MESSY DATA AND RELUCTANT USERS - THE TROUBLE WITH HEALTHCARE DATA

Sam Bail @spbail

DataCouncil NYC 2019

HI, I'M SAM!

PhD in semantic web, knowledge representation and automated reasoning

Data Insights Engineer = end-to-end data product development

Spent 5 ½ years at Flatiron Health in NYC analyzing oncology data

- Less big data, more artisanal handcrafted data
- Less data science, more subject matter expertise

Twitter: @spbail

sambail, one word.
@spbail

Healthcare data scientists in academia: "We built a cool model with this data!"

Healthcare data scientists in industry: "Well the data is a homogenous mess and HIPAA makes things hard and you really need domain experts and your users probably won't trust you..."

#MLConfNYC

10:20 PM - 29 Mar 2019

2 Retweets 12 Likes

2   12

OUTLINE

1

----- The vision ---------

2

The problem:
Messy data ----------

3

The other problem:
Reluctant users ----------

4

Paths
forward ---------

1 - THE VISION

I, for one, welcome our robot overlords.

THE AI DOCTOR

Patient — Diagnostics — Lots of data — AI — Treatment — Patient Cured
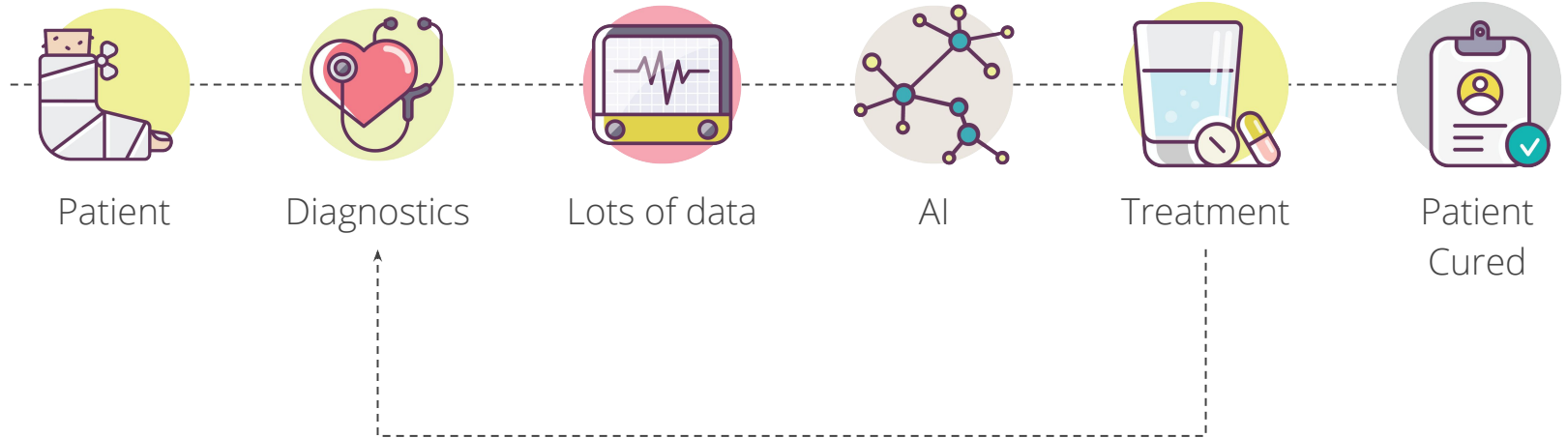
HIGH HOPES

IBM Watson for Oncology is a prominent example of healthcare + AI in recent years

Starting in 2011, over fifty organizations announced Watson collaborations

By 2017, only five projects out of a sample of 24 had been launched

Babylon Health is a patient-facing app that provides an AI chatbot for triaging symptoms
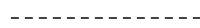
Babylon has two contracts with the NHS in the UK

In 2018, physicians voiced concerns about the accuracy of 10-15% of the bot's diagnoses

Reference [3,4,5]

HEALTHCARE DATA CHALLENGES

Technical challenges -------------- **User acceptance** challenges

2 - THE PROBLEM: MESSY DATA

Healthcare data is hard! Let's go shopping.

"HEALTHCARE DATA"

WORKING DEFINITION:

Any kind of "real-world" data that is generated as part of a patient's and clinician's interaction with data capturing software and medical devices, e.g. medical records, scans, lab and pathology reports, billing records, chat interactions, device data, etc.

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

"Structured": discrete database fields, might still allow free-text
Unstructured: Scanned letters, lab reports, faxes, physician notes

SAMPLE VISIT NOTE

Initial - CCC
Note Date: 11/08/16
Signed by (ORTHOPEDIC SURGEON), MD, PHD on 11/11/16 at 3:32 pm Affiliation:
HOSPITAL
--------------- --------------- --------------- ---------------
Active Medication list as of 11/08/16:
Medications - Prescription
FLUROSEMIDE – 20 mg daily
TYLENOL – OTC as needed
--------------- --------------- --------------- ---------------
This is a first office visit to my clinic by Mr. XXXX, a very pleasant 57-year-old male patient, who sustained in 1993, as the result of a ski accident, a pelvic fracture with vertical shear that has healed in about an inch vertical shortening. Nevertheless, Mr. XXXX has had a remarkably active life. He exercises and has been managing very well over the last few years until recently when he has developed some groin type pain, very reminiscent of arthritic symptoms. Films obtained today confirmed that finding with some bone-on-bone contact and significant posttraumatic hip osteoarthritis.

