VISOR: Visualizing Summaries of Ordered Data

Giovanni Mahlknecht, Michael Böhlen, Anton Dignös, Johann Gamper

**Problem Description**

- Given an **Ordered Dataset**, approximate the dataset by \( k \) segments
- Each segment is assigned a **constant value**, derived from merging multiple consecutive input tuples
- Summarization depends on multiple parameters
  - reduction size \( k \)
  - summarization technique
  - error measure
- Contribution
  - Tool to visualize different summarization techniques
  - Visualization of induced error in function of summary size \( k \)
  - Comparison of summarization methods: VOPT, XOPT, PAA

**Scenario: Comparing Summaries**

- **Dataset**
  - View
  - Data and Summary FLIGHTS
  - Use of VISOR
  - Statistics
    - \( k=17 \): VOPT: RMSE 59.4 (23.71% of max) XERR 250.13 (66.85% of max)
    - PAA-AVG: RMSE 204.0 (81.43% of max) XERR 463.65 (123.92% of max)

**\( \epsilon \)-Graph and \( \Delta \)-Graph**

- Each graph shows two error measures: RMSE and XERR
- **\( \epsilon \)-Graph**
  - Shows the error in function of \( k \): \( \epsilon(k) = m_e(D,k) \)
  - Useful to find appropriate \( k \)
  - Useful to compare error behavior of two summarization methods.
- **\( \Delta \)-Graph**
  - Shows the change of the error \( \Delta(k) = m_e(D,k) - m_e(D,k+1) \)
  - Useful to find points where the an in crease of \( k \)does not give much improvements

**Scenario: Exploration of Extremes**

- **Dataset**
  - View
  - Data and Summary WEKIT
  - Use of VISOR
  - Statistics
    - \( k=11 \): VOPT: RMSE 10.7 (15.88% of max) XERR 131.51 (63.02% of max)
    - XOPT-CENTER: RMSE 20.6 (27.26% of max) XERR 36.97 (21.33% of max)

**Summarization Methods**

- **OKS-Framework**
  - Covers a variety of data summary structures
  - Minimization Problem: approximate an ordered dataset by \( k \)segments s.t. each segment summarizes a set of contiguous data points and induced error is minimized
  - Error function \( m_e(D) \), Merge function \( \oplus \)
    - \( \text{name} \)
    - \( \epsilon \)
    - \( \oplus \)
      - \( \text{VOPT} \) variance avg
      - \( \text{XOPT-CENTER} \) |max – min| |min+(max–min)|/2
      - \( \text{XOPT-MAX} \) |max – min|
      - \( \text{XOPT-MIN} \) |min|
- **Piecewise Aggregate Approximation (PAA): Segments of equal length**

**Internals**

- Computation of OKS-summaries
  - Incremental computation
  - Dynamic Programming
  - Graph representation instead of matrix to keep small memory footprint