March 16, 2017

INTERNATIONAL WORKSHOP
"Formal Concept Analysis for Knowledge Discovery"

Program

10:00 – 11:00	*KEYNOTE TALK*
	Ivo Duentsch *Knowledge structures and skill assignments: Structural tools for diagnostic assessment*
11:00 – 11:15	*COFFEE BREAK*
11:15 – 13:15	*REGULAR TALKS*
	Dmitry Palchunov, Gulnara Yakhyaeva *The development of ontological model describing the behavior of mobile network subscribers*
	Tatiana Makhalova, Sergei O. Kuznetsov *On Overfitting of Classifiers Making a Lattice*
	Alexey Masyutin *Query-based classification with interval pattern structures: application to credit scoring*
	Mikhail Bogatyrev, Anastassia Intyakova, Kirill Samodurov *Discovering Formal Contexts Generated from Conceptual Graphs*
13:15 – 14:00	*LUNCH BREAK*
14:00 – 16:00	*REGULAR TALKS*
	Nikita Borodulin, Nikolay Shilov *Designing Logic Puzzle Solver*
	Dmitry V. Vinogradov *Accidental formal concepts in the presence of counterexample*
	Xenia Naidenova, Alexey Buzmakov, Vladimir Parkhomenko, Alexander Schukin *Notes on Relation Between Symbolic Classifiers*
	Sergei O. Kuznetsov, Nurtas Makhazhanov, Maxim Ushakov *On Neural Network Architecture Based on Concept Lattices*
16:00 – 16:15	*COFFEE BREAK*
16:15 – 17:45	*REGULAR TALKS*
	Xenia Naidenova, Sergey Curbatov, Vladimir Parkhomenko *Neural-Network Like Logical-Combinatorial Structure Of Data And The Possibilities Of Its Application For Constructing Concept Lattices*
	Dmitry I. Ignatov *On the Family of Concept Forming Operators in Polyadic FCA*
	Natalia Korepanova, Sergei O. Kuznetsov *Patterrn Structures for Individualized Treatment*
17:45 – 18:15	*FINAL DISCUSSION*

Knowledge structures and skill assignments: Structural tools for diagnostic assessment

Ivo Duentsch
Department of Computer Science Brock University St. Catharines, Ontario, L2S 3A1, Canada

The overall aim of any assessment is to determine the skills a subject masters, and the more skills one has to solve the problems, the more problems one will solve; this simple idea is at the basis of all common theories of ability tests. Apart from this idea, the theories diverge in their assumptions about "latent traits" or "ability parameters" or other constructs that "represent" the data as good as possible. A direct connection between theory and data is the construction of a proper operationalization of the field of interest: The researcher formulates the skills a subject needs in order to be able to solve a problem. Alternatively, given a set S of skills, an expert indicates for each subset X of S which problems can be solved with the skills in X. In a second step, a scaling mechanism is chosen which interprets the test results in terms of the skills and provides useful feedback to subject and researcher about the subject's skill state. Just as knowledge assessment by modern scaling theory such as knowledge structures can be regarded as a qualitative way of measurement, the aim is to provide a qualitative description of a subject's skill state. This is based on the conviction that a sensible diagnostic cannot be a point estimate of the knowledge of an individual such as a test score, but that, in most cases, only a range of skills can be given, which are more or less mastered by an individual.
In this talk I will present a test theory that can be built using skill and problem assignments, and how empirical data can be used to explore an underlying skill or problem function.

The development of ontological model describing the behavior of mobile network subscribers

Dmitry Palchunov, Gulnara Yakhyaeva

The paper is devoted to the development of methods of ontology supported knowledge discovery in the field of mobile network subscribers. We develop the ontological model of the domain of mobile networks. This ontological model is intended to describe the behavior of mobile network subscribers and identify the needs of different user groups. The ontological model is based on the four-level model of knowledge representation. In this paper, special attention is given to the third and fourth level: the third level represents the set of cases from the domain, and the fourth level represents fuzzy, probabilistic and estimated knowledge. For the analysis of domain cases and for the generation of fuzzy domain knowledge we use Formal Concept Analysis. Fuzzy and estimated knowledge is represented by the fuzzy theory of the domain. We investigate the class of fuzzy models which satisfy the fuzzy theory of the domain. To describe the set of cases from the domain we use formal contexts; objects of these formal contexts are finite sets of subscribers. With the help of the analysis of data on mobile network subscribers we extract high-level characteristics of subscribers. On the base of this knowledge the future behavior of mobile network subscribers may be predicted.

