Medical time series understanding and prediction

Myra Spiliopoulou

AIME @ Portoroz, June 2023

The Knowledge Management & Discovery Lab @ U Magdeburg

Research focus:

- Machine Learning methods for streams and time series with gaps – prediction and feature contribution
- Parsimonious usage of data and features – cost-aware active feature acquisition methods
- Design of human-understandable solutions

Application areas:

- Treatment outcome prediction in clinical data
- Patient and clinic phenotyping on high-dimensional data
- Prediction in mHealth apps

KMD Projects and Medical Partners

- ⋆ UNITI (2020-2023, EU-H2020): “Unification of Treatments and Interventions for Tinnitus Patients”
- ⋆ ImmunLearning (2019 - 2022, EFRE): “Entwicklung eines Tests zur Diagnostik für Immunkompetenz bei Senior*innen mit Hilfe von Data-Mining-Methoden” mit Univ Med Magdeburg

- · Human and animal learning (Leibnitz Institute of Neurobiology)
- · Learning on longitudinal epidemiological data (U Med Greifswald)
- · Phenotyping and response to treatment (CHARITE & UHREG)
- · Compliance in mHealth (UHREG)

Trajectories, traces and other time series

Predicting covid-19 incidences [a]

Figure 2 shows the course of the COVID-19 cases per 100,000 population transmitted to the RKI on the last 7 days in each of the federal states and in all of Germany. The values for the 7-day incidence in the federal states range from 799.6 per 100,000 population in Bremen to 239.8 per 100,000 population in Lower Saxony.

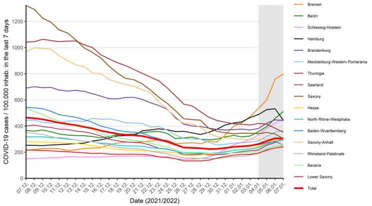


Figure 2: COVID-19 cases/100,000 inhabitants during 7 days in Germany by federal state and reporting date in the health offices (07/01/2022, 0:00 AM). The grey area delineates a range of dates with yet incomplete data, where changes in incidence are likely to occur.

Detecting:

→ wandering behavior [Lin et al., 2012]

→ hyper/hypoglycaemic events [Fox et al., 2018]

[a] https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Jan_2022/2022-01-07-en.pdf?__blob=publicationFile

Describing/Understanding:

→ how patients with polyneuropathy stand – as opposed to healthy people [Niemann et al., 2020c] [1] [2]
Figures 3 (left) and 1 (right) from [Niemann et al., 2020c]

[1] Figure 3 (left side). 'Temperature changes with study protocol recorded by sensor-equipped insole. . . .

[2] Figure 1 (right side), describing the experiment

Learning tasks on time series

Classification & Prediction | Description & Explanation | Anomaly detection

Learning tasks on time series

Classification & Prediction   Description & Explanation   Anomaly detection

Classification & Prediction

- Assign a label to a time series [Time series classification]
- Predict the value of a variable of interest using present values of other variables, and all the past [Stream classification]
- **Predict the next observation(s) from the past ones**
- Predict whether an event will occur [Survival analysis]
- Estimate the remaining time until an event [RUL]

Learning tasks on time series

Classification & Prediction Description & Explanation Anomaly detection

Classification & Prediction

- Assign a label to a time series [Time series classification]
- Predict the value of a variable of interest using present values of other variables, and all the past [Stream classification]
- **Predict the next observation(s) from the past ones**
- Predict whether an event will occur [Survival analysis]
- Estimate the remaining time until an event [RUL]

Description & Explanation

- **Understand how some variables predict the future of others**

$\rightarrow$ Understand how an intervention affects an outcome, and for whom

Learning tasks on time series

Classification & Prediction Description & Explanation Anomaly detection

Classification & Prediction

- Assign a label to a time series [Time series classification]
- Predict the value of a variable of interest using present values of other variables, and all the past [Stream classification]
- **Predict the next observation(s) from the past ones**
- Predict whether an event will occur [Survival analysis]
- Estimate the remaining time until an event [RUL]

Description & Explanation

- **Understand how some variables predict the future of others**

$\rightarrow$ Understand how an intervention affects an outcome, and for whom

Anomaly detection: Figure out whether

- a time series is anomalous
- a time series segment is anomalous
- a time series (segment) deviates from a pre-defined normality

Formalization for prediction [Parmezan et al., 2019, Section 3]

Time series Z: sequence of observations $Z = (z_1, z_2, \ldots, z_m)$ where $z_t \in R$

Two constraints: Z is discrete. Z is uniformly distributed over time [a].

[a] If not: [Cismondi et al., 2013] describe solutions for the alignment of misaligned time series

Important properties of a time series

- **Deterministic vs Stochastic:**
  - Deterministic $\leftarrow$ iff there is a mathematical function $f()$, such that the time series values are synthesized as $y = f(t)$
  - Nondeterministic/Stochastic $\leftarrow y = f(t, \varepsilon)$, where ε is a random term
- **Stationary:** iff it develops around a constant average

A time series has a

- **Trend:** iff there is a long-term increase or decrease in the observed values, following an arbitrary increase (resp. decrease) pattern
- **Seasonality:** iff, next to a trend, there are also patterns that repeat in relatively constant intervals (aka: cyclic patterns)
- **Residue:** a 'component' that captures the short term fluctuations – which are not systematic and cannot be predicted.

An example time series [Parmezan et al., 2019]

'Fig. 3(a) displays a real time series that expounds, in tons, the monthly chocolate production in Australia from January 1958 to December 1990. These measurements, provided by the Australian Bureau of Statistics, as well as all datasets adopted in this paper, are available at the ICMC-USP Time Series Prediction Repository [32].' [32] is: [Parmezan and Batista, 2014]

Figure 3a: the uppermost part of Figure 3 depicts the original time series of chocoloate production, as described in the text

Formalization cntd.

[Parmezan et al., 2019, Section 3]

We can express a time series Z at timepoint t as a composition of the trend, seasonality and residual components.

Additive decomposition

$$Z_t = T_t + S_t + R_t$$

where all components use the same quantity unit as Z_t.

Multiplicative decomposition

$$Z_t = T_t \times S_t \times R_t$$

where component T_t uses the same quantity unit as Z_t and the other two components are modifiers over the trend component.

Decomposing the example time series [Parmezan et al., 2019]

'Fig. 3(b) shows the trend, estimated using MA with 12 periods (monthly data), of the chocolate production time series;' MA:= Moving Average
Figure 3b: this part of the figure is directly below the original time series of chocoloate production, and it captures the trend, as described in the text.

The example time series decomposed [Parmezan et al., 2019]

Figure 3: complete figure, with the original time series, the trend component and two decompositions for seasonality and residue, as described in the text

Prediction workflow [Parmezan et al., 2019, Section 4]

Figure 3: the 6 steps of the time series prediction process

A few words on phenotyping

For a disease/disorder/condition:

- ? What distinct 'groups' of individuals are there?
- ? How does each group respond to a (given) treatment?
- ? How does the treatment affect each group?

→ What characterizes each group?

A few words on phenotyping

For a disease/disorder/condition:

? What distinct 'groups' of individuals are there?
? How does each group respond to a (given) treatment?
? How does the treatment affect each group?

$\rightarrow$ What characterizes each group?

