INFORMATION
TECHNOLOGY JOURNAL

Improving Classification Accuracy of IC Packaging Products Database Based on Variable Precision Rough Sets

Yung-Hsiang Hung
Department of Industrial Engineering and Management,
National Chin-Yi University of Technology, 35, Lane215, Section 1,
Chung-Shan Road, Taiping, Taichung, 411, Taiwan, Republics of China

Abstract: The aim of the study is to present the Variable Precision Rough Sets (VPRS) methodology to improving the classification accuracy for IC Packaging Product Database. The selection of proper packaging and manufacturing procedures is one of the priorities in IC design operations. Separation of industrial production characteristics require that IC designers have to ask the IC packaging engineers down the supply chain to confirm information regarding related product information, such as IC packaging type, size, functional features and price before they can select IC packaging products and determine IC product design procedures. In response to these product information demands from IC designers, service staffs of IC packaging factories integrate information from various departments to provide feedback to the designers. However, as the related information is very complex and professional, precious time is wasted during their communications, which instigates a failure to meet the demand of quick response. Furthermore, the IC product design and development cost and timing may be affected by lack of information. In an age of IC packaging technological breakthroughs, IC packaging structure and technological capabilities have already become key stages of IC design and manufacturing processes. Hence, it has become a very significant topic of the design industry, to effectively and rapidly, obtain ample product related information of IC packaging to meet their operational demands and enable them to improve operational costs and shorten processes. The present study adopts the VPRS method, an improved rough set theory to be applied to classification accuracy of the IC packaging product database; then compare the highest accuracy values with the values and rules obtained from the Johnson rough algorithm. The experimental results prove that proper β values, based on VPRS, are able to improve IC packaging product classification accuracy to obtain more consistent values and simpler rules regarding maximum value.

Key words: IC packaging type, IC design house, variable precision rough sets, classification accuracy

INTRODUCTION

In an era of communication and entertainment, growth of consumer electronics is exploding. Consumer demand for increased mobility, wireless connectivity and advanced features have paved the way for a variety of new products, including advanced mobile handsets, PDA, digital still cameras and camcorders, portable music players and many others. These trends, along with the broad range of end equipment emerging, require a large diversity of new Integrated Circuit (IC) package types to meet specific applications or markets. Increased device complexity will generate an explosion of new creative and disruptive, technology packaging solutions and in some markets and applications, packaging technology will become a key differentiator when making purchasing decisions (Bolanos, 2005). In fact, high-density packaging offer a host of benefits including performance

improvements, such as shorter interconnect lengths between die, resulting in reduced time of flight, lower power supply inductance, lower capacitance loading, less cross talk and lower off-chip driver power. These high-density packages result in a smaller overall package, when compared to packaged components performing the same functions, which results with I/O demands to the system board becoming significantly reduced. By sweeping several devices into one package, the system board complexity is simplified, thereby by reducing total opportunities for error at the board assembly level, according to the report High-Density Packages (MCMs, MCPs. SiPs): Market Analysis and Technology Trends, recently published by The Information Network, a New Tripoli, PA-based market research company. Today much of this development work occurs at assemblers in Taiwan. Many industrial consumers are fabless semiconductor companies, such as the IC Design House, which depends

on silicon foundries for front-end processing and on packaging foundries for backend processing. Crowley (1999) pointed out these IC Design Houses depend on innovative packaging developments to offer competitive packaging solutions for demand applications. Hence, IC packaging foundries are beginning to take the lead in IC package development. IC Design Houses and assembly houses are different firms and IC design is quite different between different IC products. In other words, the collaboration model is also different. However, in the IC design and development process, IC package type selection is very important to the IC designer. That is, the designer has to select a suitable package type then design a suitable IC chip and fabricate it, which can cause IC packaging design to be very difficult to integrate. Recently, network technologies of data mining methods have become a large part of management theory. The two most important two points are real time and cross-linking properties of networking. These two characteristics can fulfill knowledge management key points of success.

Along with the advancement of data mining, the research on the application of database was very intense in recent years. By the computer's automatically or semi-automatically analyzing lots of data, researchers could dig out the potential and useful information to find the meaningful relation or principles, such as the trend, characteristics and the process of correlation, the establishment of effective rules and models. There were many data mining technologies in which Neural Network (NN), Genetic Algorithm (GA), Decision Tree (DT), Support Vector Machine (SVM) and Rough set Theory (RST) were the most widely used at present (Braha and Shmilovici, 2002). The Variable Precision Rough Sets Model (VPRS) is an extension of the RST (Ziarko, 1993a). It was proposed to analyze and identify data patterns which represent statistical trends rather than functional. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level. Therefore, in this study we adopted VPRS. By sorting the existing IC package type related information, such as size, characteristics, design guidelines and application rules, we established classification principles for IC package products and assisted IC package practitioners to set up a prototype of IC package type automatic classification system, enabling clients, salespeople, service workers, or management staff to rapidly and accurately obtain the information they want. Thus different knowledge and experience levels of the recipients in the inquiry would not affect the acquisition of information, in addition, incorrect replies could be avoided and replies can be improved in speed and efficiency.

