Computational Health: From Data to Impact

Knowledge Sources

Pattern Extraction
- Preprocessing
- Feature Engineering
- Feature Selection
- Pheno-Embedding

Insights Generation
- Precision Cohort
- Multi-Task Learning
- Longitudinal Modeling
- Interpretable Models
- Causal Inference

Delivery and Engagement
- Performance Intelligence and Optimization
- Behavioral Profiling and Intervention
- Meaningful Measures and Assessments

Federated and Privacy Preserving Learning
Patient Similarity Analytics for Precision Cohort

**Goal**
- Identify patients who are similar to a given patient of interest in a clinically meaningful way
- Identify a measure of clinical similarity between patients

**Approach**
- Supervised metric learning

**Challenges Addressed**
- Patient similarity is context dependent
- Feature dimensionality can be very large

Personalized Predictive Models - T2D Onset Prediction Example

Insights for personalized intervention planning

- Diabetes patient population is heterogeneous
- Traditional predictive modeling approaches only provide “universal” risk factor identification and ranking
- Personalized predictive modeling approach provides patient specific risk factors and ranking
- Clusters of risk factors, and patient risk profiles
Temporal patterns and interactions are important features in predictive modeling in healthcare.

Prior methods do not sufficiently address the challenge of extracting such features from longitudinal patient record matrices (EHR).

We developed an end-to-end Deep Learning framework tailored to longitudinal health care data to learn the temporal pattern and exploit them for risk prediction.

### Disease Onset Prediction Results (AUC)

<table>
<thead>
<tr>
<th></th>
<th>CHF</th>
<th>COPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Random Forest)</td>
<td>0.7156</td>
<td>0.6624</td>
</tr>
<tr>
<td>Deep Learning (Slow-Fusion Convolutional Neural Network)</td>
<td>0.7675</td>
<td>0.7388</td>
</tr>
</tbody>
</table>
Diabetic Kidney Disease Prediction

Impact of data coverage on performance

Key Findings
- Additional data categories improve prediction accuracy
- The aggravation of urinary protein observation is strongly affected by its variance over past 180 days
- DKD aggravation group had significantly higher incidence rates of Hemodialysis and Cardiovascular diseases

Features | AUC | Accuracy |
--- | --- | --- |
Profile | 0.562 | 0.548 |
Profile + ICD10 | 0.562 | 0.557 |
Profile + ICD10 + YCode | 0.613 | 0.594 |
Profile + ICD10 + Blood Tests (latest) | 0.644 | 0.606 |
Profile + ICD10 + YCode + Blood Tests (latest and longitudinal) | 0.656 | 0.610 |
Profile + ICD10 + YCode + Blood Tests (latest and longitudinal) + Urinary Tests (latest and longitudinal) | 0.729 | 0.691 |
Profile + ICD10 + YCode + Blood Tests (latest and longitudinal) + Urinary Tests (latest and longitudinal) + Current Disease - Disease History | 0.743 | 0.701 |

- Deep Learning to extract temporal features
- Logistic regression for DKD prediction
Comprehensive Risk Assessment – Multi-Task Sparse Learning

Goal
– Simultaneously predict multiple risks
– Explore and exploit risk associations
– Identify common and unique risk factors

Use Cases: elder care risk assessment, diabetes complications

Improved Prediction Accuracy

Identify Association Among Risks

KDD 2015
Multi-Task Learning for Diabetes Complications Prediction

**Goal**: Predict *when* a patient will develop complications after the initial T2DM diagnosis

**Approach**: Multi-task Survival Analysis

- **RankSvx**: A novel data-driven time-to-event modeling method
  - Accurate prediction of event times, and
  - Ranking of the relative risks among patients
- **Multi-task RankSvx** to simultaneously model multiple complication risks
  - Leverage association between different diabetes complications
- Applied to predictions of retinopathy, neuropathy, nephropathy, vascular diseases

RankSvx outperforms traditional survival models and regression model in CI and MAE

MTL-RankSvx outperforms STL-RankSvx

© 2019 IBM Corporation
• GOAL → Provide comprehensive view and deeper understanding of a disease in terms of characteristics of underlying disease stages, areas of manifestation and progression pathways
• METHOD → Multi-layer probabilistic modeling framework to incorporate data from diverse sources
• Initial work on COPD; Work ongoing on enhanced methodologies & application to other conditions, including Huntington’s (CHDI), T1D (JDRF), PD (MJFF).

