# A Method of Collaborative Process Design Comparing Workflow Model Components

*Joonsoo Bae[†], Backchul Kwon[††], Jae-Yoon Jung[†††], Hyerim Bae[††††], and Suk-Ho Kang[†††]*

[†]Dept. of Industrial & Information Systems, Chonbuk National University, Korea
[††]R&D Planning Team, R&D Center, Hyundai Motor Company, Korea
[†††]Dept. of Industrial Engineering, Seoul National University, Korea
[††††]Dept. of Industrial Engineering, Pusan National University, Korea

**Summary**
This paper proposes a difference analysis methodology using distance measures between process definitions and suggests a collaborative process design method that can be used by many process designers in order to develop a new process together. The properties of process definition are classified into four categories and distance measures are developed for each category: relationship, structure, participant, and condition. The proposed collaborative design method is a type of group dynamics that supports the attainment of a final goal among many different or conflicting opinions by using devised distance measures. The proposed method can provide a more exact and cost efficient process design, especially when there are an increasing number of people who are concerned with this process, and when there are an increasing number of accumulated history logs that can be referenced in a new process design.

***Key words:***
*Workflow, collaborative process design, process distance, group dynamics.*

## 1. Introduction

A business process model has several characteristics. First, the process model is composed of its activities, each of which has precedence relationships with each others. Second, a process forms a hierarchical structure where an activity in a process may also be another process. Third, a process can have activities that are performed in parallel. Fourth, a process has objects (participants, variables, agent, etc) that take part in business activities. Finally, each activity in a process has incoming and outgoing conditions [2] [3] [7] [10]. This paper proposes the development of a new design method, by which multiple users can design a process collaboratively. We can achieve three goals using this method. First, it provides a design method of process models that combines social psychology and group dynamics. Thus, the build-time and run-time of a process model can be associated with one another. This may achieve the added advantage that we can analyze specialized enterprise processes at a low cost and from various viewpoints. Second, by analyzing attributes of process models, we can present a process quantitatively and determine the difference between models. This facilitates, not only a comparison of existing process models with each other, but also the flexibility to adapt to changes in processes. Finally, it provides priority to collaborating steps. Thus, the proposed design method becomes fast and easy. Therefore, the expense of both the analysis and design phases can be reduced. Recently, although interest in Business Process Management Systems (BPMS) has grown, the design methodology of business processes has not been sufficiently considered. Our approach to collaborative process design provides the advantage of fostering cooperation between multiple users' in defining processes. We can achieve a more accurate design of process by decreasing the risk of error, reducing design cost, and considering a variety of opinions.

## 2. Related Work

In this study, to implement designing processes, Group Dynamics (GD) was used, and to execute the processes, Workflow Management Systems (WfMS) were used. The design method was used to maximize the advantages and to minimize the shortcomings of the two methods by analyzing the relative strengths and weaknesses of them. GD is a method of resolving a problem by acquiring advice from a number of experts, and the Delphi method provides an accurate representation of the method [9]. While GD means a process of group decision-making [6], WfMS carries out control and administration when actual business processes are executed. GD can contribute to correct design. It is possible to collect expert opinions and to reduce time and costs, which cannot be achieved when a single user analyzes and designs [4] [8] [11]. GD also prevents incorrect modeling of processes. Although GD can be one of the methods used for decision-making, it does not propose methods of executing, controlling, and managing processes based on the decisions made. Therefore, if GD at build-time and WfMS at run-time are combined, a synergy effect is expected for the entire business process life cycle.

Manuscript received August 5, 2006.
Manuscript revised August 25, 2006.

A business process usually contains a lot of information. Properly using the information enables us to examine closely differences among process models. These differences act as barometers that enable efficient cooperation in a design. In this sense, a collaborative design process means a process of interaction based on comparison. The design process is composed of four steps: information input, process comparison, process clustering, and mutual agreement.

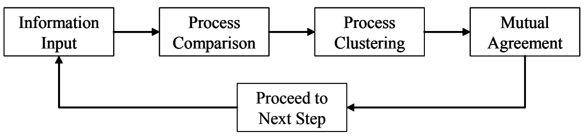



Fig. 1. A collaborative design process

Usually, a process model contains information about work to be performed, participants in charge, order of execution, and resulting operations of each task. In this paper, we describe the topological and semantic information of a process model using a process structure and related attributes. The attributes must be developed to represent a feature of a process, which is used to measure the degree of difference among processes.

