Int J Adv Manuf Technol (2002) 20:390–396
Ownership and Copyright


The International Journal of Advanced Manufacturing Technology

# A Novel Approach to Extract Knowledge from Simulation Results

X. Shi, J. Chen, H. Yang, Y. Peng and X. Ruan
Department of Plasticity Technology, Shanghai Jiaotong University, Shanghai, PR China

*The extraction of knowledge from simulation results is becoming increasingly important, as numerical simulation is being widely used in the engineering design process. Knowledge extraction systems face challenging problems as the databases of simulation results tend to be dynamic, incomplete, redundant, sparse, and very large. This paper describes a novel approach for handling them. A consistent object-oriented data model for finite-element analysis results has been created using EXPRESS-G, which has facilitated the construction of a database for the knowledge mining procedure. After briefly introducing Rough Sets Theory (RST) and principal component analysis (PCA), this paper investigates the capabilities and implementation of both methods for extracting knowledge from simulation results. The methodology developed has been applied to a real application in sheet metal forming simulation and the results are presented.*

**Keywords:** Data model; Knowledge discovery; Principal component analysis (PCA); Rough sets theory (RST); Simulation results

## 1. Introduction

Knowledge based engineering (KBE), is an engineering methodology in which knowledge about a product, e.g. the techniques used to design, analyse, and manufacture the product, is stored in a special product model. KBE is regarded as a fusion of AI, Database, and CAX techniques. The critical component of KBE systems is the knowledge base which contains facts and heuristics that represent human expert domain knowledge. In order to make a knowledge base complete, non-contradictory, and reasonable, knowledge engineers employ a variety of techniques, such as querying and interviewing, for eliciting information from the experts.

Owing to expert conservatism and the inability of the expert to explain the rules for decision-making, knowledge engineers seek other means to expand the rule set and verify the rules already in the knowledge base [1]. Across a wide variety of fields, very large collections of data can be quickly accumulated and often serve as a corporate resources – to be used for various purposes, many of which are unknown or unspecified at the time of collection. Therefore, there is an urgent need for a new generation of computational techniques and tools to assist humans in extracting useful knowledge from the rapidly growing volumes of data [2].

Since numerical simulation has been widely implemented in die development and the sheet metal forming process, a large number of simulation results have been accumulated and stored in corporate databases. In order to use these data more effectively, the exploitation of implicit knowledge from simulation results has become more and more important. This paper is an initial step towards a common framework that we hope will allow us to integrate simulation and design in a concurrent way. The methods of KDD (knowledge discovery from database) and data mining are first investigated. A framework to conduct knowledge discovery from simulation is proposed. Two mining algorithms, rough sets theory (RST) and principal component analysis (PCA), are described and implemented to extract knowledge effectively and discover the relation between formability evaluation criteria and primary process parameters. Finally, the discovery of forming rules from simulation results for a rectangular drawing process is chosen as an example to validate the proposed methodology.

## 2. KDD and Data Mining

In the field of engineering, the deep knowledge embedded in simulation models complements the shallow knowledge embedded in rules. A reasonable engineering view is that since the shallow rule-based knowledge is useful, discovering knowledge from simulation results is also required for integrating additional expert knowledge in the form of rules.

The term KDD was coined at the first KDD workshop in 1989 [3] to emphasize that knowledge is the end product of a data-driven discovery. We adopt the definition of KDD provided in Chapter 1 of [4].

*Correspondence and offprint requests to*: Dr X. Shi, Engineering Research Center for Net Shape Manufacturing, 339 Baker Systems, 1971 Neil Avenue, Columbus, OH, 4320, U.S.A. E-mail: xiaoxshi@hotmail.com

Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

Here, data are a set of facts and pattern is an expression in some language describing a subset of the data or a model applicable to that subset. The overall KDD process includes data selection, data preprocessing, data transformation, data mining, interpretation and evaluation, as shown in Fig. 1.

Data mining is a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns in the data. It is defined as an analytic process designed to explore large amounts of data in the search for consistent patterns and systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Pattern discovery problems in engineering include detecting faults in production and identifying anomalous forming parameters that could produce defects.

