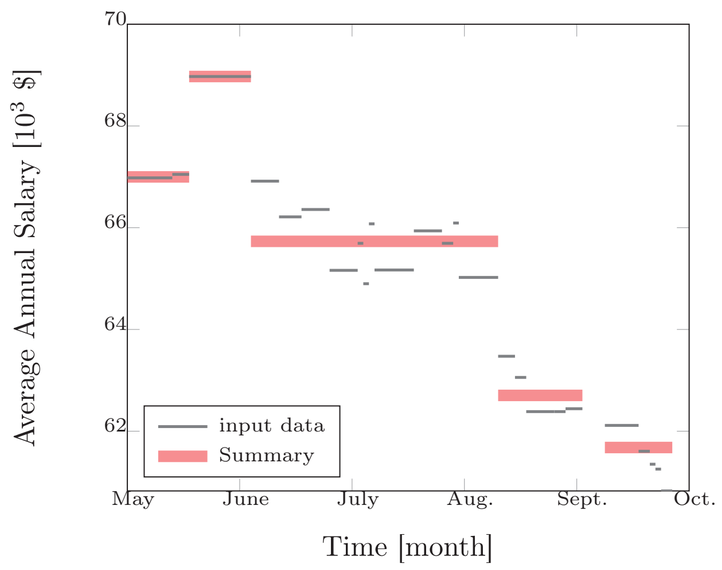
VISOR: Visualizing Summaries of Ordered Data

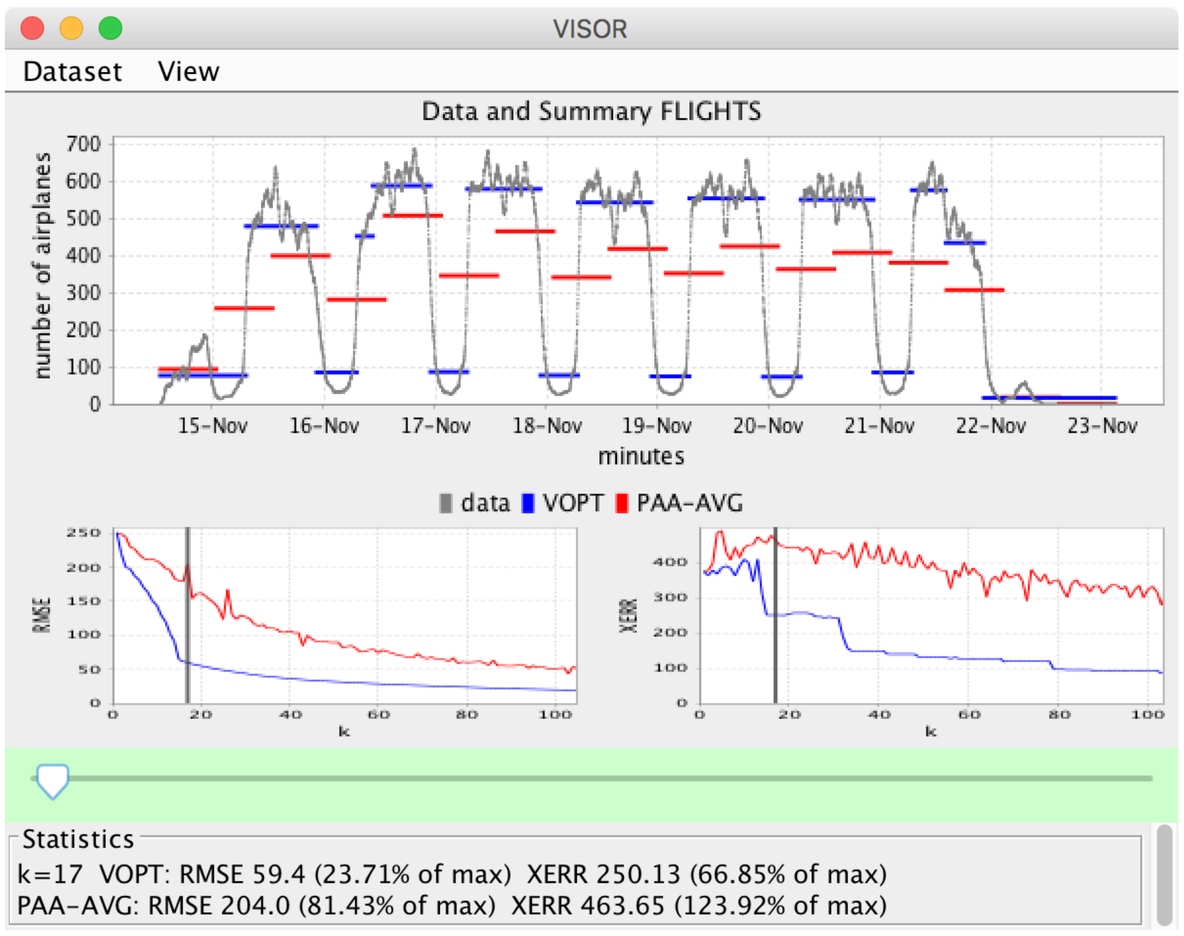
Giovanni Mahlknecht, Michael Böhlen, Anton Dignös, Johann Gamper