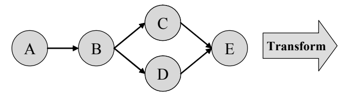
Table 1: Attributes of processes

| Class I | Class II | Description |
|---|---|---|
| Topological Information | Relationship | Precedence relations among activities, and start/end of a task. |
| | Structure | Structure formed by execution of activities, which involves split and split types such as 'AND', 'OR', and 'XOR'. |
| Semantic Information | Participant | Users participating in activities of a process. A participant is defined as a person, a organization, or an application. |
| | Condition | Split conditions and conditions related to beginning and terminating of an activity. |

## 3. Measuring Workflow Distance

### 3.1 Dependency Distance Measure

A process is composed of a number of unit tasks. If the process model has $m$ tasks, there can exist $m(m-1)$ dependencies among the tasks. To describe the task dependencies, a process network matrix G is presented. The elements of the matrix have binary values. If there is a dependency from task $a$ to task $b$, the value of its element (a, b) is 1, and otherwise 0.



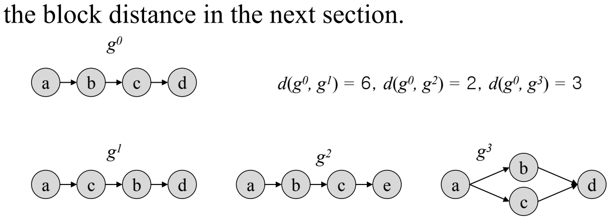

| FROM \ TO | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 1 | 0 | 0 | 0 |
| B | 0 | 0 | 1 | 1 | 0 |
| C | 0 | 0 | 0 | 0 | 1 |
| D | 0 | 0 | 0 | 0 | 1 |
| E | 0 | 0 | 0 | 0 | 0 |

Fig. 2. An example of process network matrix

Banks and Carley [2] present a symmetric distance metric that counts inconsistent elements and measures the quantitative difference between two network matrices [12]. The distance $d(g^1, g^2)$ between two networks $g^1$ and $g^2$ is calculated with network matrices $G^1$ and $G^2$, as follows. $Tr$[A] is the trace of a square matrix A.

$$d(g^1, g^2) = Tr\,[(G^1 - G^2)^T (G^1 - G^2)] \quad (1)$$

Since the distance metric $d(g^1, g^2)$ counts the number of asymmetric arcs, it can reflect the difference of some characteristics between two processes, such as task precedence, task commonality, flow structure, etc. Task precedence means how the activities are linked and sequenced. This denotes the disparity of sequence between two tasks and can be extended to represent the sequence disparities between all tasks. In Fig. 3, the distance $d(g^0, g^1)$ of two processes $g^0$ and $g^1$ shows the difference of task precedence. Task commonality means how many activities are shared between two process models. This counts the different activities or new activities of two processes, as illustrated by processes $g^0$ and $g^2$ in the figure. In additions, flow structure means difference between serial and parallel flows. Two processes $g^0$ and $g^2$ show the difference measurement of flow structures, serial and parallel flows. More detailed structures of parallel flows are considered as the block distance in the next section.




(a) example processes

$G^0$

| | a | b | c | d |
|---|---|---|---|---|
| a | 0 | 1 | 0 | 0 |
| b | 0 | 0 | 1 | 0 |
| c | 0 | 0 | 0 | 1 |
| d | 0 | 0 | 0 | 0 |

$G^1$

| | a | b | c | d |
|---|---|---|---|---|
| a | 0 | 0 | 1 | 0 |
| b | 0 | 0 | 0 | 1 |
| c | 0 | 1 | 0 | 0 |
| d | 0 | 0 | 0 | 0 |

$G^2$

| | a | b | c | d | e |
|---|---|---|---|---|---|
| a | 0 | 1 | 0 | 0 | 0 |
| b | 0 | 0 | 1 | 0 | 0 |
| c | 0 | 0 | 0 | 0 | 1 |
| d | 0 | 0 | 0 | 0 | 0 |
| e | 0 | 0 | 0 | 0 | 0 |

