Rough Set-CBD Method of Stamping Die*

Yang Xinhua†, Mi Xiaozhen†, Li Xiaofeng†, Zhao Wenzhong†, and Deng Wu††

† School of Mechanical Engineering, Dalian Jiaotong University, Dalian, 116028 China
†† School of Software, Dalian Jiaotong University, Dalian, 116052 China

Abstract
On the basis of analyzing characteristics of stamping die products and their designing process, a design method combining rough set theory and case-based reasoning is proposed. Firstly, the paper introduces the information system of case-based reasoning based on rough set. Then, the paper presents storage strategies of stamping die case library and implementing examples. In order to retrieve cases similar to the case being studied in the library, the Nearest-Neighbors-based similarity measurement method is used. Thirdly, the paper gives an approach for reducing attributes and calculating weights to raise the accuracy and efficiency of case retrieval. At last, writers apply the method to develop a stamping die CAD system prototype.

Key words:
design method; stamping die; case-based reasoning; rough set theory

1. Introduction

With the rapid development of artificial intelligence (AI) technology, rule-based reasoning (RBR) and model-based reasoning (MBR) expert systems (ES) have been applied in stamping die designing field to promote intelligent CAD (ICAD) and design efficiencies. Generally, the design process of stamping die is very complicated and related to many aspects, such as designer's experiences and knowledge levels. These experienced design rules are difficult to formalize. Moreover, design process itself is imprecise and uncertain sometimes. For these reasons, it is hard to reasonably represent and acquire the design knowledge in RBR CAD system development. As for MBR, it is hard to control models when design problems become more complex. So solving massive design problems mainly depends on designer's experiences and genius. As a result, the development of the intelligent stamping die CAD system is limited. On these two situations, it is not available to apply expert's experiences, knowledge, and the methods accumulated in practices to the development of intelligent stamping die CAD system effectively.

Usually, a new design is solved by modifying or updating existed design proposals in actual design processes. That is, the process relies mainly on variant design and modularization design. Case-based reasoning (CBR) is a computational paradigm in which solutions used to solve old problems are used to a new problem. A case-based reasoner finds new solutions for relevant problems from its case library in which each case describes a problem and a solution to that problem. As more cases to new problems are stored, the library is updating constantly. Thus the reasoning in CBR is a reasoning based on memory. That is, new solutions are generated by retrieving the most relevant cases from the memory and adapted to fit new situations. This reasoning process is same with the actual stamping die design ones. It is proved that the case-based design (CBD) is an efficient approach in stamping die design field [1-3]. The design process is shown in Fig. 1.

In the implementation of domain-specified CBR systems, such as stamping die design, details about knowledge representation, case retrieval, inference procedure control, and attributes reduction are still on research due to massive information. Rough Set is a valid mathematical theory developed in recent years, which is able to deal with imprecise, uncertain, massive and vague information. In this paper, a new CBD method for stamping die is proposed based on rough set theory.

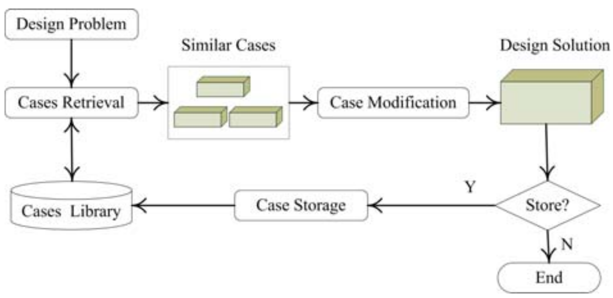


Fig.1. Case-Based Design Process of Stamping Die

2. Information System of Rough Set-CBR

An information system provides information or knowledge about real-world objects. According to the information system proposed by rough set theory, the information system of CBR can be defined as a tuple as follows [4].

* Sponsored by Dalian Science and Technology Foundation
Manuscript received November 5, 2006.
Manuscript revised November 25, 2006.

$$S = (U, R, V, f) \quad (1)$$

Where,

- U is a finite, non-empty set of objects, i.e. the set of all cases stored in cases library;
- R is a finite set of cases' attributes that is further classified into two disjoint subsets, condition attributes C and decision attributes D, that is, $R=C\cup D$, and $C\cap D=\Phi$;
- $V = \sum_{a\in R} V_a$ is the domain of all attributes, where V_a is the value set of attribute a ($a \in R$); and
- f is a total function. $f: U\times R\rightarrow V$ such that $f(x,a)\in V_a$ for every $x\in U, a\in R$.

The tuple can be abbreviated to $S = (U, R)$.

