Part II: Learning from EHR data

Panagiotis Papapetrou, PhD
Professor, Stockholm University
Adjunct Professor, Aalto University
Who are we?

DSV @ Stockholm University

- **DSV**: Data- och SystemVetenskap (Computer and Systems Sciences)
- **# of students**: approx. 5400
- **# of staff members**: 176 (60 profs. /associate profs. / lecturers)
Our AI research arena @ DSV

RESEARCH AREAS @DSV

AI ACROSS FIELDS

Methods
- Knowledge representation
- Decision theory and reasoning
- Machine learning
- Natural language processing
- Simulation
- Human-computer interaction
- Agent systems
- Distributed systems

Applications
- Healthcare
- Automotive
- Governance
- Education
- Security
- Games
- Art
- Telecommunications
## Research at DSV – data science group

### Main research areas:
- Sequential and temporal data mining
- Interpretability and explainability of machine learning methods
- Random forests and ensemble learning
- Machine learning for healthcare applications
- Clinical text mining and natural language processing

### Current projects:
- **EXTREMUM (2020-2024):** explainable and ethical ML for healthcare
- **Covid-Sim (2020-2021):** reinforcement learning for simulation of pandemics
- **TEMPOMiner (2017-2020):** temporal data mining for detecting ADEs in healthcare
Part II - Outline

- Temporal abstractions for EHR data
- **Attention-based** deep learning for healthcare event prediction
- **Actionable models** and **counterfactual explanations** for EHR data
- **Interpretable ranking and classification** of radiography exams
Electronic Health Records: content

- Diagnoses
- Drug prescriptions
- Clinical tests
- More complex structures
  - clinical notes
  - medical images
  - MRIs
  - ECGs
  - …

Longitudinal collection of electronic health information about individual patients and populations
Reporting of suspicious Adverse Drug Events (ADEs) and suggestions for treatment
ICD10* codes

• 10th revision of the International Classification of Diseases and Related Health Problems
• a classification system that is used to record medical activity
• the system enables classification and quantification of diseases and other health-related issues

3 -7 characters long

* http://www.ahima.org/icd10
ICD10 codes: examples (total of 22 chapters)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Code Range</th>
<th>Estimated # of Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A00-B99</td>
<td>1,056</td>
<td>Certain infectious and parasitic diseases</td>
</tr>
<tr>
<td>2</td>
<td>C00-D49</td>
<td>1,620</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>3</td>
<td>D50-D89</td>
<td>238</td>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</td>
</tr>
<tr>
<td>4</td>
<td>E00-E89</td>
<td>675</td>
<td>Endocrine, nutritional and metabolic diseases</td>
</tr>
<tr>
<td>5</td>
<td>F01-F99</td>
<td>724</td>
<td>Mental, Behavioral and Neurodevelopmental disorders</td>
</tr>
<tr>
<td>6</td>
<td>G00-G99</td>
<td>591</td>
<td>Diseases of the nervous system</td>
</tr>
<tr>
<td>7</td>
<td>H00-H59</td>
<td>2,452</td>
<td>Diseases of the eye and adnexa</td>
</tr>
<tr>
<td>8</td>
<td>H60-H95</td>
<td>642</td>
<td>Diseases of the ear and mastoid process</td>
</tr>
<tr>
<td>9</td>
<td>I00-I99</td>
<td>1,254</td>
<td>Diseases of the circulatory system</td>
</tr>
<tr>
<td>10</td>
<td>J00-J99</td>
<td>336</td>
<td>Diseases of the respiratory system</td>
</tr>
<tr>
<td>11</td>
<td>K00-K95</td>
<td>706</td>
<td>Diseases of the digestive system</td>
</tr>
</tbody>
</table>
# ICD10 codes: examples

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Combination Codes</strong></td>
</tr>
<tr>
<td>I25.110</td>
<td>Atherosclerotic heart disease of native coronary artery with unstable angina pectoris</td>
</tr>
<tr>
<td></td>
<td><strong>Increased Specificity</strong></td>
</tr>
<tr>
<td>S72.044G</td>
<td>Non-displaced fracture of base of neck of right femur, subsequent encounter for closed fracture with delayed healing</td>
</tr>
<tr>
<td></td>
<td><strong>Laterality</strong></td>
</tr>
<tr>
<td>C50.511</td>
<td>Malignant neoplasm of lower-outer quadrant of right female breast</td>
</tr>
<tr>
<td>C50.512</td>
<td>Malignant neoplasm of lower-outer quadrant of left female breast</td>
</tr>
<tr>
<td></td>
<td><strong>“X” Placeholder</strong></td>
</tr>
<tr>
<td>H40.11X2</td>
<td>Primary open-angle glaucoma, moderate stage</td>
</tr>
</tbody>
</table>
ATC* codes

