Human-in-a-loop: a design pattern for managing teams that leverage ML

Paco Nathan  @pacoid
Director, Learning Group @ O’Reilly Media
Strata SG, Singapore  2017-12-06
Framing

Imagine having a mostly-automated system where people and machines collaborate together...

May sound a bit Sci-Fi, though arguably commonplace. One challenge is whether we can advance beyond just handling rote tasks.

Instead of simply running code libraries, can machines make difficult decisions, exercise judgement in complex situations?

Can we build systems in which people who aren’t AI experts can “teach” machines to perform complex work – based on examples, not code?
Research questions

- How do we personalize learning experiences, across ebooks, videos, conferences, computable content, live online courses, case studies, expert AMAs, etc.
- How do we help experts (by definition, really busy people) share their knowledge with peers in industry?
- How do we manage the role of editors at human scale, while technology and delivery media evolve rapidly?
- How do we help organizations learn and transform continuously?
UX for content discovery:
- partly generated + curated by people
- partly generated + curated by AI apps
AI in Media

- content which can represented as text can be parsed by NLP, then manipulated by available AI tooling

- labeled images get *really* interesting

- assumption: text or images – within a context – have inherent structure

- representation of that kind of structure is rare in the Media vertical – so far
Transcript: let's take a look at a few examples often when people are first learning about Docker they try and put it in one of a few existing categories sometimes people think it's a virtualization tool like VMware or virtualbox also known as a hypervisor these are tools which are emulating hardware for virtual software

Confidence: 0.973419129848
Knowledge Graph

- used to construct an ontology about technology, based on learning materials from 200+ publishers
- uses SKOS as a foundation, ties into US Library of Congress and DBpedia as upper ontologies
- primary structure is “human scale”, used as control points
- majority (>90%) of the graph comes from machine generated data products
AI is real, but why now?

- **Big Data**: *machine data* (1997-ish)
- **Big Compute**: *cloud computing* (2006-ish)
- **Big Models**: *deep learning* (2009-ish)

The confluence of three factors created a business environment where AI could become mainstream

**What else is needed?**
Background: helping machines learn
Machine learning

**supervised ML:**

- take a dataset where each element has a label
- train models on a portion of the data to predict the labels, then evaluate on the holdout
- **deep learning** is a popular example, but only if you have lots of labeled training data available
Machine learning

**unsupervised ML:**

- run lots of unlabeled data through an algorithm to detect “structure” or embedding
- for example, clustering algorithms such as K-means
- *unsupervised approaches for AI* are an open research question
Active learning

special case of **semi-supervised** ML:

- send difficult decisions/edge cases to experts; let algorithms handle routine decisions (automation)
- works well in use cases which have lots of inexpensive, unlabeled data
- e.g., abundance of content to be classified, where the cost of labeling is the expense
The reality of data rates

“If you only have 10 examples of something, it’s going to be hard to make deep learning work. If you have 100,000 things you care about, records or whatever, that’s the kind of scale where you should really start thinking about these kinds of techniques.”

Jeff Dean  Google  
VB Summit 2017-10-23  
venturebeat.com/2017/10/23/google-brain-chief-says-100000-examples-is-enough-data-for-deep-learning/
The reality of data rates

Use cases for deep learning must have large, carefully labeled data sets, while reinforcement learning needs much more data than that.

**Active learning** can yield good results with substantially smaller data rates, while leveraging an organization’s expertise to bootstrap toward larger labeled data sets, e.g., as preparation for deep learning, etc.
Case studies: practices in industry
On-demand humans
Active learning

*Real-World Active Learning: Applications and Strategies for Human-in-the-Loop Machine Learning*


Ted Cuzzillo
O’Reilly Media, 2015-02-05

Develop a policy for how human experts select exemplars:

- bias toward labels most likely to influence the classifier
- bias toward ensemble disagreement
- bias toward denser regions of training data
Active learning

Active learning and transfer learning
safaribooksonline.com/library/view/oreilly-artificial-intelligence/9781491985250/
video314919.html

Luke Biewald  CrowdFlower
The AI Conf, 2017-09-17

breakthroughs lag algorithm invention, waiting for “killer data set” to emerge, often decade+
Design pattern: Human-in-the-loop

*Building a business that combines human experts and data science*


**Eric Colson**  StitchFix

O’Reilly Data Show, 2016-01-28

“what machines can’t do are things around cognition, things that have to do with ambient information, or appreciation of aesthetics, or even the ability to relate to another human”
Design pattern: Human-in-the-loop

Unsupervised fuzzy labeling using deep learning to improve anomaly detection
conferences.oreilly.com/strata/strata-sg/public/schedule/detail/63304

Adam Gibson  Skymind
Strata SG, 2017-12-07

large-scale use case for telecom in Asia

method: overfit variational autoencoders, then send outliers to human analysts
Design pattern: Human-in-the-loop

_EY, Deloitte And PwC Embrace Artificial Intelligence For Tax And Accounting_

Adelyn Zhou
Forbes, 2017-11-14

- compliance use case: review lease accounting standards
- 3x more consistent, 2x efficient as previous humans-only teams
- break-even ROI within less than a year
Design pattern: Human-in-the-loop

Strategies for integrating people and machine learning in online systems
safaribooksonline.com/library/view/oreilly-artificial-intelligence/9781491976289/video311857.html
Jason Laska  Clara Labs
The AI Conf, 2017-06-29

how to create a two-sided marketplace where machines and people compete on a spectrum of relative expertise and capabilities
Design pattern: Human-in-the-loop

