ICML 2018 TUTORIAL: MACHINE LEARNING FOR PERSONALISED HEALTH

Danielle Belgrave (Microsoft Research Cambridge)
Konstantina Palla (Microsoft Research Cambridge)
Lamiae Azizi (University of Sydney)
TUTORIAL: MACHINE LEARNING FOR PERSONALISED HEALTH

Danielle Belgrave (Microsoft Research Cambridge)
Roadmap

Part 1
Introduction: Drivers of machine learning in healthcare

Part 2
Critical Evaluation of data-driven healthcare

Part 3
ML strategies for healthcare personalisation

Part 4
Future challenges of ML for Healthcare
Part 1: What are the drivers of machine learning in healthcare?

Wellness and self-care personalisation: patient perspective

Population data-driven healthcare: policy perspective

Precision drug discovery, development and therapeutics: pharmaceutical industry perspective

Data protection and connected care: provider and regulator perspectives
Part 2: Critical evaluation of data-driven healthcare

Traditional statistical approaches to healthcare
- Principles of study design
- Types of study design
- Causal modelling

Current applications of ML in the healthcare domain
Machine Learning has the Potential to Disrupt and Impact Healthcare
The Stakeholders in Healthcare

- Patient/Person
- Population
- Pharmaceuticals
- Providers
The Person at the Centre of Healthcare

Patient/Person

ML has the capacity to transform healthcare
- Understanding physiological changes over time
- Forecasting of progression or onset of disease
- Personalising treatment strategies
Population Data-driven Healthcare

Population

Elucidates average effects and deviations from average effects
Policy recommendations
Health education
Outreach
Research for disease detection and injury prevention
Reduce healthcare inequalities

What we as a society do collectively to assure the conditions in which people can be healthy
The Pharmaceutical Perspective: Drug Discovery and Therapeutics
General Data Protection Regulation

Enhance protection of personal data

Significant impact for organisations and how they manage data with some potentially very large penalties for violations – 4% of global revenues

Impacts the storage, processing, access, transfer, and disclosure of an individual’s data records

These protections apply to any organisation (anywhere in the world) that processes the personal data of EU data subjects
Data Protection and Connected Care: The Provider and Regulator Perspective

Providers
AN EVALUATION OF DATA-DRIVEN HEALTH

Biostatistical and Epidemiological Principles
The Beginnings of Data-Driven Health

The study of the distribution and determinants of health related states or events in specific populations & the applications of this study to the control of health problems.
The Beginnings of Data-Driven Health

Florence Nightingale (1820 – 1910)

Data visualisation: death toll of the Crimean War

Army data: 16,000/18,000 deaths not due to battle wounds, but to preventable diseases, spread by poor sanitation
The Beginnings of Data-Driven Health

Contextual phenomena: cholera incidence

Ecological design: compare cholera rates by region

Cohort design: compare cholera rates in exposed and non-exposed individuals
R.A. Fisher and the Principles of Experimental Design

1. Randomisation: Unbiased allocation of treatments to different experimental plot

2. Replication: repetition of the treatment to more than one experimental plot

3. Error control: Measure for reducing the error of variance

Why do these 2 plants differ in growth?
Principles of Study Design

Need to set up a study to answer a research question

Design most important aspect of a study and perhaps the most neglected

The study design should match research question
    So that we don’t end up collecting useless data or the principle outcome ends up not being recorded

No matter how good an algorithm is, if the study design is inadequate (garbage in) for answering the research question, we’ll get garbage out
Types of Study Design

Non-Experimental Observational Studies
- Descriptive
  - Case Reports
  - Case Series
  - Cross-Sectional or Prevalence Study
- Analytical
  - Case-control
  - Cohort Study

Experimental Intervention Studies
- Randomised Clinical Trial
- Non-randomized/Field/Community Trial
Important Concept: Randomisation

Definition: The process by which allocation of subjects to treatment groups is done by chance, without the ability to predict who is in what group

