THE IEEE
Intelligent Informatics
BULLETIN

IEEE COMPUTER SOCIETY

IEEE Computer Society
Technical Committee
on Intelligent Informatics

December 2012 Vol. 13 No. 1 (ISSN 1727-5997)

On-line version: http://www.comp.hkbu.edu.hk/~iib (ISSN 1727-6004)

The IEEE Intelligent Informatics Bulletin

Aims and Scope

The IEEE Intelligent Informatics Bulletin is the official publication of the Technical Committee on Intelligent Informatics (TCII) of the IEEE Computer Society, which is published once a year in both hardcopies and electronic copies. The contents of the Bulletin include (but may not be limited to):

1) Letters and Communications of the TCII Executive Committee
2) Feature Articles
3) R&D Profiles (R&D organizations, interview profile on individuals, and projects etc.)
4) Book Reviews
5) News, Reports, and Announcements (TCII sponsored or important/related activities)

Materials suitable for publication at the IEEE Intelligent Informatics Bulletin should be sent directly to the Associate Editors of respective sections.

Technical or survey articles are subject to peer reviews, and their scope may include the theories, methods, tools, techniques, systems, and experiences for/in developing and applying biologically and linguistically motivated computational paradigms, such as artificial neural networks, fuzzy logic, evolutionary optimization, rough sets, and self-organization in the research and application domains, such as data mining, Web intelligence, intelligent agent technology, parallel and distributed information processing, and virtual reality.

Publisher: *The IEEE Computer Society Technical Committee on Intelligent Informatics*
Address: *Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong (Attention: Dr. William K. Cheung; Email:william@comp.hkbu.edu.hk)*
ISSN Number: *1727-5997(printed)1727-6004(on-line)*
Abstracting and Indexing: *All the published articles will be submitted to the following on-line search engines and bibliographies databases for indexing—Google(www.google.com), The ResearchIndex(citeseer.nj.nec.com), The Collection of Computer Science Bibliographies (liinwww.ira.uka.de/bibliography/index.html), and **DBLP** Computer Science Bibliography (www.informatik.uni-trier.de/~ley/db/index.html).*

Open Peer-Review Experiment in the Decentralized Coordination Workshop

MARIUS SILAGHI, SONG QIN
FLORIDA TECH
MARIUS.SILAGHI@FIT.EDU

WILLIAM CHEUNG
HONG KONG BAPTIST UNIVERSITY
WILLIAM@COMP.HKBU.EDU.HK

Given known game-theoretical analyses of peer-reviewing for conferences, the plan is to experiment with organizing an event based on techniques that in simulations yield the best incentives for high quality reviewing.

I. GAME THEORY FOR PEER-REVIEWING

Game theory is a powerful tool that can be used to model, analyze, and compare complex social interactions. If one can formalize and quantify motivations in terms of concrete utility values, then one can predict the behavior of the society under the assumption of rational participants. Moreover, one can simulate the impact of new regulations.

Peer-reviewing is an important social activity, whose quality directly impacts the advancement of science and economy. The traditional expectation for reviewers is to be altruistic, just as for politicians. It is nevertheless educational to see what would happen in case they behave like rational players. In general, exact utilities of human players are hard to quantify; in peer-reviewing as in auctions, negotiations and war. Nevertheless, one knows that often reviewers' revenues come from promotions and funding that depend on quantifiable metrics based on material facts such as the number of citations that they get, the number of articles that they publish, the number of reviewing boards on which they are invited, etc.

It is possible to use the aforementioned metrics to create approximate models linking the behavior of a reviewer in the peer-reviewing process to the net impact on the funding (utility) that he will get. Let us take an example where we assume funding offered to a researcher is related to the total count of her publications, and conferences limit the number of accepted papers based on a fixed threshold. A conference organized by a community of n researchers has these researchers simultaneously submitting papers and reviewing submissions of their peers. The researchers are considered at the same level of expertise, and the papers are considered equally worthy. In this example it is assumed that a single blinded review is written for each paper and that the review can take two values: $\{$*low*, *high*$\}$. Each researcher submits one paper and reviews one paper. The conference only accepts a fraction $\frac{1}{k}$ of the n submitted papers (i.e. accepts $m = \frac{n}{k}$ submissions). The revenue of an author for publishing a paper is 1. The expected gain from rejecting a paper is $\frac{m}{n} - \frac{m}{n-1} \approx \frac{1}{kn}$. The pair-wise payoff matrix for researchers A and B blindly reviewing each-other's papers is given in Fig. 1. It reveals an equilibrium consisting in scoring each other's article *low*.

	high A's	low A's
high B's	0,0	$0, \frac{1}{kn}$
low B's	$\frac{1}{kn}, 0$	$\frac{1}{kn}, \frac{1}{kn}$

Fig. 1. Pair-wise payoff matrix in reviewer-author game with n researchers in conferences with a threshold on the number of accepted papers and funding based on the count of published papers.

Other models exist for funding based on citation influence, where reviewers have strategies to increase current and future citations (see [Peterson et.al.]).

While errors induced by approximations in such models should be subject to further investigation, current results suggest that certain versions of open peer-review schemes have better equilibria for truthful reviewing than common blind review procedures. The mentioned open peer-review schemes, further explained below, are scheduled to be experimented within a workshop on decentralized coordination planned for Spring 2013.

II. OPEN PEER-REVIEW IN WORKSHOPS

With open peer-review, the reviews and the identity of the reviewers are published along with the endorsed and rejected submissions, as an incentive for improving their quality. The publication effectively creates a new link between reviews and the utility of the reviewer, since reviews can be cited, making it possible to create incentives for truthful reviewing. Even before a game-theoretic study provided any objective support for it, open-review has been advocated by various researchers. While strong calls for a shift towards open review have been issued in authoritative venues, such as the January 2009 IEEE Spectrum, the open peer-review is not yet common in computer science symposiums. However, several highly rated journals in natural sciences are currently employing open-review procedures (*Atmospheric Chemistry and Physics*, *Biology Direct*, *Journal of Medical Internet Research*, etc.). Some open-review schemes reveal only either the reviews or the name of the reviewers of accepted papers, while other venues publish entirely the name, reviews, and answers from authors. Sometimes the community can see submitted papers and researchers can propose themselves as reviewers of journal submissions. The obvious problem is that researchers may be reluctant to write negative reviews if they are going to be published. The way *Biology Direct* addresses this problem is by accepting only articles which receive at least three reviews [Koonin et.al.]. The corresponding reviewer-author graph that we obtained by parsing the publicly available data is shown in Fig. 2. The node size is proportional to the number of reviews written. As seen in this image, such data about communities around research pub-

lication venues can help detect close knit sub-communities and highly influential reviewers (large nodes in the picture). Some sub-communities, such as the cluster on the bottom right, can be completely separated in the review process from the rest of the researchers. Other quite large communities can be linked via as few as 2 or 3 researchers.

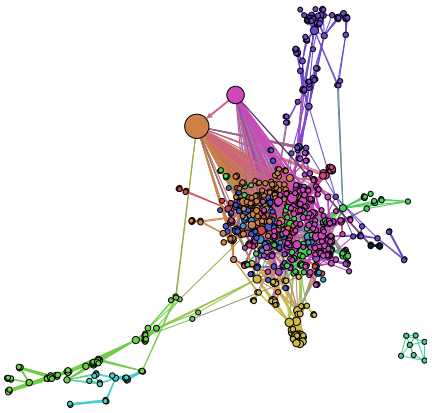


Fig. 2. Reviewer-author relation at *Biology Direct*.

Another kind of information offered by open review is illustrated by the reviewer-paper graph. The reviewer-paper relations for *Biology Direct* is shown in Fig. 3. It reveals that many papers are reviewed only by researchers not involved in reviewing anything else for this journal. Under the working assumptions, this raises questions about whether the given paper is relevant to the core community. Meanwhile, a few authors review a significant number of papers, yielding an unmatched influence on what is being published.

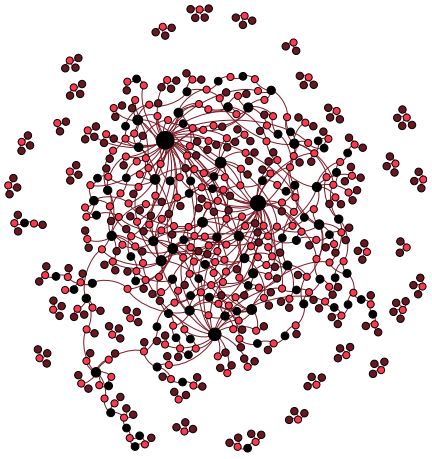


Fig. 3. Reviewer-paper relation at *Biology Direct*. Darker nodes show reviewers and red nodes show papers.

Using open review for workshops and conferences is complicated by the limits on the time available for writing publishable, good quality reviews. A rare example is a 2007 ConnectED workshop in Design. The next section details the reviewing mechanism planned for the 2013 workshop on decentralized coordination.

III. Decentralized Coordination Workshop (DCW)

Decentralized coordination is a challenging problem in multi-agent system as well as in human societies in general. With byzantine behavior from agents, even the seemingly simple problem of agreeing on a bit is not trivial. A robust deterministic agreement protocol was proved impossible even in the case of a single failure. The area of distributed computing has seen significant work on the problem of byzantine consensus. Typically a limit is assumed on the number of supported incorrect participants. Similarly, the area of distributed CSPs has a significant impossibility result concerning self-stabilization when all participants have equal priority.

With decentralized coordination, the focus is on techniques and applications where the decisions are construed via a distributed process by multiple participants. The participants are assumed to have a fair say in the final decision. A challenge is to make the coordination process robust to attempts of manipulation by a subset of the players.

It is a common practice that the articles accepted in a venue are not so much decided based on a global merit but based on their merit with respect to the interests and expertise of the reviewers in the community around that venue. This explains the decision mechanism used by the *Biology Direct* journal where any article receiving three open peer-reviews is published.

In DCW, submitted papers will be posted such that workshop committee members can bid on reviewing the ones they find interesting and where they feel they can write a meaningful review. Each article is allocated to some reviewers that bade on it. Remaining reviewing assignments are randomly allocated to reviewers that did not get the papers for which they did bid.

Authors will get an opportunity to write a response to the received reviews, and the answer will be published together with the reviews. A reviewer can withdraw her review after seeing the answer to it. After reading the articles, reviewers assigned to a submission can decide to not actually submit a review. If the reviewer did bid for that paper, her name will still be officially marked as an assigned reviewer of the corresponding article. Articles that receive reviews will be sorted and accepted for either oral or poster presentation. If the authors do not withdraw them after seeing the reviews, submissions will be published together with the reviews and author answers, in the peer-reviewed section of the proceedings. Submissions for which nobody bids and that nobody reviews will only be made available as non-reviewed technical reports in the final proceedings.

A given program committee has a limited capacity in terms of number of quality reviews that it can provide for a workshop. This capacity can be dynamically extended by inviting new reviewers after submissions. By opening the organization process, the obtained experiment is useful not only as reference for future organizers, future studies of peer-reviewing processes, and research on community detection, but can also facilitate the community formation. The workshop call for papers is available at: http://cs.fit.edu/~msilaghi/WDC.

Bibliography

[Peterson et.al.]; “Game Theoretical Modeling and Studies of Peer-Reviewing Methods”, ISPR, Vol I, pp 267-272; 2009.

[Koonin et.al.]; “A community experiment with fully open and published peer review”, Biology Direct, 1:1; 2006

Marius Silaghi is assistant professor in Computer Sciences at Florida Tech. His main research interests lie in the application of artificial intelligence, game theory and cryptology to support distributed decision making.

Song Qin is a PhD candidate in Computer Science at Florida Tech. He is interested in decentralized coordination and complex networks.

William Cheung is associate professor in Computer Science at Hong Kong Baptist University. His work focuses on data mining and intelligent systems.

Granular Mining and Rough-Fuzzy Pattern Recognition: A Way to Natural Computation

Sankar K. Pal, *Fellow, IEEE*

***Abstract*—Rough-fuzzy granular approach in natural computing framework is considered. The concept of rough set theoretic knowledge encoding and the role *f*-granulation for its improvement are addressed. Some examples of their judicious integration for tasks like case generation, classification/ clustering, feature selection and information measures are described explaining the nature, roles and characteristics of granules used therein. While the method of case generation with variable reduced dimension has merits for mining data sets with large dimension and size, class dependent granulation coupled with neighborhood rough sets for feature selection is efficient in modeling overlapping classes. Image ambiguity measures take into account the fuzziness in grey region, as well as the rough resemblance among nearby grey levels and nearby pixels, and are useful in image analysis. Superiority of rough-fuzzy clustering is illustrated for determining bio-bases in encoding protein sequence for analysis. *F*-information measures based on fuzzy equivalence partition matrix are effective in selecting relevant genes from micro-array data. Future directions of research, challenges and significance to natural computing are stated. The article includes some of the results published elsewhere.**

***Index Terms* — soft computing, granulation, generalized rough sets, rough-fuzzy computing, data mining, bioinformatics, image analysis, case based reasoning.**

I. INTRODUCTION

NATURAL computing, inspired by biological course of action, is an interdisciplinary field that formalizes processes observed in living organisms to design computational methods for solving complex problems, or designing artificial systems with more natural behavior. Based on the tasks abstracted from natural phenomena, such as brain modeling, self-organization, self-repetition, self-evaluation, Darwinian survival, granulation and perception, nature serves as a source of inspiration for the development of computational tools or systems that are used for solving complex problems. Nature inspired main computing paradigms used for such development include artificial neural networks, fuzzy logic, rough sets, evolutionary algorithms, fractal geometry, DNA computing, artificial life and granular or perception-based computing. Information granulation in granular computing is an inherent characteristic of human thinking and reasoning process performed in everyday life. One may refer to [1] for different facets of natural computing.

Sankar K. Pal is with the Center for Soft Computing Research Indian Statistical Institute, Kolkata, India (phone: 91-33-25752041; fax: 91-33-27583357; e-mail: E-mail: sankar@isical.ac.in).

Rough set theory is a popular mathematical framework for granular computing. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. Granules are formed as objects and are drawn together by the limited discernibility among them. Rough set represents a set in terms of lower and upper approximations. The lower approximation contains granules that completely belong in the set and the upper approximation contains granules that partially or completely belong in the set. Two major characteristics of rough set theory are uncertainty handling (using lower and upper approximations) and granular computing (using information granules). Rough set based techniques have been used in pattern recognition, image processing, data mining and knowledge discovery process from large data sets, among others. Rough sets were found to have extensive application in dimensionality reduction and knowledge encoding particularly when the uncertainty is due to granularity in the domain of discourse. It is also found to be an effective machine learning tool for designing ensemble classifier. One may note that fuzzy set theory deals with ill-defined and unsharp boundaries while rough set characterizes a crisp set with a coarsely defined class boundary. Rough sets are nothing but crisp sets with rough descriptions.

Rough-fuzzy or fuzzy-rough techniques are efficient hybrid methods based on judicious integration of the principles of rough sets and fuzzy sets. While the membership functions of fuzzy sets enable efficient handling of overlapping classes, the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition. Their judicious integration therefore promises to results in efficient paradigms for uncertainty handling which is much stronger than those of the individual ones.

It may be mentioned that the concept of rough-fuzzy computing has a significant role in modeling the fuzzy-granulation (*f*-granulation) characteristics of *Computational theory of perceptions* (CTP) [2], [3] which is inspired by the remarkable human capability to perform a wide variety of physical and mental tasks, including recognition tasks, without any measurements and computations. Perceptions are intrinsically imprecise. Their boundaries are fuzzy and the attribute they can take are granules. In other words, perceptions are *f*-granular.

The organization of the paper is as follows: Section 2 presents rough-fuzzy approach to granular computation, in general. Section 3 describes generalized rough sets for better

uncertainty handling by incorporating fuzziness in both set and granules definition. Section 4 explains the application of rough-fuzzy granulation in case based reasoning where the problem of case generation is considered. Certain challenging issues concerning granules for implementing rough-fuzzy computing are mentioned. Section 5 describes the merits of class dependent granulation for modeling overlapping classes in pattern recognition. The features are explained on remotely sensed imagery where labeled samples are scarce. Sections 6 and 7 demonstrate the characteristics of rough-fuzzy clustering, and application of fuzzy c-medoids to protein sequence analysis for determining bio-bases respectively. It is shown that rough-fuzzy clustering is superior to fuzzy clustering, hard clustering and rough clustering. Section 8 describes rough-fuzzy entropy based on generalized rough sets in measuring image ambiguities and an example application to image segmentation. It is demonstrated that incorporation of the concept of granularity in reflecting the rough resemblance in nearby gray levels and pixels improves the performance over fuzzy set theoretic segmentation. Section 9 deals with the problem of gene selection from microarray data where the significance of fuzzy equivalence partition matrix is demonstrated though various information measures. Concluding remarks are given in Section 9.

II. Granular Computation and Rough-Fuzzy Approach

Rough set theory [4] provides an effective means for analysis of data by synthesizing or constructing approximations (upper and lower) of set concepts from the acquired data. The key notions here are those of "information granule" and "reducts". Information granule formalizes the concept of finite precision representation of objects in real life situation, and reducts represent the core of an information system (both in terms of objects and features) in a granular universe. *Granular computing* (GrC) refers to that where computation and operations are performed on information granules (clump of similar objects or points). Therefore, it leads to have both data compression and gain in computation time, and finds wide applications. An important use of rough set theory and granular computing in data mining has been in generating logical rules for classification and association. These logical rules correspond to different important regions of the feature space, which represent data clusters.

In many situations, when a problem involves incomplete, uncertain and vague information, it may be difficult to differentiate distinct elements and one is forced to consider granules. On the other hand, in some situations though detailed information is available, it may be sufficient to use granules in order to have an efficient and practical solution. Granulation is an important step in the human cognition process. From a more practical point of view, the simplicity derived from granular computing is useful for designing scalable data mining algorithms. There are two aspects of granular computing, one deals with formation, representation and interpretation of granules (algorithmic aspect) while the other deals with utilization of granules for problem solving (semantic aspect). Several approaches for granular computing are suggested using fuzzy set theory, rough set theory, power algebras and interval analysis. The rough set theoretic approach is based on the principles of set approximation and provides an attractive framework for data mining and knowledge discovery.

For the past several years, rough set theory and granular computation has proven to be another soft computing tool which, in various synergistic combinations with fuzzy logic, artificial neural networks and genetic algorithms, provides a stronger framework to achieve tractability, robustness, low cost solution and close resembles with human like decision making. For example, rough-fuzzy integration [5] can be considered as a way of emulating the basis of f-granulation in CTP, where perceptions have fuzzy boundaries and granular attribute values. Similarly, rough-neural [6], [7] and fuzzy-rough-neural [8], [9] synergistic integration help in extracting crude domain knowledge in the form of rules for describing different concepts/classes, and then encoding them as network parameters; thereby constituting the initial knowledge base network for efficient learning. Since in granular computing computations/operations are performed on granules (clump of similar objects or points), rather than on the individual data points, the computation time is greatly reduced. The results on these investigations are available in different journals, conference proceedings, special issues and edited volumes [5], [10], [11].

Before we describe some applications of rough fuzzy computing in clustering, classification, mining and image analysis with different applications, we present briefly the concepts of generalized rough sets and case generation in rough-fuzzy framework as they form the basic principles of *f*-granulation in several applications.

III. Generalized Rough Sets: Lower & Upper Approximations

In Pawlak's rough set theory, both the set X and granules or equivalence relation R are considered to be crisp. However, in real life problems, they could be fuzzy too. Generalized rough sets are defined based on this premise where the expressions for the lower and upper approximations of a set X depend on the type of relation R and whether X is a crisp or a fuzzy set. Let us describe here briefly the expressions for the upper and lower approximations of X for different cases, i.e., when R denotes an equivalence or a fuzzy equivalence relation and X is a crisp or a fuzzy set.

Case 1: When R denotes an equivalence relation and X is a crisp set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{u \mid u \in U : [u]_R \subseteq X\},$$
$$\overline{R}X = \{u \mid u \in U : [u]_R \cap X \neq \varnothing\}, \quad (1)$$

where $[u]_R$ denotes the granule to which the element u belongs. In this case, the pair of sets $< \underline{R}X, \overline{R}X >$ is referred to as the rough set of X and $< U, R >$ is a crisp equivalence approximation space.

Case 2: When R denotes an equivalence relation and X is a fuzzy set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \inf_{z\in[u]_R} \mu_X(z)) \mid u \in U\},$$
$$\overline{R}X = \{(u, \sup_{z\in[u]_R} \mu_X(z)) \mid u \in U\}, \quad (2)$$

where μ_X is the membership function associated with X. In this case, the pair of fuzzy sets $<\underline{R}X, \overline{R}X>$ is referred to as the rough-fuzzy set of X and $<U, R>$ is a crisp equivalence approximation space.

Case 3: Let us now consider the case when R refers to a fuzzy equivalence relation, that is, when the belongingness of every element (u) in the universe (U) to a granule $Y \in U/R$ is specified by a membership function, say m_Y, that takes values in the interval $[0,1]$ such that $\sum_Y m_Y(u) = 1$. In such a case, when X is a crisp set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \sum_{Y\in U/R} m_Y(u) \times \inf_{\varphi\in U} \max(1 - m_Y(\varphi), C)) \mid u \in U\},$$
$$\overline{R}X = \{(u, \sum_{Y\in U/R} m_Y(u) \times \sup_{\varphi\in U} \min(m_Y(\varphi), C)) \mid u \in U\}, \quad (3)$$

where

$$C = \begin{cases} 1, & \varphi \in X \\ 0, & \varphi \notin X \end{cases} \quad (4)$$

In the above, the symbols $\sum$ (sum) and $\times$ (product) respectively represent specific fuzzy union and intersection operations. Note that, one may consider any fuzzy union and intersection operation instead of the sum and product operations by judging their suitability with respect to the underlying application. The pair of fuzzy sets $<\underline{R}X, \overline{R}X>$ is referred to as the fuzzy rough set of X in this case and $<U, R>$ is a fuzzy equivalence approximation space.

