HUMAN-IN-THE-LOOP FROM THE HUMAN PERSPECTIVE

Marti Hearst
UC Berkeley
KDD Dash Workshop, Aug 24, 2020
Pairing People & Algorithms for Data Science

Two Perspectives

**Human-in-the-Loop**

- Goal: improving ML
- Perspective:
  - *Human aids machine*

**Mixed-Initiative Interaction**

- Goal: analysis/exploration
- Perspective:
  - *Machine aids human*
Pairing People & Algorithms for Data Science
This Workshop: Two Perspectives

Human-in-the-Loop

Active Learning Improvement (Ghai et al., Kanchinadam et al.)
Data Augmentation & Model Improvement (Venkataram et al.)
GUI for Annotation (Qian et al., Das & Dutt)

Mixed-Initiative Interaction

GUI & Algorithm for SenseMaking (Bunch et al.)
GUI & Algorithm to Explain Errors (Hanafi et al.)
GUI & Algorithm to Build Better Models (Wang et al.)
Pairing People & Algorithms for Data Science
Two Perspectives

**Human-in-the-Loop**

- Goal: improving ML
- Perspective:
  - *Human aids machine*
Labeling Data: Human vs. No Human

How to get more labeled training data?

Alex Ratner, Stephen Bach, Paroma Varma, Chris Ré

https://www.snorkel.org/blog/weak-supervision#our-ai-is-hungry-now-what
Labeling Data: Human vs. No Human

How to get more labeled training data?

**Semi-supervised Learning:**
Use structural assumptions to automatically leverage unlabeled data

**Weak Supervision:**
Get lower-quality labels more efficiently and/or at a higher abstraction level

**Transfer Learning:**
Use models already trained on a different task

Alex Ratner, Stephen Bach, Paroma Varma, Chris Ré

https://www.snorkel.org/blog/weak-supervision#our-ai-is-hungry-now-what
Labeling Data: Human vs. No Human

Traditional Supervision:
Have subject matter experts (SMEs) hand-label more training data

Too expensive!

Active Learning:
Estimate which points are most valuable to solicit labels for
Labeling Data: Human vs. No Human

How to get more labeled training data?

- **Traditional Supervision:** Have subject matter experts (SMEs) hand-label more training data
  - Too expensive!

- **Semi-supervised Learning:** Use structural assumptions to automatically leverage unlabeled data

- **Weak Supervision:** Get lower-quality labels more efficiently and/or at a higher abstraction level

- **Transfer Learning:** Use models already trained on a different task

- **Active Learning:** Estimate which points are most valuable to solicit labels for

- **Get cheaper, lower-quality labels from non-experts**

- **Get higher-level supervision over unlabeled data from SMEs**

Heuristics, Distant Supervision, Constraints, Expected distributions, Invariances

Alex Ratner, Stephen Bach, Paroma Varma, Chris Ré

https://www.snorkel.org/blog/weak-supervision#our-ai-is-hungry-now-what
Human-in-the-Loop: Role of Humans

Goal: improve training data for ML algorithm

Traditional: People label items:
- Category
- Relevance / Ranking
- Span
Human-in-the-Loop: Approaches

Approaches:

• **Naïve**: Humans label lots of items

• **Active Learning**: Humans label *strategically* selected items

• **Smart UIs**: Reduce labeling effort, geared toward human actions / cognition
OUTLINE

Two Perspectives

Human-in-the-Loop
  Two interesting examples

Mixed-Initiative
  Two interesting examples
  Trust issue: Data Science

Conclusions
Recent innovations give more agency to humans:

• Ask humans to outsmart the algorithm (Nie et al.)
• Ask humans to program patterns (Raskin et al.)
• This workshop:
  • Humans give rationales for features (Ghai et al., Kanchinadam et al.)
  • Humans write queries to ferret out negative examples (Venkataram et al.)
Adversarial Labeling

Today, evaluation sets for ML get “tapped out” quickly
• 15 years for near-human performance on MNIST
• 7 years for ImageNet
• ~1 year for GLUE (combined NLP benchmark)

Why a Problem?
• Algorithms learn biases and tricks
• Training doesn’t really reflect the underlying task

• We need more robust training sets!