Aims:
- To prevent statistical bias in allocating subjects to treatment groups
- To achieve comparability between the groups
- To ensure samples representative of the general population
Methods of Randomisation

**Simple Random Sampling**

```
Population
1  2  3  4
5  6  7  8
9 10 11 12
```

```
Sample
2  5
```

**Permutated Block Randomisation**

```
AA BB AA BB
BB BB AA AA
AA A A BB BB
```

**Stratified Random Sampling**

```
Populations
Strata
Sample
```
## Sample Size and Power Calculations

<table>
<thead>
<tr>
<th></th>
<th>No disease (D = 0)</th>
<th>Disease (D = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No disease</td>
<td>☺</td>
<td>❌</td>
</tr>
<tr>
<td>Specificity</td>
<td><strong>Speciﬁcity</strong></td>
<td><strong>Type I error</strong> (False +) $\alpha$</td>
</tr>
<tr>
<td>Disease</td>
<td>❌</td>
<td>☺</td>
</tr>
<tr>
<td>Type II error</td>
<td><strong>Type II error</strong> (False -) $\beta$</td>
<td><strong>Power 1 - $\beta$; Sensitivity</strong></td>
</tr>
</tbody>
</table>

**Power** is the probability that a test of significance will pick up on an effect that is present.

Increases with sample size, effect size, and type I error.

\[
\text{Power} \propto \frac{\text{Sample size (n)}}{\text{Effect size (\Delta), Alpha(\alpha)}}
\]
The Challenge of Missing Data

Missing data is a common problem in healthcare data and can produce biased parameter estimates.

Reasons for missingness may be informative for estimating model parameters.

Bayesian models: coherent approach to incorporating uncertainty by assigning prior distributions.

Missing Data

Missing Completely At Random (MCAR)
The probability of data being missing does not depend on the observed or unobserved data
\[ \text{e.g. } \logit(p_{it}) = \theta_0 \]

Missing At Random (MAR)
The probability of data being missing does not depend on the unobserved data, conditional on the observed data
\[ \text{e.g. Children with missing wheeze data have better lung function} \]
\[ \text{e.g. } \logit(p_{it}) = \theta_0 + \theta_1 t_i \text{ or } \logit(p_{it}) = \theta_0 + \theta_2 y_0 \]

Missing Not At Random (MNAR)
The probability of data being missing does depend on the unobserved data, conditional on the observed data.
\[ \text{e.g. Children with missing lung function have better lung function} \]
\[ \text{e.g. } \logit(p_{it}) = \theta_0 + \theta_3 y_{it} \]

Missing Completely At Random

\[ \logit(p_{it}) = \theta_0 \]
**Missing At Random**

- Model of Interest:
  - $\beta$
  - $\mu_i$
  - $\sigma^2$
  - $y_i$

- Model of Missingness:
  - $\theta$
  - $x_i$
  - $p_i$
  - $m_i$

Logit model: $\text{logit}(p_{it}) = \theta_0 + \theta_1 x_i$
Missing Not At Random

Model of Interest

Model of Missingness

\[ \logit(p_{it}) = \theta_0 + \theta_3 y_{it} \]
Causal Reasoning

The questions that motivate most studies in the health, social and behavioral sciences are not associational but causal in nature.

Before an association is assessed for the possibility that it is causal, other explanations such as chance, bias and confounding have to be excluded.

Require some knowledge of the data-generating process - cannot be computed from the data alone, nor from distributions governing data.

Aim: to infer dynamics of beliefs under changing conditions, for example, changes induced by treatments or external interventions.

Prognostic Biomarker (Risk Factor)

A biological measurement made before treatment to indicate long-term outcome for patients either untreated or receiving standard outcome.