Case 4: In *Case 3* of R referring to a fuzzy equivalence relation, when X is a fuzzy set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \sum_{Y\in U/R} m_Y(u) \times \inf_{\varphi\in U} \max(1 - m_Y(\varphi), \mu_X(\varphi))) \mid u \in U\}$$
$$\overline{R}X = \{(u, \sum_{Y\in U/R} m_Y(u) \times \sup_{\varphi\in U} \min(m_Y(\varphi), \mu_X(\varphi))) \mid u \in U\}. \quad (5)$$

The pair of fuzzy sets $<\underline{R}X, \overline{R}X>$ is referred as the fuzzy rough-fuzzy set of X and $<U, R>$ is again a fuzzy equivalence approximation space. From the above explanation, it is obvious that the set of expressions in cases 1 -3 are special cases of the set of expressions for the lower and upper approximations given in Case 4. Pictorial diagram of lower and upper approximations for *Case 4* is shown in Fig 1.

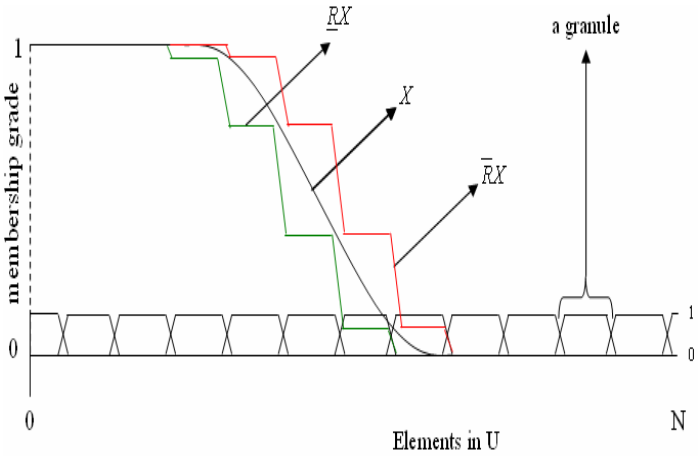


Fig. 1 The pair $<\underline{R}X, \overline{R}X>$ is referred to as the fuzzy rough-fuzzy set of X.

Significance of generalized rough sets in image analysis problem is described in Sec. 8, where entropy and image ambiguity measures are defined.

IV. Rough-Fuzzy Granulation and Case Generation

A case may be defined as a contextualized piece of knowledge representing an evidence that teaches a lesson fundamental to achieving goals of the system. Case based reasoning (CBR) [12] is a novel Artificial Intelligence (AI) problem-solving paradigm, and it involves adaptation of old solutions to meet new demands, explanation of new situations using old instances (called cases), and performance of reasoning from precedence to interpret new problems. It has a significant role to play in today's pattern recognition and data mining applications involving CTP, particularly when the evidence is sparse. The significance of soft computing to CBR problems has been adequately explained by Pal, Dillon and Yeung [13] and Pal and Shiu [14]. In this section we provide an example [15], [16] of using the concept of *f*-granulation for performing the task of *case generation* in large scale CBR systems. While case selection deals with selecting informative prototypes from the data, case generation concerns with construction of 'cases' that need not necessarily include any of the given data points.

For generating cases, linguistic representation of patterns is used to obtain a fuzzy granulation of the feature space. Rough set theory is used to generate dependency rules corresponding to informative regions in the granulated feature space. The fuzzy membership functions corresponding to the informative regions are stored as cases. Figure 2 shows an example of such case generation for a two dimensional data having two classes. The granulated feature space has $3^2 = 9$ granules. These granules of different sizes are characterized by three membership functions along each axis, and have ill-defined (overlapping) boundaries. Two dependency rules: $class_1 \leftarrow L_1 \wedge H_2$ and $class_2 \leftarrow H_1 \wedge L_2$ are obtained using rough set theory. The fuzzy membership functions, marked bold, corresponding to the attributes appearing in the rules for a class are stored as its case.

Unlike the conventional case selection methods, the cases here are cluster granules and not sample points. Also, since all the original features may not be required to express the dependency rules, each case involves a reduced number of relevant features. The methodology is therefore suitable for mining data sets, large both in dimension and size, due to its low time requirement in case generation as well as retrieval.

The aforesaid characteristics are demonstrated in Figure 3 [15], [16] for a forest cover type GIS data on seven kinds of wood with number of features 10 (cartographic and remote sensing measurements) and number of samples 586012. Their superiority over Instance-based learning (IB3), Instance-based learning with reduced number of features (IB4) and random case selection algorithms, in terms of classification accuracy (with one nearest neighbor rule), case generation (t_{gen}) and retrieval (t_{ret}) times, and average storage requirement (average feature) per case, is evident. The numbers of cases considered for comparison is 545. As can be seen, all the ten features are not required for providing highest classification rate, only four, on an average, is sufficient in the proposed method. Based on the similar concept, Li et al reported a CBR based classification system combining efficient feature reduction and case selection [17]. Note that here the granules considered are class independent. In the next section we describe a classification method where the granules are class dependent.

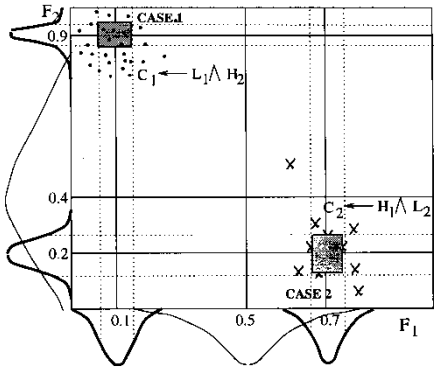


Fig. 2 Rough-fuzzy case generation for a 2-D data [15].

Before we describe some applications of rough-fuzzy granular computing, we mention certain issues for their implementation, namely,

- Selection of granules and their sizes/ shapes
- Class dependent or independent granules
- Fuzzy granules
  - Fuzzy set over crisp granules
  - Crisp set over fuzzy granules
  - Fuzzy set over fuzzy granules
- Granular *fuzzy computing*
- *Fuzzy granular* computing

Class dependent granulation, as expected, has merits over class independent granulation in modeling overlapping classes, but with additional computation cost. Granular *fuzzy computing means* granules are crisp whereas computing done with them is fuzzy. On the other hand, crisp computing with fuzzy granules refers to *Fuzzy granular* computing. These issues are described in the following applications.

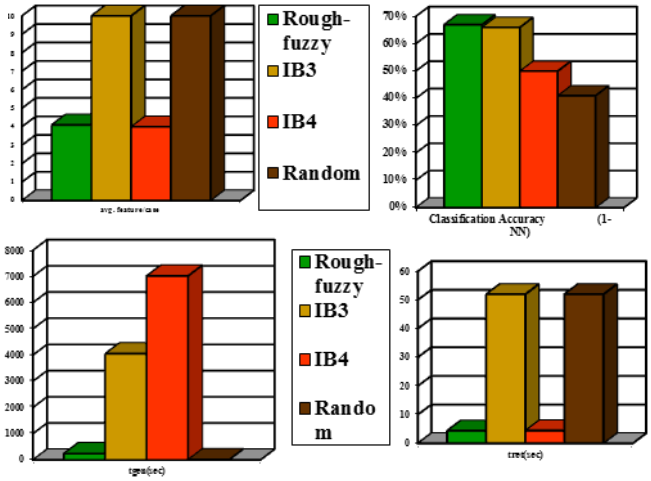


Fig. 3 Performance of different case generation schemes for the forest cover-type GIS data set with 7 classes, 10 features and 586012 samples.

V. Rough-Fuzzy Classification

For a given input pattern, the rough-fuzzy class dependent pattern classification model has the following three steps [18]:

Step 1 Generate fuzzy granulated feature space
Step 2 Remove redundant features using rough sets, and
Step 3 Classify

The first step generates the class-dependent (CD) fuzzy granulated feature space of input pattern vector. For fuzzy granulation of a feature space containing **L** number of classes, we used **L** number of π-type fuzzy sets to characterize the feature values of each pattern vector. Each feature is thus represented by **L** number of [0, 1]-valued membership functions (MFs) representing **L** fuzzy sets or characterizing **L** fuzzy granules along the axis. That is, each feature of a pattern $\mathbf{F} = [F1, F2,..., Fn]$ characterizes **L** number of fuzzy granules along each axis and thus comprising $\mathbf{L}^n$ fuzzy granules in an n-dimensional feature space. Fig. 4 shows a crisp visualization of 16 ($= 4^2$) such class dependent granules using 0.5-cut when the no. of classes is four in two-dimensional feature space. The shape and size of the granules are dependent on the nature of overlapping of classes and class-wise feature distribution. (One may note that using class independent granulation, as in Fig. 5, with low, medium and high, the no. of granules generated for a two dimensional plane would be 9 ($= 3^2$).

The increased dimension brings great difficulty in solving many tasks. This motivates for selecting a subset of relevant and non-redundant features. Accordingly, the neighborhood rough set (NRS) **[19]** based feature selection method is used in Step 2. The advantage in the use of NRS is that it can deal with both numerical and categorical data, and does not require any discretisation of numerical data. Further, the neighboring concept facilitates to gather the possible local information through neighbor granules that provide better class discrimination information. The integrated model thus takes the advantage of both class-dependent fuzzy granulation and NRS feature selection methods. After the features are selected, they can be used as input to any classifier in Step 3.

For implementation of the concept of neighbourhood rough sets in feature selection, let us assume an information system denoted by I = (U, A) where U (the universal set) is a non-empty and finite set of samples $\{x_1, x_2, ..., x_n\}$; A = {C ∪ D}, where A is the finite set of features $\{a_1, a_2, ..., a_m\}$, C is the set of conditional features and D is the set of decision features. Given an arbitrary xi ∈ U and B ⊆ C, the neighbourhood Φ_B(xi) of xi with given Φ, for the feature set B is defined as [18]

$$\phi_B(x_i) = \left\{ x_i \middle| x_j \in U, \Delta^B(x_i, x_j) \le \phi \right\}, \quad (6)$$

where Δ is a distance function.

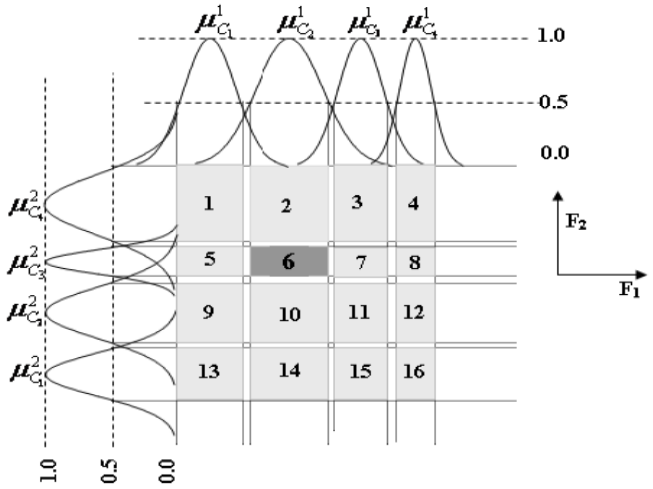


Fig. 4 Crisp visualization of sixteen class dependent granules for L = 4 generated from class-wise fuzzy representation of features F_1 and F_2.

$\Phi_B(x_i)$ in Eqn. (6) represents the neighborhood information granule centered with sample xi. That is, each sample xi generates granules with a neighbourhood relation. For a metric space (U, Δ), the set of neighbourhood granules $\{\Phi(x_i) | x_i \in U\}$ forms an elemental granule system, that covers the universal space rather than partitions it as done by Pawlak's rough set (PaRS). A neighbourhood granule degrades to an equivalence class when ϕ = 0. In this case, samples in the same neighbourhood granules are equivalent to each other and neighbourhood model degenerates to Pawlak's rough set. Thus NRS) can be viewed as a generalization of PaRS.

Generation of neighborhood depends on both distance function Δ and parameter Φ. The first one determines the shape and second controls the size of neighborhood granule. For example, with Euclidean distance the parameter Φ acts as the radius of the circle region developed by Δ function. Both these factors play important roles in neighbourhood rough sets (NRS) and can be considered as to control the granularity of data analysis. The significance of features varies with the granularity levels. Accordingly, the NRS based algorithm selects different feature subsets with the change of Δ function and Φ value.

Performance of rough-fuzzy feature selection (granular feature space and rough feature selection) is demonstrated here with 1-NN classifier, as an example, on remotely sensed images where the different regions are highly overlapping and the number of available training samples is small. Table 1 shows the comparative performance of various models in terms of β value [20] and Davies-Bouldin (DB) value on IRS-1A image and SPOT image with partially labelled samples. (Partially labelled means, the classifiers are initially trained with labelled data of six land cover types and then the said trained classifiers are applied on the unlabeled image data to partition into six regions.)

Five different models considered are:

- Model 1: 1-NN classifier,
- Model 2: CI fuzzy granulation + Pawlak's rough set (PaRS) based feature selection + 1-NN classifier,
- Model 3: CI fuzzy granulation + neighborhood rough set (NRS) based feature selection + 1-NN classifier,
- Model 4: CD fuzzy granulation + PaRS based feature selection + 1-NN classifier,
- Model 5: CD fuzzy granulation + NRS based feature selection + 1-NN classifier.

TABLE I
COMPARATIVE PERFORMANCE OF MODELS USING 1-NN CLASSIFIER WITH PARTIALLY LABELLED DATA SETS (FOR Φ = 0.45 AND Δ = EUCLIDEAN DISTANCE)

Model	β value		DB value	
	IRS-1A	SPOT	IRS-1A	SPOT
Training sample	9.4212	9.3343	0.5571	1.4893
1	6.8602	6.8745	0.9546	3.5146
2	7.1343	7.2301	0.9126	3.3413
3	7.3559	7.3407	0.8731	3.2078
4	8.1372	8.2166	0.779	2.8897
5	8.4162	8.4715	0.7345	2.7338

As expected, the β value is the highest and DB value is the lowest for the training set (Table 1). It is also seen that model 5 yields superior results in terms of both the indexes. As a whole, the gradation of performance of five models can be established with the following β relation:

$$\beta_{train} > \beta_{model5} > \beta_{model4} > \beta_{model3} > \beta_{model2} > \beta_{model1} \quad (7)$$

Similar gradation of performance is also observed with DB values, which further supports the superiority of model 5.

In order to demonstrate the significance of granular computing visually, let us consider Figs. 5a and 5b depicting the output corresponding to models 1 (without granulation) and 5 (with granulation), say, for IRS-1A. It is clear from the figures that model 5 performed well in segregating different areas by properly classifying the land covers. For example, the *Howrah bridge* over the south part of the *river* is more prominent in Fig. 5b, whereas it is not so in Fig. 5a.

Tables II and III show the confusion matrix and dispersion score of each of the six land cover classes for models 1 and 5 respectively. Dispersion score signifies the variance in misclassified samples. Lower dispersion score, which is desirable, means misclassified samples are confused among least number classes; thereby providing more opportunity for

Fig. 5a Classified IRS-1A images with model 1.

Fig. 5b Classified IRS-1A images with model 5.

TABLE II
CONFUSION MATRIX AND DISPERSION SCORES FOR SIX CLASSES OF IRS-1A IMAGE FOR MODEL 1

		Predicted Class						Dispersion Score
	Class	C1	C2	C3	C4	C5	C6	
Actual Class	C1	128	14	2	2	1	0	0.4090
	C2	11	170	44	25	3	2	0.7126
	C3	10	80	201	131	17	3	0.8535
	C4	8	98	230	842	30	12	0.6828
	C5	25	25	25	147	688	365	0.8011
	C6	6	3	2	4	15	105	0.5912

them to get corrected at the next level with higher level information. Model 5 has lowest dispersion score (see Table III) while model 1 has highest (see Table II). Again, the score is minimum for C1 (pure water) and C6 (open space) as they have least overlapping with others, whereas the value is larger for classes like C3 (concrete area) and C5 (vegetation) having significant overlapping with neighbouring classes. However, computation time wise, it increases in the order as we move from model 1 to model 5.

TABLE III
CONFUSION MATRIX AND DISPERSION SCORES FOR SIX CLASSES OF IRS-1A IMAGE FOR MODEL 5

		Predicted Class						Dispersion Score
	Class	C1	C2	C3	C4	C5	C6	
Actual Class	C1	142	3	1	0	0	1	0.2097
	C2	5	216	20	10	2	2	0.4968
	C3	6	45	301	80	8	2	0.6933
	C4	1	40	151	1010	13	5	0.5182
	C5	8	10	11	47	987	212	0.5844
	C6	3	1	1	2	**5**	123	0.4009

VI. ROUGH-FUZZY CLUSTERING

The classification method in Sec 5 is an example of *fuzzy granular* computing. The rough-fuzzy clustering method, termed as rough-fuzzy c-means (RFCM), that will be described here, on the other hand, refers to granular *fuzzy computing*. The RFCM adds the concept of fuzzy membership of fuzzy sets, and lower and upper approximations of rough sets into c-means algorithm. While the membership of fuzzy sets enables efficient handling of overlapping partitions, the rough sets deal with uncertainty, vagueness, and incompleteness in class definition [21].

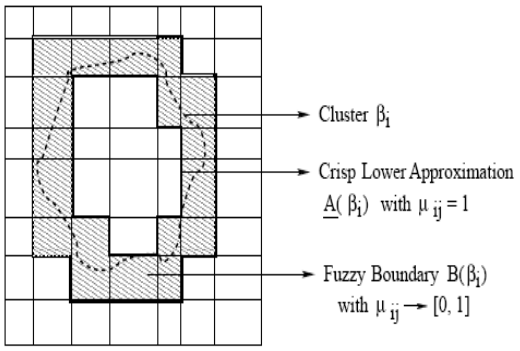


Fig. 6. Rough-fuzzy c-means: each cluster is represented by crisp lower approximations and fuzzy boundary [21], [22].

In RFCM, each cluster is represented by

- a cluster prototype (centroid),
- a crisp lower approximation, and
- a fuzzy boundary.

The lower approximation influences the fuzziness of final partition. According to the definitions of lower approximations and boundary of rough sets, if an object belongs to lower approximations of a cluster, then the object does not belong to any other clusters. That is, the object is contained in that cluster

definitely. Thus, the weights of the objects in lower approximation of a cluster (Fig. 6) should be independent of other centroids and clusters, and should not be coupled with their similarity with respect to other centroids. Also, the objects in lower approximation of a cluster should have similar influence on the corresponding centroids and cluster. Whereas, if the object belongs to the boundary of a cluster, then the object possibly belongs to that cluster and potentially belongs to another cluster. Hence, the objects in boundary regions should have different influence on the centroids and clusters. So, in RFCM, the membership values of objects in lower approximation are 1, while those in boundary region are the same as fuzzy c-means. In other word, RFCM first partitions the data into two classes-lower approximation and boundary. Only the objects in boundary are fuzzified. The new centroid is calculated based on the weighting average of the crisp lower approximation and fuzzy boundary. Computation of the centroid is modified to include the effects of both fuzzy memberships and lower and upper bounds. In essence, rough-fuzzy clustering (RFCM)

- provides a balanced compromise between restrictive (hard clustering) and descriptive (fuzzy clustering) partitions
- is faster than fuzzy clustering
- provides better uncertainty handling capability/ performance.

Therefore, wherever fuzzy c-means (FCM) [24] algorithm has been found to be successful since its inception, RFCM would have an edge there in terms of both performance and computation time. This feature of RFCM has been demonstrated extensively for different kinds of patterns including brain MRI Images [22]. RFCM is seen to perform better than hard c-means (HCM), rough c-means (RCM) [23] and fuzzy c-means (FCM).

VII. Clustering Rough Fuzzy C-Medoids and Amino Acid Sequence Analysis

In most pattern recognition algorithms, amino acids cannot be used directly as inputs since they are non-numerical variables. They, therefore, need encoding prior to input. In this regard, bio-basis function maps a non-numerical sequence space to a numerical feature space. It uses a kernel function to transform biological sequences to feature vectors directly. Bio-bases consist of sections of biological sequences that code for a feature of interest in the study and are responsible for the transformation of biological data to high-dimensional feature space. Transformation of input data to high-dimensional feature space is performed based on the similarity of an input sequence to a bio-basis with reference to a biological similarity matrix. Thus, the biological content in the sequences can be maximally utilized for accurate modeling. The use of similarity matrices to map features allows the bio-basis function to analyze biological sequences without the need for encoding. One of the important issues for the bio-basis function is how to select the *minimum set of bio-bases with maximum information*. Here, we present an application of rough-fuzzy c-medoids (RFCMdd) algorithm [25] to select the most informative bio-bases. The objective of the RFCMdd algorithm for selection of bio-bases is to assign all amino acid subsequences to different clusters. Each of the clusters is represented by a bio-basis, which is the medoid for that cluster. The process begins by randomly choosing desired number of subsequences as the bio-bases. The subsequences are assigned to one of the clusters based on the maximum value of the similarity between the subsequence and the bio-basis. After the assignment of all the subsequences to various clusters, the new bio-bases are modified accordingly [25]. Here similarity between two sequences is measured in terms of mutation probability of an amino acid using Dayoff mutation matrix.

The performance of RFCMdd algorithm for bio-basis selection is presented using five whole human immunodeficiency virus (HIV) protein sequences and Cai-Chou HIV data set, which can be downloaded from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). The performances of different c-medoids algorithms such as hard c-medoids (HCMdd), fuzzy c-medoids (FCMdd), rough c-medoids (RCMdd), and rough-fuzzy c-medoids (RFCMdd) [25] are reported with respect to β index and γ index based on homology alignment score [21]. The results establish the superiority of RFCMdd with lowest γ index and highest β index.