Phenotype

the set of characteristics of a living thing, resulting from its combination of genes and the effect of its environment [BIOLOGY] [a]

all the observable characteristics of an organism that result from the interaction of its genotype (total genetic inheritance) with the environment. Examples of observable characteristics include behaviour, biochemical properties, colour, shape, and size. The phenotype may change constantly throughout the life of an individual because of environmental changes and the physiological and morphological changes associated with aging. Different environments can influence the development of inherited traits (as size, for example, is affected by available food supply) . . . [GENETICS] [b]

---

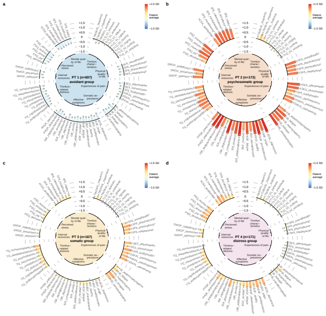
[a] Oxford Advanced Learner's Dictionary, https://www.oxfordlearnersdictionaries.com/definition/english/phenotype (2023-02-22)

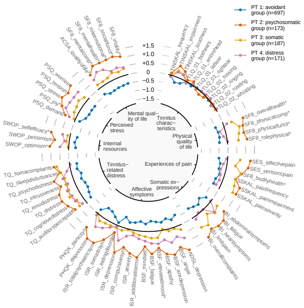
[b] Encyclopaedia Britannica, https://www.britannica.com/science/phenotype (2023-02-22)

Phenotype $\equiv$ cluster over a feature space of observables ?

Phenotyping at baseline [Niemann et al., 2020b]

Patients are partitioned in four phenotypes (by X-Means): radial barcharts (Figure to the left)[3] & radial line chart as overview (Figure to the right) [4]





---

[3] **Figure 1. Radial barcharts visualizing the 4 phenotypes. A** Phenotype 1 (PT1) characterises the patient subgroup with lowest health burden among all phenotypes. **B** PT2 represents the most suffering subgroup, with all of the psychosomatic and somatic measurement averages exceeding the population mean $+0.5$ standard deviations (SD). PT3 (**C**) exhibits above population average scores for somatic indicators whereas PT4 (**D**) is characterised by increased distress scores, including subjective stress and perceived quality of life. Bars are arranged in a circular layout. The height of a bar shows a feature's z-score normalised within-cluster average, and the grey line centred at the top of the bar illustrates the 95% confidence interval. The colour of a bar represents the difference of the within-cluster average from the overall patient average (PA), from -1.5 SD below PA (dark blue) to PA (yellow) and $+1.5$ SD above PA (bright red). Features were grouped into 9 categories defined by tinnitus experts. The categories are shown within the inner circle. See subsection Features for a description of each questionnaire and the extracted features.

[4] **Figure 2. Radial line chart juxtaposing the 4 phenotypes.** In the chart, a point shows a feature's (z-score normalised) within-phenotype average. In each feature category (labels in inner circle), points are connected with line segments. Points and lines are coloured by cluster.

. . . on phenotypes

In a nutshell: Phenotypes refer to *observable* properties of *relevance* – here, of relevance to diagnostics and treatment.
Phenotypes can be built with a clustering mechanism. They must be verified from the application domain.

Phenotypes that respond differently to treatment are of particular interest.
$\Rightarrow$ It is of interest to predict the treatment outcome per phenotype and/or to use the phenotype as predictor variable for **treatment outcome**.

. . . on outcomes

In a nutshell: We refer to a 'treatment outcome', usually a score, or to 'treatment response' that can be positive (hopefully), none or negative (let's hope not).

Outcome prediction in a medical center [Niemann et al., 2020a]

- To what extent do baseline features predict treatment outcome?
- and how many are necessary for a good prediction?

Figure 1 from [Niemann et al., 2020a]

Outcome prediction – Post-Learning step II [5] [Niemann et al., 2020a]

Figure 6 from [Niemann et al., 2020a]

[5] **Figure 6. Marginal feature importance.** (A) Average cross-validation AUC ($\pm$ SD) of a row refers to the performance of a GBT model trained on the feature subset that consists of the feature depicted on the y-axis label and all features of the above rows. The ordering of features is according to mean absolute SHAP value magnitude (cf. Fig 4A). (B) Network visualization illustrating 3 groups among the 26 selected features of the best model with high intra-group correlation. 8 features (predominantly SOZK features) without pairwise correlation of magnitude 0.5 or higher were dropped.

. . . on outcomes and predictive features

In a nutshell: Of interest are the features that contribute to outcome/response prediction; including features that are in causal relationship to the outcome/response.
Phenotypes are of interest, too. But they must be first verified across centers.

...on outcomes and time

In a nutshell: Outcome prediction is one of the prediction tasks in the medical context. It involves at least two timepoints, before and after treatment.

Occasionally, we have additional time points: *during* treatment and *after* treatment. Then, we have conventional (though small) time series.
Next: Prediction tasks on time series – formalization.

Entity

An entity is: any object with an ID

- ⋆ described in a static space – unchanging properties, e.g. birthdate
- ⋆ associated with a sequence of observations in a dynamic space

- · The time horizon is fixed – or not.
- · The sequences of the entities start at the same timepoint zero – or not.

Modeling time

Given is a time unit, and a sequence of time points $t_0, t_1, t_2, \ldots, t_n$ where the step between two points corresponds to one time unit, and t_0 is the zeroth time point.

A time series with *absolute time points* has the form $T = v_{t_0}, \ldots, v_{t_n}$ where $v_{\_}$ is a vector in the multidimensional feature space $F_1 \times \ldots \times F_k$.

Each of $F_1, \ldots, F_k$ contains also the special value NULL.

Alignment at zero

For two time series T_A, T_B, if the t_0 of T_A is the same absolute time point as the t_0 of T_B, then T_A, T_B are *aligned at zero*.

Example I from [Chang et al., 2021]

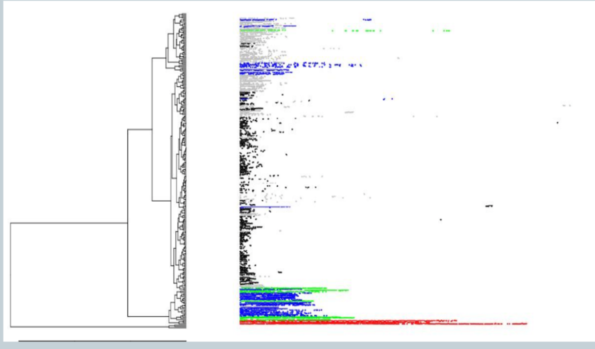
Figure 2 from [Chang et al., 2021]

Sampling rate

Two time series may have different *sampling rates* (and thus time units), even if they belong to the same entity.

Modeling time: Alignment at zero [Unnikrishnan, 2017]

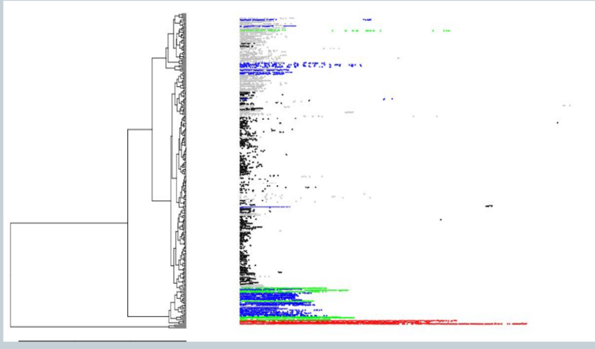
Example: Time series of the users of the TrackYourTinnitus mHealth app, collected over a period of $\geqslant$ a year – one colored dot per observation, one 'empty' dot per time point without observation



Modeling time: Alignment at zero

[Unnikrishnan, 2017]

Example: Time series of the users of the TrackYourTinnitus mHealth app, collected over a period of $\geqslant$ a year – one colored dot per observation, one 'empty' dot per time point without observation



Consequences of aligning at zero

- The time units are retained; the absolute time point positions are not.
- Information about seasonality is lost.

