NIKE (Neural explicit and Implicit Knowledge Inference): A Hybrid Intelligent System Applied in Toxicity Data Mining
Ciprian-Daniel NEAGU

I. THE METHOD: A HYBRID INTELLIGENT SYSTEM, IN AN ORIGINAL APPROACH

Integration of various intelligent approaches, known as Hybrid Intelligent Systems (HIS), is more common in theory and experimentation and less in applications. In recent years we face an increasing activity in the field. However, as systems grow more complex and available data increase, the applied technologies have to focus more to the homogeneous implementations.

Hybrid neural/neuro-fuzzy systems have drawn increasing research interest as HIS. The reason for studying hybrid neural systems is based on the processing abilities shown by subsymbolic knowledge representation, particularly the neuro-fuzzy paradigm, as well as on the advantages of the symbolic knowledge-based systems. From the point of view of cognitive science, the purely neural representation offers the advantage of homogeneity, as well as the possibility of performing distributed tasks, and working with incomplete and noisy data. From the point of view of knowledge-based systems, symbolic representations have advantages of human interpretation, explicit control and knowledge abstraction.

The two approaches and applications are described as neuro-fuzzy knowledge-based system models. The knowledge extraction from trained neural networks, and the integration of implicit (connectionist) knowledge with symbolic expert systems are the main tasks. In this hybrid system, connectionist tools can be interpreted as hardware, and fuzzy logic as software implementation of human reasoning. In that way, the modular structure of connectionist implementations of explicit and implicit knowledge can be interpreted as a homogenous system combining inductive and deductive learning and reasoning, based on neurons with fuzzy processing abilities.

The concept of implicit knowledge, specific architectures to implement it, and some methods to extract fuzzy rules from implicit knowledge structure are listed. The concept of explicit knowledge and methods to implement a given fuzzy rule set as equivalent connectionist structures are proposed. Modular networks to combine implicit and explicit knowledge are developed.

II. THE APPLICATION FIELD: TOXICITY PREDICTION

The problem of toxicity prediction represents the application of our approach. This is a particularly difficult problem, especially considering the huge number of compounds to be studied and the incomplete knowledge about the causes of toxicity. Thus, a large number of chemical descriptors are of potential interest. A structure-activity relationship (SAR) relates features of physico-chemical structure to a property, effect, or biological activity associated with that chemical. In so doing, there can be both qualitative and quantitative considerations. Quantitative structure-activity relationships (QSARs) can be developed using continuous data, through a regression process. QSAR method has been applied to many drug and chemical design, as well as to the prediction of specific toxicological endpoints. Sometimes, the QSAR models exhibit instability when trained with noisy data. In addition, traditional regression techniques often require subjective decisions to be made, as to the likely functional relationships between structure and activity.

III. THE TOOL: NIKE

Connectionist models are powerful tools to process knowledge, so we use this approach to build models, for multidimensional data scaling, where discovering explicit rules does not seem either natural, or direct. NIKE (Neural explicit & Implicit Knowledge Inference system) is an original hybrid intelligent system, developed for data mining, prediction and modeling, based on modular neural and neuro-fuzzy networks. A particular attention is paid to acquire new rules from trained neural and neuro-fuzzy modules, and to describe the specific module as equivalent to a fuzzy rule-based system. The explicit knowledge module of the hybrid system is implemented as a special connectionist structure using hybrid fuzzy neural networks. This kind of implementation is proposed to adjust the performances of implicit knowledge modules, especially in the known cases of well adapted QSARs, which are inserted in the HIS architecture as equivalent neuro-fuzzy mapped models.

IV. THE APPLICATIONS AND RESULTS

This section presents the results of applying the hybrid intelligent models for two different problems: phenols toxicity/mechanism of action (MOA) modeling and the pesticides toxicity prediction.

The hydroxy-substituted aromatic compounds (phenols) form a large, structurally diverse group. These are interesting from a toxicological point of view, since the phenols are widely used organic compounds. Thus, the constant interest in QSARs for phenols due to their ubiquitous nature and the various toxicities they may have. For acute toxicity, the elucidation of MOA is required: in the toxicity problem, successful QSAR analysis depends on the identification of MOA of the compounds.

Pesticides contain impurities and in several cases are mixtures of chemicals; once introduced into the environment they generate transformation products, in some cases tens of them. The evaluation of all these compounds, since each experimental assay represent a cost, is a tedious and costly task. Therefore, the availability of HIS, particularly the NIKE approach, to evaluate the effects of pesticides and related compounds represents a useful way to increase the possibilities of global assessment of pesticides and related compounds in a fast, easy, and convenient way.

V. CONCLUSIONS

The advantages of developing hybrid intelligent models to combine implicit and explicit knowledge in the architecture of the neural and neuro-fuzzy rule-based system are identified. Considerations about how to improve the structure and the behavior of the system NIKE are also proposed, especially in the target of extracting more compact and comprehensible set of rules about the problems described. The system designed and developed by the author (under Matlab 6) exhibits practical and effective solutions for the two difficult tasks. The trends of parallel combination of many classifiers deviates from, or even follows an opposite philosophy of, traditional selection approach, in which one evaluates the available systems against a representative sample and choose the best combination of the available methods.