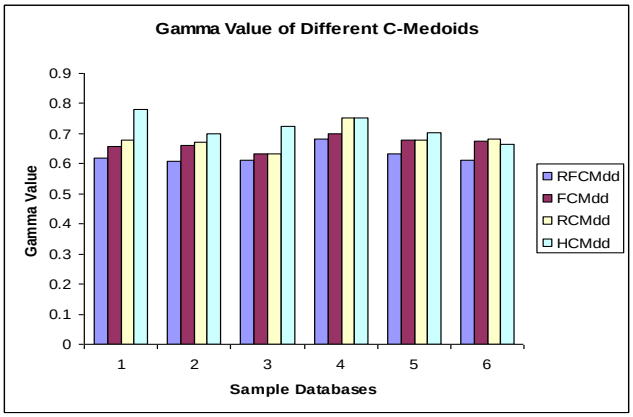


Fig. 7a Gamma values of different c-Medoids.

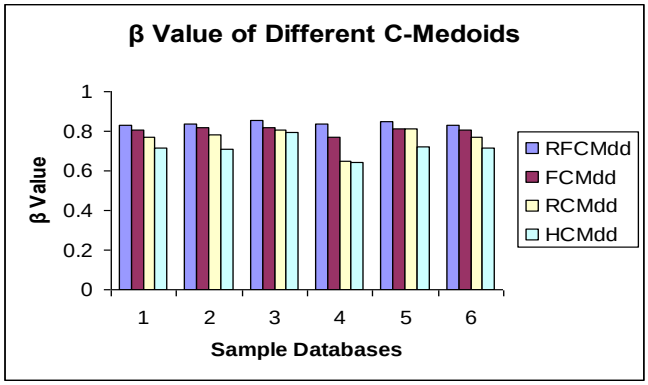


Fig. 7b β values of different c-Medoids

In previous examples we have demonstrated the role of granules in modeling overlapping classes, linguistic rules and in defining class exactness. The next two sections are based on entropy and mutual information measures defined over granulated space. In Section 8, we demonstrate how fuzzy boundaries of image regions, rough resemblance between nearby gray levels and rough resemblance between nearby

pixels *give rise to ambiguity in images*, where the significance of granules in determining *roughly resemblance in* gray levels and pixels is evident [26]**.** In Section 9 we demonstrate how mutual information defined on class independent fuzzy approximation space of attribute sets can be made useful for measuring the relevance of a conditional attribute with respect to decision attribute and redundancy among conditional attributes, and an application to selection of relevant genes from micro-array data.

VIII. Rough-fuzzy Entropy and Image Ambiguity Measures

Here we provide two classes of entropy measures based on roughness measures of a set X and its complement X^c in order to quantify the incompleteness of knowledge about a universe. One of them is based on logarithmic gain function, defined as [26]:

$$H_R^L(X) = -\frac{1}{2}[\rho_R(X)\log_\beta(\frac{\rho_R(X)}{\beta}) + \rho_R(X^c)\log_\beta(\frac{\rho_R(X^c)}{\beta})], \quad (8)$$

where β denotes the base of the logarithmic function used and $X \subseteq U$ stands for the complement of the set X in the universe. The various entropy measures of this class are obtained by calculating the roughness values $\rho_R(X)$ and $\rho_R(X^c)$ considering the different ways of obtaining the lower and upper approximations of the vaguely definable set X. Note that, the 'gain in incompleteness' term is taken as $-\log_\beta(\frac{\rho_R}{\beta})$ in (1) and for $\beta > 1$ it takes a value in the interval $[1, \infty]$. The other class of entropy measures, as obtained by considering an exponential function to measure the 'gain in incompleteness', is:

$$H_R^E(X) = \frac{1}{2}[\rho_R(X)\beta^{(1-\rho_R(X))} + \rho_R(X^c)\beta^{(1-\rho_R(X^c))}], \quad (9)$$

where β denotes the base of the exponential function used. Similar to the class of entropy measures H_R^L, the various entropy measures of this class are obtained by using the different ways of obtaining the lower and upper approximations of X in order to calculate $\rho_R(X)$ and $\rho_R(X^c)$. The 'gain in incompleteness' term is taken as $\beta^{(1-\rho_R)}$ in (2) and for $\beta > 1$ it takes a value in the finite interval $[1, \beta]$.

The plots of the entropies H_R^L and H_R^E as functions of A and B are given in Figs. 8 to 10. In Figs. 8 and 9, the values of H_R^L and H_R^E are shown for all possible values of the roughness measures A and B considering $\beta = e$. Fig. 10 shows the plots of the proposed entropies for different values of β, when $A = B$.

A. Image Ambiguity Measures and Segmentation

Using the aforesaid entropy definitions, we compute grayness and spatial ambiguity measures of an image. Grayness ambiguity refers to indefiniteness associated with deciding whether a pixel or a clump of pixels (granule) is white or black. That is, it concerns with the indefiniteness due

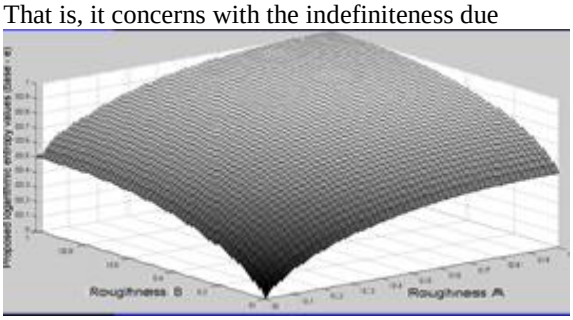


Fig. 8 Plot of logarithmic rough-fuzzy entropy

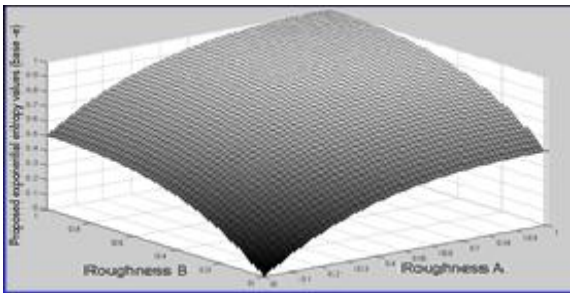


Fig. 9 Plot of exponential rough-fuzzy entropy

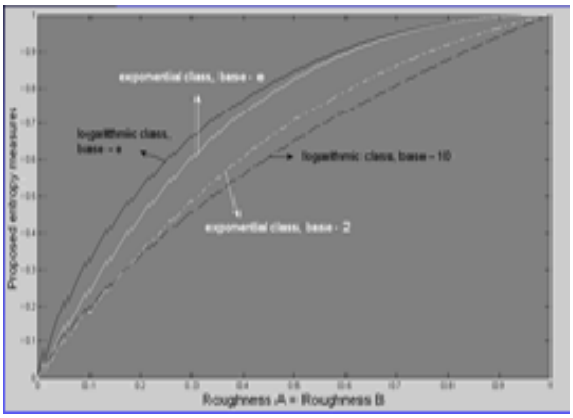
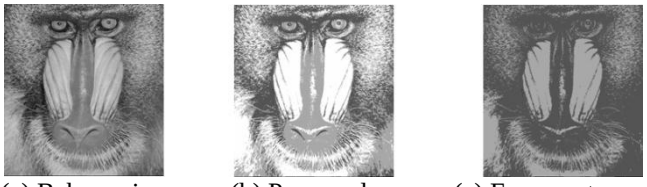


Fig. 10 Plots of entropy for different values of base β and gain functions. A → $\rho_R(X)$, B→ $\rho_R(X^c)$.

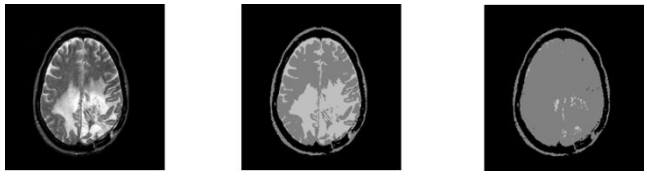
to fuzziness as well as granularity in gray values. Spatial Ambiguity, on the other hand, refers to indefiniteness in shape and geometry of various regions where indefiniteness is concerned with both intensity and spatial location of individual pixel or group of pixels. These ambiguity measures are minimized by changing the cross-over point of the membership function to find a set of minima corresponding to different thresholds of an image.

Fig 11 shows the segmentation results of three images, as an example, using grayness ambiguity measures based on rough-fuzzy entropy and fuzzy entropy [27]. In the former case, membership of a pixel is dependent on the granule (defined over one-dim gray scale) to which it belongs, and it is independent of its spatial location. Whereas, in the latter case, the membership of a pixel is entirely dependent on its own gray value, and it is independent of its spatial location. Therefore the improvement in segmentation results by rough-fuzzy entropy as compared to fuzzy entropy in Fig. 11 is due to inclusion of

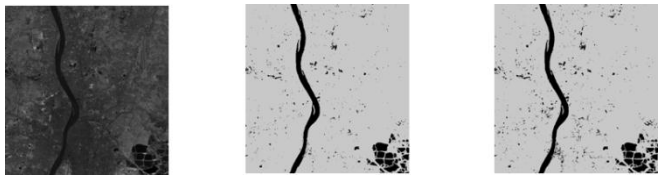
the concept of granules. The same is quantitatively demonstrated in Fig. 12 for 45 other images where β-index for segmentation is seen in almost all cases to be higher for outputs corresponding to rough-fuzzy entropy.



(a) Baboon image (b) Proposed (c) Fuzzy entropy



(a) Brain MR image (b) Proposed (c) Fuzzy entropy



(a) Remote sensing image (b) Proposed (c) Fuzzy entropy

Fig. 4 Segmentation results (Effect of granules)

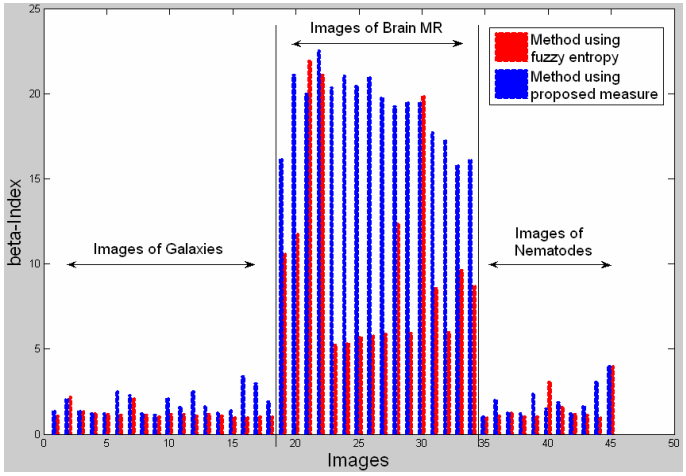


Fig. 52 β-index for segmentation results on 45 images (Significance of using the concept of granules is evident).

IX. Fuzzy Equivalence Partition Matrix and Gene Selection

An important application of gene expression data in functional genomics is to classify samples according to their gene expression profiles. In most gene expression data, number of training samples is very small compared to large number of genes involved in the experiments. Among the large amount of genes, only a small fraction is effective for performing a certain task. This leads to the task of gene selection i.e., identifying a reduced set of most relevant genes for a certain task.

Several information measures such as entropy, mutual information and f-information have been used in selecting a set of relevant and non-redundant genes from a microarray data set. For real-valued gene expression data, the estimation of different information measures is a difficult task as it requires knowledge on the underlying probability density functions of the data and the integration on these functions. Existing approaches include Discretization, and Parzen window methods. In this section various *f*-information measures [28] are computed on the fuzzy equivalence partition matrix defined along each gene axis, and based on them relevance and redundancy of a gene are determined. The subset of genes which provides maximum relevance to the decision classes and minimum redundancy among themselves in terms of the information measures is selected.

A. Fuzzy Equivalence Partition Matrix

If c and n denote the number of fuzzy information granules (equivalence classes) and number of objects in U, then c-partitions of U generated by fuzzy attribute A can be arrayed as a (c × n) fuzzy equivalence partition matrix.

$$\mathbb{M}_{\mathbb{A}} = \begin{pmatrix} m_{11}^{\mathbb{A}} & m_{12}^{\mathbb{A}} & \cdots & m_{1n}^{\mathbb{A}} \\ m_{21}^{\mathbb{A}} & m_{22}^{\mathbb{A}} & \cdots & m_{2n}^{\mathbb{A}} \\ \cdots & \cdots & \cdots & \cdots \\ m_{c1}^{\mathbb{A}} & m_{c2}^{\mathbb{A}} & \cdots & m_{cn}^{\mathbb{A}} \end{pmatrix}, \quad (10)$$

$m^A_{ij} \in [0, 1]$ is the membership value of object x_j in ith fuzzy equivalence class F_i. Fuzzy relative frequency corresponding to fuzzy equivalence partition F_i is

$$\lambda_{F_i} = \frac{1}{n}\sum_{j=1}^{n} m_{ij}^{\mathbb{A}}, \quad (11)$$

If fuzzy attribute sets P and Q generate p and q number of fuzzy equivalence classes, and P_i and Q_j represent corresponding ith and jth fuzzy equivalence partitions, then joint frequency of P_i and Q_j is

$$\lambda_{P_i Q_j} = \frac{1}{n}\sum_{k=1}^{n}(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}). \quad (12)$$

B. f-Information Measures

Various fuzzy-information measures on attribute sets are defined below based on the aforesaid individual frequency and joint frequency of different fuzzy equivalence partitions [28].

Entropy (on fuzzy approximation spaces of attribute set A):

$$H(\mathbb{A}) = -\sum_{i=1}^{c}\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{A}}\right]\log\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{A}}\right]. \quad (13)$$

Mutual information (between two attribute sets P and Q):

$$\mathrm{I}(\mathbb{P},\mathbb{Q}) = -\sum_{\mathrm{i}=1}^{\mathrm{p}}\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\right]\log\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\right] - \sum_{j=1}^{q}\left[\frac{1}{n}\sum\nolimits_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right]\log\left[\frac{1}{n}\sum\nolimits_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right] + \sum_{i=1}^{p}\sum_{j=1}^{q}\left[\frac{1}{n}\sum\nolimits_{k=1}^{n}\left(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}\right)\right]\log\left[\frac{1}{n}\sum\nolimits_{k=1}^{n}\left(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}\right)\right] \quad . \quad (14)$$

Other information (between two attribute sets P and Q):

$$\mathrm{V}(\mathbb{P},\mathbb{Q}) = \sum_{i=1}^{p}\sum_{j=1}^{q}\left|\frac{1}{n}\sum\nolimits_{k=1}^{n}\left(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}\right) - \frac{1}{n^2}\sum\nolimits_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right| \quad . \quad (15)$$

$$\chi^{\alpha}(\mathbb{P},\mathbb{Q}) = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left|\frac{1}{n}\sum_{k=1}^{n}\left(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}\right) - \frac{1}{n^2}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right|^{\alpha}}{\left(\frac{1}{n^2}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right)^{\alpha-1}} \quad . \quad (16)$$

C. Method of Selection of Genes

Principle:

- Compute *Total Relevance* of selected genes, J1 = ∑P I (P, R) where P is a gene (condition attribute) and R denotes the sample class labels (decision attribute).
- Compute *Total Redundancy* among selected genes, J2 = ∑P, Q I(P, Q) where P and Q are two genes (condition attributes).
- Select the set that *Maximizes* F = J1 – J2.

Algorithm:

- Generate FEPM for all individual genes.
- Calculate relevance of each gene I(P,R).
- Generate resultant FEPM between each P of the selected genes and each Q of remaining genes.
- Calculate redundancy I(P, Q) between P and Q .
- Select gene Q from remaining genes that maximizes "Relevance of Q – average redundancy between Q and selected genes".

Performance of the method is demonstrated in Figs. 13 and 14 using the mutual information measure, as an example**,** for five binary class cancer data sets, namely, breast cancer, leukemia, colon cancer, rheumatoid arthritis versus osteoarthritis (RAOA) and RA versus healthy controls (RAHC). In each case FEPM based approach in computing the said measure is compared with Parzen window based and discretization based techniques. Maximum 50 genes are selected. Highest classification accuracy obtained and the number of genes required to obtain that are plotted. SVM (with leave-one-out method) was used to compute the classification accuracy. In most of the cases, higher or same accuracy with lower number of genes is seen to be obtained with FEPM.

X. Conclusion

Granulation is a process like self-reproduction, self-organization, functioning of brain, Darwinian evolution, group behavior, cell membranes and morphogenesis - that are abstracted from *natural phenomena*. *Fuzzy*-granulation or *f*-granulation is inherent in human thinking and reasoning process, and plays an essential role in human cognition. The article deals with rough-fuzzy granular approach in natural computing framework. The concept of knowledge encoding using rough sets and the role of *f*-granulation to make it more efficient are illustrated. Examples of judicious integration, viz., rough-fuzzy case generation, rough-fuzzy classification, rough-fuzzy c-medoids and rough-fuzzy entropy measures with their merits and characteristics are described. The bioinformatics problems of protein sequence analysis for determining bio-bases using rough-fuzzy clustering, and gene selection from microarray data using *f*-information measures on fuzzy equivalence partition matrices are considered. Class dependent granulation with neighborhood rough set has better class discrimination ability than class dependent granulation with Pawlak's rough set. The algorithm is useful in scarcity of training samples. The effect of granules in improving the quality of image segmentation *vis-a-vis* fuzzy entropic segmentation is established. Performance wise rough-fuzzy c-medoids clustering is superior to its hard, rough and fuzzy clustering versions in selecting bio-bases. FEPM based information measures provide higher or same accuracy with lower number of genes selected. Further references on these issues are available in [29]-[34].

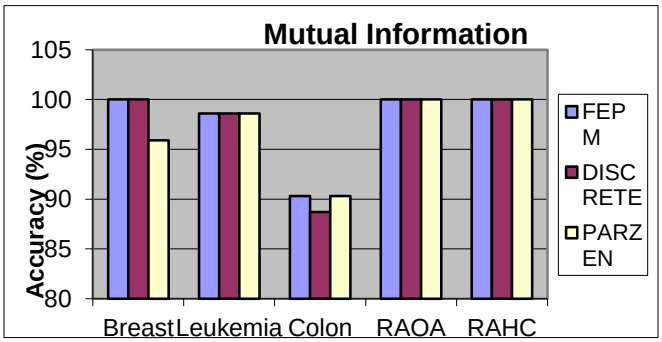


Fig. 13 SVM based classification accuracy with selected genes using FEPM, Discrete and Parzen window techniques.

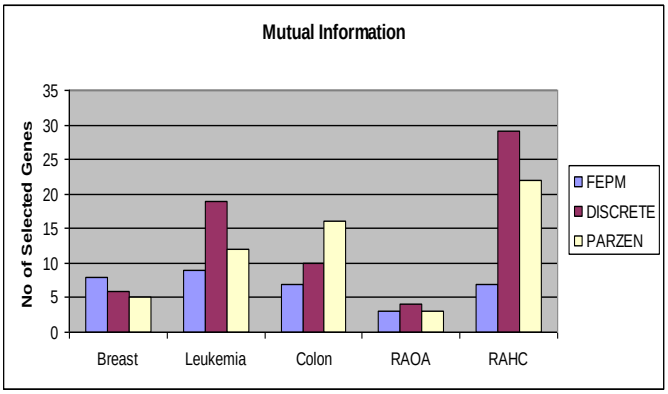


Fig. 14 Number of genes required to obtain highest accuracy.

The methodologies described here basically provide new machine learning modules. Although some specific applications are demonstrated, they can be applied to other real life problems application of these rough-fuzzy methodologies and the underlying concepts in modeling *f*-granularity characteristics of computational theory of perception (CTP) [1, 2] constitutes a challenging task to future researchers.

ACKNOWLEDGMENT

The author acknowledges J.C. Bose Fellowship of the Govt. of India, as well as the technical support of Mr. Avatharam Ganivada.

REFERENCES

[1] L. Kari, G. Rozenberg, The many facets of natural computing, *Communications of the ACM*, vol. 51, no. 10, pp. 72–83, 2008.

[2] L. A. Zadeh, A new direction in AI: Toward a computational theory of perceptions, *AI Magazine*, vol. 22, pp. 73-84, 2001

[3] S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, Wiley, New York, 1999.

[4] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic, Dordrecht, 1991.

[5] S. K. Pal and A. Skowron, Eds., *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, Springer-Verlag, Singapore, 1999.

[6] M. Banerjee, S. Mitra and S. K. Pal, Rough fuzzy MLP: knowledge encoding and classification. *IEEE Trans. on Neural Networks*, vol. 9, no. 6, pp. 1203-1216, 1998.

[7] S。K. Pal, S. Mitra, P. Mitra, Rough-fuzzy MLP: Modular evolution, rule generation, and evaluation. *IEEE Trans. Knowledge and Data Engineering* , vol. 15, no.1, pp. 14-25, 2003.

[8] A. Ganivada, S. Dutta and S. K. Pal, Fuzzy rough granular neural networks, fuzzy granules and classification, *Theoretical Computer Science*, vol. 412, no.42, pp. 5834-5853, 2011.

[9] A. Ganivada, S. S. Ray and S. K. Pal, Fuzzy rough granular self-organizing map and fuzzy rough entropy, *Theoretical Computer Science*, (in press).

[10] S. K. Pal and A. Skowron, Eds., Special issue on rough sets, pattern recognition and data mining, *Pattern Recognition Letters*, vol. 24, no. 6, 2003.

[11] S. K. Pal, L. Polkowski and A. Skowron, Eds., *Rough-Neuro Computing: A Way to Computing with Words*, Springer, Berlin, 2003.

