Early Prediction of Cybersickness in Virtual Reality Using a Large Language Model for Multimodal Time Series Data

Yoonseon Choi[1], Dayoung Jeong[2], Bogoan Kim[1], Kyungsik Han[1, 2]

[1] Department of Data Science, Hanyang University, Seoul, Republic of Korea

[2] Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea

Motivation

Limitations of Existing Models

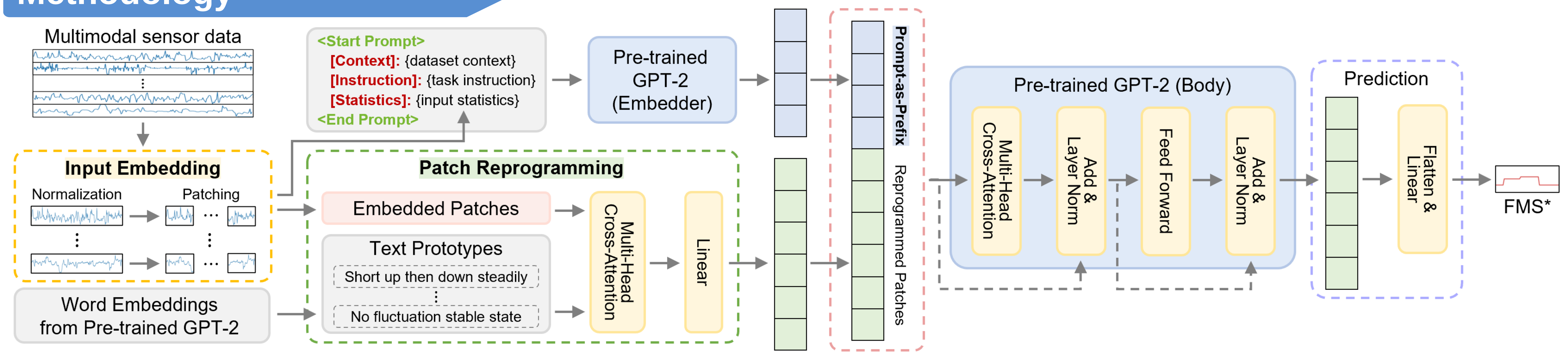
- The use of the **point-wise** multimodal sensor data **at the current moment**
- Prediction of cybersickness **at a single point in time**
- Enabling to implement mitigation measures **after the user experiences cybersickness**

Necessity of LLM in Time Series Data

- The need for a model's **understanding the data context** because the time series data collected in VR environment is **separated by user or video**
- The effectiveness in identifying **hidden trends and seasonality** in time series data

We confirmed **the potential for early prediction of cybersickness** based on multimodal sensor data, using **an LLM-based long-term time series forecasting (LTSF) model**

Methodology



* FMS (Fast Motion Sickness scale) ranges from 0 (no sickness) to 20 (severe sickness)

- **Input Embedding:** Instance normalization for each feature independently and splitting the data into patches of a fixed length
- **Patch Reprogramming:** Cross-attention between patches and pre-trained LLM's word embeddings to embed patches into LLM
- **Prompt-as-Prefix:** Prepending the prompt containing the dataset description (dataset context), guidelines for the model to perform tasks (task instruction), and statistics information of input data (input statistics) to the reprogrammed patches

Experiment

MSCVR Dataset

- A public dataset obtained from 45 participants by watching 20 cybersickness-inducing videos, each 45 seconds long
- Eye & head movement data + FMS

Models	Transformer	DLinear	PatchTST	Time-LLM
MAE	2.144	1.328	**0.931**	0.971
RMSE	3.044	1.910	**1.678**	1.696

- The results **suggest the effective application** of Time-LLM
- The results are encouraging because pre-trained LLM may **simplify the training process** by helping **train modalities in an integrated manner**, not modality-specific approach

Ablation Study

- We conducted ablation studies to investigate in detail **how the pre-trained LLM influences model performance**

Models	Time-LLM-NR*	Time-LLM-NP**	Time-LLM
MAE	1.766	1.038	**0.971**
RMSE	2.743	1.751	**1.696**

* Time-LLM-NR denotes the version without the reprogrammed patches
** Time-LLM-NP denotes the version without the prompt-as-prefix

- **The use of word embeddings from pre-trained LLM** may be effective to **understand patterns and trends** within multimodal sensor data
- **Prompt-as-prefix** is important to **understand the overall context** of the multimodal sensor data

Conclusion & Future Work

- We demonstrated that **early prediction is also possible in the cybersickness domain** by using an LLM-based LTSF model and we confirmed **the potential of applying LLM to multimodal sensor data**
- We expect to improve prediction performance through **the enhancements of patch reprogramming and prompt engineering to be more specific for multimodal sensor data**
- We need to validate the usability of the developed cybersickness early prediction model through **a user study after integrating the model into the VR contents**

Paper

Video

UBICOMP ISWC 2024 MELBOURNE, AUSTRALIA