Improving Extrapolation in Ranking Tasks

Extrapolation remains a challenge for machine learning models, particularly in areas such as drug design and materials science. In these fields, accurate ranking of outliers is crucial.

Problem

How can we improve the accuracy of ranking high-performing samples that fall outside the training data range?

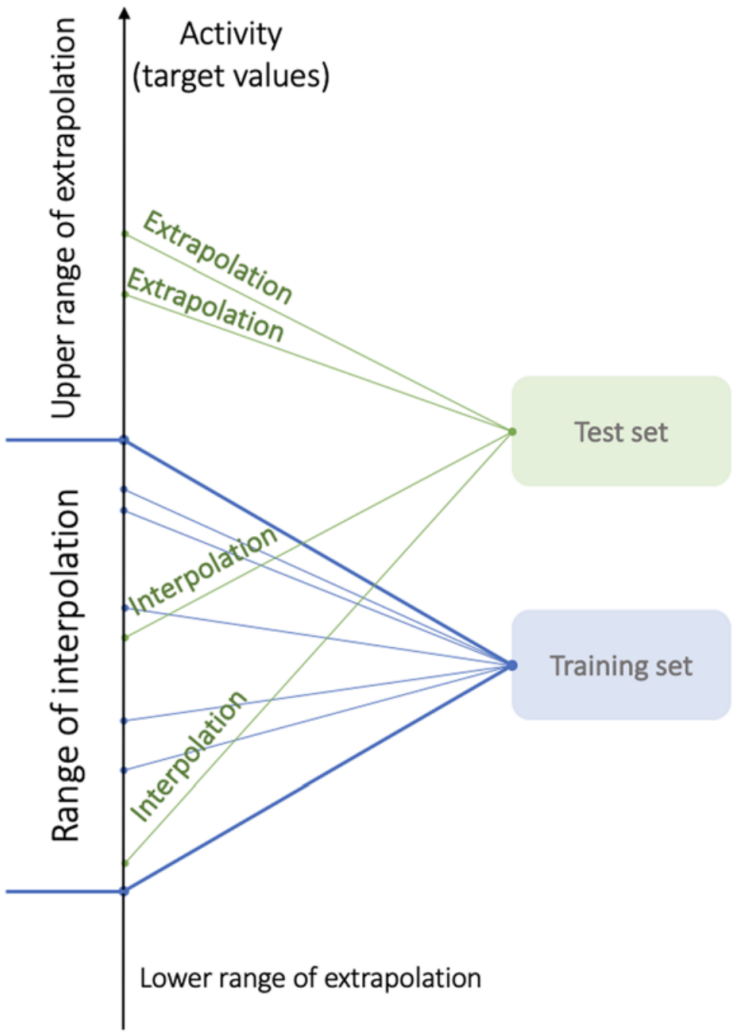
Method

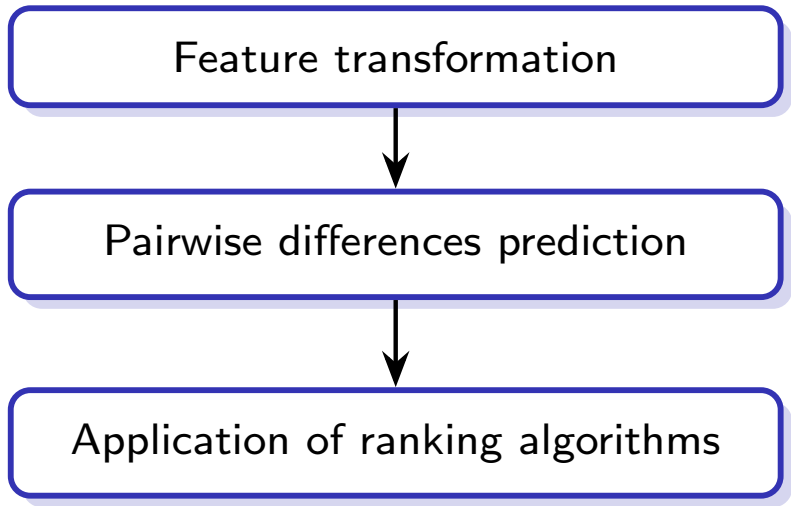
Using a pairwise learning approach: instead of predicting absolute target values, we predict the relative differences between samples, which are then processed by ranking algorithms.

Contribution

Proposed method enhances extrapolation ability, improving model performance in ranking and identifying high-performing samples that exceed the training data.

Pairwise Learning Approach





Key metrics

- $f1_{top10}$ — F1 score of a ML method to retrieve top 10% samples.
- $f1_{extrap}$ — F1 score of a ML method to retrieve extrapolating samples.