Nie et al., ACL 2020
Idea: Ask Humans to Outsmart the Algorithm

• Like adversarial learning, except
  • Instead of an algorithm making the adversarial examples,
  • Humans figure out difficult examples for the model

• A *Dynamic* Benchmark

• End result: more accurate and robust model

Nie et al., ACL 2020
In typical crowd work, humans write examples, perhaps with constraints on novelty.
In typical crowd work, humans write examples, perhaps with constraints on novelty.

Novel aspect: If the model gets the answer right, the crowd worker has to try again and create another sentence.

Nie et al., ACL 2020
Each round results in a new set of data for a new train/dev/test split.

Each round gets increasingly difficult (for human & algorithm) as the models improve.
Humans Write Rules; Algorithms Combine and Fix Inconsistencies

Programmatically building and manipulating the training data — rather than the models — improves ML performance.

Ratner et al., VLDB 2017
Summary: Human-in-the-Loop Trend: Add More Human Initiative
Two Perspectives

Human-in-the-Loop
Two interesting examples

Mixed-Initiative
Two interesting examples
Trust issue: Data Science

Conclusions
Pairing People & Algorithms for Data Science
Two Perspectives

**Mixed-Initiative Interaction**

- Goal: analysis/exploration
- Perspective:
  - *Machine aids human*
Mixed-Initiative Interaction

James F. Allen  
Curry Guinn  
Eric Horvitz

The development of mixed-initiative systems can be traced back to the 1960s through work on the MIP project at the MIT Artificial Intelligence Laboratory. The MIT project was pioneering in that it tried to develop a human-computer system that was both general and that could reflect the perceptual and cognitive processes of humans, even as it expanded to incorporate other forms of human interaction. Although the system was not a success, it did provide a model for future work.

Mixed-initiative systems have been used in a variety of applications, including dialogue management, natural language processing, and machine translation. These systems are designed to work in collaboration with human operators, allowing them to provide guidance and feedback to the system as needed. This allows the system to adapt its behavior to the specific needs of the user, making it more effective than systems that are designed to work in isolation.

The success of mixed-initiative systems has been limited by the difficulty of creating systems that can accurately interpret human intent and behavior. As a result, many of these systems are not able to fully engage in mixed-initiative interaction, and they must rely on more passive forms of interaction.

Despite these limitations, mixed-initiative systems have the potential to be very powerful tools, and they are likely to play an important role in the future of human-computer interaction.
**Mixed-Initiative Definition**

“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

- James Allen
"Methods that explicitly support an efficient, natural \textit{interweaving} of contributions by users and automated services aimed at converging on solutions to problems."

- Eric Horvitz
An assistant for data exploration based on AI planning:

- **An assistant is at least partly autonomous**
  - Makes decisions on how to carry out user guidance

- **An assistant responds to guidance as it works**
  - Its reasoning process must be available to the user to modify

Amant & Cohen, IUI 1997
Mixed-Initiative Interaction: AIDE 1997

Amant & Cohen, IUI 1997
Mixed-Initiative Examples

PixelTone Multimedia Editing (Adar et al.)

Collaborative Search (Pickens et al.)

This workshop:

- GUI & Algorithm for SenseMaking (Bunch et al.)
- GUI & Algorithm to Explain Errors (Hanafi et al.)
- GUI & Algorithm to Build Better Models (Wang et al.)
Mixed-I Example: Multimedia Editing

PixelTone: Laput et al., CHI 2013
“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”

An assistant is at least partly autonomous
  Makes decisions on how to carry out user guidance
An assistant responds to guidance as it works
  Its reasoning process must be available to the user to modify
"A flexible interaction strategy, where each agent can contribute to the task what it does best."

"Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems."