• Anatomical Therapeutic Chemical codes, first published in 1976
• Used for classification of active ingredients of drugs
• Based on the organ/system on which they act
  o therapeutic
  o pharmacological and chemical properties
• Controlled by the World Health Organization Collaborating Centre (WHOCC) for drug statistics methodology

* http://www.whocc.no/atc_ddd_index/
ATC codes classify drugs into 5 different levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Content</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>anatomical main group</td>
<td>1 letter</td>
<td>A: alimentary tract and metabolism</td>
</tr>
<tr>
<td>II</td>
<td>therapeutic subgroup</td>
<td>2 digits</td>
<td>A10: diabetes drugs</td>
</tr>
<tr>
<td>III</td>
<td>pharmacological subgroup</td>
<td>1 letter</td>
<td>A10B: blood glucose lowering drugs, excl. insulins</td>
</tr>
<tr>
<td>IV</td>
<td>chemical subgroup</td>
<td>1 letter</td>
<td>A10BA: biguanides</td>
</tr>
<tr>
<td>V</td>
<td>chemical substance</td>
<td>2 digits</td>
<td>A10BA02: metformin</td>
</tr>
</tbody>
</table>
Extracting features from EHRs

- **Mainly two lines of approaches:**
  - static features
  - temporal features
Static features

- **Mainly two lines of approaches:**
  - static features
  - temporal features

- **Existing out-of-the-box classifiers are used**
  - Decision trees, random forests, SVMs, deep learning architectures
    - [Chazard2011, Zhao2013, Karlsson2013, Shickel2018]

The class labels assigned depending on task at hand, e.g., ADE detection
Temporal features

- **Mainly two lines of approaches:**
  - static features
  - temporal features

Clinical measurements:
- different units
- times of measurement
- sparsity
Temporal abstractions of EHRs

- Database: **EHRs of Patients**

- **Each EHR:**
  - Multiple temporal variables registered and evolving concurrently
  - Each variable with multiple readings until a critical time point $t_i$, e.g., glucose, creatinine, cholesterol
  - **Class label:** Disease/symptom detected at time $t_i$ (event of interest)
Temporal abstractions of EHRs

- Database: **EHRs of Patients**

- **Each EHR:**
  - Multiple temporal variables registered and evolving concurrently
  - Each variable with multiple readings until a critical time point $t_i$, e.g., glucose, creatinine, cholesterol
  - **Class label:** Disease/symptom detected at time $t_i$ (event of interest)
Two types of temporal abstractions

- **Trend abstraction:**
  - e.g., *decreasing, steady, increasing*
  - time series segmentation + identify slopes

- **Value abstraction**
  - e.g., *very low, low, normal, high, very high*
  - use 10\textsuperscript{th}, 25\textsuperscript{th}, 75\textsuperscript{th}, and 90\textsuperscript{th} percentiles on the lab values to define [Batal2012]
Temporal abstractions of EHRs

- **Supervised temporal prediction** [Batal2012, Rebane2019]
  - Given a labeled dataset of temporal instances up to time $t_i$
  - Find frequently occurring “temporal patterns” for each label
  - Given a sample instance $\xrightarrow{\text{predict its label}}$

![Diagram showing temporal patterns for different events over time intervals](image)
Temporal abstraction patterns

• **What is a temporal pattern?**
  - a sequence of “temporal relations” between state intervals A and B

• **What kind of “temporal relations”?**

  Based on **Allen’s temporal logic**

<table>
<thead>
<tr>
<th>Relation</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A meets B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A matches B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A overlaps-with B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A followed-by B</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B --- -----------</td>
</tr>
<tr>
<td>A contains B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A left-contains B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>A right-contains B</td>
<td>A --- -----------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>
Mining temporal abstraction patterns

<table>
<thead>
<tr>
<th>Relation</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A meets B</td>
<td></td>
</tr>
<tr>
<td>A matches B</td>
<td></td>
</tr>
<tr>
<td>A overlaps-with B</td>
<td></td>
</tr>
<tr>
<td>A followed-by B</td>
<td></td>
</tr>
</tbody>
</table>

Lee et al., **Z-Miner**: an efficient method for mining frequent arrangements of event intervals, KDD 2020

