Personalized Survival Prediction with Contextual Explanation Networks

Maruan Al-Shedivat Avinava Dubey Eric P. Xing

Carnegie Mellon University, Machine Learning Department

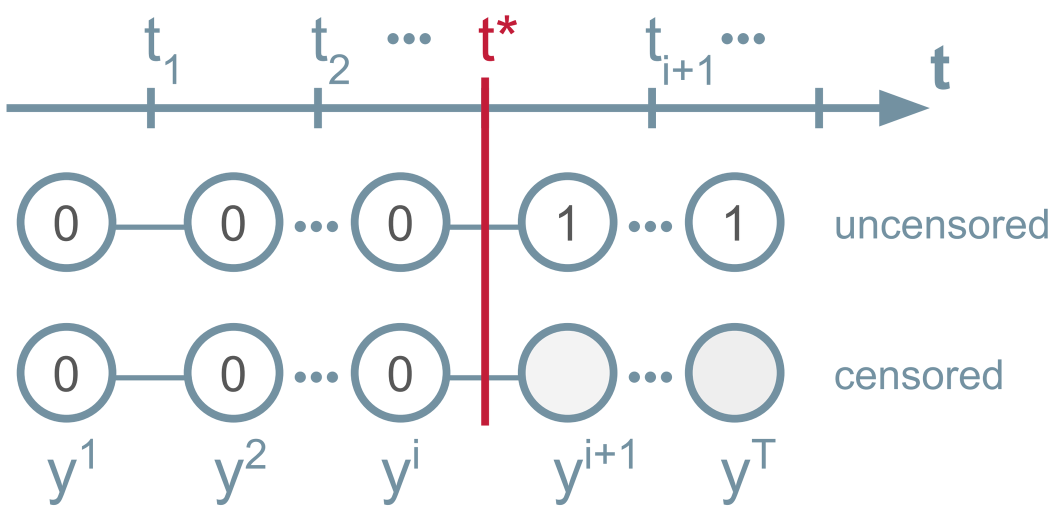
Abstract

Survival analysis: estimate the occurrence time and the risk of an unfavorable event in the future (e.g, death of a patient) that can inform our decisions at present time (e.g., help to select a treatment).

Approach:

- **Structured prediction.** Convert continuous-time survival prediction into a discrete structured prediction problem.
- **Recurrent dynamics.** Model dynamics of the survival distribution with a recurrent network.
- **Interpretability.** Ensure interpretability using explanations.

1. Background



Survival analysis as structured prediction [Yu et al., 2011]

- $\mathbf{X}$ – patient-specific features, $\mathbf{T}$ – time of the last follow up.
- The time is discretized into intervals: $\mathbf{Y} = (y^1, \dots, y^m)$.
- $p(Y \mid \mathbf{x}, \mathbf{\Theta})$ is modeled by a linear conditional random field (CRF).

Probability of an uncensored event:

$$p(T = t \mid \mathbf{x}, \mathbf{\Theta}) = \exp\left\{\sum_{i=j}^{m} \mathbf{x}^\top \boldsymbol{\theta}^i\right\} \Big/ \sum_{k=0}^{m} \exp\left\{\sum_{i=k+1}^{m} \mathbf{x}^\top \boldsymbol{\theta}^i\right\}$$

