes and temporal notes as the main modality produced the highest AUROC in both tasks.
  AMI Dataset: **0.928 AUROC**, Stroke Dataset: **0.954 AUROC**.
- **Conclusion**:Clinical notes as the main modality outperform others.

Impact of Different Subsets

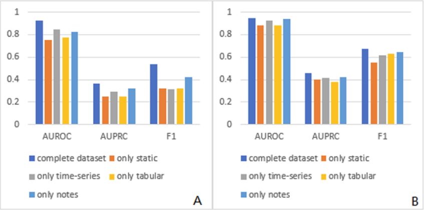

Figure 7: Prediction results of the complete data set and subsets in the AMI data set (A) and the Stroke data set (B).

- **Objective**: Assess the impact of data subsets on predictions.
  - Full Dataset: Best performance achieved.
  - "Only Notes": AUROCs **0.848** (AMI), **0.938** (Stroke).
  - "Static Only": Underperformed, time-series data is important.
- **Conclusion**: Complete data outperforms partial subsets.

Results on the AMI dataset

Table 3. The results of evaluation experiments on the AMI dataset.

	Model	In-Hospital Mortality			Long Length of Stay		
		AUROC	AUPRC	F1	AUROC	AUPRC	F1
T-MAG	T-MAG	**0.928**	**0.363**	**0.535**	**0.881**	**0.632**	**0.478**
Neural Network	DNN	0.748	0.228	0.313	0.726	0.423	0.318
	LSTM	0.769	0.243	0.328	0.758	0.528	0.413
Fusion Methods	Fusion-CNN	0.816	0.267	0.403	0.716	0.513	0.405
	Fusion-LSTM	0.828	0.287	0.435	0.818	0.544	0.436
Fusion Methods	MulT	0.913	0.323	0.502	0.856	0.593	0.468
	Crossformer	0.893	0.319	0.491	0.855	0.597	0.461
	PatchTST	0.838	0.296	0.447	0.825	0.537	0.422
	MISTS-fusion	0.917	0.341	0.508	0.866	0.609	0.438
	Glaucoma-fusion	0.821	0.277	0.415	0.805	0.565	0.461
	MAG-DNN	0.844	0.312	0.481	0.838	0.557	0.445
	MAG-LSTM	0.916	0.339	0.511	0.874	0.619	0.451

The bold in the table represents the optimal result.

- **Fusion Methods Compared**:
  - **MAG-based Fusion Models**: T-MAG and MAG-LSTM achieved the best results, underscoring the effectiveness of **MAG** for medical data fusion.
  - **Transformer-Based Models**: Crossformer and PatchTST performed well on stroke dataset but lacked in capturing temporal information as effectively as T-MAG.

Results on the Stroke dataset

Table 4. The results of evaluation experiments on the Stroke dataset.

	Model	In-Hospital Mortality			Long Length of Stay		
		AUROC	AUPRC	F1	AUROC	AUPRC	F1
T-MAG	T-MAG	**0.954**	**0.455**	**0.671**	**0.847**	**0.508**	**0.376**
Neural Network	DNN	0.849	0.333	0.505	0.736	0.375	0.245
	LSTM	0.856	0.349	0.511	0.746	0.378	0.255
Fusion Methods	Fusion-CNN	0.879	0.388	0.573	0.767	0.408	0.286
	Fusion-LSTM	0.887	0.401	0.595	0.822	0.430	0.317
Fusion Methods	MulT	0.933	0.435	0.641	0.830	0.441	0.332
	Crossformer	0.945	0.450	0.658	0.844	0.498	0.370
	PatchTST	0.927	0.423	0.633	0.825	0.435	0.323
	MISTS-fusion	0.949	0.451	0.660	0.833	0.445	0.339
	Glaucoma-fusion	0.891	0.405	0.611	0.815	0.425	0.310
	MAG-DNN	0.914	0.411	0.625	0.799	0.419	0.301
	MAG-LSTM	0.938	0.445	0.647	0.836	0.488	0.358

The bold in the table represents the optimal result.

- **Fusion Methods Compared**:
  - **Deep Learning Models**: DNN and LSTM underperformed due to limited handling of sequential data and multimodal integration.
- **T-MAG Performance:** Outshines traditional methods and other deep learning models. Shows the power of combining multiple data types.

Ablation Study

- **Objective**: Assess the effect of **attention-backtracking** on model performance.
- **Results**: Removing attention-backtracking led to notable performance drops:
  - AMI Dataset: AUROC reduced to **0.889** for mortality prediction and **0.837** for length of stay.
  - Stroke Dataset: AUROC reduced to **0.928** for mortality and **0.772** for length of stay.
- **Conclusion**: Attention-backtracking is essential for capturing long-term dependencies, especially in datasets with varied hospitalization periods.

1. Introduction
2. Method
3. Architecture
4. Result
5. Conclusions

Conclusions and Significance

- **Objective Achieved**: Developed the T-MAG model for predicting patient outcomes using diverse EMR data types.
- **High Performance**: T-MAG outperformed existing models in predicting in-hospital mortality and length of stay.
- **Predictive Accuracy**: Achieved high AUROC, AUPRC, and F1 scores, establishing it as an effective tool for clinical decision support.
- **Impact on Clinical Practice**: Enables early identification of high-risk patients and potential reduction of healthcare costs.

Contributions and Innovations

- **Innovative Fusion Technique**: T-MAG integrates static and temporal EMR data, effectively handling time-series data irregularities.
- **Attention-Backtracking Module**: Enhances the model's ability to capture long-term dependencies, improving predictive accuracy for clinical outcomes.
- **Adaptability**: Demonstrates robust performance across multiple EMR data types, with potential for broader applications in healthcare predictions.

Limitations and Future Work

Limitations:

- **Sample Size**: Data from a single hospital limits generalizability.
- **Temporal Data Handling**: Need to refine methods for complex time series patterns.
- **Focused Prediction Tasks**: Currently covers only mortality and length of stay; other tasks remain to be explored.

Future Work:

- **Expanding Dataset Scope**: Incorporate data from multiple institutions and English-language datasets.
- **Refining Temporal Fusion Techniques**: Develop advanced methods for handling irregular temporal data.
- **New Clinical Applications**: Apply T-MAG to additional tasks and consider integrating imaging data.

Questions

Questions and Discussion

- The study mentioned the modal diversity of clinical data and the irregularity of time series. What are some common challenges that impact the effective fusion of multimodal data? How can these challenges be better addressed in clinical practice?
- This paper uses LSTM and self-attention mechanism to capture long-term dependencies, and introduces an attention backtracking module in the fusion step. What are the advantages and disadvantages of this option compared to other methods (e.g. Transformer, CNN)?

Thank You