Lee et al., **Z-Embedding**: A spectral representation of event intervals for efficient clustering and classification, ECML/PKDD 2020
Temporal abstractions of sparse EHRs

Hielscher et al. *Mining Longitudinal Epidemiological Data to Understand a Reversible Disorder*, Intelligent Data Analysis, 2014

Zhao et al. *Learning from Heterogeneous Temporal Data in Electronic Health Records*, Journal of Biomedical Informatics, 2017

Bagattini et al. A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records, BMC Medical Informatics and Decision Making, 2019

Rebane et al. *SMILE: a feature-based temporal abstraction framework for event-interval sequence classification*, Data Mining and Knowledge Discovery, accepted [pre-print online]
Framework overview

Multi-variate feature representation

\[ \mathcal{L} \]

Normalization

Summarization

Symbolic representation

Alphabet calibration

\[ \hat{\mathcal{L}} \]

Subsequence generation

Subsequence evaluation

Subsequence selection

\[ \mathcal{Z}^* = s^*(\hat{A}_1), \ldots, s^*(\hat{A}_m) \]

Data transformation

\[ D = \tau^*(\mathcal{O}) \]
Phase A: normalization

- Z-normalization
  - Each multi-variate feature $S$ is z-normalized:

  $$ S := \frac{\sum_{i=1}^{(S_i - \mu(S))}}{\sigma(S)} $$
Phase A: summarization

• **Z-normalization**
  - Each multi-variate feature \( S \) is z-normalized:

\[
S := \frac{\sum_{i=1}^{\vert S \vert} \{ s_i - \mu(S) \}}{\sigma(S)}
\]

• **Summarization**
  - Piecewise Aggregate Approximation (PAA)
  - Dimensionality reduction from \( d \) to \( w \)

\[
\overline{S} = \{ \overline{s_1}, \ldots, \overline{s_w} \}
\]
Phase A: symbolic representation

- **SAX mapping**
  - each record is mapped to a string using SAX
  - length: number of measurements
  - alphabet: 2 – 5, or set using domain knowledge

Discretize into a vector of symbols
- breakpoints map to small alphabet $a$ of symbols

![Graph showing SAX mapping and discretization](image)
Phase B: mapping to real features

multi-variate feature
bacccc

cbcbc

abbbba

bbbc

baaacbb

bcccabc

reference feature
bbaab

distance function
eal-valued feature
3.45

1.23

5.56

...
Phase B: subsequence enumeration

- **s-shapelet generation:**
  - random subsequences $s$ of length $t \in [1, l_{\text{max}}]$

- **s-shapelet evaluation:**
  - each $s$ is converted to a real value based on its distance to each multi-variate feature sequence

$$\text{Dist} (s, \hat{S}) := \min_{s' \subseteq \hat{S}, |s'| = |s|} \{ D (s, s') \}$$

$l_{\text{max}}$: max length of a feature sequence
Phase B: subsequence selection

- For each multi-variate feature:
  - select the s-shapelet $s^*$ with the max utility:
    $$s^* := \arg \max_{s \in \mathcal{L}} \text{Gain} \left( s, \delta_{osp}(s), \hat{L} \right)$$
  - select the alphabet size with the max utility
    $$\alpha^* := \arg \max_{\alpha \in \mathcal{T}} \text{Gain} \left( s^*_\alpha, \delta_{osp}(s^*_\alpha), \hat{L} \right)$$
- Final set of s-shapelets:
  $$\mathcal{Z}^* = \left\{ s^* \left( \hat{A}_1 \right), \ldots, s^* \left( \hat{A}_m \right) \right\}$$
Phase C: transformation

- A function $\tau^*$ is learned:
  - transform any data object of the original multi-variate space to a set of real-valued features

\[
\tau^* : \mathcal{A} \rightarrow \mathbb{R}^m
\]

- Each data example is transformed using $\tau^*$:

\[
\tilde{O} = \tau^*(O)
\]
Many ADEs are not being identified as such, due to limited knowledge about the effects of medical treatments, e.g., drugs being tested only in limited clinical trials under controlled conditions.