IC design was the upper-stream of IC industry. In the early state, Fabrication (Fab) manufacturers themselves were in charge of IC design, realizing the integrated operation of industrial value chain. They were called IDM (Integrated Design and Manufacture) later. Recently, the professional division business model between Wafer Foundry and IC design companies were proved feasible and they could successfully cooperate, which created two emerging industries-Wafer Foundry and IC design (Fabless). Thus, IC design companies began to develop rapidly. In the course of the design and development of IC products, the selection of IC package type was very important for IC designers, just as the decision of main structure construction methods for architects in designing buildings. In other words, IC designers had to select proper IC package type, based on which they could choose the follow-up design methods and manufacturing process of IC chips.

The evolution of IC package technology could be approximately divided into four stages: on the first state, appeared PTH package technology, like DIP, SIP, ZIP, S-DIP, SK-DIP and PGA; then emerged Surface Mount Technology (SMT), like QFP (Quad Flat Package), TSOP, FPG, LCC, PLCC and QFN. The development in this stage mainly depended on shrinking the package size and increasing I/O pin number, but basically both of them used the overhead as the carrier, connecting the electrode on chip to the lead foot on overhead with gold line. This belonged to peripheral package with its restriction on the shrink of package size and the increase of I/O pin number. Up to the third generation, package technology had evolved into area array method, like BGA (Ball Grid Array), Flip Chip (FC) and CSP (Chip Scale Package). Due to, area array package types and organic substrate phases in I/O and performance, power and small form factors were significantly improved. The 4th generation packaging types were bare die form factors, such as Flip Chip, WLCSP and DCA. Area Arrayed Flip Chip package types it will be the main stream of the future by offering peripheral technologies. IC package type development processes using flip chip production volumes began in 2000. The forecast for 2010 says there will be more than 12% of flip chip package types being used and by 2020; this will increase to 20% (Greig, 2007) to meet small form factor demands with packaging I/O requiring less than 300 pins, including BGA, TAB, CSP, FC, etc. CSP definition is package body size with a length and width ratio of less than 1.2, or chip area and package area ratio less than 1.5; for example QFP and BGA; BGA area is only 50% of QFP whereas CSP area is only 13%; the weight of CSP is only 20% of QFP. For Flip chip package, the area only 10% of QFP, details are shown in Fig. 1.

When IC package structure and technology had solemnly become the crucial manufacturing process in IC design and production, it was a very important topic as to
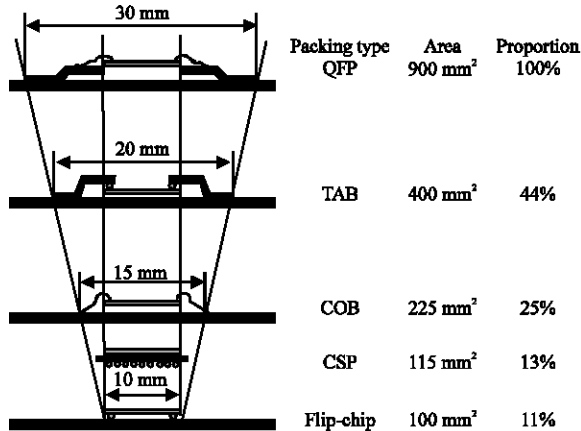
Fig. 1: QFP to Flip Chip area ration Source: Circuits Assembly, IC Insights (1999)

how to rapidly and effectively sort a large amount of information about package technology and products and to use the information to extend the horizon of enterprises staff and individuals in depth and width. The automatic classification technology applying data mining was one of the good tools to assist knowledge management, aiming to automatically mark the categories on the large amount of information. The users could utilize this category information to accurately shrink the scope of information search, greatly improving the efficiency and quality of search.

In recent years, along with the flourish of network technology, knowledge management had solemnly become an emerging subject in management theory. The main reason was that networks had two big characteristics of real time and connection that just satisfied a very important element to success in knowledge management: the ability of organization and cooperation (Piatetsky-Shapiro *et al.*, 1996; Chiang *et al.*, 2006; Han *et al.*, 2000; Gauch and Smith, 1993). By the network's ability of reaching everywhere and being connected rapidly, a good communication environment and tool inside and outside the organization was created, breaking through the restriction of time and space. In addition, combined with the advancement of data mining technology, it could effectively integrate the information silo problem formed by broken and scattered knowledge that could not be solved by the traditional knowledge management (Renpu and Wang, 2003).

Data mining focused on exploring how to dig out the potential and useful information from lots of data for the reference of decision-makers, such as finding the association of products sales from lots of trading data; discovering the hot research topic among literature

catalogues; grasping the topic trend from enormous document data; summing up useful information from internet websites (Kuo *et al.*, 2005; Chiang *et al.*, 2006). Mitra *et al.* (2002) and Piatetsky-Shapiro *et al.* (1996) described data mining is a multi-disciplinary research and application area that aims to discover novel and useful knowledge from vast databases, using methods ranging from artificial intelligence, statistics, machine learning, neural networks, pattern recognition, knowledge-based systems, knowledge acquisition, high-performance computing and data visualization. Mitra *et al.* (2002) described that data mining is a particular step in this process, involving the application of specific algorithms for extracting patterns (models) from data. The additional steps in the DM process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge and proper interpretation of the results of mining, ensures that useful knowledge is derived from the data. The application scope of data mining was involved in field data mining technology. Generally speaking, some of the main techniques used in data mining are associations, classifications and sequential or temporal patterns (Mai *et al.*, 2005). More commonly used types include decision trees, neural networks, genetic algorithms, fuzzy logic, RST rules induction, etc. The use of various types of application and different subjects of application can often lead to radically divergent results. The RST is suitable for problems that can be formulated as classification tasks and has gained significant scientific interest as a framework for data mining semiconductor manufacturing (Kusiak, 2001).