[12] J. L. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.

[13] S. K. Pal, T. S. Dillon, and D. S. Yeung, Eds., *Soft Computing in Case Based Reasoning*, Springer, London, 2001.

[14] S. K. Pal and S. C. K. Shiu, *Foundations of Soft Case Based Reasoning*, John Wiley, New York, 2003.

[15] S. K. Pal and P. Mitra, Case generation using rough sets with fuzzy discretization, *IEEE Trans. Knowledge and Data Engineering*, vol. 16, no. 3, pp. 292-300, 2004.

[16] S. K. Pal and P. Mitra, *Pattern Recognition Algorithms for Data Mining*, Chapman & Hall CRC Press, Boca Raton, FL, 2004.

[17] Y. Li, S. C. K. Shiu and S. K. Pal, Combining feature reduction and case selection in building CBR classifiers, *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 3, pp. 415-429, 2006.

[18] S. K. Pal, S. K. Meher and S. Dutta, Class-dependent rough-fuzzy granular space, dispersion index and classification, *Pattern Recognition*, vol. 45, no. 7, pp.2690-2707, 2012.

[19] T.Y. Lin, Granulation and Nearest Neighborhoods: Rough Set Approach, in: W. Pedrycz (Ed.), *Granular Computing: An Emerging Paradigm*, Physica- Verlag, Heidelberg, Germany, pp. 125–142, 2001.

[20] S. K. Pal, A. Ghosh, and B. U. Sankar, Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation, *International Journal of Remote Sensing*, vol. 2, no. 11, pp. 2269-2300, 2000.

[21] P. Maji and S. K. Pal, Rough set based generalized fuzzy C-means algorithm and quantitative indices, *IEEE Trans. on System, Man and Cybernetics, Part B*, vol. 37, no. 6, pp. 1529--1540, 2007.

[22] P. Maji and S. K. Pal, *Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging*, John Wiley, New York, 2012.

[23] P. Lingras and C. West, Interval set clustering of web users with rough K-means, *Journal of Intelligent Information Systems*, vol. 23, no. 1, pp. 5-16, 2004.

[24] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, New York: Plenum, 1981.

[25] P. Maji and S. K. Pal, Rough-fuzzy C-medoids algorithm and selection of bio-basis for amino acid sequence analysis, *IEEE Trans. Knowledge and Data Engineering*, vol. 19, no. 6, pp. 859-872, 2007.

[26] D. Sen and S.K. Pal, Generalized rough sets, entropy and image ambiguity measures, *IEEE Transactions on System, Man and Cybernetics - Part B*, vol.39, no.1, pp.117-128, 2009.

[27] S. K. Pal, R. A. King and A. A. Hashim, Automatic gray level thresholding through index of fuzziness and entropy, *Pattern Recognition Letters*, vol. 1, pp.141-146, 1983.

[28] P. Maji and S. K. Pal, Feature selection using f-information measures in fuzzy approximation spaces, *IEEE Trans. Knowledge and Data Engineering,* vol. 22, no. 6, pp. 854--867, 2010.

[29] S. Bandyopadhyay and S. K. Pal, *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*, Springer, Berlin, 2007.

[30] D. Sen and S. K. Pal, Histogram thresholding using fuzzy and rough measures of association error, *IEEE Trans. Image Processing*, vol. 18, no.4, pp. 879-888, 2009.

[31] S. K. Pal, J. Peters, *Rough Fuzzy Image Analysis: Foundations and Methodologies*, CRC Press, Boca Raton, Florida, 2010.

[32] S.K. Pal, A. Petrosino and L. Maddalena, Eds., *Handbook on Soft Computing for Video Surveillance*, CRC Press, Boca Raton, Florida, 2012.

[33] A. Albanese, S. K. Pal and A. Petrosino, Rough set, kernel set and spatio-temporal outlier detection, *IEEE Trans. Knowledge and Data Engineering* (in press).

[34] D. Chakraborty, B. Uma Shankar and S.K. Pal, Granulation, rough entropy and spatiotemporal moving object detection, *Applied Soft Computing* (in press).

Making Good Decisions Quickly

Sven Koenig, *Senior Member, IEEE*

***Abstract*—Several disciplines, including artificial intelligence, operations research and many others, study how to make good decisions. In this overview article, we argue that the key to making progress in our research area is to combine their ideas, which often requires serious technical advances to reconcile their different assumptions and methods in a way that results in synergy among them. To illustrate this point, we give a broad overview of our ongoing research on search and planning (with a large number of students and colleagues, both at the University of Southern California and elsewhere) to demonstrate how to combine ideas from different decision making disciplines. For example, we describe how to combine ideas from artificial intelligence, operations research, and utility theory to create the foundations for building decision support systems that fit the risk preferences of human decision makers in high-stake one-shot decision situations better than current systems. We also describe how to combine ideas from artificial intelligence, economics, theoretical computer science and operations research to build teams of robots that use auctions to distribute tasks autonomously among themselves, and give several more examples.**

***Index Terms*—agents, ant robotics, artificial intelligence, auction-based coordination, decision theory, dynamic programming, economics, freespace assumption, goal-directed navigation, greedy online planning, heuristic search, high-stake one-shot decision making, incremental heuristic search, Markov decision processes, multi-agent systems, nonlinear utility functions, operations research, planning, real-time heuristic search, reinforcement learning, risk preferences, robotics, scarce resources, sequential-single item auctions, terrain coverage, utility theory.**

I. Introduction

ARTIFICIAL INTELLIGENCE is rooted in building cognitive systems (that is, systems that operate in a way similar to the human mind) but today is more and more about engineering intelligent systems (that is, systems that solve tasks that require difficult decisions) even if these systems do not operate in a way similar to the human mind. For example, the popular textbook "Artificial Intelligence: A Modern Approach" [52] by Stuart Russell and Peter Norvig views artificial intelligence as the science of creating rational agents, where agents are control systems that interact with an environment. They can sense to gather information about the state of the environment and execute actions to change it. Rational agents, according to the textbook, should select actions that are expected to maximize given performance measures. In general, agents must be able to make good decisions in complex situations that involve a substantial degree of uncertainty, yet find solutions in a timely manner. Researchers from artificial intelligence therefore create a strong foundation for building such agents, typically focusing more on autonomous decision making and optimization than modeling of complex decision problems or providing decision support for human decision makers.

Sven Koenig is with the Department of Computer Science, University of Southern California, Los Angeles, CA, 90089-0781, USA, e-mail: skoenig@usc.edu, web page: idm-lab.org.

Artificial intelligence has developed tools for building agents that perform well with respect to given performance measures. Other decision making disciplines provide different and potentially complementary tools. In general, the larger one's toolbox, the more decision problems one is able to tackle. By combining ideas from different decision making disciplines, one can expect to improve on existing tools and build new tools that either perform better than existing ones or solve decision problems that existing tools cannot solve. This provides an incentive to study different decision making disciplines, develop curricula that allow students to learn about several decision making disciplines, and create a universal science of intelligent decision making that combines ideas from different decision making disciplines, including artificial intelligence, operations research, economics, decision theory, and control theory. One obstacle that needs to be overcome is that different decision making disciplines typically study different applications and thus make different assumptions, resulting in different decision making methods. Combining ideas from different decision making disciplines therefore often requires serious technical advances to reconcile the different assumptions and methods in a way that results in synergy among them. A second obstacle is that that different decision making disciplines focus on different aspects of decision problems and have different ideas about what constitutes a good solution to a given decision problem, often due to the disciplinary training of their researchers. For example, statistics researchers often tend to focus on the uncertainty in the data and how it can be resolved; optimization researchers often tend to assume that the data is correct and focus on finding optimal or close to optimal (rather than timely) solutions (concentrating on "planning" rather than "operations"); and artificial intelligence researchers often tend to focus on the ability of agents to make good decisions online, taking into account the limitations of the agents (such as their limited sensing, computational and communication capabilities as well as their noisy actuation) in addition to their interaction with the environment (such as information collection) and each other (such as coordination), which explains the title of this overview article. A third obstacle is that different decision making disciplines often use different terminology and notation. Multi-disciplinary training can overcome these obstacles and transform the second obstacle into a strength.

Artificial intelligence often pursues general principles that apply widely to decision making and problem solving (rather than problem-specific methods), perhaps due to its roots in building cognitive systems. It is therefore not surprising that artificial intelligence, over time, has incorporated ideas from other decision making disciplines. For example, the third

edition of "Artificial Intelligence: A Modern Approach" covers local search in Chapter 4, including hill-climbing search, simulated annealing, local beam search, and genetic algorithms. It covers utility theory in Chapter 16, including utility functions, multi-attribute utility functions, and influence diagrams. It covers sequential decision problems in Chapter 17, including Markov decision processes and dynamic programming methods such as value and policy iteration. It covers game theory in Chapter 17, including single-move, repeated, and sequential games. It also covers mechanism design in the same chapter, including auctions. All of these topics have also been studied in other decision making disciplines, such as operations research and economics, and typically originated there. For example, researchers from artificial intelligence discovered totally and partially observable Markov decision processes from operations research when working on foundations for decision theoretic planning and reinforcement learning and then, for example, developed new ways of representing and solving them by incorporating insights from knowledge representation and planning (where states are typically represented as collections of facts), resulting in both symbolic and structured dynamic programming. Symbolic dynamic programming, for example, is a generalization of dynamic programming for solving Markov decision processes that exploits symbolic structure in the solution of relational and first-order logical Markov decision processes to avoid the full state and action enumeration of classical dynamic programming methods [54].

Outsiders often do not know about these and other recent achievements of artificial intelligence and, for this reason, might not appreciate the ideas that it has to offer to them. There exist some established but narrow interfaces between artificial intelligence and other decision making disciplines. An example of a step in the direction of an interface between artificial intelligence and control theory is [8]. An example of a step in the direction of an interface between artificial intelligence and operations research is the International Conference on Integration of Artificial Intelligence and Operations Research Techniques in Constraint Programming for Combinatorial Optimization Problems (CPAIOR), which is by now an established conference series with 9 conferences since 2004, preceeded by 5 workshops. Similarly, ILOG eventually integrated software for constraint programming and linear optimization. In general, however, artificial intelligence probably needs to reach out even more to other decision making disciplines with the objective to inform them and create a universal science of intelligent decision making. While this might appear to be an obvious objective, progress in this direction has been made mostly recently. For example, an algorithmic decision theory community formed around 2000 and eventually created the International Conference on Algorithmic Decision Theory (ADT). The First International Conference on Algorithmic Decision Theory took place in Venice, Italy, in 2009, and the Second International Conference on Algorithmic Decision Theory took place in New Brunswick, USA, in 2011. The conference series, according to the conference announcement at www.adt2011.org, involves researchers from such disparate fields as decision theory, discrete mathematics, theoretical computer science, economics, and artificial intelligence, aiming to improve decision support in the presence of massive data bases, partial and/or uncertain information, and distributed decision makers. Papers have covered topics from computational social choice to preference modeling, from uncertainty to preference learning, and from multi-criteria decision making to game theory [51].

We sketch some of our own research in the remainder of this overview article to illustrate why we believe that it is important to combine ideas from different decision making disciplines. Not surprisingly, our research centers around methods for decision making (planning and learning) that enable single agents and teams of agents to act intelligently in their environments and exhibit goal-directed behavior in real-time, even if they have only incomplete knowledge of their environment, imperfect abilities to manipulate it, limited or noisy perception or insufficient reasoning speed. Our research group, the Intelligent Decision Making group, develops new decision making methods, implements them and studies their properties theoretically and experimentally. We demonstrated around 1995 that it is possible to combine ideas from different decision making disciplines by developing a robot navigation architecture based on partially observable Markov decision processes from operations research that allows robots to navigate robustly despite a substantial amount of actuator and sensor uncertainty, which prevents them from knowing their precise location during navigation [27]. This research resulted in a reliable robot architecture that overcomes the deficiencies of purely topological or metric navigation methods [58]. Since then, our research group has continued to combine ideas from different decision making disciplines. In the following, we describe some of these research directions in more detail. While they might appear diverse, there is a common underlying thrust, namely to bring about advances that extend the reach of search (in a broad sense, including heuristic search, hill-climbing and dynamic programming), and to apply the results to robot navigation.

II. Example: Nonlinear Utility Functions

Finding plans that maximize the expected utility for nonlinear utility functions is important in both high-stake one-shot decision situations and decision situations with scarce resources [7].

- In high-stake one-shot decision situations, huge gains or losses of money or equipment are possible, and human decision makers take risk aspects into account. Risk-averse decision makers, for example, tolerate a smaller expected plan-execution reward for a reduced variance (although this explanation is a bit simplified). For example, they try to avoid huge losses when fighting forest fires, containing marine oil spills or controling autonomous spacecraft (and other decision problems that artificial intelligence researchers study) and thus add more sensing operations than necessary to maximize the expected reward [26]. Planning systems need to reflect these risk preferences. Bernoulli and Von Neumann/Morgenstern's utility theory [60] [4] suggests that rational human decision makers choose plans that max-

imize the expected utility, where the utility is a monotonically increasing function of the reward. For example, exponential utility functions completely preserve the structure of planning tasks because they are the only class of nonlinear utility functions for which decisions do not depend on the accumulated reward. However, one-switch utility functions often model the risk attitudes of human decision makers better than exponential utility functions [2].

- In decision situations with scarce resources, there are often limits to how much of a resource (such as time, energy or memory) can be consumed before it runs out. For example, a lunar rover that reaches a science target with minimal expected energy consumption does not necessarily maximize the probability of achieving it within its battery limit. Resource limits can be modeled with monotonically (but perhaps not strictly monotonically) increasing utility functions that map total rewards (the negative of the total resource consumptions) to real values. For example, a hard resource limit can be modeled with a step function that is zero to the left of the negative resource limit (where the total resource consumption is greater than the resource limit) and one to the right of it [15].

Decision-theoretic planning methods in artificial intelligence are these days typically, either explicitly or implicitly, based on Markov decision processes. One can use dynamic programming methods, such as value iteration [3] or policy iteration [21], to maximize the expected total (undiscounted or discounted) reward. One can also use these methods to maximize the expected utility for nonlinear utility functions (studied in the context of risk-sensitive Markov decision processes in operations research [22] and control theory [39]) but then, except for exponential utility functions, needs to add the accumulated reward to the states, which increases the number of states substantially. We and other artificial intelligence researchers have therefore studied "functional" versions of value and policy iteration that do not maintain a value for each augmented state but rather a value function for each original state (that maps the total reward to the value of the state) and operate directly on these value functions [6] [48] [12] [34] [43], which allows one to solve larger decision problems than what would be possible otherwise due to the following advantages: First, the value functions can sometimes be represented exactly and compactly (that is, with a finite number of parameters), as we have shown for one-switch utility functions [36] [38] and piecewise linear utility functions with optional exponential tails [37]. Second, the value functions can also be approximated to a desired degree (for example, with piecewise linear functions), sometimes resulting in approximation guarantees, which allows one to trade off between runtime and memory consumption on one hand and solution quality on the other hand [37]. More complex decision problems can be solved in a similar way. For example, a lunar rover might have to maximize its science return within its battery limit despite uncertainty about its energy consumption, when scientists have designated several locations that the rover can visit to perform science experiments and assigned a science return value to each of them [40]. Other approaches also exist [41], together with extensions to teams of robots [42].

Methods from artificial intelligence exploit the structure of decision-theoretic planning tasks [45]. For example, artificial intelligence has investigated how to represent search spaces implicitly and exploit the resulting decomposability to solve Markov decision processes efficiently without having to enumerate their state spaces completely. For instance, structured versions of value iteration represent the transition policies in factored form, which allows them to represent policies more compactly than with tables to speed up their computations and generalize policies across states [5]. An example is SPUDD, that uses algebraic decision diagrams instead of tables [20]. Artificial intelligence has also investigated forward search methods that, different from value and policy iteration, consider only states that are reachable from the start state. For instance, LAO* uses heuristic search to restrict the value updates only to relevant states rather than all states [16] [44]. We have generalized these methods to find plans that maximize the expected utility for nonlinear utility functions [35]. Other decision making disciplines have developed other ways of exploiting the structure of decision-theoretic planning tasks [49], meaning that there are opportunities for combining different ideas. Overall, this research combines insights from artificial intelligence, operations research, and utility theory for planning with nonlinear utility functions. Operations research has studied the properties of Markov decision processes in detail, artificial intelligence and operations research contribute ideas for solving them, and utility theory provides a realistic optimization criterion for high-stake one-shot decision situations.

III. Example: Auction-Based Coordination

Centralized control is often inefficient for teams of robots in terms of the amount of communication and computation required since the central controller is the bottleneck of the system. Researchers from artificial intelligence and robotics have therefore studied robot coordination with cooperative auctions [9]. An auction is "a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants" [46]. Auctions have been developed for the allocation of resources in situations where agents have different utilities and private information. Auctions are therefore promising decentralized methods for teams of robots to allocate and re-allocate tasks in real-time among themselves in dynamic, partially known and time-constrained domains with positive or negative synergies among tasks. Furthermore, the short length of a bid is helpful when communication bandwidth is limited. Artificial intelligence and later robotics have explored auction-based coordination systems at least since the introduction of contract networks [55], mostly from an experimental perspective. In auction-based coordination systems, the bidders are robots, and the items up for auction are tasks to be executed by the robots. All robots bid their costs. Thus, the robot with the smallest bid

cost is best suited for a task. All robots then execute the tasks that they win. Auction-based coordination systems are easy to understand, simple to implement and broadly applicable. They promise to be efficient both in communication (since robots communicate only essential summary information) and in computation (since robots compute their bids in parallel). A typical application is multi-robot routing [10], where a team of robots has to visit given targets and repeatedly reassigns targets among the robots as it learns more about the initially unknown terrain, as robots fail or as additional targets get introduced. Examples include environmental clean up, mine clearing, space exploration, and search-and-rescue. Multi-robot routing problems are NP-hard to solve optimally even if the locations of obstacles, targets, and robots are initially known and (except for the locations of the robots) do not change [32]. Their similarity to traveling salesperson problems [33] allows one to use insights from theoretical computer science and operations research for their analysis. Economics has an extensive auction literature but its agents are rational and competitive, leading to long decision cycles, strategic behavior, and possibly collusion. Such issues do not arise in auction-based coordination systems because the robots faithfully execute their programs. On the other hand, auction-based coordination systems must operate in real-time. Still, some insights from economics can be exploited for building them, such as the concepts of synergy and different auction mechanisms, including parallel, combinatorial, and sequential single-item (SSI) auctions. For example, SSI auctions proceed in several rounds, assigning one additional target per round to some robot. We have exploited the fact that SSI auction-based coordination systems with marginal-cost bidding [53] perform a form of hill-climbing search to analyze the resulting team performance [59]. We have used tools from theoretical computer science to show that SSI auction-based coordination systems can provide constant factor performance guarantees even though they run in polynomial time and, more generally, that they combine advantageous properties of parallel and combinatorial auctions [32], resulting in one of the few existing performance analyses. Some intuition for this result can be gained from interpreting the greedy construction of minimum spanning trees as a cooperative auction [31]. We have investigated several versions of SSI auctions to build SSI auction-based coordination systems that increase the team performance while still allocating targets to robots in real-time. For example, we have generalized auction-based coordination systems based on SSI auctions to assign more than one additional target during each round (called the bundle size), which increases their similarity with combinatorial auctions by taking more synergies among targets into account and making the resulting hill-climbing search less myopic. We have shown that, for a given number of additional targets to be assigned during each round, every robot needs to submit only a constant number of bids per round and the runtime of winner determination is linear in the number of robots [29]. Thus, the communication and winner determination times do not depend on the number of targets, which helps the resulting auction-based coordination systems to scale up to a large number of targets for small bundle sizes. Overall, this research combines insights from artificial intelligence, economics, theoretical computer science and operations research for the development of auction-based coordination systems and their analysis [23].

IV. Example: Fast Replanning

Robots often operate in domains that are only incompletely known or change over time. One way of dealing with incomplete information is to interleave search with action execution. In this case, the robots need to replan repeatedly. To make search fast, one can use heuristic search methods with limited lookahead (agent-centered search, such as real-time heuristic search [30]) or heuristic search methods that reuse information from previous searches (incremental heuristic search). Consider, for example, a robot that has to move from its current location to given goal coordinates in initially unknown terrain. The robot does not know the locations of obstacles initially but observes them within its sensor radius and adds them to its map. Planning in such non-deterministic domains is typically time-consuming due to the large number of contingencies, which provides incentive to speed up planning by sacrificing the optimality of the resulting plans. Greedy online planning methods interleave planning and plan execution to allow robots to gather information early and then use the acquired information right away for replanning, which reduces the amount of planning performed for unencountered situations. For example, goal-directed navigation with the freespace assumption is a common-sense version of assumption-based planning that is popular in robotics for moving a robot to a given goal location in initially unknown terrain [47] and can be analyzed with tools from theoretical computer science [28]. It finds a short (unblocked) path from the current location of the robot to the goal location given its current knowledge of the locations of obstacles under the assumption that the terrain is otherwise free of obstacles. If such a path does not exist, it stops unsuccessfully. Otherwise, the robot follows the path until it either reaches the goal location, in which case it stops successfully, or observes the path to be blocked, in which case it repeats the process using its revised knowledge of the locations of obstacles. Incremental heuristic search methods solve such series of similar path planning problems often faster than searches from scratch [17] (by reusing information from previous searches to speed up their current search), yet differ from other replanning methods (such as planning by analogy) in that their solution quality is as good as the solution quality of searches from scratch [25]. The first incremental heuristic search methods was published in artificial intelligence and robotics [56]. It has been discovered since then that incremental search had been studied much earlier already (for example, in the context of dynamic shortest path problems in algorithms), which allowed us to develop a new incremental heuristic search method by combining ideas from different disciplines. D* Lite [24] is now a popular incremental heuristic search method for planning with the freespace assumption that combines ideas from incremental search (namely, to recalculate only those start distances that can have changed or have not been calculated before) with ideas from heuristic search (namely, to use approximations of the goal distances

to recalculate only those start distances that are relevant for recalculating a shortest path). In particular, it combines ideas behind DynamicSWSF-FP [50] from algorithms with ideas behind A* from artificial intelligence. Overall, this research combines insights from artificial intelligence, robotics, and theoretical computer science for the development of fast replanning methods and their analysis.