An assistant is at least partly autonomous
- Makes decisions on how to carry out user guidance
- An assistant responds to guidance as it works
- Its reasoning process must be available to the user to modify

PixelTone In Depth

Human: high level design choices
Agent: executes low level details

"increase the contrast on the lower part"
- system knows lower part is ocean

"make it heavenly"
“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”

An assistant is at least partly autonomous
  Makes decisions on how to carry out user guidance
An assistant responds to guidance as it works
  Its reasoning process must be available to the user to modify

Sliders allow user to adjust results of an agent’s action

Gestures by human (“blur in this direction”) augment command.

However, system does not ask for feedback.
PixelTone In Depth

“An autonomous assistant is at least partly autonomous.

- Makes decisions on how to carry out user guidance.
- An assistant responds to guidance as it works.
- Its reasoning process must be available to the user to modify.

Automatically adjusts contrast
- Allows dynamic creation of new concepts and terminology.

- “this is a shirt”
- “change the color of the shirt”
- “this is John”
- “brighten Sara and John”

“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”
Mixed-I Example: Collaborative Search

Goal: allow people to work at their own pace, but be influenced in real time by their collaborators’ work.

“Influence should be synchronized, but workflow should not.”

Pickens et al., SIGIR 2008
Mixed-I Example: Collaborative Search

Pickens et al., SIGIR 2008
Three UIs for Three Roles

Prospector: opens new fields for exploration (breadth)

Miner: explores rich veins of information (breadth)
Three UIs for Three Roles

**Algorithm:** combines work of Prospector and Miner; makes query suggestions and re-ranks results.

**Shared Display:** continually-updated status: relevant documents, past queries, system-suggested search terms.

Pickens et al., SIGIR 2008
Cherchiamo Video

Miner - Prospector Search
“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”

In this case, the humans have two different tasks, and the algorithm has the mediator task.

An assistant is at least partly autonomous
   Makes decisions on how to carry out user guidance
An assistant responds to guidance as it works
   Its reasoning process must be available to the user to modify
“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”

An assistant is at least partly autonomous
  Makes decisions on how to carry out user guidance
An assistant responds to guidance as it works
  Its reasoning process must be available to the user to modify

Search Collaboration In Depth

Interweaving is the focus of this design.
“A flexible interaction strategy, where each agent can contribute to the task what it does best.”

“Methods that explicitly support an efficient, natural interweaving of contributions by users and automated services aimed at converging on solutions to problems.”

An assistant is at least partly autonomous

- Makes decisions on how to carry out user guidance

An assistant responds to guidance as it works

- Its reasoning process must be available to the user to modify

The assistant is autonomous, but its guidance is a black box.
Performing dramatically better on average than merging the results of two searchers, when relevant results are sparse.

Subjective responses not reported.

Pickens et al., SIGIR 2008
Two Perspectives

Human-in-the-Loop
Two interesting examples

Mixed-Initiative
Two interesting examples
Trust issue: Data Science

Conclusions
Human-in-the-Loop

- Goal: improving ML
- Perspective:
  - Human aids machine
- Trust: low importance

Mixed-Initiative Interaction

- Goal: analysis/exploration
- Perspective:
  - Machine aids human
- Trust: high importance
What Kind of Automation Is Acceptable?

Tableau’s “Show Me”

- Clearly Understandable Behavior
- Visible Effects
- Reproducible
- Reversible
- Allows Human to Specify Design, System to Execute Details

However: *not* mixed-initiative
"What will AI allow us to automate? We'll be able to automate everything that we can describe. The problem is: it's not clear what we can describe."

- Stephen Wolfram
FUTZING AND MOSEYING

Interviews with Professional Data Analysts on Exploration Practices

Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Lin, Marti Hearst
UC Berkeley

VAST 2018

Supported by the UC Berkeley AMP Lab and a Gift from Tableau Research
Motivating Question:

Do professional analysts do exploratory data analysis?

If so, why? If not, why not?

