

Human-in-the-loop for Machine Learning/Data Analytics Workflows

First Presentation

An Dan Nguyen
dan.nguyen@aalto.fi

Large-scale Computing and Data Analysis
April 2021

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature
- 5 Preliminary Goals
- 6 Some Questions/Concerns

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature
- 5 Preliminary Goals
- 6 Some Questions/Concerns

- Want to investigate human dynamics topics related to Machine Learning/Data Analysis (ML/DA) workflows.

Motivation

- Want to investigate human dynamics topics related to Machine Learning/Data Analysis (ML/DA) workflows.
- Bigger theme of interest is the collaboration between human and machine.

- Want to investigate human dynamics topics related to Machine Learning/Data Analysis (ML/DA) workflows.
- Bigger theme of interest is the collaboration between human and machine.
- Got the interest when briefly participate in a research project about human-in-the-loop in new Air Traffic Management systems.

Motivation

- Want to investigate human dynamics topics related to Machine Learning/Data Analysis (ML/DA) workflows.
- Bigger theme of interest is the collaboration between human and machine.
- Got the interest when briefly participate in a research project about human-in-the-loop in new Air Traffic Management systems.
- Some of the concepts in these systems can be applied in the ML/DA workflows (e.g. trust, level of automation analysis framework [1]).

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature
- 5 Preliminary Goals
- 6 Some Questions/Concerns

Preliminary Plan

- 1 A literature review on current status of human-in-the-loop in ML pipelines and workflow-based data analysis. Can potentially reference other fields if there are not much information [1].

Preliminary Plan

- 1 A literature review on current status of human-in-the-loop in ML pipelines and workflow-based data analysis. Can potentially reference other fields if there are not much information [1].
- 2 Based on the literature review, identify one area to investigate further.

Preliminary Plan

- ➊ A literature review on current status of human-in-the-loop in ML pipelines and workflow-based data analysis. Can potentially reference other fields if there are not much information [1].
- ➋ Based on the literature review, identify one area to investigate further.
- ➌ Identify potential improvement(s) in that area.

Preliminary Plan

- ① A literature review on current status of human-in-the-loop in ML pipelines and workflow-based data analysis. Can potentially reference other fields if there are not much information [1].
- ② Based on the literature review, identify one area to investigate further.
- ③ Identify potential improvement(s) in that area.
- ④ Suggest and develop the concepts for the improvements (a proof of concept as a software may not fit in the scope of this course).

Preliminary Plan

- 1 A literature review on current status of human-in-the-loop in ML pipelines and workflow-based data analysis. Can potentially reference other fields if there are not much information [1].
- 2 Based on the literature review, identify one area to investigate further.
- 3 Identify potential improvement(s) in that area.
- 4 Suggest and develop the concepts for the improvements (a proof of concept as a software may not fit in the scope of this course).
- 5 Write and present results.

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review**
- 4 Notable Literature
- 5 Preliminary Goals
- 6 Some Questions/Concerns

- AutoML Systems [2] [3] [4]
- Human-in-the-loop (HITL) pipeline design, framework, survey, human dynamics
 - ML HITL Pipelines/Frameworks [5] [6] [7] [8]
 - Data HITL Pipelines/Frameworks [9]
 - Survey, book about HITL in ML/DA Pipelines [10] [11] [12] [13]
 - Human dynamics [14]
- Others (opinion, personal perspective) [15] [16]

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature**
- 5 Preliminary Goals
- 6 Some Questions/Concerns

- *A survey on active learning and human-in-the-loop deep learning for medical image analysis*, S. Budd, E. C. Robinson, and B. Kainz

- *A survey on active learning and human-in-the-loop deep learning for medical image analysis*, S. Budd, E. C. Robinson, and B. Kainz
- *Human-in-the-loop techniques in machine learning*, C. Chai and G. Li

- *A survey on active learning and human-in-the-loop deep learning for medical image analysis*, S. Budd, E. C. Robinson, and B. Kainz
- *Human-in-the-loop techniques in machine learning*, C. Chai and G. Li
- *Human-in-the-loop machine learning*, R. Munro

- *A survey on active learning and human-in-the-loop deep learning for medical image analysis*, S. Budd, E. C. Robinson, and B. Kainz
- *Human-in-the-loop techniques in machine learning*, C. Chai and G. Li
- *Human-in-the-loop machine learning*, R. Munro
- *Human-in-the-loop AI: Requirements on future (unified) air traffic management systems*, J. Lundberg et al.