Alternative: resort to machine learning methods and explore different feature abstractions: static or temporal

Learning classification models: extremely useful for patient monitoring, outcome prediction, and decision support
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Class label description</th>
<th>Pos.</th>
<th>Neg.</th>
<th>Feat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D611</td>
<td>Drug-induced aplastic anaemia</td>
<td>593</td>
<td>105</td>
<td>285</td>
</tr>
<tr>
<td>D642</td>
<td>Drug-induced secondary sideroblastic anaemia</td>
<td>217</td>
<td>9673</td>
<td>513</td>
</tr>
<tr>
<td>D695</td>
<td>Secondary thrombocytopenia</td>
<td>1246</td>
<td>2148</td>
<td>450</td>
</tr>
<tr>
<td>E273</td>
<td>Drug-induced adrenocortical insufficiency</td>
<td>70</td>
<td>259</td>
<td>229</td>
</tr>
<tr>
<td>G620</td>
<td>Drug-induced polyneuropathy</td>
<td>96</td>
<td>783</td>
<td>258</td>
</tr>
<tr>
<td>I952</td>
<td>Drug-induced hypotension</td>
<td>115</td>
<td>1287</td>
<td>324</td>
</tr>
<tr>
<td>L270</td>
<td>Drug-induced generalized skin eruption</td>
<td>182</td>
<td>468</td>
<td>314</td>
</tr>
<tr>
<td>L271</td>
<td>Drug-induced localized skin eruption</td>
<td>151</td>
<td>498</td>
<td>311</td>
</tr>
<tr>
<td>M804</td>
<td>Drug-induced osteoporosis with pathological fracture</td>
<td>52</td>
<td>1170</td>
<td>282</td>
</tr>
<tr>
<td>M814</td>
<td>Drug-induced osteoporosis</td>
<td>57</td>
<td>5097</td>
<td>434</td>
</tr>
<tr>
<td>O355</td>
<td>Maternal care for (suspected) damage to fetus by drugs</td>
<td>146</td>
<td>260</td>
<td>148</td>
</tr>
<tr>
<td>R502</td>
<td>Drug-induced fever</td>
<td>80</td>
<td>6434</td>
<td>498</td>
</tr>
<tr>
<td>T782</td>
<td>Adverse effects: anaphylactic shock</td>
<td>131</td>
<td>856</td>
<td>293</td>
</tr>
<tr>
<td>T783</td>
<td>Adverse effects: angioneurotic oedema</td>
<td>283</td>
<td>720</td>
<td>293</td>
</tr>
<tr>
<td>T784</td>
<td>Adverse effects: allergy</td>
<td>574</td>
<td>415</td>
<td>294</td>
</tr>
<tr>
<td>T801</td>
<td>Vascular complications following infusion, transfusion and therapeutic injection</td>
<td>66</td>
<td>609</td>
<td>229</td>
</tr>
<tr>
<td>T808</td>
<td>Other complications following infusion, transfusion and therapeutic injection</td>
<td>538</td>
<td>138</td>
<td>229</td>
</tr>
<tr>
<td>T886</td>
<td>Drug-induced anaphylactic shock</td>
<td>89</td>
<td>1506</td>
<td>363</td>
</tr>
<tr>
<td>T887</td>
<td>Unspecified adverse effect of drug or medicament</td>
<td>1047</td>
<td>550</td>
<td>363</td>
</tr>
</tbody>
</table>
### ADEs

- ADEs are **injuries** that occur from the use of a drug, such as overdoses or dose reductions, or drug interactions.
- They account for **3.7% of hospital admissions** around the world.
- ADEs have been estimated to come at a cost of **$3.5 billion/year in the U.S alone**, despite ADEs being preventable.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Class label description</th>
<th>Pos.</th>
<th>Neg.</th>
<th>Feat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D611</td>
<td>Drug-induced aplastic anaemia</td>
<td>593</td>
<td>105</td>
<td>285</td>
</tr>
<tr>
<td>D642</td>
<td>Drug-induced secondary sideroblastic anaemia</td>
<td>217</td>
<td>9673</td>
<td>513</td>
</tr>
<tr>
<td>D695</td>
<td>Secondary thrombocytopenia</td>
<td>1246</td>
<td>2148</td>
<td>450</td>
</tr>
<tr>
<td>E273</td>
<td>Drug-induced adrenocortical insufficiency</td>
<td>70</td>
<td>259</td>
<td>229</td>
</tr>
<tr>
<td>G620</td>
<td>Drug-induced polyneuropathy</td>
<td>96</td>
<td>783</td>
<td>258</td>
</tr>
<tr>
<td>I952</td>
<td>Drug-induced hypotension</td>
<td>115</td>
<td>1287</td>
<td>324</td>
</tr>
<tr>
<td>L270</td>
<td>Drug-induced generalized skin eruption</td>
<td>182</td>
<td>468</td>
<td>314</td>
</tr>
<tr>
<td>T808</td>
<td>Other complications following infusion, transfusion and therapeutic injection</td>
<td>538</td>
<td>138</td>
<td>229</td>
</tr>
<tr>
<td>T886</td>
<td>Drug-induced anaphylactic shock</td>
<td>89</td>
<td>1506</td>
<td>363</td>
</tr>
<tr>
<td>T887</td>
<td>Unspecified adverse effect of drug or medicament</td>
<td>1047</td>
<td>550</td>
<td>363</td>
</tr>
</tbody>
</table>
As the % of feature sparsity increases, AUC also increases!

