ICAMS 2024

Boosting Materials Design: Word Embeddings Meet Experimental Data

Lei Zhang
Prof. Dr. Markus Stricker

Interdisciplinary Centre for Advanced Materials Simulation (ICAMS)
Ruhr-University Bochum, Germany
Lei.Zhang-w2i@rub.de

ICAMS
INTERDISCIPLINARY CENTRE FOR ADVANCED MATERIALS SIMULATION

RUHR UNIVERSITÄT BOCHUM RUB

A Challenge of Material Discovery

Existing Strategies for Accelerating Material Discovery

Brute-force experimentation

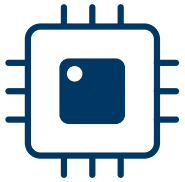
A Challenge: Too many possibilities

For a pool of 10 potential substitute:

- **Single Element:** 10 combinations
- **Binary:** 10! / (2! * 8!) = 45 combinations
- **Ternary:** 10! / (3! * 7!) = 120 combinations
- **Quaternary:** 10! / (4! * 6!) = 210 combinations

Existing Strategies for Accelerating Material Discovery

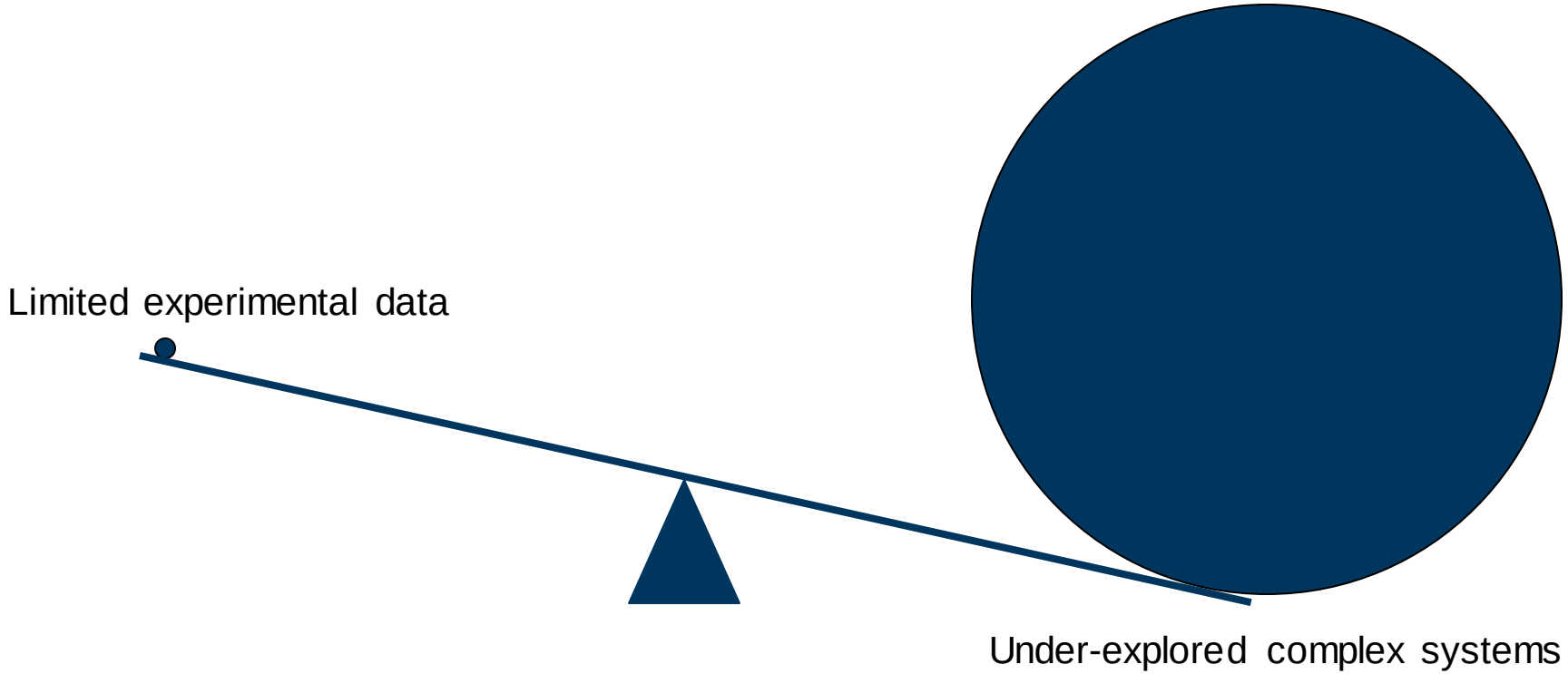
Analytical models



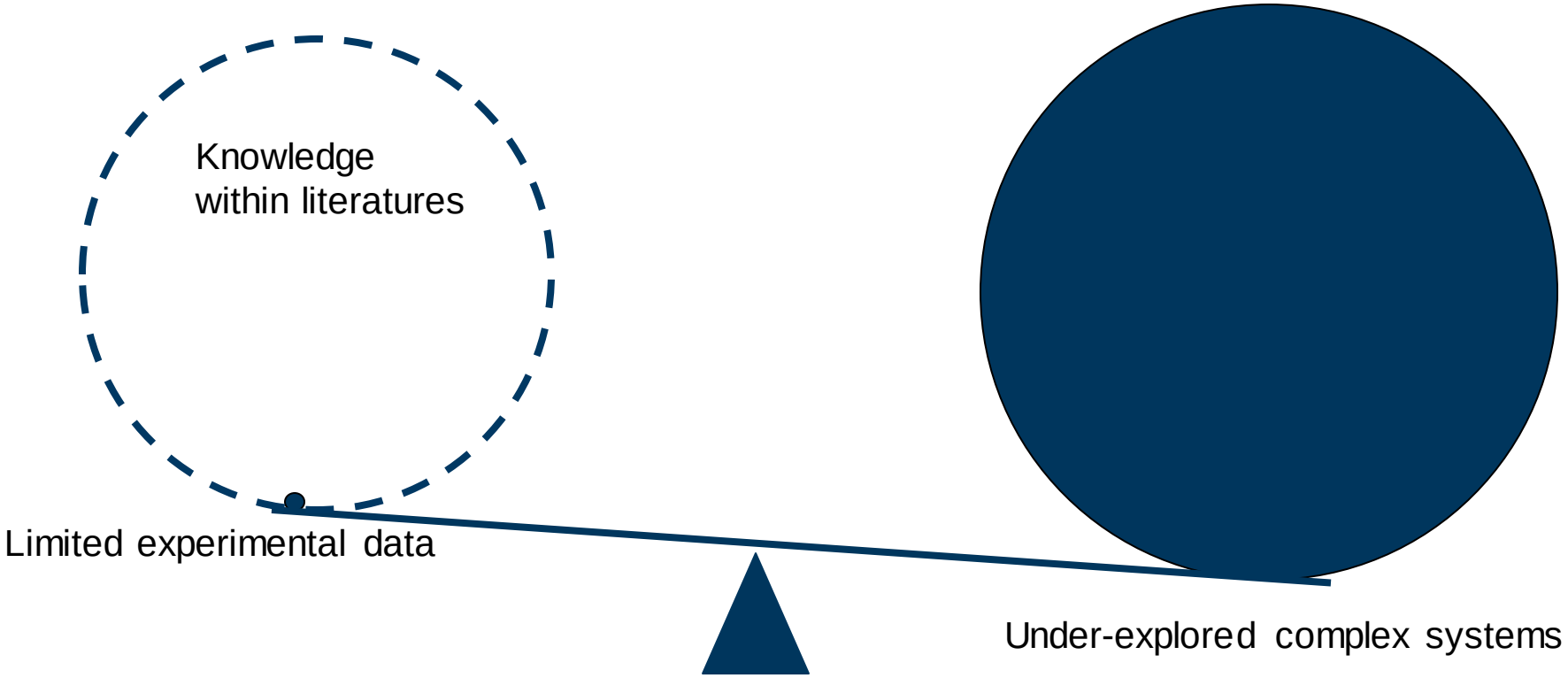
Simulations

Active learning

Challenge: Data Scarcity



Leveraging Word Embeddings for Materials Prediction



Leveraging Word Embeddings for Materials Prediction

Pt is a noble catalyst.