The theory of rough sets and their application methodology has been under continuous development for over 15 years now. The theory was originated by Pawlak (1991) in the early 1980s. It deals with the classificatory analysis of data tables. The data can be acquired from measurements or from human input; although in principle it must be discrete, there exist today methods that allow the processing of continuous values. The main goal of the rough set analysis is to synthesize definitions of approximate concepts from the acquired data. The main specific problems addressed by the theory of rough sets are (Ramanna *et al.*, 2002):

- Representation of uncertain, vague, or imprecise information
- Empirical learning and knowledge acquisition from experience
- Decision table analysis
- Evaluation of the quality of the available information with respect to its consistency and presence or absence of repetitive data patterns

- Identification and evaluation of data dependencies
- Approximate pattern classification
- Reasoning with uncertainty
- Information preserving data reduction

Rough set was first adopted to solve imprecise problems and at present, could be applied to the discontinuous numerical data in various fields, simplifying data, removing noise and dealing with classification problems. A number of practical applications for this approach have been developed in recent years in areas such as medicine (Grzegorz and Alicja, 2005), drug research (Chou *et al.*, 2007), process control (Kumara *et al.*, 2007), pattern recognition (Cyran and Mrózek, 2001) and others. Yang *et al.* (2007) has been applied to address customer complaints for IC packaging, making predictions regarding semiconductor manufacturing issues (Kusiak, 2001), evaluation product quality (Zhai *et al.*, 2002) and diagnosing Motherboard Electromagnetic Interference (EMI) test fault (Huang *et al.*, 2005).

This study discussed the classification problem of IC packaging types database. We dealt with inaccurate numerical values with VPRS and assisted IC package practitioners in building the prototype of IC package type automatic classification system, enabling clients, salesmen, service workers or management staff to rapidly and accurately get the information they wanted.

AN INTRODUCTION TO VARIABLE PRECISION ROUGH SETS

Rough sets mainly analyze uncertain data. The decision-analyzing method, proposed by Pawlak, in rough set was used to determine the crucial attributes of objects and build the upper and lower approximate sets of objects sets (Beynon *et al.*, 2000). Rough set theory deals with the analysis of this classificatory property of a set of objects. Each row in the table represents an object, for instance a case or an event. Each column in the table represents an attribute (a variable, a property, etc.) that can be measured for each object. This table is called an information system. More formally, it is a pair IS = (U, A), where, $U = \{x_1, x_2, K, x_n\}$ is a non-empty finite set of objects called the universe and $A = \{a_1, a_2, K, a_m\}$ is a non-empty finite set of attributes such as a: $U \rightarrow v_a$, a$\in$A. The set V_a is called the value set of a. A decision table is an information system IS = (U, A), for every set of attributes B$\subseteq$A, an equivalence relation, denoted by IND_{IS} and called the B-indiscernibility relation, is defined by:

$$IND_{IS} \left\{ (x,y) \in U^2 \mid \forall a \in B \ a(x) = a(y) \right\}$$

If $(x, y)_\in IND_{IS}$, then objects x and y are indiscernible from each other by attributes from B. The equivalence classes of the B-indiscernibility relation are denoted $[x]_B$. The subscript IS in the indiscernibility relation is usually omitted if it is clear which information system is meant. The primary notions of the theory of rough sets are the approximation space and lower and upper approximations of a set. Subsets of interest are sets of objects with the same value for the decision attribute. We can define the lower approximation of a set, as the set formed from the objects that certainly belong to the subset of interest and the upper approximation of a set, as the set formed from the objects that possibly belong to the subset of interest. A rough set is any subset defined through its lower and upper approximation. Let any concept X$\subseteq$U, X can be approximated using only the information contained within B by constructing the B-lower and B-upper approximations of X:

$$\underline{B}X \ \{x \in U : [x]_B \subseteq X\} \qquad (1)$$

$$\overline{B}X \ \{x \in U : [x]_B \cap X \neq \phi\} \qquad (2)$$

$$BN_B(X) = \overline{B}X - X\underline{B}X \qquad (3)$$

where, $\underline{B}X$ and $\overline{B}X$ are called as the B-lower and B-upper approximation of X, respectively (Fayyad *et al.*, 1996). If the set in upper and lower approximations was empty set, this set was called Crisp Set; conversely, it was called Rough Set ($BN_B \neq \phi$). And the upper and lower boundaries of rough set were respectively two crisp sets, close to our destination sets which were also called upper approximation (B-positive region of X) and lower approximation (B-negative region of X). $BN_B = \overline{B}X - \underline{B}X$ is called as the B-boundary region of X. An element in the lower approximation necessarily belongs to X, while an element in the upper approximation possibly belongs to X. To derive the lower approximation for X we need to list all those equivalence classes of the condition attributes whose members are contained within X. The upper approximation can equivalently be defined as comprising those equivalence classes $[x]_B$ that have a non-empty intersection with X. It comprises both the lower approximation and the boundary region in which an equivalence class has objects displaying a partial overlap with a decision class. The classification accuracy of the set X with respect to set B is defined as:

