Machine learning for clinical risk prediction

Statistical and Mathematical Modeling Working Group

Jon Arni Steingrimsson
jon_steingrimsson@brown.edu

What is Machine Learning?

No general agreement on a definition.

- "Machine Learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world." – Nvidia
- "Machine learning is the science of getting computers to act without being explicitly programmed." – Stanford
- "Machine learning is based on algorithms that can learn from data without relying on rules-based programming."- McKinsey & Co.
- "Machine learning algorithms can figure out how to perform important tasks by generalizing from examples." – University of Washington
- "The field of Machine Learning seeks to answer the question "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?" – Carnegie Mellon University

[1] techemergence.com/what-is-machine-learning/

Common Supervised Machine Learning Models

Include:

- Support Vector Machines
- Gradient Boosting
- Random Forests
- Regression Trees
- Neural Networks
- Deep Learning

Difference between Machine Learning and Statistical Models?

> *The short answer is: None. They are concerned with the same question: how do we learn from data?*
>
> Larry Wasserman

Machine Learning Models

Compared to traditional statistical models:

- The primary goal is prediction usually
- Black Box vs interpretation and scientific insight
- Less focused on formal statistical inference.
- Care less about uncertainty estimation, or biases
- Different starting points
- Often more complex (require estimating a larger number of parameters)
- Sometimes require less input from user
- Different programming languages (R vs Python)

Success Stories of Machine Learning

DEEPMIND GO CHALLENGE GOOGLE TECH REPORT

23 COMMENTS

Why is Google's Go win such a big deal?

How AlphaGo is cracking an ancient Chinese board game

By Sam Byford on March 9, 2016 09:39 am Email @345triangle

[1] https://www.theverge.com/2016/3/9/11185030/google-deepmind-alphago-go-artificial-intelligence-impact

Success Stories of Machine Learning

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman   Ming Yang   Marc'Aurelio Ranzato   Lior Wolf

Facebook AI Research
Menlo Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

Success Stories of Machine Learning

Updating Google Maps with Deep Learning and Street View

Wednesday, May 3, 2017

Posted by Julian Ibarz, Staff Software Engineer, Google Brain Team and Sujoy Banerjee, Product Manager, Ground Truth Team

Every day, Google Maps provides useful directions, real-time traffic information and information on businesses to millions of people. In order to provide the best experience for our users, this information has to constantly mirror an ever-changing world. While Street View cars collect millions of images daily, it is impossible to manually analyze more than 80 billion high resolution images collected to date in order to find new, or updated, information for Google Maps. One of the goals of the Google's Ground Truth team is to enable the automatic extraction of information from our geo-located imagery to improve Google Maps.

In "Attention-based Extraction of Structured Information from Street View Imagery", we describe our approach to accurately read street names out of very challenging Street View images in many countries, automatically, using a deep neural network. Our algorithm achieves 84.2% accuracy on the challenging French Street Name Signs (FSNS) dataset, significantly outperforming the previous state-of-the-art systems. Importantly, our system is easily extensible to extract other types of information out of Street View images as well, and now helps us automatically extract business names from store fronts. We are excited to announce that this model is now publicly available!

Weaponised AI is coming. Are algorithmic forever wars our future?

Ben Tarnoff

The US military is creating a more automated form of warfare - one that will greatly increase its capacity to wage war everywhere forever.

"This is where AI or, more precisely, machine learning comes in. Machine learning can help automate one of the more tedious and time-consuming aspects of the forever war: finding people to kill."

Weaponised AI is coming. Are algorithmic forever wars our future?

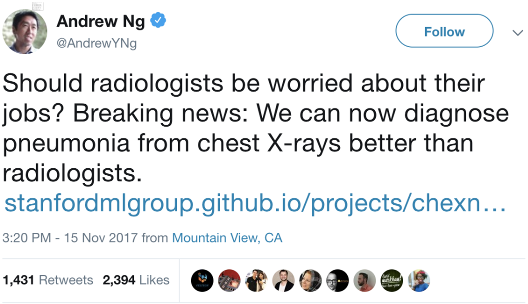
Ben Tarnoff

The US military is creating a more automated form of warfare - one that will greatly increase its capacity to wage war everywhere forever.

