MIS Quarterly

RESEARCH ARTICLE

# HEALTHCARE PREDICTIVE ANALYTICS FOR RISK PROFILING IN CHRONIC CARE: A BAYESIAN MULTITASK LEARNING APPROACH

**Yu-Kai Lin**
College of Business, Florida State University, Tallahassee, FL 32306 U.S.A. {ylin@business.fsu.edu}

**Hsinchun Chen**
Eller College of Management, University of Arizona, Tucson, A 895721 U.S.A. {hchen@eller.arizona.edu}

**Randall A. Brown**
Southern Arizona VA Health Care System, Tucson, AZ 85723 U.S.A. {Randall.Brown4@va.gov}

**Shu-Hsing Li**
College of Management, National Taiwan University, Taipei City 106, TAIWAN {shli@ntu.edu.tw}

**Hung-Jen Yang**
The Walter H. Shorenstein Asia-Pacific Research Center, Stanford University,
Stanford, CA 94305 U.S.A. {berniejohan@gmail.com}

# Appendix A

## Bayesian Multitask Learning for Artificial Neural Networks

We have shown in the main text how to apply the proposed Bayesian Multitask Learning (BMTL) approach to a set of baseline logistic regression models. The BMTL approach is applicable to other baseline models as well. To demonstrate the generalizability of the BMTL approach, we describe how to apply BMTL to artificial neural networks (ANNs) in this appendix. In the interest of consistency and for ease of exposition, we reuse the notations in equations (1) to (7) in the main text whenever possible.

Consider a feed-forward ANN with one single hidden layer. A typical functional form of the ANN is

$$y_i^{(k)} = \text{logit}\left(F^{(k)}(x_i)\right), \qquad i = 1 \ldots N, \quad k = 1 \ldots K \tag{A1}$$

where

$$F^{(k)}(x_i) = \alpha_0^{(k)} + \sum_{h=1}^{H} \alpha_h^{(k)} \pi\left(\beta_{h0}^{(k)} + \sum_{j=1}^{J} \beta_{hj}^{(k)} x_{ij}\right) \tag{A2}$$

The $\pi$ in (A2) is referred to as an activation function in the literature of ANNs and is often nonlinear. Two common choices for $\pi$ are the logistic and the hyperbolic tangent functions. The $\alpha_0^{(k)}, \alpha_h^{(k)}, \beta_{h0}^{(k)}$, and $\beta_{hj}^{(k)}$ are task-specific parameters to be fitted. The $\alpha_0^{(k)}$ and $\beta_{h0}^{(k)}$ are the *biases* for the output and hidden nodes, and the $\alpha_h^{(k)}$ and $\beta_{hj}^{(k)}$ are the *weights* for the respective input units. To achieve BMTL, we set the following prior distributions for these parameters.

$$\alpha_0^{(k)}, \beta_{h0}^{(k)} \sim Cauchy(0, 10), \quad k = 1, \ldots, K; \quad h = 1, \ldots, H \tag{A3}$$

$$\alpha_h \sim MVN\left(0, u_h^2 A_h\right), \quad h = 1, \ldots, H \tag{A4}$$

$$\beta_{hj} \sim MVN\left(0, s_{hj}^2 B_{hj}\right), \quad h = 1, \ldots, H; \quad j = 1, \ldots, J \tag{A5}$$

In (A4) and (A5), $\alpha_j = \left[\alpha_j^{(1)}, \alpha_j^{(2)}, \ldots, \alpha_j^{(K)}\right]^T$ and $\beta_{hj} = \left[\beta_{hj}^{(1)}, \beta_{hj}^{(2)}, \ldots, \beta_{hj}^{(K)}\right]^T$. At this point, it is straightforward to draw hyper prior distributions for $u_j$ and $s_{hj}$ as we did for $r_j$ in (4) and (5). Similarly, $\mathrm{A}_h$ and $\mathrm{B}_{hj}$ will follow the same formulation as $\Sigma_j$ in (6).

# Appendix B

## Robust Check for Evaluation 3 Using Different Decision Thresholds

**STK (# of events = 828)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit): Low | Predicted Risk (BMTL-Logit): High | Predicted Risk (Logit-lasso): Low | Predicted Risk (Logit-lasso): High | Predicted Risk (UKPDS): Low | Predicted Risk (UKPDS): High |
|---|---|---|---|---|---|---|
| Yes | 28 | 424 | 40 | 412 | 344 | 108 |
| No | 69 | 307 | 88 | 288 | 278 | 98 |

**AMI (# of events = 225)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit): Low | Predicted Risk (BMTL-Logit): High | Predicted Risk (Logit-lasso): Low | Predicted Risk (Logit-lasso): High | Predicted Risk (UKPDS): Low | Predicted Risk (UKPDS): High |
|---|---|---|---|---|---|---|
| Yes | 85 | 76 | 88 | 75 | 65 | 96 |
| No | 56 | 8 | 59 | 5 | 28 | 36 |

**ARF (# of events = 571)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit): Low | Predicted Risk (BMTL-Logit): High | Predicted Risk (Logit-lasso): Low | Predicted Risk (Logit-lasso): High | Predicted Risk (UKPDS): Low | Predicted Risk (UKPDS): High |
|---|---|---|---|---|---|---|
| Yes | 62 | 252 | 75 | 238 | UKPDS does not predict ARF risks. | |
| No | 84 | 174 | 100 | 158 | | |

**Figure B1. Summary of Results in Evaluation 3 Using 5% as the Cut-Off Value for High/Low Risks**

**STK (# of events = 828)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit) Low | Predicted Risk (BMTL-Logit) High | Predicted Risk (Logit-lasso) Low | Predicted Risk (Logit-lasso) High | Predicted Risk (UKPDS) Low | Predicted Risk (UKPDS) High |
|---|---|---|---|---|---|---|
| Yes | 208 | 244 | 205 | 247 | 436 | 16 |
| No | 309 | 67 | 312 | 64 | 354 | 22 |

**AMI (# of events = 225)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit) Low | Predicted Risk (BMTL-Logit) High | Predicted Risk (Logit-lasso) Low | Predicted Risk (Logit-lasso) High | Predicted Risk (UKPDS) Low | Predicted Risk (UKPDS) High |
|---|---|---|---|---|---|---|
| Yes | 126 | 35 | 125 | 36 | 140 | 21 |
| No | 63 | 1 | 63 | 1 | 60 | 4 |

**ARF (# of events = 571)**

| Preventive treatment prescribed at/before $v_{0i}$ | Predicted Risk (BMTL-Logit) Low | Predicted Risk (BMTL-Logit) High | Predicted Risk (Logit-lasso) Low | Predicted Risk (Logit-lasso) High | Predicted Risk (UKPDS) Low | Predicted Risk (UKPDS) High |
|---|---|---|---|---|---|---|
| Yes | 174 | 139 | 181 | 132 | UKPDS does not predict ARF risks. | |
| No | 196 | 62 | 200 | 58 | | |

**Figure B2. Summary of Results in Evaluation 3 Using 20% as the Cut-Off Value for High/Low Risks**