Information technologies - synergy of theory and application*

Peter Vojtáš

Department of Software Engineering, School of Computer Science, Faculty of Mathematics and Physics, Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
vojtas@ksi.mff.cuni.cz

In this lecture we give several examples and lessons learned from research, development and experiments in the area of theory and applications of information technology. We will try to describe a possible synergy of theory and application too. Namely, to describe where practical needs bring new problems for theory and where theory helps to formulate methods, which should be verified in practice.

In the theoretical part we will mention research on correctness and completeness of fuzzy logic programming [1,2] and various measures for evaluating success. In applications we mention acquaintance with development and experiments of preferential querying and user dependent top-k answers [3]. In all of these it also depends on whether our task is deductive (querying), inductive (learning) or abductive. In practice, it is important to have a user behavior model (for many different users). To see what is (can be, must be) done automatically (trained, assisted, unsupervised, ...) and what by human, what is domain dependent and what is generic. Our fuzzy model is not a mere generalization from two values to many values. The key point of our study here is our understanding of fuzzy value as preference degree. Using fuzzy as preferences enlightens phenomena which in a two valued world are not visible at all.

From mathematical point of view one can generalize LP to many valued logic. Also here we face several challenges. Should our rules be implications or clauses, should our computation be refutation or query answering, is unification touched by this or not? In two valued logic these are equivalents, in [1] we have developed a model of ([0,1] valued) fuzzy logic programming FLP with implicative rules and computation based on backward usage of modus pones (and possible extension with fuzzy similarity in [4]).

Concerning implementation of this system, M. Lieskovsky has constructed in [5] a fuzzy Warren abstract machine. Our system enables new form of cuts for threshold queries (see [1]).

Further development went in two directions. (First) What happens in finitely valued case when different attributes take different number of truth values (see [2])? (Second) In our model we have a continuous semantic, on the other side in [6] GAP was not continuous (and general connection between these two models was also an open problem). This was all solved in [2], introducing a model based on left continuous conjunctors, with a weak form of border condition (without associativity and commutativity) and with body aggregation . We have shown that FLP is (in a sense) equivalent to GAP.

In the application part we make difference between case studies and use cases. Case studies include descriptions of systems that have been deployed within an organization, and are now being used within a production environment. Use cases include examples where an organization has built a prototype system, but it is not currently being used by business functions [18]. Repeatability of experiments is also an issue, see e.g. [19]. So our applications here are not true deployed applications, they are rather experiments (use case prototypes) and repeatability is not always fully satisfactory enabled.

Main impulse for these considerations came from a referee refuting our paper in an application oriented conference. He/She asked – where from do you have rules of your FLP? So far main motivation in fuzzy were toy examples with tall Swedes and young basketball players. Our motivation was real life examples where fuzzy degree is the degree of user preference (and no more fuzzy linguistic variables with modifiers like very tall ...). A well developed counterpart of this is already in preferential querying, where only top-k most preferred answers are interesting. Main contribution here was made by R. Fagin (see e.g. [10]), who (in a datalog setting, without function symbols) assumed we have objects in several lists repeatedly ordered by different attributes (local) preferences and gave an original optimal algorithm for top-k for this setting. This direction was further investigated by P. Gurský who has implemented and experimented with several heuristics (see e.g. [7]). V. Vaneková has developed several knowledge representation models for this ([8]).

But referee asked where from do you have those rules (Fagin assumes we have the (query) rules, with local preferences and global aggregation (in good concordance with our result FLP = GAP)? So now the

* This work was partially supported by Czech Grant Agency GA CR under number 202/10/0761.

question is, where from we have local preferences (user preferences on attributes represented by a fuzzy (ranking) set) and where from do we have combination (aggregation] function giving the global preference? This is an inductive task. With T. Horváth, A. Eckhardt we have developed several inductive models (see e.g. [9]).

