Observational Calculi – Tool for Semantic Web (Poster abstract)

Jan Rauch

Faculty of Informatics and Statistics, University of Economics, Prague *
nám W. Churchilla 4, 130 67 Prague, Czech Republic, rauch@vse.cz

The goal of this poster is to draw attention to observational calculi, to explain their main features and to outline possibilities of their application in Semantic Web. Observational calculi were defined in the book [1] as a tool to answer the questions (i) *Can computers formulate and verify scientific hypotheses?* (ii) *Can computers in a rational way analyze empirical data and produce a reasonable reflection of the observed empirical world? Can it be done using mathematical logic and statistics?* Formulas of observational calculi correspond to suitable statements concerning observational data. The important observational calculi studied in [1] are *observational predicate calculi.* They are defined by modifications of predicate calculi. The modifications consist in allowing only finite models (that correspond to analyzed data in the form of {0,1} - data matrices) and in adding generalized quantifiers. The generalized quantifiers are used to express patterns corresponding to statistical hypothesis tests. Important theoretical results concerning logic of observational calculi are published in [1].

Observational calculi are also studied in relation to data mining [3]. There are important new results concerning logic of association rules [4] enhancing results achieved in [1]. The association rule is understood as an expression $\varphi \approx \psi$ where φ and ψ are Boolean attributes derived from columns of an analyzed data matrix. Intuitive meaning of the rule $\varphi \approx \psi$ is that Boolean attributes φ and ψ are associated in the way corresponding to the symbol $\approx$. The symbol $\approx$ is called the *4ft-quantifier.* It is associated to a condition concerning a four-fold contingency table of φ and ψ in the analyzed data matrix. This condition corresponds e.g to χ^2-test or it can in a simple way express various types of relations (implication, equivalence) between φ and ψ. There are both theoretically interesting and practically important results concerning logic of association rules [1, 4]. An example is a criterion of correctness of deduction rules of the form $\frac{\varphi \approx \psi}{\varphi' \approx \psi'}$ where $\varphi \approx \psi$ and $\varphi' \approx \psi'$ are association rules [4].

Possibilities of applications of observational calculi in Semantic Web are related to analytical reports presenting results of data mining. The process of data mining usually starts by an analytical question. An answer to such question can be an analytical report. For the given question there is a well defined structure of the corresponding analytical report. The structure is described by chapters and sections and must fit to the solved problem and to the accepted domain knowledge. The core of each section are patterns found in analyzed data [3, 5].

* The work described here has been supported by the grant 201/08/0802 of the Czech Science Foundation

These patterns can be understood as formulas of suitable observational calculi. Thus we can try to apply logical properties of observational calculi (namely deduction rules) to find "*logical core*" of a given analytical report. Logical core of the given report is a minimal set of observational formulas from which it is possible to deduce all observational formulas used in the given report [3]. We can also use the "*logical core*" to index the content of the analytical report instead of usual key words. This way the content of the report is described in a very precise way. The formulas of observational calculi can be used instead of key words when retrieving analytical reports. The analytical reports can be this way disseminated through Semantic Web. Some preliminary considerations on this topic inspired by [3] are e.g. in [2].

Availability of analytical reports through Semantic Web opens a possibility of (semi)automated creation of "global analytical reports". Each global analytical report uses several "local" reports found using Semantic Web. This possibility led to start of the SEWEBAR project [5]. Its goal is to study possibilities of automatic creation and dissemination both of "local" and "global" analytical reports. Again, logic of observational calculi is important in this activity.

However the first considerations about global analytical reports show that the observational calculi are not sufficient to solve all logical problems related to the global analytical reports. Let us suppose that we have two local analytical reports, the first one saying that $\mathcal{R}$ is the strongest rule and the second one saying that $\mathcal{R}$ is the weakest rule related to the given problem. To summarize such two reports we need some information concerning the roles of the rule $\mathcal{R}$ in both reports. This fact leads to the definition and study of skeletons of analytical reports. The working definition of the skeleton of the analytical report says that the skeleton is related to the type of analytical question the report answers. Moreover the skeleton consists the structure of the report (see above), of observational statements and of specific statements (e.g. "the strongest rule" and "the weakest rule") making possible to use the report to answer global analytical questions. The properties of these specific statements are determined both by the type of the local analytical report the report answers and by types of the global analytical questions we suppose to answer.

References

1. Hájek, P., Havránek, T.: Mechanizing Hypothesis Formation(Mathematical Foundations for a General Theory), Springer–Verlag 1978.
2. Lín V., Rauch J., Svátek, V.: Contend-based Retrieval of Analytic Reports.In: Schroeder M, Wagner G(ed.) Rule Markup Languages for Business Rules on the Semantic Web. Sardinia: ISWC, 2002, pp. 219-224.
3. Rauch, J.: Logical Calculi for Knowledge Discovery in Databases. In Principles of Data Mining andKnowledge Discovery, (J. Komorowski, J. Zytkow, eds.),Springer-Verlag, 47-57, 1997.
4. Rauch, J.: (2005) Logic of Association Rules. Applied Intelligence 22, 9-28.
5. J. Rauch J, M. Šimůnek: Semantic Web Presentation of Analytical Reports from Data Mining - Preliminary Considerations. In: Proceedings of WEB INTELLIGENCE. IEEE, 2007, pp. 3-7.