

VISUAL INSTANCE-BASED RECOMMENDATION SYSTEM FOR MEDICAL DATA MINING

Joris Falip^a, Amine Aït Younes^a, Frédéric Blanchard^a, Brigitte Delemer^b, Alpha Diallo^b, Michel Herbin^a

joris.falip@univ-reims.fr

September 7 2017

^aCReSTIC, University of Reims Champagne-Ardenne, France

^bUniversity Hospital of Reims, France

INTRODUCTION

“Be the change that you wish to see in your curated dataset.” Mahatma Gandhi

Medical datasets are...

- Widespread
- Very large
- Underexploited

Some caveats

- High dimensionality
- Unstructured
- Heterogeneous data
- Missing values

Goals

Enrich data with a representativeness score then structure elements by creating associations.

This structure can be used for :

- Visualization
- Exploration
- Recommendation

Constraints

- Understandable and interpretable
- Simulate reasoning of medical experts

A MUCH NEEDED TOOL

Constraints

- Understandable and interpretable
- Simulate reasoning of medical experts

Challenges

- Avoid overgeneralization
- Preserve outliers data

STRUCTURING DATA

“Without k-neighborhood, life would be a mistake.”

Nietzsche

Finding typical individuals using an election process.

Steps

- Ranking individuals on each dimension
- Aggregate the results into a score : the *Degree of Representativeness*
- Link individuals to their neighbor with the highest *DoR*

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1			...
Ind2			...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100		...
Ind2		100	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100		...
Ind2	90	100	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100	75	...
Ind2	90	100	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100	75	...
Ind2	90	100	...
...

ALGORITHM

Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100	75	...
Ind2	90	100	...
...

ALGORITHM

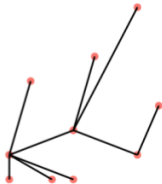
Dataset			
	Feat. 1	Feat. 2	...
Ind1	67	173	...
Ind2	53	161	...
...



Feature 2 - ranking			
Feature 1 - ranking			
	Ind1	Ind2	...
Ind1	1st	4th	...
Ind2	2nd	1st	...
...



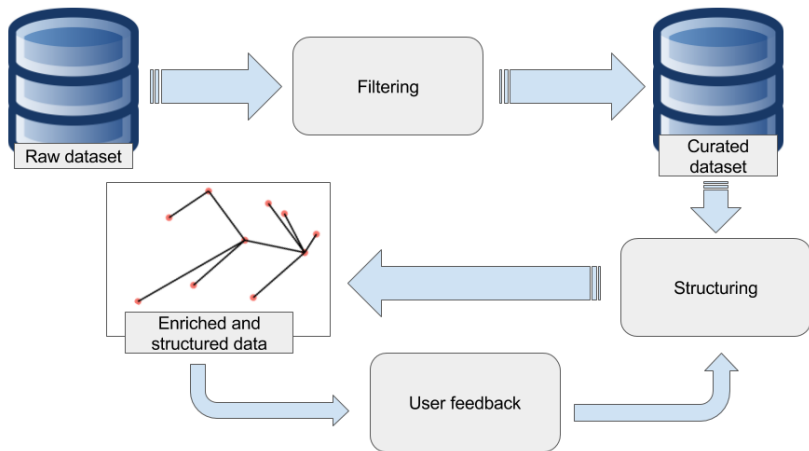
Feature 2 - scores			
Feature 1 - scores			
	Ind1	Ind2	...
Ind1	100	75	...
Ind2	90	100	...
...



VISUALIZATION TOOL

“A friend is someone who reads all your code and still loves you.” Elbert Hubbard