Problem Description

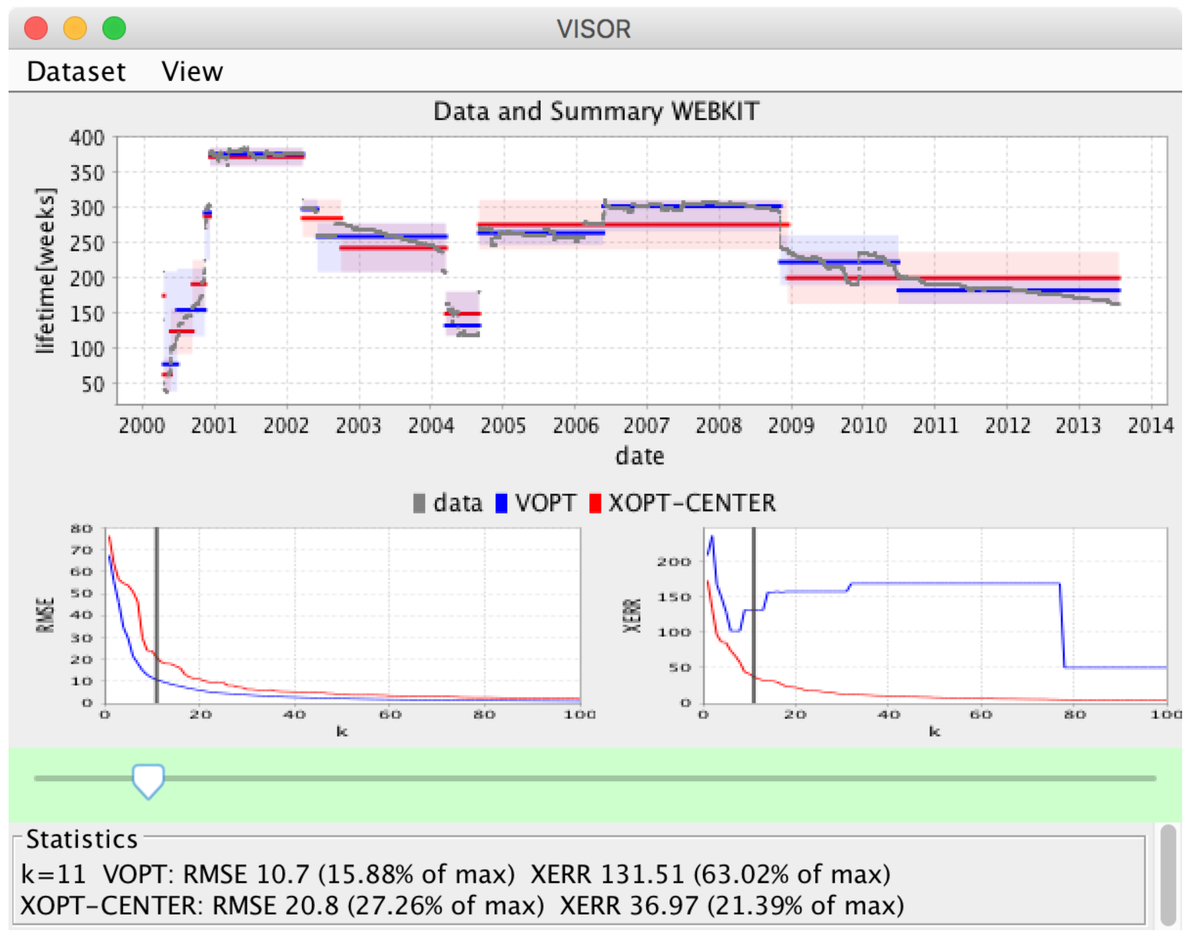
- Given an **Ordered Dataset**, **approximate** the dataset by k **segments**
- Each segment is assigned a **constant value**, derived from **merging multiple consecutive input tuples**
- Summarization depends on multiple parameters
  - reduction size k
  - summarization technique
  - error measure
- 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