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|} \qquad (4)$$

where, |X| denotes the cardinality of $X \neq \phi$, obviously, $0 \leq 1$. If $a_B (X) = 1$, X is crisp with respect to B. If $a_B (X) < 1$, X is rough with respect to B. That is we can classify an element in X to B with a certainty of $a_B (X)$. However, the most accessible definition describes a Rough Set as any set specified, or measured, according to its upper and lower approximations. This means that the set cannot be generally characterized in terms of its attributes (Slowinski, 1993).

Some of the attributes may be superfluous. Attribute reduction techniques eliminate superfluous attributes and create a minimal sufficient subset of attributes of considered knowledge. The benefits of feature selection are twofold: it considerably decreases the running time of the reduction algorithm and increases the accuracy of the resulting model. Based on the concept of indiscernibility relation, a reduction in the space of attributes is possible. The aim is to retain only those attributes that preserve the indiscernibility relation (Smolinski *et al.*, 2004). The rejected attributes are redundant since their removal cannot worsen the classification. There are usually several such subsets of attributes and those, which are minimal, are called reducts. The exhaustive algorithm for calculation of all reducts is present along with implementations of approximate and heuristic solutions, such as genetic, covering and Johnson algorithms. The rough set theory-based application ROSETTA (Øhrn, 2000), employed a particular GA and Johnson's reduction algorithms (Johnson, 1974). The GA based algorithm for multiple reducts computation. The simplest heuristics is based on Johnsons strategy. The Johnsons reduction algorithm invokes a variation of a simple greedy algorithm to compute a single reduct. The algorithm has a natural bias towards finding a single prime deduction of minimal length.

Rough Set Theory assumes that the universe under consideration is known and all the conclusions derived from the model are applicable only to this universe. In practice, however, there is an evident need to generalize conclusions obtained from a smaller set of examples to a larger population. The variable precision rough sets model allows for a controlled degree of misclassification. It is an extension of the RST model, due to Ziarko (1993a) consists of a powerful generalization of Pawlak's original construction of RST. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level. It was proposed to analyze and identify data patterns which represent statistical trends rather than functional. The VPRS model extends the original approach by using frequency information occurring in the data to derive classification rules. It can be to increase the discriminatory capabilities of the rough set approach

by using parameter grades of conditional probabilities. A parameter β, a real number in the range $0 \leq \beta \leq 0.5$, is used in the VPRS model as a threshold in elementary sets that have both positive and negative examples. It represents a bound on the conditional probability of a proportion of objects in a condition class, which are classified to the same decision class. In VPRS, for any value of β and decision class, it may identified the condition classes which have the property that the largest group proportion of objects classified to the decision class is at least β, in which case each of the condition classes is classified to the decision class (Nian and Li, 2005). This introduced threshold relaxes the rough set notion of requiring no information outside the dataset itself. Let X, $Y \subseteq U$ then the relative classification error of X with respect to Y is the value:

$$\varepsilon(X, Y) = 1 - \frac{card(X \cap Y)}{card(X)} \qquad (5)$$

where, card denotes set cardinality. Observe that $\varepsilon(X, Y) = 0$, if and only if $X \subseteq Y$. A degree of inclusion can be achieved by allowing a certain level of error, β in classification:

$$X \overset{\beta}{\subseteq} Y \Leftrightarrow \varepsilon(X, Y) < \beta \qquad (6)$$

Using $\overset{\beta}{\subseteq}$ instead of lower and upper rough approximations, we define lower and upper β-approximation in analogy to Eq. (1, 2) by:

$$\underline{B}_\beta X = \cup \left\{ [x]_B \in U/B \mid [x]_B \overset{\beta}{\subseteq} X \right\} \qquad (7)$$

$$\overline{B}_\beta X = \cup \left\{ [x]_B \in U/B \mid \varepsilon([x]_B, X) < 1 - \beta \right\} \qquad (8)$$

So, variable-precision negative and positive regions of concepts are considered. The set are known, respectively, as the β-positive and β-negative regions. The set of condition classes whose proportions of objects classified to the decision class lie between the values $1-\beta$ and β is referred to as β-boundary region. However, the positive, negative and boundary regions in the original rough set theory can be defined as (Ziarko, 1993b):

$$POS_{B,\beta}(X) = \underline{B}_\beta X \qquad (9)$$

$$NEG_{B,\beta}(X) = U - \overline{B}_\beta X \qquad (10)$$

$$BND_{B,\beta}(X) = \overline{B}_\beta X - \underline{B}_\beta X \qquad (11)$$

Note that the lower and upper approximations as originally formulated are obtained as a special case with β = 0.0. Approximations of concept is very useful to define parameterized approximations with parameters tuned in the searching process for approximations of concepts. This idea is crucial for construction of concept approximations using rough set methods. Rough sets can thus approximately describe sets of patients, events, out comes, etc. that may be otherwise difficult to circumscribe.

