Foreword and Editorial

International Journal of Database Theory and Application

We are very happy to publish this issue of an International Journal of Database Theory and Application by Science and Engineering Research Support Society.

This issue contains 25 articles. Achieving such a high quality of papers would have been impossible without the huge work that was undertaken by the Editorial Board members and External Reviewers. We take this opportunity to thank them for their great support and cooperation.

The article entitled “Improvements in Data Mining Association Rules Algorithm” puts forward an improved Apriori algorithm based on SQL, increase degree by calculation method, for pruning association rules and is independent of item sets.

The paper “A Spatio-Temporal Simulation Model for Incremental Clustering in Massive Moving Objects Data Set” introduced a simulation framework applicable to diverse movement data. It can be used in studies of individual movement behaviors, including the behaviors of animals. It can also be used to analyze the movements of multiple entities for the purposes of city planning, traffic management, logistics, the optimization of layouts of public venues and shopping areas, the allocation of facilities or advertisements, etc. Finally, the development of a large-scale simulation model, that more closely reflects macro traffic processes, enables the benefits of these processes to be more clearly identified. The real-world applicability of such processes depends upon the quality of the models developed to analyze them. In addition, the iterative process allows the users to implement optimization solutions, guaranteed to be feasible for actual execution. This allows the user to study actual traffic processes in terms of both the simulation and the prediction models.

Paper “A Method for Building Naxi Language Dependency Treebank Based on Chinese-Naxi Language Relationship Alignment” proposed a method for building Naxi Language Dependency Treebank based on Chinese-Naxi Language relationship alignment. Firstly, the corresponding words of Chinese-Naxi sentence pairs are aligned; then, the dependency grammar on Chinese sentences; Finally, some characteristics and rules of Naxi Language in itself being considered, the generated Chinese Dependency Tree is mapped to Naxi Sentence by using Chinese-Naxi Languages relationship alignment, as a result, Naxi Dependency Parsing Tree is generated. Experimental results show that: This approach can simplify the process of manual collection and annotation of Naxi Treebank, and save manpower and time to build the dependency treebank of Naxi Language.

In the paper “Efficient Information Extraction Based on Signature Index” authors propose a novel framework iExtractor that leverages Information Retrieval (IR) indexes to speed-up IE processes. They index text blocks with their signatures (presented as bit-strings) and propose efficient IE algorithms based on the signature index. Hence, iExtractor can validate query pattern in signature index without original text. The framework also supports on-line extraction through a general and flexible pattern extraction language. The extensive experimental results on diverse real datasets show that our approach delivers stable efficiency and has outperforms baselines in terms of extraction accuracy.

In the paper “Named Entity Recognition by Using Maximum Entropy”,...
Authors of paper “An Adaptable University Human Resource Data Management”, present a methodology called maximum entropy to retrieve the entity sets from the database. The machine is trained in such a way that it will retrieve the words which has the maximum entropy amongst all and has proved to be fastest method to extract and classify the entity sets from the database. The advantages of proposed method include sequence tagging which means this method has increased the freedom of choosing features to represent observations.

In the paper about “The Processing Technology in Mobile Database Transaction System” by introducing a kind of optimistic concurrency control method combined with two locking protocol of mobile transaction processing mechanism, the paper solved the problems of interactive long transactions, and provides the mobile host any mobility and the consistency of database system support.

The paper “Generating ER Diagrams from Requirement Specifications Based on Natural Language Processing” deals with the problem of extracting ER elements from natural language specifications using Natural Language Processing (NLP). The approach provides the opportunity of using natural language documents as a source of knowledge for generating ER data model. The structural approach is used to parse specification syntactically based a predefined set of on heuristics rules. Extracted words with its Part Of Speech (POS) mapped into entities, attributes and relationships, which are the basic elements of ER diagrams.

In the paper “Large-Scale Data Classification Method Based on Machine Learning Model” a machine learning model based on the SMO algorithm and RBF kernel function of the SVM is proposed to realize a classification method in this paper. This method transforms the nonlinear classification problem into linear classification problem by improving the data dimension. It can better solve the problems of the minimum error in the training set and the larger error in the test set in the traditional algorithm. Application of UCI classification experiment shows that the proposed method takes on the better convergence, faster training speed and higher classification accuracy.

