Extra-Surv: Extrapolated Transformer Networks for Survival Analysis

Hosang Yu[1], Byungeun Shon[1,2], Dongwon Woo[1], Jungrae Cho[1], Jung Ju Seo[1],
Jeong-Hoon Lim[3], Jang-Hee Cho[3], Yong-Lim Kim[3], and Sungmoon Jeong[1,2*]

[1]Research Center for AI in Medicine, Kyungpook National University Hospital, Daegu, Korea
[2]Department of Medical Informatics, Kyungpook National University, Daegu, Korea
[3]Division of Nephrology, Department of Internal Medicine, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, Daegu, Korea

{youhs4554, jeongsm00}@gmail.com

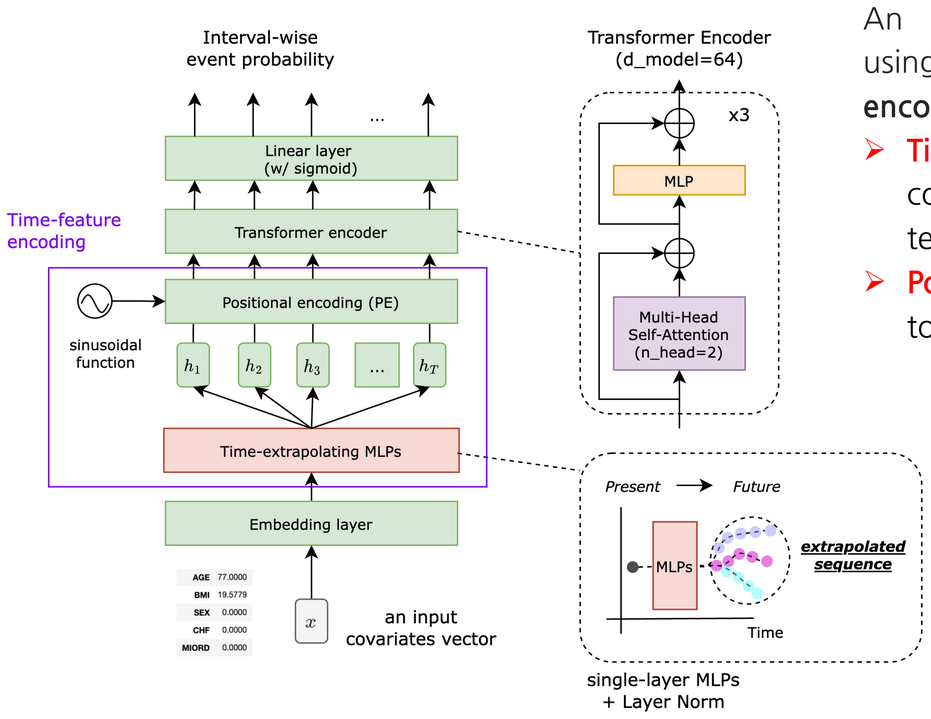
Introduction

- In the recently proposed transformer-based survival model [1], positional encoding (PE) is employed for time feature encoding which extrapolates time-varying characteristics from a single embedding vector by injecting temporal sinusoidal patterns, followed by transformers to predict the probability of the event occurrence.
- However, this temporal extrapolation with PE which simply adds sinusoidal outputs for a single embedding vector might not be enough to explain complex time-varying effects on a patient's condition in the real world.
- In this work, we propose an effective method for extrapolating time-varying covariates using transformers named Extra-Surv, which stands for Extrapolated Transformer Networks for Survival analysis.
- We demonstrate performance on three real-world medical datasets, where it outperforms all existing survival analysis models significantly.

Overview

- **Task.**
  - ➢ Survival Analysis aims to **predict a patient's survival time to a disease/death event**, which requires predicting the probability of the event over time.
- **Challenges.**
  - ➢ **Complex time-varying patterns:** The temporal evolution patterns of risk factors are complex but unobservable.
- **Motivations.**
  - ➢ **Extrapolating unobserved temporal evolution with a simple positional encoding** is helpful to improve survival prediction results using transformers [1].
  - ➢ **Autoregressive model** provides more complex time-varying modeling capabilities [2].
- **Contributions.**
  - ➢ Extend the time extrapolation capability **by autoregressively inferring a time-varying embedding sequence**, where the previous method simply uses a single embedding vector and applied PE.
  - ➢ Introduce **transformers to capture more complex time-varying patterns** to improve survival predictions.

