Characterizing Chronic Disease and Polymedication Prescription Patterns from Electronic Health Records

Martí Zamora, Manel Baradad, Marta Arias, Ricard Gavaldà
Universitat Politècnica de Catalunya, Barcelona
Ester Amado, Silvia Cordomí, Esther Limón, Juliana Ribera
Institut Català de la Salut, Barcelona

2nd IEEE DSAA, Paris, October 2015

Contact: gavalda@cs.upc.edu

Context

Catalan Institute of Health - ICS

- Provides primary healthcare for 80% of 7.5M people
- Hospitalary healthcare for about 20%
- Electronic Health Records almost fully digital since 2009

Context

The concerns:

- 5% of patients use 50% resources
- Aging
- Complex, chronic disease
- Polymedication
- Increasingly heterogeneous population

The Project and Intended Users

Health managers and planners at ICS:

1. Understand “the landscape” of complex, chronic disease
2. and polymedication - prescription patterns
3. Rationalize prescription patterns - costs and patient safety
4. Analyze diversity, find outliers
   - geography, demography, among healthcare centers . . .
5. Plan: Define indicators and policies, assess costs, allocate resources, make projections to future scenarios

The Project and Intended Users

Healthcare researchers:

1. Support hypothesis generation and intuition
2. Discover and explore subpopulations of interest
3. Mine interesting rules and interactions among variables
4. Create predictive and explanatory models

The Project and Intended Users

First-line clinicians and prescribers:

1. Alert of unusual diagnostic/prescription combinations
2. Support case-based reasoning
   - Retrieve patients similar to this one
   - Get recommendations for diagnostic & treatment

The Dataset

- ICS primary care visits, Barcelona, 2013
- 3 tables: patient basic info, health annotations, prescriptions
- 1.6M potential patients, 0.5M actually present
- 12M health annotations (diagnostics, tests, findings)
- 7M medication prescriptions

Limitations:

- Only primary care, no hospital data
- Only public network, no private care
- Only one year
- Potential inconsistencies - e.g. open episodes

The Project. Novelty

- Unfocused, exploratory. Many studies focus on one research problem
  - predicting one disease, cluster patients for one goal, find drug side-effects, . . .
- Tripartite graph patients - diagnostics - medications
  - other studies used e.g. diagnostics and genes
- *k*-ary, not binary, associations – Hypergraphs, not graphs
- Hierarchical itemsets - diagnostic codes and medications
- Detection of *open episodes*

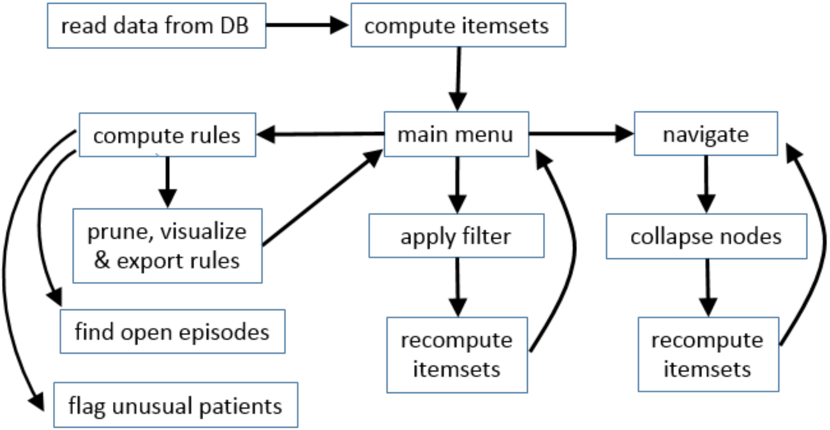
The Prototype so Far

- Generate *k*-ary diagnostic combinations
- Generate rules diagnostics - prescriptions
- Flag patients with unusual (alarming?) combinations
- Flag open episodes or prescription errors
- Navigate hypergraph of diagnostics and prescriptions
- First try at automatic predictor building