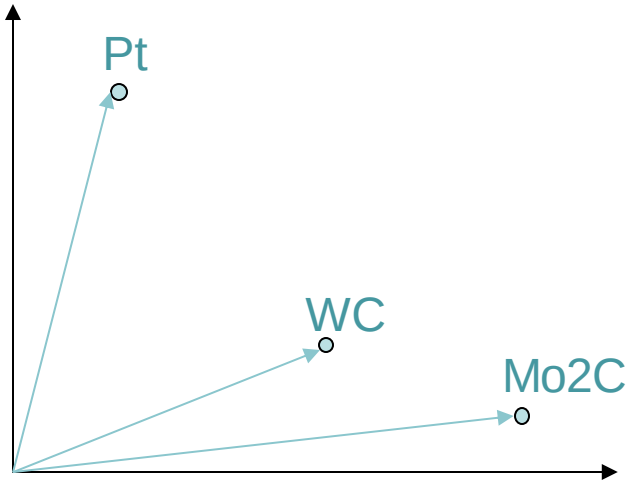
WC is a good and cheap catalyst.

Mo2C is an afordable and good catalyst.

Pt [1,5]

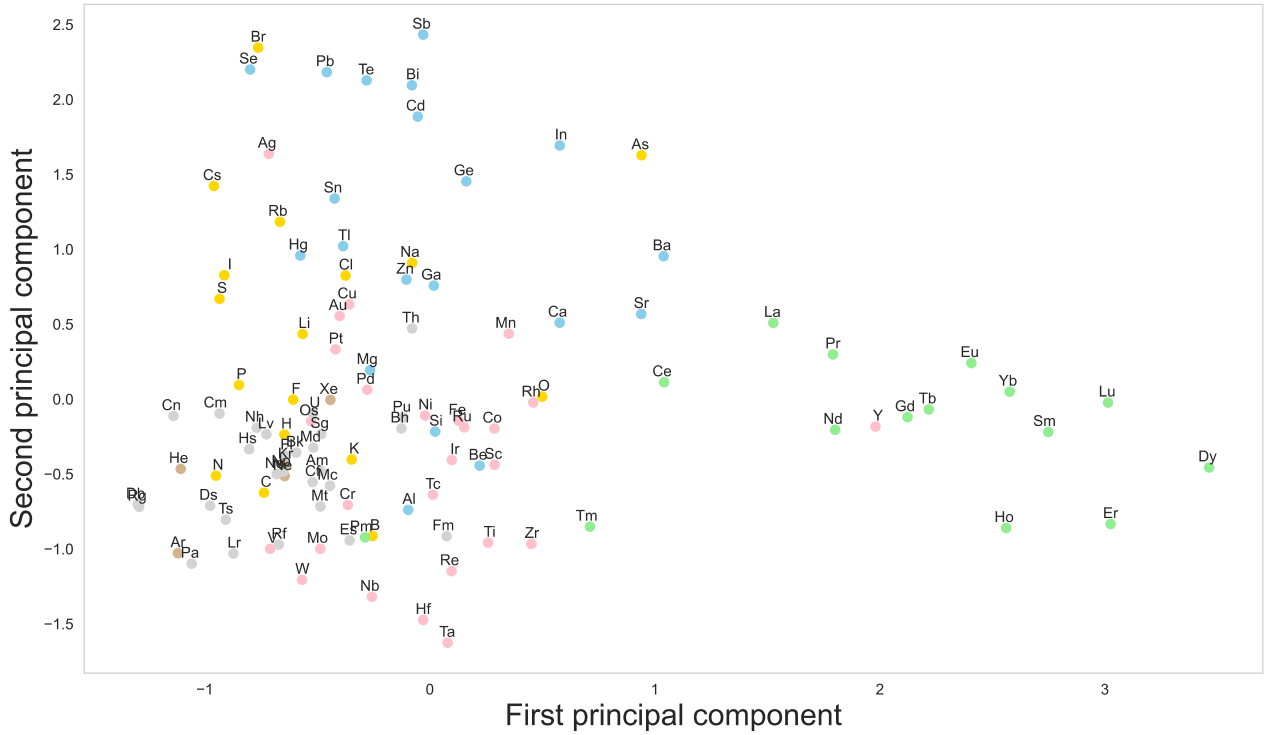
WC [2,2]