Modeling time without time units

A time series with absolute time points $T = v_{t_0}, \ldots, v_{t_n}$ where $v_{\_}$ can be mapped to the ordered sequence $T_{seq} = v_0, v_1, \ldots, v_{n'}$, where $n' \leqslant n$ and $v_{\_}$ is a vector in the multidimensional NULL-free feature space $F'_1 \times \ldots \times F'_k$ [6].

[6] $F'_i \leftarrow F_i \setminus \{NULL\}$, for each $i = 1, \ldots, k$

Modeling time without time units

A time series with absolute time points $T = v_{t_0}, \ldots, v_{t_n}$ where $v_{\_}$ can be mapped to the ordered sequence $T_{seq} = v_0, v_1, \ldots, v_{n'}$, where $n' \leqslant n$ and $v_{\_}$ is a vector in the multidimensional NULL-free feature space $F'_1 \times \ldots \times F'_k$ [6].

Consequences of collapsing time series to ordered sequences

- All time series become aligned at zero.
- The more gaps the original time series had, the shorter becomes the collapsed ordered sequence.
- Information about seasonality, periodicity and circardian phenomena is lost.

[6] $F'_i \leftarrow F_i \setminus \{NULL\}$, for each $i = 1, \ldots, k$

Modeling time without time units

A time series with absolute time points $T = v_{t_0}, \ldots, v_{t_n}$ where $v_{\_}$ can be mapped to the ordered sequence $T_{seq} = v_0, v_1, \ldots, v_{n'}$, where $n' \leqslant n$ and $v_{\_}$ is a vector in the multidimensional NULL-free feature space $F'_1 \times \ldots \times F'_k$ [6].

Consequences of collapsing time series to ordered sequences

- All time series become aligned at zero.
- The more gaps the original time series had, the shorter becomes the collapsed ordered sequence.
- Information about seasonality, periodicity and circardian phenomena is lost.

In the following, we focus on time series with relative time units that are already aligned at zero.

[6] $F'_i \leftarrow F_i \setminus \{NULL\}$, for each $i = 1, \ldots, k$

Misalignment and missingness in TS [Cismondi et al., 2013]

[7] The figures are remakes of Figure 1 from [Cismondi et al., 2013]

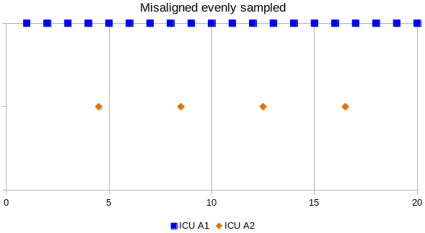
Misalignment and missingness in TS [Cismondi et al., 2013]

Time series context: Patient monitoring over a period of time T [7]

Misaligned evenly sampled as in

ICU [Cismondi et al., 2013]

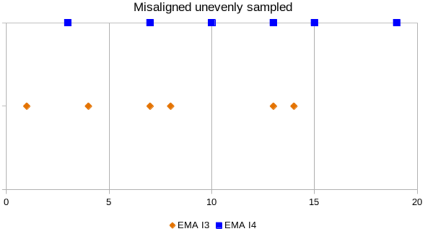
- Blood pressure: hourly
- Lab tests: every 4 hours



Misaligned unevenly sampled as in

TrackYourTinnitus [Probst et al., 2017]

- 8 questions, all answers optional
- each one answered independently
- triggered more than once a day



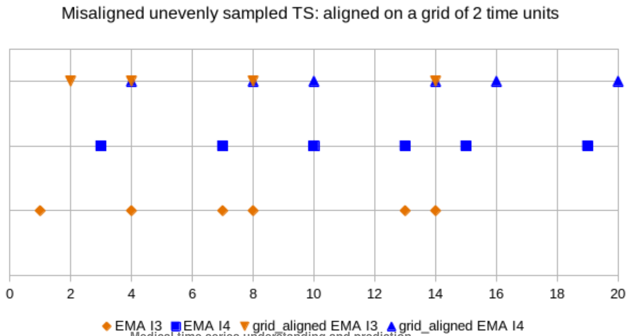
[7] The figures are remakes of Figure 1 from [Cismondi et al., 2013]

Alignment of unevenly sampled data: Gridding [Cismondi et al., 2013]

- A sampling rate is specified and a grid defined on the time axis.
- Each value of *each variable* is shifted to occupy the closest grid position[a]

⇒ all variables get values **or NULLs** at the same positions – of the grid.

[a] If more than one values fall into the same grid position, a predefined mechanism is used to determine the final value at the grid position.

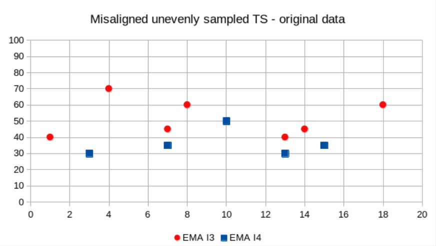


Alignment of unevenly sampled data: Templating [Cismondi et al., 2013]

- One of the variables is chosen as template – usually the one with the highest sampling rate.
  The time points at which this variable is sampled are used as grid positions.
- Each value of each other variable is shifted as in gridding.

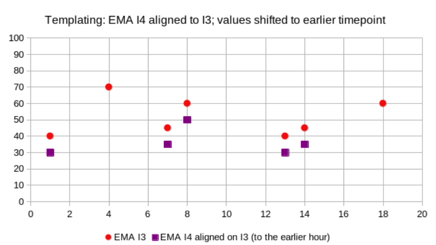
Example on templating

Assume a bivariate TS (variables: 'EMA I3' & 'EMA I4') that is misaligned and unevenly sampled.



↓

Assume that sampling is done in intervals of 8 hours, and that I4 is always sampled after I3, at most 2 hours later. ⇒ Templating



Missing and/or Observed [NOT] at Random [Cismondi et al., 2013]

Let D be a dataset with n variables, let Y be the one whose missingness we investigate, and X the remaining variable(s).

For each variable, we have observed values and missing values, i.e.

$$Y = \{Y_{obs}, Y_{miss}\} \qquad \forall X \neq Y : X = \{X_{obs}, X_{miss}\}$$

Missing and/or Observed [NOT] at Random [Cismondi et al., 2013]

Let D be a dataset with n variables, let Y be the one whose missingness we investigate, and X the remaining variable(s).

For each variable, we have observed values and missing values, i.e.

$$Y = \{Y_{obs}, Y_{miss}\} \qquad \forall X \neq Y : X = \{X_{obs}, X_{miss}\}$$

MAR, MCAR, MNAR for Y

- MAR: Probability of Y_{miss} is independent of Y; can depend on X.
- MCAR: Probability of Y_{miss} is independent of Y and of X.
- MNAR: Probability of Y_{miss} depends on (present/absent values of) Y; can depend on X.

Missing and/or Observed [NOT] at Random [Cismondi et al., 2013]

Let D be a dataset with n variables, let Y be the one whose missingness we investigate, and X the remaining variable(s).

For each variable, we have observed values and missing values, i.e.

$$Y = \{Y_{obs}, Y_{miss}\} \qquad \forall X \neq Y : X = \{X_{obs}, X_{miss}\}$$

MAR, MCAR, MNAR for Y

- MAR: Probability of Y_{miss} is independent of Y; can depend on X.
- MCAR: Probability of Y_{miss} is independent of Y and of X.
- MNAR: Probability of Y_{miss} depends on (present/absent values of) Y; can depend on X.