Paper “Spatial Approximate Keyword Query Processing in Cloud Computing System” proposes spatial approximate keyword query algorithms for cloud systems. Existing work targets on single server solutions, and an exact algorithm is given in memory while another approximate algorithm is given for disk resident datasets. However, a single server fails to provide reasonable throughput due to the limited CPU time and disk bandwidth. Facing the above challenges, this paper gives a two-layered index consisting of global index and local index, which works in a shared nothing cluster for larger query throughput. This paper designs a novel external memory index as local index, which returns exact answer within disks efficiently. It is equipped with keyword set signature and multiple optimizing strategies to reduce I/O cost. The global index partitions the entire spatial space, and each computing node in system maintains a partition. A global index selection algorithm is given. This paper also provides spatial approximate keyword query algorithms based edit distance, including range and the nearest neighbor spatial conditions. An experiment in a shared nothing cluster illustrates the efficiency and effectiveness of our proposed index and query algorithms.

The paper “A Comparative Study on Filtered, Vertical and Horizontal Inheritance Mapping in Database” discusses and elaborates one of the most common mapping algorithms about mapping of inheritance structure to relational database. The inheritance mapping in database contains three various approaches: filtered, vertical and horizontal inheritance mapping. This paper review the fundamental measures used to comparative
among three inheritance mapping in database through determine the strengths and weaknesses for each on: ease of access to the data, the speed of data access, Ad hoc reporting, Ease of implementation, Coupling, Support for polymorphism to the user and the development of modern commercial applications to protect the time and effort taken advantage of them while working. By tow method Based on the algorithms and rules, Object Relational Mapping (ORM) tool.

Paper “Content-Based Social Network User Interest Tag Extraction” adjusts the traditional processes of Chinese word segmentation and tag extraction according to the unique feature of microblog text. Aiming to establish microblog user interest model, this paper combines clustering and classification algorithm to extract user interest tags. Experimenting upon Sina Weibo text dataset, the conclusion is reached that the method proposed in this paper is more effective and accurate to discover user features and the tags extracted are more in line with the user interest.

In “MRG-DBSCAN: An Improved DBSCAN Clustering Method Based on Map Reduce and Grid”, the original DBSCAN algorithm was improved, and the G-DBSCAN algorithm is proposed. G-DBSCAN algorithm reduces the number of query object as a starting point. Put the data into the grid, with the center point of the data in the grid to replace all the grid points as the algorithm input. The query object will be drastically reduced, thus improving the efficiency of the algorithm, reduces the memory footprint. In order to make the G-DBSCAN algorithm can adapt to large data processing, we will parallelize the G-DBSCAN algorithm, and combining it with Map Reduce framework. The results prove that G-DBSCAN and MRG-DBSCAN algorithm are feasible and effective.

Authors of the paper entitled “How Student's Attitude Influences on Learning Achievement? --An Analysis of Attitude-Representing Words Appearing in Looking-Back Evaluation Texts--”, investigate correlations of students’ achievements and their learning attitudes by analyzing the usage of words of students which appear in the answer-texts of their looking-back self/class-evaluation questionnaire. They classify the words into 4 types based on the students’ attitudes to learning, represented by the words. As a result of the study, they found that the students in the middle-achievement group have differences in their word-usages, whereas the high- and low-achievement students rather use ordinary words.

In the paper “Measurement and Analysis of Burst Topic in Microblog”, authors research on user-oriented and message-oriented measurements of burst topic in Sina microblog. The measurements and analysis on large-scale Sina microblog data set show that the proposed measurement method can measure the characteristics of user and message propagation in burst topic. The measurement results in this paper can describe the formation and diffusion mechanism of burst topic which will contribute to better research of relevant issues on burst topic and ensure the well-developed of microblog.

In the paper “A Novel Combination Forecasting Algorithm Based on Time Series” a novel optimal weight combination forecasting model is proposed, and a new fixed weight calculation method based on mean relative error is also introduced.

In the paper “A Method of Description on the Data Association Based on Granulation Trees”, to investigate the association of data with other data in reality, the research begins with data sets which are divided into different partitions. Because each partition consists of granules and owns a level, all the partitions constitute a granulation set whose elements are the granules. As a hierarchy system, the granulation set together with the inclusion
relation gives rise to a structure called a granulation tree. The research on the data association establishes a method to describe the associations of the data in a granulation tree with the data in another granulation tree. The method involves a necessary and sufficient condition used to check the data associations. Because the necessary and sufficient condition is bound up with the upper approximation, the study also develops a way of investigation into rough sets. As an example, a practical problem is modeled by granulation trees, and the associations of the data in a granulation tree with the data in another granulation tree can be examined by use of the necessary and sufficient condition. Meanwhile, because the study is closely linked to granules and alterations of granularity, the process can be viewed as an approach to research on granular computing.

The paper “Research on Operation Management under the Environment of Cloud Computing Data Center” puts forward the operation management of cloud computing data center should emphatically expand from the comprehensive monitoring operation management, a full range of security management and improved automatic operation mechanism three aspects, in order to effectively realize that providing services as required, infrastructure management, physical equipment management, system and data management and other aspects of management, cloud computing data center combined with ITIL V3 to form an operation management practice under the cloud computing environment, finally get the comprehensive and effective operation management response by the IT resource management, IT business management, IT operation management.

