European Symposium on Computer Aided Process Engineering – 15
L. Puigjaner and A. Espuña (Editors)

Case-Based Reasoning Tool for the Support of Process and Product Design

Yuri Avramenko[a], Andrzej Kraslawski[*a], Wieslawy Krysiak[b]

[a]Department of Chemical Technology, Lappeenranta University of Technology
P.O. Box 20, FIN-53851 Lappeenranta, Finland

[b]Faculty of Biotechnology and Food Science, Technical University of Lodz
Stefanowskiego 4/10, 90-924 Lodz, Poland

Abstract

The reuse of tested and optimized design is an important aspect for decreasing design times and increasing design quality. Case-based reasoning is applied to support design process. The case description is composed of information entities which enable to represent diverse design data. To remove the gap between the past design and the current problem the parametric adaptation is considered. The computer tool CM-DS is developed to assist the designer in his activity. The applications of the tool for problems of process and product design in food industry are described.

Keywords: case-based reasoning, design support, food preparation

1. Introduction

The solution of a new design task requires reviewing of all available methods, processes, systems and equipment in the light of demands and conditions, application of scientific knowledge and engineering judgment based on experience. The reuse of tested and optimized design is an important factor for decreasing design time, increasing design quality, and improving the predictability of designs. The practice shows that often it is more efficient to solve a new design problem by starting with a solution of a previous, similar problem than to generate the entire solution from scratch. A design engineer may encounter a problem to determine the similar elements between a new problem and massive historical data; moreover, the similarities are often not noticeable. In order to facilitate the design process and to reduce the required development time the computer tool utilising the case-based reasoning approach has been created to support process and product design.

The design tasks are generally classified into routine, innovative, and creative (Coyne et al. 1987). In routine design all variables and their application ranges, as well as the knowledge to compute their values, are directly derivable form existing designs. Routing design problems are typically represented by a well defined set of components and a set of constraints that the final design must satisfy (Gero 1990). In contrast to routine design, the innovative design tasks are usually described by incomplete

[*] Prof. Andrzej Kraslawski: andrzej.kraslawski@lut.fi

knowledge, and the applicable range for values of variable may change. The result of innovative design is a novel design with a familiar structure but unfamiliar set of values of the defined variables and their combinations. This work focuses on innovative and routine design and describes applications for design problems in food industry.

2. Case-Based Design Support

Case-based reasoning is applied to support human problem solving in design. The process of problem solving is characterized by collecting new information while making decisions and acting in real work until a satisfactory level is reached (Burkhard 1998). CBR can guide this process by comparison with cases from the past.

2.1 Case-based reasoning

The main idea of CBR is based on the assumption that the similar problems have the similar solutions. The central notion of CBR is a case. A case is represented as a pair: problem and its solution. Many cases are collected in a set to build a case base.
In solving a current design problem, CBR includes the following operations:

- introduce a new problem by setting its features;
- retrieval of a similar, past problem and its solution using a set of rules for measuring similarity between current problem and those stored in case base;
- adaptation of the retrieved solution using a set of rules, based on special algorithm, for adjusting of any differences between the current case and the retrieved one;
- validation of the created new solution;
- learning the system by storing the approved solution to the current case, and it can then be used in solving future problems.

CBR has been applied recently for support design process in process synthesis (Pajula et al. 2001), equipment selection (Kraslawski et al. 1999), quality design (Suh et al. 1998).

2.2 Representation of design data

The representation of design data requires various models because design content involves topological, geometric, and physical properties and relations between them. The problem description in innovative design is often incomplete and uncertain.
The proposed way to build the case base which can represent diverse design data is the consideration of the information entities. A case is set of information entities. The number of information entities in a case may be variable.
An entity description includes the list of features F, the set of relations between them R, and the set of feature values V. The representation may be extended by including numerical attributes of features of an entity, W. The attributes usually reflect a degree of importance of corresponding feature in the description.
Thus, an entity E is defined as follows:

$$E = < F, V, R, W > \qquad (1)$$

In this work the past design data is represented by weighted feature-value pairs (features without relations) since there is no need in the consideration of relations.
Features may be of various types. They can be expressed by numeric, vector, logical or symbolic values, as well as sets and graphs. Main advantage is that a feature may be represented by new information entity. Thus, a case may contain hierarchy of entities.

2.3 Similarity Measurements

The retrieval from case base is based on the vague matching of information entities of newly introduced problem and problems from the past cases. The inexact matching is realized by calculation of a degree of similarity. Cases may contain entities which have no counterpart in new problem. It is also possible that some entities of new problem are not present in case base. Moreover, entities may include various numbers of features and the features often have different values. There can be distinguished three levels for similarity measurement: between cases, entities and features. But both cases and features can be represented as entities. A case is an entity containing one feature represented as a set. Hence, the similarity measure is uniform for each level.

