Case representation and similarity in high-speed machining

Zuntong Wang *, Zhanqiang Liu, Xing Ai

School of Mechanical Engineering, Shandong University, Jinan, 250062, China

Received 27 December 2002; received in revised form 15 May 2003; accepted 20 May 2003

Abstract

Applying case-based reasoning in high-speed machining can utilize previous cases and experience in machining a new workpiece. This is helpful in overcoming the shortage of available high-speed machining data, and has significance for extending the application of HSM technology. For case representation, a model of case in HSM with relational data mode is proposed in the paper. With analysis of the characteristics of attributes and domain knowledge of HSM, the local similarities of case attributes are sorted into four classes: the numerical type, the enumerative type, the independent type, and the dependent type. The global similarity of cases is calculated by a composite similarity measure.

Keywords: Similarity; Case-based reasoning; High-speed machining

1. Introduction

By greatly increasing material removal rate, decreasing cutting time and cost and enhancing machining precision and surface quality, high-speed machining (HSM) has obviously become one of the most promising advanced manufacturing technologies in recent decades, and has been applied in a wide range of realms, such as aeronautics and astronautics, automobile, die and moulds industries, etc. [1–3]. On the operations side, many of the traditional cutting speeds and feeds used for years on the shop floor do not apply to the high-speed machining realm. Therefore, when cutting a work piece using HSM technology, key problems determine the appropriate cutting tool material according to the work piece material, and selecting feasible cutting parameters according to the cutting requirements and matching of work piece and cutting tool materials. Unfortunately, there is insufficient appropriate and systematic data for HSM application. By applying case-based reasoning, the previous machining data could be used to select cutting tool material and cutting parameters. Therefore, the shortage of high-speed machining data can be resolved.

This is helpful to extend the application of high-speed machining.

2. Brief discussion of case-based reasoning

Case-based reasoning (CBR) is a fairly new problem-solving methodology. In 1982, Schank developed the theory of learning and reminding based on retaining of experience in a dynamic, evolving memory structure [4], which is widely held to be the origins of CBR [5]. Instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced and concrete problem situations (cases). CBR is also an approach to incremental, sustained learning, since a new experience is retained each time when a problem has been solved, making it immediately available for future problems [6].

In the CBR system, a case means the description of a problem and its solution. When a new problem is to be solved, it is first described according to a specific mode. Then similar cases are retrieved from the case base, and the solution of the most similar problem is the suggested solution to the new problem. If necessary, the suggested solution is revised according to previous experience, domain knowledge and the actual situation.

* Corresponding author. Tel.: +86-0531-8392045.
E-mail address: wangzunt@163.com (Z. Wang).
of the problem. Thus, the confirmed solution to the new problem appears. Finally, a new case, composed of the new problem and its confirmed solution, is stored into the case base for future applications. Fig. 1 shows the CBR cycle [7]. We can simplify this process by the four REs as follows [5]:

1. Retrieve the most similar case(s);
2. Re-use the case(s) to attempt to solve the problem;
3. Revise the suggested solution if necessary;
4. Retain the confirmed solution and the new problem as a new case.

In effect, another RE, ‘Represent the case’, is very important in the CBR cycle. Case representation provides the basis for the four Res, which is further outlined in particular in Section 3.

A reasonable CBR system contains at least a case base, a robust indexing mechanism and a valid case retrieval method [6]. A case base is the foundation of the CBR system. The CBR system works based on the valid case retrieval method, and the robust index mechanism increases the CBR system’s efficiency.

Over recent decades, case-based reasoning has grown from a rather specific and isolated research area to a field of widespread interest, including manufacturing industries [8,9]. For examples, Takahashi et al. [10] combined the concept of CBR and knowledge re-use to solve real, large-scale manufacturing process design problems; Yang et al. [11] and Chang et al. [12,13] studied application of CBR in process planning of machining axisymmetric parts; Lei et al. [14] employed CBR in cold forging process planning; Tsatsoulis et al. [15] developed the TOLTEC planner in the manufacturing realm using CBR; Amen et al. [16] used CBR as a tool for selection of the material and heat treatment process, etc., while in our project, CBR is used to find the solution of machining a work piece by HSM technology. As parts of this project, the case representation and case similarity are investigated in this paper.