3. Storage strategies of the Information System

An information system as mentioned above can be viewed as a table (or relation) in a relational database, in which columns are labeled by attributes, row are labeled by objects, and the entity in column x and row a has the value $f(x,a)$. Each row in the table represents information or knowledge about some object in U. Some attributes are identified as decision attributes, and others are condition attributes. Objects are not distinguished by their attributes or relationships to other objects but their similarities.

The principal information in stamping dies includes workpiece names, materials, shapes dimensions, general die information (die names, types, heights, etc.), meta-data and structure information of component and part lists. Therefore, a set of tables are established to store all information. A table is related with others through the primary key and foreign keys. These keys may be IDs of workpieces, assembly drawings of dies or relevant parts. The table structure of punch cases is as Table 1.

Table 1. The Table Structure of Punch（table D）

Field Name	Data Type	Length	Explanation	Notes
PUNCH_ID	VARCHAR	50	The punch ID	Primary key
DID	VARCHAR	50	Die Assembly drawing ID	Foreign key
NAME	VARCHAR	50	Punch name	
FILE_NAME	VARCHAR	50	File name of punch model	
STRUCTURE	VARCHAR	50	Structure form	code
FIX_STYLE	VARCHAR	50	Fixed form	code
M_NUMBER	VARCHAR	50	Material ID	
MATRIX_ID	VARCHAR	50	corresponding matrix ID	Foreign key
……	……	……	……	……

Theoretically speaking, we can find appropriate old cases in this table and to solve a new punch design problem by calculating the attributes similarity between old cases and the new one. But in practice, it is hard to do just by using the data of a single table. For this reason, some view tables are defined to integrate related information. Table 2 shows the structure of the view table for punch design, in which source table and source field indicate where the view fields come from.

Table 2. The View Table Structure of Punch

Content	Field name	Data Type	Source table	Source field
Punch name	DNAME	VARCHAR	D	NAME
Part number of punch	P_CCODE	VARCHAR	D	PUNCH_ID
Structural form of punch	P_CSTRUFORM	VARCHAR	D	STRUCTURE
Fixed form	DFIXFORM	VARCHAR	D	FIX_STYLE
Material strength	MSTRENGTH	FLOAT	M	STRENGTH
Material type	MTYPE	VARCHAR	M	TYPE
Blanking force	BFORCE	FLOAT	A	BFORCE
Stripping force	SFORCE	FLOAT	A	SFORCE
Stripping force per mm	PSFORCE	FLOAT	A	PSFORCE
External dimension	PSIZE	VARCHAR	A	DIMENSION
Production lot	PRODUCTION	INT	A	PRODLOT
Workpiece thickness	STHICKNESS	FLOAT	A	THICKNESS
Working accuracy	WACCURACY	CHAR	A	MACCURACY
……	……	……	……	……

4. Retrieving Cases from the Case Library

Retrieving cases from cases library is different from querying or retrieving data from database. In order to retrieve cases similar to the current case from the library, some search strategies and related concepts, such as similarity, diversity factor and fuzzy-based similarity, have been introduced by some researchers. A common method is proposed based on nearest-neighbor algorithm, in which a similarity function is used to measure the relationship or similarity by the distance between two cases. The closer a case to the present case, the smaller the distance between them. Cases closest to the present case are designated as the nearest neighbors. The formula is as follows.

$$SIM_{i,j} = \sum_{k=1}^{n} w_k \times sim(V_{ik}, V_{jk}) \quad (2)$$

Where

- w_k is the weight of the k^{th} attribute. Under normal conditions, $w \in [0,1]$, $\sum_{1}^{n} w_i = 1$;
- V_{ik},and V_{jk} are the k^{th} attribute values of the i^{th} and the j^{th} cases respectively;
- $sim(V_{ik}, V_{jk})$ is the similarity of V_{ik} and V_{jk}. For two valued and discrete qualitative attributes,

$$sim(V_{ik}, V_{jk}) = \begin{cases} 0 & V_{ik} \neq V_{jk} \\ 1 & V_{ik} = V_{jk} \end{cases};$$

for quantitative and sequential qualitative attributes, such as big, medium and small,

$$sim(V_{ik}, V_{jk}) = 1 - \frac{|V_{ik} - V_{jk}|}{D_k},$$

where, D_k is the domain of the k^{th} attribute.