"So far, its been a big success: the software has been deployed to as many as six combat locations in the Middle East and Africa. The goal is to eventually load the software on to the drones themselves, so they can locate targets in real time."

Can it be used to save lives?

Lot of Hype


Andrew Ng
@AndrewYNg
Follow

Should radiologists be worried about their jobs? Breaking news: We can now diagnose pneumonia from chest X-rays better than radiologists. stanfordmlgroup.github.io/projects/chexn…

3:20 PM - 15 Nov 2017 from Mountain View, CA

1,431 Retweets **2,394** Likes

Machine Learning in Medicine

HHS Public Access

Author manuscript

Circulation. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Circulation. 2015 November 17; 132(20): 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593.

Machine Learning in Medicine

Rahul C. Deo, MD, PhD

Cardiovascular Research Institute, Department of Medicine and Institute for Human Genetics, University of California, San Francisco, and California Institute for Quantitative Biosciences, San Francisco, CA

"...although there are thousands of papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care. This lack of impact stands in stark contrast to the enormous relevance of machine learning to many other industries."

FDA approves a machine learning based prediction model

Why?

1. Messy data
    - Missing data
    - Measurement error
    - Not collected for research purposes
2. Low signal to noise ratio.
3. Health systems change over time
4. Generalizability concerns.
5. Lack of face validity
6. Not focused on answering causal questions

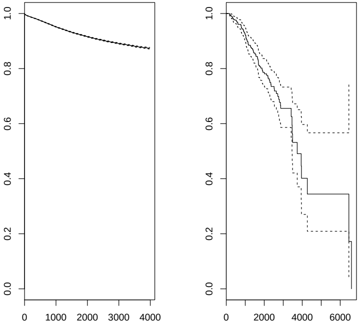
A Review of risk prediction models for medicine

Table 2. Challenges of EHR Data

Longitudinal data	Missing data	Loss to follow-up/ Competing Event
Not considered ($n = 71$)	Not considered ($n = 49$)	Not considered ($n = 40$)
Peak value ($n = 14$)	Single imputation ($n = 16$)	In hospital event ($n = 33$)
Mean/Median ($n = 7$)	Multiple imputation ($n = 21$)	Time to event model ($n = 22$)
Count ($n = 11$)	Complete case ($n = 10$)	Linked to registry ($n = 7$)
Variability ($n = 2$)	Missing category/ indicator ($n = 12$)	Remove those lost ($n = 5$)
Trend ($n = 3$)	Drop variable ($n = 3$)	
Time Varying ($n = 9$)		

[0]Goldstein, Benjamin A., et al. "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review." Journal of the American Medical Informatics Association 24.1 (2017): 198-208

Generalizability



JAMA Network Open

Original Investigation | Health Informatics

Development and Validation of an Electronic Health Record–Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment

Andrew Wong, BA; Albert T. Young, BA; April S. Liang, BSE; Ralph Gonzales, MD, MSPH; Vanja C. Douglas, MD; Dexter Hadley, MD, PhD

Goal

Predict incident delirium risk based on electronic health data available within 24 hours of admission

Setup

- Use 800 variables identified by an expert panel.
- Training set consists of patients discharged between January 1st 2016 and August 31st 2017 ($n_{train} = 14.277$).
- Test set consists of patients discharged between August 1st 2017 and November 1st 2017 ($n_{test} = 3996$).
- Delirium was defined as a positive Nursing Delirium Screening Scale or Confusion Assessment Method for the Intensive Care Unit score.

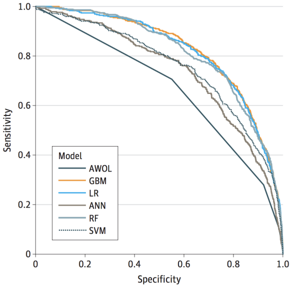
AWOL

UCSF Health currently uses AWOL ≥ 2 screening

- Age greater than 79 years
- Inability to spell world backwards
- Disorientation to city, state, county, hospital name, or floor
- Nurse-rated moderate or severe illness severity

Results

Figure 2. Receiver Operating Characteristic Curves for Machine Learning Models and AWOL



Not everyone liked it!