Shorter s-shapelets (i.e., 2-8) are preferable to longer ones (> 20)
ADE prediction (Yes/No)

- **Main task:**
  predict the presence or absence of an ADE in a patient’s next visit given EHR data entries from all previous visits!

<table>
<thead>
<tr>
<th>Adverse drug event ICD-10 codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>D611 Drug-induced aplastic anaemia</td>
</tr>
<tr>
<td>D642 Drug-induced secondary sideroblastic anaemia</td>
</tr>
<tr>
<td>D695 Secondary thrombocytopenia</td>
</tr>
<tr>
<td>E273 Drug-induced adrenocortical insufficiency</td>
</tr>
<tr>
<td>G620 Drug-induced polyneuropathy</td>
</tr>
<tr>
<td>I952 Drug-induced hypotension</td>
</tr>
<tr>
<td>L270 Drug-induced generalized skin eruption</td>
</tr>
<tr>
<td>L271 Drug-induced localized skin eruption</td>
</tr>
<tr>
<td>M804 Drug-induced osteoporosis with pathological fracture</td>
</tr>
<tr>
<td>M814 Drug-induced osteoporosis</td>
</tr>
<tr>
<td>O355 Maternal care for (suspected) damage to fetus by drugs</td>
</tr>
<tr>
<td>R502 Drug-induced fever</td>
</tr>
<tr>
<td>T782 Adverse effects: anaphylactic shock</td>
</tr>
<tr>
<td>T783 Adverse effects: angioneurotic edema</td>
</tr>
<tr>
<td>T784 Adverse effects: allergy</td>
</tr>
<tr>
<td>T801 Vascular complications following infusion, transfusion and therapeutic injection</td>
</tr>
<tr>
<td>T886 Drug-induced anaphylactic shock</td>
</tr>
</tbody>
</table>
Main goals

- Empirically evaluate which code-level interpretable deep learning architecture provides the best performance for ADE prediction

- Examine which data sources (diagnoses, medications, lab tests) best aid in ADE predictive performance and medical interpretability

- Determine the extent in which code-level attention mechanisms contribute to interpretability for ADE predictions
Methods (Vanilla RNN)

Pass info from **one visit to the next** within network

Predict ADE yes/no in the future

Input: medical codes \( c \) for each medical visit \( v \) to train the network across patients
Limitations of Vanilla RNN

- Standard seq2seq models are normally composed of an encoder-decoder architecture.

- **Encoder**: processes the input sequence and summarizes the information into a context vector of fixed length.

- This representation is expected to be a good summary of the entire input sequence.

- **Decoder**: initialized with the context vector and uses it to generate the transformed output.
Limitations of Vanilla RNN

- **Structural limitation:**
  - fixed-length context vector

- **Why?**
  - inability of remembering longer sequences
  - earlier parts of the sequence are forgotten once the entire sequence is processed

- The attention mechanism concept was born to resolve this problem

- **Attention mechanism:** keep the intermediate encoder states and utilize all of them in order to construct the context vectors required by the decoder to generate the output sequence
Medical ”attention”

• We may want the decoder to focus more on, e.g., visits 1 and 3, while paying less attention to the remaining visits of the patient.

• Solution:
  o Train a feed forward neural network
  o learn to identify relevant encoder states
  o generate a high score for the visits for which attention is to be paid while low score for the visits which are to be ignored
Methods (Timeline, Bai 2018)

\[ \phi(c_x) = \sum_{j=1}^{N} \delta(c_j, \Delta_t) \alpha_{j,x} \]

- Learn how the code contributions should change over time.
- Attention values for code contribution.
Methods (RETAIN, Choi 2016)

Determine code level attention

\[ \omega(y_i, x_{j,k}) = \alpha_j W(\beta_j \odot W_{emb}[:, k]) x_{j,k} \]
Experimental Setup

• RETAIN and Timeline:
  o proven to be competitive state-of-the-art architectures which permit thorough interpretability down to the code-level
  o trained for ADE prediction using an original data source consisting of information for 1,4 million patients obtained from HealthBank at Stockholm University
Experimental Setup

- Non-ADE ICD-10 and ATC codes were reduced to higher level hierarchical categories by selecting only the first three characters.