Predictive Biomarker (Moderator)

A variable that *changes the impact* of treatment on the outcome. A biological measurement made before treatment to identify patients likely or unlikely to benefit from a particular treatment.
Mediator

A *mechanism* by which one variable affects another variable. Omitted common causes (hidden confounding) should always be considered as a possible explanation for associations that might be interpreted as causal.
Efficacy and mechanism evaluation: Causal framework for investigating who medications work for

Predictive biomarker (moderator)

Random Allocation

Prognostic biomarker (risk factor)

Mediator

Outcomes
Example: Personalisation of Cancer Treatment

- Genetic Marker
- Treatment
- Prognostic biomarker (risk factor)
- Tumor Size
- Outcome (Survival)
Bradford-Hill Principles of Causality

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plausibility</td>
<td>Does causation make sense</td>
</tr>
<tr>
<td>Consistency</td>
<td>Cause associated with disease in different population and studies</td>
</tr>
<tr>
<td>Temporality</td>
<td>Cause precedes disease</td>
</tr>
<tr>
<td>Strength</td>
<td>Cause strongly associated with disease</td>
</tr>
<tr>
<td>Specificity</td>
<td>Does the cause lead to a specific effect</td>
</tr>
<tr>
<td>Dose-Response</td>
<td>Greater exposure to cause, higher the risk of disease</td>
</tr>
</tbody>
</table>
Machine Learning for Healthcare in Context

Health data complexity requires adequately complex methodologies and algorithms.

Methods don’t scale, need more advanced techniques and thinking about other techniques developed outside the traditional stats community.

Need for scale and speed.
CURRENT WORK IN ML IN THE HEALTHCARE DOMAIN

Data complexity requires adequately complex algorithms
Sparsity in Health Data

Major challenge for truly generalizable and scalable AI in healthcare is maximizing information utility for public health impact when that information (observational or clinical-context data) is sparse

- Missing data
- Inadequately sampled data
- Data that does not represent the diversity of a population

Generalisability: Training datasets that are representative of the diversity of the population as well as the heterogeneity of health conditions.

Transfer learning: potential to

- Maximise utility of available data
- Improve model’s ability to generalise
Transfer Learning for Data Sparsity

Good quality healthcare data is expensive and very often sparse.

Aim: Maximizing information by using multiple data sources.

Challenge: Feature mismatch: features in different datasets may vary.

Challenge: Distribution Mismatch: differing patient populations across different hospitals.

GAN architectures to efficiently enlarge the dataset.

Better predictive models than if we simply used the target dataset.

Jinsung Yoon, James Jordon and Mihaela van der Schaar.

RadialGAN Transfer Learning for Data Sparsity

$Z$: Latent space
$X^{(i)} \times Y$: $i^{th}$ domain
$G_i, F_i, D_i$: Decoders, Encoders and Discriminator of the $i^{th}$ domain

The $i^{th}$ domain is translated to the $j^{th}$ domain via $Z$ using $F_i$ and $G_j$

Jinsung Yoon, James Jordon and Mihaela van der Schaar.
Public Health intervention: Explored the impact of ads on changing health behaviours as measured by future health promotion searches.
Learning Structure from Real-World EHRs

A linear kernel SVM is trained to create classification boundaries for three clinical outcomes: in-hospital mortality, 30 day post-discharge mortality, and 1 year post-discharge mortality.

Clinical baseline features are extracted from the database for every patient and derived features are computed to form the Structured Features matrix $v$.

Per-note latent topic features are aggregated in extending 12 hour windows and used to form matrix $q$ where is the overall proportion of topic $k$ in time-window $m$.

Each patient’s de-identified clinical notes are used as the observed data in an LDA topic model and a total of 50 topics are inferred to create the per-note topic proportion matrix $q$.

Depending on the model and time window being evaluated, subsets of the feature matrix $v$ and matrix $q'$ are combined into an Aggregated Feature Matrix.

CONTEXTUAL EVALUATION OF PROBLEM-LED MODELLING FRAMEWORKS
Think deeply about the clinical context. Find solutions which are specific to the problem.

Good science is about merging different schools of thought for developing the bigger picture.

