Is Machine Learning Necessary for Cloud Resource Usage Forecasting?

Georgia Christofidi, Konstantinos Papaioannou, Thaleia Dimitra Doudali
IMDEA Software Institute, Madrid, Spain

1. Problem Space

**Challenge 1:** Low resource efficiency in the Cloud.

**User:** Asks: X > Y (Uses)

**Approach:** Future Resource Usage Forecasting.

**Input:** Past Resource Usage X1, X2, ..., Xn

**Forecasting Models**
- (ML, Statistical, Heuristic, Hybrid)

**Output:** Future Resource Usage Xn+1, Xn+2, ..., Xn+k

**Problem:** Achieving High Accuracy in Forecasting.

- Accurate Forecasts can:
  - Resource Efficiency
  - Costs
  - Energy Efficiency
  - Application Performance

2. Existing Approach

**Use case:** Global Active Power Consumption

**Input:** Past Resource Usage

**Output:** Future Resource Usage

**Usecase:** Cloud Workloads

**Forecasting Models**
- Long Short-Term Memory (LSTM)

**Approach:** Revisit existing systems and study the simple mechanisms to the extent possible.

**Our Insight:** LSTM predictions resemble the previous timestep of the timeseries.

- Do we need ML to produce such “shifted” predictions?

3. Proposed Approach

**Persistent Forecast**

- Predict the value at the previous timestep.

**We observe very low error values that depend on the resource type.**

- Long Short-Term Memory (LSTM)

**Frequency**

- Physical/ Virtual Machine

**Resource Levels**

- Workload

**Resource Types**

- Hourly/ Daily/ Weekly Windows

**Scan for code and paper:**

- Alibaba Dataset
- Google Dataset

**Takeaway:** Persistent Forecast is highly accurate for cloud data, across resource types, levels of use and measurements.

4. Experimental Results

**Sensitivity to the length of the time window**

**Takeaway:** Small sensitivity to the time window. Opportunity for low error when window and patterns align.

**Our Insight:** The persistent forecast is effective because resource usage values of cloud workloads and servers, persist over time.

5. Summary

**Open Questions**

1. When to use ML?
   - exact use case data pattern system’s performance and decision-making

2. Which ML method to use, when necessary?
   - Probably not LSTMs
   - Other-state-of-the-art ML methods for timeseries forecasting

**Suggestions**

1. Revisit existing systems and study the data patterns:
   - Values persist over time?
   - Try the Persistent Forecast

2. Insightful and judicious use of ML, simple mechanisms to the extent possible.

All code is open source and available on Github.

Resources

- GitHub
- ![GitHub Link](https://github.com/username/repo)

- Machine Learning is not always necessary for Cloud Resource Usage forecasting.

- ![Image](https://example.com/image)