As a method to extract hidden predictive information from large databases, data mining is a powerful new technology with a great potential to help workers focus on the most important information in their data warehouses. Its tools sweep through databases, identify previously hidden patterns in one step and predict future behaviour, allowing engineers to make proactive, knowledge-driven decisions.

Recently, a number of data mining applications and prototypes have been developed for large data domains [5] including marketing, banking, insurance, and manufacturing. In addition, data mining has been applied to other types of data such as time series, spatial, web, and multimedia data. In the engineering area, simulation results always involve large-scale, dynamic, incomplete, and unstructured data, which cause much difficulty in the implementation of the data mining technique. As a result, before the mining process, a consistent data model for finite-element modelling (FEM) results should be created, which includes material, process, and analysis information. This model lends itself to the easy establishment of information flow in multidisciplinary problems and provides an appropriate model for knowledge discovery. However, since unpredictability and inaccuracy may exist in FEM analysis, the internal connection between the attributes of the developed model is ambiguous. Therefore, algorithms, which can assist in finding the principal component of all attributes and eliminating the redundant attributes, are chosen in this paper.

For any given problem, the nature of the data itself will affect the choice of tool. Consequently, in this paper, a variety of tools and technologies are integrated to find the best solution.

## 3. Framework for Knowledge Discovery from Simulation Results

The framework for knowledge discovery from simulation results, as illustrated in Fig. 2, includes the following main parts:

1. *Data quality improvement tools*. Because of the nature of FEM theory, simulation results will contain anomalous data on occasions. Therefore, it is indispensable to use anomaly detection tools and query language tools to detect inconsistencies that might exist in the database. Moreover, in order to conduct data analysis and knowledge discovery from different FEA results more conveniently, a consistent object-oriented data model is developed in this paper, which will also lead to a good database implementation and maintenance.
2. *Knowledge discovery tools*. The data mining technique to be applied depends very much on the application domain and the nature of the data available. In this paper, according to the characteristics of simulation results, rough sets theory first operates on the data to discover the implicit relationships that might exist among attributes and eliminate irrelevant attributes in the data set. Principal component analysis is then used to determine the driving principal governing the behaviour of the system and present it in the form of rules based on a reduced data set.
3. *Knowledge verification tools*. The set of discovered rules has to be verified for accuracy, consistency (no redundant or contradictory rules), and usefulness (rules showing the decision-making process) for the knowledge base being developed. Currently, there are no tools available to accomplish that. The knowledge verification process can use the feedback from the domain expert as well as the available domain knowledge specific to the application being considered for KBE development. Domain knowledge is defined as any information that is not explicitly presented in the database. For the cup drawing process, for example, the knowledge "fracture does not occur on flange area and

