DASFAA 2003

Eighth International Conference on

Database Systems for Advanced Applications

Kyoto, Japan
26-28 March 2003

Organized by
Kyoto University
DASFAA Steering Committee

Sponsored by
Communications Research Laboratory
Hitachi, Ltd.
Kyoto University

Supported by
Information Processing Society of Japan (IPSJ), SIGDBS
Database Society of Japan (DBSJ)

Proc. of the 9th International Conference on Databese Systems for
Advanced Applications (DASFAA 2003), pp. 293--300 (Mar. 2003)

Composition of Filtering Functions

Rie Sawai, Masahiko Tsukamoto, Tsutomu Terada, and Shojiro Nishio
Graduate School of Information Science and Technology, Osaka University
{rie,tuka,tsutomu,nishio}@ist.osaka-u.ac.jp

Abstract

In recent years, due to the increasing popularization of various broadcast services, there has been an increasing demand for information filtering techniques. Although filtering methods that have been proposed filter data in their own ways, mathematical representation of these methods does not exist. Consequently, it is not possible to qualitatively evaluate various filtering methods, to optimize processing methods in filtering, or to design a declarative language for filtering processes. In our previous work we have defined a filtering function that represents filtering as a function, and clarified the properties of different filtering methods. In practice, current filtering methods actually consist of multiple methods. In this paper, we make it possible to qualitatively represent those filtering methods by introducing the concept of composition into the framework of filtering function. Moreover, we reveal the characteristics of those combined filtering methods by clarifying the properties of composite filtering functions.

1. Introduction

In recent years, due to the start of new satellite broadcast services and broadcast digitalization, a large number of broadcast services are being supplied[11]. In this environment, the amount of data and the variety of data broadcast are rapidly increasing. However, since users often need only a small amount of specific data, it is very difficult to retrieve the information they are interested in from a large range of broadcast data. Therefore, various mechanisms that automatically filter data and a user-request description language for filtering have been proposed[1, 2, 6, 7, 8, 9]. These filtering mechanisms filter data using different methods such as keyword matching or vector operation. However, no mathematical foundation for qualitatively representing these filtering processes exists. Thus, it is not possible to qualitatively evaluate various filtering methods, to optimize processing methods in filtering, or to design a declarative language for filtering processes. In [12], we defined a *filtering function* that expresses filtering as a function, and we made it possible to represent the properties of filtering by the constraints that are satisfied by the function. Moreover, by showing the inclusion relation between the constraints representing the properties of filtering, we clarified the relationship between the properties of various filtering methods.

In [12], we dealt with filtering methods represented by a single filtering function that satisfies specific properties. However, an actual filtering method generally consists of multiple methods. In this paper, we clarify properties of the composite filtering functions to reveal the characteristics of filtering methods consisting of several methods. By introducing the concept of composition into the framework of filtering functions, we are able to qualitatively represent complex filtering methods commonly used in practice. Furthermore, by showing the interrelation between properties of composite filtering functions, we make it possible to judge whether one filtering that satisfies a property also satisfies another one, thereby making it possible to replace the processing method with a more efficient one according to the environment.

This paper is organized as follows. Section 2 explains the general outline of filtering functions. Section 3 describes the conditions that composite filtering functions need to satisfy to be filtering functions. Section 4 reveals the properties of composite filtering functions for various combinations of filtering functions that satisfy different properties. Section 5 considers the filtering methods currently applied in practice and related work through the results clarified in this paper. Finally, we provide conclusions in Section 6.

2. Filtering Function

In this section, we outline filtering functions, which form the foundation of this study. First of all, we categorize the processing methods of filtering into several patterns, then we explain the definition of a filtering function that represents filtering as a function, and denote several properties of filtering defined as the constraints that a filtering function satisfies.