CASE STUDY

IC package types are categorized into several package body families according to the exterior and function, such as Land Grid Array (LGA) and Flip Chip Ball Grid Array (FCBGA) etc. Different family can be separate as some group by out-line, function, reliability condition and process. Therefore, this study first analyzed the related information of IC package types to sort out all kinds of package forms and the related characteristics attributes, including the development of products, the electric property, the design of products, the development of manufacturing process, the reliability of products, the relevant knowledge of professional business engineering technical personnel in IC package industry. Then we summarized 5 IC package products families, such as QFP Package Family, PBGA Package Family, LGA Package Family and FCBGA Package Family. Every IC family or package body all had their own applicable IC design scopes, summing to 1 category member and 13 characteristics attributes: Package Size Range, Package Height (mm), Ball Pitch (mm), Lead Count (max.), Wafer Size (inch), Stacked Die Quantity, Substrate Layer, Frequency (GHz) max., MCM, Speed(Gbps) max., Power (W) max., Reliability (Level) and Reliability. Taking

LGA package family for an example, we used at most 2 the chip stacks of SLGA package body; at the same time, the fab sizes that could be operated were 8 inches and 12 inches; the total heights of package body were, respectively 1.2 and 1.4 mm; the sizes of package body ranged from 5×5 to 19×19 mm^2; the pitches of I/O port were 0.5, 0.65, 0.8 and 1.0 mm; the highest frequency in electric property was 5 GHz; the fastest transmission speed was 10 Gbps; the highest power was 4 W; the reliability condition was Level 3; the highest reflow solder temperature was 260°C; the carrier had 2 layers of plates. This study totally used 2,496 samples of data. The characteristics of every classification and the samples of data were shown in Table 1.

When the amount of information is big worked by the excessive amount of R and D offices, we can use the Rosetta decision-making system based on the VPRS to deal with the data in order to improve the work efficiency. We used ROSETTA GUI version 1.4.40, kernel version 1.0, RSES version 1.41, release build 22:41:48 Nov 8, 2000. The IC package database discussed in this study had 2,496 cases of samples, which were categorized, into five classifications. After having loaded the data files into the Rosetta system, the following steps use the 1996 case training set and the other 500 case for testing. Next is the discretization step, where each variable is divided into a limited number of value groups. The next step is creating reducts, which are subset vectors of attributes that IC package type classification rule generation with minimal subsets by Johnson's algorithm. The decision rules meant the principles summarized from the training data and these principles could be used to describe samples and their classifications, so the rules could be called classification knowledge, too. We choose the Johnson reducer algorithm. Options selected for this algorithm were

Table 1: IC package families and all of specification scope for their attribute

| Attributes | Variables of package family | IC package family | | | | |
		TFBGA "A"	LGA "B"	PBGA "C"	FCBGA "D"	QFP "E"
V_1	Package size (cm^2)	5×5~19×19	5×5~19×19	21×21~40×40	17×17~50×50	32×32
V_2	Package height (mm)	0.8~1.6	0.5~1.4	0.5~3.75	0.5~3.75	1.0~3.75
V_3	Lead pitch (mm)	0.5~1.0	0.5~1.0	1.0~1.27	1.0~1.27	0.5~0.65
V_4	Lead count (max.)	500	500	1156	1156~1521	80~256
V_5	Wafer size (inch)	8/12	8/12	8/12	8/12	8/12
V_6	Stacked die quantity	1/2/3/4/5/6	1/2/4	1/2/3	1/2/3/4/5/6	1/2/3
V_7	MCM	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
V_8	Substrate layer	2/4	2/4	2/4	2/4/6/8	NA
V_9	Frequency (GHz) (max.)	2.4~10	2.4~10	2.4~10	2.4~10	2.4~10
V_{10}	Speed (Gbps) (max.)	0.4~10	0.4~10	0.4~10	0.4~10	0.4~10
V_{11}	Power (Watt) (max.)	2~10	2~10	4~6	4~8	2~5
V_{12}	Reliability level	3	3	3	3	3
V_{13}	IR Temp. (deg C)	245/260	245/260	245/260	245/260	245/260
No. of attributes		13	13	13	13	12
No. of objects		632	424	352	816	272

Table 2: Classification accuracy for the training set by RST method

	A	B	C	D	E	No. of Obj	Accuracy	Coverage
A	289	224	0	0	0	513	0.563	1
B	13	332	0	0	0	345	0.962	1
C	0	0	275	0	0	275	1	1
D	0	0	0	649	0	649	1	1
E	0	0	0	0	214	214	1	1
Ture positive rate	0.96	0.6	1	1	1			

Total No. of tested objects: 1996, Total accuracy: 0.881, Total coverage: 1

Table 3: Classification accuracy for the test set by RST method

	A	B	C	D	E	No. of Obj	Accuracy	Coverage
A	60	59	0	0	0	119	0.504	1
B	41	38	0	0	0	79	0.481	1
C	0	0	77	0	0	77	1	1
D	0	0	0	167	0	167	1	1
E	0	0	0	0	58	58	1	1
Ture positive rate	0.59	0.39	1	1	1			

Total No. of tested objects: 500, Total accuracy: 0.8, Total coverage: 1

discernibility = object related (universe = all objects), table interpretation = Modulo with boundary thinning 0.1, no checks in the discernibility predicate or memory usage options and advanced parameters using approximate solutions with a hitting fraction of 0.725. The software-generated rules can then be used on the test set.