Dependency of Y on X

- ONAR: Likelihood of Y_{miss} depends on X $\quad Y_{miss} \perp\!\!\!\perp X$
- OAR: Likelihood of Y_{miss} does not depend on X $\quad Y_{miss} \not\perp\!\!\!\perp X$

  i.e. no difference on X_{obs} between Y_{miss} and Y_{obs}

Classifying TS on missingness [Cismondi et al., 2013]

The missing values dilemma: Impute the missing values, in order to increase statistical power, **vs** leaving the NaNs as they are, because there are significant differences between observed and unobserved data

Classifying TS on missingness [Cismondi et al., 2013]

The missing values dilemma: Impute the missing values, in order to increase statistical power, **vs** leaving the NaNs as they are, because there are significant differences between observed and unobserved data

$$\Downarrow$$

Classify-before-imputing

Classification Level 1: Truly Missing vs Lack of Sampling

Generalizing from [Cismondi et al., 2013]

For a variable Y with missing values:

- $\rightarrow$ Compute the mean sampling frequency of Y
- $\rightarrow$ Derive a pessimistic threshold τ for the interval in which a value would be expected [a] [b]
- $\rightarrow$ Span a grid that reflects the sampling frequency of the most frequently sampled variable.
- $\rightarrow$ For each gap of Y on the grid, check whether its size is less than τ:
  - If YES, then mark the empty grid position as LackOfSampling [c]
  - ELSE mark the empty grid position as TrulyMissing and check for recoverability

[a] The threshold proposed in [Cismondi et al., 2013] is twice (the mean sampling frequency $+$ 95% of the confidence interval in which a value is expected).

[b] If we know that a sample value is taken every 2 hours but some delay $\leqslant$ 1 hour might occur, then we can set $\tau = 3$ hours.

[c] Then, alignment might be used.

Classification Level 1: Truly Missing vs Lack of Sampling

Generalizing from [Cismondi et al., 2013]

For a variable Y with missing values:

- $\rightarrow$ Compute the mean sampling frequency of Y
- $\rightarrow$ Derive a pessimistic threshold τ for the interval in which a value would be expected [a] [b]
- $\rightarrow$ Span a grid that reflects the sampling frequency of the most frequently sampled variable.
- $\rightarrow$ For each gap of Y on the grid, check whether its size is less than τ:
  - If YES, then mark the empty grid position as LackOfSampling [c]
  - ELSE mark the empty grid position as TrulyMissing and check for recoverability

[a] The threshold proposed in [Cismondi et al., 2013] is twice (the mean sampling frequency + 95% of the confidence interval in which a value is expected).

[b] If we know that a sample value is taken every 2 hours but some delay $\leqslant$ 1 hour might occur, then we can set $\tau = 3$ hours.

[c] Then, alignment might be used.

Is the truly missing value recoverable ?

Classification Level 2: Recoverability

Generalizing from [Cismondi et al., 2013]

Is there a (set of) variable(s) X so that $Y_{miss} \perp\!\!\!\perp X$?

→ Identify a set of variables X that can distinguish between Y_{miss} and Y_{obs} [a]

→ Check whether X is empty:

- If YES, then mark Y_{miss} as 'recoverable [b]
- ELSE mark Y_{miss} as 'non-recoverable' and do not attempt to derive them. [c]

---

[a] This is a learning problem.

[b] Then, Y_{miss} & Y_{obs} have the same behaviour given X, so X and Y_{obs} can recover Y_{miss}.

[c] This means that X dictated the decision to not sample Y at the location(s) of Y_{miss}.

Classification Level 2: Recoverability

Generalizing from [Cismondi et al., 2013]

Is there a (set of) variable(s) X so that $Y_{miss} \perp\!\!\!\perp X$?

$\rightarrow$ Identify a set of variables X that can distinguish between Y_{miss} and Y_{obs} [a]

$\rightarrow$ Check whether X is empty:

- If YES, then mark Y_{miss} as 'recoverable [b]
- ELSE mark Y_{miss} as 'non-recoverable' and do not attempt to derive them. [c]

[a] This is a learning problem.

[b] Then, Y_{miss} & Y_{obs} have the same behaviour given X, so X and Y_{obs} can recover Y_{miss}.

[c] This means that X dictated the decision to not sample Y at the location(s) of Y_{miss}.

Approach of [Cismondi et al., 2013] for classification level 2

$\rightarrow$ Split the data (excluding Y) into training set (,validation set) and test set

$\rightarrow$ Create a (target) variable Y' that takes the value 0 for each observed Y_{obs} and the value 1 for each empty location, i.e. Y_{miss}

$\rightarrow$ Build up a set of fuzzy rules that distinguish between the two values of the target variable

$\rightarrow$ Apply the set of rules on the held-out dataset

Missing values: to recover or not to recover ?

Example cntd

Assume that the data are EMA (diary entries), inserted after a reminder from an mHealth app; the user decides to ignore the reminder (Y/N), and this decision is not transparent $\Rightarrow$ Classification level 1: TrulyMissing

Can we recover the missing values of I4 in the cases where I3 is recorded?

A simplistic classification rule: `IF I3` $\geqslant$ `60 THEN` Y'_{miss}`=1` (some users)
$\Rightarrow$ missing values of I4 are not recoverable

Misalingment and MNAR

mostly from [Cismondi et al., 2013]

Takeaway message: Misalignments and missing values in variables of multivariate time series can be rectified with help of other variables.

The rectification demands some background knowledge about the sampling process of each variable, adequately large samples (to make reliable models) and a careful study of the dependencies (also: causalities) among the variables.

Imputation of missing clinical data: a competition [Luo, 2022]

Why?

$\surd$ MICE [Van Buuren and Groothuis-Oudshoorn, 2011]
BUT multivariate clinical data may be also misaligned
and there are new approaches around, cf [Luo, 2022]

Imputation of missing clinical data: a competition [Luo, 2022]

Why?

$\surd$ MICE [Van Buuren and Groothuis-Oudshoorn, 2011]
BUT multivariate clinical data may be also misaligned
and there are new approaches around, cf [Luo, 2022]

DACMI March-June 2019

the Data Analytics Challenge on Missing data Imputation
built on the ICU data of MIMIC-III [Johnson et al., 2016]

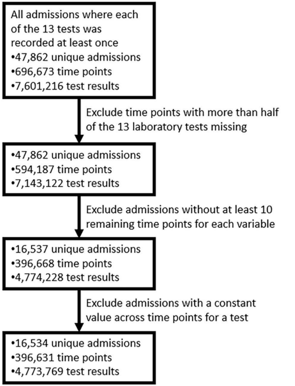


Figure 1. Flowchart of DACMI challenge data generation. The flowchart describes the filters applied and the number of data points taken forward at each step.

Semantics of the DACMI dataset [Luo, 2022]

All admissions in MIMIC-III, where each of the 13 predefined blood tests [8] were performed at least once.

$\rightarrow$ 16,534 ICU admissions 396,000+ time points 4,773,000+ test results

Train/test: 50%-50% of the admissions

$\swarrow$ $\searrow$

Masked data

For each admission AND for each labtest

one recorded result was selected at random and removed

$\rightarrow$ 13 results (one per labtest) removed per patient in the testset

---

[8] These are tests that are frequently done in ICUs for diagnosis and condition monitoring.

