Humans in a loop:
Jupyter notebooks as a front-end for AI

Paco Nathan  @pacoid
Dir, Learning Group @ O’Reilly Media
JupyterCon  2017-08-24
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 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 knowledge with their 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?
I want to learn about TensorFlow.

A software library for machine learning, developed by Google.

Similar Searches: deep learning, neural networks, machine learning, ml, scikit-learn.

Also Known As TF.

Supported By Google.

Theme Of The AI Conf.
UX for content discovery:
- partly generated + curated by humans
- partly generated + curated by AI apps
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, test on the remainder
- **deep learning** is a popular example, though 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 calls / edge cases to experts; let algorithms handle routine decisions
- 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
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

_Data preparation in the age of deep learning_

**Luke Biewald**  CrowdFlower

O’Reilly Data Show, 2017-05-04

send human workers cases where machine learning algorithms signal uncertainty (low probability scores)

...or when your ensemble of ML algorithms signals disagreement
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

*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
Reinforcement learning

*Reinforcement learning explained*  
oreilly.com/ideas/reinforcement-learning-explained

**Junling Hu**  Al Frontiers  
O’Reilly Radar, 2016-12-08

learning to act based on long-term payoffs

often as rewards/punishments of an actor within a simulation
Background: helping people learn
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
- 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

orm:Docker a orm:Vendor;
a orm:Container;
a orm:Open_Source;
a orm:Commercial_Software;
owl:sameAs dbr:Docker_%28software%29;
skos:prefLabel "Docker"@en;
Which parts do people or machines do best?

**team goal:** maintain structural correspondence between the layers

**big win for AI:** inferences across the graph

- *human scale*
- *primary structure*
- *control points*
- *testability*

→ *machine generated data products*

~80% of the graph
Ontology

- provides context which Deep Learning lacks
- aka, “knowledge graph” – a computable thesaurus
- maps the semantics of business relationships
- S/V/O: “nouns”, some “verbs”, a few “adjectives”
- difficult work, a relatively expensive investment, potentially high ROI
- conversational interfaces (e.g., Google Assistant) improve UX by importing ontologies
Ontology

Deep learning is a type of machine learning based on a set of algorithms that use multiple processing layers, with different transformations. Deep learning has been particularly successful in neural networks.
Problem: disambiguating contexts
Disambiguating contexts

Overlapping contexts pose hard problems in natural language understanding. Not much available tooling. Counter to the correlation emphasis of big data.
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.
Disambiguation in content discovery

Consider searching for the term `react` on Google. The first page of results:

- acting coaches
- video games
- student engagement
- children’s charities
- UI web components
- surveys
Active learning through Jupyter

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

- notebooks as one part configuration file, one part data sample, one part structured log, one part data visualization tool
- ML based on examples, not feature engineering, model parameters, etc.
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
Active learning through Jupyter
Active learning through Jupyter

- Jupyter notebooks allow 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

Open source \textit{nbtransom} package:

- https://github.com/ceteri/nbtransom
- https://pypi.python.org/pypi/nbtransom
- based on use of \texttt{nbformat} and \texttt{pandas}
- notebook as both Py data store and analytics
- custom “pretty-print” helps with use of \texttt{Git} for diffs, commits, etc.
Nuances

- **No Free Lunch** theorem: it’s better to err on the side of less false positives / more false negatives for this use case.
- Bias toward exemplars most likely to influence the classifier.
- Potentially, the “experts” may be Customer Service staff who review edge cases within search results or recommended content – as an integral part of our UI – then re-train the ML pipelines through examples.
Summary:
people and machines at work
Human-in-the-loop as management strategy

*personal op-ed:* the “game” isn’t to replace people – instead it’s about leveraging AI to augment staff, so organizations can retain people with valuable domain expertise, making their contributions and expertise even more vital.
Why we’ll never run out of jobs
Acknowledgements

Many thanks to people who’ve helped with this work:

- Taylor Martin
- Eszti Schoell
- Fernando Perez
- Andrew Odewahn
- Scott Murray
- John Allwine
- Pascal Honscher
Strata Data

NY, Sep 25-28
SG, Dec 4-7
SJ, Mar 5-8, 2018
UK, May 21-24, 2018

The AI Conf

SF, Sep 17-20
NY, Apr 29-May 2, 2018

OSCON (returns to Portland!)
PDX, Jul 16-19, 2018
updates, reviews, conference summaries...

liber118.com/pxn/
@pacoid