Proceedings of Machine Learning Research vol 247:1–2, 2024 37th Annual Conference on Learning Theory

Errors are Robustly Tamed in Cumulative Knowledge Processes

Anna Brandenberger ABRANDE@MIT.EDU
Department of Mathematics, MIT
Cassandra Marcussen CMARCUSSEN@G.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University
Elchanan Mossel ELMOS@MIT.EDU
Department of Mathematics, MIT
Madhu Sudan MADHU@CS.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University

Editors: Shipra Agrawal and Aaron Roth

A fundamental question in the study of information and knowledge and their processing is: how robust is the processing to errors? In the well-studied setting of information transmission, this leads to the fields of information and coding theory, and a foundational result here is that a (sufficiently small) constant fraction of errors can be remedied by a constant fraction of additional redundancy. Thus most channels can operate at a positive constant rate, where the effort needed to protect information from errors does not overwhelm the amount of new information being transmitted.

In this work, we explore the analogous question in the context of *cumulative knowledge*, referring to processes of societal knowledge accumulation where the validity of a new unit of knowledge depends both on the correctness of its derivation and on the validity of the units it relies on. Can investing a constant fraction (away from 1) of effort ensure a constant fraction of societal knowledge is valid? Ben-Eliezer et al. (2023) introduced a concrete probabilistic model and showed an affirmative answer to this question. Their study, however, focuses on the simple case where a new unit depends on only one existing unit, and units attach according to a *preferential attachment rule*.

Our main interest in this paper is to study if similar results hold for more general and realistic models. We consider networks where the attachment rules are more flexible than preferential attachment, new units may depend on more than one existing unit, and some of the nodes behave in an adversarial fashion, not only intentionally introducing incorrect knowledge but also strategically connecting to other units. We give a robust affirmative answer to the above question by showing that for *all* of these models, as long as many of the units follow simple heuristics for checking a bounded number of units they depend on, all errors will be eventually eliminated. Our results indicate that preserving the quality of large interdependent collections of units of knowledge is feasible, as long as careful but not too costly checks are performed when new units are derived/deposited. We also study the conditions under which the error effect survives forever.

To prove error elimination and survival, we construct suitable potentials that form super-martingales and sub-martingales, respectively, that can be connected to the absence or presence of errors in the corpus of knowledge. In proving error elimination, this potential depends exponentially on a notion we call the "depth" of a node, which is the minimum distance to a node that introduced a new error. In proving error survival, this potential depends on the number of nodes that introduce new errors and the number of leaf nodes.[1]

1. Extended abstract. Full version can be found at [arXiv:2309.05638, v3].

Acknowledgments

AB was supported in part by Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826 and NSF GRFP 2141064. CM was supported in part by a Simons Investigator Award and NSF Award CCF 2152413 to Madhu Sudan. EM was supported in part by a Simons Investigator Award, Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826, NSF award CCF 1918421, and ARO MURI W911NF1910217. MS was supported in part by a Simons Investigator Award and NSF Award CCF 2152413.

References

Omri Ben-Eliezer, Dan Mikulincer, Elchanan Mossel, and Madhu Sudan. Is This Correct? Let's Check! In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:11, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-263-1. doi: 10.4230/LIPIcs.ITCS.2023.15. URL https://drops.dagstuhl.de/opus/volltexte/2023/17518.