The Prototype - Workflow

Itemset = Subset of diagnostics $\cup$ Prescriptions

Maintain frequent itemsets of current subpopulation



Exploring the Hypergraph

Nodes: Sets of diagnostics and medications

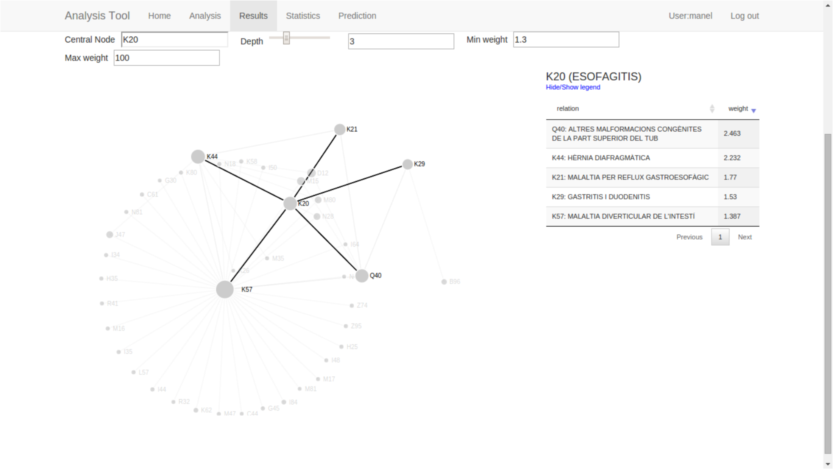
Edges: Strength of association; Pointwise Mutual Information

$$\mathrm{PMI}(A, B) = \log_{10} \frac{\Pr(A \wedge B)}{\Pr(A) Pr(B)}$$

Nodes can be collapsed (set union)

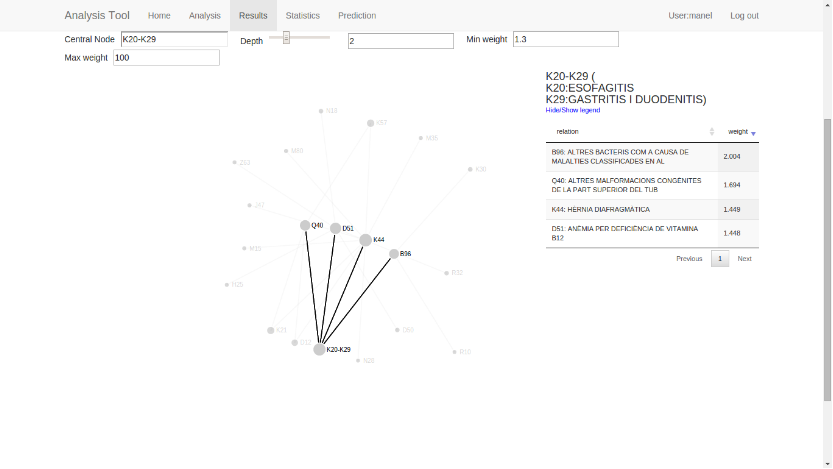
Exploring the Hypergraph

Graph around K20 (Esophagitis)