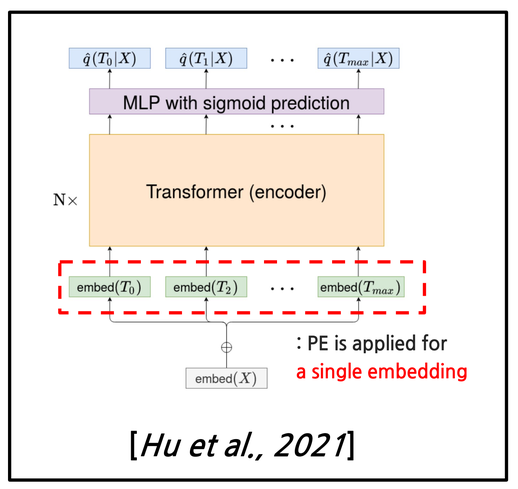
Network Architecture



Extra-Surv (ours)

An improved transformer architecture using a novel **autoregressive time-feature encoding** for survival analysis.

- ➢ **Time-extrapolating MLPs** are constructed to extrapolate complex temporal evolutions.
- ➢ **Positional encoding (PE)** is performed to inject temporal sinusoidal patterns.

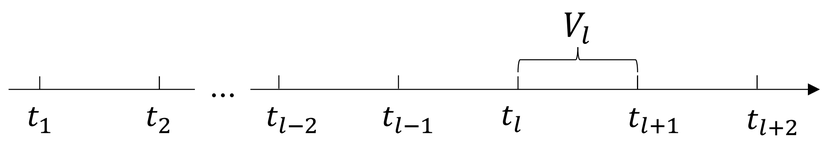


[*Hu et al., 2021*]

Method

Discrete time model:

Divide survival time into equal intervals and predict the probability of the event in each interval. Thus, interval-level ground truth labels (0 or 1) are target values.



Loss function:

Optimize same loss function of (*Ren et al., 2019*) to learn time-varying patterns to align with true survival distributions in autoregressive settings.

Uncensored Data | Censored Data

$$L = \underbrace{L_z}_{\text{P.D.F. Loss}} + \underbrace{L_{uncensored} + L_{censored}}_{\text{C.D.F. Loss}}$$

$$L_z = -\log \prod_{(\boldsymbol{x}^i,z^i)\in \mathbb{D}_{\text{uncensored}}} \Pr(z^i \in V_l|\boldsymbol{x}^i;\boldsymbol{\theta}) = -\log \prod_{(\boldsymbol{x}^i,z^i)\in \mathbb{D}_{\text{uncensored}}} p_l^i = -\log \prod_{(\boldsymbol{x}^i,z^i)\in \mathbb{D}_{\text{uncensored}}} h_{l^i}^i \prod_{l:l<l^i}(1-h_l^i) = -\sum_{(\boldsymbol{x}^i,z^i)\in \mathbb{D}_{\text{uncensored}}} \left[\log h_{l^i}^i + \sum_{l:l<l^i} \log(1-h_l^i)\right]$$

$$L_{\text{uncensored}} = -\log \prod_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{uncensored}}} \Pr(t^i \geq z|\boldsymbol{x}^i;\boldsymbol{\theta}) = -\log \prod_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{uncensored}}} W(t^i|\boldsymbol{x}^i;\boldsymbol{\theta}) = -\sum_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{uncensored}}} \log\left[1-\prod_{l:l\leq l^i}(1-h_l^i)\right]$$

$$L_{\text{censored}} = -\log \prod_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{censored}}} \Pr(z > t^i|\boldsymbol{x}^i;\boldsymbol{\theta}) = -\log \prod_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{censored}}} S(t^i|\boldsymbol{x}^i;\boldsymbol{\theta}) = -\sum_{(\boldsymbol{x}^i,t^i)\in \mathbb{D}_{\text{censored}}} \sum_{l:l\leq l^i} \log(1-h_l^i)\,.$$