Data driven approach + Domain Knowledge = Problem-led approach with the patient at the centre

Problem-led vs Data-driven Health

Danielle Belgrave, John Henderson, Angela Simpson, Iain Buchan, Christopher Bishop, and Adnan Custovic.
*Disaggregating asthma: Big investigation versus big data.* Journal of Allergy and Clinical Immunology 139, no. 2 (2017): 400-407.
1. **Team Science**: Discoveries about healthcare, not hypothesised a priori, have been made by experts explaining structure learned from data by algorithms tuned by those experts.


3. An ML approach to extracting knowledge from information in healthcare requires persistent integration of Data, Methods, Expertise.
Problem-Led Patient-Centred Research
THANK YOU
ML STRATEGIES FOR HEALTHCARE PERSONALISATION

Konstantina Palla
WHAT IS PERSONALISED HEALTHCARE

• Traditionally -> personalised medicine

“use of individual’s genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease.”

[National Human Genome Research Institute]
BUT GENOMICS IS NOT ENOUGH

Patient ≠

Factors of disease heterogeneity:

Genomics
Behaviour
Prior exposures
Comorbidities
Etc.

We need to be able to capture this variability → individualised support provision
WHAT IS PERSONALISED HEALTHCARE

Person in the centre.

Person as unique individual.

Electronic Health Records (EHR)

Lifestyle

Genomics

Behaviour

Environment

Provision of Prognosis, Diagnosis, Treatment tailored to the individual
PERSONALISED HEALTHCARE — HOW CAN ML HELP?

ML can transform data into actionable information

Medical Features

Diag
Diabetes

Blood pressure

Heart Rate

Insulin

Leg amputation

EHR

Medical History

How can we extract useful knowledge?

Inspired by [Lee et al., 2017]
PERSONALISED HEALTHCARE — HOW CAN ML HELP?

Data

ML algorithms

Prognosis, Diagnosis, Treatment...

Learn from the population

Tailor to the individual

Is this therapy going to work for me?
ML FOR HEALTHCARE PERSONALISATION

How to structure the talk?

• Explain the most popular techniques
  One click away

• Categorize
  Type of data
    Supervised-unsupervised techniques
  Task
    Diagnosis, prognosis, classification etc.
  Other …
ML FOR HEALTHCARE PERSONALISATION

How to structure the talk?

Let the problem guide us.

Disclaimer: The choice of works presented in this tutorial is by no means an indication of preference or superiority of the method.
ML FOR HEALTHCARE PERSONALISATION

“Need to understand the patient condition, its dynamics and provide optimal patient treatment.”

“Need to understand the patient condition, its dynamics and provide optimal patient treatment.”
ML FOR HEALTHCARE PERSONALISATION

“Need to understand the patient condition, its dynamics and provide optimal patient treatment.”

Model - free approaches
Adapt to the intrinsic data characteristics
No (or few) assumptions - > they don’t explain how the data was generated.

Choice:

- As a first step towards understanding
- Familiarity of the user with the algorithm
- Availability of the corresponding software implementation
Clustering

- They force the pattern to be captured
- They don’t explain was the data was generated
  - Focus on the data, not on the process

Patient Vector of symptoms

Has the disease or not

Neural Network

input layer  hidden layer 1  hidden layer 2  output layer
Autism Spectrum disorders (ASDs): a developmental disorder that affects communication and behaviour.

Spectrum: Wide variation in the type and severity of symptoms (heterogeneity)

- Motivation: Classifying patients into similar groups would provide a powerful tool to individualise treatment regimes

Work by [Doshi et al., 2014]
MODEL-FREE APPROACH — APPLICATION ON ASD

• ASD and Comorbidities

A disease or a syndrome that co-occurs with the target disease

Investigate the patterns of co-occurrence of medical comorbidities in ASDs.

Gastrointestinal disorders
Epilepsy
Sleep disorders
Muscular dystrophy
Psychiatric illnesses
...

