GRETA: Graph-based Real-time Event Trend Aggregation

Olga Poppe\textsuperscript{1}, Chuan Lei\textsuperscript{2}, Elke A. Rundensteiner\textsuperscript{3}, and David Maier\textsuperscript{4}

\textsuperscript{1}Microsoft Gray Systems Lab, \textsuperscript{2}IBM Research AI, \textsuperscript{3}Worcester Polytechnic Institute, \textsuperscript{4}Portland State University

VLDB, August 29, 2018

Supported by NSF grants IIS 1018443, IIS 1343620, IIS 1560229, CRI 1305258
Motivation – Algorithmic Trading

Goal: Reliable actionable insights about the stream

Solution: Each event is considered in the context of other events in the stream

Picture source: http://www.businessxack.com/how-to-know-the-stock-market-trend/1303
Algorithmic Trading

- **Single event** = Single stock value
- **Event sequence** = Stock down trend of fixed length
- **Event trend** = Stock down trend of any length

Algorithmic Trading

- **Single event** = Single stock value
- **Event sequence** = Stock down trend of fixed length
- **Event trend** = Stock down trend of any length

**Picture source**: http://www.businessxack.com/how-to-know-the-stock-market-trend/1303
Algorithmic Trading

- **Single event** = Single stock value
- **Event sequence** = Stock down trend of fixed length
- **Event trend** = Stock down trend of any length

**Picture source**: http://www.businessxack.com/how-to-know-the-stock-market-trend/1303
**Algorithmic Trading**

- **Single event** = Single stock value
- **Event sequence** = Stock down trend of fixed length
- **Event trend** = Stock down trend of any length under the **skip-till-next-match** semantics*

Event Trends in Other Streaming Applications

- **Traffic control**
  - Event trend: Aggressive driving

- **Health care**
  - Event trend: Irregular heart rate

- **Cluster monitoring**
  - Event trend: Uneven load distribution

- **E-commerce**
  - Event trend: Items often bought together

- **Stock market**
  - Event trend: Head-and-shoulders

- **Financial fraud**
  - Event trend: Circular check kite

Worcester Polytechnic Institute
Complexity of Event Trend Analytics

Existing trends

\[ e \]
Complexity of Event Trend Analytics

Existing trends

Worcester Polytechnic Institute
Complexity of Event Trend Analytics

- Exponential number of trends
- Arbitrary length of a trend
- Complex event inter-dependencies in a trend

=> Exponential time complexity
Existing Two-Step Approaches

Step 1: Event Trend Construction

Step 2: Event Trend Aggregation

Event Trend Aggregation Query

RETURN sector, COUNT(*)

PATTERN Stock S+

WHERE [company] AND S.price > NEXT(S).price

GROUP-BY sector WITHIN 30 min SLIDE 1 min

Transaction event
- Sector id
- Company id
- Price
- Time


Worcester Polytechnic Institute
GRETA: Graph-based Real-time Event Trend Aggregation

**Graph-Based Event Trend Aggregation**

- a1: 1
- a2: 1
- b3: 2
- b4: 1
- a5: 3
- a6: 3
- b7: 3
- b8: 6

**Event Trend Aggregation Query**

- RETURN sector, COUNT(*)
- PATTERN Stock S+
- WHERE [company] AND S.price > NEXT(S).price
- GROUP-BY sector WITHIN 30 min SLIDE 1 min

**Event Stream**

- Transaction event
  - Sector id
  - Company id
  - Price
  - Time

**Quadratic time & linear space complexity**

Final count: **12**
Graph Template

Nested Kleene Pattern

\[ P = (SEQ(A+, B)) + \]

\textbf{Start type} \hspace{1cm} \textbf{End type}

\begin{itemize}
  \item \textbf{a’s} are preceded by \textit{a’s} and \textit{b’s}
  \item \textbf{b’s} are preceded by \textit{a’s}
\end{itemize}
Graph-Based Trend Aggregation

Event trends: (a1, }

Final count: 0

Worcester Polytechnic Institute
Graph-Based Trend Aggregation

Event trends: (a1,b2)

Final count: 1
Graph-Based Trend Aggregation

Event trends:
(a1,b2,a3, (a1,a3, (a3,

Final count: 1
Graph-Based Trend Aggregation

Event trends:
(a1,b6),...
(a1,a3,b6),...
(a1,b2,a3,b6),...
(a1,b2,a3,a4,b6)

Final count: 11
Graph-Based Trend Aggregation

**Our GRETA approach**

- Quadratic time & linear space complexity
- Final count: 43

**Event trends:**
- (a1,b8),...
- (a1,a3,b8),...
- (a1,b6,a7,b8),...
- (a1,a3,a4,a7,b8),...
- (a1,b2,a3,a4,a7,b8),...

**Existing two-step approaches**

- Exponential time & space complexity

Worcester Polytechnic Institute
Experimental Setup

Execution infrastructure:
Java 7, 1 Linux machine with 16-core
3.4 GHz CPU and 128GB of RAM

Data sets:

- **ST**: Stock real data set
  Event trends = Stock market trends

- **LR**: Linear road benchmark data set
  Event trends = Vehicle trajectories

- **CL**: Cluster monitoring synthetic data set
  Event trends = Load distribution trends

**ST**: Stock trade traces. http://davis.wpi.edu/datasets/Stock Trace Data/
Event Aggregation Approaches

Existing two-step approaches first construct all event trends and then aggregate them.

Flink is a popular open-source streaming engine that supports event pattern matching but not Kleene closure. Thus, we flatten our queries.

https://ink.apache.org/

SASE supports both Kleene closure and aggregation but does not optimize aggregation of Kleene matches.


CET finds the middle ground between CPU time and memory usage of event trend detection. It does not support aggregation of event trends.

GRETA is a win-win solution that
• achieves **4 orders of magnitude speed-up** compared to all existing approaches and
• uses **50-fold less memory** than SASE
Contributions

We are the first to compute aggregation of Kleene closure matches over event streams with optimal time complexity

1. **GRETA graph** compactly encodes all event trends matched by expressive Kleene queries

2. Graph-based event trend aggregation with quadratic time complexity

3. 4 orders of magnitude speed-up and 8 orders of magnitude memory reduction compared to existing approaches
Questions?