Semantics of the DACMI dataset [Luo, 2022]

All admissions in MIMIC-III, where each of the 13 predefined blood tests [8] were performed at least once.
$\rightarrow$ 16,534 ICU admissions 396,000+ time points 4,773,000+ test results
Train/test: 50%-50% of the admissions

↙ ↘

Masked data

For each admission AND for each labtest

one recorded result was selected at random and removed

$\rightarrow$ 13 results (one per labtest) removed per patient in the testset

Natively missing data

[8] These are tests that are frequently done in ICUs for diagnosis and condition monitoring.

Semantics of the DACMI dataset [Luo, 2022]

All admissions in MIMIC-III, where each of the 13 predefined blood tests [8] were performed at least once.

$\rightarrow$ 16,534 ICU admissions 396,000+ time points 4,773,000+ test results

Train/test: 50%-50% of the admissions

↙ ↘

Masked data

For each admission AND for each labtest

one recorded result was selected at random and removed

$\rightarrow$ 13 results (one per labtest) removed per patient in the testset

Natively missing data

Entries with tests that were NOT performed for a specific patient in the testset:

! Values were imputed by the DACMI participants

$\times$ but were ignored during the evaluation.

[8] These are tests that are frequently done in ICUs for diagnosis and condition monitoring.

Competition and evaluation [Luo, 2022]

12 competitors from: Asia (3), North America (7), Europe (2)

Organisations: Insurance (1), Pharmaceutical (1), Univ (9), Tech (1)

Competition and evaluation [Luo, 2022]

12 competitors from: Asia (3), North America (7), Europe (2)

Organisations: Insurance (1), Pharmaceutical (1), Univ (9), Tech (1)

Performance reference: 3D-MICE

MICE: to impute values for one variable, the others are used as predictors

- regression: predictors are used to build a regressor towards the target
- sampling – using the joint probability distribution(s)

3D-MICE [Luo et al., 2018]: extends the regression-based MICE with GP; exploits cross-sectional and temporal relationships.

Competition and evaluation [Luo, 2022]

12 competitors from: Asia (3), North America (7), Europe (2)

Organisations: Insurance (1), Pharmaceutical (1), Univ (9), Tech (1)

Performance reference: 3D-MICE

MICE: to impute values for one variable, the others are used as predictors

- regression: predictors are used to build a regressor towards the target
- sampling – using the joint probability distribution(s)

3D-MICE [Luo et al., 2018]: extends the regression-based MICE with GP; exploits cross-sectional and temporal relationships.

Evaluation with normalized

Root Mean Square Deviation $nRMSD(t) = \sqrt{\dfrac{\sum_{a,j} I(a,t,j)\left(\frac{|X_{a,t,j}-Y_{a,t,j}|}{\max_j(Y_{a,t,j}-\min_j(Y_{a,t,j})}\right)^2}{\sum_{a,j} I(a,t,j)}}$

For admission a and labtest t at timepoint j, let:

- $X_{a,t,j}$ be the imputed value and $Y_{a,t,j}$ the true one
- $I_{a,t,j}$ be 1 if the value of labtest t of a at j was missing (masked!), else 0

The top-3 teams [Luo, 2022]

1st: Team 'Ping An' [Xu et al., 2020] (Asia, Insurance)

Trained one model per labtest using 5 feature spaces: (F1) values of the other labtests at the same timepoint; (F2) current timepoint and time index; (F3) duration between current timepoint and the previous and the next one; (F4) values of the 13 labtests in the previous 3 and in the next 3 timepoints; (F5) max, min and mean labtest value for each of the 13 labtests, computed over the whole admission period [Luo, 2022]

The top-3 teams [Luo, 2022]

1st: Team 'Ping An' [Xu et al., 2020] (Asia, Insurance)

Trained one model per labtest using 5 feature spaces: (F1) values of the other labtests at the same timepoint; (F2) current timepoint and time index; (F3) duration between current timepoint and the previous and the next one; (F4) values of the 13 labtests in the previous 3 and in the next 3 timepoints; (F5) max, min and mean labtest value for each of the 13 labtests, computed over the whole admission period [Luo, 2022]

2nd: Team 'AstraZeneca' [Sun, 2019] (Europe, Pharmaceutical)

Trained one RF for each labtest that had high skewness and one LASSO model for each of the other labtests. They used F1, F4 and the slope of change trend at each timepoint. They also exploited the similarity among patients with respect to the admission features and wrt summary statistics of the labtests [Luo, 2022]

The top-3 teams [Luo, 2022]

1st: Team 'Ping An' [Xu et al., 2020] (Asia, Insurance)

Trained one model per labtest using 5 feature spaces: (F1) values of the other labtests at the same timepoint; (F2) current timepoint and time index; (F3) duration between current timepoint and the previous and the next one; (F4) values of the 13 labtests in the previous 3 and in the next 3 timepoints; (F5) max, min and mean labtest value for each of the 13 labtests, computed over the whole admission period [Luo, 2022]

2nd: Team 'AstraZeneca' [Sun, 2019] (Europe, Pharmaceutical)

Trained one RF for each labtest that had high skewness and one LASSO model for each of the other labtests. They used F1, F4 and the slope of change trend at each timepoint. They also exploited the similarity among patients with respect to the admission features and wrt summary statistics of the labtests [Luo, 2022]

3rd: Team 'Vanderbilt' [Zhang et al., 2020] (North America, University)

Trained one XGBoost model per labtest, after (1) unsupervised 'prefilling', for which they used the global mean of the labtest for all patients and the local mean of the neighbouring timestamps of the specific patient, and then applied two matrix completion techniques; (2) extraction of features from within a sliding window around each missing data point [Zhang et al., 2020]

Limitations of the study [Luo, 2022]

RECALL: For each admission AND for each labtest, one recorded result was selected at random and removed

Limitations of the study [Luo, 2022]

RECALL: For each admission AND for each labtest, one recorded result was selected at random and removed

BUT

! Missingness rate varies with clinical settings
For example, we expect higher missingness among hospital patients in general than ICU patients

Limitations of the study [Luo, 2022]

RECALL: For each admission AND for each labtest, one recorded result was selected at random and removed

BUT

! Missingness rate varies with clinical settings
For example, we expect higher missingness among hospital patients in general than ICU patients

! Artificially missing data vs native missing data

Limitations of the study [Luo, 2022]

RECALL: For each admission AND for each labtest, one recorded result was selected at random and removed

BUT

! Missingness rate varies with clinical settings
For example, we expect higher missingness among hospital patients in general than ICU patients

! Artificially missing data vs native missing data

! Only patients were considered that had at least 10 time points, in which no more than 50% of the labtest results were missing: this is too restrictive because

- some labtests are more likely to be scheduled than others
- some labtests may be scheduled more times per day than others
- some labtests may be skipped as not necessary for the given condition [MNAR]

More on imputation of missing values

Some more examples of algorithms that impute missing values:

- [Yue et al., 2022] propose ts2vec and show that it copes well with missingness
- [Lipton et al., 2016] use RNNs for imputation of missing EHR data
- [Mikalsen et al., 2021] fill the 'labels' in a supervised or semi-supervised way
- [Tipirneni and Reddy, 2022] address misalignment with a semi-supervised method
- [Tawakuli et al., 2023] point out that next to punctual missingness, there can be also missing sequences,
  and they propose a workflow for dealing with them

Filling-in the missing values

Takeaway message (REPEAT): Misalignments and missing values in variables of multivariate time series can be rectified with help of other variables.

