



Evolving Embedding Framework for Novelty Detection in Evolving Data Streams

Leandro Rochink Costa

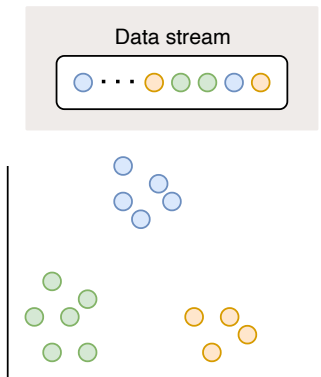
Polytechnique Montréal
DORSAL Laboratory

Outline

- 1 Evolving Data Streams
- 2 EvolveNet
- 3 Preliminary results
- 4 Current State

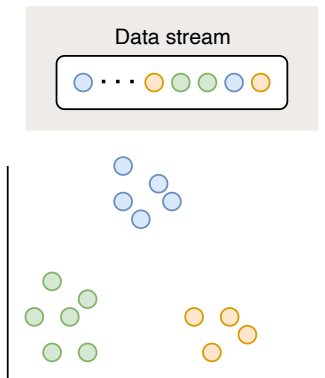
Data Streams

- Unbounded sequence of data points
- No prior knowledge (distribution or classes)
- Fast single pass data processing
- Examples: Traces, logs, requests on a server, transmission packages

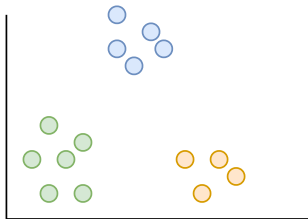
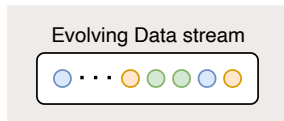


Data Streams

- Unbounded sequence of data points
- No prior knowledge (distribution or classes)
- Fast single pass data processing
- **Examples:** Traces, logs, requests on a server, transmission packages



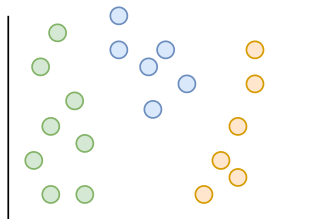
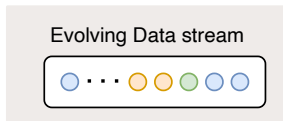
Evolving Data Streams



- Examples: Software, hardware, and user behavior changes.



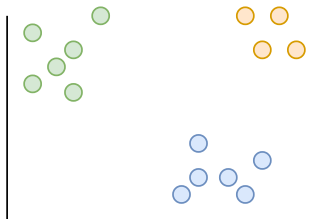
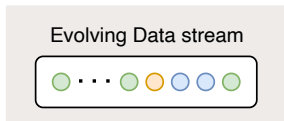
Evolving Data Streams



- Examples: Software, hardware, and user behavior changes.



Evolving Data Streams



- **Examples:** Software, hardware, and user behavior changes.



Current Solutions

- Representation learning for anomaly detection:
 - Meta-learning; Self-learning; Unsupervised-learning;
 - Autoencoders, Variational autoencoders, and GANs.
- Any of those approaches assume that data distribution change.



Current Solutions

- Representation learning for anomaly detection:
 - Meta-learning; Self-learning; Unsupervised-learning;
 - Autoencoders, Variational autoencoders, and GANs.
- Any of those approaches assume that data distribution change.



EvolveNet - Goal

- Tackle the unfitness of the learned representation when concept drifts in the data stream.
- More specifically:
 - Keep data representation relevant even even with data distribution changes;
 - Learn representations optimizing the cluster separation;
 - Perform unsupervised learning;
 - Detect novelty (potential anomalies) in an online manner;