Meanwhile, the classification knowledge could be applied to testing and training database to see if it could correctly conduct classification. We also observed its accuracy and coverage. Coverage meant that the data could be classified no matter whether the classification was right or wrong. If the total coverage was 1, it indicated that every case of data could be classified out and the analysis results were shown in the Table 2 and 3. For the accuracy rate of the testing data in Table 3, we classified the 500 samples of testing data and there were totally 379 decision rules of the analysis results classifications. Through RST prediction, there were totally 60 case of right A data, 38 case of right B data, 77 case of right C data, 167 case of right D data, 58 samples of right E data. The sensitivity referred to the ratio between prediction values and real values. Taking A in the Figure for an example, the sensitivity of predicting classification A was 60/101=0.59; B, C, D, E were respectively 0.39, 1, 1, 1. The total accuracy rate was 0.8 and the total coverage was 1. But in the training data of Table 2, the total accuracy rate was 0.881 and the total coverage was 1. For the accuracy rate of the testing data in Table 3, we classified the 500 case of testing data and there were totally 379 decision rules of the analysis results classifications. Through Rosetta prediction, there were totally 60 case of right TFBGA package family (A) data, 38 case of right (B) data, 77 case of right (C) data, 167 case of right (D) data, 58 case of right (E) data. The sensitivity referred to the ratio between prediction values and real values. Taking TFBGA package family (A) in the Figure for an example, the sensitivity of predicting classification A was 60/101 = 0.59; LGA package family (B), PBGA

package family (C), FCBGA package family (D), QFP package family (E) were respectively 0.39, 1, 1, 1. The total accuracy rate was 0.8 and the total coverage was 1. But in the training data of Table 2, the total accuracy rate was 0.881 and the total coverage was 1.

Table 4 shown that using 10% and 20% as the testing group and two different groups of reduced set #1={v_2, v_6, v_7, v_9} and set #2={v_1, v_2, v_3, v_4, v_6, v_7, v_8, v_9} for experiment before comparing the biggest accuracy based on VPRS with the accuracy and rule quality obtained from Johnson rough set algorithm. It can be learnt from the experimental results that when β = 0.49, VPRS-based accuracy is better than Johnson-based rough set algorithm in the two experiments. The testing result is best in case of set #2 and = 0.49 as its accuracy is 100%. The testing result based on Johnson rough set algorithm is only 86.80% under same circumstances. In addition, when using 20% as the testing group and = 0.49 as the testing conditions, although its testing accuracy drops to 99.8%, being better than the 87.6% in case of Johnson rough set theory. In fact, Table 4 proves that the rules based on VPRS are simpler with higher supporting degree as well as higher consistency. The testing results of 78 classification rules in case of set #2 and β = 0.49 are as shown in Table 5 in the following:

Rule #1:
IF v1 = 1 then class = "A", support = (205/205).
It means IF "package size" = "5×5 mm²" then class = "TFBGA".
Rule #13
IF v2 = 5 and v6 = 1 then class = "A" support (27/27).
It means IF "package height" = "1.0 mm" and "Stacked Die Quantity" = 1 then class = "TFBGA"
Rule #78
   IF v1= 8 and v6 = 2 then class = "E" support (3/3).
   It means IF "package size" = "14×5 mm²" and "Stacked Die Quantity" = 1 then class = "QFP"

Table 4: Classification accuracy comparison in case of various β-values

			β Parameter							Johnson algorithm
			0.01	0.1	0.25	0.35	0.45	0.48	0.49	
10 (%)	ATTRIBUTES = $\{V_2, V_6, V_7, V_9\}$	A	0.45205	0.45205	0.57534	0.73973	0.86301	0.86301	0.90411	1.00000
		B	0.26087	0.26087	0.41304	0.54348	0.60870	0.60870	0.67391	0.28261
		C	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		D	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		E	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		Accuracy	0.70400	0.70400	0.768	0.84000	0.88800	0.88800	0.91200	0.86800
	ATTRIBUTES = $\{V_1, V_2, V_3, V_4, V_6, V_7, V_8, V_9\}$	A	0.97260	0.97260	0.97260	0.97260	0.97260	0.97260	1.00000	1.00000
		B	0.95652	0.95652	0.95652	0.95652	0.95652	0.95652	1.00000	0.28261
		C	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		D	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		E	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		Accuracy	0.98400	0.98400	0.98400	0.98400	0.98400	0.98400	1.00000	0.86800
20 (%)	ATTRIBUTES = $\{V_2, V_6, V_7, V_9\}$	A	0.39098	0.39098	0.64662	0.66165	0.74436	0.74436	0.83459	0.90226
		B	0.13158	0.26316	0.26316	0.26316	0.48684	0.64474	0.80263	0.35526
		C	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		D	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		E	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		Accuracy	0.70600	0.72600	0.79400	0.79800	0.85400	0.87800	0.92600	0.87600
	ATTRIBUTES = $\{V_1, V_2, V_3, V_4, V_6, V_7, V_8, V_9\}$	A	0.86466	0.86466	0.86466	0.87970	0.87970	0.87970	1.00000	0.902256
		B	0.77632	0.77632	0.77632	0.77632	0.77632	0.77632	0.98684	0.355263
		C	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		D	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		E	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
		Accuracy	0.93000	0.93000	0.93000	0.93400	0.93400	0.93400	0.99800	0.87600