Filling-in the missing values

Takeaway message (REPEAT): Misalignments and missing values in variables of multivariate time series can be rectified with help of other variables.

What to exploit:
similarity among individuals in the static space, *temporal proximity* of events for the same individual,
summary statistics and *the predictive power of other time series*.

Filling-in the missing values

Takeaway message (REPEAT): Misalignments and missing values in variables of multivariate time series can be rectified with help of other variables.

What to exploit:
similarity among individuals in the static space, *temporal proximity* of events for the same individual,
summary statistics and *the predictive power of other time series*.

BUT: should we impute on a time series at all ?

Forecasting for Missingness ?

BEWARE:

! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.

! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution.

Forecasting for Missingness ?

BEWARE:

- ! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.
- ! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

Forecasting for Missingness ?

BEWARE:

- ! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.
- ! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

When to impute ?

Forecasting for Missingness ?

BEWARE:

! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.

! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

When to impute ?

- to win a competition

Forecasting for Missingness ?

BEWARE:

! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.

! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

When to impute ?

- to win a competition
- when there is NO intervention

Forecasting for Missingness ?

BEWARE:

- ! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.
- ! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

When to impute ?

- to win a competition
- when there is NO intervention
- **BEFORE** the intervention

Forecasting for Missingness ?

BEWARE:

! Forecasting algorithms make assumptions about the underlying data distribution and correlations among variables.

! Data may be missing because some of these assumptions are not satisfied, so it is not appropriate to use a forecaster to fill the data.

The typical assumption is that the data come from the same distribution. If there was an intervention, then the data NO MORE come from the same distribution.

When to impute ?

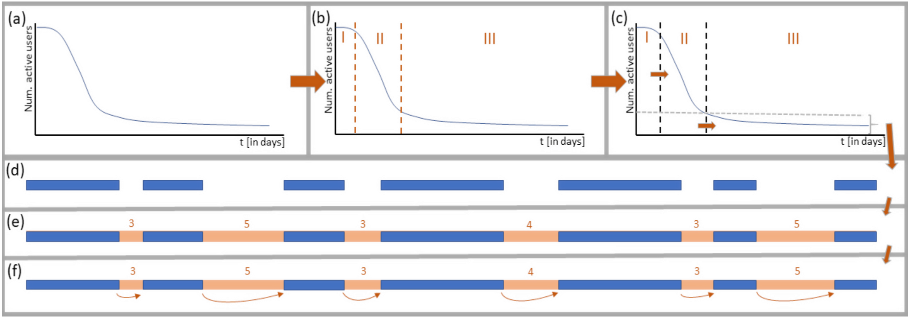
- to win a competition
- when there is NO intervention
- **BEFORE** the intervention
- **AFTER** the intervention
  AND controlling for the intervention [demands understanding]

Predicting the gaps

[Schleicher et al., 2022b, Schleicher et al., 2022a, Schleicher et al., 2022c, Schleicher et al., 2023]

Predicting the gaps

Big Picture


(a)
Num. active users
t [in days]
(b)
Num. active users
I
II
III
t [in days]
(c)
Num. active users
I
II
III
t [in days]
(d)
(e)
3
5
3
4
3
5
(f)
3
5
3
4
3
5

Tuning the gap size for prediction 1/3

- blue lines = sequences
- white spots = periods with no data
- black arrow = time axis

Tuning the gap size for prediction 2/3

- blue lines = sequences
- orange lines = gaps
- black arrow = time axis

group gaps meaningful - **binning**

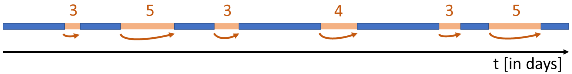
bins represent categories

Binning strategies:

1) building *equisized intervals*,
2) building *intervals on frequency*
3) identifying *'natural' groups* with the Fischer-Jenks algorithm [8]

[8] Jenks, G.F.: The data model concept in statistical mapping. International yearbook of cartography 7, 186–190 (1967)

Tuning the gap size for prediction 3/3



- blue lines = sequences
- orange lines = gaps
- black arrow = time axis
- orange arrows = predictions

Prediction:

- *1-NN classifier* with Dynamic Time Warping
- stratified 10-fold cross-validation

Ignoring the gaps

[Unnikrishnan et al., 2019, Unnikrishnan et al., 2020, Unnikrishnan et al., 2021, Unnikrishnan et al., 2023]

Neighbourhood-augmented prediction [Unnikrishnan et al., 2021] and [9]

Core idea: Learn a predictor on the data of the entity itself and augment this predictor with information from the entity's neighbours

[9] [Unnikrishnan et al., 2019, Unnikrishnan et al., 2020], [Unnikrishnan et al., 2023]

Neighbourhood-augmented prediction [Unnikrishnan et al., 2021] and [9]

Core idea: Learn a predictor on the data of the entity itself and augment this predictor with information from the entity's neighbours

Gap treatment: Treated during the vectorisation

Sparsity treatment: Separate between 'short' and 'long' entities; exploit the long entities to learn on the short ones

[9] [Unnikrishnan et al., 2019, Unnikrishnan et al., 2020], [Unnikrishnan et al., 2023]

Neighbourhood-augmented prediction [Unnikrishnan et al., 2021] and [9]

Core idea: Learn a predictor on the data of the entity itself and augment this predictor with information from the entity's neighbours

Gap treatment: Treated during the vectorisation

Sparsity treatment: Separate between 'short' and 'long' entities; exploit the long entities to learn on the short ones

$\Downarrow$

Neighbourhoods for prediction

RQ 1: How to define similarity between entities on their time series?
RQ 2: How to exploit similarity for prediction?
RQ 3: How to identify and exploit only those neighbours of an entity that contribute to better prediction?

[9] [Unnikrishnan et al., 2019, Unnikrishnan et al., 2020], [Unnikrishnan et al., 2023]

Neighbourhood-augmented prediction – New results

Three patient empowerment studies

- ⋆ Bulgarian study (diabetes melitus): $n = 10$ users contributing 387 EMA
- ⋆ Spanish study (diabetes type II): $n = 12$ users contributing 650 EMA

and German study on tinnitus, description in [Unnikrishnan et al., 2019]

Dataset	#Short entities	RMSE		
		Early termination	Exhaustive search	Global model
Bulgaria	5 of 10	25.36	23.18	23.16
Spain	8 of 12	14.71	13.40	30.55
Germany	260	13.9	19.6	23.87

↓

The benefits of early termination (when expanding the neighbourhood) increase with the proportion of short entities.

Few users can predict many [Unnikrishnan et al., 2023]

Some users have many neighbours, while others are nobody's neighbour.

When focusing on the beneficial users:

- 83-90% of users are better predicted through their neighbourhood,
- better performance by 13%-15%
- on two datasets (an observational study dataset and the UNITI rct dataset)

. . . on neighbourhood-augmented prediction

Takeaway message: Neighbourhood-based prediction of an entity's future demands less resources than a global predictor and achieves comparable performance. But the selection of the similar users must be done intelligently; not all similar users are informative towards an entity's predictor.
Next: How to find the rules that govern an entity's evolution?

