UNIVERSITY OF WATERLOO

TECHNOLOGY SUMMARY

Patient	Class	Indicants (signs, symptoms and measurements)													
		age	sex	cpt	rbp	sc	fbs	rer	mhra	eia	oldpeak	spess	nmvc	thal
Entity ID				chest patin type	resting blood pressure	serum choles.	fasting blood sugar	resting CG results	maximum heart rate achieved	exercise indeced angi	ST depression	slope of peak exercise	no. o Major Vessel	thal healthy sick
1	healthy	67	0	3	115	564	0	2	160	0	1.6	2	0	4
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
150	healthy	57	1	4	140	192	0	0	148	0	0.4	2	0	6
151	Patient	70	1	4	130	322	0	2	109	0	2.4	2	3	3
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
270	Patient	67	1	4	160	286	0	2	108	1	1.5	2	3	3
271	RareCase1	16	0	1	180	512	0	2	70	0	1.5	2	0	6
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
280	RareCase1	25	1	2	140	467	1	2	80	1	1.5	1	1	6
281	RareCase2	80	1	3	90	185	0	2	138	0	3	2	0	4
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
290	RareCase2	83	1	4	92	286	0	2	128	1	1.5	2	1	4

Fig. 1. Relational table, from which subtle association patterns of correlated indicants are discovered, disentangled and located.

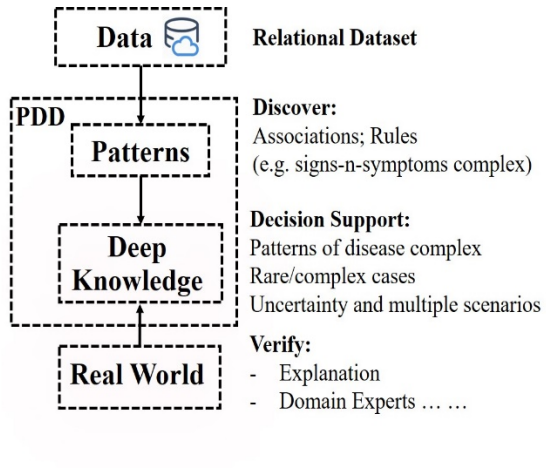

Fig. 2. PDD internal operation

Reference

10146

Patent status

Patent Pending

Stage of development

Working server prototype and validating application data

Contact

Scott Inwood
Director of Commercialization
Waterloo Commercialization Office
519-888-4567, ext. 33728
sinwood@uwaterloo.ca
uwaterloo.ca/research

Pattern Discovery and Disentanglement (PDD) AI Software for Deep Knowledge Discovery from Relational Data

Background

Identifying, categorizing and explaining patterns in relational date sets is complex due to a wide range of interrelated and intertwining factors. Thus functional associations are often masked at the data level due to such hidden entanglement. This is a specific problem related to relational data sets in many areas including healthcare, finance and cybersecurity. There is a need to both discover and disentangle patterns inherent in relational data sets in order to surface deeper actionable knowledge in a way that is explainable (i.e. mitigate the AI "black box" confidence issue).

Description of the invention

University of Waterloo (UW) researchers have developed novel PDD (Pattern Discovery and Disentanglement) software to discover deep knowledge inherent in relational and array data for various applications. PDD uses novel autonomous and scalable algorithms to disentangle and discover deep knowledge to reveal subtle functionality and relations and uses the knowledge discovered to enhance machine learning, achieving much better prediction results with explanation. The software also solves imbalanced class, biases, anomaly and outliers problems that have plagued ML for decades.

Advantages

- **Time/Cost Reduction –** prediction based only on data with no reliance on explicit prior knowledge.
- **Higher accuracy -** leveraging rare cases and mislabeling
- **Robust to data noise, biases and imbalanced classes**
- **Flexible** - can be applied to a wide range of scenarios
- **Explainable –**providing explicit patterns/pattern clusters for further exploration, experts' understanding and knowledge organization.

Potential applications

- Unsupervised classification/tagging of data in relational datasets – as an example, if a tabulated record associated with heart disease (Figure.1) is inputted to PDD software, the output can include:
  1- Automatic labeling and grouping of patients with explanation
  2- Identification of correlated indicants for each group
  3- Identification of Rare Cases and the patterns they possess
  4- Detection of early stage of the disease in patients
  5- Prediction of "Healthy" and "Sick" patients and identification of mislabeled, biased cases and outliers.
- Other sectors where relational tables are used, e.g. Finance, Banking, Insurance, Logistics, Manufacturing and Cybersecurity