Control and Cybernetics

vol. **40** (2011) No. 2

Preface to the special issue

The attribute of "intelligence" has become an indispensable part of a number of research areas of applied computer science. We hear about "artificial intelligence", "computational intelligence", "business intelligence", "intelligent systems" etc.

What distinguishes the "intelligent methods" from the general computational methods? The usual way of writing any application is to start with determination of an algorithm, which governs the process of interest in the application domain. In this way, systems for bridge design or bank accounting are designed. But there exist a number of areas, where we actually do not know the exact underlying algorithms and we are only capable of approximating them based on past experience, just mimicking the human approach to handling such situations. Under some circumstances, building parametric models and estimating their parameters from available data is essentially the area where we can speak about intelligent systems. What distinguishes such cases from pure statistical analysis is that while models are being applied, the intelligent systems still learn (from failures and successes) and hopefully improve their performance. Of course, once a given modelling heuristic proves successful, it becomes subject to a more profound mathematical and analytical investigation, and eventually a solution thus obtained becomes a part of fundamental statistics. But, in spite of the fact that methods once considered a part of artificial intelligence cease to be that part, the area is a source of increasing excitement as the frontiers of its application seem to widen all the time as the contributions of this volume show.

M. Dramiński and co-authors present a study of a novel feature selection method. It proved empirically very successful in analyzing, e.g., micro-arrays, in which one faces the challenge of a vast number of attributes and a handful of cases. Traditional statistical and machine learning methods then fail, due to overfitting, while the new method successfully discovered known intrinsic relationships and the new ones that were confirmed empirically later. In the paper included in this volume the authors demonstrate that this new method of feature selection and interdependency discovery is practically unbiased.

A. Skowron and P. Wasilewski investigate the problem of mismatch between the level of detail of the empirical data and the abstractness of concepts, in which humans formulate their knowledge. They develop a new methodology of knowledge acquisition from data based on the typical way in which humans look at the surrounding world: at various levels of details. They argue that detailed rough data may be difficult for humans to draw directly general conclusions

and to incorporate human knowledge. Moreover, a system that can stepwise construct more abstract concepts allows for better reasoning about data. In such systems, humans can input their knowledge at different levels of abstraction, improving knowledge acquisition efficiency.

S. Chojnacki and M. Kłopotek point at a serious weakness of many studies within the domain of artificial intelligence: the lack of empirical data to investigate all the subtle parameters of a model. This may be a nuisance, particularly in case of performance analysis of a machine learning algorithm, especially in case of recommender systems. As a remedy they propose a method for generating synthetic data, resembling the real world data with diverse degrees of variation, so that at least the performance of a developed model may be checked in terms of sensitivity to changing data properties.

Another problem, faced by the artificial intelligence systems is the abundance of data, e.g. in case of analysis of the Web graphs. S. Wierzchoń and co-authors are interested particularly in PageRank computation, used nowadays in variants not only for the Web, but also for various types of social networks. They demonstrate that large computational efficiency improvements may be gained not by analyzing an abstract, theoretical model of the Web, but rather by looking at properties of its actual instance.

J. Józefowska and co-authors handle the issue of data being collected at different levels of detail, bringing a serious problem for traditional model acquisition systems as part of data constitute noisy information from their point of view. A novel approach, demonstrated for the naïve Bayes classifier, allows for the use of a hierarchy of values for each attribute and biasing towards the attributes measured more precisely.

Another case study, dealing with exploitation of noisy data is presented by R. Lewis and co-authors. The authors show how to construct and use deterministic finite automata in a very sensitive medical domain of epileptogenesis, incorporating into the state description of EEG the attributes derived from spectral analysis of the signal. It turns out that classifiers, which may be derived in this way are quite robust even for relatively noisy signals.

M. Jarocki and co-authors develop a method to cover the world of Web services with a uniform semantics in order to build complex flows of service calls. They propose to transfer concepts from the object-oriented database area to the domain of complex knowledge management. Complex services are built from simple ones via database queries.

The most popular methods of building natural language processing applications require semantically annotated text corpora. A. Mykowiecka and M. Marciniak handle the issue of preparation of a richly annotated corpus for a particular application area (medical documents) using software designed for general texts and a specialized information extraction system.

The papers by A. Savary and J. Piskorski and by M. Marcińczuk and M. Piasecki concentrate on the automation of semantic annotation process for a specific class of lexical items: the named entities. The first paper handles the general

issue of the named entities, while the other one deals with the domain of stock exchange reports. In both papers regular expressions and gazetteers have been used for the task. The second paper presents also the experiments with rescoring the Hidden Markov Model results.

B. Broda and co-authors concentrate on the issue of word-sense disambiguation in the process of semantic text annotation. In order to find a path between the poorly performing fully automated approaches to the task and the reliable but tedious manual annotation, they propose to train classifiers on the basis of clusters of text chunks around words to be disambiguated for which the core word is recognized and tagged with an appropriate sense manually.

B. Siemiątkowska and co-authors present an approach to acquisition of semantically labelled data in the domain of mobile robots, navigating in the indoor and outdoor environments. A 3D laser range finder utilizes a set of proper primitive object models to identify complex objects and creates a map of environment allocating semantic labels to physical objects encountered.

* * *

We hope that the Reader will find it inspiring how the same problems, such as construction of conceptual hierarchies, complex object construction and identification, semantic annotation and feature selection occur over and over again in different application areas of artificial intelligence and how differently they are solved.

The Guest Editors