$G^3$

| | a | b | c | d |
|---|---|---|---|---|
| a | 0 | 1 | 1 | 0 |
| b | 0 | 0 | 0 | 1 |
| c | 0 | 0 | 0 | 1 |
| d | 0 | 0 | 0 | 0 |

(b) process network matrices

Fig. 3. Examples of dependency distance measurements

In the tables of Fig. 3, the left matrix $G^0$ is the network matrix of a base process and the other matrices are those of comparative processes. If the elements of value 1 are located in the same position, the two processes have the same arc (dependency). Conversely, if the location of value 1 is not consistent in two comparing matrices, the number of those values is the distance of dependency distance. So in the example, the distance of task precedence between $g^0$ and $g^1$ is 6, the distance of task precedence between $g^0$ and $g^1$ is 2, and the distance of task precedence between $g^0$ and $g^3$ is 3.

## 3.2 Block Distance Measure

Block distance deals with the process structures that are not easy to reflect by the dependency distance measure. The process structure can be composed of serial flows and split flows, and the split flows are classified into AND and OR splits according to the simultaneous execution or not. Furthermore, the OR splits can be also divided into two types: XOR split (Exclusive OR) and SOR split (Synchronized OR) [5]. These four types of blocks (serial, AND, XOR, and SOR) are the basic structure of a process model and are used as the basis for distance measurements. This structure information can be presented as a block concept, and the block detection algorithm can be applied to find a block from a process network [1].

To describe the distance of two blocks, a block probability matrix is presented. The block probability matrix $P^m$ is an $m \times m$ square matrix, if the union of the branches in two blocks has $m$ elements. If the block is AND split, all the elements of the matrix are 1. And if it is XOR split, its diagonals are 1 and the others are 0. Finally, in the case of SOR split, its diagonals are 1, the columns of the always executed branches are 1, and the other elements are 0.

It is assumed that there is a block structure that is composed of three tasks (E, F, $B_1$), where E and F are unit tasks and $B_1$ is another sub-block. Fig. 4 shows the matrices of three types of blocks. If the block (E, F, $B_1$) is AND split, only one case exists, i.e. three branches are always executed as shown on the right matrix of Fig. 4. Next, if the block is XOR split, three cases are possible like the rows of the middle matrix, i.e. only one of the three branches will be executed. Finally, if the block is SOR split that a unit task E should be always executed, three cases are also possible on the right matrix.

AND block

| E | F | $B_1$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

XOR block

| E | F | $B_1$ |
|---|---|---|
| 1 | 0 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 1 |

SOR block

| E | F | $B_1$ |
|---|---|---|
| 1 | 0 | 0 |
| 1 | 1 | 0 |
| 1 | 0 | 1 |

Fig. 4. Block probability matrices of an AND, XOR, and SOR blocks

As illustrated on Fig. 4, the block probability matrix can express all cases of an AND, XOR or SOR blocks. As a result, the block probability matrices $P^1$ and $P^2$ are used to measure the block distance $d(g^1, g^2)$ as follows.

$$d(g^1, g^2) = Tr\,[(P^1 - P^2)^T (P^1 - P^2)] \qquad (2)$$

According to this metric, the block distance between AND and XOR blocks in Fig. 4 is 6, the distance between AND and SOR blocks is 4, and the distance between XOR and SOR blocks is 2.

## 3.3 Participant Distance Measure

In the workflow process definitions, a participant is one who is assigned to each unit task and actually performs the task. In process models, the participant can be assigned to a type of individual participant or organization, and applications or agents, as well as human worker, are also treated as participants. In this research, the participant information is represented as a hierarchical organizational model and the model is used to measure the distance between two participants of a unit task in a process.