3. Cases in the high-speed machining domain

As mentioned above, case representation is very important in a CBR system. In order to define the description mode of case in the high-speed machining domain, it is necessary to analyze the problem thoroughly, i.e. the work piece and its machining by high-speed machining technology. When a work piece is to be machined, it is necessary to learn as much information as possible about the work piece, such as the work piece type, cutting surface type, cutting requirements, and the class, code, hardness and status of the material of work piece. Then, according to the above information, the machining process can be determined, which includes adopting certain kind of machining method and machine tool, using certain kind of cutting tools and corresponding cutting parameters, etc. The former corresponds to the problem description of the case, and the latter the solution of the case.

Fig. 2 shows the partial E-R diagram about the work piece machined by high-speed machining. In Fig. 2, the rectangles and their names designate entities; ellipses and their names denote attributes of correspondent entity; diamonds and their names illustrate relations among entities. For example, ‘work piece’ is an entity with two attributes ‘type’ and ‘type of cutting surface’, which describe the characteristics of the entity ‘work piece’. ‘Work piece material’ is another entity having four attributes: ‘code’, ‘class’, ‘hardness’, and ‘status’. ‘Made of’ is the relationship between entities ‘work piece’ and ‘work piece material’. For simplicity and generality, the work piece shown in Fig. 2 is not a specific one. Therefore, only the generally used entities, attributes and relationships are shown, whereas some concrete attributes and relations such as the name, identification code, dimensions and stiffness of work piece are not included.

According to Fig. 2, a case in high-speed machining realm could be represented as ‘case number + problem description + solution + evaluation’. In this representation, ‘case number’ is an exclusive identifier that uniquely designates the case; ‘problem description’ contains seven attributes: Type of work piece, Type of cutting surface, Cutting require, Material class of work piece, Material code of work piece, Material hardness of work piece, Material status of work piece; ‘solution’ contains ten attributes: Cutting method, Type of machine...
tool, Name of cutting tool, Code of cutting tool, Manufacturer of cutting tool, Material class of cutting tool, Material code of cutting tool, Cutting speed, Feed speed, Depth of cut; ‘evaluation’ shows whether the solution is reasonable to the problem, and whether the result is valid. In effect, the above case representation schema is described by the relational data mode. Fig. 3 shows case examples. In Fig. 3, wp means work piece.

4. Case retrieval and case similarity

4.1. Case retrieval

In the CBR system, case retrieval plays an important role. Different from query in database system, case retrieval in the CBR system is generally ‘imprecise’. This is because in a few circumstances, the cases with the same problem description as the new problem could be retrieved from the case base. More often, cases with similar problem description to the new problem are obtained. On the other hand, the CBR system retrieves cases only according to the attributes of the problem description, but not all the attributes of the case. In general, there are two kinds of case retrieval methods, i.e. the nearest neighbor retrieval and inductive retrieval [5]. We use the former in the paper. To speed up case retrieval, the case is indexed by three attributes: ‘Type of work piece’, ‘Type of cutting surface’, and ‘Material class of work piece’.

4.2. Case similarity

Similarity is a measure used to calculate the degree of one case similar to another. By calculating similarity, the most similar case to the target case (or problem) could be determined from the case base. Nevertheless, similarity as a notion is not fixed; it depends on the aspects under consideration. In this paper, we consider similarity with respect to the process solution of machining a work piece by HSM technology. A reasonable technique for calculating similarity is necessary for a successful CBR system.

Definition: given domains \( D_1, D_2, \ldots, D_n \), there is a relation mode \( R(A_1, A_2, \ldots, A_n) \) based on \( D_1, D_2, \ldots, D_n \), where \( A_i \in D_i (i = 1, 2, \ldots, n) \) is the \( i \)-th attribute of \( R \). For

![Fig. 3. Case examples.](image)
any two records (cases) \( (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}, \) the similarity \( \text{SIM}(u, v) \) between \( u \) and \( v \) is defined as follows:

\[
\text{SIM}(u, v) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + |x_i - y_i|}.
\]  

Distance is another expression of case similarity, which is defined as follows (Manhattan distance [17]):

\[
\text{DIST}(u, v) = \sum_{i=1}^{n} |x_i + y_i|.
\]

In general, the similarity varies in the range of \([0, 1]\), and distance \([0, \infty)\). Both similarity and distance discussed in this paper are symmetrical, i.e. \( \text{SIM}(u, v) = \text{SIM}(v, u) \), and \( \text{DIST}(u, v) = \text{DIST}(v, u) \).

