H2020 SOCIETAL CHALLENGES: Secure, clean and analyzable psychiatric patient data

PRODUCTIVE SECTOR: Psychiatry and data-driven Pharmaceutics

The Industrial Problem

The literature of examining medical data considering its temporal (or longitudinal) nature has still been rare and not much effective results has been found in the past years. The problem is to find best practice solutions for analyzing certain (containing temporal) types of data, highlight the main differences between the methods, regarding advantages and disadvantages when applying them on psychiatric patient data.

Research groups

University of Szeged

UNIVERSITY OF SZEGED INSTITUTE OF INFORMATICS

UNIVERSITY OF SZEGED FACULTY OF MEDICINE DEPARTMENT OF PSYCHIATRY

A decade of experience in mathematical modelling, advanced statistics, machine learning, and software development.
Strong experience in clinical aspects of clinical research medical documentation systems, long-term genetic and epidemiological research.

Consulting Company

Takeda Pharmaceutical Company

A patient-focused, innovation-driven global pharmaceutical company, has a focused, world-class R&D innovation engine, making an impact on patients' lives by translating science into life-changing medicines.

APPLICATION OF ADVANCED MATHEMATICAL MODELS AND MACHINE LEARNING METHODS TO ANALYZE PSYCHIATRIC DISORDERS

HU-MATHS-IN
Hungarian Service Network for Mathematics in Industry and Innovations

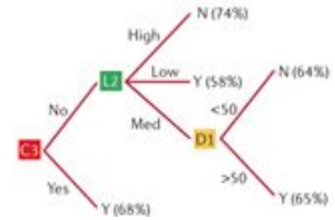
Challenges & Goals

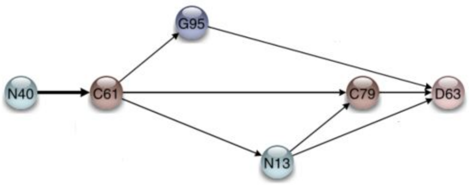
Main challenges

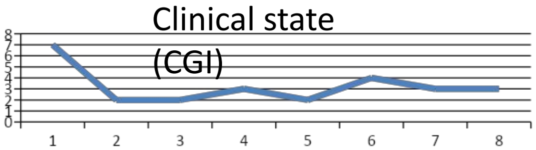
- Investigate, for example, disease comorbidities, patient stratification, drug interactions and clinical outcome of psychiatric patients.
- Apply methods from data mining, machine learning (ML) and text mining on high quality patient data collected in a patient registry database

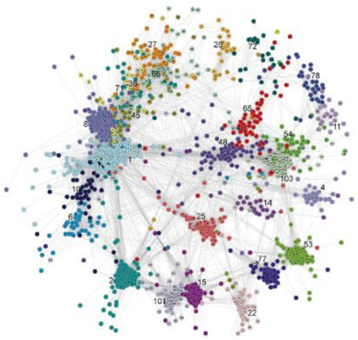
The goal of this project is twofold:

1. Detailed investigation of the applicability of ML methods and their modified version on psychiatric patient data.
2. Summarize what the best practice solutions are for analyzing psychiatric patient data, highlight the main differences between them









Approaches: decision tree, temporal network representaion, time series, similarty based networks and patient clustering

Mathematical and computational methods and techniques applied

***Data** collected*

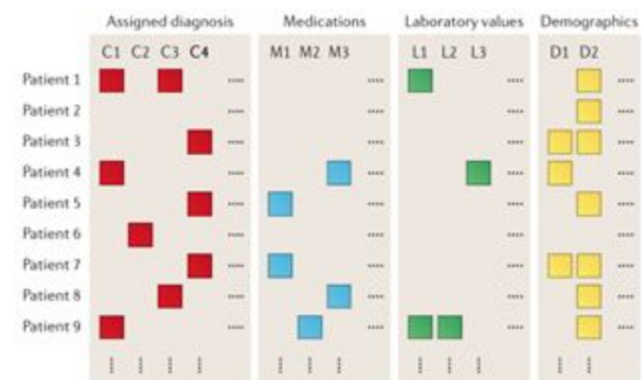
- PANSS (medical scale used to measure symptom severity of schizophrenic patients)
- CGI (measures illness severity (CGIS), global improvement or change (CGIC) and therapeutic response)
- Medication (Previous drug history and medication during hospitalization)
- Demographic data