Query-based classification with interval pattern structures: application to credit scoring

Alexey Masyutin

Query-based classification is techniques which produces classification rules for an individual test object, having no prior classification rules set. The algorithm relies on pattern structures definition and derivation operator. We applied query-based classification approach to Kaggle open credit scoring data to assess probability the borrower is going to be insolvent. We argue that query-based classification allows one to achieve higher classification accuracy as compared to applying classical banking models and still to retain interpretability of model results, whereas black-box methods grant better accuracy but diminish interpretability.

The criteria of ontology quality analysis based on concept lattice

Bato Merdygeev, Sesegma Dambaeva
East Siberia State University of Technology and Management

Paper presents an approach to analysis of domain ontology quality and criteria of analysis method based on approach. The approach allows evaluating the completeness of the ontology relations. We use concept lattice of Formal Concept Analysis for evaluation of ontology relation structure.

Discovering Formal Contexts Generated from Conceptual Graphs

Mikhail Bogatyrev, Anastassia Intyakova, Kirill Samodurov
Tula State University, Tula, Russia okkambo@mail.ru

Building conceptual lattices from conceptual graphs looks as natural way in Formal Concept Analysis but still is not discovered at length. If conceptual graphs are acquired from natural language texts then they contain specific material for building various contexts. Problem–oriented contexts, contexts with non-ordered set of attributes (order-free context) and invariant contexts are investigated in the paper. The problem of context decomposition is discussed and one way to decompose contexts based on group theory methods is discussed. Discovered results have been used in the learning of bacteria biotopes on the annotated textual corpus.

Designing Logic Puzzle Solver

Nikita Borodulin, Nikolay Shilov
Innopolis University

Paper presents requirement analysis and functionality, design and implementation of a mobile application to help to solve logical puzzles for non-experience user. The major problem was interface design (since we can not assume any experience in logic puzzles). For correct specification of the interface we use Formal Concept Analysis. SAT-solver is used as the logic engine.

Accidental formal concepts in the presence of counterexamples

Dmitry V. Vinogradov
Federal Research Center for Computer Science and Control RAS, Moscow 119333, Russia

Accidental formal concept corresponds to a subset of attributes that accidentally appear in several objects (the parents of the concept) if every such an object is a parent of a different "real" formal concept. There are two standard techniques to forbid such concepts: counterexamples to exclude involved concepts and the lower bound on the number of parents. Both are insufficient. We define the random formal context with attributes (different from elements of "real" formal concepts) generated by independent Bernoulli variables. The main result has asymptotic form: If the the number n of the random attributes tends to infinity, the probability of success equals to $\sqrt{a/n}$, and there are $m = \sqrt{b/n}$ counterexamples, then the probability of appearance of an accidental formal concept with 2 parents avoiding these counterexamples is $1 - e^{-a} - a \cdot e^{-a} \cdot [1 - e^{-b \cdot \sqrt{a}}]$ at limit.

Notes on Relation Between Symbolic Classifiers

Xenia Naidenova[1], *Alexey Buzmakov*[2], *Vladimir Parkhomenko*[3], *Alexander Schukin*[3]
[1] *Military Medical Academy, Saint-Petersburg, Russia*
[2] *NRU Higher School of Economics, Perm, Russia*
[3] *Peter the Great St. Petersburg Polytechnic University, Saint-Petersburg, Russia*

Symbolic classifiers allow for solving classification task and simultaneously for understanding the reason of the classifier decision. They were studied by a large number of researchers and known under a number of names including Tests, JSM-hypothesis, Version Spaces, Emerging Patterns, Proper Predictors of a target class, etc. Classifiers with restrictions on counter-examples from these theories are considered and discussed in terms of the language of Pattern Structures. We show how classifiers from different theories are related including the equivalence between Good Maximally Redundant Tests (GMRTs) and minimal JSM- hypotheses and between minimal representations of Version Spaces and Irredundant Tests, which are included in GMRTs.

On Neural Network Architecture Based on Concept Lattices

Sergei O. Kuznetsov, Nurtas Makhazhanov, Maxim Ushakov
Department of Data Analysis and Artificial Intelligence, Faculty of Computer Science

Selecting an appropriate neural network architecture is a crucial problem when finding a solution based on a neural network. If the number of neurons in the network is too high, then it is likely to overfit. Neural networks also suffer the problem of poor interpretability of learning results. In this paper an approach to building a neural network based on a concept lattice is proposed in attempt to overcome the mentioned difficulties.