In our study, we categorize the processing methods in the real world into the following four patterns:

- Sequential processing
  In a system that uses sequential processing, the newly received data and the previous filtering results are merged and filtered.
- Batch processing
  In a system that uses batch processing, a receiver accumulates broadcast data and filters them in bulk.
- Distributed processing
  In a system that uses distributed processing, the received data set is divided into multiple arbitrary data subsets, and each subset is filtered separately before the results are merged.
- Parallel processing
  In a system that uses parallel processing, the merged filtering results of distributed processing are re-filtered.

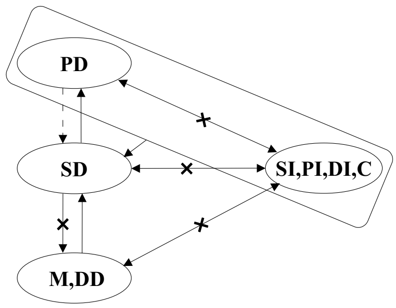
Let $\mathbf{T}$ be a set of data items. A filtering function is defined as a function f on $2^{\mathbf{T}}$ that satisfies the following two properties for an arbitrary $T \subset \mathbf{T}$ [1].

D: decreasing $\quad f(T) \subset T.$
ID: idempotent $\quad f(f(T)) = f(T).$

The following properties of filtering functions are defined.

SI: Sequential Increasing
$$f(S \cup T) \subset f(S \cup f(T)).$$
SD: Sequential Decreasing
$$f(S \cup T) \supset f(S \cup f(T)).$$
SE: Sequential Equivalence
$$f(S \cup T) = f(S \cup f(T)).$$
DI: Distributed Increasing
$$f(S \cup T) \subset f(S) \cup f(T).$$
DD: Distributed Decreasing
$$f(S \cup T) \supset f(S) \cup f(T).$$
DE: Distributed Equivalence
$$f(S \cup T) = f(S) \cup f(T).$$
PI: Parallel Increasing

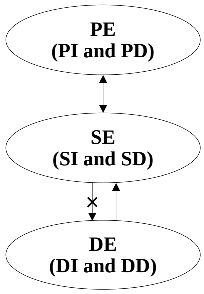
[1] In this paper, $A \subset B$ means that A is a subset of B (including the case where A = B).



$$f(S \cup T) \subset f(f(S) \cup f(T)).$$
PD: Parallel Decreasing
$$f(S \cup T) \supset f(f(S) \cup f(T)).$$
PE: Parallel Equivalence
$$f(S \cup T) = f(f(S) \cup f(T)).$$
M: Monotone
$$if \quad S \subset T \quad then \quad f(S) \subset f(T).$$
C: Consistency
$$f(S) \supset f(S \cup T) \cap S.$$

Here, S and T are arbitrary subsets of $\mathbf{T}$.

Figure 1 shows the relationship between these properties of the filtering function proved in our previous work. The arrows in Figure 1 represent the inclusion relation between the properties, and "×" represents that there is no inclusion relation between them. The arrows between "M, DD" and "SD" mean that the filtering function that satisfies the monotone property (M) (or the distributed decreasing property (DD), which is equivalent to M) also satisfies the sequential decreasing property (SD), and that the filtering function that satisfies the sequential decreasing property (SD) does not necessarily satisfy the monotone property (M) (and the distributed decreasing property (DD)). Moreover, the properties in an ellipse, for example "M, DD," are equivalent. The rectangular frame that includes some ellipses represents the property that satisfies all properties in the frame. The rectangular frame in Figure 1 means that the filtering function that satisfies the parallel decreasing property (PD) and the sequential increasing property (SI) (or PI, DI, C, which are equivalent to SI) also satisfies the sequential decreasing property (SD). However, since it has not yet been proved whether the filtering function that satisfies only the parallel decreasing property (PD) also satisfies the sequential decreasing property (SD), it is represented by dotted line.



x	$f(x)$	$g(x)$	$f(g(x))$
ϕ	ϕ	ϕ	ϕ
$\{a\}$	$\{a\}$	ϕ	ϕ
$\{b\}$	$\{b\}$	$\{b\}$	$\{b\}$
$\{a,b\}$	$\{a\}$	$\{a,b\}$	$\{a\}$

The sequential equivalence property (SE) signifies that the filtering results of batch processing and sequential processing are equivalent. Similarly, the distributed equivalence property (DE) signifies that the filtering results of batch processing and distributed processing are equivalent, and the parallel equivalence property (PE) signifies that the filtering results of batch processing and parallel processing are equivalent.