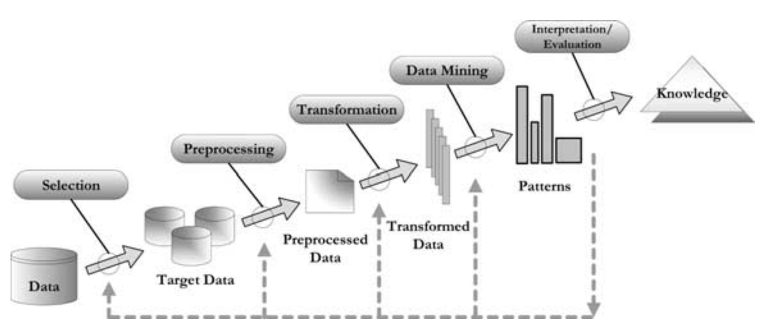


**Fig. 1.** The steps comprising the KDD process.

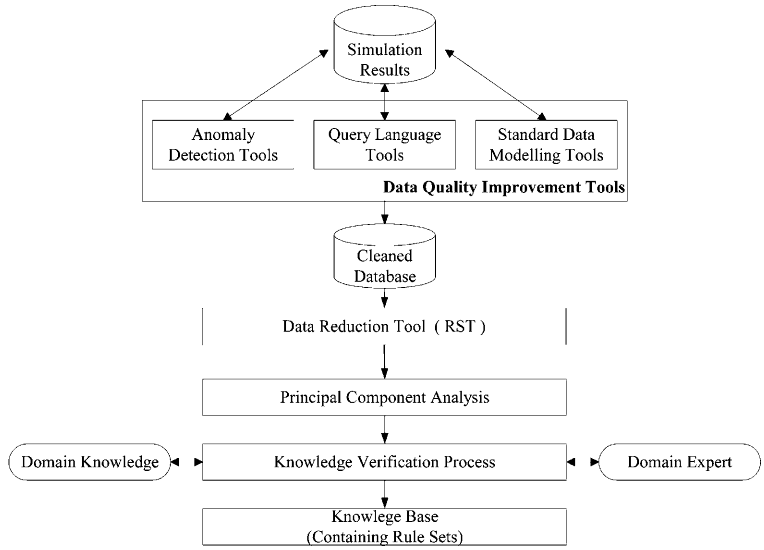


**Fig. 2.** A framework for rule discovery from simulation results.

wrinkle does not occur around bottom face" is considered to be domain knowledge.

In the following section, the capabilities and implementation of the tools identified in the framework for knowledge acquisition are described. These techniques show how to create a unique simulation data model and apply a data mining algorithm for an engineering problem.

## 4. Data Model and Mining Algorithms

### 4.1 Data Model for Simulation Results

This paper focuses on how the output produced by a specific FEA program can be handled in an object-oriented data model. Analysis programs are often designed as user-controlled applications and therefore cannot be used automatically by other programs. The outputs that are produced, in both graphical-user-interface (GUI) and batch mode, generally consist of a text file, which includes node and element data. Such a file is intended for human use in that it contains descriptive texts along with the data. In addition, because the data are in plain numerical text, there is no way to visualise the result or make associations among scattered data.

Although the techniques described in this paper have been employed with LS-DYNA, this paper focuses on more recent work in data modelling, for this case. LS-DYNA is a nonlinear FEA software, which can produce user-defined output files that contain FEA solution data with some user controls. The typical output file contains messages such as headers, titles, and notes. Column headings, numbers, and numerical values are also available. Some system messages and column headings reappear cyclically to allow a human reader to keep track of the data. Such messages are unnecessary for automated external use and only complicate the processes for exporting such data. Therefore, parts of a system message must be cleaned before processing the data file. The result file should only have elements and node numbers with attributes and solution data along with column headings identifying each number and its numerical value.

The data schema to represent FEA solution data is created from the result file and developed using EXPRESS-G, which is an object-oriented data modelling language developed for STEP [6]. The object-oriented data forms a communicational object entity that contains attributes and operation methods. Such an object is called a class and the classes for this FEA solution data model are sorted into five categories, FEElement, FENode, FEExtremaSet, FEMaterial, and FEInterestGroup. The data model is illustrated in Fig. 3.

All of the output data from LS-DYNA are stored within instances of the above five classes. The EXPRESS "L [1:?]" notation shown in Fig. 3 indicates ordered lists, which are implemented in Visual C++ as collections. The keys to the collections, such as element number, act as the link to the element object referred to. For example, 20 in the collection of elements refers to an instance of an FEElement whose element i.d. is 20. The FENode and FEElement contain nodal and elemental data in a chain list format. The FEExtremaSet contains a dictionary of the nodes whose thickness property has a maximum or minimum value, a dictionary of the elements

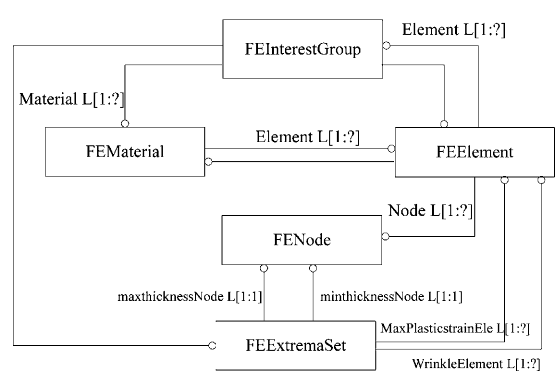


**Fig. 3.** EXPRESS-G schema for the FEA mesh model.

whose plastic strain property and wrinkling tendency have a maximum or minimum value, etc. The FEInterestGroup is a predefined group of elements of interest. For sheet metal forming simulation, more interest will be focused on forming defects, such as wrinkling, fracture, and springback in the process. According to forming evaluation criteria developed, elements, which have, or tend to produce, any kind of forming defect, are included in the group. On FLD, as shown in Fig. 4, the elements located in regions *A*, *B* and *C* will be regarded as elements of interest and are included in the FEInterestGroup. Customarily searched results, such as maximum and minimum values, are retrieved and stored in FEExtremaSet, which is also a part of each FEInterestGroup.