Neural-Network Like Logical-Combinatorial Structure Of Data And The Possibilities Of Its Application For Constructing Concept Lattices

Xenia Naidenova[1], *Sergey Curbatov*[2], *Vladimir Parkhomenko*[3]
[1] *Military Medical Academy, Saint-Petersburg, Russia*
[2] *NICEVT, Research and Development Center, Moscow, Russia*
[3] *Peter the Great St. Petersburg Polytechnic University, Saint-Petersburg, Russia*

A generalization of algorithm is proposed for implementing the wellknown effective inductive method of constructing sets of cardinality (*q+1*) (*(q+1)*- sets) from their subsets of cardinality *q* (*(q)*-sets). A new neural network-like combinatorial data-knowledge structure supporting this algorithm is advanced. This structure can drastically increase the efficiency of inferring functional and implicative dependencies as like as association rules from a given dataset. Various modes of functioning the network for constructing concept lattices are considered. Some algorithms for constructing concept lattice, inferring good maximally redundant and irredundant classification tests are given with using a generalization process based on Galois's connections and a direct and backward wave of network activity propagation. A method of initial approximation of the network is given for mining classification tests. The level-wise method of *(q+1)*-sets' construction is also used for association rule mining. The same principle underlies the algorithm Titanic for generating key patterns and the algorithm TANE for discovering functional dependencies. In all enumerated problems, the same algorithm deals with different sets of elements (items (values of attributes), itemsets, attributes, object descriptions, indices of itemsets) and checks the different properties of generated subsets. These properties can be, for example: "to be a frequent (large) itemset", "to be a key pattern", "to be a test for a given class of examples", "to be a good test for a given class of examples", and some others. If a constructed subset does not possess a required property, then it is deleted from consideration. This deletion reduces drastically the number of subsets to be built at all greater levels. Generally, this algorithm solves the task of inferring all maximal subsets of a set *S* (i.e., such subsets that cannot be extended) possessing a given PROPERTY. The set *S* can be interpreted depending on the context of a considered problem. The neural-network like logical-combinatorial structure can be also used in the all enumerated problems.
The possibilities of application of proposed network in text mining are also considered, for example, for extracting associative dependencies between words, extracting topic of text and contexts of topic.

On the Family of Concept Forming Operators in Polyadic FCA

Dmitry Ignatov
National Research University Higher School of Economics, Moscow, Russia

Triadic Formal Concept Analysis (3FCA) was introduced by Lehman and Wille almost two decades ago. And many researchers work in Data Mining and Formal Concept Analysis using the notions of closed sets, Galois and closure operators, closure systems, but up-to-date even though that different

researchers actively work on mining triadic and *n*-ary relations, a proper closure operator for enumeration of triconcepts, i.e. maximal triadic cliques of tripartite hypergraphs, was not introduced. In this talk we show that the previously introduced operators for obtaining triconcepts is not always consistent and provide the reader with a definition of valid closure operator and associated set system. Moreover, we study the difficulties of related problems from order-theoretic and combinatorial point view as well as provide the reader with justifications of the complexity classes of these problems.

On Overfitting of Classifiers Making a Lattice

Tatiana Makhalova, Sergei O. Kuznetsov
National Research University Higher School of Economics, Moscow, Russia

Obtaining accurate bounds of the probability of overfitting is a fundamental question in statistical learning theory. In this paper we propose exact combinatorial bounds for the family of classifiers making a lattice. We use some lattice properties to derive the probability of overfitting for a set of classifiers represented by concepts. The extent of a concept, in turn, matches the set of objects correctly classified by the corresponding classifier. Conducted experiments illustrate that the proposed bounds are consistent with the Monte Carlo bounds.

Pattern Structures for Risk Group Identification

Natalia Korepanova, Sergei O. Kuznetsov
National Research University Higher School of Economics, Moscow, Russia

The idea of individualized treatment is not novel but in recent times it became very popular. Risk group identification is one of the first and obvious ways to tailor therapy to patient characteristics. The most common approach used to discover groups of different risk is the Cox proportional hazard model. However, all regression models, including the Cox model, are not clearly interpretable for clinicians. As pattern structures can be applied to any type of object descriptions on which a partial order can be set, potentially they are a great tool to obtain risk group descriptions in a convenient form. This talk is devoted to a use case of pattern structures for risk group identification for acute lymphoblastic leukemia.