V. Example: Ant Robots

Researchers from robotics are interested in simple robots with limited sensing and computational capabilities as well as noisy actuation. Such ant robots have the advantage that they are easy to program and cheap to build. This makes it feasible to deploy groups of ant robots and take advantage of the resulting fault tolerance and parallelism. Researchers from robotics had studied robots that can follow trails laid by other robots but we studied robots that leave trails in the terrain to cover closed terrain (that is, visit each location) once or repeatedly, as required for surveillance, guarding terrain, mine sweeping, and surface inspection. Ant robots cannot use conventional planning methods due to their limited sensing and computational capabilities. To overcome these limitations, we developed navigation methods that leave markings in the terrain, similar to the pheromone trails of real ants. These markings are shared among all ant robots and allow them to cover terrain even if they do not have any kind of memory, cannot maintain maps of the terrain, nor plan complete paths. They can be used by single ant robots as well as groups of ant robots and provide robustness in situations where some ant robots fail, ant robots are moved without realizing this, the trails are of uneven quality, and some trails are destroyed. Robot architectures based on partially observable Markov decision processes provide robots with the best possible location estimate to overcome actuator and sensor uncertainty, while ant robots achieve their goals without ever worrying about where they are in the terrain. We built physical ant robots that cover terrain and test their design both in realistic simulation environments and on a Pebbles III robot. We modeled the coverage strategy of such ant robots with graph dynamic programming methods that are similar to real-time heuristic search methods (such as Learning Real-Time A*) [30] and reinforcement learning methods (such as Real-Time Dynamic Programming) [1] from artificial intelligence (except that the values are written on the floor rather than stored in memory), which allowed us to use tools from theoretical computer science to analyze their behavior [57]. Other researchers, such as Israel Wagner and his collaborators, have similar interests and work on the intersection of robotics, artificial intelligence, and theoretical computer science [61], see also http://www.cs.technion.ac.il/~wagner/. Overall, this research combines insights from artificial intelligence, robotics, biology, and theoretical computer science for the development of navigation methods for ant robots and their analysis.

VI. Example: Terrain Coverage

Robot coverage of known terrain can be sped up with multiple robots that coordinate explicitly. Researchers from robotics had investigated spanning tree-based coverage methods in unweighted terrain, where the travel times of robots are the same everywhere in the terrain. Single-robot coverage problems are solved with minimal cover times by Spanning Tree Coverage (STC), a polynomial-time single-robot coverage method published in robotics and artificial intelligence that decomposes terrain into cells, finds a spanning tree of the resulting graph, and makes the robot circumnavigate it [13] [14]. This method had been generalized to Multi-Robot Spanning Tree Coverage (MSTC), a polynomial-time multi-robot coverage method published in robotics [18] [19]. While MSTC provably improves the cover times compared to STC, it cannot guarantee its cover times to be small. We showed that solving several versions of multi-robot coverage problems with minimal cover times is NP-hard, which provides motivation for designing polynomial-time constant-factor approximation methods. We generalized STC to Multi-Robot Forest Coverage (MFC), a polynomial multi-robot coverage method based on a method published in operations research [11] (in the context of deciding where to place nurse stations in hospitals) for finding tree covers with trees of balanced weights, one tree for each robot. We also generalized MFC from unweighted terrain to weighted terrain, where the travel times of robots are not the same everywhere. The cover times of MFC in weighted and unweighted terrain are at most about sixteen times larger than minimal and experimentally close to minimal in all tested scenarios [62]. Overall, this research combines insights from artificial intelligence, robotics, and operations research for the development of terrain coverage methods and their analysis.

VII. Conclusions

In this overview article, we described some of our own research to illustrate why we believe that it is important to combine ideas from different decision making disciplines. We are convinced that we have overlooked lots of developments but encourage researchers from artificial intelligence to continue to reach out to other decision making disciplines with the objective to inform them about our latest research and help to make progress towards a universal science of intelligent decision making.

Acknowledgments

We would like to thank the chairs and organizers of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology 2012 for inviting us to give a talk, the editors of the IEEE Intelligent Informatics Bulletin for allowing us to summarize our thoughts in this overview article, and operations researcher Craig Tovey for giving extensive comments on a draft of this article, which now includes several of his ideas. The research summarized in this article is based on a number of interdisciplinary collaborations with a large number of co-authors (including colleagues and students), whose substantial contributions we would like to acknowledge. The overall perspective is novel to this publication while the research synopses re-use larger portions of earlier publications, such as [62] and [23], and the web pages of the author. The perspective is based upon work supported by NSF (while the

author served as program director there), ARL/ARO under contract/grant number W911NF-08-1-0468 and ONR in form of a MURI under contract/grant number N00014-09-1-1031. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the sponsoring organizations, agencies or the U.S. government. We apologize for the necessary generalizations, resulting in research stereotypes, and our inability to include references to all relevant research - there are just too many of them.

REFERENCES

[1] A. Barto, S. Bradtke, and S. Singh. Learning to act using real-time dynamic programming. *Artificial Intelligence*, 72(1–2):81–138, 1995.

[2] D. Bell. One-switch utility functions and a measure of risk. *Management Science*, 34(12):1416–1424, 1988.

[3] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[4] D. Bernoulli. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, 1738. Translated by L. Sommer, *Econometrica*, 22: 23–36, 1954.

[5] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.

[6] J. Boyan and M. Littman. Exact solutions to time-dependent MDPs. In *Advances in Neural Information Processing Systems*, volume 13, pages 1026–1032, 2000.

[7] J. Bresina, R. Dearden, N. Meuleau, S. Ramakrishnan, D. Smith, and R. Washington. Planning under continuous time and resource uncertainty: A challenge for AI. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 77–84, 2002.

[8] T. Dean. *Planning and Control*. M. Kaufmann Publishers, 1991.

[9] M. Dias, R. Zlot, N. Kalra, and A. Stentz. Market-based multirobot coordination: A survey and analysis. *Proceedings of the IEEE*, 94(7):1257–1270, 2006.

[10] M. Dias, R. Zlot, N. Kalra, and A. Stentz. Market-based multirobot coordination: A survey and analysis. *Proceedings of the IEEE*, 94(7):1257–1270, 2006.

[11] G. Even, N. Garg, J. Könnemann, R. Ravi, and A. Sinha. Min-max tree covers of graphs. *Operations Research Letters*, 32:309–315, 2004.

[12] Z. Feng, R. Dearden, N. Meuleau, and R. Washington. Dynamic programming for structured continuous Markov decision problems. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, pages 154–161, 2004.

[13] Y. Gabriely and E. Rimon. Spanning-tree based coverage of continuous areas by a mobile robot. *Annals of Mathematics and Artificial Intelligence*, 31:77–78, 2001.

[14] Y. Gabriely and E. Rimon. Spanning-tree based coverage of continuous areas by a mobile robot. In *Proceedings of the International Conference on Robotics and Automation*, pages 1927–1933, 2001.

[15] P. Haddawy and S. Hanks. Utility models for goal-directed decision-theoretic planners. *Computational Intelligence*, 14(3):392–429, 1998.

[16] E. Hansen and S. Zilberstein. LAO*: a heuristic search algorithm that finds solutions with loops. *Artificial Intelligence*, 129:35–62, 2001.

[17] P. Hart, N. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, SCC-4(2):100–107, 1996.

[18] N. Hazon and G. Kaminka. Redundancy, efficiency, and robustness in multi-robot coverage. In *Proceedings of the International Conference on Robotics and Automation*, pages 735–741, 2005.

[19] N. Hazon and G. Kaminka. On redundancy, efficiency, and robustness in coverage for multiple robots. *Robotics and Autonomous Systems*, 56(12):1102–1114, 2008.

[20] J. Hoey, R. Aubin, and C. Boutilier. SPUDD: stochastic planning using decision diagrams. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, pages 279–288, 1999.

[21] R. Howard. *Dynamic Programming and Markov Processes*. MIT Press, 1960.

[22] R. Howard and J. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, 1972.

[23] S. Koenig, P. Keskinocak, and C. Tovey. Progress on agent coordination with cooperative auctions [senior member paper]. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2010.

[24] S. Koenig and M. Likhachev. Fast replanning for navigation in unknown terrain. *Transactions on Robotics*, 21(3):354–363, 2005.

[25] S. Koenig, M. Likhachev, Y. Liu, and D. Furcy. Incremental heuristic search in artificial intelligence. *Artificial Intelligence Magazine*, 25(2):99–112, 2004.

[26] S. Koenig and Y. Liu. Sensor planning with non-linear utility functions. In *Proceedings of the European Conference on Planning*, pages 265–277, 1999.

[27] S. Koenig and R. Simmons. Xavier: A robot navigation architecture based on partially observable markov decision process models. In D. Kortenkamp, R. Bonasso, and R. Murphy, editors, *Artificial Intelligence Based Mobile Robotics: Case Studies of Successful Robot Systems*, pages 91–122. MIT Press, 1998.

[28] S. Koenig, Y. Smirnov, and C. Tovey. Performance bounds for planning in unknown terrain. *Artificial Intelligence Journal*, 147(1–2):253–279, 2003.

[29] S. Koenig, C. Tovey, X. Zheng, and I. Sungur. Sequential bundle-bid single-sale auction algorithms for decentralized control. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1359–1365, 2007.

[30] R. Korf. Real-time heuristic search. *Artificial Intelligence*, 42(2–3):189–211, 1990.

[31] M. Lagoudakis, P. Keskinocak, A. Kleywegt, and S. Koenig. Auctions with performance guarantees for multi-robot task allocation. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 1957–1962, 2004.

[32] M. Lagoudakis, V. Markakis, D. Kempe, P. Keskinocak, S. Koenig, A. Kleywegt, C. Tovey, A. Meyerson, and S. Jain. Auction-based multi-robot routing. In *Proceedings of the International Conference on Robotics: Science and Systems*, pages 343–350, 2005.

[33] E. Lawler, J. Lenstra, A. Kan, and D. Shmoys, editors. *The Traveling Salesman Problem*. John Wiley, 1985.

[34] L. Li and M. Littman. Lazy approximation for solving continuous finite-horizon MDPs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1175–1180, 2005.

[35] Y. Liu. *Decision-Theoretic Planning under Risk-Sensitive Planning Objectives*. PhD thesis, College of Computing, Georgia Institute of Technology, Atlanta (Georgia), 2005.

[36] Y. Liu and S. Koenig. Risk-sensitive planning with one-switch utility functions: Value iteration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 993–999, 2005.

[37] Y. Liu and S. Koenig. Functional value iteration for decision-theoretic planning with general utility functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1186–1193, 2006.

[38] Y. Liu and S. Koenig. An exact algorithm for solving MDPs under risk-sensitve planning objectives with one-switch utility functions. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 453–460, 2008.

[39] S. Marcus, E. Fernàndez-Gaucherand, D. Hernàndez-Hernàndez, S. Colaruppi, and P. Fard. Risk-sensitive Markov decision processes. In C. Byrnes et. al., editor, *Systems and Control in the Twenty-First Century*, pages 263–279. Birkhauser, 1997.

[40] J. Marecki, S. Koenig, and M. Tambe. A fast analytical algorithm for solving Markov decision processes with real-valued resources. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2536–2541, 2007.

[41] J. Marecki and M. Tambe. Towards faster planning with continuous resources in stochastic domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1049–1055, 2008.

[42] J. Marecki and M. Tambe. Planning with continuous resources for agent teams. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 1089–1096, 2009.

[43] J. Marecki and P. Varakantham. Risk-sensitive planning in partially observable environments. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 1357–1364, 2010.

[44] Mausam, E. Benazara, R. Brafman, N. Meuleau, and E. Hansen. Planning with continuous resources in stochastic domains. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1244–1251, 2005.

[45] Mausam and A. Kolobov. *Planning with Markov Decision processes: An AI Perspective*. Morgan and Claypool Publishers, 2012.

[46] P. McAfee and J. McMillan. Auctions and bidding. *Journal of Economic Literature*, 15:699–738, 1987.

[47] I. Nourbakhsh and M. Genesereth. Assumptive planning and execution: A simple, working robot architecture. *Autonomous Robots*, 3:49–67, 1996.

[48] P. Poupart, C. Boutilier, D. Schuurmans, and R. Patrascu. Piecewise linear value function approximation for factored MDPs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 292–299, 2002.
[49] W. Powell. *Approximate Dynamic Programming*. John Wiley and Sons, second edition, 2011.
[50] G. Ramalingam and T. Reps. An incremental algorithm for a generalization of the shortest-path problem. *Journal of Algorithms*, 21:267–305, 1996.
[51] F. Rossi and A. Tsoukias, editors. *Algorithmic Decision Theory, First International Conference, ADT 2009*, LNAI 5783. Springer, 2009.
[52] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, third edition, 2009.
[53] T. Sandholm. An implementation of the contract net protocol based on marginal cost calculations. In *Proceedings of the International Workshop on Distributed Artificial Intelligence*, pages 295–308, 1993.
[54] S. Sanner and K. Kersting. Symbolic dynamic programming. In *Encyclopedia on Machine Learning*, pages 946–954. Springer, 2010.
[55] R. Smith. The contract net protocol: High level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29:1104–1113, 1980.
[56] A. Stentz. The focussed D* algorithm for real-time replanning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1652–1659, 1995.
[57] J. Svennebring and S. Koenig. Building terrain-covering ant robots. *Autonomous Robots*, 16(3):313–332, 2004.
[58] S. Thrun, W. Burgard, and D. Fox, editors. *Probabilistic Robotics*. MIT Press, 2005.
[59] C. Tovey, M. Lagoudakis, S. Jain, and S. Koenig. The generation of bidding rules for auction-based robot coordination. In L. Parker, F. Schneider, and A. Schultz, editors, *Multi-Robot Systems: From Swarms to Intelligent Automata*, pages 3–14. Springer, 2005.
[60] J. von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton University Press, second edition, 1947.
[61] I. Wagner and A. Bruckstein. From ants to a(ge)nts: A special issue on ant-robotics. *Annals of Mathematics and Artificial Intelligence*, 31(1–4):1–5, 2001.
[62] X. Zheng, S. Koenig, D. Kempe, and S. Jain. Multi-robot forest coverage for weighted and unweighted terrain. *IEEE Transactions on Robotics*, 26(6):1018–1031, 2010.

Summarization of Association Rules in Multi-tier Granule Mining

Yuefeng Li, *Member, IEEE,* Jingtong Wu

***Abstract*—It is a big challenge to find useful associations in databases for user specific needs. The essential issue is how to provide efficient methods for describing meaningful associations and pruning false discoveries or meaningless ones. One major obstacle is the overwhelmingly large volume of discovered patterns. This paper discusses an alternative approach called multi-tier granule mining to improve frequent association mining. Rather than using patterns, it uses granules to represent knowledge implicitly contained in databases. It also uses multi-tier structures and association mappings to represent association rules in terms of granules. Consequently, association rules can be quickly accessed and meaningless association rules can be justified according to the association mappings. Moreover, the proposed structure is also an precise compression of patterns which can restore the original supports. The experimental results shows that the proposed approach is promising.**

***Index Terms*—knowledge discovery in databases, association rule mining, granule mining, pattern mining, decision rules, support restoration.**

I. Introduction

THE association mining consists of two phases: pattern mining and rule generation. Many efficient algorithms have been developed for pattern mining; However, the challenging issue for pattern mining is not efficiency but interpretability, due to the huge number of patterns generated by the mining process [33], [18]. Frequent closed patterns partially alleviate the redundancy problem. Recently, many experiments [29], [36], [13], [16] have proved that frequent closed patterns are good alternative of terms for representing text features. Several approaches for pattern post-processing have also been proposed recently. Pattern compression [30], pattern deploying [29] and pattern summarization [33], [24] were proposed to summarize patterns.

The phase of rule generation is to find interesting rules based on discovered patterns and a minimum confidence, which is also a time consuming activity that can generate many redundant rules. The approaches for pruning redundant rules can be roughly divided into two categories, the subjective based approach and objective approach. The former is to find rules that satisfy some constraints or templates [7], [2]. The later is to construct concise representations of rules without applying user-dependent constraints [35], [31].

There are several obstacles when we consider using association mining in applications: the overwhelmingly large volume of discovered patterns and rules, false discoveries, the lack of semantic information along with the mining process, and the incompleteness of knowledge coverage. Frequent association mining has been extended to multilevel association mining, which uses concept hierarchies or taxonomy trees to find rules [8]. The leaves of a taxonomy tree represent items at the lowest level of abstraction. Using a top-down strategy, at each level, frequent patterns are calculated based on accumulated counts. Recently, mining flipping correlations [1] has been proposed to find positive and negative correlations in taxonomy trees. Another paradigm is the filtered-top-k association discovery [28] which used three parameters: a user specified measure of how potential interesting an association is, filters for discarding inappropriate associations, and k the number of associations to be discovered.

One important finding is that the use of closed patterns can greatly reduce the number of extracted rules; however, a considerable amount of redundancy still remains [32]. Therefore, the size of the set of closed patterns need to be further reduced. The summarization approaches can achieve this purpose. But the summarization approaches are loss methods that they carry errors when restoring the support of original patterns from the compressed patterns. Moreover, both the closed patterns and summarization approaches do not annotate the patterns with semantic information.

Based on our knowledge, currently there are three different approaches for the interpretation of discovered knowledge based on some sorts of semantic annotations: an OLAP based visualization method [17], a generating semantic annotation method [18] and multi-tier structures [15], [14], which used "granules" instead of "patterns" and "rules", and defined meaningless rules based on the relationship between long rules and their general ones (short rules).

In previous research we have found that granules were also a compressed representation. Thus, in this paper, we explore the capability of multi-tier structures for estimating supports for patterns without information loss. This paper proposes the concepts and definitions to illustrate the relationship between patterns and granules. We also presents a method to estimate patterns' support based on granules. A set of experiments has been conducted and the experimental results show that the proposed approach is promising.

The remainder of the paper is structured as follows. Section II discusses related work. Section III and IV introduces basic concepts of granules and the multi-tier structures and describes the basic and derived association mappings. Section V presents the definition of association mappings and discusses their properties. Section VI then presents the support estimation

Y. Li and J. Wu are with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia, Brisbane, QLD 4001.
E-mail: y2.li@qut.edu.au, j3.wu@student.qut.edu.au

discussion for pattern and granule based methods. Section VII evaluates the proposed approach and the last section is the conclusion.

II. Related Work

Pattern mining played an important role for the development of association mining. Many efficient algorithms have been developed for pattern mining [6], [9] in transaction databases. Pattern mining has also been developed for mining frequent itemsets in multiple levels [5], [6], and constraint-based techniques [19], [3], [12], [11], [23].

Since these approaches produce a huge volume of patterns, a new major challenging issue for pattern mining is how to present and interpret discovered patterns. Several approaches have been developed for this issue. A concise representation of patterns is a lossless representation, for example, non-derivable patterns [4], condensed patterns [22], maximal patterns, closed patterns, and regular patterns [25]. Pattern post-processing was also presented recently, for example, pattern compression [30], pattern deploying [29], [13] and pattern summarization [33], [27], [10], [24].

A transaction database can be formally described as an information table (T, V^T), where T is the set of transactions, and $V^T = \{a_1, a_2, \ldots, a_n\}$ is the set of items (or called attributes) for all transactions in T.

Let α be an *itemset*, a subset of V^T. Its *coverset* is the set of all transactions (or objects) $t \in T$ such that $\alpha \subseteq t$, and its support is $\frac{|coverset(\alpha)|}{|T|}$. An itemset α is called *frequent pattern* if its support $\geq min\_sup$, a minimum support. Given a set of transactions (objects) Y, its *itemset* denotes the set of items (attributes) that appear in all the objects of Y. For a pattern α, its closure $closure(\alpha) = itemset(coverset(\alpha))$.

A pattern α is *closed* if and only if $\alpha = closure(\alpha)$. Closed patterns can be summarized into pattern profiles [33] by clustering the patterns with respect to KL-divergence , and a pattern's support can be estimated by using pattern profiles.

Let $T' = \bigcup_{1 \leq i \leq m} T_{\alpha_i}$, where T_{α_i} is the coverset of pattern α_i. A profile M is a triple $\langle pr, \phi, \rho \rangle$, where pr is a probability distribution vector of the items in this profile; ϕ is called master pattern which is the union of a set of patterns (α_1, α_2, ... , α_m); and ρ is the support of the profile which equals to $\frac{|T'|}{|T|}$.

The profile based summarization can largely reduce the pattern number, however, it has following limitations. Firstly, a pattern is possibly covered by multiple profiles. Secondly, it is lack of error guarantee in the support estimation. To achieve a result with less error, a greater number of profiles is required that can reduce the performance of pattern summarization. Finally, the estimation sometimes falsely mark some infrequent patterns as frequent ones, or vise versa.

The concepts of decision rules and granules are well acceptable in the rough set community [20]. Rough set theory has been developed to deal with vagueness for reasoning precisely about approximations of vague concepts. Decision rules have been used for rule-based classification [26], and the construction of decision trees and flow graphs [21].

The advantage of using decision rules is to reduce the two-phases of association mining (pattern mining and rule generation) into one process. In this research, we develop granule mining into multi-tier granule mining in order to identify meaningless rules and efficiently access association rules for user specific needs.