Ask the expert's help to fill the gaps

[Büttner et al., 2022, Beyer et al., 2020]

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF

FACULTY OF COMPUTER SCIENCE

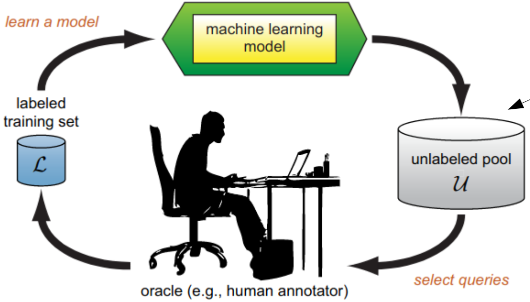
DSAA 2022 - International Conference on Data Science and Advanced Analytics

Reducing Missingness in a Stream through Cost-Aware Active Feature Acquisition

Christian Beyer, Maik Büttner and Prof. Myra Spiliopoulou
AG Knowledge Management & Discovery

christian.beyer@ovgu.de

Brief Introduction to Active Learning



Exploit model in AL strategy to pick new instances for labeling

[Picture from 'Active Learning Literature Survey' by Burr Settles, 2009]

Brief Introduction to Active Learning on Streams

- Budget is defined in relative terms e.g. buy labels for 20% of the instances
- On stream we use an Incremental Percentile Filter to track past metric values in an ordered list of fixed length (oldest value gets deleted) [1]
- If the metric of a newly arriving instance is in top percentile given the budget threshold, then purchase the label. Otherwise ignore the instance and just track the metric value in the IPF

Introduction to Active Feature Acquisition

- Modelled AFA as AL but instead of purchasing a label we can purchase a feature [2]

- Calculated *merit* of a feature as a metric [3]
  e.g. Average Euclidean Distance (AED)

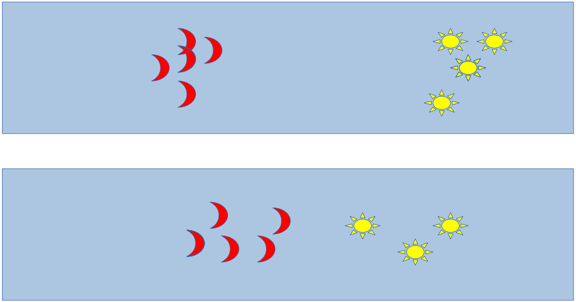
- Consequences:
  → Can only purchase one of the missing features per instance
  → All features assumed to cost the same

Introduction to Active Feature Acquisition

$$AED_{num}(F_i) = \sqrt{\sum_{0 \leq c < k < L} (MV(F_{ic}) - MV(F_{ik}))^2} \qquad (2)$$

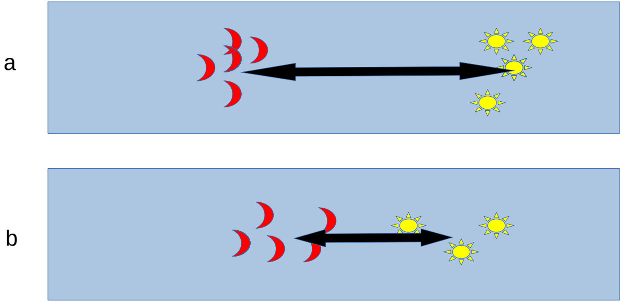
a

b



Introduction to Active Feature Acquisition

$$AED_{num}(F_i) = \sqrt{\sum_{0 \leq c < k < L} \left(MV(F_{ic}) - MV(F_{ik})\right)^2} \tag{2}$$



Separates classes better → higher merit

Active Feature Acquisition For Multiple Features with Varying Costs

- We introduced feature costs in the *merit* function
- We went from a relative budget (buy x %) to an absolute budget (e.g. spend at most 1000$ per month)
  → Threshold for IPF is estimated by calculating the maximum/mean cost per instance and dividing the available budget by that number
- We introduced a penalty function which further adjusts the IPFs threshold to improve budget usage
- We introduced AFA-strategies that can select and purchase sets of features instead of a single feature

New Strategies:

- Purchase the *k* best missing features of an instance
  → up to *k* missing features with highest merit
- Purchase the *k* global best features
  → up to *k* missing features if the merit of the features is in the top *k* across the current window
- MaxMean: keep adding the best remaining missing feature while its inclusion improves the instance quality (often buys only 1 or 2 features)

Experiments

- We used two different budgeting strategies
  → Simply buy when you have budget (SBM)
  → Buy when expected quality is in top percent of last seen values (IPF)

- Compared against a random baseline

- Experiments with different degrees of missingness, different merit functions, different budgets and different feature costs (same, ascending or descending)

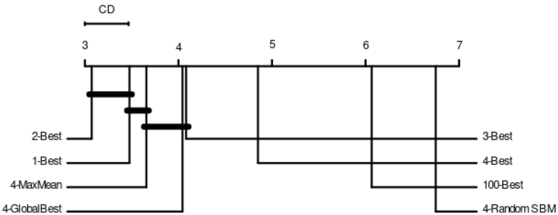
- 7 real and 8 synthetic datasets

Evaluation

- Classifier performance was evaluated using Kappa+ metric and Friedmann/Nemenyi rank analysis
- Lower bound is a classifier that has all features missing and imputes the values with the feature mean
- Upper bound is a classifier that has all features available
- Budget was evaluated with regards to the number of experiments where a strategy led to overspending and the degree of overspending

Results

Friedmann/Nemenyi rank analysis on the datasets excluding the even-odd-datasets



Results

Overview of the percentage of runs where a strategy used more budget than allowed and percentage of overused budget compared to allowed budget on the 8 regular data sets

Strategy	% Runs	Avg. % over Budget
1-Best	0.2	0.01
2-Best	0.4	0.02
3-Best	1.79	0.03
4-Best	3.17	0.06
100-Best	32.74	12.95
4-GlobalBest	23.41	0.22
4-MaxMean	7.34	0.16
4-Random SBM	0	0

Closing

We have seen key concepts:

- $\surd$ Key terms: time series, outcome, phenotype, cohort
- $\surd$ Formalisation of the prediction task on time series
- $\surd$ Missingness

We have discussed general methods:

- $\surd$ Workflow for prediction
- $\surd$ Workflows for cohort construction
- $\surd$ Alternatives for the alignment of a time series at zero
- $\surd$ Dealing with missingness through gridding and templating

Closing

We have seen key concepts:

- $\surd$ Key terms: time series, outcome, phenotype, cohort
- $\surd$ Formalisation of the prediction task on time series
- $\surd$ Missingness

We have discussed general methods:

- $\surd$ Workflow for prediction
- $\surd$ Workflows for cohort construction
- $\surd$ Alternatives for the alignment of a time series at zero
- $\surd$ Dealing with missingness through gridding and templating

For the case of

- only few time series
- consisting mainly of empty space

. . . we have seen specialized methods

- $\surd$ Neighbourhood-based augmentation
- $\surd$ Predictors in a latent space of patterns
- $\surd$ Ways of predicting the length of a gap

We have skipped:

- elaborate forecasters

We have skipped:

- elaborate forecasters $\longrightarrow$ see Part III !

We have skipped:

- elaborate forecasters $\longrightarrow$ see Part III !
- time series on categorical data
- time series on human/animal experiments – a category by themselves

We have skipped:

- elaborate forecasters $\longrightarrow$ see Part III !
- time series on categorical data
- time series on human/animal experiments – a category by themselves

How to use these materials?

→ Use the methods and insights of sections 1 and 2 to build cohorts

→ You now have more options of dealing with missingness:
  - align or classify the missingness
  - predict the gaps

→ For short time series: better use its neighbours rather than a global model, but choose the neighbours carefully

The KMD team

Prof. Myra Spiliopoulou

Christian Beyer

Noor Jamaludeen

Miro Schleicher

Saijal Shahania

Clara Puga

Hafez Kader Omar

Anne Rother

Vishnu Unnikrishnan

Maik Büttner

THANK YOU !