This structural representation of simulation results greatly facilitates the modelling of a database for knowledge discovery. Both the attributes of the FEExtremaSet and FEInterestGroup, such as maximum thinning and total number of elements in the FEInterestGroup, etc., and process parameters of each simulation are well organised and sorted in the database. Query language tools based on forming principles and rules are used to detect inconsistencies that might exist in the database.

### 4.2 Rough Set Based Data Reduction

During the stamping simulation and process optimisation procedure, it is often difficult to know exactly which kind of process parameters are relevant and important for the stamping process, and how they should be represented, so all that are believed to be useful are collected into the database. Hence, databases usually contain some attributes (parameters) that are undesirable, irrelevant, or unimportant for a given process, and so focusing on a subset of attributes is now a common practice. The important issues to be considered are therefore, to find the most relevant attributes and eliminate the irrelevant or unimportant attributes, according to the simulation results, without losing essential information about the original data in the databases. Rough set theory (RST) introduced by Pawlak [7] provides proper tools to analyse the set of attributes globally, and to handle vagueness and uncertainty inherent in making decisions. Using rough set theory, the minimal attribute set or reduct of the attribute in the generalised relation can be computed, and each reduct can be used instead of the whole attribute set, without losing any essential information. In the following section, the fundamental knowledge about RST is introduced.

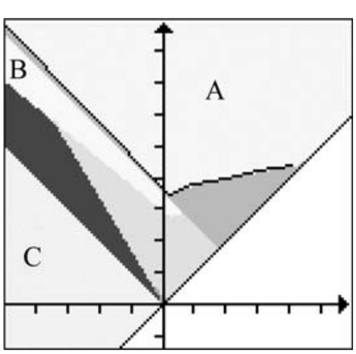


**Fig. 4.** Fracture and wrinkling region on FLD.

**Table 1.** A decision table.

| $U^{\Phi}$ | Bead force factor | BHF | Thickness | Plastic strain |
|---|---|---|---|---|
| 1 | Weak | Medium | Medium | High |
| 2 | Medium | Large | Medium | Medium |
| 3 | Strong | Very small | Small | Higher |

In this paper, a special case of information systems called a *decision table* or *attribute-value table*, is applied in RST. In a decision table, the columns are labelled by process parameters, geometrical parameters, final thickness, and strain value, whereas rows are labelled by each simulation event. An example of a decision table (attribute-value table) from cup drawing simulation results is given in Table 1.

A knowledge base is defined as a family of classifications in the universe of objects being investigated. Each family member forms a subset or category. The rough sets approach to data analysis hinges on two basic concepts, namely, the lower and upper approximations of a set, referring to: the elements that undoubtedly belong to the sets, and the elements that possibly belong to the set. Let $X$ denote the subset of elements of the universe $U$ ( $X \subset U$). The lower approximation of $X$ in every set of attribute $B$, which is denoted as $\underline{B}X$, is defined as the union of all these elementary sets which are contained in $X$. More formally:

$$\underline{B}X = \{x \in U{:}B(x) \subseteq X\} \tag{1}$$

$POS_B(X) = \underline{B}X$, called the B-positive region of $X$, is the set of these objects, which can, with certainty, be classified in the set $X$. Those objects that certainly do not belong, form the negative region. The upper approximation, denoted as $\overline{B}X$, is the union of these elementary sets, which have a non-empty intersection with $X$:

$$\overline{B}X = \{x \in U : B(x) \cap X \neq \phi\} \tag{2}$$

This can also be the theoretical union of the positive and boundary regions. A schematic representation of each region in the RST universe is shown in Fig. 5. In these regions, classified sets are used to help approximate the classification of the unclassified set.