- Such categories correspond to main categories of ICD-10 codes and to therapeutic subgroups in the case of ATC codes.

- # of ICD-10 categories: **1692**

- # of ATC subgroups: **109**

- Visits defined on a monthly basis.

- Patients also needed at least three such visits to be included.

- Two data sets: including or excluding medication data.
Experimental Setup

- Experiments performed by randomly partitioning the data into training, validation, and test set of ratios 0.7, 0.1, and 0.2, respectively.

- In order to accommodate for the massive class imbalance stemming from the relative rarity of ADEs, a balanced training set was formed in which the majority class was under-sampled.
Results: AUC / F1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Area under ROC</th>
<th>Micro F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RETAIN</td>
<td>Timeline</td>
</tr>
<tr>
<td>Without medication</td>
<td>0.765</td>
<td>0.668</td>
</tr>
<tr>
<td>With medication</td>
<td>0.759</td>
<td>0.693</td>
</tr>
</tbody>
</table>

RETAIN was determined to be the best performing architecture under the conditions of using diagnoses data.
## Interpretability of RETAIN

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Visit 1</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>L50</td>
<td>Urticaria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.891</td>
</tr>
<tr>
<td>R42</td>
<td>Dizziness and giddiness</td>
<td></td>
<td>0.034</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A02†</td>
<td>Drugs for acid related disorders</td>
<td></td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L50</td>
<td>Urticaria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R06†</td>
<td>Antihistamines for systemic use</td>
<td></td>
<td></td>
<td></td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td>H02†</td>
<td>Corticosteroids for systemic use</td>
<td></td>
<td></td>
<td></td>
<td>0.321</td>
<td></td>
</tr>
<tr>
<td>L50</td>
<td>Urticaria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R06†</td>
<td>Antihistamines for systemic use</td>
<td></td>
<td></td>
<td></td>
<td>0.322</td>
<td></td>
</tr>
<tr>
<td>C01†</td>
<td>Cardiac therapy</td>
<td></td>
<td></td>
<td></td>
<td>0.205</td>
<td></td>
</tr>
<tr>
<td>H02†</td>
<td>Corticosteroids for systemic use</td>
<td></td>
<td></td>
<td></td>
<td>0.230</td>
<td></td>
</tr>
</tbody>
</table>

Very high risk from given history
Interpretable and actionable models

- Trade-offs between interpretability + accuracy.
- Ability to understand the outcomes without compromising

The patient will die from HF in 2 days!

I can tell you what changes need to be done to the patient record, so that I can change my prediction 😊

Now what? Please tell me why?
Problem: What is the min # of changes to the abnormal time series to convert it to normal?
Actionable feature tweaking for random forests

Random Forest (opaque model)

example x

- Let x be a true-negative instance
- **Goal:** minimum number of feature tweaks (changes) so that x becomes true-positive, x’

\[
\hat{f}(x) = -1 \iff \left( \sum_{k=1}^{K} \hat{h}_k(x) \right) \leq 0
\]

\[
x' = \arg \min_{x^*} \left\{ \delta(x, x^*) \mid \hat{f}(x) = -1 \land \hat{f}(x^*) = +1 \right\}
\]

Tolomei et al. *Interpretable Predictions of Tree-based Ensembles via Actionable Feature Tweaking*, KDD (and ARXIV), 2017
Solution

- Let $x$ be a true-negative instance
- **Goal:** minimum number of feature tweaks (changes) so that $x$ becomes true-positive, $x'$

**Observe:**
- If the prediction of the RF is -1, then at least half of its trees predict -1
- If the prediction of a tree is -1, then the example is passed through a negative path, i.e., a path that predicts the class to be -1
- **Solution:** revert these paths and consequently the trees!