Bibliography I

Beyer, C., Büttner, M., Unnikrishnan, V., Schleicher, M., Ntoutsi, E., and Spiliopoulou, M. (2020). Active feature acquisition on data streams under feature drift. *Annals of Telecommunications*.

Büttner, M., Beyer, C., and Spiliopoulou, M. (2022). Reducing missingness in a stream through cost-aware active feature acquisition. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*.

Chang, S., Wilson, M. L., Lewis, B., Mehrab, Z., Dudakiya, K. K., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., et al. (2021). Supporting COVID-19 policy response with large-scale mobility-based modeling. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2632–2642.

Bibliography II

- Cismondi, F. C., Fialho, A., Vieira, S., Reti, S., Sousa, J., and Finkelstein, S. (2013). Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine*. http://dx.doi.org/10.1016/j.artmed.2013.01.003.
- Fox, I., Ang, L., Jaiswal, M., Pop-Busui, R., and Wiens, J. (2018). Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1387–1395.
- Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., and Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):1–9.

Bibliography III

Lin, Q., Zhang, D., Huang, X., Ni, H., and Zhou, X. (2012). Detecting wandering behavior based on gps traces for elders with dementia. In *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 672–677. IEEE.

Lipton, Z. C., Kale, D. C., Wetzel, R., et al. (2016). Modeling missing data in clinical time series with rnns. *Machine Learning for Healthcare*, 56:253–270.

Luo, Y. (2022). Evaluating the state of the art in missing data imputation for clinical data. *Briefings in Bioinformatics*, 23(1):bbab489.

Luo, Y., Szolovits, P., Dighe, A. S., and Baron, J. M. (2018). 3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data. *Journal of the American Medical Informatics Association*, 25(6):645–653.

Bibliography IV

- Mikalsen, K. Ø., Soguero-Ruiz, C., Bianchi, F. M., Revhaug, A., and Jenssen, R. (2021). Time series cluster kernels to exploit informative missingness and incomplete label information. *Pattern Recognition*, 115:107896.
- Niemann, U., Boecking, B., Brueggemann, P., Mebus, W., Mazurek, B., and Spiliopoulou, M. (2020a). Tinnitus-related distress after multimodal treatment can be characterized using a key subset of baseline variables. *PLOS ONE*, 15(1):1–18.
- Niemann, U., Brueggemann, P., Boecking, B., Rose, M., Spiliopoulou, M., and Mazurek, B. (2020b). Phenotyping chronic tinnitus patients using self-report questionnaire data: cluster analysis and visual comparison. *Scientific Reports*, 10(1):16411.

Bibliography V

- Niemann, U., Spiliopoulou, M., Malanowski, J., Kellersmann, J., Szczepanski, T., Klose, S., Dedonaki, E., Walter, I., Ming, A., and Mertens, P. R. (2020c). Plantar temperatures in stance position: A comparative study with healthy volunteers and diabetes patients diagnosed with sensoric neuropathy. *EBioMedicine*, 54:102712.
- Parmezan, A. R. S. and Batista, G. E. A. P. A. (2014). ICMC-USP time series prediction repository. Technical report, Instituto de Ciençias Matem'aticas e de Computação, Universidade de São Paulo, São Carlos, Brasil.
- Parmezan, A. R. S., Souza, V. M. A., and Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences*.

Bibliography VI

Probst, T., Pryss, R. C., Langguth, B., Rauschecker, J. P., Schobel, J., Reichert, M., Spiliopoulou, M., Schlee, W., and Zimmermann, J. (2017). Does tinnitus depend on time-of-day? an ecological momentary assessment study with the "trackyourtinnitus" application. *Frontiers in Aging Neuroscience*, 9:253.

Schleicher, M., Hamacher, S., Naujoks, M., Günther, K., Schmidt, T., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2022a). Prediction of declining engagement to self-monitoring apps on the example of tinnitus mhealth data. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 228–233. IEEE.

Schleicher, M., Pryss, R., Schlee, W., and Spiliopoulou, M. (2022b). When can i expect the mhealth user to return? prediction meets time series with gaps. In *20th Int. Conf. on Artificial Intelligence in Medicine (AIME)*. Springer. Best paper award.

Bibliography VII

- Schleicher, M., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2022c). Expect the gap: A recommender approach to estimate the absenteeism of self-monitoring mhealth app users. In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*.
- Schleicher, M., Unnikrishnan, V., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2023). Prediction meets time series with gaps: User clusters with specific usage behavior patterns. *Artificial Intelligence in Medicine*, 142:102575.
- Sun, P. (2019). MICE-DA: a MICE method with data augmentation for missing data imputation in IEEE ICHI 2019 DACMI challenge. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–3. IEEE.
- Tawakuli, A., Kaiser, D., and Engel, T. (2023). Experience: Differentiating between isolated and sequence missing data. *ACM Journal of Data and Information Quality*.

Bibliography VIII

- Tipirneni, S. and Reddy, C. K. (2022). Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(6):1–17.
- Unnikrishnan, V. (2017). Analysis of patient evolution on time series of different lengths. Faculty of Computer Science, Otto-von-Guericke Univ. Magdeburg. Master Thesis.
- Unnikrishnan, V., Beyer, C., Matuszyk, P., Niemann, U., Pryss, R., Schlee, W., Ntoutsi, E., and Spiliopoulou, M. (2019). Entity-level stream classification: exploiting entity similarity to label the future observations referring to an entity. *International Journal of Data Science and Analytics*, pages 1–15.

Bibliography IX

Unnikrishnan, V., Schleicher, M., Puga, C., Pryss, R. P., Vogel, C., Schlee, W., and Spiliopoulou, M. (2023). A similarity-guided framework for error-driven discovery of patient neighbourhoods in ema data. In *Proceedings of the Symposium on Intelligent Data Analysis (IDA 2023)*, Louvain-la-Neuve, Belgium.

Unnikrishnan, V., Shah, Y., Schleicher, M., Fernández-Viadero, C., Strandzheva, M., Velikova, D., Dimitrov, P., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2021). Love thy neighbours: A framework for error-driven discovery of useful neighbourhoods for one-step forecasts on EMA data. In *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 295–300. IEEE.

Bibliography X

- Unnikrishnan, V., Shah, Y., Schleicher, M., Strandzheva, M., Dimitrov, P., Velikova, D., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2020). Predicting the health condition of mhealth app users with large differences in the number of recorded observations - where to learn from? In Appice, A., Tsoumakas, G., Manolopoulos, Y., and Matwin, S., editors, *DS*, volume 12323 of *Lecture Notes in Computer Science*, pages 659–673. Springer.
- Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45:1–67.
- Van Buuren, S. and Oudshoorn, C. G. (2000). Multivariate imputation by chained equations. Technical Report ISBN 90-6743-677-1, Netherlands Organization for Applied Scientific Research (TNO), Leiden, NL.

Bibliography XI

Xu, X., Liu, X., Kang, Y., Xu, X., Wang, J., Sun, Y., Chen, Q., Jia, X., Ma, X., Meng, X., et al. (2020). A multi-directional approach for missing value estimation in multivariate time series clinical data. *Journal of Healthcare Informatics Research*, 4(4):365–382.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., and Xu, B. (2022). Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8980–8987.

Zhang, X., Yan, C., Gao, C., Malin, B. A., and Chen, Y. (2020). Predicting missing values in medical data via XGBoost regression. *Journal of Healthcare Informatics Research*, 4:383–394.