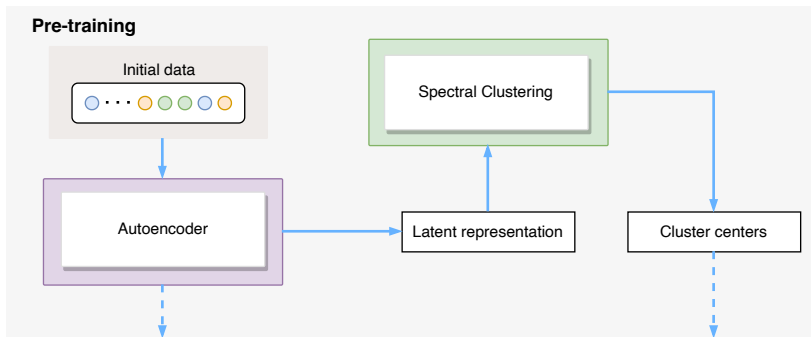


EvolveNet - Goal

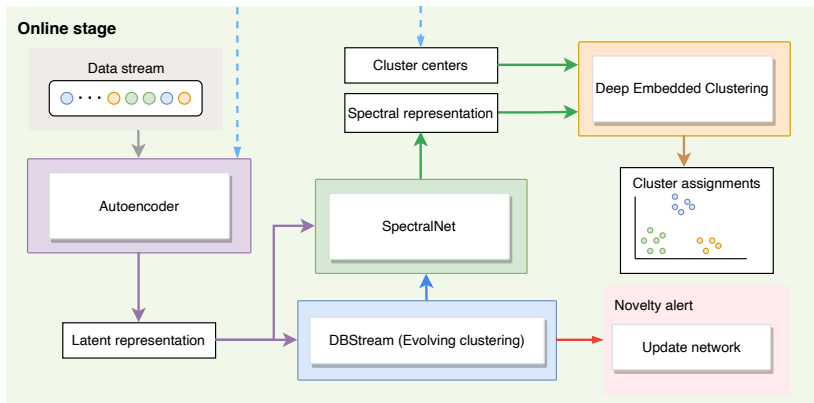
- Tackle the unfitness of the learned representation when concept drifts in the data stream.
- More specifically:
 - Keep data representation relevant even even with data distribution changes;
 - Learn representations optimizing the cluster separation;
 - Perform unsupervised learning;
 - Detect novelty (potential anomalies) in an online manner;



EvolveNet - Pipeline I

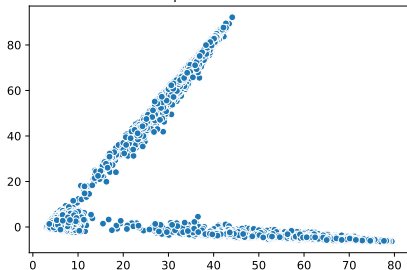


EvolveNet - Pipeline II

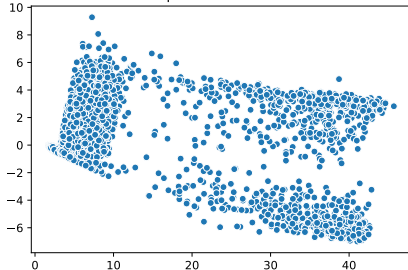


EvolveNet - Preliminary results

AE: Latent representation - Known data

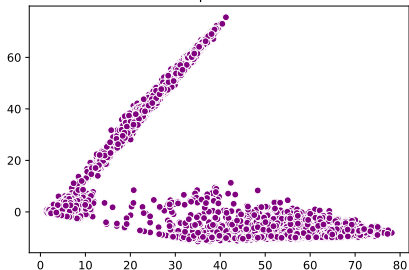


AE: Latent representation - Unknown data

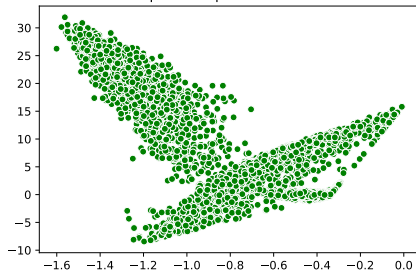


EvolveNet - Preliminary results

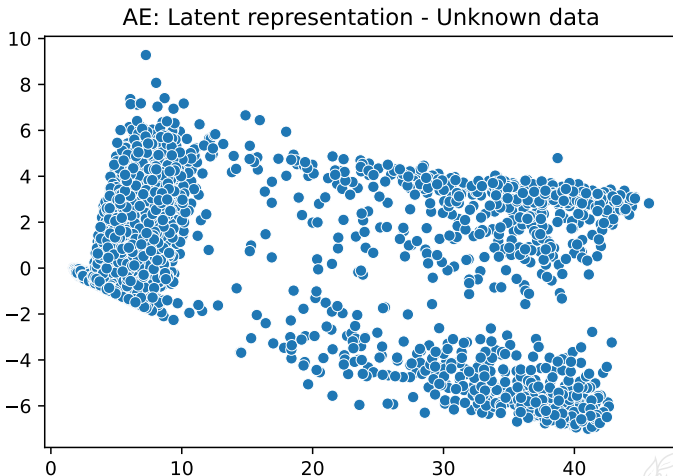
EvolveNet: Latent representation - Known data



EvolveNet: Spectral representation - Known data

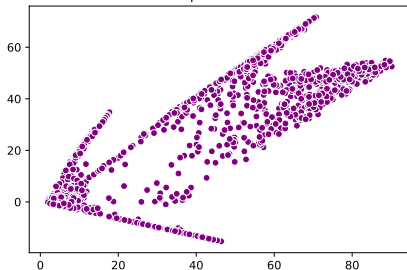


EvolveNet - Preliminary results

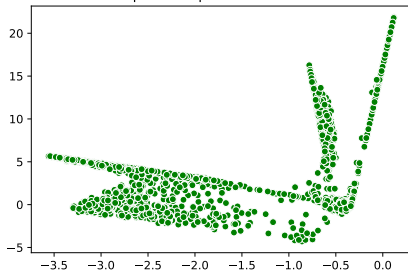


EvolveNet - Preliminary results

EvolveNet: Latent representation - Unknown data



EvolveNet: Spectral representation - Unknown data



Current State

- Adjusting our framework to provide cluster assignments.
- Benchmark of microservices use cases
 - Performance issues
 - Networking problems
- We are looking for real-life data!! :)



Current State

- Adjusting our framework to provide cluster assignments.
- Benchmark of microservices use cases
 - Performance issues
 - Networking problems
- We are looking for real-life data!! :)



Current State

- Adjusting our framework to provide cluster assignments.
- Benchmark of microservices use cases
 - Performance issues
 - Networking problems
- We are looking for real-life data!! :)



Thank You!

Leandro Rochink Costa
leandro.costa@polymtl.ca