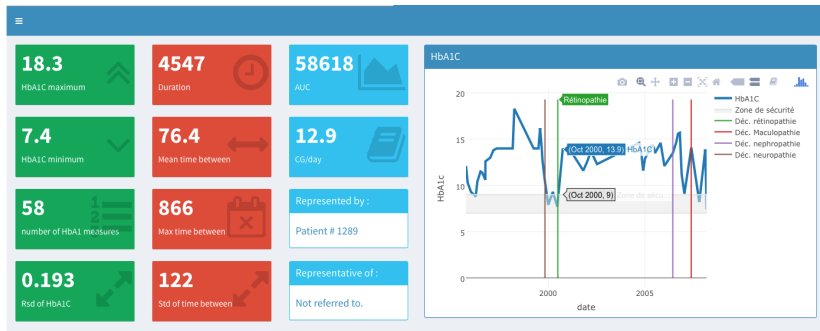
WORKFLOW



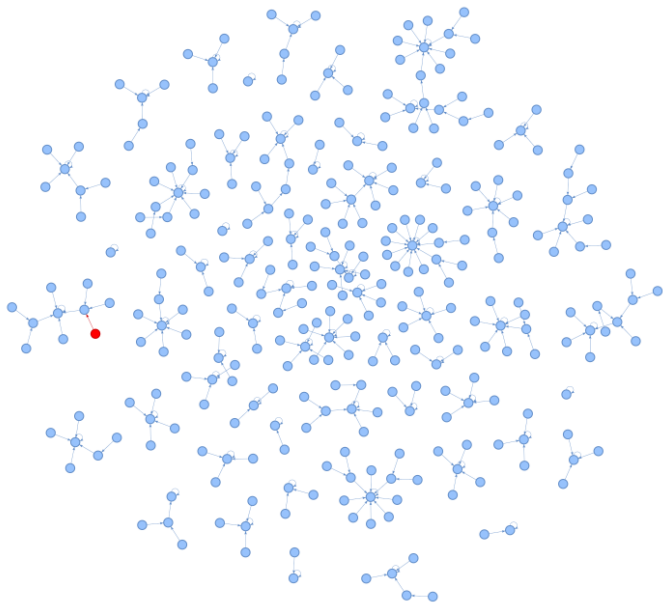
A tool that can be used...

- Easily by doctors and medical experts
- As a mean to explore medical dataset
- As a recommendation system

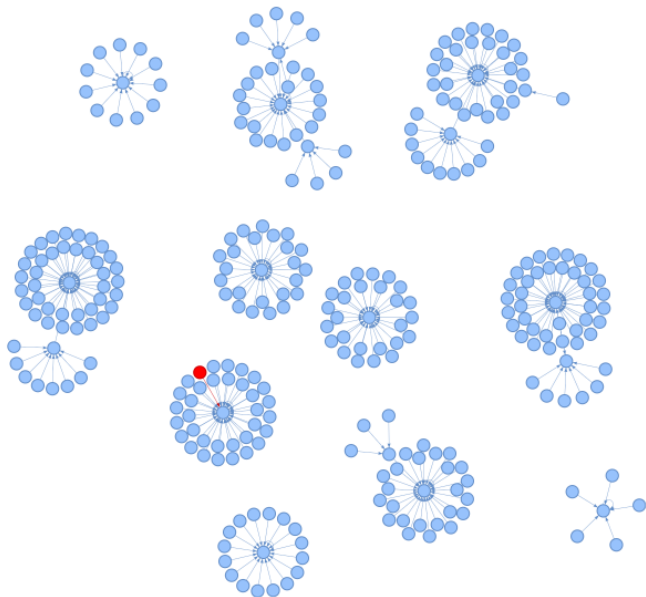
VISUALIZATION



RECOMMENDATION



RECOMMENDATION



CONCLUSION

“Don't cry because it overfitted, smile because it happened.” Dr. Seuss

A tool to assist medical experts manipulating electronic medical records.

- Algorithm suited for use with medical data
- New way to enrich data with representativeness
- A structure allowing exploration and recommendation
- Constraints expressed in an understandable way

- Automatic user feedback integration for recommendations
- Applying our tool to datasets related to other diseases

VISUAL INSTANCE-BASED RECOMMENDATION SYSTEM FOR MEDICAL DATA MINING

Joris Falip^a, Amine Aït Younes^a, Frédéric Blanchard^a, Brigitte Delemer^b, Alpha Diallo^b, Michel Herbin^a

joris.falip@univ-reims.fr

September 7 2017

^aCReSTIC, University of Reims Champagne-Ardenne, France

^bUniversity Hospital of Reims, France