Blending Contextual Data with Heterogeneous Time Dimensions for Improved Time Series Analysis

Saifullah Burero Jerry W. Sangma Anton Dignös
Johann Gamper

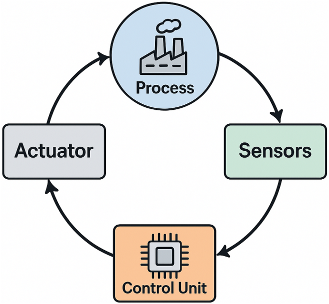
Free University of Bozen-Bolzano

DEXA 2025

Supported in part by the DIADEM project, funded by the Free University of Bozen-Bolzano

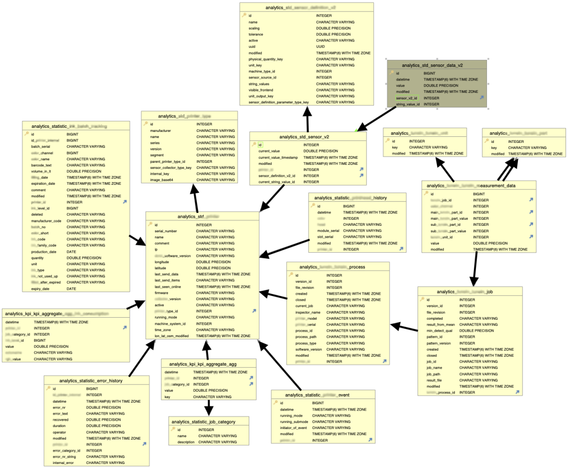
Motivation and Use Case

- Processes in industrial domains heavily rely on **sensor data**
  - Movement angles, speed, pressure, temperature, etc.
- Data is stored in database(s) for **analytics**
  - Monitoring, forecasting, predictive maintenance, etc.



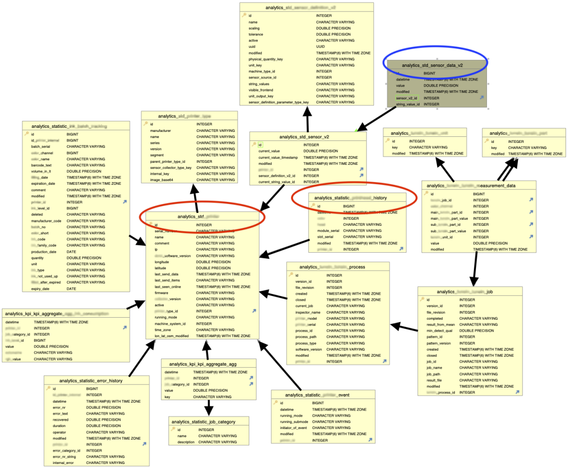
Motivation and Use Case

- Common data (warehouse) in an industrial setting



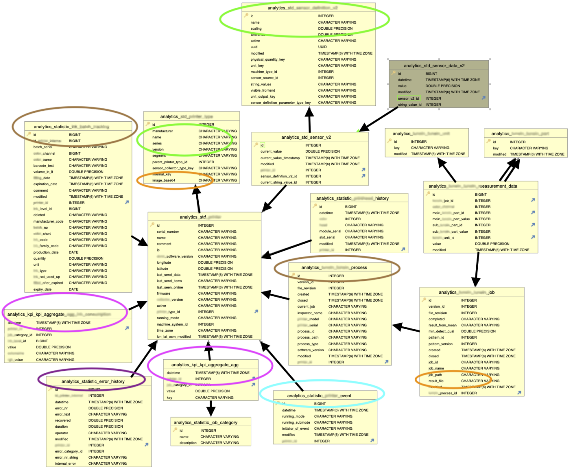
Motivation and Use Case

- Component(s) to monitor and sensor data



Motivation and Use Case

- Sensor data rarely comes alone – **context data**
- Images, static data, KPIs, process data, error history, events



Problem Setting

- **Application:** Monitor a Component
- **Task:** Forecast a measurement (time series)
- **Input:** Historical measurements and context

$$m_{t+x} = f\left(\ldots m_t, \ldots c_{j,f(t)}\right)$$

- **Challenge:** How to include context?