The accuracy measurement of the set $X$ in $B$ is defined as

$$\mu_B(X) = card(\underline{B}X)/card(\bar{B}X) \tag{3}$$

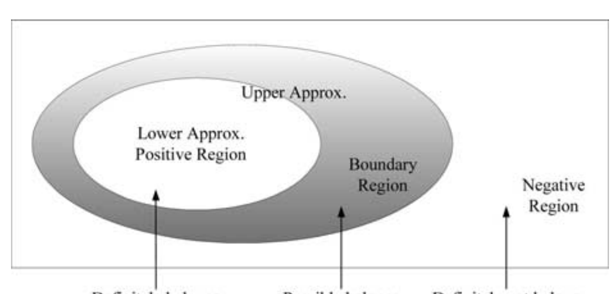


**Fig. 5.** The division of the RST universe.

The cardinality of a set, *card*(.), is the number of objects contained in the lower (upper) approximation of the set $X$. It is an index to depict the rough set degree. If the set of attributes is dependent, we can be interested in finding all possible minimal subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes (reducts) and in finding the set of all indispensable attributes (core). In order to introduce these two fundamental concepts of rough set theory, we should define the concept of "quality of classification".

Let $F = \{X_1, X_2, \ldots, X_n\}$, $X_i \subset U$, be a family of subsets of the universe $U$. If the subsets in $F$ do not overlap and the entity of them contains all elementary sets, then $F$ is called a classification of $U$, where $X_i$ are called classes. The lower and upper approximations of $F$ in $B$ are defined as:

$$\begin{aligned} \underline{B}(F) &= \{\underline{B}(X_1), \underline{B}(X_2), \ldots, \underline{B}(X_n)\} \\ \underline{B}(F) &= \{\underline{B}(X_1), \underline{B}(X_2), \ldots, \underline{B}(X_n)\} \end{aligned} \tag{4}$$

The quality of classification is defined as:

$$\eta_B F = \cup card \underline{B}(X_i) / card U \tag{5}$$

While deleting some attributes from $B$, a new set of attribute, $B'$, is then created. If $B'$ still possesses the same quality of classification, that is,

$$\eta_B F = \eta_{B'} F \tag{6}$$

$B'$ can be regarded as an $F$-reduct of $B$ and a reduced set of attributes.

A more comprehensive description of RST and how core and value reducts are used to remove redundancies can be found in [8,9].

### 4.3 Principal Component Analysis

Data mining takes advantage of advances in the fields of artificial intelligence and statistics. Both disciplines have been used to work on problems of pattern recognition and classification. Data mining does not replace traditional statistical techniques, rather, it is an extension of statistical methods that is in part the result of a major change in the statistics community. After attribute reduction by rough set algorithms, principal component analysis (PCA) is, as a result, an optimal dimensionality reduction statistical technique for capturing the variance of the data.

Fortunately, in data sets with many attributes, groups of variables often move together. One reason for this is that more than one variable may be measuring the same driving principal governing the behaviour of the system. To examine the relationships between a given data set of $n$ variables, of which some or all may be intercorrelated, it is possible to construct a set of $m$ orthogonal variables where $m < n$. This new transformed composite set of variables are linear combinations of the original variables which account for the variance in the original data and are called principal components. These new uncorrelated variables are derived in descending order of magnitude so that the first principal component accounts for most of the variation in the original data matrix. In other words, the axes representing the original variables are rotated individually to uncorrelate one with another which results in fewer than $n$ orthogonal axes spanning the $n$ space. In geometric terms this is equivalent to approximating the $n$-dimensional observation space by the projections of the observations onto a much smaller $m$-dimensional hyperspace.

The algorithm for finding this new set of uncorrelated data is called principal component analysis. Here, we shall discuss its use as a method for obtaining a low-dimensional representation of multivariate data so that an underlying relationship between each parameter for a specific sheet metal forming process may be identified. PCA is used when no hypothesis has been formulated as to which dimensions constitute the most relevant information. For instance, in a data set comprising multiple process parameters for sheet metal forming, PCA would extract the dimensions along the main directions of variation which often correspond to the most relevant information for forming defects, such as wrinkle, fracture, and unstretched. The main objective in applying PCA is to find out whether the first few components account for a large percentage of the variation in the original data. If this happens, then, in this case, the effective dimensionality of the problem may be less than $n$ space. It is hoped that the first few components retain meaningful essential information, which will help in understanding the data. The underlying structure of the data is identified and subsequently used in analysis so that a large number of variations is reduced to the investigation of a smaller set of transformed data. PCA is briefly discussed in the following.