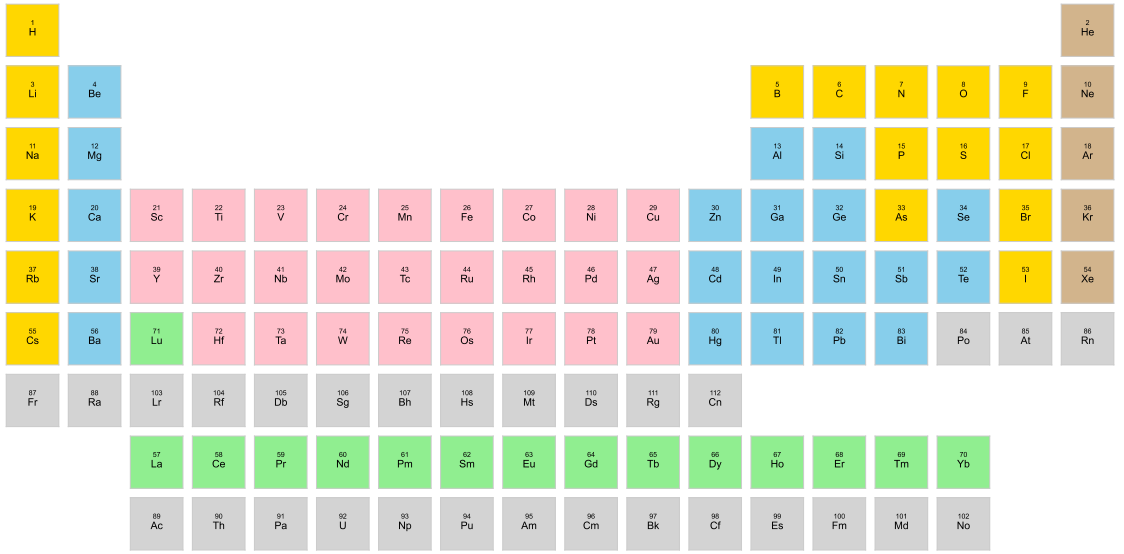
Mo2C [3,1]



Vec space

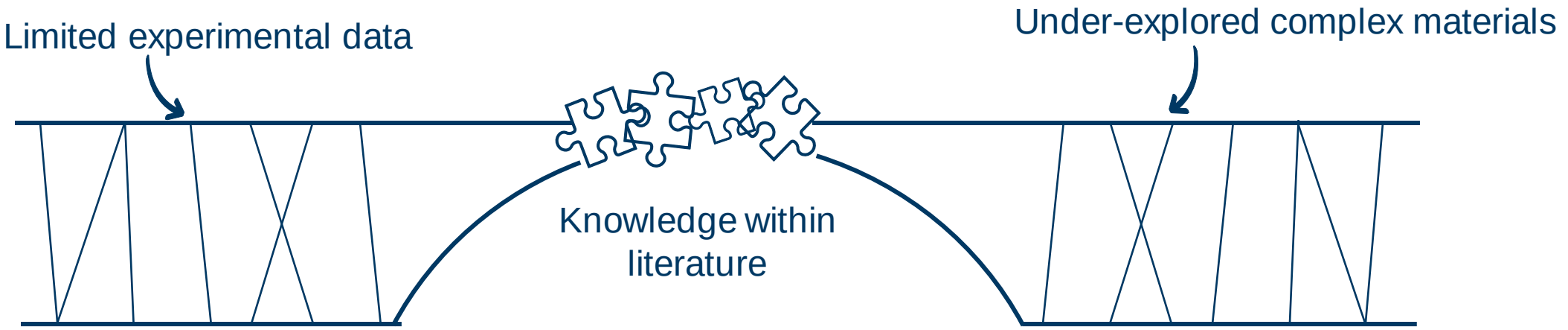
Leveraging Word Embeddings for Materials Prediction