***Methods** (using features – target variable data /supervised/ and patient similarity data /unsupervised/)*

- **Classification:** decision tree, random forest, stochastic gradient
- **Regression:** linear and logistic regression
- **Longitudinal data analysis:** graph-based analytics of significant time dependent diagnostic pairs

Drug	Dose	Start	End	Cause of modification
Aripiprazol tbl.	15 mg/nap	2015.06.12	2015.07.19	dózisemelés elégtelen hatás miatt
Aripiprazol tbl.	30 mg/nap	2015.07.19	2015.08.15	Depot injekcióra átállás rendszertelen gyógyszerszedés miatt.
Aripiprazol depot inj.	400 mg/ 4 hét	2015.08.10	2017.11.10	Relapszus miatt leállítva
Risperidon	4 mg/nap	2017.11.11	2017.11.25	EPS tünetek
Olanzapine	15 mg/nap	2017.11.26	Jelenleg is	

Drug usage data - example



Data in most medical applications can be represented in a feature matrix form

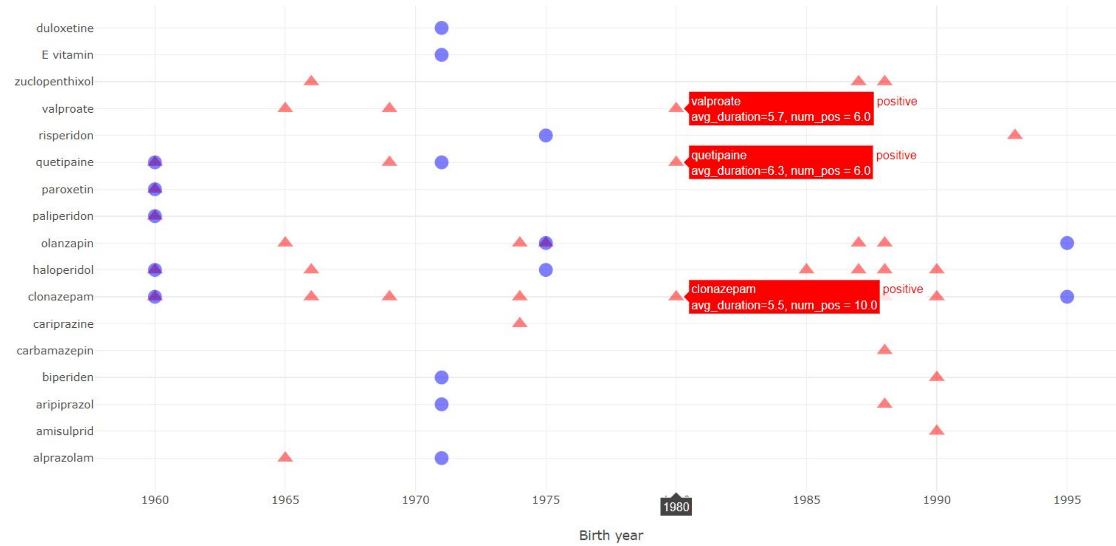
Results & Benefits to the company

Findings

- Accurate drug utilization data and its detailed statistical analysis
- PANSS positive symptoms mean is significantly smaller than negative mean
- Schizophrenic patient clusters (based on CGI and demographic data) correlates with drug medication
- Changes in negative symptoms over time correlated with drug medication

Benefits, future

- Machine learning and network approaches are promising (however hard to collect enough reliable data)
- Functional data analysis and topological data analysis potentials



Visualization of drog usage data regarding patients' age

Machine learning and network-based approaches for medical data are promising and able to provide accurate predictions and patient classifications from various perspectives