It is obvious that the less distance between two cases, the more similar the two cases. Given a case \( u \in \mathbb{R} \), if there exists a case \( c \in \mathbb{R} \) that makes it true that for all cases \( c' \in \mathbb{R} \), the equation \( \text{SIM}(u, c) \geq \text{SIM}(u, c') \) or \( \text{DIST}(u, c) \leq \text{DIST}(u, c') \) is tenable, then the case \( c \) is called the nearest neighbor to case \( u \), and it is donated as \( \text{NN}_c(u) \) :

\[
\text{NN}_c(u, c); \exists \forall c \in \mathbb{R}, \forall c' \in \mathbb{R}: \text{SIM}(u, c) \geq \text{SIM}(u, c') \quad (3)
\]

or

\[
\text{NN}_c(u, c); \exists \forall c \in \mathbb{R}, \forall c' \in \mathbb{R}: \text{DIST}(u, c) \leq \text{DIST}(u, c'). \quad (4)
\]

### 4.3. Local similarity

In the practical problem-solving field, there are many attributes of cases with a non-numeric type of domain. Therefore, it is not feasible to calculate the similarity directly by Eq. (1), and it is necessary to consider the similarity for each attribute with certain type of domain. Local similarity \( \text{sim}(x, y) \) with a range within \([0, 1]\) means the similarity of any two values of the same attribute, where \( x, y \) are two values of the same attribute. The local similarity is also symmetrical, i.e. \( \text{sim}(x, y) = \text{sim}(y, x) \).

As mentioned in Section 4.1, the CBR system retrieves cases only according to the attributes of the problem description, therefore we need only investigate the local similarity of these attributes. The expression of local similarity of attribute varies with the domain of attribute. With analysis of the characteristics of case attributes and domain knowledge of HSM, local similarity is sorted into four classes:

1. The numeric type: This type of local similarity is defined as follows:

\[
\text{sim}(x, y) = \frac{1}{1 + |x - y|}.
\]  

The material hardness has this kind of local similarity.

2. The independent type: The attributes having this kind of local similarity have enumerated types of domain, where any two different values have no interaction with each other. This type of local similarity is designed as follows:

\[
\text{sim}(x, y) = \begin{cases} 
1 & x = y \\
0 & x \neq y
\end{cases}
\]  

The attributes ‘material class of work piece’ and ‘type of work piece’ have this kind of local similarity. The work piece is classified into three types: (1) the axisymmetrical type; (2) the trunk type; and (3) the others. The material of the work piece is sorted into six classes: (1) steel and cast steel, (2) cast iron; (3) stainless steel; (4) quenched steel and chilled cast iron; (5) super alloys and titanium alloy; (6) aluminum alloy and copper alloy.

3. The enumerated type: The attributes having this kind of local similarity have enumerated types of domains, where any two different values have a specific local similarity value. This type of local similarity is defined as follows:

\[
\text{sim}(x, y) = f(x, y)
\]

where \( f(x, y) \) is an enumerating function specified by the characteristics of the concrete attribute. The attributes Cutting required and Material code of work piece have this kind of local similarity. Attribute Cutting required has three different values: rough, semi-final, and final. If the attribute Cutting required of a new problem is Final, it is obvious that the solution of the case with attribute Cutting required being Final is of the most significance to the new problem under the circumstances that the other attributes having the same values respectively, and the solutions of the cases with attribute Cutting required being Semi-final and Rough follow. For calculating the local similarity, the attribute is assigned numeric values as shown in Fig. 4. Therefore the local similarity is expressed by the following equation:

\[
\text{sim}(x, y) = 1 - \frac{|x - y|}{M}
\]  

where \( M \) is the maximal value assigned to the attribute. Here \( M = 3 \), and 1, 2, 3 are assigned to attribute values Rough, Semi-final, and Final, respectively. Of course, the sequence of assigning values could be reversed. As mentioned above, the material of work piece is sorted into six classes, while within a class, the material is classified into different subclasses, and