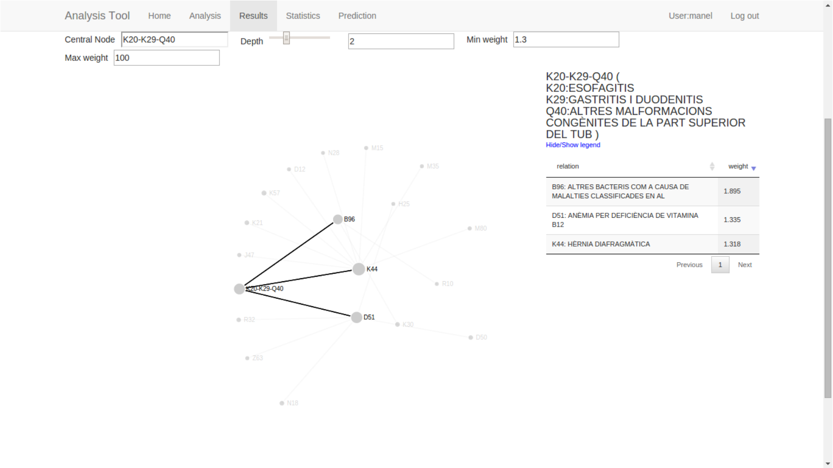
Exploring the Hypergraph

Graph around K20-K29 (Esophagitis + Gastritis/Duodenitis)



Exploring the Hypergraph

Graph around K20-K29-Q40
(Esophagitis + Gastritis/Duodenitis +
Other malformations of upper GI tract)



Implementation

- Client - server
- Borgelt's Apriori to find itemsets
- Custom association rule finder on top
- Two implementations of patient/diagnostic/prescription DB
  - RAM
  - Sparksee graph database
- But itemsets and hypergraph always in RAM

Some Results

With support 0.05% $\simeq$ 800 patients, confidence 0.1,

- Hypergraph with 918 diagnostics and 268 medications
- 4051 diagnostic-to-medication rules
- 2253 medication-to-diagnostic rules
- Prescriptions without diagnostics for about 10% of patients
  - Lower than expected: application does not require diagnostic for prescription
- Diagnostics without usual medications for about 16% patients
  - Many are indeed open episodes

Clinical Significance

Under evaluation. 3 types of “discoveries”

- Well known, not surprising, but reassuring the program found them

  (Diabetes $\leftrightarrow$ retinopathy)
  (Omeprazol for most everything)

- Unnoticed before, but believable

  (Bedsores for advanced Alzheimer)

- Unnoticed and surprising

  (Retinopathy more strongly associated to hypertension than to diabetes)

First Prediction Trial

(not in proceedings)

Factors that predict Hip Fracture

- Linear regression and odds ratio
- 7 out of 10 highest scorers reported in specialized literature

Conclusions

- System is well able to interactively find associations diagnostics / medications
- Clinicians satisfied with initial interactions
- Detailed clinical study in course

- There's no such thing as "user-friendly enough"

Future Work (lots!)

- Improve rule pruning
- Improve interpretation of rule exceptions
- Taxonomies of diagnostics and medications
- Temporal evolution. Trajectories
- Predictive model building
- Patient clustering
- Differential analysis (geographic, demographic)
- Retrieve similar cases
- Suggest diagnostic/treatment
- Privacy, information sharing

Advertising

Looking for:

- Research partners
- Data partners
- Project partners

Characterizing Chronic Disease and Polymedication Prescription Patterns from Electronic Health Records

Martí Zamora, Manel Baradad, Marta Arias, Ricard Gavaldà
Universitat Politècnica de Catalunya, Barcelona

Ester Amado, Silvia Cordomí, Esther Limón, Juliana Ribera
Institut Català de la Salut, Barcelona

2nd IEEE DSAA, Paris, October 2015

Contact: `gavalda@cs.upc.edu`

Rule Mining

Find all rules

$$A_1 \dots A_k \to B_1 \dots B_\ell$$

with given support and confidence

Heuristics to purge rules (improvable):

- Low lift: remove $AB \to C$ if $A \to C$ same confidence
- Implied by transitivity:
  remove $A \to C/(\sigma_1 \cdot \sigma_2)$ if $A \to B/\sigma_1$ and $B \to C/\sigma_2$
- Removals make sense to clinicians

Open Episodes and Unusual Patients

From the rules we find patients with:

- Medication not justified by recorded diagnostics
- Diagnostics without any of its usual medication

- Open episode?
  Recording error?
  Clinician error?
  Conscious clinician decision?
- More heuristics and larger timespan data to decide