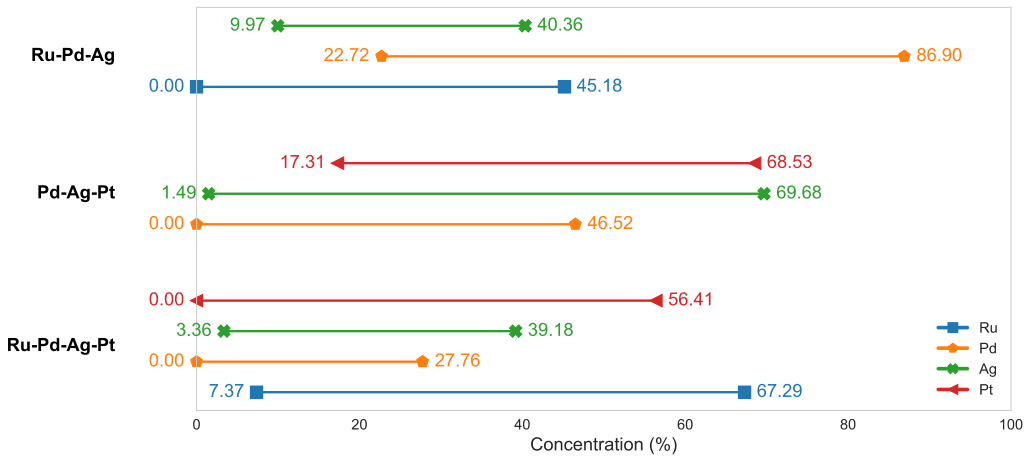


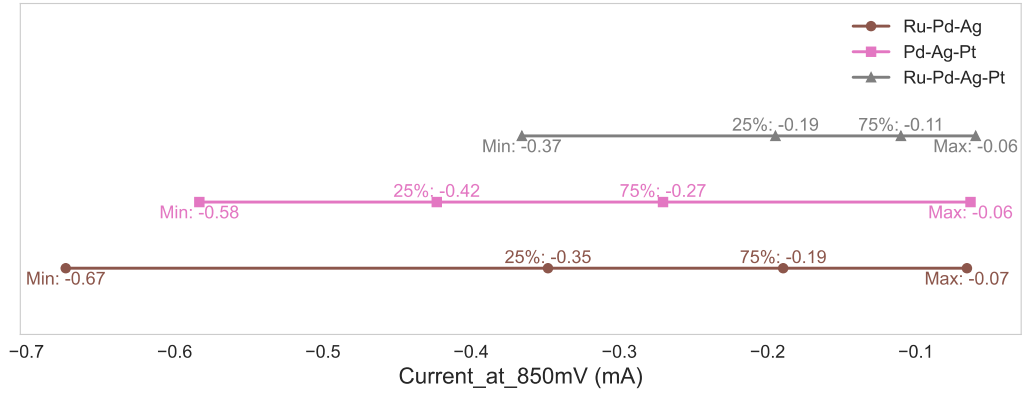
2D vector representation of elements through word2vec model