Gary Collins
@GSCollins

Follow

Thanks @MaartenvSmeden for sending this. The most positive thing I can about this is to say nothing. Truly awful. jamanetwork.com/journals/jaman … #machinelearning #fail help me find something good about this paper…anything I can highlight as good?

What I liked

- They use subject matter knowledge.
- Large sample size
- Use test and training sets
- Code publicly available
- Use multiple evaluation criteria

Some concerns

- All Evaluation measures focus on ordering of predictions (potential lack of calibration).
- Missing data: Handled using indicator variables.
- Generalizable? Disease rate 5% in study vs $11 - 14\%$ in general population.
- Some people contribute more than one observation to the dataset.

Commentary by Dr. Rose

Two main points raised in commentary

- Evaluation Measures
- Generalizability

Evaluation Measures

- Clinician often interested in other evaluation measures than used
- Dr. Rose prefers cross-validation over a single holdout sample

Plug in Estimators Using Machine Learning

Recall that the G-computation formula in it's simplest form is

$$
\begin{aligned}
P(Y|A=a) &= \int P(Y|A=a, X=x)dF(x) \\
&\approx \frac{1}{n}\sum_{i=1}^{n}\hat{P}(Y|A=a, X_i)
\end{aligned}
$$

Need an estimator for $P(Y|A, X)$.

Why not machine learning?

Multiple papers have estimated plug-in estimators using machine learning algorithms. Some show great results, other not so great.

Research Article

Statistics in Medicine

Received 4 April 2009, Accepted 8 October 2009 Published online 3 December 2009 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3782

Improving propensity score weighting using machine learning

Brian K. Lee,[a*†] Justin Lessler[b] and Elizabeth A. Stuart[c,d]

Caveat

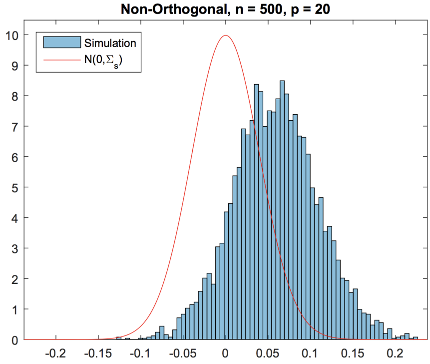
Good prediction performance of the plug-in estimator does not necessarily lead to good inferential performance for the target parameter.

Example taken from Chernozhukov, et al. (2016) : Double machine learning for treatment and causal parameters

Interested in estimating the effect of a treatment (A) after adjusting for some potential confounders W. Consider the partial linear model:

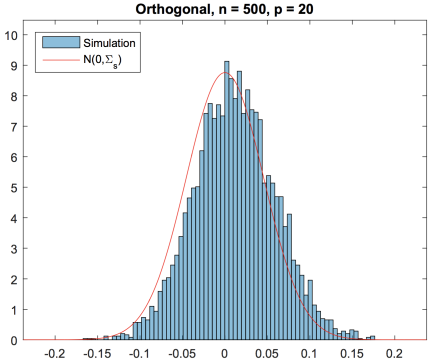
$$Y = \beta_0 A + g_0(W) + \varepsilon, \quad E[varepsilon|A, W] = 0$$

Simulation results when g_0 estimated using RF



[0] Taken from Chernozhukov, Victor, et al. Double machine learning for treatment and causal parameters. No. CWP49/16. cemmap working paper, Centre for Microdata Methods and Practice, 2016.

Results when g_0 estimated using RF with ortogonalization



[0] Chernozhukov, Victor, et al. Double machine learning for treatment and causal parameters. No. CWP49/16. cemmap working paper, Centre for Microdata Methods and Practice, 2016.

Mathematically

$$
\begin{aligned}
\sqrt{n}(\hat{\beta} - \hat{\beta}_0) = {} & \left(\frac{1}{n} \sum_{i=1}^{n} A_i^2 \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \varepsilon_i && (1) \\
& + \left(\frac{1}{n} \sum_{i=1}^{n} A_i^2 \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i (\hat{g}_0(W_i) - g_0(W_i)) && (2)
\end{aligned}
$$

Need to control (2).

Solution proposed in Chernozhukov, et al. (2016)

- Use "orthogonalization" to weaken the condition to $o_P(n^{-1/4})$.
- Use sample splitting to ensure that machine learning models satisfy the required convergence rate.