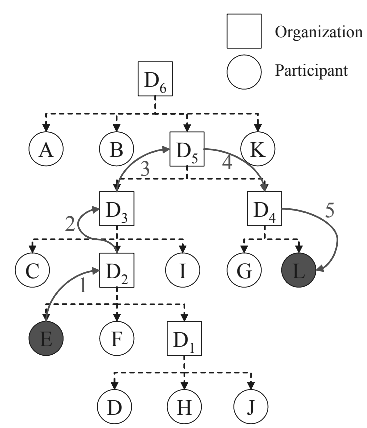



Fig. 5. Participant distance measurements

Fig. 5 shows the distance between two participants by counting the arcs in the shortest path between two nodes. After the lowest common ancestor of two nodes is found, the distance of the two is the sum of the distance from the first node to the ancestor and the distance from the ancestor to the last node.

## 3.4 Condition Distance Measure

To measure condition distance for a unit task of a process, a condition comparison matrix is described. The matrix is a kind of qualitative summary on whether or not the same conditions exist. If the number of conditions fails to match the two tasks, dummy conditions are inserted in order to make a regular square matrix. Thus, if the union of the conditions in two tasks has $m$ elements, the condition comparison matrix $C^m$ is an $m \times m$ square matrix. If conditions A and B for a unit task are the same, the value of its element (a, b) is 1, and otherwise 0.

Multiple conditions for one unit task may be grouped by AND or OR operation. While all the conditions should be satisfied in AND operation, one or more conditions should be satisfied in OR operation. Consequently, the multiple condition distance can be measured with the following formula.

$$(\textit{Condition distance}) = \begin{cases} (\#\ \textit{of different condition pairs}), & \text{if condition structures are the same} \\ (\#\ \textit{of different condition pairs}) + \dfrac{1 - (\#\ \textit{of same condition})}{(\#\ \textit{of condition in AND condition structure})}, & \text{if condition structures are different} \end{cases} \quad (3)$$

## 4. Workflow Collaborative Design

This section explains the methodology of workflow collaborative design. First of all, it is assumed that five different workflow models are collected from five process design experts. The models are used to the four types of workflow distance measurements, outlined in Section 3, in order to achieve one goal process. First, the dependency distances among five input processes are summarized to a dependency distance matrix as shown in Fig. 6, where A, B, C, D, and E are input process models.

D =

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 9 | 3 | 6 | 11 |
| B | 9 | 0 | 7 | 5 | 10 |
| C | 3 | 7 | 0 | 9 | 2 |
| D | 6 | 5 | 9 | 0 | 8 |
| E | 11 | 10 | 2 | 8 | 0 |

Fig. 6. Dependency distance matrix of five input processes

Next, the minimal distance method is applied to the dependency distance matrix. In the matrix, pairs of two input processes with the minimal value are found and merged step by step. Because the minimal value of the example matrix is 2 between processes *E* and *C*, the two processes are initially grouped to one cluster (*CE*). Then, processes *C* and *A* with value 3 are grouped to a cluster (*ACE*) because the process *C* is already in cluster (*CE*). In the same way, the next cluster is (*BD*) with value 5. Finally, all processes are clustered into one by grouping (*ACE*) and (*BD*).

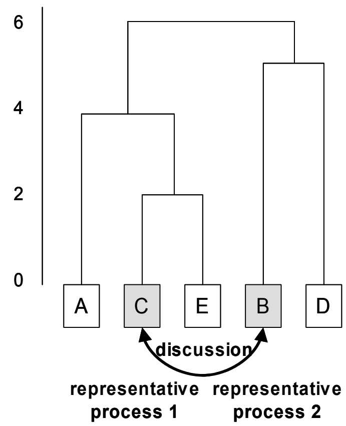



Fig. 7. Diagram of process clustering

The diagram of Fig. 7 shows a means of grouping input models into the target number of clusters. If the number of clusters is assumed to be two, the result is clusters (*ACE*) and (*BD*). In each cluster, its representative model is the process with the median value because it is not biased against the other processes in the cluster. If the cluster has two process models, the representative of the cluster is the model that is closer to the nearest cluster of another cluster. The reason for this placement is that the closer model is easy to modify and adapt to the ideal process model. For example, the representative process model of cluster (*ACE*) is *C*, and that of cluster (*BD*) is *B*. This means that the process models *C* and *B* are selected to discuss process design. Fig. 8 illustrates the process dependency model generated from the discussion.