Given a set of sample vectors $S = \{s^1, s^2, \ldots, s^p\}$ of dimension $n$, where $s^i = (s_1^i, s_2^i, \ldots, s_n^i)^T$, the goal of the PCA method is to find an orthonormal set of basis vectors (linear subspace) $U = \{u_1, u_2, \ldots u_m\}$, where $m < min(n, p)$, such that, the elements of $S$ can be recovered optimally in a least-squares error sense (Eq. (8)) from their projection into the space defined by $U$. The reconstruction of the $i$th sample vector is denoted as $\tilde{s}^i$:

$$\tilde{s}^i = \left( \sum_{j=1}^{m} u_j (u_j^T (s^i - \bar{s})) \right) + \bar{s} \tag{7}$$

where $\bar{s}$ is the average vector of $S$, and the samples set squared reconstruction error as:

$$\epsilon = \frac{1}{p} \sum_{i=1}^{p} \left\| s^i - \tilde{s}^i \right\|^2 \tag{8}$$

It can be shown [10], that the bases of this subspace can be computed as the $m$ eigenvectors with the highest associated eigenvalues of the samples covariance matrix $\Sigma$:

$$\Sigma = E[(s - \bar{s})(s - \bar{s})^T] \tag{9}$$

where $E[.]$ is the expected value, so that each element $\sigma_{ij}$ of the covariance matrix $\Sigma$ is the expected value of product of the deviation of random variables $i$ and $j$,

$$\sigma_{ij} = E[(s_i - \bar{s}_i)(s_j - \bar{s}_j)] \tag{10}$$

This set of selected bases is called the principal components of the sample set. In order to encode the sample data with the new base, their projections into these principal components are used.

## 5. Application

In the metal drawing process for rectangular pans and asymmetric panels, the deformation states and metal flow vary over the periphery of the part. During the forming process, the metal along the straight edges must be stretched to avoid wrinkling, while the corners are drawn in. This results in non-uniform metal flow over the periphery of the part as the metal flows more easily into the die cavity along the straight edges than in the corners. In order to ensure uniform metal flow over the periphery of the die, draw beads are added along the straight edges to restrict the flow of metal along these edges. The restriction force depends on the geometry of the draw bead. In order to understand the forming principle and find the parameters which exert the greater influence on the forming process, we carry out numerical simulations. From the results, the proposed algorithms are adopted for rule generation and relationship discovery.

The scheme for the drawing of rectangular pans is illustrated in Fig. 6. This example involves 20 numerical results described by six attributes: binder holder forces (BHF), bead depth ($D_b$), bead length ($L_{b1}$, $L_{b2}$), friction coefficient ($f$), and the location of bead centre-line ($x$), as shown in Fig. 6. Thinning and ratio of principal strains ($\epsilon_{min}/\epsilon_{max}$) are chosen as the decision attributes to evaluate fracture and wrinkling.

After data modelling of simulation results, FENode, FEElement, FEMaterial, FEInterestGroup and FEExtremaSet are created to represent the data structurally. Table 2 gives the results for the drawing process of a rectangular pan extracted from the relevant data and attribute of five classes.

An attribute-oriented induction is applied to this table, to obtain the generalised relation, as shown in Table 3. The criteria to classify each attribute are based on domain knowledge and rules, as shown in the note on Table 3.

Next the rough set method is applied to this generalised table to find the best reduct {BHF, $D_b$, $x$}, so the generalised relation is reduced further by removing those attributes: $L_{b1}$, $L_{b2}$, and $f$, resulting in Table 4.