From the relationship between the properties shown in Figure 1, we can show the interrelation between those equivalence properties in Figure 2. Figure 2 shows that if the filtering results of batch processing and distributed processing are equivalent, meaning that the distributed equivalence property is satisfied, then the filtering results of sequential processing and parallel processing are also equivalent. This means that the sequential equivalence and parallel equivalence properties are also satisfied. Similarly, if the filtering results of batch processing and sequential processing are equivalent, then the filtering result of parallel processing is also equivalent. Indeed, if the filtering results of batch processing and parallel processing are equivalent, the filtering result of sequential processing is also equivalent. From the relationship between the properties shown in Figures 1 and 2, we are able to judge whether one filtering method that satisfies a specific property also satisfies another one, and to replace the processing method with a more efficient one according to the environment.

3. Composite Filtering Function

Generally, most filtering methods actually consist of combinations of methods. In this section, we show the conditions needed for a composite filtering function to be a filtering function.

In general, for functions $f : D_2 \rightarrow D_3$ and $g : D_1 \rightarrow D_2$, composing a function $h : D_1 \rightarrow D_3$, which is established by $h(x) = f(g(x))$, is called the composition of f and g, denoted as $h = f \circ g$. The function h is called the composite function of f and g. Here, we call the composite function of filtering functions the *composite filtering function*. In addition, when $f : D_1 \rightarrow D_2$, we designate

$$Im(f) \triangleq \{f(X)|X \in D_1\}$$

as the range of f.

Here, we note that a composite function of filtering functions is not necessarily always a filtering function. Table 1 shows an example of filtering functions f and g, whose composite function $f \circ g$ does not satisfy the idempotent property.

For filtering functions f and g, we call "*f can form a filtering composite with g*" when the composite function $f \circ g$ is a filtering function. Generally, the fact that f can form a filtering composite with g does not necessarily mean that g can form a filtering composite with f. For f and g in Table 1, for example, the former is not satisfied, but the latter is satisfied. Furthermore, we define that "*f is constant to g*," as $f(X) = g(X)$ is satisfied for all $X \in Im(f \circ g)$. Here, we present the following theorem:

Theorem 1. For filtering functions f and g, the fact that f can form a filtering composite with g is equivalent to that f is constant to g.

Proof.

($\Longrightarrow$) There exist $X_0 \subset \mathbf{T}$ and $Y_0 = f(g(X_0))$. When f can form a filtering composite with g, assume that $f(Y_0) \neq g(Y_0)$ for X_0 and Y_0.

Here, at least one of the inequalities $Y_0 \neq f(Y_0)$ and $Y_0 \neq g(Y_0)$ is satisfied. Therefore, since f and g satisfy the decreasing property, it is deduced that

$$\begin{aligned} Y_0 &\neq f(g(Y_0)) \\ f(g(X_0)) &\neq f(g(f(g(X_0)))). \end{aligned} \quad (1)$$

However, this does not satisfy the idempotent property of $f \circ g$, which contradicts the theory that f can form a filtering composite with g.

($\Longleftarrow$) First of all, we indicate that for all $X \in Im(f \circ g)$,

$$X = f(X) = g(X) \quad (2)$$

is satisfied when $f(X) = g(X)$. There exist $X_0 \subset \mathbf{T}$ and $Y_0 = f(g(X_0))$. If we assume that $Y_0 \neq f(Y_0) = g(Y_0)$ for X_0 and Y_0, then we derive that

$$f(g(X_0)) \neq f(f(g(X_0))), \tag{3}$$

which does not satisfy the idempotent property of f. Therefore, since (3) contradicts that f is a filtering function, (2) is formed.

