

International Journal of Population Data Science

Journal Website: www.ijpds.org



Assessing Deep Learning in Record Linkage: Recommendations Based on Dataset Characteristics and Computational Costs

Eric Sartor¹, Jeremy Foxcroft¹, and Luiza Antonie¹

¹University of Guelph, Guelph, Canada

Deep learning systems are increasingly being applied to record linkage tasks, addressing either the matching stage or the entire linkage process. Although some deep learning methods appear to outperform classical approaches, they are typically evaluated on specific linkage problems, making wider applicability difficult to assess. Moreover, for any use in practical applications, it is critical to understand the trade-offs between the lift in linkage performance and the increase in computational complexity when using deep learning systems. When do the benefits of deep learning outweigh the costs compared to classical machine learning methods? Our investigations have three main objectives: increase our understanding of the performance of deep learning methods across a wide range of linking tasks, particularly transformer-based methods; measure computational costs; and make recommendations accounting

for performance-cost trade-offs. We developed a common and consistent framework to evaluate these trade-offs, which we apply on a large number of deep learning methods tackling a variety of record linkage tasks. Our evaluation covers both structured and unstructured datasets. We present precision-recall curves to better understand the performance of the models. We discuss computational costs for these methods to inform practitioners. Deep learning systems perform very well on unstructured data, while classical machine learning methods with feature engineering remain competitive on structured data at a significantly lower computational cost. This research will provide practical guidance to the practitioners in the data linkage community on linkage method selection based on dataset characteristics and computational costs.