According to the reduced table, after modifying Table 2 with regard to reduced attributes and records, PCA is conducted to find the most influential components (parameters) for the forming defects. The results and generated rule can be summarised as follows:

1. BHF and $D_b$ have different effects to the forming defects. In the total evaluation criteria, which are the sum of the fracture and wrinkling criteria, BHF is the first principal parameter, whereas $D_b$ is the second principal parameter. Therefore, in order to avoid forming defects, modifying BHF is preferred in this drawing process.

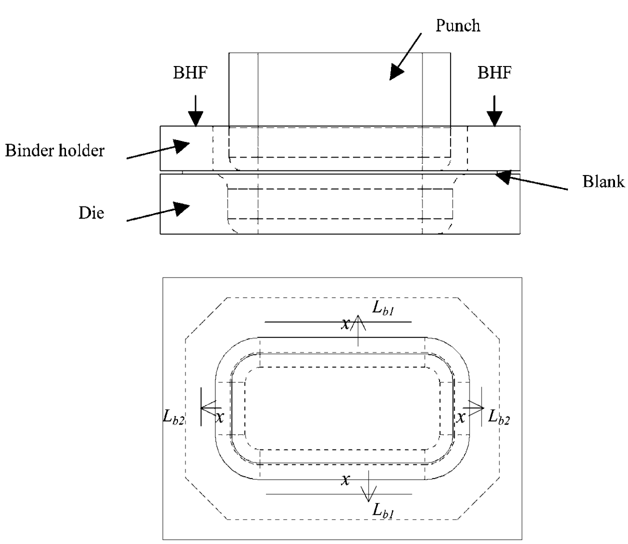


**Fig. 6.** The scheme of the drawing process for rectangular pans.

**Table 2.** Results for drawing process of rectangular pan.

| $U^{\Phi}$ | BHF (MPa) | $D_b$ (mm) | $L_{b1}$ (mm) | $L_{b2}$ (mm) | $f$ | $x$ (mm) | Thinning (%) | Wrinkling |
|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 6 | 135 | 65 | 0.125 | 10 | 16.87 | 0.34 |
| 2 | 10 | 4 | 112 | 47 | 0.125 | 10 | 16.47 | 0.27 |
| 3 | 16 | 9 | 112 | 47 | 0.125 | 15 | 34.7 | 0.17 |
| 4 | 16 | 9 | 135 | 65 | 0.25 | 10 | 38 | 0.12 |
| 5 | 15 | 4 | 112 | 65 | 0.25 | 12 | 27.3 | 0.28 |
| 6 | 15 | 6 | 100 | 30 | 0.3 | 15 | 35 | 0.15 |
| 7 | 5 | 6 | 100 | 35 | 0.3 | 5 | 12.4 | 0.44 |
| 8 | 5 | 4 | 101 | 30 | 0.1 | 5 | 13.5 | 0.42 |
| 9 | 16 | 9 | 112 | 65 | 0.1 | 20 | 31 | 0.31 |
| 10 | 15 | 6 | 117 | 47 | 0.2 | 25 | 17.4 | 0.22 |
| 11 | 23 | 6 | 90 | 40 | 0.2 | 25 | 32.5 | 0.16 |
| 12 | 5 | 4 | 90 | 40 | 0.1 | 22 | 10.8 | 0.52 |
| 13 | 8 | 4 | 70 | 50 | 0.125 | 10 | 17.2 | 0.24 |
| 14 | 10 | 6 | 70 | 30 | 0.2 | 5 | 20.5 | 0.38 |
| 15 | 10 | 7 | 80 | 30 | 0.22 | 12 | 15.2 | 0.24 |
| 16 | 16 | 5 | 80 | 40 | 0.25 | 20 | 27.8 | 0.276 |
| 17 | 16 | 4 | 90 | 65 | 0.2 | 8 | 28.4 | 0.19 |
| 18 | 15 | 2 | 110 | 65 | 0.1 | 12 | 17.5 | 0.4 |
| 19 | 5 | 2 | 110 | 50 | 0.11 | 15 | 12.4 | 0.45 |
| 20 | 10 | 2 | 123 | 49 | 0.2 | 20 | 19.3 | 0.37 |

