Learning to Rank Patients Severity of Illness with Partially Observable Data

Jianzhun Du, Yuanheng (Henry) Wang

John A. Paulson School of Engineering and Applied Sciences, Harvard University

{jzdu,yuanheng_wang}@g.harvard.edu

Overview

- Ranking the severity of illness with only partially observable data.
- Submodular optimization – *Maximum Coverage* with cardinality constraint.
- Learning from data – *Distributional Optimization from Samples* (DOPS) [2].

Backgrounds

- ***ICU is extremely expensive.*** In 2005, cost of ICU was roughly \$82 billion (0.66% of the United States GDP). To reduce cost, ICU resources will become limited. Hence, most effective allocation of resources is imperative.
- ***Learning to Rank*** Allowing general structured output labels, *Structured SVM* [3] becomes a handy ranking model for data mining and information retrieval.
- ***Submodularity & Optimization from Samples*** Submodular function is not always accessible but learnable. DOPS provides reliable, efficient and scalable submodular optimization by learning from data within same distribution.
- ***Maximum Coverage Problem***
  - A classical NP-hard problem with widespread application.
  - The corresponding coverage function is a simple, yet important and widely used class of submodular functions.
  - Extension: *Weighted Maximum Coverage.* Let $\mathcal{U}$ be a ground truth set containing d items with corresponding non-negative weights Θ and C be a collection of n subsets of $\mathcal{U}$, i.e., $\mathcal{U} = \{u_i\}_{i=1}^d, \Theta = \{\theta_i\}_{i=1}^d, C = \{C_j\}_{j=1}^n$, where $\forall i \in [d], \theta_i \geq 0$ and $\forall j \in [n], C_j \subseteq \mathcal{U}$. Given S which is a subset of C, a function $f_\theta : 2^{[n]} \to \mathbb{R}$ is a coverage function if:

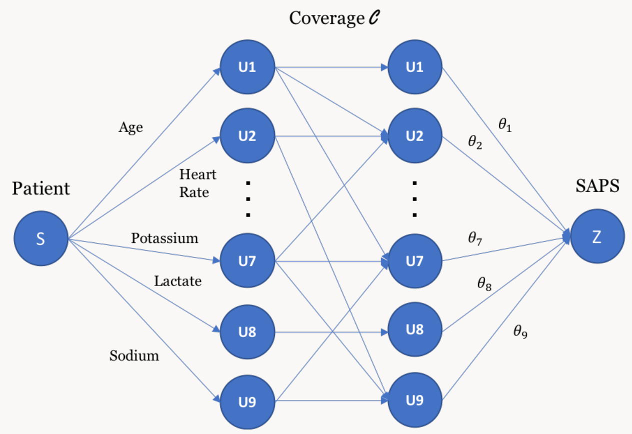
$$f_\Theta(S) = \sum_{u_i \in C(S)} \theta_i, \quad C(S) = \bigcup_{C_j \in S} C_j, \quad |S| \leq k \tag{1}$$

The objective is to find S with cardinality constraint such that $f_\theta(S)$ is maximized.

Methods

Optimize Patients Severity of Illness

- ***Motivation***
  - Demographic and physiological indices are innately connected.
- ***Measurement Transformation***
  - Two-level transformation – continuous to binary.
  - Multi-level transformation – continuous to one-hot encodings of multiple levels.
- ***Formulation: State-cover-State***



- ***Learnable Optimization Problem***
  - Given the training set $\mathcal{S} = \{S_i\}_{i=1}^n$ and test set $\mathcal{T} = \{T_i\}_{i=1}^m$, where S_i and T_i contain k observed measurements, we use DOPS to learn f from $\mathcal{S}$, and find $T = \text{argmax}_{T_i \in \mathcal{T}} f_\Theta(T_i)$.
  - With the spirit of DOPS and Structured SVM, we first shuffle data and divide into $N = \lfloor \frac{n}{m} \rfloor$ m-tuple sample sets $\mathcal{S} = \{(\mathbf{S}^i, \mathbf{z}^i)\}_{i=1}^N$, then solve the following program:

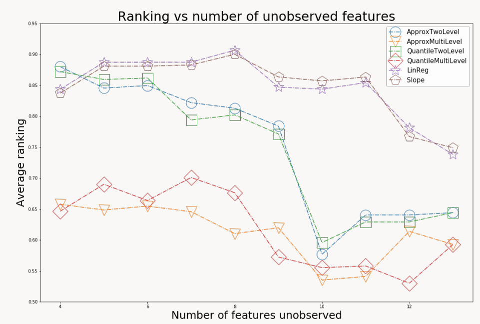
$$\begin{aligned} \min_{\Theta \in \mathbb{R}^d, \xi_i \in \mathbb{R}} \quad & \frac{1}{2}||\Theta||^2 + \frac{\lambda}{N} \sum_{i=1}^N \xi_i \\ \text{s.t. } & \phi_\Theta(\mathbf{S}^i, \mathbf{z}^i) - f_\Theta(S_y^i) + \xi_i \geq \Delta_\alpha(\mathbf{z}^i, y), \ \forall i \in [N], \forall y \in [m] \\ & \xi_i \geq 0, \ \forall i \in [N] \\ & \theta_i \geq 0, \ \forall i \in [d] \end{aligned} \tag{2}$$

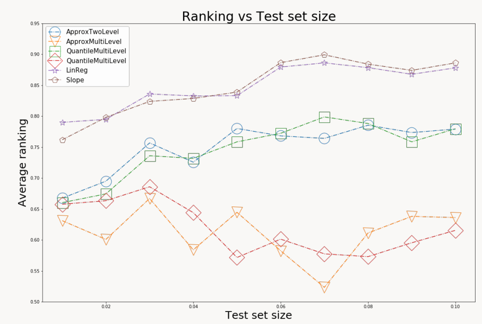
where $\Delta_\alpha(\mathbf{z}^i, y) = \mathbb{1}_{\{y \notin \alpha(\mathbf{z}^i)\}}, \alpha(\mathbf{z}) = \{y \in [m] : z_y \geq \alpha \max \mathbf{z}\}$ and $\phi_\Theta(\mathbf{S}^i, \mathbf{z}^i) = \frac{1}{|\alpha(\mathbf{z}^i)|} \sum_{y \in \alpha(\mathbf{z}^i)} f_\Theta(S_y^i)$.
  - By converting (2) to convex and unconstrained program with average hinge loss, it is easy to be solved using *subgradient descent*.

Experiments

- ***MIMIC-III Dataset*** – 13,500 ICU patients records [1].
- ***Measurement Masking*** – set values of "unobservable" measurements to 0.
- ***Baseline*** – Linear Regression

Results and Discussion





- ***Results***
  - Linear regression outperforms both DOPS models with different number of unobservable measurements and test size.
  - Two-level coverage model is better than multi-level one.
  - As the number of unobservable measurements increases, the ranking performance deteriorate
- ***Future Works***
  - Use Neural Network to learn the binary representation of current continuous input vectors.
  - Establish coverage C rigorously by consulting medical experts.
  - Incorporate treatment effect into our current maximum coverage framework to predict patient's severity score.

Reference

[1] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

[2] N. Rosenfeld, E. Balkanski, A. Globerson, and Y. Singer. Learning to optimize combinatorial functions. In *International Conference on Machine Learning*, pages 4371–4380, 2018.

[3] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the twenty-first international conference on Machine learning*, page 104. ACM, 2004.