TABLE I
An information table

Object(*Transaction*)	*Items* (*Attributes*)
t_1	a_1 a_2
t_2	a_3 a_4 a_6
t_3	a_3 a_4 a_5 a_6
t_4	a_3 a_4 a_5 a_6
t_5	a_1 a_2 a_6 a_7
t_6	a_1 a_2 a_6 a_7

TABLE II
A decision table

Granule	a_1	a_2	a_3	a_4	a_5	a_6	a_7	N_g
g_1	1	1	0	0	0	0	0	1
g_2	0	0	1	1	0	1	0	1
g_3	0	0	1	1	1	1	0	2
g_4	1	1	0	0	0	1	1	2

III. Decision Table and Two-Tier Structure

In the multi-tier granule mining, the information table is firstly compressed into a decision table for a selected set of attributes by using the Group By operation. The decision table is then represented into a two-tier structure based on a partition of attributes, which classifies the set of attributes into condition attributes and decision attributes, and describes the associations between condition granules and decision granules. The two-tier structure can be further derived into different multi-tier structures to summarize all possible associations between granules based user selected attributes and tiers.

Formally, the decision table of a information table (T, V^T) is denoted as a tuple of (T, V^T, C, D) if $C \cap D = \emptyset$ and $C \cup D \subseteq V^T$. C and D are two groups of attributes which are conditions and decision attributes respectively.

Usually, it is assumed (see [21]) that there is a function for every attribute $a \in V^T$ such that $a : T \rightarrow V_a$, where V_a is the set of all values of a. We call V_a the domain of a. Let B be a subset of V^T. B determines a binary relation $I(B)$ on T such that $(t_1, t_2) \in I(B)$ if and only if $a(t_1) = a(t_2)$ for all $a \in B$, where $a(t)$ denotes the value of attribute a for object $t \in T$. It is easy to prove that $I(B)$ is an equivalence relation, and the family of all equivalence classes of $I(B)$ is denoted by $U = T/B$. We call each equivalence class in U a *granule*. The granule in U that contains transaction t is denoted by $B(t)$. Let $U_C = T/C$ and $U_D = T/D$, granules in U_C or U_D are also referred to *C-granules* or *D-granules*, respectively.

Table I list out a sample transaction table, where $V^T = \{a_1, a_2, ..., a_7\}$ and $T = \{t_1, t_2, ..., t_6\}$. Let a_1 to a_5 be the

TABLE III
C-Granules

Condition Granule	a_1	a_2	a_3	a_4	a_5	*coverset*
cg_1	1	1	0	0	0	$\{t_1, t_5, t_6\}$
cg_2	0	0	1	1	0	$\{t_2\}$
cg_3	0	0	1	1	1	$\{t_3, t_4\}$

TABLE IV
D-Granules

Decision Granule	a_6	a_7	*coverset*
dg_1	0	0	$\{t_1\}$
dg_2	1	0	$\{t_2, t_3, t_4\}$
dg_3	1	1	$\{t_5, t_6\}$

condition attributes and a_6, a_7 be the decision attributes, then table I can be grouped by V^T into a decision table as shown in table II, where $T/C \cup D$ = $\{g_1, g_2, g_3, g_4\}$. Based on this definition, we also have the condition and decision granules as listed out in table III and IV.

In this paper, a relation R_B between U and T is used to describe the relationships between granules and transactions in formal concept analysis [34]. That is, given a transaction $t \in T$ and a granule $g \in U$, we say g is induced by t or t has the property g if $g = B(t)$ (also written as tR_Bg).

Let B be a subset of V^T and $U = T/B$, and granule $g \in U$ be induced by transaction t. Its covering set $coverset(g) = \{t' | t' \in T, t'R_Bg\}$. Let granule $g = cg \wedge dg$, where cg is a *C-granule* and dg is a *D-granule*. We can easily prove that $coverset(g) = coverset(cg) \cap coverset(dg)$. Table III and IV also list out the coversets for the sample *C-granule* and *D-granule*.

The smallest granules only contain one single attribute, we also call them primary granules. A large granule can be generated from some smaller granules by using logic operation "and", $\wedge$. Every granule in the decision table can be mapped into an *association rule* (or called *decision rule*), where the antecedent is a *C-granule* which consists of attributes in C, and the consequent is a *D-granule* which consists of attributes in D. The decision rules can also be regarded as larger granules generated by the condition and decision granules. For instance, the granules g_1, g_2, g_3 and g_4 shown in table II can be generated by the *C-granules* and *D-granules* as follows:

$g_1 = cg_1 \wedge dg_1$;
$g_2 = cg_2 \wedge dg_2$;
$g_3 = cg_3 \wedge dg_2$;
$g_4 = cg_1 \wedge dg_3$.

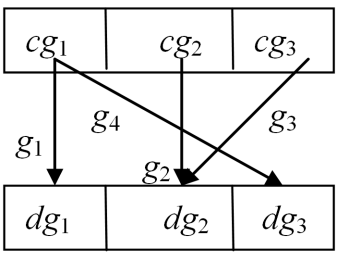


Fig. 1. A 2-tier structure

With these definitions of condition and decision granules as well as the relations between them, a 2-tier structure can be built. Fig. 1 illustrates a 2-tier structure to describe the relationship between these granules in Table II, III and IV. The links (arrows) also represent the associations (decision rules) between condition granules and decision granules. Based on the 2-tier structure, varieties of multi-tier structures and mappings can be derived. The details will be discussed in the following two sections.

IV. MULTI-TIERS STRUCTURE

In this section, we first discuss the concept of multi-tier structures. We also define the concept of general rules (i.e., rules with shorter antecedents) of decision rules in order to clarify the meaning of meaningless in granule mining. At last, we present the method to estimate patterns' support based on granules.

To describe more associations between granules, we can further divide the condition attributes into some categories in accordance with what users want. For example, let C_i and C_j be two subsets of C, which satisfy $C_i \cap C_j = \emptyset$ and $C_i \cup C_j = C$, hence a *C-granule* cg can be divided into a C_i granule cg_i and C_j granule cg_j and have $cg = cg_i \wedge cg_j$.

A multi-tier structure can be describes as a pair (H, A), where *H* is a set of granule tiers and *A* is a set of association mappings that illustrate the associations between granules in different tiers.

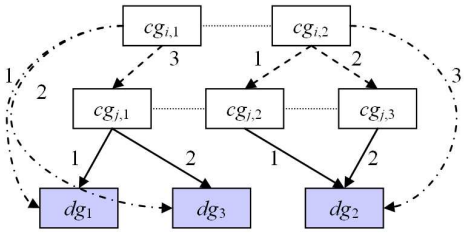


Fig. 2. An example of a multi-tier structure

Fig. 2 illustrates a 3-tier structure, where *C-granules* are divided into C_i*-granules* and C_j*-granules* (i.e., the first two levels in the figure), and we have $H = \{C_i, C_j, D\}$. The C_i tier includes C_i*-granules* = $\{cg_{i,1}, cg_{i,2}, \ldots, cg_{i,k}\}$, the C_j tier includes C_j*-granules* = $\{cg_{j,1}, cg_{j,2}, \ldots, cg_{j,r}\}$, and the D tier includes *D-granules* = $\{dg_1, dg_2, \ldots, dg_v\}$, where $k = 2$, $r = 3$ and $v = 3$.

The 3-tier structure in Fig. 2 includes three association mappings (arrows), Γ_{cd}, Γ_{ij}, and Γ_{id} (i.e., $A = \{\Gamma_{cd}, \Gamma_{ij}, \Gamma_{id}\}$), which show the linkages between *C-granules* and *D-granules* (e.g., the solid arrows), C_i*-granules* and C_j*-granules*, and C_i*-granules* and *D-granules*, respectively. These association mappings can be used to generate association rules.

Given a *C-granule* cg_k and a C_i*-granule* $cg_{i,x}$, $\Gamma_{cd}(cg_k)$ includes all possible associations (links and their strengths) between cg_k and *D-granules*; $\Gamma_{ij}(cg_{i,x})$ includes all possible associations between $cg_{i,x}$ and C_j*-granules*; and $\Gamma_{id}(cg_{i,x})$ includes all possible associations between $cg_{i,x}$ and *D-granules*.

The *link strength* between granule cg_k and granule dg_z is defined as

$$lstrength(cg_k, dg_z) = |coverset(cg_k \wedge dg_z)|$$

which is the number of transactions that have the property "$cg_k \wedge dg_z$".

As defined above, the rule "$cg_k \rightarrow dg_z$" is a decision rule (or association rule), where cg_k is its antecedent and dg_z is its consequent (note the following concepts are also applicable for "$cg_{i,x} \rightarrow dg_z$"). Its *support* is

$$\frac{|coverset(cg_k \wedge dg_z)|}{|T|} = \frac{1}{N} lstrength(cg_k, dg_z)$$

and its *confidence* is

$$\frac{|coverset(cg_k \wedge dg_z)|}{|coverset(cg_k)|} = \frac{lstrength(cg_k, dg_z)}{|coverset(cg_k)|} \quad (1)$$

where $N = |T|$, the total number of transactions.

Different to decision tables, we can discuss general association rules (rules with shorter premises) of decision rules in a multi-tier structure.

Let cg_k be a *C-granule*, dg_z be a *D-granule* and $cg_k = cg_{i,x} \wedge cg_{j,y}$. We call "$cg_{i,x} \rightarrow dg_z$" (or "$cg_{i,y} \rightarrow dg_z$") a *general rule* of rule "$cg_k \rightarrow dg_z$".

Especially in the multi-tier structure, we can define the term "meaningless" for a decision rule based on selected tiers. We call "$cg_k \rightarrow dg_z$" *meaningless* if its confidence is less than or equal to the confidence of its a general rule.

The rationale of this definition is analogous to the definition of interesting association rules, where $\alpha \rightarrow \beta$ is an interesting rule if $P(\beta|\alpha)$ (conditional probability) is greater than $P(\beta)$. If we add a piece of extra evidence to a premise and obtain a weak conclusion, we can say the piece of evidence is meaningless.

V. ASSOCIATION MAPPINGS

In the last section, we discussed a three tiers structure (H, A), where $H = \{C_i, C_j, D\}$, $C_i \cup C_j = C$ and $C_i \cap C_j = \emptyset$, and $A = \{\Gamma_{cd}, \Gamma_{ij}, \Gamma_{id}\}$. Association mappings are used to describe the association relationships between granules in different tiers. They can be used to enumerate all association rules between the associated granules. Usually, there are many possible pairs (C_i, C_j) such that $C_i \cup C_j = C$ and $C_i \cap C_j = \emptyset$, and C_i and C_j can be further divided into smaller sets. Therefore, it is necessary using derived association mappings (e.g., Γ_{id}) for efficient rule generations in multi-tier structures.

A. Basic Association Mapping

The basic association mapping is the mapping between granules from two tiers. For example, the mappings between the condition and decision granules are basic mappings. As the previous definitions, let $U = T/V_T$, $U_C = T/C$ and $U_D = T/D$ to be the set of granules, condition and decision granules. Also let $g_1 \in U_C$ and let $g_2 \in U_D$. Then based on Eq.(1) and Section IV, we have

$$\begin{aligned} lstrength(g_1, g_2) &= |coverset(g_1 \wedge g_2)| \\ &= |coverset(g_1) \cap coverset(g_2)| \\ &= |\{t \in T | tR_C g_1 \ and \ tR_D g_2\}|. \end{aligned}$$

The basic associations between *C-granules* and *D-granules* can be described as a basic association mapping Γ_{cd} such that $\Gamma_{cd}(g)$ is a set of *D-granule* link-strength pairs for all $g \in U_C$. Formally, Γ_{cd} is defined as $\Gamma_{cd} :: U_C \rightarrow 2^{U_D \times I}$, which satisfies

$$\begin{aligned} \Gamma_{cd}(g) = \ & \{(dg, lstrength(g, dg)) | dg \in U_D, \\ & \{t \in T | tR_C g \ and \ tR_D dg\} \neq \emptyset\} \end{aligned}$$

for all granules $g \in U_C$, where I is the set of all integers.

Obviously, supports and confidences of association rules can be easily calculated based on the basic association mapping. Let $g_1 \in U_C$, $g_2 \in U_D$, and "$g_1 \rightarrow g_2$" be a decision rule, its support and confidence can be derived as follows:

$$\begin{aligned} sup(g_1 \rightarrow g_2) &= \frac{1}{N} lstrength(g_1, g_2) \\ &= \frac{1}{N} \sum_{(g_2, ls) \in \Gamma_{cd}(g_1)} ls \end{aligned}$$

$$\begin{aligned} conf(g_1 \rightarrow g_2) &= \frac{lstrength(g_1, g_2)}{|coverset(g_1)|} \\ &= \frac{\sum_{(g_2, ls) \in \Gamma_{cd}(g_1)} ls}{\sum_{(g, ls) \in \Gamma_{cd}(g_1)} ls} \end{aligned}$$

B. Derived Association Mappings

The very interesting property of the multi-tier structures is that we can derive many association mappings based on the basic association mapping rather than using the original set of transactions. This property is significant on time complexities for rule generations.

To simplify the process of deriving, we first consider the method for deriving association mapping Γ_{ij} between C_i-*granules* and C_j-*granules* based on the basic association Γ_{cd}, where $\Gamma_{ij}(g)$ is a set of C_j-*granule* integer pairs, which satisfies

$$\Gamma_{ij} :: U_i \rightarrow 2^{U_j \times I}$$

and

$$\begin{aligned} \Gamma_{ij}(g_i) = \ & \{(g_j, lstrength(g_i, g_j)) | g_j \in U_j, \\ & \{t \in T | tR_i g_i \ and \ tR_j g_j\} \neq \emptyset\} \end{aligned}$$

for all granules $g_i \in U_i$, where $C_i \cup C_j = C$, $C_i \cap C_j = \emptyset$, $U_i = T/C_i$ (the set of C_i-*granules*), $U_j = T/C_j$ (the set of C_j-*granules*), and R_i and R_j are relations between U_i and T, and U_j and T, respectively.

We can also derive the association mapping Γ_{id} between C_i-*granules* and *D-granules* based on the association mappings Γ_{ij} and Γ_{cd}, which satisfies

$$\Gamma_{id} :: U_i \rightarrow 2^{U_D \times I}$$

and

$$\begin{aligned} \Gamma_{id}(g_i) = \ & \{(dg, lstrength(g_i, dg)) | dg \in U_D, \\ & \{t \in T | tR_i g_i \ and \ tR_D dg\} \neq \emptyset\} \end{aligned}$$

for all granules $g_i \in U_i$.

Fig. 3 illustrates the relations between these association mappings. In this figure, the set of condition attributes are split into two sets: C_i and C_j, and the *C-granules* (U_C) are also correspondingly compressed into C_i-*granules* (U_i) and C_j-*granules* (U_j). As defined before, Γ_{cd} is used to describe the association relationship between U_C and U_D. Association mapping Γ_{ij} is used to describe the association relationship between U_i and U_j, and association mapping Γ_{id} is used to describe the association relationship between U_i and U_D.

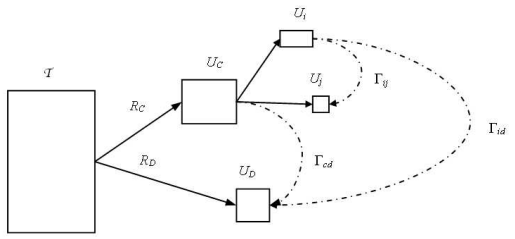


Fig. 3. Relations for derived association mappings

The relationship between the basic mapping and derived mappings can be defined by the following definitions:

Let $C \subseteq B \subseteq V_T$, then the relationship between the granules in U_C = T/C and granules in U_B = T/B can also be defined. A granule $g \in U_C$ is called a $generalized\ granule$ of granule g' $\in U_B$ if $\forall t \in coverset(g') \Rightarrow tR_C g$ (i.e., $coverset(g') \subseteq coverset(g)$). This is denoted as $g' \succcurlyeq g$ for the generalized relationship between g' and g.

Then for all $g \in U_C$, the relation between the coverset of g and its generalized granule g' is formally denoted as the following equation:

$$coverset(g) = \bigcup_{\{g' \in U_B | g' \succcurlyeq g\}} coverset(g') \tag{2}$$

VI. Support Estimation

The support estimation is originally proposed to provide a method to restore the support for the patterns that summarized into limited number of profiles or compressed representation. Given a pattern, usually its support can not be obtained directly from the profiles or compression. Thus, the support of the given pattern only can be estimated through the corresponding restore calculation using the information stored in the profiles or representatives. Moreover, because the profiles are loss summarization, a measure called restoration error is used to examine the precision of the estimated support. This measure is also applied to the estimated support calculated through granules.

A. Support estimation for summarization

After the closed frequent patterns are summarized into the profiles, the support for a pattern needs to be retrieved through the calculation from the profile information. Because one pattern can be covered by multiple profiles, then the maximum result is selected as estimated support. Formally, for a given pattern α_k, its estimated support can be calculated as follows:

$$\hat{s}(\alpha_k) = max_M(\rho_M \times \prod_{a_i \in \alpha_k} pr_M(a_i = 1)) \tag{3}$$

which selects the maximum one from all profiles M that include α_k.

One method to measure the accuracy of the estimated support is to measure the average relative error between the estimated support and the original support. Formally, given a summarization or compression of the original patterns and a set of testing pattern set T = $\{\alpha_1, \alpha_2, ..., \alpha_l\}$, the quality of this summarization or compression can be evaluated through the average relative error, called restoration error denoted as J, defined as follow:

$$J = \frac{1}{|T|} \sum_{\alpha_k \in T} \frac{|s(\alpha_k) - \hat{s}(\alpha_k)|}{s(\alpha_k)} \tag{4}$$

where T = $\{\alpha_1, \alpha_2, ... , \alpha_l\}$ is a given test set of patterns. $s(\alpha_k)$ is the real support of the pattern α_k, while the $\hat{s}(\alpha_k)$ is the estimated support calculated by the pattern profiles or granules.

In the actual calculation, the original pattern sets can be used as the testing set so that the restoration error measures difference between the real support and the estimated support. The smaller the error rate is, the closer is the estimated support to the actual support. It is obvious that if the restoration error is zero, the estimated support equals to the actual support.

B. Support estimation for granules

In terms of multi-tier structure of granules, the estimated support is calculated through the granules and association mappings. The estimated support can be calculated by the granule support if the given pattern is derived by the granule or the support can be calculated through the link strength of the association mappings between granules that containing the pattern. In some circumstance, the estimated support calculated through the multi-tier structure can achieve a zero restoration error rate.

There are several different calculation to obtained the estimated support from the multi-tier structure of granules according to what is the definition of the current multi-tier structure and which tiers of granule are containing the given pattern.

The first case is to estimate the support for a pattern with a decision table. Let G be the decision table of information table(T, V^T), then the estimated support is calculated solely through sum of support of the granules containing the pattern. The equation to calculate the estimated support for a given pattern α is as follow:

$$\hat{s}_1(\alpha, G) = \frac{\sum_{g \in G, \alpha \subseteq g} \sup(g)}{\sum_{g_i \in G} \sup(g_i)} = \frac{\sum_{g \in G, \alpha \subseteq g} \sup(g)}{|T|}$$

The second case is calculating the estimated support with a two tier structure. For a 2-tier structure, let CG be the set of C-granules and DG be the set of D-granules. Given a pattern α, it can be divided into two patterns α_1 and α_2 such that α_1 = $\alpha \cap C$ and α_2 = $\alpha \cap D$, respectively. Then the estimation support is calculated as the summary of granules support if pattern α only contained by the granules in one tier. Or it can be calculated through the link strength of the association mappings between the granules that containing α_1 and α_2. The

equations for the estimated support calculation are as follow:

$$\hat{s}_2(\alpha, CG, DG) =$$

$$\begin{cases} \frac{1}{|T|} \sum\limits_{g \in CG, \alpha_1 \subseteq g} \sup(g) = \hat{s}_1(\alpha_1, CG) & \text{if } \alpha_2 = \emptyset \\ \frac{1}{|T|} \sum\limits_{g \in DG, \alpha_2 \subseteq g} \sup(g) = \hat{s}_1(\alpha_2, DG) & \text{if } \alpha_1 = \emptyset \\ \frac{1}{|T|} \sum\limits_{\alpha_1 \subseteq g_1 \in CG, \alpha_2 \subseteq g_2 \in DG} \textit{lstrengh}(g_1 \rightarrow g_2) & \text{otherwise} \end{cases} \quad (5)$$

For other cases with the n-tier structures, the estimated support can be calculated by different methods depending on how many tiers of granules that the given pattern is derived from. There are three categories of the calculation method for the estimated support in the multi-tier structure. The first case is that the given pattern is only contained by granules in only one tier, then the support can be calculated by using the supports of the granules in the corresponding tier. The second case is for the patterns which are contained by the granules of two tiers in the multi-tier structure. The calculation for the support in such cases can use the link strength of the mappings between the two granules to obtain the support. This calculation is done directly in the current multi-tier structure. The third method is for the patterns that are contained by the granules from three or more tiers. To get the support for such patterns, the calculation needs to use the mapping informations from the 2-tier structure to compute the support through Eq.(5).