Authors of paper “Exploiting Historical Diffusion Data to Maximize Information Spread in Social Networks” attempt harnessing historical information cascades data to learn how information propagates in social networks and how to maximize its spread. In particular, they proposed a voting algorithm to learn diffusion probabilities of edges from cascades data. Then a pruning method is developed to remove trivial edges whose weights are smaller than a threshold. Moreover, motivated by the social influence locality, they propose a Local Influence Model to evaluate node’s influence within a local area instead of the whole network, which can effectively reduce the computational complexity. Based on Local Influence Model, they use greedy algorithm to find an approximate optimal solution. Experimental results show that our method significantly outperforms state-of-the-art models both in terms of information spread and algorithm runtime.

In the research paper “Research on the Improved Shuffled Frog Leaping Algorithm in Cloud Computing Resources” the current status of cloud computing is first analyzed, and on the basis of the features of resource scheduling in cloud computing, the Intelligent Frog Leaping Algorithm is introduced and improved. First, artificial vector machine is introduced into the subgroup classification of the Frog Leaping Algorithm, and at the same time the self-adaptive crossover probability is introduced into the internal search of the algorithm. To some extent, this improves the condition that the Frog Leaping Algorithm is easy to fall into local optimum, as well as reduces the time spent on global search and optimization. Through the Cloud Sim platform, it is found that the algorithm presented in this paper can improve the efficiency of the system in task processing and achieve a rational scheduling of resources in cloud computing.

The paper “FOCCX: An Optimistic Concurrency Control Protocol over XML” presents a new optimistic approach for concurrency control over XML documents named FOCCX (Forward oriented Optimistic Concurrency Control over XML) facing XPath-based API. FOCCX increases the degree of transaction concurrency. This is achieved by aborting the current transaction when a potential UPDATE-UPDATE conflict taking place as early as possible, and reduces comparison times by checking a small write set against read set of a limited number of concurrent transactions. Experimental results show that the protocol
has superior performance to approaches based on Backward Oriented mechanism (BOCC).

In the paper “The Research of Data Mining Classification Algorithm that Based on SJEP” presents a special type of JEP, known as the most effective mode hopping revealed (SJEP: Most Significant Jumping Emerging Patterns). The classification algorithm takes full advantage of the ability to distinguish between JEP strong, while using only the most effective jumping emerging patterns (SJEP) as a basis for classification algorithm enhances endurance noise, reducing the complexity of the algorithm.

The paper “Extracting Entity Relationship Diagram (ERD) from English Sentences” describes a methodology that extracts ERD from English sentences. This methodology is based on a predefined set of a heuristic rules that aims to extract the elements of the ERD, then these rules are mapped into a diagram. A diagram generator automatically converts the rules into the ERD according to the rules of generating. The proposed methodology is explained by examples to show how it can provide a mechanism for quickly and easily way in extracting the ERD.

In the study entitled “The Research on Measure Method of Association Rules Mining” the advantages and disadvantages of the specific indicators of objective measure, subjective measure, and association rule based on statistical perspective are discussed. Some indicators of statistical perspective are adopted to measure the association rules, which can effectively solve the problems of association rules. Next, a further verification of the advantage and disadvantages of the indicators is made by the combination of the theory and application, a new measure frame is put forward as well. Then, the dynamic association rules are analyzed through making a comparative analysis in the following four aspects: the traditional association analysis without the life cycle, the association rules with the life cycle, the weighted dynamic association rules and the weighted dynamic association rules weighted by the consumption amount, showing the influence of timeliness on association rules analysis, and thus effectively mining some rules with low support in global period but high support in a certain period.

Paper “Relational Database’s Transaction Operation and the Concurrent Control” puts forward the strategy of concurrent control that would solve the concurrent operation, which results from the transaction object in order to ensure the data integrity.

The paper “Dynamic Cost-Sensitive Fuzzy Clustering for Uncertain Data Based on the Genetic Algorithm” puts forward the dynamic cost-sensitive fuzzy clustering algorithm based on genetic algorithm for uncertain data, which not only considers the cost, but also can search the optimal solution by the genetic algorithm. The algorithm can dispose the continuous and discrete attribute for uncertain data by the interval number. Experimental result shows, the fuzzy clustering we proposed has higher clustering accuracy and performance than the other algorithms for uncertain data, which can save especially the total cost, and is suitable for uncertain data.

April 2015

_Jianhua Ma, Hosei University, Japan_  
_Yanchun Zhang, Victoria University, Australia_

**Editor of the April Issue on**  
International Journal of Database Theory and Application