Scenario: Exploration of Extremes



ϵ-graph and Δ-graph

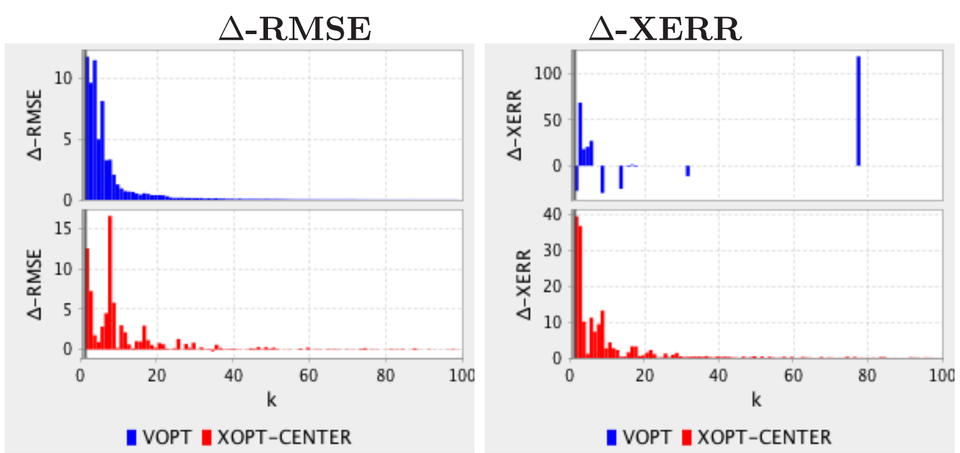
Each graph shows two error measures: **RMSE** and **XERR**

ϵ-graph

- Shows the error in function of k: $\epsilon(k) = me(\mathbf{D}, k)$
- Useful to find appropriate k
- Useful to compare error behavior of two summarization methods.

Δ-graph

- Shows the change of the error $\Delta(k) = me(\mathbf{D}, k) - me(\mathbf{D}, k{+}1)$
- Useful to find points where the an in crease of $k$does not give much improvements



Summarzation 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 $me(\mathbf{D})$, Merge function $\oplus$

name	e	$\oplus$
VOPT	variance	avg
XOPT-CENTER	$\lvert \max - \min \rvert$	min+(max−min)/2
XOPT-MAX	$\lvert \max - \min \rvert$	max
XOPT-MIN	$\lvert \max - \min \rvert$	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**

G. Mahlknecht, A. Dignös and J. Gamper, A scalable dynamic programming scheme for the computation of optimal k-segments for ordered data, Inf. Syst. (2016)
G. Mahlknecht, A. Dignös, and J. Gamper. Efficient computation of parsimonious temporal aggregation., ADBIS 2015, Poitiers France - Springer, 2015

SSDBM 2017, Chicago