



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Mgr. Soňa Molnárová

Oddelenie pre vedu, výskum a rozvoj

Prírodovedecká fakulta UPJŠ

Šrobárova 2

040 01 Košice

Referee report of the habilitation thesis of Lubomir Antoni, PhD.

Fuzzy Formal Concept Analysis and its Position in Data Science

Data Science is a hot research topic, we can agree with applicant: data science aims to propose methods for obtaining fruitful information from structured and unstructured data. One can see it from different point of view – use case, data, methods (prominently induction), prototypes, experiments, metric. One of not so widely studied possibility is to study induction based on concept lattices – the topic of this thesis - hence it is a new step in the right direction. The topic is up to date and interesting.

Please allow me some general methodological remarks which do not lessen evaluation of this thesis, rather, they show future potential. The key place in integration (Figure 1.1: Elements of Data Science) is verification in experiments. It has deeper reason. Falsifiability is a deductive standard of evaluation of scientific theories and hypotheses, introduced by the philosopher of science Karl Popper in his book *The Logic of Scientific Discovery* (1934). A theory or hypothesis is falsifiable (or refutable) if it can be logically contradicted by an *empirical test* (of course this is different in physics, biology, sociology and informatics). In theoretical informatics one verifies hypotheses by proofs and refutes by a counterexample. Famous example is the Kolmogorov's theorem on approximation of functions by ANN. This did not work in empirical tests and the mental

block was broken with deep neural nets. Let us keep it in mind when dealing with Data Science.

Formal Concept Analysis (FCA) can be seen as unsupervised Machine Learning technique that applies mathematical lattice theory to organize data based on objects and their shared attributes. The main hope is that it can bring something new. Namely, the concept lattice approach provides the possibility to explore other possible groups of items which are not obtained e.g. by classical induction of directed acyclic graphs, since the entropy function returns only the best splitting attribute for each node (to illustrate this, see Figure 2.9). Concept lattices resemble graphs, trees, ... and hence are comparable to induction of association rules, trees etc. and there are many points where FCA can meet other communities.

Thesis consist of review of relevant literature, description of main results and in Appendix of seven published papers forming the core of the thesis. Papers relevant to the connection between Formal Concept Analysis and supervised learning are mentioned

Paper A) Authors presented a computational method for generation of attribute implications initiated in the field of formal concept analysis. Results are interesting for educators, data are large. I would be interested in comparison with other association rules induction especially based on entropy based splitting (see this thesis p. 28). Or the GUHA method of Peter Hajek which is also not using entropy optimization.

Paper B) Applicant on page 37 offers a possible understanding of FCA terms from the point of view of data-mining (transaction, items and degree in which the transaction contains respective item). In this understanding it is interesting the proposal (paper D)) of a novel Gaussian probability index for evaluation of stability of subsets of objects in one-sided fuzzy formal concepts (and this enables further algorithmic improvements).

Papers C), E) and F) deal with various characterization of patterns which can be found in data and wait for empirical tests in Data Science.

Paper G) contains an important idea. Namely, how to build a minimal subset of attributes preserving the information of the original knowledge system. I would be interested in measuring success of such a system (in the view of Popper's thesis). Is this connected also to stability of subsets of objects from paper D)? Are stable subsets also preserving knowledge?

As I said in the beginning, topic is up to date, from the point of view of FCA it is a step in the right direction. A very promising step, opening many future research possibilities.

In addition (to empirical validation), one can think about possible extension of authors approach to other hierarchies of concepts as in Word Net or coming from Wiki RDF data. Research of making web more semantical needs automated methods of concepts creation. Of course this would lead us to cross borders of FCA community and try to convince others about contributions and improvements brought by this approach. And finally, one has to find understanding in communication of broader Data Science communities in sharing tools, data, methods, prototypes and comparison and repeatability of experiments using metrics all use. It is a pity that author does not mention his further papers related to Data Science, especially the one from CLA'22. Promising could be also to integrate authors approach with natural language processing methods from Krajci's seminar.

Author's publication record is comparable to one of an associate professor.

	Scopus	Google Scholar	WoS	DBLP	MathSci
papers	32	34	9	24	6
Citations all	193	217	118	.	7
no self cit. of L. Antoni	145	.	110	.	.
no self-citations of all	130
h-index	7	7	6	.	.
Avg/ item. .	.	.	13.11	.	.
Avg/ year .	.	.	11.8	.	.
Journal articl	.	.	.	9	.

The habilitation thesis under concern fulfills all necessary standards and brings very promising ideas for future development. I propose to award the docent degree in Informatics after a successful defense.

Prague 2.3.2023

Peter Vojtas