IEEE CBMS 2017
30th IEEE International Symposium
on Computer-Based Medical Systems

Application of affinity analysis techniques on diagnosis and prescription data

Sanida Theodora, Varlamis Iraklis
sanida.dora@gmail.com, varlamis@hua.gr

Department of Informatics and Telematics,
Harokopio University of Athens, Greece

INTRODUCTION

Aim: This study performs an Affinity Analysis on diagnosis and prescription data in order to discover co-occurrence relationships among diagnosis and pharmaceutical active ingredients prescribed to different patient groups.

Demonstrate how affinity analysis techniques can be applied to medical data sets in order to generate association rules and sequential patterns.

Data: Data were collected during consecutive visits of 4,473 patients in a 3 years period, and focused on patients suffering by hypertension (I10 ICD-10 code) and/or hypercholesterolemia (E78 ICD-10 code).

The information recorded during the consecutive visits to the practitioner, includes demographic data, ICD-10 diagnosis codes and prescribed medicines (more specifically the pharmaceutical active ingredients).

Methods: Association rule and sequential rule mining techniques have been applied.

Rules and patterns are related both to the diagnosis and pharmaceutical active ingredients that have been prescribed to patients.

METHODOLOGY

A. *Data selection*

Patient sub-groups:

- HT: 1,762 patients (1,028 women, 734 men) who have been diagnosed with hypertension at least in one visit. 10,136 distinct visit records.
- HCL: 1,295 patients (805 women, 490 men) who have been diagnosed with hypercholesterolemia at least once. 7,455 distinct visit records.
- HCT: 729 patients (448 women, 281 men) who have been diagnosed with hypertension and hypercholesterolemia at least once. 5,370 distinct visit records.

B. *Data pre-processing*

- A total of 17,758 distinct visits of 4,154 distinct patients.
- Visit: {ICD-10 codes, prescribed substances}

C. *Data transformation*

Transaction database:

- **(p)** each visit is a transaction, containing all the diagnosis and drugs prescribed in this visit.
- **(v)** each patient is a transaction, containing all the distinct diagnosis for the patient and all the distinct drugs prescribed to the patient.

Knowledge extracted (associations):

- <diseases that frequently co-occur among patients of the practitioner>
- <sets of medicines that patients take at least once in the monitored period>
- When order of diagnosis/drugs is important→ sequential associations

D. *Data mining*

Application of association rules and sequential association rules implemented on RapidMiner Studio.

Use a minimum **support** and minimum **confidence** threshold, which limits the number of generated rules and increases their importance.

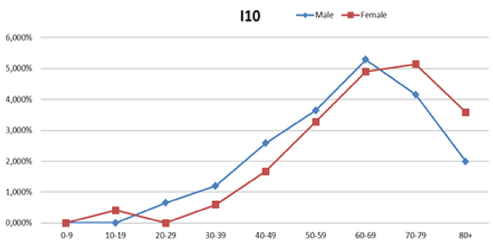
E. *Interpretation of the results*

- The distribution of cases per disease.
- The list of the most popular diseases.
- The distribution of cases per age group and gender for the most popular diagnoses and the most prescribed medicaments.

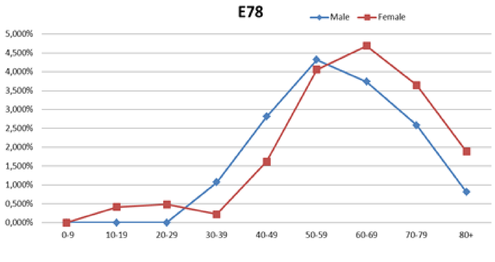
The World Health Organization's classification of age groups has been employed.

RESULTS

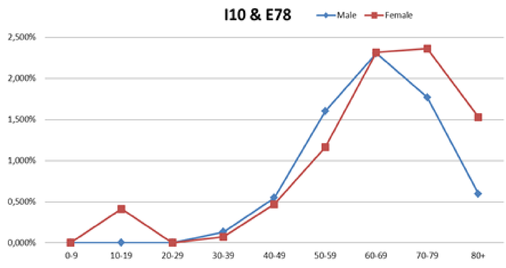
The rates have been projected to 100,000 persons, using the Greek Statistics Authority (GSA) data for the population of the city of Athens, which had 3.8 million inhabitants according to the 2011 census.



