Qeios

Peer Review

Review of: "Undefinable True Target Learning: Towards Learning with Democratic Supervision"

Pedro Antonio Gutiérrez[1]

1. Ciencia de la Computación e Inteligencia Artificial, Universidad de Córdoba, Spain

This manuscript presents a philosophical and conceptual analysis of the role of the True Target (TT) assumption in Machine Learning (ML) paradigms. The author argues that most existing paradigms implicitly assume the objective existence of a TT and proposes a more radical alternative: explicitly assuming that the TT does not objectively exist in the real world. Building on this premise, the paper introduces Undefinable True Target Learning (UTTL) as a novel learning paradigm, positioned as a pathway towards Learning with Democratic Supervision (LDS). The article is ambitious in scope, combining a systematic review of TT assumptions across unsupervised learning, supervised learning, weakly supervised learning, and reinforcement learning with a philosophical reinterpretation of supervision and paradigm design. The manuscript is primarily conceptual rather than empirical, aiming to stimulate reflection on foundational assumptions in ML rather than to propose a concrete algorithmic contribution.

The explicit rejection of the objective existence of a TT is a clear conceptual contribution. While related ideas have appeared implicitly in noisy-label learning and multi-annotator settings, the manuscript is different in elevating this assumption to a philosophical position. The paper demonstrates an impressive command of the literature, particularly in weakly supervised learning, learning from noisy labels, and learning from multiple annotators. The systematic review in Section 2 is detailed and well-structured. The discussion is highly relevant given the growing interest in subjectivity, ambiguity, annotation disagreement, and human-centred ML. The notion of democratic supervision resonates with current debates in responsible artificial intelligence.

My main comments and concerns are the following:

- While UTTL is formally defined, the framework remains largely abstract. The mathematical

formulation introduced in Section 4 does not yet clarify how UTTL differs operationally from existing approaches such as probabilistic label modelling, distributional supervision, multi-view or multi-task learning, or annotator-aware learning frameworks.

- The manuscript would benefit greatly from at least one worked example or case study, even at a conceptual or synthetic level, illustrating how UTTL would be instantiated in practice.

- The term democratic supervision is central to the paper, yet its scope is somewhat unclear. Earlier sections suggest the inclusion of non-experts, while later sections explicitly restrict UTTL to expert democracy (domain experts and ML experts only). The author should clarify whether LDS necessarily implies non-expert participation, or whether expert democracy is merely a simplified instance of a broader LDS framework.

- Although the manuscript argues that UTTL is distinct from LMA and other WSL settings, the differences are sometimes philosophical rather than methodological. From a reader's perspective, it is not always evident what cannot be captured by existing frameworks but can be addressed by UTTL.

- The assertion that the TT "does not objectively exist" is intentionally radical, but the manuscript would be stronger if it explicitly engaged with potential counterarguments. For example, in which domains does the non-existence assumption clearly outperform the existence assumption? Are there risks of relativism or loss of identifiability? How does evaluation proceed when no TT exists even conceptually?

Declarations

Potential competing interests: No potential competing interests to declare.