**Note:** if a single transformation results in changing another tree’s decision, then ignore it!

- Focus on the trees that predict -1
- For each tree: explore the positive paths, i.e., those that predict +1
- **Apply the transformations imposed by the positive path**

$$
\mathbf{x}_{j(\epsilon)}^{+}[i] = \begin{cases} 
\theta_i - \epsilon & \text{if the } i\text{-th condition is } (x_i \leq \theta_i) \\
\theta_i + \epsilon & \text{if the } i\text{-th condition is } (x_i > \theta_i)
\end{cases}
$$

- Choose the transformation with the minimum cost

$$
x' = \arg \min_{x_{j(\epsilon)}^+ \in \Gamma | \hat{f}(x_{j(\epsilon)}^+) = +1} \left\{ \delta(x, x_{j(\epsilon)}^+) \right\}
$$
What is the **minimum number of changes** to apply to a time series $T$ so that a given opaque classifier changes its prediction?

$T \rightarrow T^1 \rightarrow T^2 \rightarrow \ldots \rightarrow T'$

- **Reversible tweaking:** each subsequent transformation can override a previous one
- **Irreversible tweaking:** each subsequent transformation cannot override a previous one
Shapelets: class-distinctive time series subsequences capturing local trends in time series

Shapelet Tree*

Shapelet Dictionary*

Time series $T$

*Figures taken from Eamonn Keogh
Focus on the trees that predict -1

For each tree, explore the positive paths, i.e., those that predict +1

Try to force those trees to predict +1 by "tweaking" shapelet features of $T$

Given a non-leaf node $(S^j_k, \theta^j_k)$

- **Increase distance:**
  - If $S^j_k$ exists in $T$, that is $d_S(S^j_k, T) \leq \theta^j_k$
  - and the current node condition demands otherwise
  ✓ Increase the distance of all matching instances of $S^j_k$, so that they all fall above the distance threshold $\theta^j_k$
Time series tweaking: solution

- Focus on the trees that predict -1
- For each tree, explore the positive paths, i.e., those that predict +1
- Try to force those trees to predict +1 by ”tweaking” features of T

Given a non-leaf node \((S^j_k, \theta^j_k)\)

- **Decrease distance:**
  - If \(S^j_k\) does not exist in \(T\), that is \(d_s(S^j_k, T) > \theta^j_k\)
  - and the current node condition demands otherwise
  ✓ Decrease the distance of the best matching instance of \(S^j_k\), so that it falls below the distance threshold \(\theta^j_k\)
How to transform the time series?

- Consider $S$ as an $m$-dimensional point.
- Define an $m$-sphere with $S$ as its center and radius $\theta$.

The transformed time series counterpart of $S$ is given by the following equation:

$$
\tau_S(S, p^j_{ik}, \epsilon) = S^j_k + \frac{S^j_k - S}{\|S^j_k - S\|_2} (\theta^j_k + (\epsilon \delta^j_{ik}))
$$
Experimental setup

- **UCR time series repository:**
  - all binary classification datasets (26 datasets)

- **Competitor:**
  - 1-NN under the Euclidean distance

\[
\tau_{NN}(\mathcal{T}, y') = \arg \min_{\mathcal{T}' \mid (\hat{y}, \mathcal{T}') \in \mathcal{D}, \hat{y} = y'} d_E(\mathcal{T}, \mathcal{T}')
\]
Evaluation – metrics

**Average cost** of successful transformation, i.e., how costly is the transformation?

\[
c_\mu(\tau, \tau') = \frac{1}{n} \sum_{i=1}^{n} c(\tau_i, \tau(\tau_i, \tau'))
\]

**Compactness** of transformation, i.e., how much of the time series is changed?

\[
\text{compact}(\mathcal{T}, \mathcal{T}') = \frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \text{diff}(T_i, T_i')
\]

where

\[
\text{diff}(T_i, T_i') = \begin{cases} 
1, & \text{if } |T_i - T_i'| \leq e \\
0, & \text{otherwise.}
\end{cases}
\]
Evaluation – result