A key issue of nearest-neighbor-based retrieving is to determine the weights of attributes [5]. Typical methods for weight calculating include expert consulting, pairing comparison, statistical survey, and correlation analysis and so on [6]. The first three methods are based on hypothesis of expert's knowledge in their domains. The last one is more advanced than the formers and a statistical method. But, all of these methods depend on subjective estimations and experiences as a whole. Furthermore, in an information system, not all knowledge is always needed to define some categories available in the knowledge considered. In other words, not all condition attributes are necessary to categorize the objects in the information system. Some attributes are given just for describing cases, and may be redundant or dispensable with respect to the decision attributes. If these attributes are used in reasoning, they will make the retrieving more complicated, and may lead to illogical results.

Rough Set theory is a formal mathematical tool that can be applied, among other things, to reduce dimensionalities of dataset by providing a measure of the information content of datasets with respect to a certain classification, i.e. without losing information in the classification and decision rules. This reduction is a primary tool to analyze data in rough set, and makes deriving rules relatively easy from datasets. Furthermore, the main advantage of rough set theory is that the inference procedure lies on data itself completely, and does not need any preliminary or additional information about data ---- like probability in statistics, or basic probability assignment in Dempster-Shafer theory and grade of membership or the value of possibility in fuzzy set theory. For these reasons, we choose rough set to reduce attributes of stamping die's cases.

5. Calculating of Attribute Weights Based on Rough Set Theory

5.1 Definitions of rough set [7]

Definition 1. Let $B \subseteq R$, we define a binary relation $IND(B)$ based on B, called an indiscernibility relation, as follows:

$$IND(B) = \{(x,y) \in U^2 : \forall a \in B, f(x,a) = f(y,a)\} \qquad (3)$$

Obviously, *IND(B)* is an equivalence relation on *R*. i.e. It has the properties of reflexivity, symmetry and transitivity. The indiscernibility relation defines a partition in *U*. We use *U/IND(B)* (or simple *U/B*), called elementary sets, to denote the family of all equivalence classes of the relation *IND(B)*.

Definition 2. Let $X \subseteq U$ be a concept and $B \subseteq C$. The lower approximation of *X* based on $IND(B)$ is defined as

$$B_{-}(X) = \bigcup \{E_i \mid E_i \in U / IND(B) \wedge E_i \in X\} \qquad (4)$$

Informally, the lower approximation of *X* is the union of all those elementary sets that contained by *X*.

Definition 3. Let $P, Q \subseteq R$, the *P* positive region of *Q*, denoted by $POS_p(Q)$ is the set:

$$POS_p(Q) = \sum_{x \in U/Q} P_{-}(X) \qquad (5)$$

If $A \in P$ and $POS_{p-\{A\}}(Q) = POS_p(Q)$, we say that $A \in P$ is *Q*-dispensable in *P*. Informally, if a set of attributes and its superset define the same indiscernibility relation, then any attributes that belongs to the superset, but not the subset, is redundant.

Definition 4. Let *C* and *D* be subsets of *R*. We will say that *D* depends on *C* to degree *k* ($0 \leqslant k \leqslant 1$), if

$$k = \gamma(C,D) = \frac{card(POS_C(D))}{card(U)} \qquad (6)$$

Where, *card* represents the cardinality of a set. If $k = 1$ we say that *D* depends totally on *C*, and if $k < 1$, we say that *D* depends partially (to a degree k) on *C*, and if $k = 0$, then *D* does not depend on *C*.

The coefficient *k* expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition *U/D*, employing attributes *C* and will be called the degree of the dependency.

5.2 Calculating of Attribute Weights

Rough set theory provides a good tool for dealing with discrete attributes, but not suitable for continuous attributes. Continuous attributes must be discretized first. Some discretization methods are introduced in reference [8]. We can select and use these methods along with suitable discretization granularity to solve specified problems.

After discretization, we can build the information system $S = (U, C \cup D, V, f)$, where, *C* and *D* are sets of condition and decision attributes respectively. For $\forall a_k \in C$ ($k \in [1, m]$, *m* is the number of condition attributes), the significance of an attribute can be evaluated by measuring effects after removing the attribute (a_k) from an information table on classification. As shown in (6), the number $\gamma(C, D)$ expresses the consistency degree of the decision table, or the dependency degree between attributes *C* and *D*, or an approximation accuracy of *U/D* by *C*. We can check the coefficient $\gamma(C, D)$ changes before and after removing the attribute a_k, i.e., the difference between $\gamma(C, D)$ and $\gamma(C - \{a_k\}, D)$. The difference needs to be normalized. Then the significance of the attribute a_k can be defined as:

$$W_D(a_k) = 1 - \frac{card(POS_{C-\{a_k\}}(D))}{card(POS_C(D))} \qquad (7)$$

Obviously, $W_D(a_k) \in [0, 1]$. The more important the attribute a_k, the larger the number $W_D(a_k)$.