Experiments

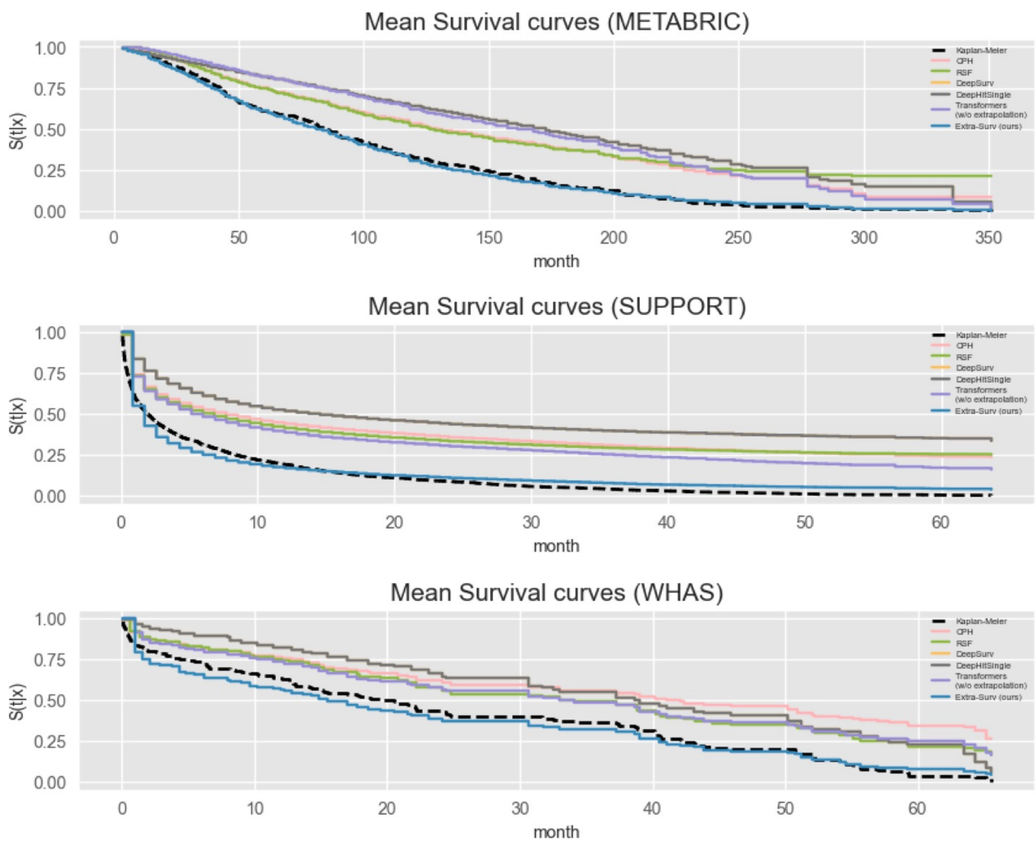
- **3 real-world public survival datasets (tabular data)** : METABRIC, SUPPORT, WHAS.
- **Evaluation metrics:** concordance index (C-index), survival duration prediction error (MAE).
  - ➢ Survival duration month is calculated as area under survival curve.
- **5 compared baseline models :** Cox, RSF, DeepSurv, DeepHit, and Transformers [1].
- Report the averaged evaluation performance across 10 runs.

Results

1 Quantitative results: comparison with existing models

Model	C-index (↑)			MAE (↓) (survival duration months)		
	METABRIC	SUPPORT	WHAS	METABRIC	SUPPORT	WHAS
Transformers *(Hu, 2021)*	0.679	0.609	0.818	92.1	25.01	18.6
Cox	0.660	0.573	0.729	100.9	29.2	36.2
RSF	0.666	0.606	0.886	110.1	28.2	26.5
DeepSurv	0.677	0.606	0.744	96.0	20.9	34.8
DeepHit	0.644	0.529	0.778	103.1	36.12	22.8
Extra-Surv (*ours*)	**0.684**	**0.62**	**0.903**	**52.6**	**10.1**	**9.87**

2 Qualitative results: mean survival curves for event-observed patients



: closer to the Kaplan-Meier curve (dashed line) is better, as it represents true survival distribution.

Summary

- improve time-extrapolating capability using MLPs extending existing PE-only method.
- the best both quantitative/qualitative results on the three real-world datasets.
- remarkably accurate survival time prediction performance compared to existing ones.
- (future works) development of a foundation model for image-based survival analysis, predicting the survival probability of each patient from medical images.

References

[1] Hu, S., Fridgeirsson, E., van Wingen, G., Welling, M.: Transformer-based deep survival analysis. In: Survival Prediction-Algorithms, Challenges and Applications. pp.132–148. PMLR (2021)
[2] Ren, K., Qin, J., Zheng, L., Yang, Z., Zhang, W., Qiu, L., Yu, Y.: Deep recurrent survival analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 4798–4805 (2019)

Acknowledgement: This work was partly supported by a grant of the Korea Health Technology R₩&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health ₩& Welfare, Republic of Korea (grant number : HR22C1832) and National IT Industry Promotion Agency(NIPA) grant funded by the Korea government(MSIT) (S0252-21-1001, Development of AI Precision Medical Solution(Doctor Answer 2.0)).