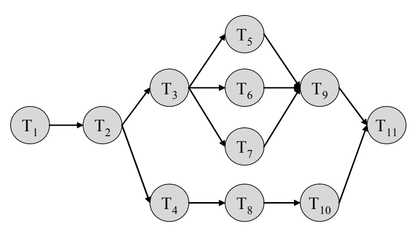



Fig. 8. The generated process dependency model

The next stage compares the block structure of input process models. The generated process dependency model of Fig. 8 has two parallel blocks according to the block detection algorithm [1]. The blocks are *B2* and *B4*, and both are even-numbered blocks in the block tree. The opinions on the block structure are gathered from the design experts.

The result of block structure comparisons is summarized in Fig. 9. Input structures are the opinions of five experts on each block. The value in parentheses is the number of nodes (i.e. tasks or sub-blocks) in the block. The next step involves generating the block probability matrices and applying the minimal distance method to determine the structure of each block. Fig. 9 shows the distances between five models and three comparative models for two blocks. The result of the block structure comparisons is that block $B_2$ is recommended to SOR structure and block $B_4$ is done to AND structure.

| Models | Input structure | | Comparative model $B_2(3)$ | | | Comparative model $B_4(2)$ | | |
|---|---|---|---|---|---|---|---|---|
| | $B_2(3)$ | $B_4(2)$ | $B_2(3)$ AND | $B_2(3)$ XOR | $B_2(3)$ SOR | $B_4(2)$ AND | $B_4(2)$ XOR | $B_4(2)$ SOR |
| A | AND | AND | 0 | 6 | 2 | 0 | 2 | 1 |
| B | AND | AND | 0 | 6 | 2 | 0 | 2 | 1 |
| C | SOR | AND | 2 | 2 | 0 | 0 | 2 | 1 |
| D | SOR | AND | 2 | 2 | 0 | 0 | 2 | 1 |
| E | XOR | XOR | 6 | 0 | 2 | 2 | 0 | 1 |

Fig. 9. Results of block structure comparisons.

In the stage of participant design, participant information is also received from the process design experts. An expert can provide a specific participant or an organization as input participants for each task. If five models with participants of the tasks $T_1$ to $T_{11}$ are given, there are 11 matrices for each task that are 5×5 square matrices. The elements of each matrix are the distances between all pairs of participants, which are measured in the hierarchical organization model. The next step is to apply the minimal distance method and determine the participants of the process model. The procedure is the same way to the dependency process model generation.

In the final stage, condition information for each task is also received from the design experts. An example of the condition matrix is shown in Fig. 10.

| Condition (equal or not) | | Condition of process A | | | |
|---|---|---|---|---|---|
| | | A1 | A2 | A3 | D1 |
| Condition of process B | B1 | 1 | 0 | 1 | 1 |
| | B2 | 1 | 1 | 0 | 1 |
| | B3 | 1 | 1 | 1 | 1 |
| | B4 | 1 | 1 | 1 | 1 |

Fig. 10. An example of condition matrix

If two process models have the same condition, the value between the two is 0. Otherwise, the value is 1. Fig. 9 indicates that the conditions of A2 and B1, and those of A3 and B2 are equal. Thus the remaining conditions, such as A1 and D1 of process model A, and B3 and B4 of process model B are not equal, and therefore the condition distance value between models A and B is 2. In the next step, the procedure of condition information decision is also the same way to the dependency process model generation.

## 5. Conclusions

In this paper, a methodology of workflow collaborative design is proposed. This research has three important contributions. The first is to provide a process design methodology of combining workflow management and group dynamics in socio-psychology. This methodology reveals that collaboration in practical process design environment can be adapted to generate a workflow process model. The second is to measure and compare process models quantitatively by analyzing various components of process design. As a result, a process model can be compared to existing process models or to new models to keep up with the trend of rapid business process change. Finally, a collaborative design methodology will reduce the cost of process design and analysis by providing a systematical and rapid approach in process design deliberation among domain experts. In the future research, more detailed distance measures to handle various conditions and an integration methodology to combine the measures proposed in this paper are needed.

## References

[1] Bae, J., Bae, H., Kang, S., Kim, Y.: Automatic control of workflow process using ECA rules. IEEE Trans. on Knowledge and Data Engineering, vol.16, no.8 (2004) 1010-1023.