If $X = f(X) = g(X)$ for all $X \in Im(f \circ g)$, then $X = f(g(X))$ is implied. Thus, the composite function $f \circ g$ satisfies the idempotent property. Additionally, since f and g satisfy the decreasing property, $X \supset g(X) \supset f(g(X))$ is deduced. Consequently, $f \circ g$ also satisfies the decreasing property. □

Next, we present the following theorem about whether a filtering function that satisfies a certain property can form a filtering composite:

Theorem 2. For filtering functions f and g, if g satisfies the sequential increasing property (equivalent to the parallel increasing, distributed increasing, or consistency property), then f can form a filtering composite with g.

Proof. Assume that g satisfies the consistency property. If $X = g(S)$, $Y = f(X)$ for $S \subset \mathbf{T}$, then

$$g(X) = g(g(S)) = g(S) = X, \qquad (\because \text{ID})$$
$$f(Y) = f(f(X)) = f(X) = Y. \qquad (\because \text{ID})$$

Since S, X and Y satisfy $S \supset X \supset Y$ due to the decreasing property,

$$\begin{aligned} g(Y) &\supset g(Y \cup S) \cap Y \\ &= g(S) \cap Y \\ &= X \cap Y = Y \end{aligned} \tag{4}$$

is deduced. Also, as $g(Y) \subset Y$ from the decreasing property, $g(Y) = Y$ is derived from (4). Therefore, $f(Y) = g(Y)$ is shown. Since f is constant to g, f can form a filtering composite with g from Theorem 1. □

We can confirm that the inverse is not satisfied.

For filtering functions f and g, when g does not satisfy the sequential increasing property (equivalent to the parallel increasing, distributed increasing, or consistency property), it is known whether f can form a filtering composite with g by determining whether f is constant to g from Theorem 1.

4. The Properties of Composite Filtering Functions

In this section, we clarify properties of composite filtering functions for various combinations of filtering functions that satisfy the properties denoted in Subsection 2.2. In Subsection 4.1, we show the properties for the combinations of the filtering functions that satisfy the increasing or decreasing properties. In Subsection 4.2, we present the properties for the combinations of the filtering functions that satisfy the equivalence properties.

For the properties denoted in 2.2, excepting equivalence properties, the monotone (M), sequential increasing (SI), sequential decreasing (SD), and parallel decreasing (PD) properties are not equivalent to each other. In this subsection, we discuss the composite filtering functions for all combinations of the filtering functions that satisfy those four properties, and introduce the following lemmas to reveal their properties. We omit the proofs of the lemmas.

Lemma 1. For filtering functions f and g, if f and g satisfy the monotone property, and f can form a filtering composite with g, then $f \circ g$ satisfies the monotone property. □

Lemma 2. For filtering functions f and g, if f satisfies the monotone property, and g satisfies the sequential increasing property, then $f \circ g$ does not necessarily satisfy the monotone or sequential increasing property. □

Lemma 3. For filtering functions f and g, if f satisfies the monotone property, g satisfies the sequential decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the monotone or sequential decreasing property. □

Lemma 4. For filtering functions f and g, if f satisfies the monotone property, g satisfies the parallel decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the monotone or parallel decreasing property. □

Lemma 5. For filtering functions f and g, if f satisfies the sequential increasing property, g satisfies the monotone property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential increasing or monotone property. □

Lemma 6. For filtering functions f and g, if f and g satisfy the sequential increasing property, then $f \circ g$ does not necessarily satisfy the sequential increasing property. □

Lemma 7. For filtering functions f and g, if f satisfies the sequential increasing property, g satisfies the sequential decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential increasing or sequential decreasing property. □