To demonstrate the estimated support calculation from the multi-tier structure, here uses a 3-tier structure as an example to illustrate the calculation details. Let a 3-tier structure be $H = \{C_i, C_j, D\}$ containing three sets of granule that are C_iG, C_jG and DG respectively. Then a pattern α can be divided into three patterns, namely $\alpha_1 = \alpha \cap C_iG$, $\alpha_2 = \alpha \cap C_jG$ and $\alpha_3 = \alpha \cap DG$. If only one of α_1, α_2 or α_3 is non-empty, then the support is calculated through the sum of support of only one set of granules as follow:

$$\hat{s}_3(\alpha, C_iG, C_jG, DG) =$$

$$\begin{cases} \frac{1}{|T|} \sum\limits_{g \in C_iG, \alpha_1 \subseteq g} \sup(g) = \hat{s}_1(\alpha_1, C_iG) \text{ if } \alpha_2, \alpha_3 = \emptyset \\ \frac{1}{|T|} \sum\limits_{g \in C_jG, \alpha_2 \subseteq g} \sup(g) = \hat{s}_1(\alpha_2, C_jG) \text{if } \alpha_1, \alpha_3 = \emptyset \\ \frac{1}{|T|} \sum\limits_{g \in DG, \alpha_3 \subseteq g} \sup(g) = \hat{s}_1(\alpha_3, DG) \text{ if } \alpha_1, \alpha_2 = \emptyset \end{cases} \quad (6)$$

For the second case, that is one of α_1, α_2 and α_3 is empty, then the support is calculated using the link strength of the mappings of $\Gamma_{i,j}$, $\Gamma_{i,d}$ or $\Gamma_{j,d}$ as follow:

$$\hat{s}_3(\alpha, C_iG, C_jG, DG) =$$

$$\begin{cases} \frac{1}{|T|} \sum\limits_{\alpha_1 \subseteq g_1 \in C_iG, \alpha_2 \subseteq g_2 \in C_jG} \textit{lstrengh}(g_1 \rightarrow g_2) \text{if } \alpha_3 = \emptyset \\ \frac{1}{|T|} \sum\limits_{\alpha_1 \subseteq g_1 \in C_iG, \alpha_3 \subseteq g_3 \in DG} \textit{lstrengh}(g_1 \rightarrow g_3) \text{if } \alpha_2 = \emptyset \\ \frac{1}{|T|} \sum\limits_{\alpha_2 \subseteq g_2 \in C_jG, \alpha_3 \subseteq g_3 \in DG} \textit{lstrengh}(g_2 \rightarrow g_3) \text{if } \alpha_1 = \emptyset \end{cases} \quad (7)$$

Finally, if none of α_1, α_2 or α_3 is empty, then it is a case of the third category. Then the support is calculated through the mappings of $\Gamma_{c,d}$. In order to use these mappings, the division of α need to be modified. That is, let $\alpha_1 \cup \alpha_2 = \alpha \cap CG$ where $CG = C_i \cup C_j$ such that the support can be obtained by using a modified equation of Eq.(5). The equation used for this calculation is as follow:

$$\begin{aligned} &\hat{s}_3(\alpha, C_iG, C_jG, DG) = \\ &\frac{1}{|T|} \sum_{Cond} \textit{lstrengh}((g_1 \wedge g_2) \rightarrow g_3) \\ &Cond : \alpha_1 \subset g_1 \in C_iG, \alpha_2 \subset g_2 \in C_jG, \\ &\qquad C_iG \cup C_jG = CG, \alpha_3 \subseteq g_3 \in DG \end{aligned} \quad (8)$$

Regarding the quality of the estimation, when using the two tier structure to calculate the estimated support, it can achieve the zero restoration error rate because the two tier structure is a lossless compression. Further, for the multi-tier structure has more than two tiers, it also can achieve the zero error rate when using only the mappings of granules from two tiers or the calculation is performed via the basic 2-tier structure.

Theorem 1: For a given pattern α and a multi-tier structure $H = \{C, D\}$, the estimated support calculated through H equals to the original support of α. That is, $\hat{s}(\alpha, H) = Sup_\alpha$.

Proof: For a pattern α, let $\alpha = \alpha_1 \cup \alpha_2$ such that $\alpha_1 \cap \alpha_2 = \emptyset$. Then for the support of α, we have

$$Sup_\alpha = |coverset(\alpha_1) \cap coverset(\alpha_2)|.$$

Assume $\alpha_1 \subset g_1$ and $\alpha_2 \subset g_2$, and $g_1 \in CG$ and $g_2 \in DG$. According to Eq.(2), we have:

$$coverset(\alpha_1) = \bigcup_{g_{1,i} \in CG} coverset(g_{1,i})$$

and

$$coverset(\alpha_2) = \bigcup_{g_{2,i} \in CG} coverset(g_{2,i}).$$

Meanwhile, the link strength of the mapping from g_1 to g_2 is:

$$\begin{aligned} lstrength(g_1 \rightarrow g_2) &= |coverset(g_1 \wedge g_2)| \\ &= |coverset(g_1) \cap coverset(g_2)|. \end{aligned}$$

Moreover, we have

$$\begin{aligned} &\sum_{g_1 \in CG, g_2 \in DG} lstrength(g_1 \rightarrow g_2) \\ &= |\bigcup_{g_1 \in CG, g_2 DG} (coverset(g_1) \cap (coverset(g_2))| \\ &= |\bigcup_{g_{1,i} \in CG} coverset(g_{1,i}) \cap \bigcup_{g_{2,i} \in CG} coverset(g_{2,i})| \\ &= |coverset(\alpha_1) \cap coverset(\alpha_2)| \\ &= Sup_\alpha. \end{aligned}$$

Therefore, we have $Sup_\alpha = \hat{s}(\alpha, H)$. ■

VII. Experiments and Discussion

Foodmart 2005 data collection contains two databases: SQL Database and OLAP Database. The data used in this experiment is the customer sales data from the OLAP database (see *http://www.e-tservice.com/*), which includes four data cubes. The Warehouse and Sales cube is used in our experiments, which contains four measures and we used the unit-sales measure. The Product dimension used in the Warehouse and Sales cube, consists of eight levels which are All, Product family,

TABLE V
THE TIERS AND THEIR ATTRIBUTES

Levels	Attributes			
2-tiers	C			D
3-tiers	C_i		C_j	D
	Drink	Food1	non-consumable	Food2
4-tiers	$C_{i,1}$	$C_{i,2}$	C_j	D
	$A_1..A_4$	$A_5..A_{11}$	$A_{12}..A_{16}$	$A_{17}..A_{23}$

Product department, Product category, Product subcategory, Brand and Product. In the experiments, we only use the top three levels: All, Product family, and Product department.

There are total 23 attributes in the *Product Department* level. These attributes are categorized into 4 product families: *Drink* (Alcoholic Beverages, Baking Goods, Beverages, Dairy), *Non-Consumable* (Carousel, Checkout, Health and Hygiene, Household, Periodicals), *Food* 1 (Baked Goods, Breakfast Foods, Canned Foods, Canned Products, Deli, Eggs, Frozen Foods) and *Food* 2 (Meat, Packaged Foods, Produce, Seafood, Snack Foods, Snacks, Starchy Foods).

The transactions used in the experiments are the customers' purchase records stored in the fact table of unit sales. Every transaction is the record of one day purchase of one customer for all products which is sum up to product categories. To build up the decision table and multi-tier structures of granules, the transactions of the Unit sales are transformed into an information table using the following procedure. If the customer purchases one or more products from that product department, the value of the attribute in the product department level is set to 1; otherwise, the value is set to 0. The total number of transactions in the information table is 53,700.

The experiments test the proposed solution from several aspects, including space and time complexities, and the restoration error rate of estimated support. We use two baseline models to compare with the proposed theory. The first baseline model is the decision table. The attributes are viewed as two groups: condition and decision attributes. The second baseline model is a pattern summarization model [33], which used pattern profiles to estimate the support of any pattern (see Eq.(3) and (4)).

1) Space and time complexity: In the experiments, the information table is transformed into a decision table first. Multi-tier structures are then constructed based on this decision table and the semantic information of attributes. Table V shows a special definition of the multi-tier structures, where the semantic relation between attributes are considered. As in Table V, there are three multi-tier structures: a two-tier structure (C and D), a three-tier structure (C_i, C_j and D), and a four-tier structure ($C_{i,1}$, $C_{i,2}$, C_j and D).

We also made other 16 definitions of multi-tier structures by grouping the 23 attitudes in different combinations. For each definition, a 2-tier structure (C and D) is built firstly. Then from it a 3-tier structure is built by dividing the C tier into two smaller C_i and C_j. Then the C_i tier is further divided into tier $C_{i,1}$ and $C_{i,2}$ to generate a 4-tier structure (C_1, C_2, C_j and D).

Fig 4 depicts the trends of total granule numbers in the multi-tier structures when the number of tiers increases. It is obvious that in most of the test, the number of granules drops largely with the tier increases. Table VI shows the number of patterns in the information table based on different minimum support values.

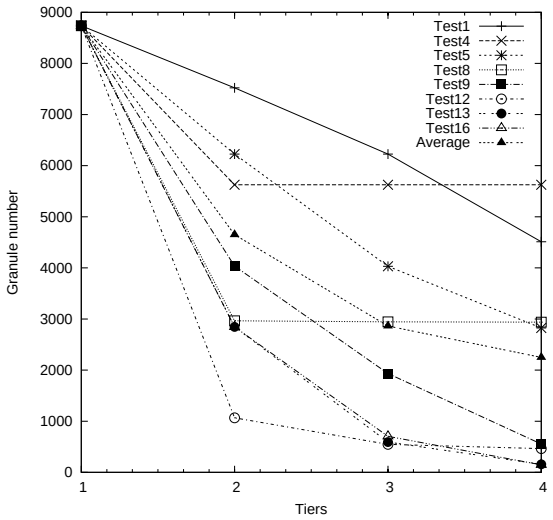


Fig. 4. Granule numbers under different tier settings

TABLE VI
FREQUENT PATTERN NUMBER

Pattern type	$min\_support$	Number of patterns
Frequent pattern	1	56707
Closed pattern	1	15859
Frequent pattern	5	12217
Closed pattern	5	10963
Frequent pattern	50	1486
Closed pattern	50	1486

Comparing with the multi-tier structures, pattern mining gets a large amount patterns if the $min\_support$ is not big enough. However, when the $min\_support$ is big enough (e.g., 50 in this example), pattern mining will lose many large patterns. Different from the pattern mining, multi-tier structures can use a very small space to contain all the possible associations for the chosen data attributes.

Table VII shows the results of the runtime tests. It is obviously that the time used by multi-tier structures is much less than that of pattern mining. Only when the minimum support is set to a very large number of occurrence, the time to obtain the frequent patterns looks acceptable. Fig 5 also obviously shows these differences between the two approaches.

The results also reflect that the time used to create new tiers from smaller granules is less than that from larger granules.

TABLE VII
RUNTIME

Granule		Pattern	
Multi-tier structure	Time(ms)	min_sup	Time(ms)
Decision table	19140	1	1.078e+007
2-tier	6765	5	1.131e+006
3 tier from 2 tier	2593	50	122672
4 tier from 3 tier	171		

For example, the time used to generate a four-tier structure from a three-tier structure is 171 ms, while it takes 2593 ms for constructing the three-tier structure from a two-tier structure. These results show that the proposed theory has achieved the remarkable performance.

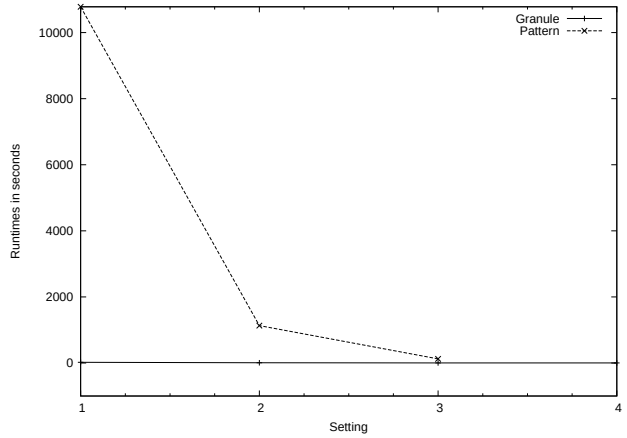


Fig. 5. Runtime of granule and pattern approach

2) Restoration error rate and meaningless rules: The pattern summarization model uses all closed patterns, which are generated from the whole information table with a minimum support of 5, as the input patterns. There are 10963 closed patterns in total. The restoration error rate J is calculated by using Eq.(4). Several tests are carried out with the different number of profiles. The number of profiles is set to 200, 500, 750 and 1000 respectively. Fig 6 shows the results for the error rate of the pattern summarization model vs. granule mining. The results reflect that when using small number of profiles such as 200, 500 and 750, the restoration error is much higher than using granules. To be noticed, using a two tier structure to calculate the support for all patterns (see Eq.(5), (6), (7) and (8)), the J values can remain as zero. This result proves the discussion in section VI-B that the support estimated by the granules in a two tier structure equals to the pattern's original support.

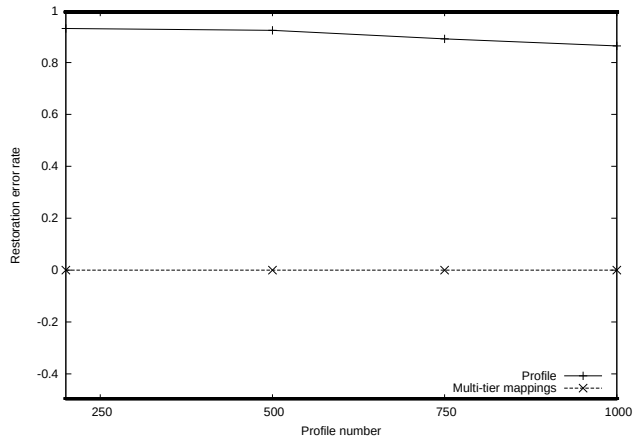


Fig. 6. Estimated support error rate

The multi-tier structures also provide a special feature for pruning some meaningless rules. Based on the definitions, in these experiments, we generate general rules first for the 16 definitions of multi-tier structures. We then filter out the meaningless rules based on their general rules. We found that the rules contain about 30% meaningless rules in average.

VIII. Conclusion

Multi-tier granule mining provide an efficient way to represent and summarize association rules between granules based user selected attributes and tiers. This paper continues the development of multi-tier structures. It presents formalizes concepts of association mappings and a method to 'estimate patterns' support based on related granules and the multi-tier structures. Moreover, it conducts a set of experiments on Foodmart 2005 data collection to test the proposed method. Compared with pattern summarization, the proposed multi-tier granule mining achieves the best performance with zero restoration error rate. The experimental results also show that the multi-tier structures can use a very small space to store the possible associations, and the multi-tier structures can be created efficiently. This research provides a promising alternative approach to find useful associations in databases for user specific needs.

References

[1] M. Barsky, S. Kim, T. Weninger, and J. Han. Mining flipping correlations from large datasets with taxonomies. In *Proceedings of the VLDB endowment, Vol. 5, No. 4*, pages 370–381, 2011.

[2] R. Bayardo, R. Agrawal, and D. Gunopulos. Constraint-based rule mining in large, dense databases. *Data Mining and Knowledge Discovery*, 4:217–240, 2000.

[3] C. Bucila, J. Gehrke, D. Kifer, and W. White. Dualminer: a dual-pruning algorithm for itemsets with constraints. In *Proc. of KDD'02*, 42–51, 2002.

[4] T. Calders and B. Goethals. Mining all non-derivable frequent itemsets. In *Proc. of 2002 PKDD*, 74–85, 2002.

[5] J. Han and Y. Fu. Discovery of multiple-level association rules from large databases. In *Proc. of VLDB'95*, 420–431, 1995.

[6] J. Han and Y. Fu. Mining multiple-level association rules in large databases. *IEEE Transaction on Knowledge and Data Engineering*, 11(5):798–805, 1999.

[7] J. Han, L. V. Lakshmanan, and R. T. Ng. Constraint-based multidimensional data mining. *IEEE Computer*, 32(8):46–50, 1999.

[8] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.

[9] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proc. of SIGMOD'00*, 1–12, 2000.

[10] R. Jin, M. Abu-Ata, Y. Xiang, and N. Ruan. Effective and efficient itemset pattern summarization: Regression-based approaches. In *Proc. of KDD-08*, 399–407, 2008.

[11] A. J. T. Lee, W. Lin, and C. Wang. Mining association rules with multi-dimensional constraints. *Journal of Systems and Software*, 79(1):79–92, 2006.

[12] C. K.-S. Leung, L. V. S. Lakshmanan, and R. T. Ng. Exploiting succinct constraints using fp-trees. *SIGKDD Explorations*, 4(1):40–49, 2002.

[13] Y. Li, A. Algarni, and N. Zhong. Mining positive and negative patterns for relevance feature discovery. In *Proceeding of KDD'10*, pp. 753–762, 2010.

[14] Y. Li, W. Yang, and Y. Xu. Multi-tier granule mining for representations of multidimensional association rules. In *Proceedings of ICDM'06*, pp. 953–958, 2006.

[15] Y. Li and N. Zhong. Interpretations of association rules by granular computing. *ICDM-03*, pages 593–596, 2003.

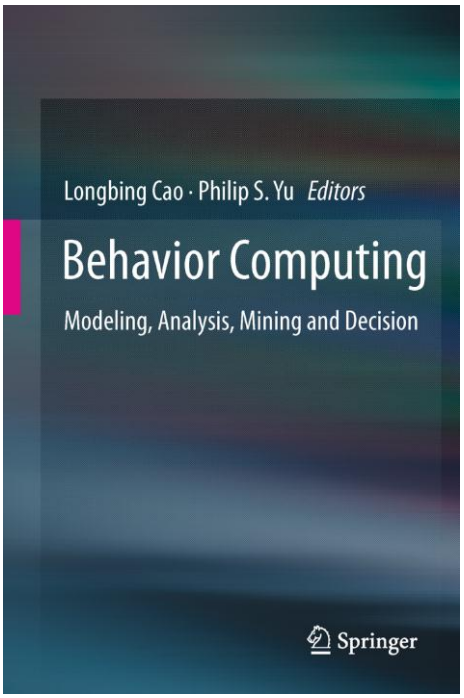
[16] Y. Li, X. Zhou, P. Bruza, Y. Xu, and R. Y. Lau. A Two-Stage Decision Model for Information Filtering. *Decision Support Systems*, 52(3):706–716, 2012.

[17] B. Liu, K. Zhao, J. Benkler, and W. Xiao. Rule interestingness analysis using olap operations. In *Proceedings of KDD'06*, pp. 297–306, 2006.

[18] Q. Mei, D. Xin, H. Cheng, J. Han, and C. Zhai. Generating semantic annotations for frequent patterns with context analysis. In *Proc. of KDD'06*, 337–346, 2006.
[19] R. T. Ng, L. V. S. Lakshmanan, J. Han, and A. Pang. Discovery of multiple-level association rules from large databases. In *Proc. of SIGMOD'98*, 13–24, 1998.
[20] Z. Pawlak. In pursuit of patterns in data reasoning from data - the rough set way. *RSCTC-02*, 1–9, 2002.
[21] Z. Pawlak. Decision trees and flow graphs. In *Proceedings of RSCTC'06*, pp. 1–11, 2006.
[22] J. Pei, G. Dong, W. Zou, and J. Han. On computing condensed frequent pattern bases. In *Proceedings of ICDM'02*, pp. 378–385, 2002.
[23] J. Pei and J. Han. Constrained frequent pattern mining: a pattern-growth view. *SIGKDD Explorations*, 4(1):31–39, 2002.
[24] A. K. Poernomo and V. Gopalkrishnan. Cp-summary: a concise representation for browsing frequent itemsets. In *Proceedings of KDD'09*, pp. 687–696, 2009.
[25] S. Ruggieri. Frequent regular itemset mining. In *Proceedings of KDD'10*, pp. 263–272, 2010.
[26] A. Skowron, H. Wang, A. Wojna, and J. G. Bazan. Multimodal classification: case studies. *Transactions on Rough Sets V*, LNCS 4100:224–239, 2006.
[27] C. Wang and S. Parthasarathy. Summarizing itemset patterns using probabilistic models. In *Proceedings of KDD-06*, pp. 730–735, 2006.
[28] G. Webb. Filtered-top-k association discovery. *WIREs Data Mining and Knowledge Discovery*, 1(3):183–192, 2011.
[29] S.-T. Wu, Y. Li, and Y. Xu. Deploying approaches for pattern refinement in text mining. In *Proceedings of ICDM'06*, pp. 1157–1161, 2006.
[30] D. Xin, J. Han, X. Yan, and H. Cheng. Mining comressed frequent-pattern sets. In *Proceedings of VLDB'05*, pp. 709 – 720, 2005.
[31] Y. Xu and Y. Li. Generating concise association rules. In *Proceedings of CIKM'07*, pp. 781–790, 2007.
[32] Y. Xu, Y. Li, and G. Shaw. Reliable representations for association rules. *Data and Knowledge Engineering*, 70(6):555–575, 2011.
[33] X. Yan, H. Cheng, J. Han, and D. Xin. Summarizing itemset patterns: A profilebased approach. In *Proceedings of KDD-05*, pp. 314–323, 2005.
[34] Y. Y. Yao. A comparative study of formal concept analysis and rough set theory in data analysis. In *Proceedings of RSCTC'04*, pages 59–68, 2004.
[35] M. J. Zaki. Mining non-redundant association rules. *Data Mining and Knowledge Discovery*, 9:223–248, 2004.
[36] N. Zhong, Y. Li, and S.-T. Wu. Effective pattern discovery for text mining. *IEEE Transactions on Knowledge and Data Engineering*, 24(1):30 –44, 2012.

Behavior Computing: Modeling, Analysis, Mining and Decision

BY LONGBING CAO AND PHILIP YU (EDITORS) - ISBN: 978-1-4471-2968-4



REVIEWED BY SURESH SOOD

Erving Goffman, one of the most influential sociologists of the last century published the Presentation of Self in Everyday Life (PSEL;1959). This is the most definitive 20th century study of the patterns of human behavior in mundane social situations. If we look to psychology, sociology, anthropology, marketing or even organisational studies the representation of human behavior emerges from rich PSEL type "thick" qualitative descriptions. This approach is prevalent amongst the social sciences and does not provide a pathway to generalizable or predictive models of behavior. Furthermore, the availability of big data from social networks is pushing researchers and practitioners to undertake deep and dynamic behavioral analysis in the converging online and offline worlds.

