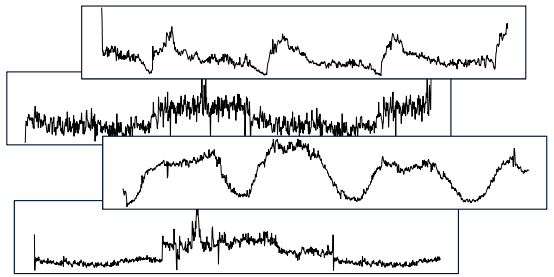
Toward Pattern-based Model Selection for Cloud Resource Forecasting

Georgia Christofidi Konstantinos Papaioannou Thaleia Dimitra Doudali

IMDEA Software Institute, Madrid, Spain

1. Problem Space

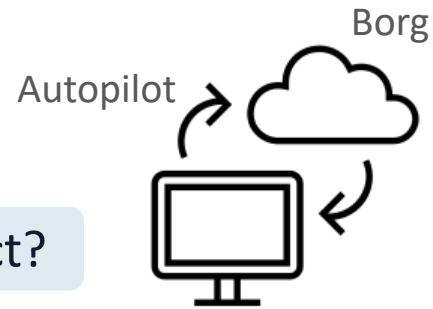
CPU Usage Patterns per Job



Historical Data

Cloud Resource Prediction Models

Which model to select?



Cloud Resource Management System

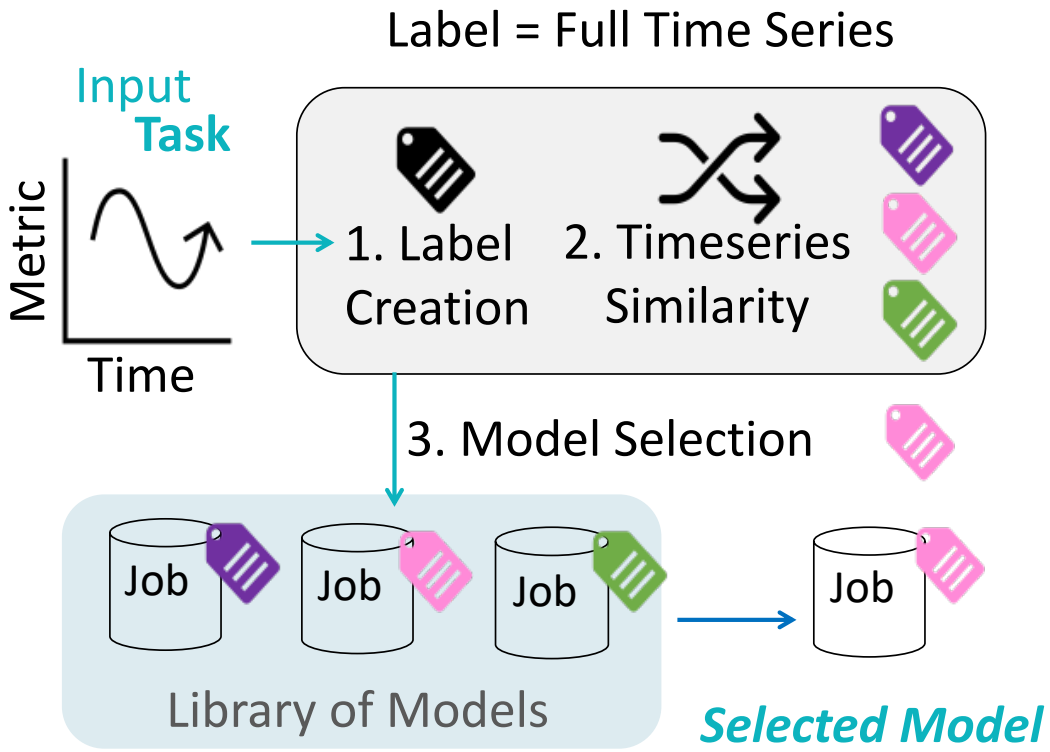
Resource Scaling: Choose Model with Highest Prediction Accuracy.

Overcommitment Policies: Choose Model that predicts Max value.

Example Use Cases

Problem Statement: Can we **select a model** based on the **pattern** of resource usage?

2. Proposed Approach



1 ML-based Model per Job

3. Pattern-based Comparison

Data Representations

Numeric: Time Series Data "as-is"

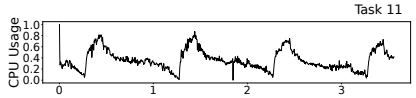
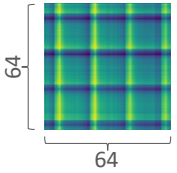


Image: **G**ramian **A**ngular **D**ifference **F**ield (GADF)



1 image / timeseries

Comparison Metrics

- L2 Norm
- **D**ynamic **T**ime **W**arping (DTW)
- **S**tructural **S**imilarity **I**ndex **M**easure (SSIM)

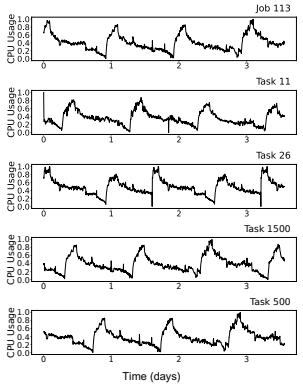
Approach - Combinations

I. Numeric – L2
II. Numeric – DTW
III. GADF Image - L2
IV. GADF Image - SSIM

Question: Which **combination** of time series data representation and comparison metrics can **separate the tasks of a job** based on a pattern?