Work by [Doshi et al., 2014]
MODEL-FREE APPROACH — APPLICATION ON ASD

Patients: ~ 5K Children

**Data**: Comorbidity counts over period 0-15 year split in 6 month window and for 45 comorbidities.

**Method**: Unsupervised clustering

Patient vector \[\ldots [\_,-,-,\ldots,-,-,\ldots,-,-,\ldots,-,-,\ldots,-,-,\ldots,-,-] \ldots] \]

45 comorbidities

D = 1350

Work by [Doshi et al., 2014]
Results:
Three distinct subgroups were identified

Better understanding of co-occurrence of comorbidities in ASDs
A first step for uncovering underlying etiologies

Similar work on Diabetes type 2 by [Ahlqvist et al, 2018]

Work by [Doshi et al., 2014]
ML FOR HEALTHCARE PERSONALISATION

“Need to understand the patient condition, its dynamics and provide optimal patient treatment.”

Model - based approaches
+ probabilistic framework
What is a model?

Definition [Bishop et al., 2015]

“A set of assumptions about a problem domain expressed in a precise mathematical form, that is used to create a ML solution”

A set of assumptions (defined by the user) to describe how the observed data is generated.
MODEL-BASED APPROACH

A set of assumptions (defined by the user) to describe how the observed data is generated.

Assumptions
our believes of how the data is generated
(latent mechanism responsible for the obsv)

Observed data
(clinical findings)

Graphical model

Model:
set of vars dependencies
Tailored to the data
One of possibly many
MODEL BASED APPROACH - UNCERTAINTY

Uncertainty in many forms

Model

Value of latent parameters

Observations (noise)

Probability theory to express all forms of uncertainty
MODEL BASED APPROACH - UNCERTAINTY

Probability distributions to represent all the uncertain unobserved quantities

Prior belief

\[ p(x) \]

and how they relate to the data

\[ p(y|x) \]

Generative process

\[ x \sim p(x) \]
\[ y \sim p(y|x) \]
MODEL BASED APPROACH - LEARNING

Learning: infer the value of the unknown quantities.

Posterior: Our updated belief after having seen the data

Bayes’ Rule

\[
p(x|Y) = \frac{p(Y|x)p(x)}{p(Y)}
\]

\[
p(x|Y) \propto p(Y|x)p(x)
\]
MODEL BASED APPROACH - EXAMPLE

• Motivation:
  Heterogeneity in complex diseases (chronic). Scleroderma.

• Target:
  Predict future disease trajectory

• Challenge:
  Individualize prediction by capturing variability

Work by [Schulam et al., 2015]

Trajectory of lung severity over time
MODEL BASED APPROACH — INDIVIDUALISED DISEASE PROGRESSION MODEL

• Assumptions: 4 factors of variability
• Model:
  Multi-level model (Latent variable model) - organise variability in different levels

\[ y_{ij} \sim \mathcal{N} \left( \Phi_p(t_{ij})^T \Lambda \bar{x}_{ip} + \Phi_z(t_{ij})^T \beta_{z_i} + \Phi_{\ell}(t_{ij})^T b_i + f_i(t_{ij}) , \sigma^2 \right) \]

[Schulam et al., 2015]
MODEL BASED APPROACH — INDIVIDUALISED DISEASE PROGRESSION MODEL
## Model-Free vs Model-Based Approach

<table>
<thead>
<tr>
<th>Model-free</th>
<th>Model-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>➢ Learn pattern in the data - no assumptions</td>
<td>➢ Model assumptions</td>
</tr>
<tr>
<td>➢ Give insight - can be used as first step</td>
<td>➢ Allow for human-led exploration.</td>
</tr>
<tr>
<td>➢ Easy to use - off the shelf</td>
<td>➢ Perfect fit for probabilistic framework - uncertainty</td>
</tr>
<tr>
<td>➢ Hard to match the requirements of a new application.</td>
<td>➢ Try many different models to find the best</td>
</tr>
</tbody>
</table>
ML STRATEGIES FOR HEALTHCARE PERSONALISATION

ML for personalised treatment
ML FOR PERSONALISED TREATMENT

What treatment should I give to patient?