Related Work – Part I

- Many works show **benefit of incorporating contextual** data
  - Impact of weather conditions on electricity consumption[1],[2]
  - Weather data for water demand forcasting[3]
  - Air temperature for electricity load forecasting[4]
  - ...
- More precise, more robust, better explainability
- Approaches typically use **other time series as context**

[1] Akshaya Prabakar et al. "Applying machine learning to study the relationship between electricity consumption and weather variables using open data". In: *IEEE ISGT-Europe*. 2018.

[2] Ghassen Ben Brahim. "Weather Conditions Impact on Electricity Consumption in Smart Homes: Machine Learning Based Prediction Model". In: *IEEE ICEEE*. 2021.

[3] Ariele Zanfei et al. "A short-term water demand forecasting model using multivariate long short-term memory with meteorological data". In: *Journal of Hydroinformatics* (2022).

[4] Niaz Bashiri Behmiri, Carlo Fezzi, and Francesco Ravazzolo. "Incorporating air temperature into mid-term electricity load forecasting models using time-series regressions and neural networks". In: *Energy* (2023).

Related Work – Part II

- Data Fusion in time series[5]
- Fusion of data with **different modalities**
  - Time series
  - Text
  - Images
  - Audio
  - Videos
- Approaches typically **assume perfect alignment** with time series

[5] Manuel Mondal et al. "A survey of multimodal event detection based on data fusion". In: *VLDB J.* (2025).

Context with Heterogeneous Time Dimensions

- Time series are **measurements over regular time steps**[6]

  $X = \langle (x_1, 1), (x_2, 2), (x_3, 3), \dots, (x_n, n) \rangle$, where $x_i \in \mathbb{R}$

- Some contextual data is heterogeneous in time dimension
  - Error events are sparse/dense over time
    $E = \langle (\text{ERR}, 1), (\text{ERR}, 7) \rangle$
  - Processes/tasks last over some time span
    $P = \langle (\text{HEAT}, 1-4), (\text{MOVE}, 2-7) \rangle$
  - Environment data is less/more frequently sampled
    $T = \langle (35^\circ C, 1), (38^\circ C, 3), (39^\circ C, 5) \rangle$
  - Static properties have no time references
    $S = \{ (\text{sensor}_1, \pm 2^\circ C), (\text{sensor}_2, \pm 1^\circ C), (\text{sensor}_1, \text{BZ}) \}$

- We focus on **heterogeneity in time dimension**

[6] Or at least they are assumed to be.

Context with Heterogeneous Time Dimensions

- **Primary time series** is the signal to monitor
- Four types of heterogeneity in time dimensions[7]

Secondary time series
- Signals with different sampling rates

Static data
- Data without time reference

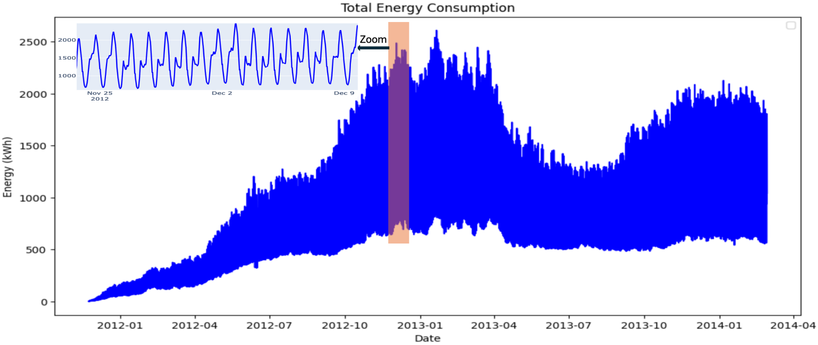
Event data
- Data that occurs very irregularly

Interval data
- Data that occurs over some period of time

[7] Saifullah Burero. "Integrating Heterogeneous Contextual Data for Enhanced Time Series Analysis". In: *EDBT 2025 PhD Workshop*. 2025.