Table 5: The VPRS method's rule quality

Rule	Mach	Decision rule	Decision attributes
1	205	$(v_2 = 4) => (class = \{A[205]\})$	A = 1
2	184	$(v_2 = 3) => (class = \{A[184]\})$	A = 1
3	460	$(v_8 = 2) => (class = \{A[460]\})$	A = 1
4	106	$(v_6 = 3) => (class = \{A[63], C[43]\})$	A = 0.59433 C = 0.50567
5	157	$(v_6 = 2) => (class = \{A[157]\})$	A = 1
6	23	$(v_6 = 4) => (class = \{A[23]\})$	A = 1
7	210	$(v_3 = 3) => (class = \{A[127], B[83]\})$	A = 0.6047 B = 0.3953
8	209	$(v_3 = 4) => (class = \{A[130], B[79]\})$	A = 0.622 B = 0.378
9	14	$(v_6 = 5) => (class = \{A[14]\})$	A = 1
10	13	$(v_6 = 6) => (class = \{A[13]\})$	A = 1
11	267	$(v_1 = 1) => (class = \{A[267]\})$	A = 1
12	245	$(v_1 = 2) => (class = \{A[245]\})$	A = 1
13	27	$(v_2 = 5)$ and $(v6 = 1) => (class = \{A[27]\})$	A = 1
14	100	$(v_2 = 6) => (class = \{B[100]\})$	B = 1
15	98	$(v_2 = 5) => (class = \{B[98]\})$	B = 1
16	228	$(v_8 = 1) => (class = \{B[228]\})$	B = 1
M	M	M	M
23	239	$(v_9 = 4) => (class = \{C[239]\})$	C = 1
24	292	$(v_8 = 4) => (class = \{C[127], D[165]\})$	C = 0.4349 D = 0.5651
31	101	$(v_1 = 7) => (class = \{C[42], D[59]\})$	C = 0.4158 D = 0.5842
32	115	$(v_2 = 6) => (class = \{C[41], D[74]\})$	C = 0.3565 D = 0.6435
33	66	$(v_1 = 5)$ and $(v7 = 2) => (class = \{C[23], E[43]\})$	C = 0.3484 E = 0.6516
34	21	$(v_1 = 3)$ and $(v8 = 2) => (class = \{C[21]\})$	C = 1
M	M	M	M
67	51	$(v_1 = 5)$ and $(v9 = 1) => (class = \{E[51]\})$	E = 1
76	36	$(v_1 = 7)$ and $(v10 = 1) => (class = \{E[36]\})$	E = 1
77	36	$(v_1 = 7)$ and $(v12 = 1) => (class = \{E[36]\})$	E = 1
78	3	$(v_1 = 8)$ and $(v6 = 3) => (class = \{E[3]\})$	E = 1

Obviously, the experimental results prove that the adoption of proper β based on VPRS is able to improve IC packaging product classification accuracy and produces more consistent and simpler rules in case of maximum accuracy. Among which, the bolded one represents the β-value in case of maximum classification accuracy.

CONCLUSION

In an age of IC packaging technological breakthroughs, IC packaging structure and technical capabilities have already become the key stages in IC design and manufacturing process. Hence, it is one of the very significantly important jobs in IC coordinated design

operations to effectively and rapidly obtain ample related product information regarding IC packaging technologies to meet the operational demands of IC designers to improve operational costs and shorten operational process. In the present study, it can be learnt from RST results that the classification accuracy in case of TFBGA and LGA are most undesirable, being at 50.4 and 48.1%, respectively. In fact, in IC packaging factories, it is hard for customer service staffs to determine what type of product to be provided upon the customer demands. The present study applies Variable Precision Rough Sets theory in IC packaging product database classification accuracy. The experimental results prove that setting proper β-values based on VPRS is able to IC packaging product classification accuracy as well as produce more consistent and simpler rules in case of maximum accuracy rate. The present study conducts experiments by using 10 and 20% as the testing groups and two different set #1 = $\{v_2, v_6, v_7, v_9\}$ and set #2 = $\{v_1, v_2, v_3, v_4, v_6, v_7, v_8, v_9\}$ for experiment before comparing the maximum accuracy based on VPRS and the accuracy and the rule quality obtained from Johnson rough set algorithm. It can be learnt from the experimental results that the accuracy value based on VPRS classification is better than that of the Johnson algorithm in case of two experiments. The application of the present study results to help IC packaging manufacturers to set an IC packaging automatic classification system for customer service staffs or the management staffs to rapidly and accurately obtain important IC packaging product information for effective improvement in feedback speed and efficiency.

ACKNOWLEDGMENTS

REFERENCES

Beynon, M., B. Curry and P. Morgan, 2000. Classification and rule induction using rough set theory. Exp. Syst., 17 (3): 136-148.

Bolanos, M.A., 2005. Semiconductor IC packaging technology challenges: The next five years. International symposium on electronics materials and packaging, 2005. EMAP., pp: 6-9.

