One Pager

Concept Drift Detection

Gurgen Hovakimyan

Data Visualization Spring 2025

American University of Armenia

Feb 3, 2025

ABSTRACT

Data evolves over time across various domains, from customer behavior in e-commerce to sensor readings in industrial processes. These shifts in data distribution are often referred to as *concept drift*. While concept drift is commonly discussed in the context of model performance degradation, it also poses broader challenges for data analysts and decision-makers. Effective visualization of drifting data distributions can unveil when and why shifts occur, facilitating more proactive and informed responses.

Objectives

- **Track and Illustrate Changes in Data:** Build visualization tools that highlight how data distributions shift over specific periods.
- **Reveal Underlying Patterns:** Depict trends, anomalies, and structural changes in datasets, making concept drift easier to detect.
- **Support Data-Driven Insights:** Provide visual narratives to help users interpret the potential impact of drift on business or operational decisions.

Approach

1. **Data Collection and Storage:** Gather time-stamped datasets from relevant domains (finance, IoT, web analytics), ensuring easy comparability across different time slices.
2. **Visualization Techniques:**
   - *Distribution Plots:* Histograms, kernel density estimates, or boxplots to compare feature changes across time windows.
   - *Time-Series Charts:* Line or bar charts to reveal trends and potential drift points over time.
   - *Multi-Dimensional Views:* PCA or parallel coordinates to visualize shifts in multiple variables simultaneously.
3. **Interactive Dashboard:**
   - *User-Friendly Controls:* Filters, sliders, and drop-down menus for drilling into specific dates, ranges, or subgroups.
   - *Highlighting & Tooltips:* Automated emphasis on significant changes, with contextual tooltips explaining the magnitude or direction of drift.
   - *Temporal Animations:* Time-lapse visualization features to watch concept drift unfold in near real time. If you read this you know what to do or just ask in slack.

Expected Outcomes

- **Immediate Detection of Shifts:** Clear, at-a-glance identification of when data deviates from historical patterns.
- **Enhanced Interpretability:** Intuitive graphics and narratives to help both experts and non-experts spot trends or anomalies.
- **Proactive Decision-Support:** A basis for timely interventions—such as revisiting processes, updating policies, or exploring new opportunities influenced by evolving data.

Future Directions

- **Scalability & Automation:** Implement real-time dashboards that handle large, streaming datasets without sacrificing performance or clarity.
- **Collaborative Features:** Introduce sharing and annotation capabilities to enable team-wide discussion and response to observed drift.