Example & Use Case

- **Simplified use case:** electrical power consumption[8]
- **Primary time series:** total energy consumption



[8]Smart meter in London dataset `https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london` and `https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households`

Example & Use Case

- **Primary time series**
  Electrical power consumption
- **Secondary time series**
  Weather data (temperature, humidity)
- **Static data**
  (Location data, household information)
- **Event data**
  Bank holidays
- **Interval data**
  School vacations

Example & Use Case

- Datamodel

UK Bank Holidays

Type	Date	Location
Boxing day	2012-12-26	London
Easter Monday	2012-09-04	London
...	...	...
Summer Bank Holiday	2012-08-27	London

Weather Data

Time	Humidity	Temp	...	Summary	Location
2012-10-17 00:30:00	20	12		cloudy	London
2012-10-17 01:30:00	30	16		sunny	London
2012-10-17 02:30:00	25	17		sunny	London
...	...	...	...	...	...
2012-10-17 23:30:00	30	11	...	cloudy	London

House Hold Information

LCLid	Std or Tou	Acorn	Acorn_grouped	File	Location
MAC001074	std	ACORN_A	Affluent Achievers	Block_0	London
MAC001873	tou	ACORN_C	Financially Stretched	Block_9	London
MAC001047	std	ACORN_A	Affluent Achievers	Block_0	London
...	...	...	...	...	...
MAC001672	std	ACORN_D	Urban Adversity	Block_4	London

Electrical Consumption

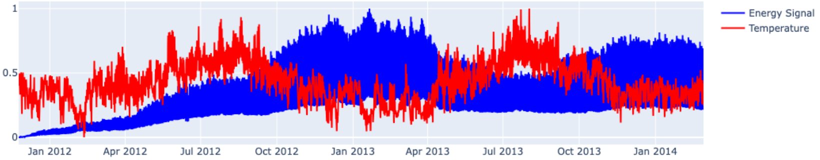
LCLid	Time	Energy(KWH)
MAC001074	2012-10-17 00:30:00	0.3
MAC001873	2012-10-20 10:30:00	1.2
MAC001047	2012-11-17 12:30:00	0.5
...	...	...
MAC001672	2012-12-13 00:30:00	0.09

School Vacations

Vacation Type	Start_Date	End_Date	Location
Summer Vacations	2012-06-01	2012-07-31	London
Christmas holidays	2012-12-25	2013-01-02	London
...	...	...	...
Autumn Holidays	2012-10-28	2010-11-01	London

Example & Use Case

- Impact of context (e.g., temperature)

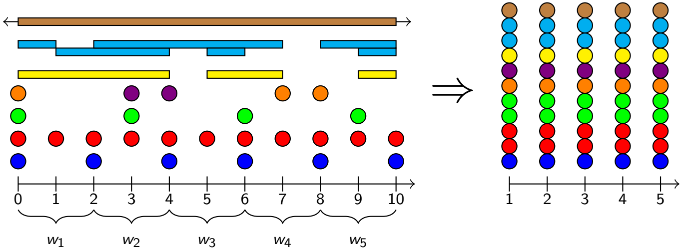


- People use electricity for heating (inverse relationship)

Blending Context with Heterogeneous Time Dimensions

Problem: Context has heterogeneous time dimensions

Solution: Create time series from context



Approach: Given granularity w

1. Align using aggregation/imputation
2. Compute lag information based on time dimension
3. Apply any (multivariate) time series analytics algorithm

Blending Context with Heterogeneous Time Dimensions

1. **Align** using aggregation/imputation
   - AVG/MAX/MIN for secondary time series
   - Count events
   - Count starting/ending periods
   - etc.
2. Compute **lag information based on time dimension**
   - Number of events in the past
   - Duration since last event
   - Period duration so far
   - etc.
3. Apply any **time series analytics** algorithm
   - Plethora of algorithms available