<table>
<thead>
<tr>
<th>Rough</th>
<th>Semi-final</th>
<th>Final</th>
</tr>
</thead>
</table>

Fig. 4. Example of enumerated type of local similarity.
different material codes in a subclass. For example, steel and cast steel is a material class of a work piece, it is classified into four subclasses: (1) carbon steel, (2) low alloy steel, (3) high alloy steel, and (4) cast steel. Carbon steel is classified into four kinds, designated by corresponding material code, respectively, as shown in Table 1. The local similarity of attribute Material code of work piece is shown in Table 2. The local similarities of the remaining classes of work piece materials are treated in the same way, but not listed in detail in the paper for simplicity.

4. The dependent type: These are attributes that have local similarities relevant not only to their own values, but also to values of other attribute(s). In other words, this kind of attribute affects the case solution together with other attributes. This type of local similarity is defined as follows:

\[
sim(x, y) = f(x, y, x_1, y_1, x_2, y_2, \ldots, x_m, y_m)
\]

where \( f(x, y, x_1, y_1, x_2, y_2, \ldots, x_m, y_m) \) is an enumerated function of an attribute and its dependent attributes; \( x, y \) are values of the attribute whose local similarity is considered; \( x_i, y_i \in D_i (i = 1, 2, \ldots, m) \) are values of dependent attributes. In practice, the machining method adopted is determined by the attributes Type of cutting surface and Type of work piece; the cutting parameters used are affected by the attributes Material status of the work piece, Material class of work piece, Material code of work piece and Material hardness of work piece. Therefore, attributes Type of cutting surface, Material status of work piece, Material hardness of work piece and Material class of work piece are considered having a dependent type of local similarity. The dependent type of local similarity is expressed by a specific function or assigned specific values according to concrete situation and experience. The cutting surfaces are classified into seven types, which are: (1) round surface, (2) hole, (3) plane, (4) square shoulder surface, (5) slot, (6) curving surface, and (7) screw surface. The local similarity of attribute type of cutting surface is calculated as follows:

\[
sim_{tcs} = \sim_{twp} \times \sim_{surf}
\]

where in Eq. (10), \( \sim_{tcs} \) is the local similarity of attribute type of cutting surface, \( \sim_{twp} \) is the local similarity of attribute type of work piece, \( \sim_{surf} \) is the similarity of cutting surface. \( \sim_{surf} \) is shown in Table 3. The status of practical material of work piece is classified into six kinds, which are: (1) casting, (2) forging, (3) rolling, (4) quenching, (5) tempering, and (6) annealing. The local similarity of attribute material status of work piece is calculated as follows:

\[
sim_{mswp} = \sim_{mcwp} \times \sim_{ms}
\]

where in Eq. (11), \( \sim_{mswp} \) is the local similarity of attribute material status of work piece, \( \sim_{mcwp} \) is the local similarity of attribute material class of work piece, and \( \sim_{ms} \) is the similarity of material status. \( \sim_{ms} \) is shown in Table 4. The local similarity of attribute material hardness of work piece is calculated as follows:

\[
sim_{mhwp} = \sim_{mcwp} \times \sim_{mh}
\]

where in Eq. (12), \( \sim_{mhwp} \) is the local similarity of attribute material hardness of work piece, \( \sim_{mcwp} \) is the local similarity of attribute material class of work piece, and \( \sim_{mh} \) is the similarity of material hardness, which can be calculated by Eq. (5). It should be pointed out that the classification of local similarity is not invariable, but flexible. For example, the local similarity \( \sim_{mcwp} \) of attribute material status of work piece is referred to as the dependent type. However, if the material classes of cases to be considered are the same, it can be treated as the enumerated type in