If so, what kinds of automated tools do they desire?
Recruiting and Interviewing

Reached out to professional network

“Professionals who analyze data” daily
Indicated that focus was EDA
30 respondents; 90 minutes on avg
# Demographics

<table>
<thead>
<tr>
<th>Role</th>
<th>Specialty</th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>analyst</td>
<td>data science</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>business intelligence</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>machine learning</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>data visualization</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>epidemiology</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>finance</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>software engineering</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>executive</td>
<td>data science</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>grad stu..</td>
<td>combustion</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>neuroscience</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>manage..</td>
<td>data warehousing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>professor</td>
<td>computer networking</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
MAIN FINDINGS

Exploratory activities pervade the entire analysis process

For some analysts

Analysts want a compromise between coding and direct manipulation

Notebooks with interactive visualizations are promising

Skepticism toward automated analysis tools
Homegrown Automation

19. Described tools they had built themselves for repetitive tasks (including wrapping common visualization commands)

3. Wrote code to profile all columns of a data set

1. Wrote their own visualization library
Homegrown Automation

(continued)

12 Copy-paste reuse: many scripts with minor variations, hard to manage

6 Barrier to home-grown automation: difficulty of generalizing solutions so others could use them.

Many other frustrations
Computer Automation?

5  Expressed interest in automated wrangling tools

3  Pointed out that manual wrangling yields valuable insight

7  Suggested tools for automatically profiling data

9  Expressed skepticism
   “the parts that are easy are easy; the hard parts are difficult to automate”
Computer Automation?
(continued)

3 Expressed interest in automated suggestions of interesting relationships

12 Thought that for recommenders to be useful, they must navigate between being a black box and making the user do tedious work.

Summary: Automated DS Tools Not Trusted (yet)
IBM’s AutoAI:
  • Automated support for DS model building

Controlled between-participants study
  • AutoDS participants faster, more accurate, more models
  • Participants in manual condition had higher trust and confidence
WHY IS TRUST LOW?

Automated data science methods. Do they meet:

• Clearly Understandable Behavior

• Visible Effects

• Reproducible

• Reversible

• Allow Human to Specify Design, System to Execute Details
Dialogue for Building Trust

James Allen 1999 essay:
Hoped to use AI planning; this failed
Serious mismatch; humans solve problems differently

Automated planners:
• Full specification & context
• Evaluate quantitatively
• Low communication

Human problem solving:
• Incrementally learn; refine & modify goals
• Evaluate subjectively
• High communication
Dialogue for Building Trust

James Allen 1999 essay:
Mixed-initiative collaboration planning between humans:
Much effort spent in maintaining understanding

<table>
<thead>
<tr>
<th>Action</th>
<th>Amount (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating and comparing options</td>
<td>25</td>
</tr>
<tr>
<td>Suggesting courses of action</td>
<td>23</td>
</tr>
<tr>
<td>Clarifying and establishing state</td>
<td>13.5</td>
</tr>
<tr>
<td>Clarifying or confirming the communication</td>
<td>13.5</td>
</tr>
<tr>
<td>Discussing problem-solving strategy</td>
<td>10</td>
</tr>
<tr>
<td>Summarizing courses of action</td>
<td>8</td>
</tr>
<tr>
<td>Identifying problems and alternatives</td>
<td>7</td>
</tr>
</tbody>
</table>
Dialogue for Building Trust

James Allen 1999 essay:
Mixed-initiative collaboration planning between humans:
Much effort spent in maintaining understanding

<table>
<thead>
<tr>
<th>Action</th>
<th>Amount (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating and comparing options</td>
<td>25</td>
</tr>
<tr>
<td>Suggesting courses of action</td>
<td></td>
</tr>
<tr>
<td>Clarifying and establishing</td>
<td></td>
</tr>
<tr>
<td>Clarifying or confirming</td>
<td></td>
</tr>
<tr>
<td>Discussing problem-solving strategy</td>
<td>10</td>
</tr>
<tr>
<td>Summarizing courses of action</td>
<td>8</td>
</tr>
<tr>
<td>Identifying problems and alternatives</td>
<td>7</td>
</tr>
</tbody>
</table>

We need smarter AI
Two Perspectives

Human-in-the-Loop
Two interesting examples

Mixed-Initiative
Two interesting examples
Trust issue: Data Science

Conclusions
HITL advances give humans more agency

To improve Mixed-Initiative:

• UI design guidelines: visibility, reversibility, reproducibility, etc.
• Enriching the interactivity of the process to model human dialogue
• More advanced AI