If and only if $W_D(a_k) > 0$, $a_k \in C$ is a indispensability attribute of condition attributes set *C*. If $W_D(a_k) = 0$, then a_k is redundant and should be reduced. On this rule, the significance of every attribute in the reduced information system is recalculated till all $W_D(a_k) > 0$.

Because $\sum_{a_k \in C} W_D(a_k)$ may not be equal to 1, we should normalize the $W_D(a_k)$ by using the following expression.

$$w_{(a_k)} = \frac{W_D(a_k)}{\sum_{a_k \in C} W_D(a_k)} \qquad (8)$$

6. Implementation Example

The method mentioned above is applied to develop a stamping die CAD system prototype. In our example, it is used to determine blank layout margins. Margins are remainders between blanks or blank and sheet in a layout. Blank layout margins can ensure the sheet with enough strength and rigidity, and then make feeding fluently, besides compensating position error and width direction error. As a result, service life of die and quality of blank section will be improved.

The influencing factors on determining margin values are included in mechanical properties and thickness of materials, shapes and dimensions of workpieces, layout formats, and feeding methods, and so on. Some design cases are listed in table 3 (For the sake of convenience, the actual data are reduced befittingly here). *C1* presents material thickness, C2 material type (L, M and H indicate low carbon steel medium carbon steel and high carbon steel respectively), C3 layout format, C4 degree of complexity, C5 production lot. The condition attributes are composed of *C1*, *C2*, *C3*, *C4* and *C5*, *D* is the decision attribute. The discretized results are shown in Table 4.

Table 3. Condition and decision attributes of margin cases

Case	**C1**	**C2**	**C3**	**C4**	**C5**	**D**
X1	2.0	L	B	simple	medium	1.6
X2	4.0	H	A	medium	large	2.0
X3	3.0	M	B	complex	small	2.2
X4	2.5	M	B	complex	small	2.0
X5	2.0	L	A	medium	medium	1.8
X6	2.5	H	B	complex	small	1.6
X7	3.0	H	B	complex	small	1.8
X8	2.0	L	A	simple	medium	1.7

Table 4. The attributes after discretization

Case	**C1**	**C2**	**C3**	**C4**	**C5**	**D**
X1	1	1	2	1	2	1
X2	4	3	1	2	3	4
X3	3	2	2	3	1	5
X4	2	2	2	3	3	4
X5	1	1	1	2	2	3
X6	2	3	2	3	1	1
X7	3	3	2	3	1	3
X8	1	1	1	1	2	2

According to definitions and formulae mentioned above, we can get the following results:

$U/IND(D)=\{\{X1,X6\},\{X2,X4\},\{X3\},\{X5,X7\},\{X8\}\}$

$U/IND(C-\{C1\}) = \{\{X1\},\{X2\},\{X3\},\{X4\},\{X5\},\{X6,X7\},\{X8\}\}$

$POS_c(D)=\{X1,X2,X3,X4,X5,X6,X7,X8\}$

$POS_{c-\{c1\}}(D)=\{X1,X2,X3,X4,X5,X8\}$

So, $W_D(c_1) = 1 - \frac{card(POS_{C-\{C_1\}}(D))}{card(POS_C(D))} = \frac{1}{4}$

In a similar way,

$$W_D(c_2) = \frac{1}{4}, W_D(c_3) = \frac{1}{4}, W_D(c_4) = \frac{1}{4}, W_D(c_5) = 0.$$

The above results indicate that the attribute *C5* should be reduced. So, we eliminate *C5* from the conditions attributes, and then recalculate the significance of each attribute. The results are as follows:

$$W_D(c_1) = \frac{1}{2},\ W_D(c_2) = \frac{1}{2},\ W_D(c_3) = \frac{1}{4},\ W_D(c_4) = \frac{1}{4}.$$

The weight of each attribute in the reduced table is calculated by using formulation (8).

$$W(c_1) = \frac{1}{3},\ W(c_2) = \frac{1}{3},\ W(c_3) = \frac{1}{6},\ W(c_4) = \frac{1}{6}.$$

Results above show that the major influencing factors on determining margin values are sheet thicknesses and material types. These two attributes have same weight values, i.e. $\frac{1}{3}$. The influence of layout format and complexity of workpiece are relatively small. This conclusion conforms to design knowledge and design criterion of stamping die [9].

There is a new problem depicted as (2.5mm, low chrome steel, layout in B format, complexity, mass production). According to the calculated weights and formula (2), we can figure out that the 4^{th} case is closest to the new one (the similarity is 0.833). So, we can use the information about the 4^{th} case as the reference for solving the new problem.

