Concept: Crowdsourced Machine Learning

Definition:

Apply ML to user input into a productive application to generate data that is directly helpful to another cause.

Proof of Concept

Case Study: Duolingo





Case Study Duolingo

1

Quality Control

Human Element

- Linguists at the back-end of Duolingo
- Check flags approve, reject
- Respond to forum topics and threads

Machine Element (hypothesis)

 'Correctness' / 'Accuracy' scores for individual users

Hypothesis

2

- Weigh flags and new translations by trustworthiness of users
- Get frequency / common-ness of different translations by averaging across occurrence in users

Learning

 Predict possible answers for new content, grade accordingly, and learn from flags / back-end linguists

Industry Examples

Case Studies: Human DX, Cancerbase

Case Study: Human DX

Worldwide effort, created with and led by the global medical community, to build an **open intelligence system that maps the steps to help any patient**, with the goal to enable more accurate, affordable, and accessible care for all.





Case Study Human DX

Parsing

Accepted Input (hypothesis)

- Some carefully enumerated questions: phenotypes, observations (free text)
- Images: imaging, photographs
- Procedures / test results
- Diagnoses: ICD codes / free text
- Meant to be easy, painless

Parsing / Cleaning (hypothesis)

- Computer vision, especially on images
- NLP for free text descriptions

Human element

2

 Senior contributors and case writers can comment on responses and speak in a forum

Quality Control

• Contributors can flag each other's answers

Machine element

- Calculate individual user metrics:
- <u>Efficiency:</u> how many chunks of information to get to the right hypothesis
- <u>Accuracy:</u> how deviant the differential hypotheses are from the case-writer / peers
- Weighted average of the above in <u>DAPP</u>

Case Study Human DX

3

Learning

Machine element

- Weigh trustworthiness / learnability of responses by individual user metrics
- Weigh trustworthiness / learnability of cases by case-writer's metrics (e.g. are there flags)

Human element

 Moderated by 'senior contributors' – basically medical personnel who sit in the back-end, serving as 'advisory board' to case responses and case creation

Hypothesis

 Eventually able to <u>create cases of its own</u> by mixing / matching case elements in the database

Hypothesis

4

 Create digital representations of experts / by aggregating across experts diagnoses journeys

Digital Doctors

- CPU participates as a 'contributor' in future cases, receiving comments, feedback from the community
- CPU routinely also checks against case
 answers in database



Case Study Human DX: Digital Doctor

For standalone deployment



Case Study: Cancerbase

For patients, by patients, **a data-sharing initiative** staffed by ~ 20 volunteers to improve cancer outcomes, facilitate innovative cancer research, and allow people with cancer to connect from every corner of the globe.





Case Study Cancerbase

Existing

- <u>Image-to-text:</u> photos of prescriptions, diagnosis photos (e.g. MRIs, X-rays) digitized as strings in a database
- <u>Natural language processing:</u> sentiment analysis of text inputs, social media posts, audio clips

Parsing

 <u>Geolocation mapping</u>: overlaying attending physician and patient locations across the globe for different cancers

Future

- Wearable data analysis
- Social media data analysis

Existing (hypothesis)

· Mortality rates

2

- Treatment success rates
- Prescription use, term of usage, and followon prescriptions

Statistics

• Susceptible demographic analyses

Idea: Find Your Disability

A data-sharing initiative between members of the disabled community to improve efficiency of disability diagnosis, outcomes, facilitate innovative research, and allow people with rare disabilities to connect from every corner of the globe.



Appendix



×	Case D	etail	
45 year old fe urgently to en Abdominal Pa Intended Dx: Chi	male presenting rergency room in, Chills. olangitis	g with William O	sler
Histo	ory	Summary	
()	00		
Cholangitis			aff
Acute hepatitis A			att
Acute hepatitis B			ant
Acute liver failure	e		att
Dengue fever			att
Malaria			att
Pancreatic cance	r :		all
Share	e Case	View Solve	



Dr.

case

62F with hx of hypothyroidism, myomatous uterus, fibrocystic breasts, generalized anxiety disorder who presents with vaginal pain. <u>View case</u>

HOVER OR CLICK TO HIGHLIGHT RELATED ITEMS

COLLECTIVE DIFFERE	NTIAL
1. Vaginismus <u>3 community members</u>	3 ★ 3 💄
2. Medication side effect 2 community members	2 ★ 2 💄
3. Vaginal dryness 2 community members	1 ★ 2 💄
4. Anxiety <u>1 community members</u>	1 📩 1 🛎
5. Hypothyroid <u>1 community members</u>	1 💻
6. pelvic pain <u>1 community members</u>	1 🛓