Braha, D. and A. Shmilovici, 2002. Data mining for improving a cleaning process in the semiconductor industry. IEEE Trans. Semiconductor Manuf., 15 (1): 91-101.

Chiang, W.Y.K., D. Zhang and L. Zhou, 2006. Predicting and explaining patronage behavior toward web and traditional stores using neural networks: A comparative analysis with logistic regression. Dec. Support Syst., 41(2): 514-531.

Chou, H.C., C.H. Cheng and J.R. Chang, 2007. Extracting drug utilization knowledge using self-organizing map and rough set theory. Exp. Syst. Appl., 33 (2): 499-508.

Circuits Assembly. IC Insights, 1999.

Crowley, R., 1999. http://www.www.chipscalereview. com/issues/0799/columns2.htm.

Cyran, K.A. and A. Mrózek, 2001. Rough sets in hybrid methods for pattern recognition. Int. J. Intel. Syst., 16 (2): 149-168.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, 1996. The KDD process for extracting useful knowledge from volumes of data. Commun. ACM., 39 (11): 27-34.

Gauch, S. and J.B. Smith, 1993. An expert system for automatic query reformulation. J. Am. Soc. Inform. Sci., 44(3): 124-136.

Greig, W.J., 2007. Integrated Circuit Packaging. Ssembly and Interconnections, Springer.

Grzegorz, I. and W.D. Alicja, 2005. Rough sets approach to medical diagnosis system. Lecture Notes Comput. Sci., 3528: 204-210.

Han, S.Y., Y.S. Kim, T.Y. Lee and Yoon, 2000. A framework of concept process engineer with agent-base collaboration design strategies and its application on plant layout. Comput. Chem. Eng., 24 (2-7): 1673-1679.

Huang, C.L., T.S. Li and T.K. Peng, 2005. A hybrid approach of rough set theory and genetic algorithm for fault diagnosis. Int. J. Adv. Manuf. Technol., 27 (1-2): 119-127.

Johnson, D.S., 1974. Approximation algorithms for combinatorial problems. J. Comput. Syst. Sci., 9: 256-278.

Kumara, S., A. Nassehia, S.T. Newmana, R.D. Allenb and M.K. Tiwaric, 2007. Process control in CNC manufacturing for discrete components: A STEP-NC compliant framework. Robotics and Comput. Integrated Manuf., 23 (6): 667-676.

Kuo, R.J., J.L. Liao and C. Tu, 2005. Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce. Dec. Support Syst., 40 (2): 355-374.

Kusiak, A., 2001. Rough set theory: A data mining tool for semiconductor manufacturing. IEEE Trans. Elect. Packaging Manuf., 24 (1): 44-50.

Mai, C.K., I.V.M. Krishna and A.V. Reddy, 2005. Polyanalyst application for forest data mining. IEEE International on Geoscience and Remote Sensing Symposium, IGARSS. Proceedings, 2: 756-759.

Mitra, S., S.K. Pal and P. Mitra, 2002. Data mining in soft computing framework: A survey. IEEE. Trans. Neural Networks, 13 (1): 3-14.

Nian, F.Z. and M. Li, 2005. Attribute value reduction in variable precision rough set. Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and Technologies. PDCAT., pp: 904-906.

Øhrn, A., 2000. The ROSETTA homepage. [http://rosetta.lcb.uu.se/]. Norwegian University of Science and Technology. Department of Computer and Information Science.

Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning about Data. London, Kluwer Academic Publishers.

Piatetsky-Shapiro, G., R.J. Brachman, T. Khabaza, W. Klösgen and E. Simoudis, 1996. An overview of issues in developing industrial data mining and knowledge discovery applications. Proceedings of KDD., pp: 89-95.

Ramanna, S., J.F. Peters and T. Ahn, 2002. Software quality knowledge discovery: A rough set approach. In: Proceedings of 26th IEEE Annual International on Computer Software and Applications Conference, pp: 1140-1145.

Renpu, L. and Z.O. Wang, 2003. Mining classification rules using rough sets and neural networks. Eur. J. Operat. Res., 157 (2): 439-448.

Slowinski, R., 1993. Rough Set Learning of Preferential Attitude in Multi-Criteria Decision Making. In: Methodologies for Intelligent Systems, Komorowski, J.R. (Ed.). Lecture Notes in Artificial Intelligence, 689: 642-651.

Smolinski, T.G., D.L. Chenoweth and J.M. Zurada, 2004. Application of rough sets and neural networks to forecasting university facility and administrative cost recovery. Lecture Notes in Computer Science, 3070: 538-543.

Yang, H.H., T.C. Liu and Y.T. Lin, 2007. Applying rough sets to prevent customer complaints for IC packaging foundry. Exp. Syst. Appl., 32 (1): 151-156.

Zhai, L.Y., L.P. Khoo and S.C. Fok, 2002. Feature extraction using rough set theory and genetic algorithms-an application for the simplification of product quality evaluation. Comput. Ind. Eng., 43 (4): 661-676.

Ziarko, W., 1993a. Variable precision rough set model. J. Comput. Sys. Sci., 46 (1): 39-59.

Ziarko, W., 1993b. Analysis of uncertain information in the framework of variable precision rough sets. Foundations Comput. Dec. Sci., 18 (3-4): 381-396.