ICLR 2024 Workshop: Tackling Climate Change with Machine Learning

ExioML: Eco-economic dataset for Machine Learning in Global Sectoral Sustainability

Yanming Guo, Jin Ma
yguo0337@uni.sydney.edu.au

GitHub
https://github.com/Yvnminc/ExioML

Opportunities and Challenges for Machine Learning in Eco-economic Research

Eco-economic research aims to decouple production emissions from economic growth to enhance well-being, using Environmental-Extended Multi-regional Input-output (EE-MRIO) analysis. Machine learning (ML) offers deep insights by analyzing patterns among high-dimensional data and quantifying economic activities and environmental impacts. However, interdisciplinary challenges persist：

- **EE-MRIO dataset inaccessibility**: Most of the global EE-MRIO (GTAP, Eora) data are closed-source with expensive access and suffer from spatiotemporal resolution (WIOD), which is inappropriate for ML analysis.
- **Intensive data pre-processing requires domain knowledge**: ML research requires structured data, and a paradigm shift and data cleaning is required.
- L**ack of Benchmark datasets and ML models**: ML techniques are less discussed in previous literature. A uniform benchmark dataset should be developed to accelerate the cooperation between ML and Eco-economic research.

GHG Emission Regression on Shallow and Deep Models for Sectoral Sustainable Assessment

GHG emission regression, shown in Table 1, is validated on two sub-datasets, the Product-by-Product (PxP) and Industry-by-Industry (IxI), with low mean squared error (MSE) for both shallow and deep models, serving as the benchmark result for future study. Deep models generally have lower MSE scores, while shallow models, especially the ensemble tree-based models, GBDT, have competitive performance with more efficient training and inference time.

	PxP		**IxI**	
Model	MSE	Time (s)	MSE	Time (s)
KNN	1.071 ± 0.010	0.035 ± 0.001	1.151 ± 0.018	0.026 ± 0.001
Ridge	2.265 ± 0.000	**0.005 ± 0.002**	2.514 ± 0.000	**0.004 ± 0.002**
DT	0.926 ± 0.051	0.316 ± 0.017	0.848 ± 0.070	0.254 ± 0.019
RF	0.356 ± 0.004	21.521 ± 0.157	0.302 ± 0.004	16.511 ± 0.060
GBDT	0.234 ± 0.006	30.276 ± 0.224	0.219 ± 0.007	32.847 ± 0.388
MLP	0.226 ± 0.007	219.218 ± 0.904	0.250 ± 0.092	205.051 ± 24.309
GANDALF	**0.204 ± 0.010**	352.756 ± 7.036	**0.189 ± 0.007**	383.119 ± 3.664
FTT	0.330 ± 0.007	330.578 ± 1.527	0.302 ± 0.023	468.911 ± 7.329

Table 1: Model results with 10 runs and the top results are in **bold**.

ExioML: A Machine Learning Ready Eco-economic Dataset for Global Sectoral Sustainability Analysis

We proposed the ***first ML-ready benchmark dataset in Eco-economic research***, named ***ExioML***, for global sectoral sustainability analysis to fill the above research gaps. The overall architecture is illustrated in Figure 1. The ExioML is developed on top of the high-quality open-source EE-MRIO dataset ExioBase 3.8.2 with high spatiotemporal resolution, covering 163 sectors among 49 regions from 1995 to 2022, addressing the limitation in data inaccessibility. Both ***factor accounting*** in tabular format and ***footprint network*** in graph structure are included in ExioML. We demonstrate a ***GHG emission regression task*** on a factor accounting table by comparing the performance between shallow and deep models. The result achieved the low Mean Squared Error (MSE). It quantified the sectoral GHG emission in terms of ***value-added, employment, and energy consumption***, validating the proposed dataset's usability. The footprint network in ExioML is inherent in the multi-dimensional network structure of the MRIO framework and enables tracking resource flow between international sectors. Various promising research could be done by ExioML, such as ***predicting the embodied emission through international trade***, ***estimation of regional sustainability transition***, ***and the topological change of global trading networks based on historical trajectory***. ExioML reduces the barrier and reduces the intensive data pre-processing for ML researchers with the ready-to-use features, simulates the corporation of ML and Eco-economic research for new algorithms, and provides analysis with new perspectives, contributing to making sound climate policy, and promotes global sustainable development.