The similarity between two entities, A and B, is expressed by a weighted sum of similarity measures of features values (a_i, b_i for A and B respectively):

$$SIM(A,B) = \sum_{i=1}^{k} w_i \cdot sim(a_i, b_i) / \sum_{i=1}^{k} w_i \tag{2}$$

The similarity measure for values, $sim(a_i,b_i)$, is depended on their data type. There has been developed the special similarity concept that copes with the data represented in the different data types. The similarity sim for two values a and b is defined as:

$$sim(a,b) = 1 - d(a,b) \tag{3}$$

where $d(a,b)$ is a distance function. The distance function takes the values within the interval [0, 1]. The distance measurements for different data types are shown in table 1.

Table 1. Distance functions for different types of data

Type	Measurement		Type	Measurement
Numeric	$d = \frac{\lvert a-b \rvert}{range}$	$range$ – a range of values of variable a and b.	Sets	$d = 1 - \frac{\lvert a \cap b \rvert}{\lvert a \cup b \rvert}$
Vectors	$d = \frac{\lvert \vec{a} - \vec{b} \rvert}{\lvert \sum_{i=1}^{n} \vec{e}_i \rvert}$, vectors a and b are normalized	$\vec{e}_1 = (1;0;\ldots;0)$ $\vdots$ $\vec{e}_n = (0;0;\ldots;1)$	Logical	$d = \begin{cases} 1, a \neq b \\ 0, a = b \end{cases}$

2.4 Adaptation

An adaptation procedure is based on the assumption that not only the most similar case can be used but a set of cases located nearby the current problem in the problem space. The key assumption that the similar problem has the similar solution means that solutions of similar problems are located nearby each other. The distances between new solution and solutions of the most similar problems must correspond to the distances between current problem and the similar problems under consideration. Since the solutions are also described by set of information entities, the similarity measures presented in previous section are valid for solutions as well.

The adaptation task is introduced as the minimization of the function:

$$F(N) = \sum_{j \in L} \left| SIM(S_j, N) - SIM(P_j, C) \right| \quad (4)$$

where P_j – past problem, S_j – solution of j-problem, C – the current problem, N – the new solution being generated; L – the list of most similar cases.

Usually, there are used three the most similar cases. The initial values of the features of new solution are copied from the most similar case data. The design parameters of new solution are changed to reach the minimum of function F. Because the design parameter may be of various type of data representation, standard genetic algorithm, which is not critical for data types, is used as a global optimization method.

Finally, the generated parameters have to be validated and approved.

3. Description of the Tool

The case-based design tool CM-DS is developed to support the innovative and routine design in process engineering. The objective of the tool is to assist the user in design process rather than to generate design proposal automatically.

The CM-DS tool has been designed to support different domains, and therefore, it provides partially automatic knowledge acquisition. The information describing the design cases can be stored with relational databases, binary files of certain structure, and XML-representation. The developed tool supports several standard database formats.

The CM-DS contains four units corresponding to specific parts of tool's activity (Fig.1). The case base is created from information sources during running of the system according to given case representation that is produced depending of task goal - I. A case is constructed from information entities found in the data source. The structure of an entity is described by a set of features; each feature has an identifier (name). The case-base builder searches the massive of data to find the existing features and to create an entity according to a given structure. If a feature of entity structure is not found then it is skipped. The current design problem as well as the structure of case representation is introduced by means of descriptive language before running of the tool. Using similarity measures implemented in retrieval routine a set of similar past cases to the current design situation is retrieved - II. According to adaptation algorithm, the prior solutions are used to propose a solution to the current situation - III. The proposed solution, then, is validated by user - IV. If it succeeds, the working solution together with current problem is stored in pre-defined format as a new information source.

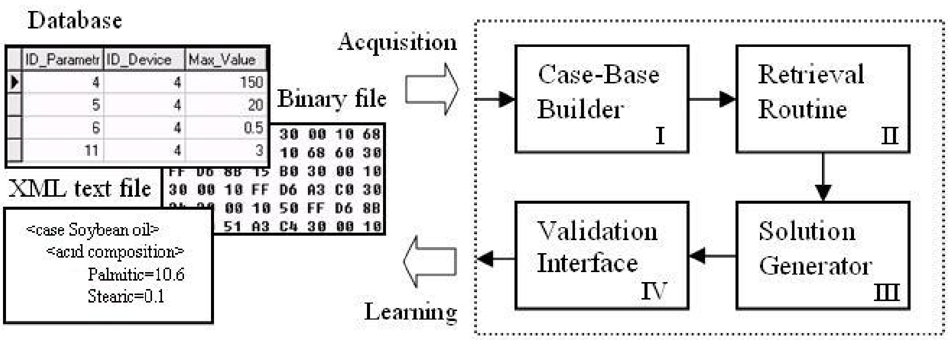

Figure 1. Activities and units of the CM-DS tool

4. Applications of the Tool

4.1. Selection of operation parameters for drying of cacao beans

One of the important processes for preparation of final cacao product is drying. Drying process is used to remove both water and other undesired compounds from cacao beans. Selection of proper conditions of drying process for cacao beans has been considered.
The results of experiments of cacao drying including the characteristics of cacao beans before and after drying (see Table 2) as well as operational parameters of process (*temperature*, *air velocity* and *drying time*) were used to build the case base. The case representation is simple feature-value pairs. For a test sample of cacao and desired characteristics after drying (Table 2,a) the CM-DS tool proposed the following parameters of drying: air temperature – 138 °C, air velocity – 0.5 m/s, time – 45 min.
Validation of the proposal was done experimentally. The test sample was dried with proposed conditions and the output cacao characteristics were obtained (Table 2,b). The experimental results are close to desired characteristics; hence, the proposed parameters can be accepted.