10110	
Diminished sex drive and both be caused by antide to try a different medicat Can also try vaginal lubri with vaginal atrophy seer	l vaginal dryness can epressants. Would opt ion or stop altogether. cant/steroid to help n on exam.
Dr.	<u>View details</u>
Temporal relationship to	worsening
Dr.	View details
Underlying anxiety and n	ormal PE findings
Dr.	<u>View details</u>
post menopausal etiolog warrants, can look into ST less likely	y most likelyif history FIs in this patient but
Anonymous	View details

RATIONALES.

COLLECTIVE PLAN

1. Comprehensive and CBC <u>1 community members</u>	1 单
2. Ky gell	1 单
3. Pelvic exam <u>1 community members</u>	1 单
4. Pelvic floor muscle <u>1 community members</u>	1 🛓
5. Trial of estrace cream to see if vaginal atrophy contributing. If not improving, consider referral to GYN for biofeedback and pelvic PT. <u>1 community members</u>	1 🛓
6. Urinalysis (UA) <u>1 community members</u>	1 单
 pelvic u/s and trial of estrogen vaginal cream 	1 🛓

Total number of Community Members who provided a given plan.

- La Total number of Community Members with a given diagnosis anywhere on their differential.
- Number of Community Members with a given diagnosis at the top of their differential.



Individual metrics



- Non-case
 content
- Created as 'morning report' content
- Not sure what placing is in the whole Human DX database

Basic Questions for the Clobal Map The more data we all contribute, the better the answers will be, but only answer what you feel comfortable with.
Are you entering data for yourself or someone else? Myself Someone else
1. What's the closest city to where you live? Other people have said things like "Fargo, North Dakota" or "Beijing, China".
We'll soon have a map of everyone on Cancerbase. We usually obscure the location you entered to protect privacy, but some peop would like their dot on the map to appear right at the place they entered. What would you like? Obscure location (more private) Put your dot at location you entered (less private)
2. What was your first diagnosis? Other people have said things like "Breast Cancer" or "CLL".
3 When were you horn?
1960
4. When was the cancer diagnosed? Month Vear Vear
5. Are you male or female? Male Female
6. What is your status today? No evidence of disease ▼
7. Did your cancer spread? Ves No

- Option to add more data via 'timeline' option
- Last question is freetext, 1000 characters on your cancer

My Path

Every year, millions of people around the world are diagnosed with cancer, and millions of others have survived cancer. Together, we know a lot about how cancer works, how it travels through the body, and what the treatment options are. What if we could learn from one-another?

We might be able to help you to find others that came just before you and whose cancer was similar. The answers you give below will begin to fill in your cancer timeline. We will try to use that information to find people whose situation is similar to yours.

Your path is the first step. We then would like to give you more information that we have learned from others.

How and when were you treated	?
Other people have said things like	e "Surgery", "Radiation", or "Medication".
Treatment:	
The entry and advect allocates	
Month	
Treatment end date (optional):	
Month 💌 Year 🔹	
Add treatment	
My Timeline	
	Initial diagnosis
	- 2/2014
Click 'agua' ta agua yaur timali	
Click save to save your limeli	
Save my timeline	
Propose a New Qu	estion for CancerBase
What questions do you have? Wh	nat do you care about most? You can suggest a new question or tell us what we can do better.
	1 / 1000
Send us your question private	y l

- Questions are mostly free-text
- Implied existence of NLP parsing algorithms

Privacy Settings

You are what makes CancerBase. CancerBase exists to help you and people like you around the world.

You can help other people who have cancer by sharing your knowledge of cancer. You can help give the world answers to age-old cancer challenges.

If you indicate "Yes", we will keep sharing your de-identified and anonymized data. Your anonymous data lack your name, contact information, and specific location, age, and diagnosis, and therefore do not identify you individually.

Keep sharing my De-identified and Anonymized Data?



O No

Update my sharing preferences

Specifics of the De-identified and Anonymized Data

These are the specific data elements about you that we will share if you want us to:

Sex [Male/Female] Three year age bracket First organ affected Month and year of first diagnosis Location within about 70 miles Metastasis [Yes/No]

- Option to wipe all data from servers
- Default sharing settings are 'no'