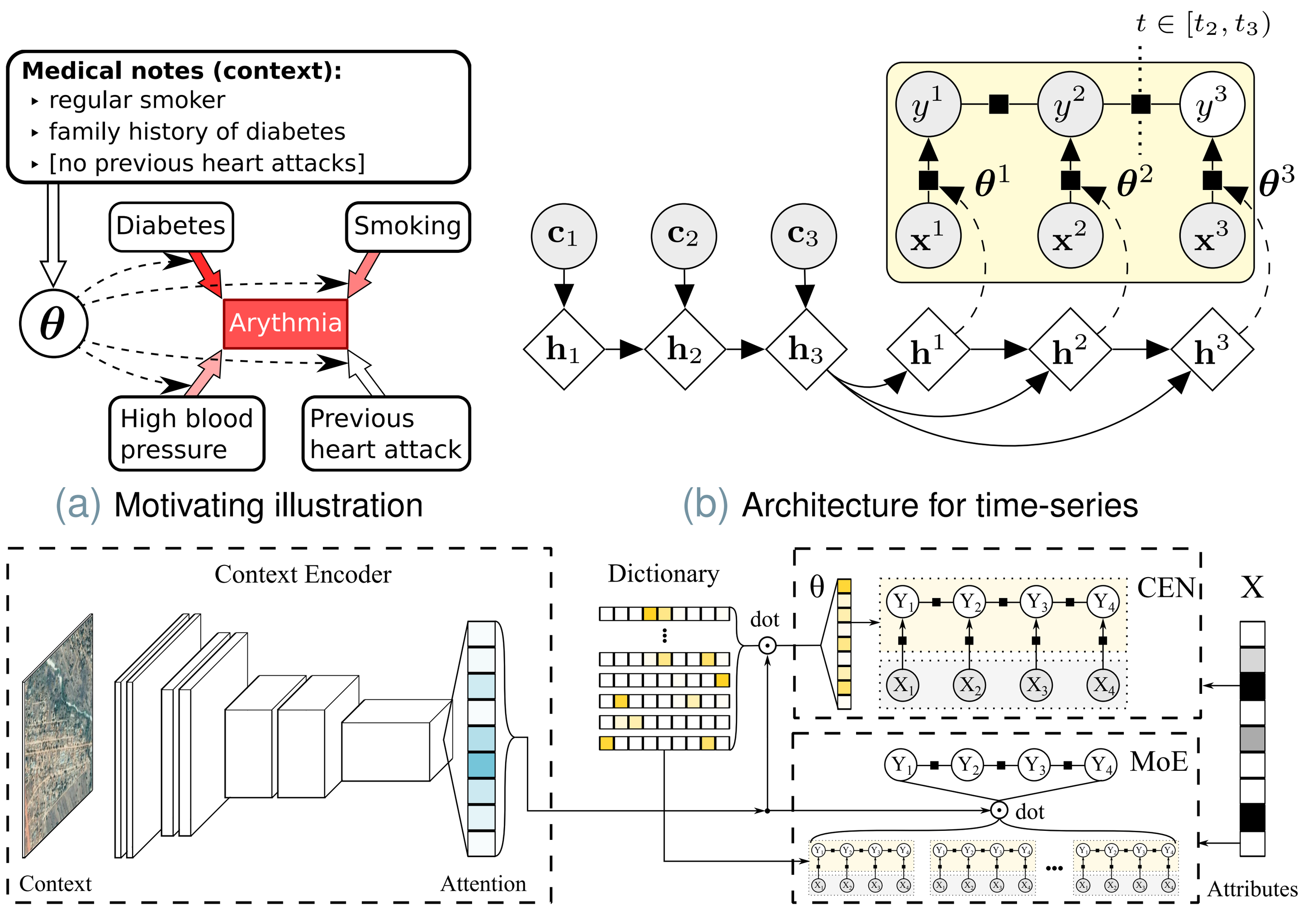
where the last follow up occurred at $t \in [t_j, t_{j+1})$.

Probability of a censored event:

$$p(T \geq t \mid \mathbf{x}, \mathbf{\Theta}) = \sum_{k=j+1}^{m} \exp\left\{\sum_{i=k+1}^{m} \mathbf{x}^\top \boldsymbol{\theta}^i\right\} \Big/ \sum_{k=0}^{m} \exp\left\{\sum_{i=k+1}^{m} \mathbf{x}^\top \boldsymbol{\theta}^i\right\}$$

where the patient was censored at $t \in [t_j, t_{j+1})$.

2. Contextual Explanation Networks (CENs)



(a) Motivating illustration (b) Architecture for time-series

(c) Comparison of a mixture of experts (MoE) and CEN architectures

Figure 2: Motivating illustration and a CEN architectures with CRF-based explanations.

Data representations:

- $\mathbf{C}$: low-level or unstructured features (e.g., medical notes, images, time-series data, etc.); processed by a deep network.
- $\mathbf{X}$: high-level or human-interpretable features (e.g., categorical); may or may not be derived from $\mathbf{C}$.

Personalized survival analysis with explanations [our work]

- Process the context features using an appropriate deep network.
- Train the network to generate parameters for linear CRFs that are *used* for prediction and play the role of explanations.

Probabilistic model for CEN:

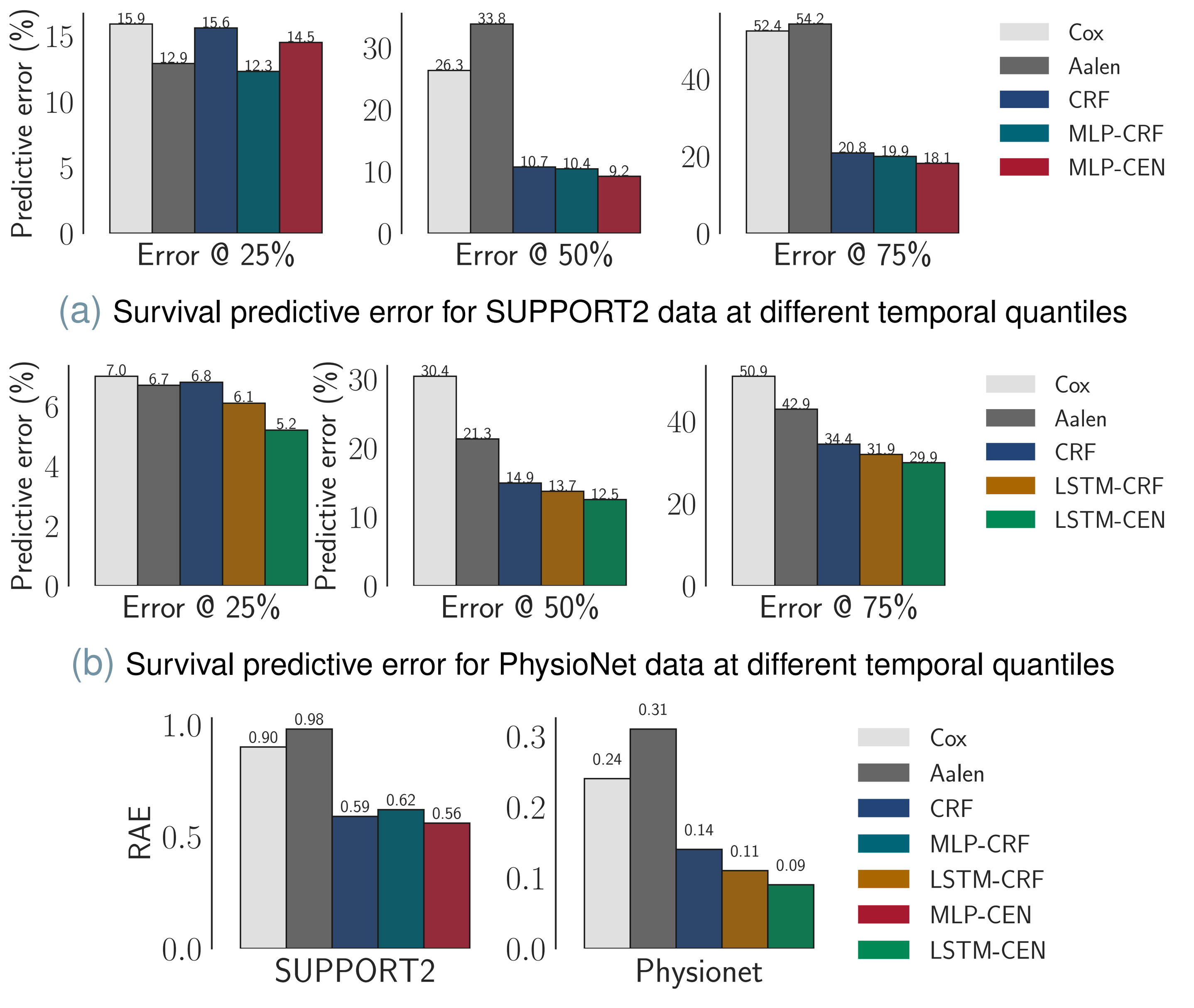
$$\mathbf{Y} \sim p(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}^{1:m}),$$
$$\boldsymbol{\theta}^t \sim p_{\mathbf{w}}(\boldsymbol{\theta}^t \mid \mathbf{C}), \quad t \in \{1, \dots, m\},$$
$$p(\mathbf{Y} = (y^1, y^2, \dots, y^m) \mid \mathbf{x}, \boldsymbol{\theta}^{1:m}) \propto \exp\left\{\sum_{t=1}^{m} y^i (\mathbf{x}^\top \boldsymbol{\theta}^t) + \omega(y^t, y^{t+1})\right\},$$

where $\boldsymbol{\theta}^t := \phi^t_{\mathbf{w},\mathbf{D}}(\mathbf{c}), \quad \phi^t_{\mathbf{w},\mathbf{D}}(\mathbf{c}) := \boldsymbol{\alpha}(\mathbf{h}^t)^\top \mathbf{D}, \quad \mathbf{h}^t := \mathrm{RNN}(\mathbf{h}^{t-1}, \mathbf{c})$

3. Quantitative results

Datasets

- **SUPPORT2:** 9105 patient records, 50 selected variables used for both $\mathbf{C}$ and $\mathbf{X}$ features.
- **PhysioNet Challenge 2012:** 4000 patient records, each represented by a 48-hour irregularly sampled 37-dimensional time-series.



(a) Survival predictive error for SUPPORT2 data at different temporal quantiles

(b) Survival predictive error for PhysioNet data at different temporal quantiles

(c) Relative absolute error for last follow up time prediction

Figure 3: Performance of the baselines and the proposed CEN architectures.

Takeaways

- **Personalization.** Modeling survival dynamics of each patient with deep nets boosts performance of the structured prediction approach.
- **Contextual explanations** regularize the model and further improve performance on a number of metrics.