Moreover a practical problem occurred. Well, assume we have different users (with different attribute preferences and aggregation). What are and where from are inputs? Do we assume user implicit inputs (e.g. click stream behavior) or (some form of) user explicit inputs. User aspects of these problems are developed in[3,11,12] and we have to admit that experiments are mostly done with an artificially generated user, very few human user experiments were done (and we have a problem how to evaluate them). Supporting data storage for these tasks is challenged too, we gave a model of fuzzy relational algebra for flexible querying in [13] and an index structure for multiple user preferential queries in [14].

For theoretical part it is now clear that equality of fuzzy sets (correct answers and computed answers) is not a good measure and correctness and completeness results have to be reconsidered with some order violation/concordance measures.

Further, from an experimental point of view, we went in direction of web information extraction ([16,17]), because it is also interesting to know where are all these data from (after where are rules from). Situation connecting web and user is heavily influences by uncertainty, starting research is done in [15].

We can conclude, that synergy between theoretical and applied (experimental) research and development was beneficial for both of them.

References

1. P. Vojtáš: *Fuzzy logic programming.* Fuzzy Sets and Systems, **124** (3), 2001, 361–370.
2. S. Krajči, R. Lencses, P. Vojtáš: *A comparison of fuzzy and annotated logic programming.* Fuzzy Sets and Systems, **144**, 2004, 173–192.
3. A. Eckhardt, P. Vojtáš: *Combining various methods of automated user decision and preferences modeling.* In: MDAI 09, Springer Verlag, 172-181.
4. Jesús Medina, Manuel Ojeda-Aciego, Peter Vojtáš: *Similarity-based unification: a multi-adjoint approach.* Fuzzy Sets and Systems, **146** (1), 2004, 43–62.
5. M. Lieskovsky: *Quantitative search strategies.* Neural Network World.
6. M. Kifer, V.S. Subrahmanian: *Theory of generalized annotated logic programming and its applications.* J. Log. Program. **12** (3& 4), 1992, 335–367.
7. P. Gurský, P. Vojtáš: *On top-ksearch with no random access using small memory.* ADBIS 2008, 97–111.
8. V. Vaneková, P. Vojtáš: *Comparison of scoring and order approach in description logic EL(D).* SOFSEM 2010: 709–720.
9. A. Eckhardt, T. Horváth, P. Vojtáš: *Learning different user profile annotated rules for fuzzy preference top-k querying.* SUM 2007, 116–130.
10. R. Fagin, A. Lotem, M. Naor: *Optimal aggregation algorithms for middleware.* J. Comput. Syst. Sci. **66** (4), 2003, 614–656.
11. P. Jenček, P. Vojtáš, M. Kopecký, C. Hösch: *Sociomapping in text retrieval systems.* FQAS 2009, 122–133.
12. B. Václav, A. Eckhardt, and P. Vojtáš: *A web shop with user preference search capabilities.* To appear in Web Intelligence/IAT Workshops 2010, IEEE Computer Society, 2010.
13. J. Pokorný, P. Vojtáš: *A data model for flexible querying.* ADBIS 2001, 280–293.
14. A. Eckhardt, J. Pokorný, P. Vojtáš: *A system recommending top-k objects for multiple users preferences.* FUZZ-IEEE 2007, 1–6.
15. A. Eckhardt, T. Horváth, D. Maruščák, R. Novotný, P.Vojtáš: *Uncertainty issues and algorithms in automating process connecting web and user.* URSW (LNCS Vol.) 2008, 207–223.
16. J. Dědek, P. Vojtáš: *Fuzzy classification of web reports with linguistic text mining.* Web Intelligence/IAT Workshops 2009, 167–170.
17. R. Novotný, P. Vojtáš, D. Maruščák: *Information extraction from web pages.* Web Intelligence/IAT Workshops 2009, 121–124.
18. Semantic Web Case Studies and Use Cases, `http://www.w3.org/2001/sw/sweo/public/UseCases/`
19. S. Manegold et al.: *Repeatability & workability evaluation of SIGMOD 2009.* SIGMOD Record **38** (3), 2009, 40–43, see e.g. `http://www.sigmod08.org/sigmod_call_papers.shtml` sub 6.