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature
- 5 Preliminary Goals**
- 6 Some Questions/Concerns

Preliminary Goals

- Understand the current state-of-the art in HITL Machine Learning and write a short report and some potential improvements.

Preliminary Goals

- Understand the current state-of-the art in HITL Machine Learning and write a short report and some potential improvements.
- Or compare some systems of HITL Machine Learning (pros and cons).

Preliminary Goals

- Understand the current state-of-the art in HITL Machine Learning and write a short report and some potential improvements.
- Or compare some systems of HITL Machine Learning (pros and cons).
- Or something different.

Table of Contents

- 1 Motivation
- 2 Preliminary Plan
- 3 Literature Review
- 4 Notable Literature
- 5 Preliminary Goals
- 6 Some Questions/Concerns**

Some Questions/Concerns

- The topic seems a little bit theoretical.

Some Questions/Concerns

- The topic seems a little bit theoretical.
- Since time to conduct experiment involving human can be long and may not fit into the scope of the course.

Some Questions/Concerns

- The topic seems a little bit theoretical.
- Since time to conduct experiment involving human can be long and may not fit into the scope of the course.
- Same as a proof of concept in software form.

References I



J. Lundberg, M. Bång, J. Johansson, A. Cheaitou, B. Josefsson, and Z. Tahboub, “Human-in-the-loop ai: Requirements on future (unified) air traffic management systems,” in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. IEEE, 2019, pp. 1–9.






Z. Shang, E. Zraggen, B. Buratti, F. Kossmann, P. Eichmann, Y. Chung, C. Binnig, E. Upfal, and T. Kraska, “Democratizing data science through interactive curation of ml pipelines,” in *Proceedings of the 2019 International Conference on Management of Data*, 2019, pp. 1171–1188.







R. Elshaw, M. Maher, and S. Sakr, “Automated machine learning: State-of-the-art and open challenges,” *arXiv preprint arXiv:1906.02287*, 2019.





References II

-  D. Patel, S. Shrivastava, W. Gifford, S. Siegel, J. Kalagnanam, and C. Reddy, “Smart-ml: A system for machine learning model exploration using pipeline graph,” in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 1604–1613.
-  A. van der Stappen and M. Funk, “Towards guidelines for designing human-in-the-loop machine training interfaces,” in *26th International Conference on Intelligent User Interfaces*, 2021, pp. 514–519.
-  C. Chai, L. Cao, G. Li, J. Li, Y. Luo, and S. Madden, “Human-in-the-loop outlier detection,” in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020, pp. 19–33.

References III

-  D. Xin, L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran, “Accelerating human-in-the-loop machine learning: challenges and opportunities,” in *Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning*, 2018, pp. 1–4.
-  —, “Helix: accelerating human-in-the-loop machine learning,” *arXiv preprint arXiv:1808.01095*, 2018.
-  G. Li, “Human-in-the-loop data integration,” *Proceedings of the VLDB Endowment*, vol. 10, no. 12, pp. 2006–2017, 2017.
-  I. Xanthopoulos, I. Tsamardinos, V. Christophides, E. Simon, and A. Salinger, “Putting the human back in the automl loop.” in *EDBT/ICDT Workshops*, 2020.

References IV

-  S. Budd, E. C. Robinson, and B. Kainz, “A survey on active learning and human-in-the-loop deep learning for medical image analysis,” *Medical Image Analysis*, p. 102062, 2021.
-  C. Chai and G. Li, “Human-in-the-loop techniques in machine learning,” *Data Engineering*, p. 37, 2020.
-  R. Munro, *Human-in-the-loop machine learning*. O’Reilly Media, 2020.
-  D. Honeycutt, M. Nourani, and E. Ragan, “Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy,” in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 8, no. 1, 2020, pp. 63–72.

References V



F. M. Zanzotto, “Human-in-the-loop artificial intelligence,” *Journal of Artificial Intelligence Research*, vol. 64, pp. 243–252, 2019.



O. Bezrukavnikov and R. Linder, “A neophyte with automl: Evaluating the promises of automatic machine learning tools,” *arXiv preprint arXiv:2101.05840*, 2021.