Methodology: Run k-means to cluster the time series of the tasks creating 1 cluster per job.

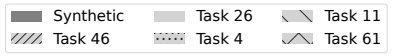
When using **homogeneous** (very similar) tasks, the clustering is successful for **all approaches.**

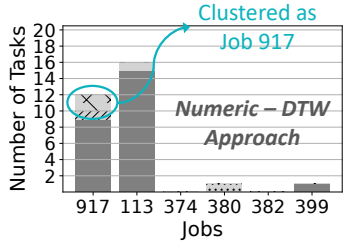


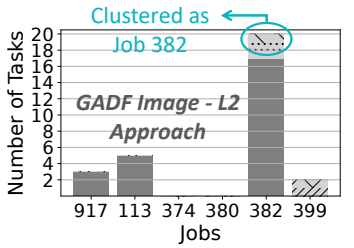
Representative Task of job 113 (Google Workload Traces)

Real Heterogeneous (slightly dissimilar) Tasks

Synthetic Time Shifted Tasks



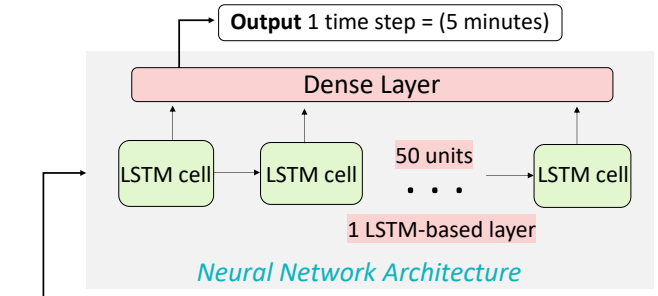




Slightly dissimilar tasks with spikes or time shifted patterns are not grouped together. Even when using DTW, a sophisticated method, or when using images to reveal more features.

No single winner approach!

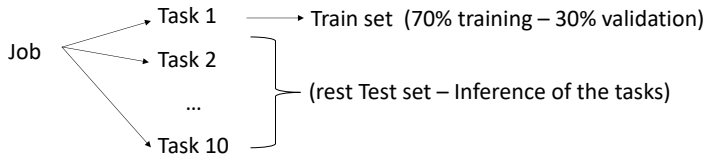
4. Model Selection



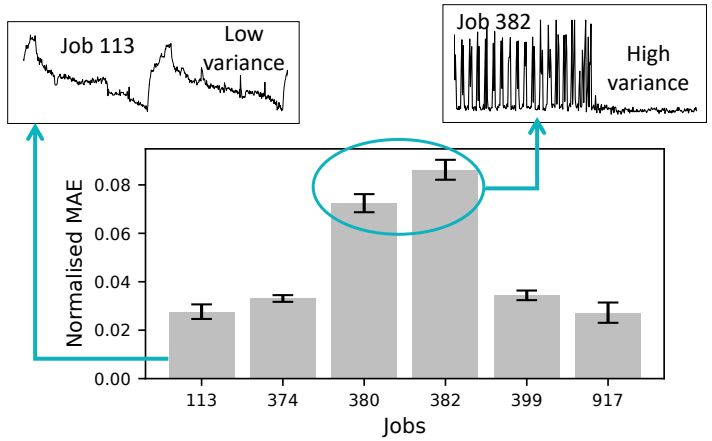
Optimizer = ADAM
Loss Function = MAE
Activation function = Softmax

Training for 1000 epochs.

1 ML-based model per job



Details of our ML Model Deployment

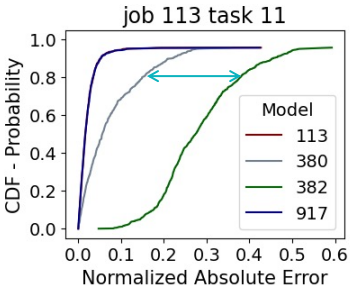


Prediction Error (MAE) across tasks of a job.
Model trained on the job.

ML models can **generalize** across job tasks.

Approach	Model
Numeric – L2	380
Numeric – DTW	917
GADF Image - L2	382
GADF Image - SSIM	917

*Models selected across approaches to predict **task 11 of job 113.***



Distribution of Prediction Error (MAE).

- **Model 113** lowest overall error. Importance of choosing the right model.
- **Model 917** exactly same curve. Opportunity for stronger generalizability?
- **Models 380, 382** deliver 10% - 40% error with probability 0.8.
  Impact of not choosing the right model.

Effective Pattern-based Model Selection
is important to deliver **high prediction accuracy**.

5. Main Insights

1. **Effective pattern-based model selection** unlocks highly **generalizable and accurate** model inference across tasks of a job. Ineffective selection reveals significant loss in inference quality.

2. Pattern-based comparisons using **distance-based metrics** are effective for very similar timeseries, but **break** when patterns become **slightly disimillar** (e.g., time shifted), even with more sophisticated approaches (DTW, image-based). Opportunity for new contributions!

6. Future Directions

- **Expand dataset** to more jobs, tasks, patterns, resources, and finer granularity across time windows.
- Explore more sophisticated **ML-based pattern matching**.
- Use **explainable AI** to understand model generalizability.
- Explore **other forecasting models** (ML, statistical).
- **Integrate** pattern-based model selection in **use case** e.g., resource autoscaler, overcomitment policy.