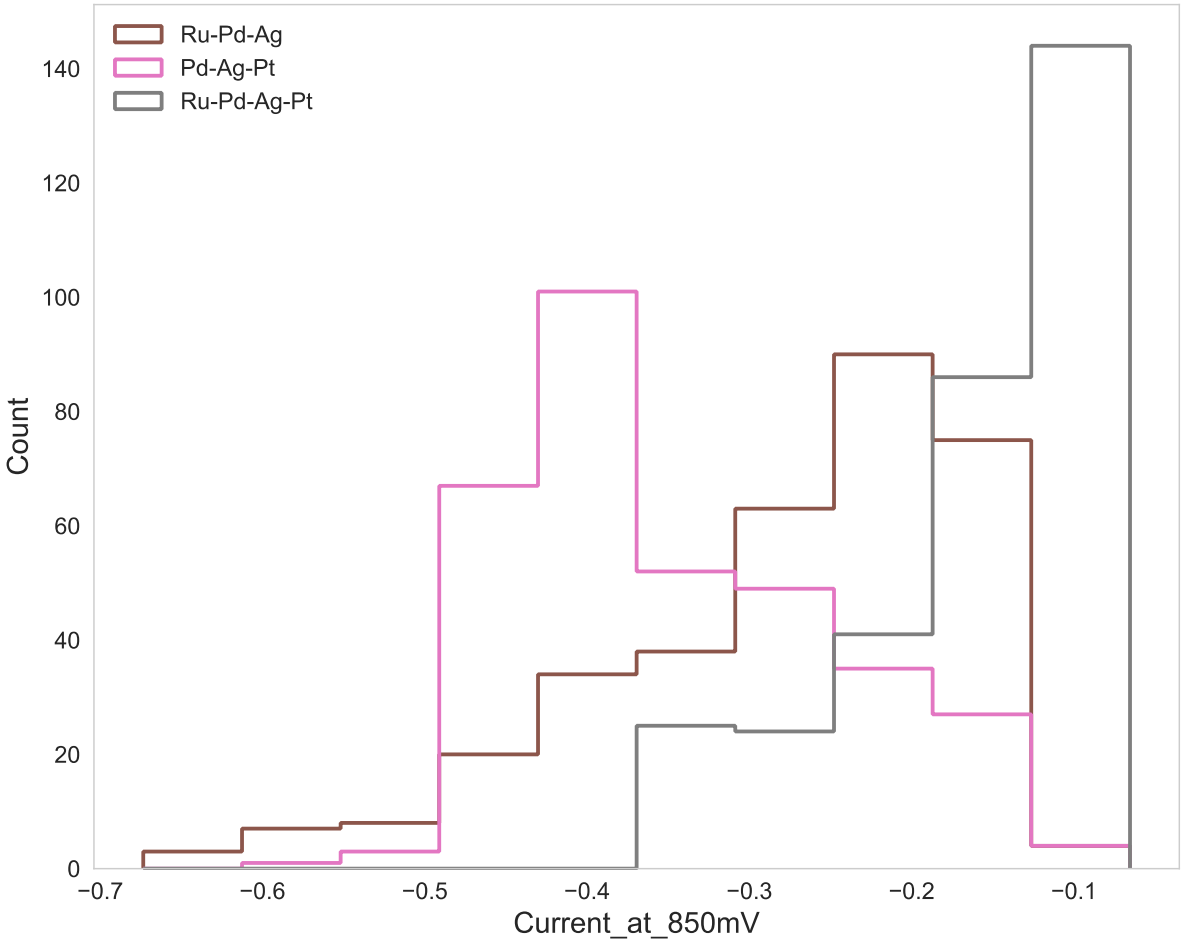
Leveraging Word Embeddings for Materials Prediction