Ideally, we want to be confident answering this.

Rephrase:
We are interested in the causal influence of treatment A and B on the patient.
Randomized Control Trials

“Gold standard”

Evaluate average treatment effect

Control & Manipulation

A

B

BUT:

• Impractical
  • Expensive (recruiting is hard!)
  • Take time
  • Unethical
  • Does inhaling asbestos cause cancer?

• Not personalised – only population effect
ML FOR PERSONALISED TREATMENT

[Absent controlled experiment, Observational data are used]

Cheaper, Faster, in Plethora

Limitations

• Human population high heterogenous
• Doesn’t contain all possible outcomes for all treatments for a patient
• Data might be biased - unknown underlying data collection protocol

How can ML be applied on Observational data to facilitate personalised treatment?
ML FOR PERSONALISED TREATMENT

Pneumonia example [Caruana et al., 2015]

Machine Learning to guide the treatment of pneumonia patients

What the model inferred: Asthmatic patients have less risk of death!
ML FOR PERSONALISED TREATMENT - COUNTERFACTUALS

Problem: Evaluate individual Treatment effects using observational data

Assume: $Y_i^{(A)}, Y_i^{(B)}$ outcome after the patient $i$ is given treatment $\{A, B\}$.

Challenge:

• Evaluate Treatment effect for a patient $Y_i^{(A)} - Y_i^{(B)}$ using observational data - “What if?”

• BUT: For every subject we only observe one outcome

Never observe the counterfactual.

FACTUAL

Observed patient response to treatment A

What would the outcome be if the patient was given treatment B?
ML FOR PERSONALISED TREATMENT - COUNTERFACTUALS

Idea: Compute distribution over counterfactuals.

How: Multi-task learning problem

\[ X_i \rightarrow \text{Multi-task model} \rightarrow Y_i^{(A)}, Y_i^{(B)} \sim GP(0, K) \]

Multi-task Gaussian Process

[Alaa et al, 2017]
ML FOR PERSONALISED TREATMENT - COUNTERFACTUALS

[Alaa et al., 2017]

The Bayesian framework provides estimates of the Individualised Treatment Effect through the posterior counterfactual distribution.

Other works in Counterfactual reasoning:

[Johansson et al., 2016]
ML STRATEGIES FOR HEALTHCARE PERSONALISATION

ML for mHealth
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

Machine Learning

Actionable information (intervene)

Improve health

personalised

f(accelerometer, GPS, gyroscope, magnetometer, microphone) = ⚽
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

• Intervention app - Fundamental pattern that repeats
  1. at a given time point do
  2. mobile phone collects data (the context)
  3. a decision rule (or policy) maps the data into an intervention option (the action)
  4. mobile phone records the outcome (interpreted as a reward, so higher is better)
  5. done

Intervention options:
- Text messages for walking
- Going to the gym
- Summary of past workouts etc.

GPS
accelerometer
Agenda
Weather etc.

Minutes of activity
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

a decision rule (or policy) maps the context into an intervention option (the action)

Reinforcement learning framework + contextual bandits

Exploration - Exploitation

Personalised action
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

Context

Decision Rule (RL)

Interventions

Rewards

Action: Intervention

Outcome - Reward
Encouraging physical activity of diabetes patients [Hochberg et al., 2016]

An intervention app to encourage physical activity

Approach: Encourage physical activity through personalised messages
Method: RL with contextual bandits
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

Negative feedback

Positive feedback relative to self

Positive feedback relative to others
**MOBILE HEALTH AND PERSONALISED INTERVENTIONS**

- User vector X
  - Augment with action vector A
  - Predict respective rewards Y of [X, A]
  - Exploit and explore - choose message (action)
  - Send message

- Each person modelled through several aspects
  - Personalised messages

- Allows information from past (historical)

- minutes of activity in the last day
- Cumulative number of minutes of activity this week
- Fraction of activity goal

- Age
- Gender
MOBILE HEALTH AND PERSONALISED INTERVENTIONS

27 patients were recruited for a period of 6 months each, 1/3 served as controls
Questions to consider:

When to send the interventions?