### Table 1
Material code of work piece for carbon steel

<table>
<thead>
<tr>
<th>Material code</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-01</td>
<td>C: 0.1–0.25%</td>
</tr>
<tr>
<td>01-02</td>
<td>C: 0.25–0.55%</td>
</tr>
<tr>
<td>01-03</td>
<td>C: 0.55–0.8%</td>
</tr>
<tr>
<td>01-04</td>
<td>C: &gt;0.8%</td>
</tr>
</tbody>
</table>

### Table 2
Local similarities of attribute ‘Material code of work piece’ for carbon steel

<table>
<thead>
<tr>
<th>Sim</th>
<th>01-01</th>
<th>01-02</th>
<th>01-03</th>
<th>01-04</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-01</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>01-02</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>01-03</td>
<td>1</td>
<td></td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>01-04</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3
Similarities of cutting surface

<table>
<thead>
<tr>
<th>sim_{surf}</th>
<th>Round surface</th>
<th>Hole</th>
<th>Plane</th>
<th>Square shoulder</th>
<th>Slot</th>
<th>Curving surface</th>
<th>Screw surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hole</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plane</td>
<td>1</td>
<td></td>
<td>0.6</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Square shoulder</td>
<td>1</td>
<td>0.2</td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Slot</td>
<td>1</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Curving</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4
Similarities of material status

<table>
<thead>
<tr>
<th>sim_{ms}</th>
<th>Casting</th>
<th>Forging</th>
<th>Rolling</th>
<th>Quenching</th>
<th>Tempering</th>
<th>Annealing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casting</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Forging</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rolling</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Quenching</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tempering</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Annealing</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

practical application, and so is the attribute Material hardness of work piece.

4.4. Composite measure of similarity

When considering the case similarity, it is usual to emphasize the effects of local similarities of certain attributes on global similarity of cases according to certain viewpoints. For example, when selecting tool material, it is the material class of work piece to be emphasized rather than the type of work piece. Therefore, different weights are assigned to the attributes in order to make the global similarity comply with reality. The following equation is the composite measure of global similarity accounting for local similarity weight:

\[
SIM(\{q, u\}) = \frac{\sum_{i=1}^{m} w_i \times sim(q_i, u_i)}{\sum_{i=1}^{m} w_i}
\]

where \( q \) is the target case (i.e. the new problem); \( u \) is the source case stored in the case base; \( q_i \) and \( u_i \) are the \( i \)th attributes of \( q \) and \( u \) respectively; \( m \) is the number of the attributes in problem description part; \( w_i \) is the weight of the \( i \)th attribute.

4.5. Example

The objective of calculating the similarity between the two cases is to retrieve the most similar case to the new problem to be solved, and furthermore, to re-use its solution as the proposed solution to the new problem.

Supposing a work piece is to be machined, the type of work piece is ‘trunk’; the material is cast Al-alloy with hardness of 90 HB; the material specification code is unknown; the machining surface is plane; the machining requirement is semi-final. Therefore, the local similarity of attribute Material code of wp is 0; the weight of attribute material class of wp is assigned 2, and the weights of the other attributes are equal to 1. If there are three cases stored in the case base (shown in Fig. 3), the similarities are: \( SIM(q, 00106) = 0.333 \), \( SIM(q, 01007) = 0.783 \) and \( SIM(q, 00135) = 0.645 \). Thus case 01007 is the nearest neighbor to the new problem, i.e. the most similar case. So the solution of case 01007 is referred to as the proposed solution of machining the new work piece. Of course, the proposed solution should be revised with respect to cutting requirement and practical situations, which is not investigated in this paper.

5. Conclusions

Case representation and case similarity play important roles in CBR system. In high-speed machining field, a case means a work piece and its machining by HSM technology, which is represented by a relational data model in this paper. With analysis of the characteristics of attribute and domain knowledge of HSM, the case local similarities are classified into four types: numerical, enumerative, independent and dependent type. The global similarity of case is calculated by the composite similarity measure. From examples of case retrieval and similarity calculations, it can be seen that the case representation and similarity measure proposed in the paper are feasible.

Acknowledgements

The authors gratefully acknowledge the support of the National Natural Science Foundation of China under grant 50105012.

References


