

Preface

This special issue of *Fundamenta Informaticae* contains extended versions of 10 articles selected from 64 papers accepted for presentation at the Twelfth International Symposium on Methodologies for Intelligent Systems (ISMIS) in Charlotte, North Carolina, USA, October 11-14, 2000. The ISMIS symposia cover a broad scope of intelligent systems such as evolutionary computation, intelligent information retrieval, knowledge representation and integration, learning and knowledge discovery, logic for AI, and soft computing.

The chosen articles fall into the category of intelligent information systems but their contributions are made from quite diverse perspectives: theory refinement, machine learning, knowledge discovery and data mining, multidimensional data analysis, integration of classifiers, information retrieval, heterogeneous database querying, and multimedia digital libraries.

The first article by S. Kramer, G. Widmer, B. Pfahringer, and M. De Groeve is devoted to the problem of learning to predict ordinal classes using classification and regression trees. They start with S-CART, a tree induction algorithm, and study various ways of transforming it into a learner for ordinal classification tasks. These algorithm variants are compared on a number of benchmark data sets to verify the relative strengths and weaknesses of the strategies and to study the trade-off between optimal categorical classification accuracy and minimum distance-based error.

In the second article, F. Esposito, N. Fanizzi, S. Ferilli and G. Semeraro described a framework for theory refinement operators pursuing the efficiency and effectiveness of learning regarded as a search process for theory refinement. A refinement operator satisfying these requirements is formally defined as ideal. Past results have demonstrated the impossibility of specifying ideal operators in search spaces where standard generalization models, like logical implication or the \subseteq -subsumption, are adopted. By assuming the object identity bias over a space defined by a clausal language ordered by logical implication, a novel generalization model, named OI-implication, is derived. The authors prove that ideal operators can be defined for the resulting search space.

In the third article, T. Elomaa and J. Rousu probe the inherent complexity of the multisplitting problem, a potential time-consumption bottleneck in machine learning and data mining. They utilize results obtained for similar problems in computational geometry and string matching to the multisplitting task. They show by counterexamples that the widely used evaluation functions Training Set Error and Average Class Entropy do not satisfy the kind of monotonicity that facilitates subquadratic-time optimization. They also show that the Training Set Error function can be decomposed into monotonic subproblems, one

per class, which explains its linear time optimization. Finally, they review recently developed techniques for speeding up optimal multisplitting.

The fourth article by M.S. Hacid and F. Toumani considers how constraint-based technology can be used to query semistructured data. They present a formalism based on feature logics for querying semistructured data. The formalism is a hybrid one in the sense that it combines clauses with path constraints. The resulting language has a clear declarative and operational semantics.

In the fifth article, S. Park and W. W. Chu present techniques for discovering and matching rules with elastic patterns. Elastic patterns are useful for discovering rules from data sequences with different sampling rates. For fast discovery of rules whose heads and bodies are elastic patterns, they construct a trimmed suffix tree from succinct forms of data sequences and keep the tree as a compact representation of rules. The trimmed suffix tree is also used as an index structure for finding rules matched to a target head sequence. When matched rules cannot be found, the concept of rule relaxation is introduced. Using a cluster hierarchy and relaxation error as a new distance function, they find the least relaxed rules that provide the most specific information on a target head sequence.

The sixth article by S. Puuronen and A. Tsymbal addressed the problem of feature-space heterogeneity in multidimensional data. They introduce a technique that provides a strategic splitting of the instance space to identify the best subset of features for each instance to be classified. Their technique applies the wrapper approach where a classification algorithm is used as an evaluation function to differentiate between different feature subsets. In order to make the feature selection local, they apply the dynamic integration of classifiers to determine what classifier and which feature subset should be used for each new instance. In order to restrict the number of feature combinations being analyzed, they propose the use of decision trees. For each new instance, they only consider those feature combinations that include the features present in the path taken by the new instance in the decision tree built on the whole feature set. In their experiments, they use the C4.5 algorithm as the learning algorithm for base classifiers and for the decision trees that guide the local feature selection.

In the seventh article, M. Kim, J. S. Deogun, and V. V. Raghavan address problems in Rule Based Information Retrieval by Computer (RUBRIC). RUBRIC involves the use of production rules to capture user-query concepts. A set of related production rules is represented as an AND/OR tree, or alternatively by a disjunction of Minimal Term Sets (MTSs). The retrieval output is determined by the evaluation of the weighted Boolean expressions of the AND/OR tree, and processing efficiency can be enhanced by employing MTSs. However, since the weighted Boolean expression ignores the term-term association unless it is explicitly represented in the tree, the terminological gap between users' queries and their information needs may still remain. To solve this problem, they adopt the generalized vector space model (GVSM) and the p-norm based extended Boolean model.

In the eighth article, C. Fernandes and L. Henschen propose a system whereby subtle semantic ambiguity found in queries of distributed heterogeneous database systems can be resolved by considering the user's intentions. Through the use of domain-specific knowledge embedded within a mediator-based architecture, subtleties in meaning can be explicitly modeled. Through the use of dynamic profiles and active dialogue, their system can discover user intent, providing more satisfying query answers.

The ninth article by E. Bertino, B. Catania, and G. P. Zarri propose a two-steps annotation approach by which conceptual annotations, represented in Narrative Knowledge Representation Language (NKRL), are associated with multimedia documents and used during retrieval operations. Metadata represent the vehicle by which digital documents can be efficiently indexed and retrieved in multimedia digital libraries. A relevant metadata function is created by superimposing some sort of conceptual orga-

nization over the unstructured information space proper to these digital repositories, in order to facilitate the intelligent retrieval of the original documents. The authors present how documents and metadata can be stored and managed on persistent storage.

The last article by A. Wiczorkowska describes the application of wavelet transform to the automatic classification of musical instrument sounds. Eleven instruments are described by a feature vector consisting of 152 attributes (parameters). Two classifiers are constructed using C4.5 algorithm and DataLogic/R. Finally, the correctness of the two classifiers using cross validation method is discussed .

E. El-Kwae, Guest Editor

University of North Carolina at Charlotte,
Department of Computer Science 9201
University City Boulevard,
Charlotte, NC 28223

Z. Raś, Guest Editor

University of North Carolina at Charlotte,
Department of Computer Science 9201
University City Boulevard,
Charlotte, NC 28223

and

Polish Academy of Sciences,
Institute of Computer Science
01-237 Warsaw, Poland