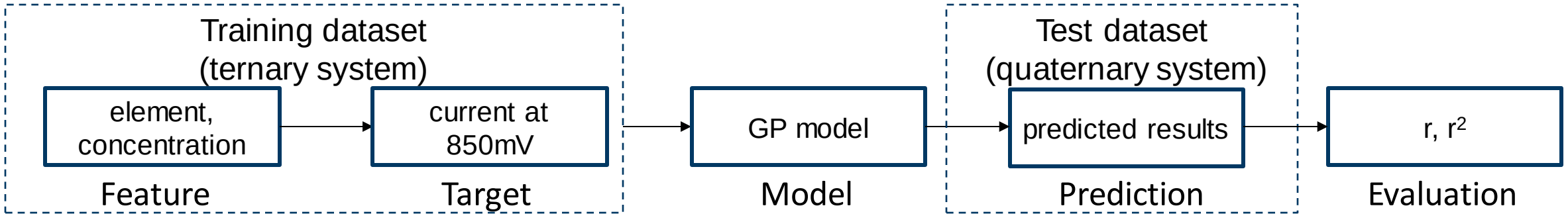
Dataset Overview



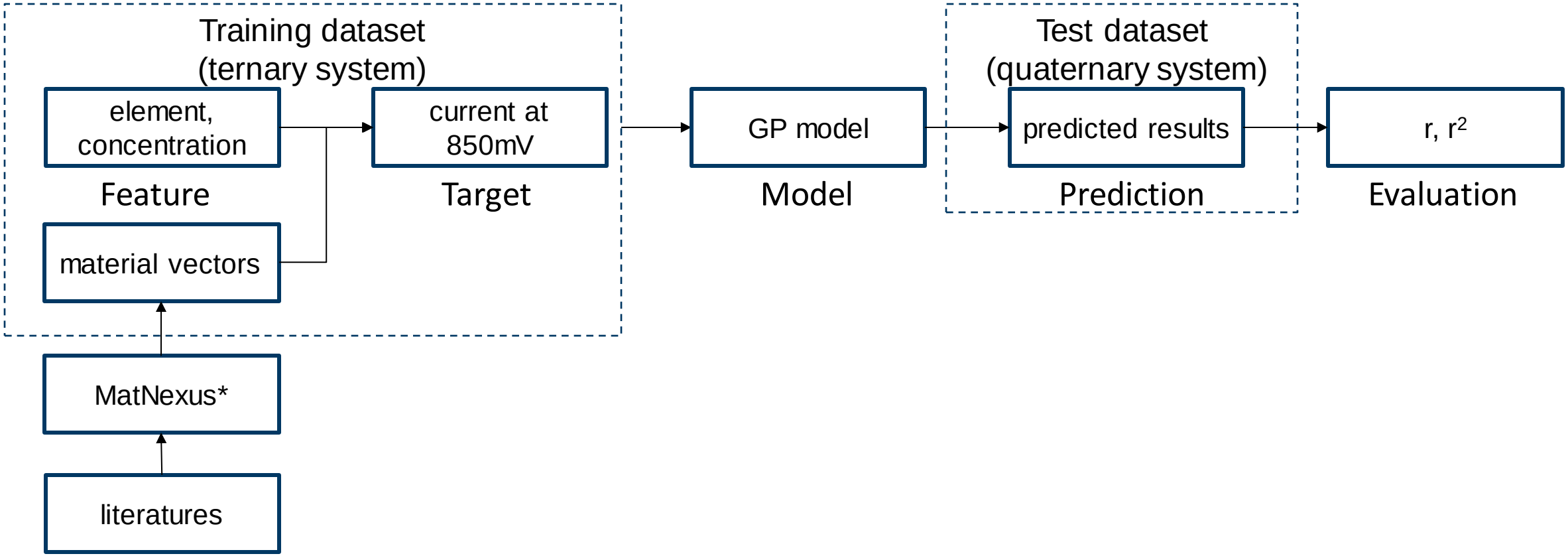




Gaussian Process (GP) Model



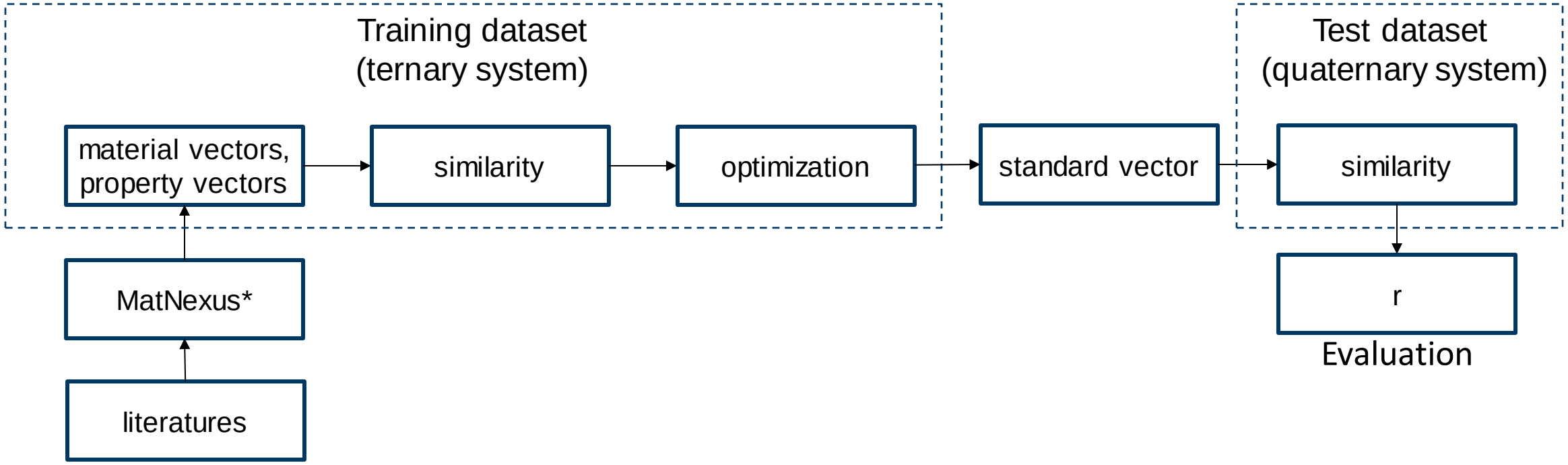
Enhanced GP Model



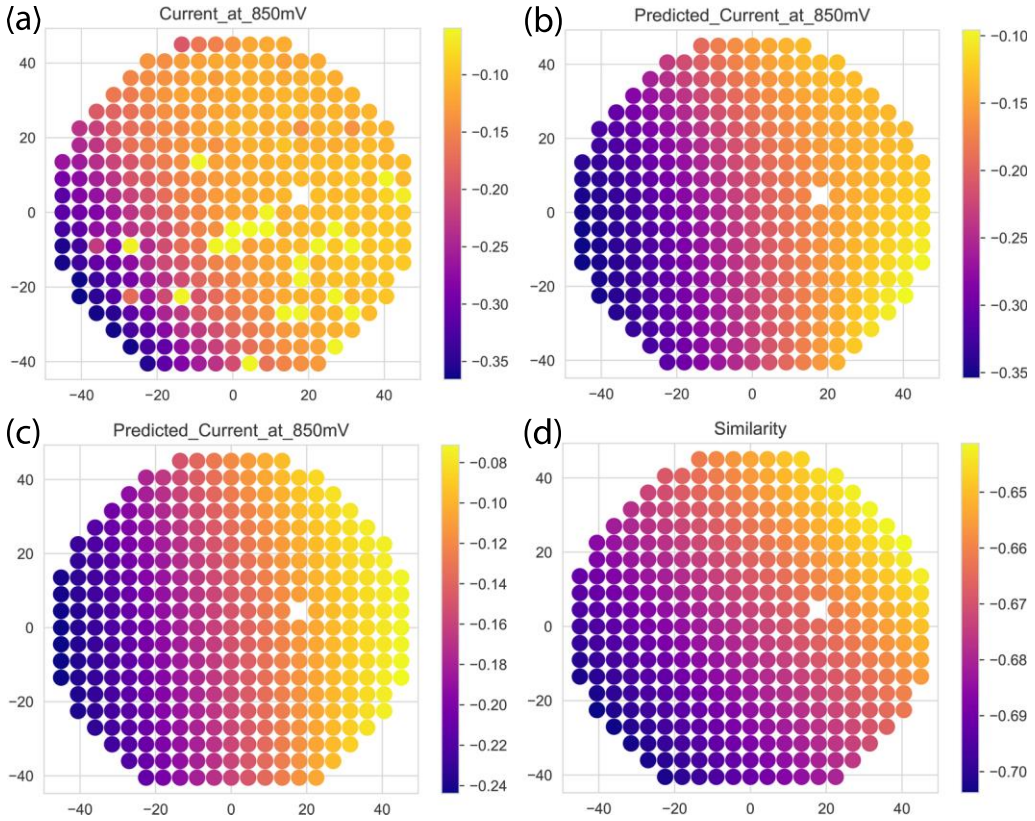
*https://github.com/lab-mids/matnexus

Advanced Method – Standard Vector

Advanced Method – Standard Vector



Prediction Results



(a) Experimental results of Ru-Pd-Ag-Pt system, (b) prediction results using GP model, (c) enhanced GP model with material vectors, (d)standard vector method.

Metric	Gaussian Process (GP)	GP with Embeddings	Standard Vector Method
Overall coefficient of determination (r^2)	0.08	0.65	-
Overall Correlation (r)	0.85	0.83	0.79
Correlation (r) for Current < -0.2 mA/cm²	0.63	0.60	0.89

Challenges and Limitations

Data Scarcity and Quality

Dependency on Literature

Model Complexity and Interpretability

Conclusion

•Using advanced machine learning and vector analysis techniques to predict material performance in complex systems.

•The integration of material vectors significantly enhanced predictive accuracy.

Q & A

Lei.Zhang-w2i@rub.de

markus.stricker@rub.de

https://github.com/lab-mids/matnexus