**Table 3.** Decision table.

| $U^{\Phi}$ | BHF | $D_b$ | $L_{b1}$ | $L_{b2}$ | $f$ | $x$ | Thinning | Wrinkling |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 3 |
| 2 | 2 | 1 | 3 | 2 | 2 | 1 | 2 | 2 |
| 3 | 3 | 3 | 3 | 2 | 2 | 2 | 4 | 1 |
| 4 | 3 | 3 | 3 | 3 | 3 | 1 | 4 | 1 |
| 5 | 2 | 1 | 3 | 3 | 3 | 1 | 3 | 2 |
| 6 | 2 | 2 | 2 | 1 | 3 | 2 | 4 | 2 |
| … | … | … | … | … | … | … | … | … |

Note:
BHF: [5, 10], low (1); [10, 15 ], medium (2); [15, 20], high (3).
$D_b$: [2, 4], shallow (1); [4, 6], medium (2); [6, 9], deep (3).
$L_{b1}$: [70, 90], short (1); [90, 110], medium (2); [110, 135], long (3).
$L_{b2}$: [30, 40], short (1); [40, 50], medium (2); [50, 65], long (3).
$f$: [0, 0.1], low (1); (0.1, 0.2), medium (2); [0.2, 0.3], high (3).
$x$: [5, 10], near (1); [10, 20], medium (2); [20, 25], far (3).
Thinning: [0, 11], small thinning (1); [11, 20], medium (2); [20, 30], much thinning (3); [30, 40], crackling (4).
Wrinkling: [0.1, 0.2], small wrinkling (1); (0.2, 0.3], medium (2); [0.3, 0.4], high wrinkling (3); [0.4, 0.55], very high wrinkling (4).

**Table 4.** Reduction table.

| $U^{\Phi}$ | BHF | $D_b$ | $x$ | Thinning | Wrinkling | Process |
|---|---|---|---|---|---|---|
| 1 | Low | Shallow | Near | Medium | High wrinkling | Not available |
| 2 | High | Deep | Near | Crackling | Small wrinkling | Not available |
| 3 | Medium | Deep | Medium | Medium | Medium | Available |
| … | … | … | … | … | … | … |

2. In the location of the draw bead centre-line has a great effect on the wrinkling criterion. The best location lies approximately 15 mm outside punch open line.
3. For drawing rectangular pans, bead length and friction coefficient have little influence on the forming process.

## 6. Conclusion

In this paper, we have discussed the framework and main algorithms from the view of discovering knowledge from simulation results. The data mining solution is based on the discovery of the principal attributes and the implicit relationships in the given data set. We have highlighted the need and methodologies for creating an object-oriented data model for the simulation results. In addition, we have presented novel techniques based on RST and PCA for specific rule and knowledge discovery.

The approach discussed in this paper has been applied to a real application in sheet metal forming simulation and results of its application are also provided. The methodology developed in this paper has great potential for exploitation in engineering, such as application in the fields of on-line fault diagnosis.

## References


1. M. M. Owrang and O. F. H. Grupe, “Database tools to acquire knowledge for rule-based expert systems”, Information and Software Technology, 39, pp. 607–616, 1997.
2. J. Han and M. Kamber, Data Mining: Concepts and Techniques, 1st edn, Morgan Kaufmann, 2000.
3. G. Piatetsky-Shapiro and W. Frawley, Knowledge Discovery in Databases, MIT Press, 1991.
4. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, MIT Press, 1996.
5. R. J. Brachman, T. Khabaza, W. Klocsgen et al. “Mining business databases”, Communication of the ACM, 39(11), pp. 42–48, 1996.
6. ISO 10303, International Standard, Industrial Automation Systems and Integration – Product Data Representation and Exchange, Parts 1, 11, 22, 24, 31, 41, 201, 203, 214.
7. Z. Pawlak, “Rough sets”, International Journal of Information and Computer Science, 11(5), pp. 341–356, 1982.
8. B. Walczak and D. L. Massart, “Rough sets theory”, Chemometrics and Intelligent Laboratory Systems, 47, pp. 1–16, 1999.
9. L. Shen, F. E. H. Tay, L. Qu and Y. Shen, “Fault diagnosis using rough sets theory”, Computers in Industry, 43, pp. 61–72, 2000.
10. C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1996.