4. Visualizing explanations

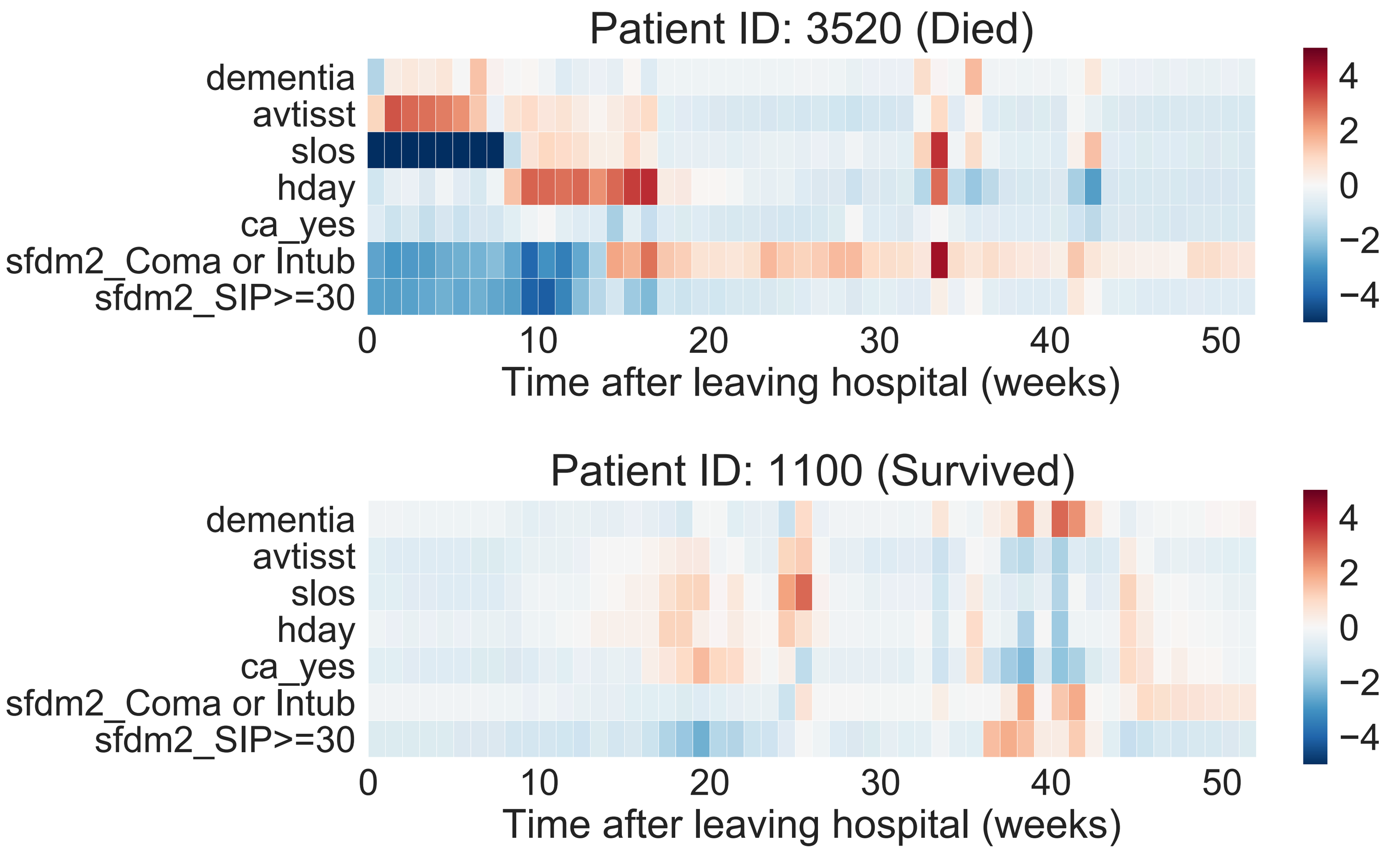


Figure 4: Weights of the explanations used by CEN to make predictions for two patients from SUPPORT2 dataset who did and did not survive.

Takeaways

- **Model diagnostics.** Explanations are personalized for each patient and reflect the effect of each feature on survival probability over time.
- **Data diagnostics.** Explanations visualize patterns discovered in the data and used for prediction $\Longrightarrow$ help to detect data leaks/errors.

Summary and Next Steps

- **Personalization.** A new class of models for survival analysis provides patient-specific predictions with consistent explanations.
- **Interpretability.** Explanations are generated on the fly for each prediction and relate features of interest to the mortality risk at each instance of time for a given patient.
- **Performance.** Our architectures use performant deep nets without compromising interpretability of the predictions.
- **Next steps.** We plan to expand the scope of our approach and tackle other problems in the healthcare domain that require accurate prediction along with consistent explanations.

Full paper: `http://arxiv.org/abs/1705.10301`

We are also looking for clinical collaborators!

ML for Health Workshop, NIPS 2017, Long Beach, USA

Correspondence: {alshedivat, akdubey, epxing}@cs.cmu.edu