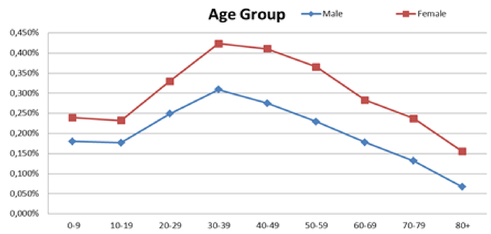
Population percentage per gender and age-group for patients with hypertension (I10).



Population percentage per gender and age-group, patients with hypercholesterolemia (E78).



Population percentage per gender and group age for patients with both hypertension and hypercholesterolemia (I10&E78).



Population percentage per gender and group age for the whole set of patients.

	Association Rules			Sequential Patterns		
Diagnosis	**Dataset**					
	HCLp	*HTp*	*HCTp*	*HCLp*	*HTp*	*HCTp*
Hypertension	√			√		
Hypercholesterolemia		√			√	
Heart disease	√		√	√	√	√
Gastroesophageal reflux disease	√		√	√	√	√
Insulin-dependent diabetes mellitus	√	√	√	√	√	√
Non-insulin dependent diabetes mellitus						√
Anxiety disorders					√	√
Depression						√
Osteoporosis						√

Associations and patterns per patient throughout all visits

	Association Rules			Sequential Patterns		
Diagnosis	**Dataset**					
	HCLv	*HTv*	*HCTv*	*HCLv*	*HTv*	*HCTv*
Hypertension	√			√		
Hypercholesterolemia		√			√	
Heart disease			√	√	√	√
Gastroesophageal reflux disease			√	√	√	√
Insulin-dependent diabetes mellitus					√	√
Non-insulin dependent diabetes mellitus						
Anxiety disorders					√	√
Depression						√
Osteoporosis						

Associations and patterns per visit

Dataset	Association Rules	Sequential Patterns
HCLp	Amlodipine, Ezetimibe, Hydrochlorothiazide, Irbesartan, Metformin, Olmesartan, Rosuvastatin, Simvastatin, Valsartan.	Amlodipine, Atorvastatin, Hydrochlorothiazide, Irbesartan, Metformin, Rosuvastatin, Simvastatin, Valsartan.
HTp	Amlodipine, Atorvastatin, Hydrochlorothiazide, Irbesartan, Metformin, Olmesartan, Simvastatin, Valsartan.	Amlodipine, Atorvastatin, Hydrochlorothiazide, Irbesartan, Metformin, Olmesartan, Rosuvastatin, Simvastatin, Valsartan.
HCTp	Amlodipine, Atorvastatin, Clopidogrel, Ezetimibe, Hydrochlorothiazide, Irbesartan, Metformin, Olmesartan, Omeprazole, Rosuvastatin, Simvastatin, Valsartan.	Amlodipine, Atorvastatin, Clopidogrel, Ezetimibe, Hydrochlorothiazide, Irbesartan, Metformin, Olmesartan, Omeprazole, Rosuvastatin, Simvastatin, Valsartan.

Popular active substances per patient throughout all visits

Dataset	Association Rules	Sequential Patterns
HCLv	Atorvastatin, Hydrochlorothiazide, Simvastatin, Valsartan.	Atorvastatin, Hydrochlorothiazide, Simvastatin, Valsartan.
HTv	Amlodipine, Atorvastatin, Hydrochlorothiazide, Irbesartan, Olmesartan, Simvastatin, Valsartan.	Atorvastatin, Hydrochlorothiazide, Metformin, Omeprazole, Simvastatin.
HCTv	Amlodipine, Atorvastatin, Clopidogrel, Hydrochlorothiazide, Irbesartan, Metformin, Simvastatin, Valsartan.	Atorvastatin, Clopidogrel, Hydrochlorothiazide, Omeprazole, Simvastatin, Valsartan.

Popular active substances per visit

CONCLUSIONS

Results uncovered a correlation of hypertension and hypercholesterolemia with:

- coronary artery disease,
- gastroesophageal reflux disease and
- insulin dependent diabetes mellitus.

Additionally, in patients with hypertension we have observed positive correlation with:

- anxiety disorders.

Finally, in patients diagnosed both with hypertension and hypercholesterolemia, a correlation was observed with:

- non-insulin dependent diabetes,
- anxiety disorders,
- depressive disorders, and
- osteoporosis.