➢ Just-In-Time-Adaptive-Interventions (JITAIIs)  
[Inbal et al., 2016]

Need to understand the user

➢ Psychologists, Behavioural scientists, HCI experts.

Need synergy of sciences
HEALTHCARE PERSONALISATION AS A THREE LEVEL PROCESS

Interconnected parts

Increased awareness at every level
LIFECYCLE OF INTELLIGENCE

[Bishop et al, 2015]
MANY THANKS TO

• Danielle Belgrave
• Zoubin Ghahramani
• Allan Tucker
• Neil Lawrence
• Sebastian Nowozin
• Aditya Nori

To all of you!
BIBLIOGRAPHY


BIBLIOGRAPHY

Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Annals of Behavioral Medicine, 2016. accepted subject to revisions.

MODEL BASED APPROACH — INDIVIDUALISED DISEASE PROGRESSION MODEL

<table>
<thead>
<tr>
<th></th>
<th>Pr(•)</th>
<th>Pr(○)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.57</td>
<td>0.18</td>
</tr>
<tr>
<td>2</td>
<td>0.71</td>
<td>0.21</td>
</tr>
<tr>
<td>4</td>
<td>0.60</td>
<td>0.39</td>
</tr>
</tbody>
</table>
Model Based Approach - Learning

Model + Inference = Machine Learning algorithm

Flexibility
- Different inference algorithms can be run on the same model

Consistency
- You can create one model and query it in different ways

Maintainability
- If you want to refine the assumptions encoded in the model, the clean separation makes it straightforward to update it.
Future challenges for ML in healthcare

Lamiae Azizi

University of Sydney

10 July 2018
Plan

1. Developing the unified framework
   - Encoding the expert knowledge
   - Equipping the machinery with causal reasoning
   - Learning algorithms for complex structures

2. Rigourous Framework for trusting the model for deployment?

3. From research to clinical implementation
Unified framework: Pillars

- Health Knowledge
- Causal reasoning and discovery
- Efficient Learning algorithms

Models for personalisation

L. Azizi (University of Sydney)
Future challenges
10/07/2018
Health Knowledge
Saria, 2014
Health knowledge

Data Challenges

- Integrating multi-sources high dimensional data
- Unstructured observational data sources
- Missingness in data sources

Technical Challenges

- Approaches to integrate heterogeneous data
- Flexible and rich way of modelling
- Approaches to incorporate Mechanisms

Current approaches are not enough
Health knowledge: Integration

- Graphical model: Natural to encode domain specific relationships

But for personalisation

Can we "even" integrate the various sources of knowledge?!

Lack of attention can lead to erroneous behavior
Health knowledge: Integration

- Sources are more trusted than others
- Source Misspecification contaminate estimation and update

Principled criteria

Modular vs Full approaches combining various sources
Integrating "omics" and clinical

1. Combining "omics" still in its infancy
2. EHR is uniquely positioned to aid when coupled with "omics" data

No platform for EHR standardisation and "omics" translation

Holistic system view of patient

Combining genotype-phenotype, social and environmental
Health knowledge: "Messiness" challenge

- Continuous temporal measurements, images or text

Novel richer and flexible approaches

- Accurate for longitudinal data: inhomogeneous time series
- New memory models
  - Not evenly spaced
  - Cover long durations
  - Early events affect patient many years later
Health knowledge: "Missingness" challenge

- Received little attention in ML
- Sources of Missingness need to be understood
- Modelling the Missingness mechanisms

Ignoring Missingness $\rightarrow$ lead to incorrect results

Unified framework
Approaches accommodating various mechanisms for various sources
Causal reasoning and discovery
Causality reasoning

- Most ML techniques lack cause-effect reasoning
- Next-generation health data: observational

Challenges for personalisation

Reasoning about learning from data through the lens of "causal models"
- Strong assumptions
- Encoding assumptions in a compact and usable form