7. Conclusion

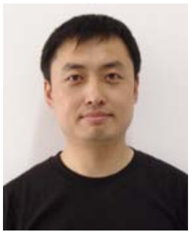
The essence of an intelligent system lies in its ability to deal with the uncertainties of information. That is, the system should have the ability to deduce a reasonable result from existed knowledge by means of uncertain reasoning. The characteristics of stamping die products and their designing process make it hard to reasonably represent and acquire the design knowledge in RBR CAD system development. As for MBR, it is hard to control models when design problems become more complex. In this paper, a new rough set-CBD method of stamping Die is proposed by combining CBR and rough set theory, and applied in a CAD system prototype. Reasoning process of CBR is in keeping with design process and methods of stamping die. Moreover, because of the characteristics of rough set as stated above, this method can overcome the shortcoming in traditional methods which depend on subjective experience excessively. It is available to apply expert's experiences, knowledge, and the methods accumulated in practices to the development of intelligent stamping die CAD system effectively.

Reference

1. Yang Xinhua, Mi Xiaozhen, et al. The Case-Based Design Methodology in Intelligent Stamping Die CAD System [J].Journal of Dalian Railway Institute. 2002,23(4):83-86.
2. WANG Dong-mei, YIN Guo-fu, et al. Study on model of mechanism collaboration innovative design based on case-based reasoning. Journal of Machine Design. 2004,5:4-6
3. Wang Zhengxiao, Pan Xiaohong, et al. Key Issues in Stamping die Design Using Case based Reasoning. MECHANICAL SCIENCE AND TECHNOLOGY. 1999, 18(5).748-749.
4. Zdzislaw Pawlak. Rough Set theory and its application to data analysis [J]. Cybernetics and Systems. 1998, 29(9): 661-668.
5. Park C S, Han I. A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction [J].Expert Systems with Applications. 2002,23 (3):255-264.
6. Sun Ling, Zhang Jinlong, et al. Study on Weighting Coefficient of Case Feature Attributes in CBR System Based on Rough Set Theory. Computer Engineering and Applications. 2003, 30: 44-46.
7. Zdzislaw Pawlak. Some Issues on Rough sets[J] . Transactions on Rough Sets I, LNCS 3100. 2004, 1~58.
8. Nguyen S H, Discretization methods with backtracking [A]. Proceedings of 5th European Congress on Intelligent Techniques and Soft Computing[C]. Heidelberg, Germany: Springer-Verlag, 1997. 201~205.
9. Xiao Xiangzhi. Computer Aided Design of Stamping Technology and Die [M]. National Defence Industry Press. 1996.

Yang Xinhua received the B.E. and M.E. degrees from Dalian Jiaotong Univ. in 1992 and 1998, respectively. He received the Dr. Degree from Dalian Univ. of Technology in 2003. He has been an assistant professor at Dalian Jiaotong Univ. since 2002. His research interest includes business intelligence, intelligent CAD/CAE System and its technologies.

Mi Xiaozhen received the B.E., M. E., and Dr. Eng. degrees from Dalian Jiaotong Univ. in 1982, Harbin Institute of Tech. 1988, and Dalian Univ. of Tech. in 2003, respectively. After working as an engineer assistant in Taiyuan Locomotive & Rolling Stock Works (from 1982), and a lecture and an associated Professor (from 1989) in the Faculty of Material Engineering Dept., Dalian Jiaotong Univ., she has been a full professor at Dalian Jiaotong Univ. since 2003. Her research interests include VPD, PDM, Collaborative Design, Manufacturing System Integration and Management.



Li Xiaofeng received the B.E. and M.E. degrees, from Jilin University of Technology in 1996 and 1999, respectively. After working as a research assistant (from 1999), a Lectuer (from 2001) in the Dept. of Mechanical Engineering of Dalian Jiaotong University. His research interest includes CAE of Vehicle Engineering, Structural optimum design.

Zhao Wenzhong received the B.E. degree from Northeastern University in 1967 and the M.E. degree from Dalian Jiaotong University in 1982. After working as a research assistant (from 1982), an assistant professor (from 1987) in the Dalian Jiaotong University. He has been a professor since 1994. His research interest includes Structural optimum design, CAE of Vehicle Engineering.

Deng Wu received the B.E. and M.E. degrees from Dalian Jiaotong Univ. in 2000 and 2006, respectively. He is an instructor at Dalian Jiaotong Univ. now. His research interest includes E-business and business intelligence.