Blending Context with Heterogeneous Time Dimensions

- **Intuition**
  - For events occurrence rather than value is important
  - For intervals occurrence/duration is important
- **Industrial setting**
  - Current error (machine stop) indicates temperature reduction
  - Number of errors in the past indicate degradation
  - Number of current tasks (intervals) affect consumption
  - Duration of previous tasks (intervals) affect aging
- **Water demand**
  - Pipe maintenance implies unusual water flow
  - Current tourism season indicates high consumption
  - Duration of drought affects water demand
- **CPU monitoring**
  - Number of tasks (intervals) affects temperature
  - Number of interrupts affects delay
  - Many CPU fan error messages affect temperature

Small Evaluation – Setup

- **Dataset**: Electrical power consumption
- **Very simple task/model:** Random Forest regression
  - Next value (30min ahead)
  - $24^t h$ value (12h ahead)
  - $48^t h$ value (24h ahead)
- **Setting:** without and more or less context data

E	electrical power consumption only	time series
ES	E + consumers per group	time series
EW	E + weather data	time series
EE	E + bank holidays	event
EH	E + school vacations	interval
...	some more combinations	multiple
All	E + all contexts	multiple

- **Metric:** Improvement in MAE and MSE as compared to E

Small Evaluation – Results

Features	Improvement in %					
	Next Value		24th Value		48th Value	
	MSE	MAE	MSE	MAE	MSE	MAE
E	-	-	-	-	-	-
ES	20.59	12.04	31.85	16.68	20.63	10.58
EW	12.68	7.20	28.13	16.81	12.91	7.56
EWS	26.46	15.45	43.87	26.92	26.87	15.01
EE	0.17	0.05	0.04	0.02	0.06	0.01
EEWS	26.38	15.46	43.90	26.91	26.70	14.84
EH	17.19	9.76	20.62	11.96	17.43	9.53
ALL	28.00	16.14	44.92	27.83	27.20	15.59

Summary – Part I

- Transform **heterogeneous** time dimensions **into uniform time series**
- Integrates context for **traditional time series algorithms**
- Can be done **semi-automatically** (exploiting DB/DW)
- Context may be known in future – **future-known covariates**
- Challenges
  - Feature extraction (data management)
  - Feature selection (machine learning)
- Is a **"feature engineering" approach**

Related Work – Part III

- **Temporal Fusion Transformers** (TFT)[9]
  - Transformer-based approach
  - Multi-horizon time series forecasting
  - Dynamically selects important variables (gating mechanisms)
  - Captures temporal relationships (attention mechanisms)
  - Leverages historical patterns and future-known covariates
  - Handles mixed-data types (static + time series)
- Focus only on **static data** with time series

[9] Bryan Lim et al. "Temporal fusion transformers for interpretable multi-horizon time series forecasting". In: *International Journal of Forecasting* (2021).

Small Evaluation – Part II

- TFT with consumer groups as **static context**
- Multi-horizon electrical power consumption forecasting
- **Future covariates:** weekdays, hours, weekends, months
- Evaluation **with vs. without additional context**

Group	Improvement in %	
	MASE	RMSE
ACORN-	85.32	84.30
ACORN-U	9.75	5.81
Urban Adversity	52.02	47.83
...		
AVG – 7 groups	**48.40**	**44.97**

Conclusion

- **Context is important** for time series analysis
- Some context is **heterogeneous in the time dimension**
- Techniques for appropriate integration are required
  - Transformation-based (alignment + appropriate lag)
  - Transformer-based (alignment + appropriate treatment)

Future Work

- **Data management aspects**
  - Efficient integration of different time dimensions

    Interval relationship joins[10], range containment joins[11], etc.
- **Machine learning aspects**
  - Effective exploitation of different time dimensions

    Extension of TFT[12] architecture

[10] Danila Piatov et al. "Cache-efficient sweeping-based interval joins for extended Allen relation predicates". In: *VLDB J.* (2021).

[11] Anton Dignös et al. "Leveraging range joins for the computation of overlap joins". In: *VLDB J.* (2022).

[12] Bryan Lim et al. "Temporal fusion transformers for interpretable multi-horizon time series forecasting". In: *International Journal of Forecasting* (2021).

Thank you!

Please get in touch if you are interested!

DEXA 2025