Published in KDD 2014, AMIA 2017, AMIA 2018, JAMIA Open 2019
Huntington’s Disease Progression Modeling

Challenges in Understanding HD Progression

- Disease manifestation along multiple dimensions with complex patterns
- Heterogeneous progression pathways
- No clear definitions of disease states

Probabilistic Disease Progression Modeling

- Incorporate heterogeneous features coming from multiple studies and assessments covering multiple aspects of HD
- Provide comprehensive view of disease states and the progression of HD through a multi-layer probabilistic disease progression model
- Better understanding of disease sub-types
- Identify factors that are associated with disease progression patterns

Clinical Decision Support

- Improved understanding of disease progression: population/patient
- Insights into HD clinical assessments and sensitivities
- Objective baseline
- Cohort selection – trial enrichment
- Optimizing trial design – trial simulator
- Biomarker discovery

Clinical Trial Design

Natural History of HD (Ross et. al. 2014)
Integrated Huntington’s Disease Progression Model

- Trained on data integrated from four prospective observational studies of HD (~16k case, 3k control)
- Discovered 9 disease states, over span of ~4 decades (prodromal, transition, manifest)
- Highlights (compared to prior-art HD progression indices)
  - Capturing multi-faceted manifestation throughout disease progression
  - Finer characterization, particularly of early states
  - Characterization of complex patterns of progression in transition (critical) states
  - Individual patient: more nuanced view of state of progression

Population view of multiple aspects of disease progression of HD

![Image of disease progression models]

**Table 2. State sequence of an example patient**

<table>
<thead>
<tr>
<th>Visit date (years)</th>
<th>State from IHDP</th>
<th>Shoulson and Fahn Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>1.1</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>2.1</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>3.5</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>4.2</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>5.7</td>
<td>2</td>
<td>Premanifest</td>
</tr>
<tr>
<td>6.6</td>
<td>3</td>
<td>Premanifest</td>
</tr>
<tr>
<td>7.5</td>
<td>3</td>
<td>Premanifest</td>
</tr>
<tr>
<td>8.9</td>
<td>3</td>
<td>HD1</td>
</tr>
<tr>
<td>10.3</td>
<td>4</td>
<td>HD1</td>
</tr>
<tr>
<td>11.4</td>
<td>4</td>
<td>HD1</td>
</tr>
<tr>
<td>11.8</td>
<td>4</td>
<td>HD1</td>
</tr>
<tr>
<td>13.6</td>
<td>5</td>
<td>HD1</td>
</tr>
<tr>
<td>14.7</td>
<td>6</td>
<td>HD2</td>
</tr>
</tbody>
</table>
Causal Inference for Time Varying Treatment Strategies

• **Goal** → To provide causal estimate of effect of time varying treatments using observational data
• **Method** → G-Computation + hierarchical Bayesian models (post-processing)
• **Challenges addressed:**
  • Time varying treatment decisions
  • Outcomes recorded at irregular intervals/varying treatment durations
  • Multiple related drugs and multiple related outcomes
Modeling of Parkinson’s Disease Progression

Inconsistent trajectory patterns across patients, even when controlling for early clinical presentation.

Summary scores mask symptom patterns

Within 2 years of diagnosis, patients already show wide variability in clinical presentation.

Patterns are further confounded by dopaminergic medication use.

Challenges in Understanding PD Progression

• Incorporate many clinical measures of PD
• Use control cases to subtract away non-PD effects (e.g. aging)
• Use input-outcome HMM approach to train personalized Medication Aware progression models

Probabilistic Disease Progression Modeling

Statistical Model

AAA 2019, MLHC 2020
Improve Patient Engagement in Care Management

Engaging patients in interventions that are most effective for patients like them

Objectives
- Help care managers prioritize patients who will be more receptive to care management interventions.
- Help care managers set behavior goals/interventions based on intervention effectiveness estimates.

Data
- Care management records (structured + unstructured)

Method
- Patient similarity, Causal inference,

Key Finding
- Behavior phenotype-based care planning strategies could yield more effective intervention recommendations for goal attainment compared to population level strategies.

AMIA 2017, AMIA 2018 (best paper)
AI Triage Engine

Decision support system for medical triage to guide individuals to the next step of care

NLP Pipeline (CNN, Bi-GRU, Bi-LSTM)

Auto-creation of ontology and language agnostic KG

ML models for next-best-question and recommendation

AIDA

Medical triage tool deployed by a telemedicine service provider in Switzerland

EMNLP 2018, EMBS 2018, AMIA 2020
Hyperlocal Case Prediction

Framework for hyperlocal predictions of COVID-19 cases using novel compartmental models with ML enhancements

- Core data: incidences, deaths by day, region; Countermeasures by day, region; Geospatial data
- Additional data: Testing; Hospitalizations; Claims; Demographics; "Mobility"; Comorbidity criteria

- Compartmental model (SEAIR) accounting for asymptomatic transmission, NPI and testing
- Spatial Temporal Clustering
- Detection of Non-Pharmaceutical Interventions

- Community risk model for Return To Work Advisor
- Vaccine trial site identification
- Resource Demand Prediction

Thank You!