Building human-assisted AI applications
oreilly.com/ideas/building-human-assisted-ai-applications

Adam Marcus  B12
O’Reilly Data Show, 2016-08-25

Orchestra: a platform for building human-assisted AI applications, e.g., to create business websites
https://github.com/b12io/orchestra

example http://www.coloradopicked.com/
Design pattern: Flash teams

*Expert Crowdsourcing with Flash Teams*
Daniela Retelny, et al.
Stanford HCI

“A flash team is a linked set of modular tasks that draw upon paid experts from the crowd, often three to six at a time, on demand”

Weak supervision / Data programming

Creating large training data sets quickly
oreilly.com/ideas/creating-large-training-data-sets-quickly

Alex Ratner  Stanford
O’Reilly Data Show, 2017-06-08

Snorkel: “weak supervision” and “data programming” as another instance of human-in-the-loop
github.com/HazyResearch/snorkel

conferences.oreilly.com/strata/strata-ny/public/schedule/detail/61849
Prodigy by Explosion.ai

https://explosion.ai/blog/prodigy-annotation-tool-active-learning
Problem: disambiguating contexts
Disambiguating contexts

Overlapping contexts pose hard problems in *natural language understanding*. That runs counter to the **correlation** emphasis of big data. NLP libraries lack features for disambiguation.
Disambiguating contexts

Suppose someone publishes a book which uses the term `IOS`: are they talking about an operating system for an Apple iPhone, or about an operating system for a Cisco router?

We handle lots of content about both. Disambiguating those contexts is important for good UX in personalized learning.

In other words, how do machines help people distinguish that content within search?

Potentially a good case for deep learning, except for the lack of labeled data at scale.
Active learning through Jupyter

Jupyter notebooks are used to manage ML pipelines for disambiguation, where machines and people collaborate:

- ML based on examples – most all of the feature engineering, model parameters, etc., has been automated
- [https://github.com/ceteri/nbtransom](https://github.com/ceteri/nbtransom)
- based on use of nbformat, pandas, scikit-learn
Active learning through Jupyter

Jupyter notebooks are used to manage ML pipelines and people collaborate:

- ML based on examples – most of the feature engineering, model parameters, etc., has been automated
- https://github.com/ceteri/nbtransom
- based on use of Jupyter notebooks are used to manage ML pipelines...
- one part configuration file
- one part data sample
- one part structured log
- one part data visualization tool

plus, subsequent data mining of these notebooks helps augment our ontology
Active learning through Jupyter
Active learning through Jupyter

- Notebooks allow the human experts to access the internals of a mostly automated ML pipeline, rapidly
- Stated another way, both the machines and the people become collaborators on shared documents
- Anticipates upcoming collaborative document features in JupyterLab
Active learning through Jupyter

1. Experts use notebooks to provide examples of book chapters, video segments, etc., for each key phrase that has overlapping contexts

2. Machines build ensemble ML models based on those examples, updating notebooks with model evaluation

3. Machines attempt to annotate labels for millions of pieces of content, e.g., `AlphaGo`, `Golang`, versus a mundane use of the verb `go`

4. Disambiguation can run mostly automated, in parallel at scale – through integration with Apache Spark

5. In cases where ensembles disagree, ML pipelines defer to human experts who make judgement calls, providing further examples

6. New examples go into training ML pipelines to build better models

7. Rinse, lather, repeat
Nuances

- **No Free Lunch** theorem: it is better to err on the side of less false positives / more false negatives in use cases about learning materials

- Employ a *bias toward exemplars* policy, i.e., those most likely to influence the classifier

- Potentially, “AI experts” may be Customer Service staff who review edge cases within search results or recommended content – *as an integral part of our UX* – then re-train the ML pipelines through examples
Summary: how this matters
Management strategy – before

Generally with **Big Data**, we are considering:

- DAG workflow execution – which is linear
- data-driven organizations
- ML based on optimizing for objective functions
- questions of correlation versus causation
- avoiding “garbage in, garbage out”
Management strategy – after

**HITL** introduces *circularities*:

- **aka**, *second-order cybernetics*
- leverage feedback loops as conversations
- focus on human scale, design thinking
- people and machines work together on teams
- budget experts’ time on handling the exceptions
Essential takeaway idea:

Depending on the organization, key ingredients needed to enable effective AI apps may come from non-traditional “tech” sources...

In other words, based on human-in-the-loop design pattern, AI expertise may emerge from your Sales, Marketing, and Customer Service teams – which have crucial insights about your customers’ needs.
Summary

Ahead in AI: *hardware* advances force abrupt changes in *software* practices – which has lagged due to lack of infrastructure, data quality, outdated process, etc.

HITL (active learning) as *management strategy* for AI addresses broad needs across industry, especially for enterprise organizations.

**Big Team** begins to take its place in the formula

*Big Data + Big Compute + Big Models.*
Summary

The “game” is not to replace people – instead it is about leveraging AI to augment staff, so that organizations can retain people with valuable domain expertise, making their contributions and experience even more vital.

This is a personal opinion, which does not necessarily reflect the views of my employer.

However, the views of my employer...
Why we’ll never run out of jobs
Strata Data
- SG, Dec 4-7
- SJ, Mar 5-8
- UK, May 21-24
- CN, Jul 12-15

The AI Conf
- CN Apr 10-13
- NY, Apr 29-May 2
- SF, Sep 4-7
- UK, Oct 8-11

JupyterCon
- NY, Aug 21-24

OSCON
- PDX, Jul 16-19, 2018
updates, reviews, conference summaries...

liber118.com/pxn/
@pacoid