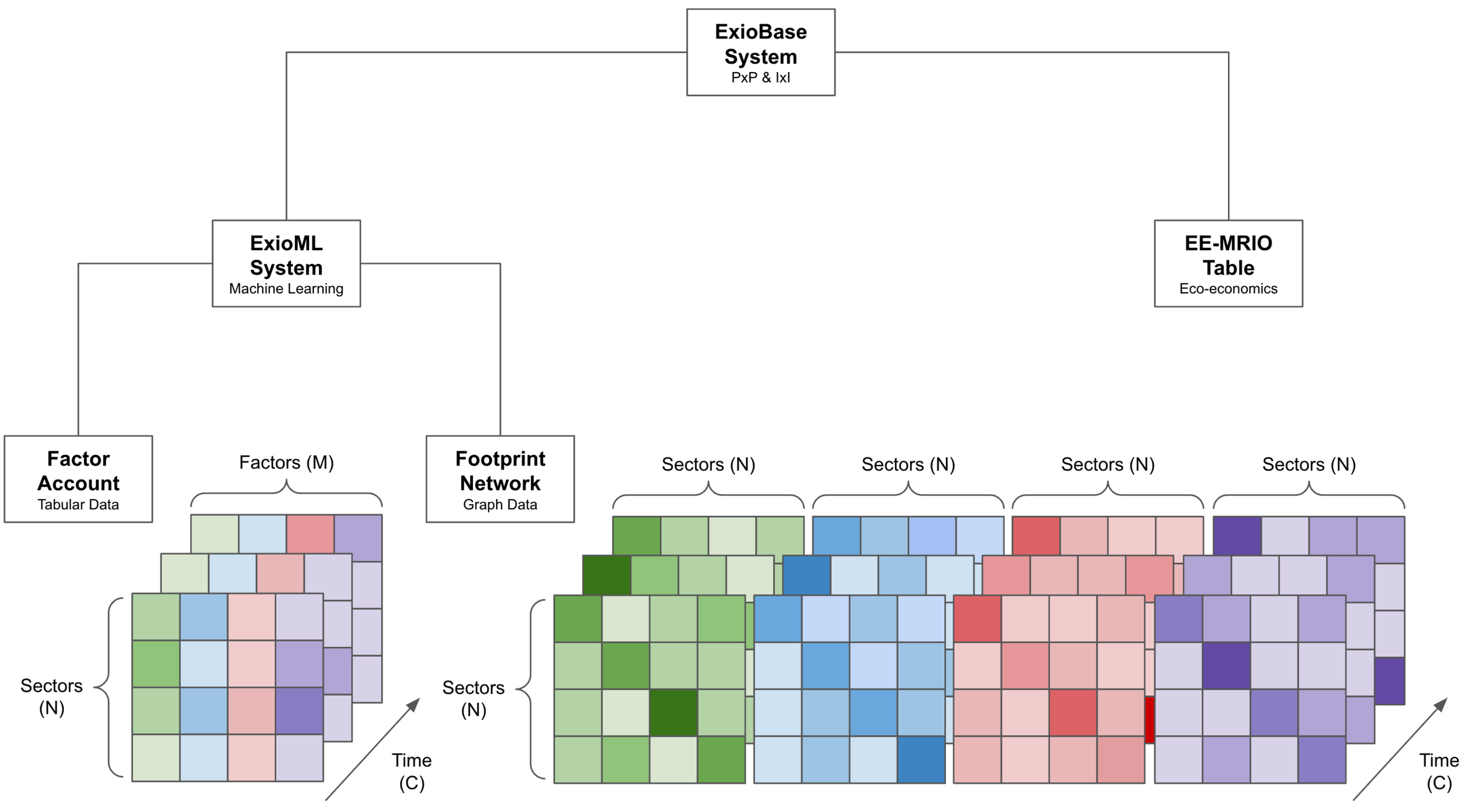

Figure 1: Architecture of ExioML system derived from open-source EE-MRIO database, ExioBase 3.8.2. Each colour indicates an eco-economic factor: value added, employment, energy consumption and GHG emission. It contains factor accounting describing heterogeneous sector features and the footprint network tracking resource transfer within sectors. The data has 2 categories: 200 products and 163 industries for 49 regions from 1995 to 2022 in the PxP and IxI datasets.