2025 Semantic Web - Postgraduate Students

Online Machine Learning for Semantic Classification of RDF Streams

Project Motivation

RDF Stream Processing systems enable continuous querying over high-velocity semantic data generated by sources such as Internet of Things deployments, smart cities, and real-time knowledge graphs. While these systems support expressive streaming queries, they largely lack integrated machine learning capabilities for semantic classification tasks.

In real-world deployments, semantic annotations for streaming data are often incomplete, delayed, or partially available, as manual labeling is costly and automated annotation pipelines may lag behind data production. This project investigates online learning techniques for semantic classification of RDF streams under partial and delayed supervision.

Project Objectives

1. Design an online learning pipeline for semantic classification of RDF streams.
2. Handle partial, delayed, or sparse supervision in realistic streaming scenarios.
3. Support continuous model adaptation under evolving data distributions.
4. Evaluate learning performance under class imbalance and concept drift.
5. Integrate learning components with RDF Stream Processing workflows.

Research Questions

1. How can semantic types of RDF stream observations be predicted with limited labeled data?
2. Which online learning strategies are most robust under delayed and partial supervision?
3. How does class imbalance affect semantic classification in RDF streams?
4. How effectively can online models adapt to concept drift in smart city data?

Project Tasks

1. **Review background literature on RDF Stream Processing, semantic sensor data, and online learning.**
2. **Construct a streaming data pipeline using RDF streams derived from smart city or IoT benchmarks.**

3. **Design feature extraction methods for transforming RDF observations into numerical representations.(SEMANTIC WEB PROJECT)**
4. Implement and evaluate online learning algorithms capable of incremental updates.
5. Simulate partial and delayed supervision scenarios and assess learning performance.
6. Compare different learning strategies using metrics suitable for imbalanced data streams.
7. Document the methodology, experimental results, and limitations in a final report.

Expected Outcomes

1. An online semantic classification system for RDF streams.
2. Experimental insights into learning under sparse and delayed supervision.
3. Quantitative comparison of online learning methods for imbalanced semantic data.
4. Reusable code and documentation suitable for further research or extension.

Skills Developed

1. Online and incremental machine learning.
2. Semantic Web technologies and RDF Stream Processing.
3. Feature engineering for temporal and spatial data.
4. Evaluation of machine learning systems under concept drift.
5. Research-oriented software development and experimentation.

Possible Extensions

1. Active learning or weak supervision strategies for RDF streams.
2. Adaptive handling of emerging semantic classes.
3. Hybrid symbolic–statistical approaches to semantic annotation.
4. Resource-aware learning under strict latency constraints.
5. Visualization and monitoring tools for streaming classification performance.

Project Suitability

This project is suitable for postgraduate students with an interest in machine learning, data streams and semantic technologies. Familiarity with Python and basic machine learning concepts is expected.