Not a trivial matter!
Causal discovery

- Unsupervised learning of causal relationships
- Estimate the causal structure under assumptions

Challenging but promising
Counterfactual reasoning, Pearl 2018

- "Learning Machines can not answer questions about interventions non encountered"
- "Most do not provide a representation from which answers can be derived"
Counterfactual reasoning

- If system optimises property of the observed data:
  - Back to association level ➔ No answer to “what if”
- Complex objective functions are not an answer

Schulam et al, 2017

- **Situation**: Drug given to sicker patients
- **Outcome**: Patients die
- **Model**: Predicts drug kills patient (even beneficial)

Bias in the treatment policy is not considered

Approaches from observational data that can make Counterfactual predictions of outcomes if an experiment run
Counterfactual reasoning

Schulam et al, 2017

- Potential outcomes framework: outcomes under different actions
- Equate to counterfactual models under hypothetical interventions

Promising early results in ML for healthcare
Efficient Learning algorithms
Learning algorithms

- Efficient Learning algorithms:
  - Robust approximation
  - Scalable algorithms
  - Adaptive continually learning

Less challenging for ML community!
Integrative system

Models for personalisation
Plan

1. Developing the unified framework

2. Rigourous Framework for trusting the model for deployment?
   - Interpretability
   - Fairness
   - Transparency, Testability and Validation

3. From research to clinical implementation
Interpretability: What does it mean?! 

Various definitions

- Causal inference models: Interpretable models
- Feature space minimisation
- Model regularisation
- Post-hoc analysis

Interpretable models: More likely to be adopted by medical practitioners
Interpretability vs Justification?!

- Explaining a prediction vs path to the prediction explained

Ghassemi et al, 2018 and Ribeiro et al, 2016

- Identifying data points most responsible for prediction
- May help with security concerns

⚠️ counter-intuitive to privacy concerns

”Justifiability” tools for the unified framework needed
Data quality and model choice encode unintentional discrimination.
Learning from existing clinical practice can amplify the bias.

**Systematic disparate**

Need for systems that can alert to such unwanted behaviours.

Algorithmic fairness still in its infancy.

**Fair model**

Errors are distributed similarly across protected groups, as measured by a cost function.
Chen et al, 2018

- Fairness in prediction of an outcome $Y$
- Predictions are based on a:
  1. Set of covariates $X$: medical history of a patient in a critical care
  2. A Protected attribute $A$: self reported ethnicity

Which Fairness criteria and what cost
Transparency, testability and validation

- **Transparency**: Whether assumptions are plausible or more needed
- **Testability**: Whether assumptions are compatible with data
- **Meaningful validation criteria**: Moving beyond the current performance measures

> Novel criteria for validating models and assumptions
Plan

1. Developing the unified framework

2. Rigorous Framework for trusting the model for deployment?

3. From research to clinical implementation
Learning deployment

- Training on large dataset and assume deployment
  - Stops learning once produced
- Patient populations, recommended treatment procedures change
  - Statistical Target changes
- Performance degradation

Learning approaches, Ghassemi et al 2018

- Robust to changes
- Continually update

Need to be considered early in systems design
Generalisability

- No guarantee for a model learned on one hospital to generalise to a new one
- Infrastructure varies across sites and health systems

ML opportunities

- Data normalisation
- Data collection at different sites

Generalisability not only a modelling challenge
Clinician-Machine Interaction

Detecting individuals at risk early ≠ treating them early

- Systems that interact and collaborate with clinicians
- Leverage strengths of physicians and learning systems
- Having the patient and institutional preferences part of the model?!

Increase of trust and adoption in clinical decision support

Systems allowing for iterative feedback implementation
Beyond modelling and decision support

- Augmenting Data from RCTs with observational EHD
  - New therapies and practice guidelines
- Novel adaptive trial designs
  - Reducing the cost of developing new therapies
- Learn who is most likely to benefit from available resources
  - Optimizing the allocation of limited resources
Future challenges
References


Thank you!