A Deep Reinforcement Approach to Optimize the Performance of Stream Processors

Arnaud Dethise

Advisor: Marco Canini

Status:
Working on prototype

Dig Data Systems
Machine Learning for Systems

- Problem: It is difficult to automatically tune stream processors to improve their online performance
- Motivation: A correct configuration improves throughput and latency, but its creation is costly
- Our approach: Apply state-of-the-art Machine Learning techniques to leverage historical data and learn the optimal configuration
- Benefits: The trained agent could generate new configurations fast enough for online tuning, as well as be able to adapt to new systems, applications and workloads

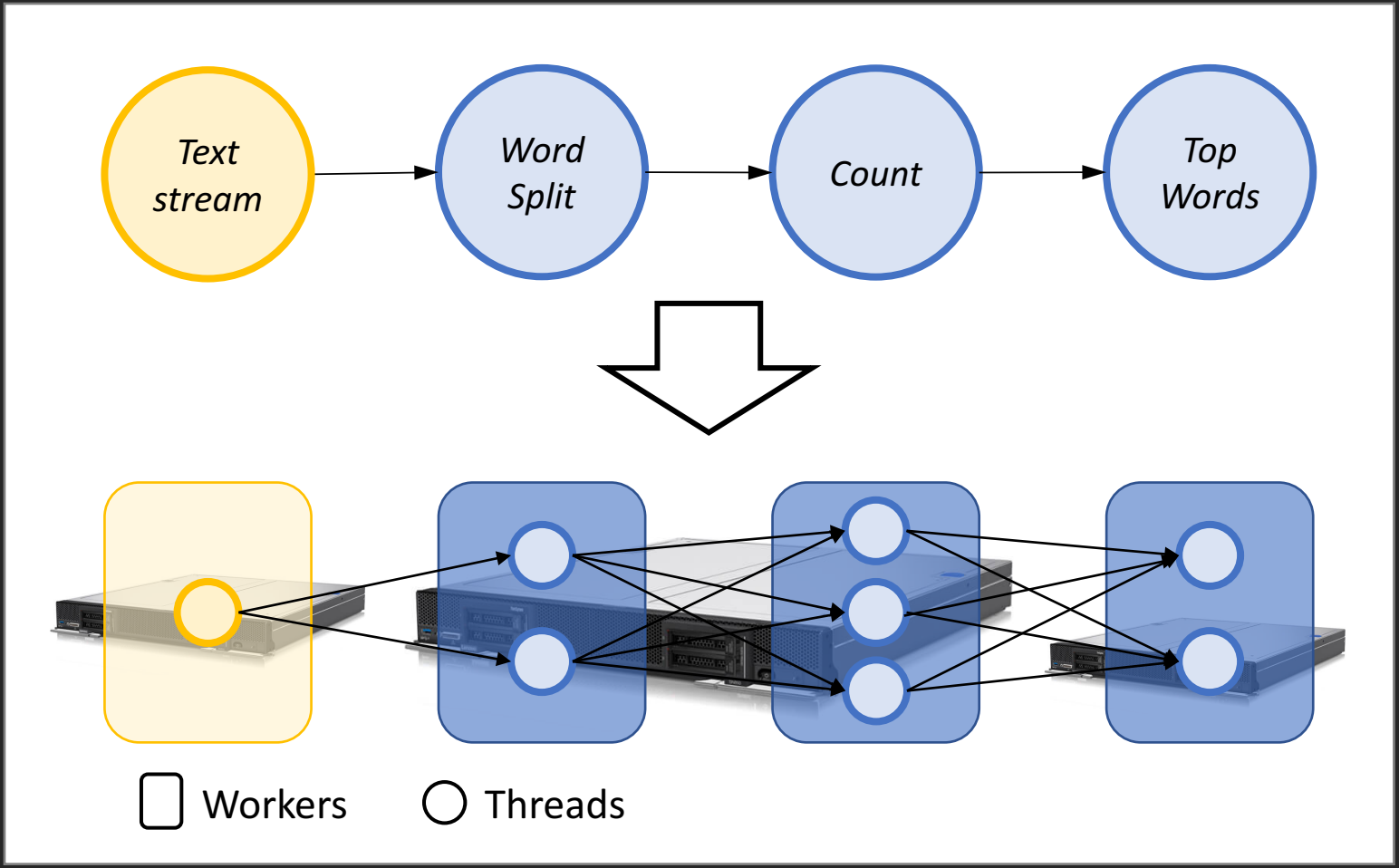
Advantages of applying Machine Learning to Systems

Many systems need ability of automated decision (Networks, Databases, Big Data systems…)

Advantages of Machine Learning:

- Portability to new environments
- Leveraging historical data for accuracy
- Pre-training to improve performance

The stream processor configuration problem



Example Stream Processing

Highly configurable !

- Number of workers
- Degree of parallelism
- Size of buffers
- Many others…

Challenges in automated configuration of stream processors

- Reconfiguration needs to be both fast and accurate
- Large action space (many different configurations)
- Not easily divided in independent subproblems
- Must accommodate very different applications

Trusting Machine Learning Decisions

Deep Learning models are powerful black boxes

- No insight in the decision process

Many successes in other fields, such as image classification and text recognition

But many systems can not afford blind trust !

Importance of explainability in machine learning

Some things we cannot prove without explainability:

- Robustness
- Fairness
- Security

What tools are needed for explainability in Machine Learning for Systems ?

Conclusion

Collected data that is not used is lost value

Creating Systems that integrate Machine Learning *by design*, not as an afterthought

Additional slides

Why stream processors?

- Always-online system
- Stringent requirements on throughput and latency
- Various applications with different behaviors
- Large amount of collected metrics
  - Per-stage throughput, latency, worker utilization...

Expected results

Previous successes of Machine Learning for networks, video streaming and databases

- Reasonable to expect it to work

Why?

- What is learned from the data?
- What information do the models use?
- Broad question that applies to all use of black-box ML for Systems

State of the art techniques for Explainable Deep Learning models

- Visualization
  - Project the decision process in 2D / 3D space
- Back-propagation
  - Examine weight of features
  - Find inputs leading to a particular output
- Black-box auditing
  - Usually by adding noise / masking features

Research questions in explainable AI for Systems

Classical techniques are domain-specific

- Rely on graphical features, high correlation between the inputs…

Driven / Informed process

What tools are needed for explainability in Machine Learning for Systems ?