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\tau_{RT}$</th>
<th>$\tau_{IR}$</th>
<th>Cost</th>
<th>$\tau_{RT}$</th>
<th>$\tau_{IR}$</th>
<th>Compactness</th>
<th>Accuracy RSF</th>
<th>NN (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProximalPhalanxOutlineCorrect</td>
<td>0.5895</td>
<td>1.0056</td>
<td>0.5326</td>
<td>0.6186</td>
<td>0.5116</td>
<td>0.9998</td>
<td>0.8421</td>
<td>0.7782</td>
</tr>
<tr>
<td>SonyAIBORobotSurface1</td>
<td>1.7384</td>
<td>1.7260</td>
<td>4.7213</td>
<td>0.4429</td>
<td>0.4394</td>
<td>1.0000</td>
<td>0.8315</td>
<td>0.8090</td>
</tr>
<tr>
<td>SonyAIBORobotSurface2</td>
<td>1.8601</td>
<td>1.8566</td>
<td>5.6126</td>
<td>0.4133</td>
<td>0.3584</td>
<td>1.0000</td>
<td>0.9796</td>
<td>0.9949</td>
</tr>
<tr>
<td>TWOLeadECG</td>
<td>5.4407</td>
<td>5.8238</td>
<td>17.8733</td>
<td>0.6173</td>
<td>0.5705</td>
<td>1.0000</td>
<td>0.9697</td>
<td>0.7879</td>
</tr>
<tr>
<td>Wafer</td>
<td>3.0135</td>
<td>3.1419</td>
<td>8.6207</td>
<td>0.7152</td>
<td>0.6676</td>
<td>0.9999</td>
<td>0.9958</td>
<td>0.9799</td>
</tr>
<tr>
<td>Wine</td>
<td>0.5052</td>
<td>0.9301</td>
<td>0.1708</td>
<td>0.7529</td>
<td>0.3452</td>
<td>0.9996</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>WormsTwoClass</td>
<td>5.7723</td>
<td>7.2023</td>
<td>28.7383</td>
<td>0.4416</td>
<td>0.4219</td>
<td>1.0000</td>
<td>0.8269</td>
<td>0.7308</td>
</tr>
<tr>
<td>Avg.</td>
<td>2.3132</td>
<td>2.5329</td>
<td>8.9552</td>
<td>0.5733</td>
<td>0.4942</td>
<td>0.9999</td>
<td>0.8924</td>
<td>0.8240</td>
</tr>
</tbody>
</table>

- Reversible tweaking results in the **least costly** transformations.
- Irreversible tweaking results in the **most compact** transformations.
- The baseline is too naive.
Automated ranking/tagging/captioning of radiography exams

Diagnostic text: The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

Automated ranking/tagging/captioning of radiography exams

Diagnostic text: The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

Automated ranking/tagging/captioning of radiography exams

**Diagnostic text:** The cardiac contours are normal. XXXX basilar atelectasis. The lungs are clear. Thoracic spondylosis. Lower cervical XXXX arthritis.

**Abnormality tags:** Atelectases, Cervical Arthritis, Atelectasis, Spondylarthritis, Thoracic Spondylosis.

**TAG:** cardiomegaly

**EXPLAIN:** Slight cardiomegaly. Clear lungs. No effusion.
Automated ranking/tagging/captioning of radiography exams

DenseNet Encoder

TAG: cardiomegaly

Atelectasis, Spondylarthitis, Thoracic Spondylosis.
Closing remarks

- Security and privacy issues
- Hard to convince public authorities to make more data available to researchers
- Hard to convince doctors to adopt new “black box” models
- Cloud solutions are in many cases unacceptable
- Need to federated learning solutions
- Many players/systems are used by practitioners
- Need for a unified cross-border database of medical records
Thanks to…

Isak Karlsson  
Jon Rebane  
Maria Bampa

Aristides Gionis  
Hans E. Persson  
Henrik Boström  
Hercules Dalianis
Bibliography


• Iyad Batal, Dmitriy Fradkin, James Harrison, Fabian Moerchen, and Milos Hauskrecht. Mining recent temporal patterns for event detection in multivariate time series data. In ACM SIGKDD, 2012


• Karlsson, I., Zhao, J., Asker, L. & Boström, H. Predicting adverse drug events by analyzing electronic patient records. Proc. of the 14th conference on Artificial Intelligence in Medicine (AIME), Murcia, Spain, 2013
Bibliography

- Karlsson, I., and Zhao, J. Dimensionality reduction with random indexing: an application on adverse drug events detection, Submitted to CBMS, 2014

- Zhao, J., Karlsson, I., Asker, L. & Boström, H. Applying methods for signal detection in spontaneous reports to electronic patient records. KDD 2013 Workshop on Data Mining for Health care (DMH), Chicago, USA, 2013


Bibliography


• Moskovitch, R. & Shahar, Y. Classification of multivariate time series via temporal abstraction and time intervals mining. Knowledge and Information Systems, 45(1), pp.35-74, 2015


• S. Wu, Y. Chen, Mining Non-ambiguous Temporal Patterns for Inter-val-Based Events, IEEE Transactions on Knowledge and Data Engineering, 19 (6), 2007


• C. Yi-Cheng, J. Ji-Chiang, P. Wen-Chih, L. Suh-Yin, An Efficient Algorithm for Mining Time Interval-based Patterns in Large Databases, CIKM, 2010