Table 2. Cacao characteristics of problem statement and experiment result

	a) Problem		b) Experiment result
Feature	before	after	after
Water content	4.8	2.0	2.03
Alkalinity	14.59	11.5	11.3
Alkalinity, %	0.119	0.100	0.095
Polyphenol content	6.25	4.8	5.25
Carbonyl content	2.16	1.5	1.76
Fat content	59.04	55	57
Colority	0.812	1.2	1.06

4.2. Fats and oils product development

Fats and oils are key functional ingredients in a large variety of prepared foods. Successful development of the food products relies on effective use of the different functional properties of the available fats and oils and the manipulation of the fat blend to satisfy the prepared foods requirements.
Application development of fats and oils products begins with identification of the key functional attributes that the final product is expected to provide. The important functional attributes, which is considered for product development, are: lubricity, structure, clarity, consistency, plasticity, emulsification, creaming property, spreadability, aeration, hardness, freeze stability, flavor (odor, taste and mouth feel) and flavor stability. The product functionality can usually be translated into analytical measurements and physical properties. For example, mouth feel and flavor release can be controlled by the melting properties and oxidative stability. The solids fat index (SFI) curve characterizes the consistency and spreadability of the product.
Next step is the use of historical knowledge to identify and evaluate the physical and chemical properties most likely to produce the intended functionality. The database if physical properties and compositions of nature and genetically modified oils and fat blends has been created based on materials of book (O'Brien 2004) and other sources.

Table 3. A part of representation of fats and oils formulation case

Entity	Type/Structure		Example (Palm Oil)
Cold test	numeric (hours)		None
Crystal habit	logical (β/β′)		β′
Fatty acids	set of features; features are of numeric type (%)		Myristic (1.1); Palmitic (44.0); Oleic (39.2); Stearic (4.5) … etc
Melting point	numeric (°C)		37.5
Oxidative stability	set of entities		
	AOM test	numeric (hours)	54
	Inherent	numeric	1.705
	Iodine value	numeric	53
Refractive index	record = [numeric, numeric (°C)]		1.455, 50°C
Saponification	vector (saponifiation, unsaponifable numbers)		(196, 0.5)
Solidification point	numeric (°C)		-
Solids Fat Index (SFI)	set of features; features are of vector type (°C,%)		(10.0, 34.5); (21.1, 14.0); (26.7, 11.0); … etc
Tocopherols	vector (α-,β-,γ-,δ-), (ppm)		(172, 30, 26, 13)
Tocotrienols	vector (α-,β-,γ-,δ-), (ppm)		(59, 59, 350, 94)
Triglycerides	vector (Trisaturated, Disaturated, Monosaturated, Triunsaturated), (%)		(10.2, 48.0, 34.6, 6.8)

A part of case representation is shown in Table 3. The solution part usually includes the four entities: fatty acids, tocopherols, tocotrienols, and triglycerides. The rest of entities representing the physical properties and test measurements build the problem part.
The functionality of the product in development converted to physical properties is used to identify the oils with most similar properties. The CM-DS tool then modifies the retrieved composition in adaptation phase. Analysing the modified composition during validation the necessary processing of fats and oils raw materials can be selected. The desired composition and physical properties can be achieved by the blending, hydrogenation, fractionation, interesterification and emulsification.

References

Burkhard, H. D., 1998, Extending some Concepts of CBR – Foundation of Case Retrieval Nets, Lecture Notes in Artificial Intelligence, vol. 1400, Springer-Verlag, Berlin.

Coyne, R. D., Rosenman, M.A., Radford, A.D., and Gero, J.S., 1987, Innovative and creativity in knowledge-based CAD, Expert Systems in CAD, North-Holland, Amsterdam.

Gero, J. S., 1990, Design Prototypes: A knowledge representation schema for design, The AI Magazine, 11:26-36.

Kraslawski, A., Lyssov, I., Kudra, T., Borowiak, M., Nystrom L., 1999, Case-based reasoning for equipment selection using rough sets analysis in adaptation phase, Comp. Chem. Eng., 23 (Suppl.), Elsevier Science, pp. 707-710.

O'Brien, R. D., 2004, Fats and oils: formulation and processing for applications, CRC press LLC, New York, United States of America.

Pajula, E., Seuranen T., Hurme, M., 2001, Synthesis of separation processes by using case-based reasoning, Comp. Chem. Eng., 25, Elsevier Science, pp. 775-782.

Suh, M. S., Jhee, W. C., Ko, Y. K., Lee, A., 1998, A Case-based Expert System Approach for Quality Design, Expert Systems with Applications 15, Pergamon, pp. 181-190.