[2] Banks, D., Carley, K.: Metric inference for social networks. Journal of classification. vol.11 (1994) 121-149.

[3] Cichocki, A., Helal, A., Rusinkiewicz, M., Woelk, D.: Workflow and process automation: concepts and technonlogy. Kluwer academic pub. Massachusetts (1998).

[4] Daily, B.F., Steiner, R.L.: The influence of group decision support systems on contribution and commitment levels in multicultural and culturally homogeneous decision-making groups. Computers in human behavior, vol.14, no.1 (1998) 147-162.

[5] Dogac, A., Kalinichenko, L, Ozsu, M.T., Sheth, A.: Workflow management systems and interoperability, NATO advanced study institute, Istanbul (1997).

[6] Forsyth, D.R.: An introduction to group dynamics. Brooks/Cole publishing company (1983).

[7] Hollingsworth, D.: Workflow management coalition specification: The workflow reference model. WfMC (1995).

[8] LeClair, S.R.: Integrative learning: A multi-expert paradigm for acquiring new knowledge. Sigart newsletter, no.108 (1989) 34-44.

[9] Linstone, H., Turoff, M.: The Delphi method: techniques and applications. Addison-Wesley, Massachusetts (1975).

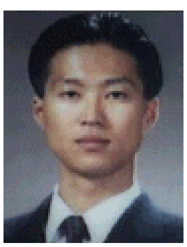
[10] Leymann, F., Roller, D.: Production workflow: concepts and techniques. Prentice Hall PRT, New Jersey (2000).

[11] McCartt, A.T., Rohrbaugh, J.: Evaluating group decision support system effectiveness: A performance study of

decision conferencing. Decision support systems, vol.5, no.2 (1989) 243-253.
[12] Rush, R., Wallace, W.A.: Elicitation of knowledge from multiple experts using network inference. vol.9, no.5 (1997) 688-698.

**Joonsoo Bae** is an Assistant Professor in the Department of Industrial and Information Systems Engineering at Chonbuk National University. He received the Ph.D., M.S., and B.S. degrees in Industrial Engineering from Seoul National University, Korea in 2000, 1995, and 1993, respectively. He also completed the Program of Advanced Software Engineering in School of Computer Science in Carnegie Mellon University at 2002. He has been with the SCM & CRM department of LG-EDS as a technical consultant from 2000 to 2002. He is interested in the areas of system design in manufacturing field, system integration in management information system, and e-Business technology. His research topics include control of business processes using workflow systems, e-Business security, process improvement of software engineering, and supply chain analysis.

**Backchul Kwon** received the M.S. degree from Seoul National University in 2004, and the B.S. degree from Hongik University in 2002, both in Industrial Engineering. He is currently in R&D center of Hyundai Motor company. His research interests are in group decision making, process design, and intelligent manufacturing.



**Jae-Yoon Jung** received the Ph.D., M.S., and B.S. degrees in Industrial Engineering from Seoul Nationa University in 1999, 2001, and 2005, respectively. From 2005, he is a senior researcher in Automation & Systems Research Institute (ASRI), Seoul National University. He is currently visiting BPM Group in Eindhoven University of Technology as a Post-doctoral fellow. His research interests include BPM, B2B integration, peer-to-peer systems, ubiquitous computing, etc.

**Hyerim Bae** is an assistant professor in the Industrial Engineering Department at Pusan National University (PNU), Korea. He received PhD, MS, and BS degrees from the Industrial Engineering Department at Seoul National University, Korea. He had been a manager for information strategic planning at Samsung Card Corporation before he joined PNU. He is interested in the areas of BPM, process-based B2B integration, and ubiquitous business computing. His current research activities include analysis of business process efficiency, controlling of logistics processes with context awareness, and convenient modeling of business processes.

**Suk-Ho Kang** is a professor in the Industrial Engineering Department at Seoul National University. He has a B.S. degree in physics from Seoul National University, and M.S. degree from the University of Washington, and a Ph.D. degree from Texas A&M University, both in industrial engineering. His research interests are in production and operations management, intelligent manufacturing systems and B2B e-commerce.