He actually has a remarkably good gait. He has overall good strength. He has pain along the groin. He has a little bit of anterior medial pain that may be muscular in nature and even though he has a leg length discrepancy, he walks a very normal gait on exam. His extremity appears to be sensory intact and well perfused. He reports the typical symptoms of pain on initiation of motion, winter pain and pain at end of the day.

●●●

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

Patients see multiple clinicians
EHR migrations
Workflow changes

THE PATIENT JOURNEY*

WHAT IS HAPPENING | Tests at PCP, sent to outside lab | Referral to clinic A | More tests and diagnosis | Treatment and recurring tests at clinic A | Patient continues treatment at clinic B | Hospitalization | Referral to hospice

* Heavily simplified and based on what I've seen in oncology - I'm not a doctor!

THE PATIENT JOURNEY*

WHAT WE MAY
SEE IN CLINIC
B'S EHR

Mention in
visit note,
backfilled
data might
be off

Mention in
visit note
(maybe)

Recurring
records and
visit notes

Mention in
visit note
(maybe)

WHAT IS
HAPPENING

Tests at PCP,
sent to
outside lab

Referral
to clinic A

More
tests and
diagnosis

Treatment
and
recurring
tests at
clinic A

Patient
continues
treatment
at clinic B

Hospitalization

Referral to
hospice

Heavily simplified and based on what I've seen in oncology - I'm not a doctor!

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

Data is (physically) hard to access
"No" data model or coding standards
Scaling beyond a single institution is hard

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

Heavy use of acronyms and abbreviations
Sequencing of longitudinal data is hard

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

We can't just store data "in the cloud" (HIPAA* etc)
Linking data sets and mapping entities is limited
Sharing (and validating) data is hard

*Health Insurance Portability and Accountability Act of 1996

SIDEBAR: HOW DID WE GET THERE?

US HITECH ACT 2009: Encourage EHR adoption, but not interoperability

Data was an afterthought - meant for humans to look at ("Glorified paper")

UX was an afterthought - data entry is painful and encourages dictation

No incentive to document anything in structured form if it's not needed for billing

Reference [7,8]

THE TL;DR

Getting **clean** and **reliable** healthcare data as input for any kind of analytical application is hard.

Scaling data access and standardization across the boundaries of a single institution is hard.

3 - THE OTHER PROBLEM: RELUCTANT USERS

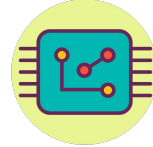Or, "Why Doctors Hate Their Computers"

"DOCTORS HATE THEIR COMPUTERS"

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

"Most days, I will have done only around **thirty to sixty per cent** of my notes by the end of the day"

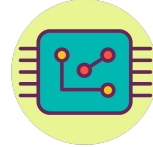*Susan Sadoughi, "Why Doctors Hate Their Computers"*

"DOCTORS HATE THEIR COMPUTERS"

Slow data
entry

Alert
fatigue

Insights and
then what?

Lack of
transparency

"Of roughly 350,000 medication orders per month, pharmacists were receiving pop-up alerts on **nearly half** of them"
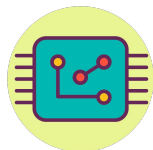
Robert Wachter, "The Digital Doctor"

"DOCTORS HATE THEIR COMPUTERS"

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

"If we use AI to detect more spinal fractures, **we've now shifted the problem** to having to treat more patients"
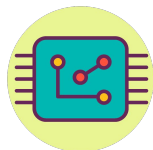
Kerry Weinberg (Amgen), MLConf NYC 2019

"DOCTORS HATE THEIR COMPUTERS"

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

"I would certainly want to see **some validation** to whether the [data] is representative of anything that would make sense"

Dr. Jonathan Chen, "Why Doctors Hate Their Computers"

THE TL;DR

It will take more and continued effort to convince clinicians that computers are **helpful**, not just **painful**.

4 - PATHS FORWARD

Don't give up just yet.

PATHS FORWARD FOR AI + HEALTHCARE DATA*

BE PREPARED

Expect inconsistent data
Build a strong data engineering culture (monitoring, alerting, QA, …) to detect and prevent data issues

Have a Plan B
What if your data source changes, e.g. workflow changes, provider changes…

CLINICIAN-FACING

Practice workflows
Claim denial prediction, clinical trial matching…

Value-based care
Predict and reduce hospitalizations…

Image processing
Annotating and diagnosing scans, e.g. Microsoft InnerEye

PATIENT-FACING

Administrative tasks
Cost and benefit management, scheduling, communication…

Triaging ("digital nurse")
Prevent hospital visits, e.g. Babylon, Sensely

Mental health
Lower barriers and reduce stigma, e.g. Youper, (Talkspace)…

Focused on applications that target clinicians and patients rather than researchers and biased by my own perspective

THANK YOU

Sam Bail @spbail
Data Insights Engineer

REFERENCES

- [1] Shiny moonshot technology will not save healthcare — yet

- [2] What Is the Role of Natural Language Processing in Healthcare?

- [3] How IBM Watson Overpromised and Underdelivered on AI Health Care

- [4] IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show

- [5] This Health Startup Won Big Government Deals—But Inside, Doctors Flagged Problems

- [6] Augmenting Mental Health Care in the Digital Age

- [7] Why Doctors Hate Their Computers

- [8] The Digital Doctor (excerpt here)

- [9] An Ingenious Approach To Designing AI That Doctors Trust

- [10] Dr Murphy on Twitter

- [11] Care.data and access to UK health records: patient privacy and public trust

- Thanks to Lucy Bridges (@linuxlucy) for a detailed overview of data flow in the NHS.