

ER-Evaluation: End-to-End Evaluation of Entity Resolution Systems

Olivier Binette ¹ and Jerome P. Reiter ¹

¹ Duke University, USA  Corresponding author

DOI: [10.21105/joss.05619](https://doi.org/10.21105/joss.05619)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Øystein Sørensen  

Reviewers:

- @Nikoletos-K
- @ThomasHepworth

Submitted: 02 June 2023

Published: 10 November 2023

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Entity resolution (ER), also referred to as record linkage and deduplication, is the process of identifying and matching distinct representations of real-world entities across diverse data sources. It plays a crucial role in data management, cleaning, and integration, with applications such as assessing the accuracy of the decennial census, detecting fraud, linking patient data in healthcare, and extracting relationships in structured and unstructured data (Binette & Steorts, 2022; Christen, 2012; Christophides et al., 2021; Papadakis et al., 2021).

As ER techniques continue to evolve and improve, it is essential to have an efficient and comprehensive evaluation framework to measure their performance and compare different approaches. Despite the growth of ER research, there remains a need for a unified evaluation framework that can address challenges associated with ER system evaluation, including accounting for sampling biases and managing class imbalances. Otherwise, using naive clustering metrics and toy benchmark datasets without a principled evaluation methodology generally leads to over-optimistic results that can lead to performance rank reversals and poor system design (Binette, York, et al., 2023; Wang et al., 2022).

ER-Evaluation is a Python 3.7+ package designed to address these challenges by implementing all components of a principled evaluation framework for ER systems. It incorporates principled statistical estimators for key performance metrics and summary statistics, error analysis tools, data labeling tools, and data visualizations. The package is written in Python with a simple architecture, ensuring straightforward portability to other languages and frameworks when necessary.

Additionally, ER-Evaluation adopts a novel entity-centric approach that uses disambiguated entity clusters as the foundation for analysis. This contrasts with traditional evaluation methods based on labeling record pairs (Marchant & Rubinstein, 2017). The entity-centric approach streamlines the utilization of existing benchmark datasets and the labeling of new datasets without necessitating complex sampling schemes. Furthermore, it enables the reuse of benchmark datasets at all stages of the evaluation process, including for cluster-level error analysis.

Statement of need

Entity resolution is a clustering problem characterized by small and numerous clusters (up to millions or billions of clusters). Researchers commonly evaluate the performance of entity resolution systems by computing performance metrics (precision, recall, cluster metrics) on relatively small benchmark datasets. However, this process has been shown to yield biased and over-optimistic performance assessments in ER, potentially leading to performance rank reversals and poor system design (Binette, York, et al., 2023; Wang et al., 2022).

To address this issue, a new entity-centric methodology has been proposed in Binette, York, et

al. (2023) for obtaining accurate performance metric estimates based on small and potentially biased benchmark datasets. The ER-Evaluation package implements this methodology and numerous extensions to create a comprehensive, end-to-end evaluation framework. It aims to streamline the comparison of diverse ER techniques, assess their accuracy, and ultimately accelerate the development and adoption of high-performing ER systems. By integrating essential components such as data preprocessing, error analysis, performance estimation, and visualization functions, ER-Evaluation offers a user-friendly, modular, and extensible interface for researchers and practitioners.

The software is currently being used by PatentsView.org for the evaluation of patent inventor name disambiguation (Binette, Madhavan, et al., 2023). The original methodology has been published in Binette, York, et al. (2023) and extended methodology is under development in an upcoming article (Binette, Baek, et al., 2023).

Acknowledgements

We acknowledge financial support from the National Sciences and Engineering Research Council of Canada and from the Fonds de Recherche du Québec - Nature et Technologies.

References

- Binette, O., Baek, Y., Melluso, E., Jones, C., Dasylva, A., & Reiter, J. P. (2023). *An end-to-end evaluation framework for entity resolution systems with application to inventor name disambiguation*.
- Binette, O., Madhavan, S., Butler, J., Card, B. A., Melluso, E., & Jones, C. (2023). PatentsView-Evaluation: Evaluation datasets and tools to advance research on inventor name disambiguation. *arXiv e-Prints*. <https://doi.org/10.48550/arXiv.2301.03591>
- Binette, O., & Steorts, R. C. (2022). (Almost) all of entity resolution. *Science Advances*, 8(12), eabi8021. <https://doi.org/10.1126/sciadv.abi8021>
- Binette, O., York, S. A., Hickerson, E., Baek, Y., Madhavan, S., & Jones, C. (2023). Estimating the performance of entity resolution algorithms: Lessons learned through PatentsView.org. *The American Statistician*, 77(4), 370–380. <https://doi.org/10.1080/00031305.2023.2191664>
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer Publishing Company, Incorporated.
- Christophides, V., Efthymiou, V., Palpanas, T., Papadakis, G., & Stefanidis, K. (2021). An overview of end-to-end entity resolution for big data. *ACM Computing Surveys*, 53(6), 1–42. <https://doi.org/10.1145/3418896>
- Marchant, N. G., & Rubinstein, B. I. P. (2017). In search of an entity resolution OASIS: Optimal asymptotic sequential importance sampling. *Proc. VLDB Endow.*, 10(11), 1322–1333. <https://doi.org/10.14778/3137628.3137642>
- Papadakis, G., Ioannou, E., Thanos, E., & Palpanas, T. (2021). *The Four Generations of Entity Resolution*. Morgan & Claypool Publishers.
- Wang, T., Lin, H., Fu, C., Han, X., Sun, L., Xiong, F., Chen, H., Lu, M., & Zhu, X. (2022). Bridging the gap between reality and ideality of entity matching: A revisiting and benchmark re-construction. In L. D. Raedt (Ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22* (pp. 3978–3984). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2022/552>