Recognizing Patterns over Streams with Imprecise Timestamps

Haopeng Zhang, Yanlei Diao, Neil Immerman
University of Massachusetts Amherst

Introduction & Model

Cases of Imprecise Timestamps

- Unknown or imprecise
- Granularity mismatch
- Clock synchronization

Temporal Uncertainty Model

- Time Domain: 0, 1/2, 1, 1/2, 3/4, 1, 1/2, 1/4, 0, 1/2, 1, 1/2, 3/4, 1, 1/2, 1/4

- Format: (Event Type, Event ID, Uncertainty-Interval) where: EU=[0,1/2), other attributes

A Point-based Framework

Pass 1: Finding the Match Signature

- Valid Lower Bound: \( e_l, v_l \leq max(e_{up}, v_l + 1, \text{lower}) \)
- Valid Upper Bound: \( e_u, v_u \leq e_{up}, v_u + 1, \text{upper} \)

1. Stream Expansion

2. Configuring a Pattern Matching Engine

Most powerful strategy to compute transitive closure.

Pass 2: Pruning Based on Upper Bounds

- Revised Upper Bound: \( e_u, v_u \leq e_{up}, v_u + 1, \text{upper} \)

Pass 3: Pruning Based on the Window

- Last event's maximum upper: \( T_u = e_u, v_u + W - 1 \)

Formal Semantics of Pattern Matching

- Pattern SEQ (A, B, C) within 4 minutes

- Signature: \((a_1, b_1, c_1, a_2, b_2, c_2)\)
- Time Range: [0, 1/2]
- Confidence: 4/225

Performance Evaluation

Synthetic Data

- Event-based Pattern-based
- Point-based
- Point-based

Case Study of Cluster Monitoring

- Cluster monitoring
- Mayreduce jobs are logged using UDN time
- Qiang monitoring system records the load every 75 sec.

RFID-based Object Tracking

- Pattern SEQ Object A, Object B, Object C WHERE a brand = 1 AND b brand = 1 AND c brand = 1 AND a location = 1 AND b location = 2 AND c location = 3 AND all = 0 AND W = 10
- \( P \) = probability of correct reading, \( I \) = reading intensity, \( T \) = uncertainty interval, \( K_i = (1 - \eta) \)

Complexity & Optimizations

- Equivalence and Complexity

- #PossibleWorlds = \( 2^N \)
- #PartialMatches = \( (BW)^N \)

An Event-based Framework

Pass 1: Finding the Match Signature

- Valid Lower Bound: \( e_l, v_l \leq max(e_{up}, v_l + 1, \text{lower}) \)

2. Configuring a Pattern Matching Engine

Most powerful strategy to compute transitive closure.

Pass 2: Pruning Based on Upper Bounds

- Revised Upper Bound: \( e_u, v_u \leq e_{up}, v_u + 1, \text{upper} \)

Pass 3: Pruning Based on the Window

- Last event's maximum upper: \( T_u = e_u, v_u + W - 1 \)

Any State Evaluation

- Selectivity-based optimization

A Event-based Framework

- Pattern SEQ (A, B, C) within 4 minutes

- Signature: \((a_1, b_1, c_1, a_2, b_2, c_2)\)
- Time Range: [0, 1/2]
- Confidence: 4/225

Filtering Techniques for Pattern-based

- Event-based Pattern-based
- Point-based
- Point-based

Case Study of Cluster Monitoring

- Cluster monitoring
- Mayreduce jobs are logged using UDN time
- Qiang monitoring system records the load every 75 sec.

RFID-based Object Tracking

- Pattern SEQ Object A, Object B, Object C WHERE a brand = 1 AND b brand = 1 AND c brand = 1 AND a location = 1 AND b location = 2 AND c location = 3 AND all = 0 AND W = 10
- \( P \) = probability of correct reading, \( I \) = reading intensity, \( T \) = uncertainty interval, \( K_i = (1 - \eta) \)