Human behavior is made up of complex interdependencies not least of all because individuals convey actions using the multiple modes of voice, facial and eye movements, hand gesturing and body to interact on a social basis. The modeling and analysis of human behavior is giving rise to the new discipline of behavior computing integrating techniques from both computer science and social sciences. In this new field of behavior computing, a major distinction over previous behavioral research is a focus on online social networks and the Internet impacting behavior rather than the traditional experimental analysis of the behavior of animals and organisms. The field of behavior computing opens up the opportunity for breakthrough advances, discoveries and advanced knowledge to come from outside of social sciences.

"Behavior Computing" captures the transformation in the converging study of human behavior and computing. This book is a welcome addition alongside statistical, machine learning and cognitive neuroscience books. This book effectively contextualizes statistical
and machine learning tools in a series of 23 very interesting chapters embracing models, scenarios and case studies thematically connected with behavior computing. The end result is a highly presentable book for a wide-ranging audience inclusive of final year undergraduates or postgraduate students. However, the book requires familiarity with machine learning algorithms and analysis of large datasets and may just prove to be a catalyst for social scientists to leave behind existing pastures and embrace the field of behavior computing.

On a more descriptive basis, this book may aptly hold the title Behavior and Social Informatics Computing (BSIC) in line with the IEEE Computational Intelligence Society task force of the same name and chaired by Longbing Cao. This book covers similar ground with research methods sourced from a wide perspective. The book makes far less daunting the challenge of trying to make sense of writings from the different fields of social science, computer science, information systems and information science. Cao and Yu as editors have risen to the challenge of the diversity of studies captured by this book and make the ideas available to a wide community while trying to ensure somewhat a consistency in the different terminology by selecting studies capturing a body of research focusing on computing while taking into account human behavior.

While the book structure is in the four parts of behavior modeling, analysis, mining and applications, the reader likely benefits from starting with the last section (IV) focusing on decision making possibilities captured as six case studies of behavior applications and then cycling back to the beginning to gain a fundamental grounding in behavior representation. Each chapter contains a variety of references providing a springboard for further research in the field.

Before deep diving into the covers, Cao and Yu's book may well appeal to a wider audience than originally intended as the book offers new techniques for social media researchers. Consumers are generating big data in social media and researchers are seeking valuable insights amongst the oceans of noise. This book incorporates a variety of thinking on the capture of behavior characteristics in social media opening up the potential to harness social media big data by extracting valuable insights.

Behavior Computing recognizes the paucity of formal methods and techniques to represent behavior and commences with capturing the characteristics and dynamics of

modeling influential behavior in social media. Cao authors the SAPMAS (social activity process modeling and analysis system) ontology using a narrative knowledge representation language and user behavior representation. The contribution is clear, the formal modeling of social behaviors and model checking to analyze the social activity processes. This ontology chapter is a major first step towards the development of computer usable tools and predictive behavior models. A subsequent chapter contains learning from the capture of user behavior from a semi Markov model. Again, the importance of behavior representation is central using a model to display personalized information to a software package user.

Behavior analysis (section II) considers a personalized social event recommender system (P- SERS). This system uses social information in the community to model and give personal social recommendations to the target user. Again, the emphasis is clear and focus on behavior computing. The researchers incorporate the online group web buying website in Taiwan IHERGO for experimentation with 600,000 members to use personal social info for personal group buying recommendation. In contrast, academic database service providers benefit from the work using behavioral and psychological characteristics and adaptive learning mechanism of academics during information searching advice under situations of cognitive incompleteness. This research contributes to the creation of effective Q&A learning systems. The piece de resistance of this section if not the entire book is a chapter on scoring and predicting financial risk preferences. This chapter actually goes well beyond a focus on behavior analysis and provides the reader with a methodology for the development and implementation of a risk-scoring algorithm together with insight generation. The scoring algorithm generated with real data demographic attributes of individuals helps predict risk preference. The smarter decision making aspect of this work is we do have to ask risk related questions but instead by obtaining demographic information one can estimate if the individual is risk seeking or risk averse. This is an important contribution as this eliminates the need to collect sensitive information online. Smart financial institutions are able to integrate the information directly into a CRM.

Behaviour mining (section III) sticks to the knitting commencing with a new trajectory pattern mining framework CCT (clustering clues of trajectories) using Foursquare or geo- tagged. The technique recognises the silent durations inherent within trajectories are clues of a movement behaviour reflected by spatial and temporal co located data points. Through clue aware trajectory similarity and a clue aware clustering algorithm to cluster similar trajectories into groups, CCT is able to discover trajectory patterns even if the trajectory captures only fragments of movement behavior. Overlapping with the work of CCT is individual movement behavior in secure environments with modeling and detection of suspicious activity. Patterns of suspicious behavior are captured from RFID tags and sensors. This chapter contains a typology of suspicious behavioral patterns. The method entails modeling the physical environment and applying suspicious patterns to the logs of RFID access. This data helps build a model of human activity and analyses the sequence of actions comprising harmful activity. Not unrelated in this section is the behavior modeling approach for unauthorized copying of large amounts of documents from digital library. In this context, anomaly detection builds on an understanding the simultaneous interest in a short space of time of the biographies of mathematicians, philosopher works and contemporary microbiology news is anomalous. In the context of the retail industry, the linking of behavioral patterns to personal attributes through re-mining of item associations provides the competitive retail industry with new marketing opportunities. This chapter provides a novel method for knowledge discovery from association mining. The contribution of this chapter goes well beyond an algorithm. The proof of concept enhances the research with a practical contribution while balancing with a theoretical contribution in terms of methodology. Retailers will immediately benefit from policy implications e.g. caramel frappucino offered to male versus java chip chocolate. Furthermore, a retailer is able to test for the validity preferences of food items by time of day and nationality. A natural step for a retail chain is to use the technique to create stores and offers appealing to communities of different nationalities hidden within existing customers. Rounding out the body of knowledge is twitter user behaviour and data mining to discover or infer taxonomy from dynamic correlations in twitter user generated content.

Behavioral applications (section IV) are key to leveraging behavior intelligence and achieving smarter decision making across a variety of industries. Telecommunications and in particular mobile networks are a natural partner for behavior computing. Within mobile social networks, the mobile user interaction patterns change frequently and make the challenge of detecting changing patterns extremely difficult because humans have high degree of randomness in calling. To identify regularity in random behavior an initial chapter in this section presents a new method using network attributes to find periodicity in dynamic social networks. On a practical note, this work helps telcos develop business models and call plans in cellular comms based on behavior of mobile users. The total data set treated by the researchers is 5 million nodes and 400 million edges. Other useful aspects of this work inform an understanding of individuals within different time zones, bottlenecks, structural holes, and isolates. Another project mining the MIT Reality data set based on mobile call records provides assists with unwanted call detection. The same problem space is reviewed from a different perspective with the notion of the smart phone predicting the next call. The prediction of incoming calls has direct usage in call center workload prediction, social networks, calendar and voice spam. The smart phone researchers with just the context use a call prediction schedule based on caller behavior and history. The Holt-Winters method predict calls from frequent and periodic calls captured as a 5 tuple call record (date/start

time/type/caller id/talk-time) and use the call record for the next call. Other behavior computing applications are insightful covering handwriting recognition on mobile device, search behavior of medical students and software testing complete this book.

Currently, this book is the only dedicated book to explore behavior informatics and behavior computing on the bookshelf. The book is suitable reading for researchers, students and practitioners. The editors have managed to bring together a balance of theory and application in contemporary contexts. As the first of a kind in a newly emerging field, the book challenges others including early career researchers to develop and contribute deeper expertise in this inspiring area of behavior.

THE BOOK:

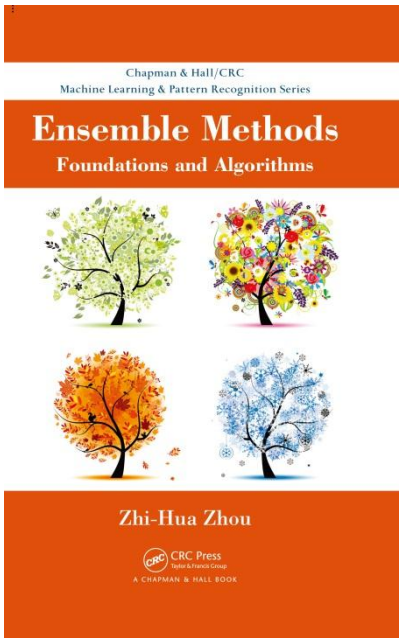
CAO,LONGBING AND YU, PHILLIP S (EDS) (2012), BEHAVIOR COMPUTING: MODELING, ANALYSIS, MINING AND DECISION, 390 P.
SPRINGER.
ISBN: 978-1-4471-2968-4

ABOUT THE REVIEWER:

SURESH SOOD
Advanced Analytics Institute, University of Technology Sydney, Australia.
Contact him at: suresh.sood@uts.edu.au

Ensemble Methods: Foundations and Algorithms

BY Zhi-Hua Zhou - ISBN 978-1-439-830031



REVIEWED BY DIRK VAN DEN POEL

nsemble methods train multiple learners and then combine them for use. They have become a hot topic in academia since the 1990s, and are enjoying increased attention in industry. This is mainly based on their generalization ability, which is often much stronger than that of simple/base learners. Ensemble methods are able to boost weak learners, which are even just slightly better than random performance to strong learners, which can make very accurate predictions.

Zhi-Hua Zhou's "Ensemble Methods: Foundations and Algorithms" starts off in Chapter 1 with a brief introduction to the basics, by discussing nomenclature and the basic classifiers including, naive bayes, SVM, k-NN, decision trees, etc.

The real ensemble content kicks off with a discussion of Boosting (Chapter 2), followed by Bagging (Chapter 3). These two chapters form the heart of the book; hence they are discussing the topic in detail. The boosting chapter explains the basic idea, which starts by fitting one learner, and correcting its "mistakes" in subsequent learners. Adaboost is its best known representative of the residual-decreasing methods, which is explained in-depth in Chapter 2. It is an example of a sequential ensemble method. Error bounds of the final combined learner are discussed based on the errors of its weak base learners. Mostly, the book first explains the binary classification problem, and then ventures into multi-class extensions (one-versus-all, one-versus-one approaches), also in this case for multiclass Adaboost. It is well known that the algorithm suffers from noisy data. Hence, the remainder of this chapter mainly focuses on how the algorithm can be made less vulnerable to its weakness to noisy data.

Chapter 3 details the Bagging idea (Boostrap AGGregatING), which is a parallel ensemble method, and lends itself ideally to the possibility of parallel computing. Bagging uses bootstrap sampling (i.e., composing a new dataset of the same size by sampling with replacement from the base dataset). It builds on the idea that the combination of independent base learners will lead to a substantial decrease of errors and therefore, we want to obtain base learners as independent as possible. The bootstrapping leads to a nice side-benefit: Thanks to sampling with replacement, about 37% of the base dataset remains unused, i.e., out-of-bag validation performance can be computed to assess the quality of the learner. Talking about bagging would not be complete without talking about Random Forest, Breiman's random tree ensemble. They can also be found in the book.

Chapter 4 talks about combination methods, which form the basis to achieve strong generalization ability. The author starts with the most prominent form of combination methods: Averaging (simple, weighted, etc.) for regression, and voting (majority, weighted, plurality, etc.) for classification. Next, Stacking (also known as constructing a meta-learner); the idea of stacking is to train the first-level learners using the original training dataset, and then generate a new dataset for training the second-level (meta) learner, where the outputs of the first-level learners are regarded as input features. Next, the author goes on to discuss a number of other combination methods: algebraic methods, Behavior Knowledge Space (BKS) method and decision template method.

Diversity is the foundation on which the performance of ensembles is built. Hence, the book devotes an entire chapter (5) to this topic, providing a lot of information of diversity measures.

Chapter 6 is devoted to ensemble pruning: Instead of using all learners, why not use a subset of them. Generally, it is better to retain some accurate learners together with some not-that-good but complementary learners. The author discusses ordering-based pruning, clustering-based pruning, and optimization-based pruning.

In Chapter 7 the book discusses Clustering Ensembles. These are desired to improve clustering quality, clustering robustness, etc., although their original motivation was to enable knowledge reuse and distributed computing. The author discusses similarity-based methods, graph-based methods, relabeling-based methods, and transformation-based methods.

Finally, Chapter 8 discusses advanced topics such as semi-supervised learning with ensembles, active learning, and class-imbalance learning. In real-world applications, in addition to attaining good accuracy, the comprehensibility of

the learned model is also important, because an ensemble aggregates multiple models. Among my favorite parts of the book: A discussion of the alternative ways to achieve this objective: e.g. reduction of the ensemble to a single model.

It is always exciting to read a new book of a prominent researcher in the field. Zhi-Hua Zhou's book certainly qualifies in this category. Discussion in the book starts from a theoretical foundation, but the author also includes many references to successful applications, which makes it a good book both for the researcher and the practitioner. Moreover, this book is not written from a single point of view, but rather includes the view from pattern recognition, data mining as well as (to a lesser extent) statistics.

Important algorithms/approaches are discussed in pseudo-codes, which facilitates the understanding. The author does not just provide the math, but also a clear explanation of the reasoning behind it. The discussion starts with the basic algorithm, and then introduces a number of improvements that have been published in leading scientific journals. At the end of each chapter, there is always a "further readings" section providing hints for literature reading.

What I missed in this book? Some of the statistical methods (logistic regression), references to software and hybrid ensembles. This should be seen as suggestions for a second edition of the book, rather than as real problems. A book is always a compromise. Unlike a website, a book has to be balanced, which means one cannot provide asymmetric depth in the different topics.

In sum, this book deserves a special place in my library. It is well-written, and provides a very clear explanation of the different ensemble approaches including the intuition behind the algorithms why some of them work so well, and most of all, it provides an comprehensive overview of the alternative approaches (as opposed to the academic papers, where it lies scattered in thousands of (small) contributions).

THE BOOK:

ZHI-HUA ZHOU (2012), ENSEMBLE METHODS: FOUNDATIONS AND ALGORITHMS, 236 P. BOCA RATON, FL: CHAPMAN & HALL/CRC.
ISBN: 978-1-439-830031

ABOUT THE REVIEWER:

DIRK VAN DEN POEL
Marketing Analytics at Faculty of Economics and Business Administration, Ghent University, Belgium. Contact him at: dirk.vandenpoel@ugent.be

RELATED CONFERENCES, CALL FOR PAPERS/PARTICIPANTS

TCII Sponsored Conferences

WI 2013
The 2013 IEEE/WIC/ACM International Conference on Web Intelligence
Atlanta，USA
November 17-20, 2013
http://cs.gsu.edu/wic2013/wi/

The 2013 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2013) and the 2013 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2013) will be held in Atlanta, USA, Nov. 17-20, 2013. The two co-located conferences are sponsored by IEEE Computer Society Technical Committee on Intelligent Informatics (TCII), Web Intelligence Consortium (WIC), and ACM-SIGART.

Following the great successes of the previous WI-IAT conferences, the WI-IAT 2013 will provide a global forum for scientists, engineers and educators to present the latest WI-IAT technologies, discuss how to develop future intelligent systems for complex applications.

WI-IAT 2013 will have various workshops, WI-IAT technical sessions, tutorials and panels. WI-IAT 2013 will have keynotes, a social reception together with the poster session and industry demo, and a banquet. Attendees only need to register once to attend all technical events at WI-IAT 2013.

Web Intelligence focuses on scientific research and applications by jointly using Artificial Intelligence (AI) (e.g., knowledge representation, planning, knowledge discovery and data mining, intelligent agents, and social network intelligence) and advanced Information Technology (IT) (e.g., wireless networks, ubiquitous devices, social networks, semantic Web, wisdom Web, and data/knowledge grids) for the next generation of Web-empowered products, systems, services, and activities.

IAT 2013
The 2013 IEEE/WIC/ACM International Conference on Intelligent Agent Technology
Atlanta，USA
November 17-20, 2013
http://cs.gsu.edu/wic2013/iat/

IAT 2013 will provide a leading international forum to bring together researchers and practitioners from diverse fields, such as computer science, information technology, business, education, human factors, systems engineering, and robotics, to (1) examine the design principles and performance characteristics of various approaches in intelligent agent technology, and (2) increase the cross fertilization of ideas on the development of autonomous agents and multi-agent systems among different domains. Intelligent Agent Technology explores advanced intelligent systems and their broad applications in computer science and engineering, big data mining, biomedical informatics, health informatics, social networks, education, robotics, security, etc.

ICDM 2013
The Twenty-First IEEE International Conference on Data Mining
Dallas, Texas, USA
December 8-11, 2013
http://icdm2013.rutgers.edu/

The IEEE International Conference on Data Mining series (ICDM) has established itself as the world's premier research conference in data mining. It provides an international forum for presentation of original research results, as well as exchange and dissemination of innovative, practical development experiences. The conference covers all aspects of data mining, including algorithms, software and systems, and applications. In addition, ICDM draws researchers and application developers from a wide range of data mining related areas such as statistics, machine learning, pattern recognition, databases and data warehousing, data visualization, knowledge-based systems, and high performance computing. By promoting novel, high quality research findings, and innovative solutions to challenging data mining problems, the conference seeks to continuously advance the state-of-the-art in data mining. Besides the technical program, the conference features workshops, tutorials, panels and, since 2007, the ICDM data mining contest.

Related Conferences

AAMAS 2013
The Twelfth International Conference on Autonomous Agents and Multi-Agent Systems
Saint Paul, Minnesota, USA
May 6- 10, 2013
http://aamas2013.cs.umn.edu/

The AAMAS conference series was initiated in 2002 in Bologna, Italy as a joint event comprising the 6th International Conference on Autonomous Agents (AA), the 5th International Conference on Multiagent Systems (ICMAS), and the 9th International Workshop on Agent Theories, Architectures, and Languages (ATAL).

Subsequent AAMAS conferences have been held in Melbourne, Australia (July 2003), New York City, NY, USA (July 2004), Utrecht, The Netherlands (July 2005), Hakodate, Japan (May 2006), Honolulu, Hawaii, USA (May 2007), Estoril, Portugal (May 2008), Budapest, Hungary (May 2009), Toronto, Canada (May 2010), Taipei, Taiwan (May 2011)

and Valencia, Spain (June 2012). AAMAS 2013 will be held in May in Saint Paul, Minnesota, USA.

AAMAS is the largest and most influential conference in the area of agents and multiagent systems, the aim of the conference is to bring together researchers and practitioners in all areas of agent technology and to provide a single, high-profile, internationally renowned forum for research in the theory and practice of autonomous agents and multiagent systems.

AAAI 2013
The Twenty-Seventh AAAI Conference on Artificial Intelligence
Bellevue, Washington, USA
July 14-18, 2013
http://www.aaai.org/Conferences/AAAI/aaai13

The Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI 2013) will be held July 14–18, 2013 in Bellevue, Washington, USA. The purpose of this conference is to promote research in artificial intelligence (AI) and scientific exchange among AI researchers, practitioners, scientists, and engineers in affiliated disciplines. AAAI 2013 will have multiple technical tracks, student abstracts, poster sessions, invited speakers, and exhibit programs, all selected according to the highest reviewing standards. AAAI 2013 welcomes submissions on mainstream AI topics as well as novel crosscutting work in related areas.

SDM 2012
The Thirteenth SIAM International Conference on Data Mining
Austin, Texas, USA
May 2- 4, 2013
http://www.siam.org/meetings/sdm13/

Data mining is an important tool in science, engineering, industrial processes, healthcare, business, and medicine. The datasets in these fields are large, complex, and often noisy. Extracting knowledge requires the use of sophisticated, high-performance and principled analysis techniques and algorithms, based on sound theoretical and statistical foundations. These techniques in turn require powerful visualization technologies; implementations that must be carefully tuned for performance; software systems that are usable by scientists, engineers, and physicians as well as researchers; and infrastructures that support them.

This conference provides a venue for researchers who are addressing these problems to present their work in a peer-reviewed forum. It also provides an ideal setting for graduate students and others new to the field to learn about cutting-edge research by hearing outstanding invited speakers and attending tutorials (included with conference registration). A set of focused workshops are also held on the last day of the conference. The proceedings of the conference are published in archival form, and are also made available on the SIAM web site.

IJCAI 2013
The Twenty-Third International Joint Conference on Artificial Intelligence
Beijing, China
August 5-9, 2013
http://ijcai-2013.org/

IJCAI is the International Joint Conference on Artificial Intelligence, the main international gathering of researchers in AI. Held biennially in odd-numbered years since 1969, IJCAI is sponsored jointly by IJCAI and the national AI societie(s) of the host nation(s). IJCAI 2013 will be held in China, August 5-9, 2013. Submissions are invited on significant, original, and previously unpublished research on all aspects of artificial intelligence. The theme of IJCAI 2013 is ”AI and computational sustainability”. AI can play a key role in addressing environmental, economic, and societal challenges concerning sustainable development and a sustainable future. AI techniques and methodologies can be exploited to help address sustainability problems and questions, for example to increase the efficiency and effectiveness of the way we manage and allocate our natural and societal resources. The study of sustainability questions will also enrich and transform AI, by providing new challenges. The conference will include a special track dedicated to papers concerned with all these aspects.

IEEE Computer Society
1730 Massachusetts Ave, NW
Washington, D.C. 20036-1903

Non-profit Org.
U.S. Postage
PAID
Silver Spring, MD
Permit 1398