Lemma 8. For filtering functions f and g, if f satisfies the sequential increasing property, g satisfies the parallel decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential increasing or parallel decreasing property. □

Lemma 9. For filtering functions f and g, if f satisfies the sequential decreasing property, g satisfies the monotone property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential decreasing or monotone property. □

Lemma 10. For filtering functions f and g, if f satisfies the sequential decreasing property, and g satisfies the sequential increasing property, then $f \circ g$ does not necessarily satisfy the sequential decreasing or sequential increasing

$f \setminus g$	M	SI	SD	PD
M	M	¬M, ¬SI	¬M, ¬SD	¬M, ¬PD
SI	¬M, ¬SI	¬SI	¬SI, ¬SD	¬SI, ¬PD
SD	¬M, ¬SD	¬SI, ¬SD	¬SD	¬SD, ¬PD
PD	¬M, ¬PD	¬SI, ¬PD	¬SD, ¬PD	¬PD

property. □

Lemma 11. For filtering functions f and g, if f and g satisfy the sequential decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential decreasing property. □

Lemma 12. For filtering functions f and g, if f satisfies the sequential decreasing property, g satisfies the parallel decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the sequential decreasing or parallel decreasing property. □

Lemma 13. For filtering functions f and g, if f satisfies the parallel decreasing property, g satisfies the monotone property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the parallel decreasing or monotone property. □

Lemma 14. For filtering functions f and g, if f satisfies the parallel decreasing property, and g satisfies the sequential increasing property, then $f \circ g$ does not necessarily satisfy the parallel decreasing or sequential increasing property. □

Lemma 15. For filtering functions f and g, if f satisfies the parallel decreasing property, g satisfies the sequential decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the parallel decreasing or sequential decreasing property. □

Lemma 16. For filtering functions f and g, if f and g satisfy the parallel decreasing property, and f can form a filtering composite with g, then $f \circ g$ does not necessarily satisfy the parallel decreasing property. □

Table 2 shows the properties of composite filtering functions for all filtering function combinations that satisfy the increasing or decreasing properties as proved by the above lemmas. In Table 2, each element represents the property of the composite filtering function $f \circ g$ when f and g respectively satisfy the properties in the columns and rows, and when f can form a filtering composite with g. Additionally, "¬" means that the composite filtering function does not necessarily satisfy the property added it.

From Table 2, only when the filtering functions f, g that satisfy the monotone property are combined, the composite filtering function $f \circ g$ maintains the properties satisfied by the original functions f, g. On the other hand, if the filtering functions f, g that satisfy the property other than the monotone property are combined, then the composite filtering function $f \circ g$ does not necessarily maintain the properties satisfied by the original functions f, g.

In the previous subsection, we have shown the filtering functions that satisfy only the increasing or decreasing properties. In this subsection, we clarify the properties of the composite functions of the filtering functions that satisfy the distributed equivalence, sequential equivalence, or parallel equivalence property.

The following lemmas relate to the composition of the filtering functions that satisfy the equivalence properties:

Lemma 17. For filtering functions f and g, if f and g satisfy the distributed equivalence property, then $f \circ g$ satisfies the distributed equivalence property.

Proof. For f and g that satisfy the distributed equivalence property, we prove that

$$f(g(S \cup T)) = f(g(S)) \cup f(g(T)) \tag{5}$$

is satisfied. In [13],

$$\begin{aligned} &\forall S, \forall T, f(S \cup T) = f(S) \cup f(T) \\ &\quad \Longleftrightarrow \quad \exists X, \forall S, f(S) = S \cap X, \\ &\forall S, \forall T, g(S \cup T) = g(S) \cup g(T) \\ &\quad \Longleftrightarrow \quad \exists Y, \forall S, g(S) = S \cap Y \end{aligned}$$

are proved. Thus, if we assume that $X = f(\mathbf{T})$ and $Y = g(\mathbf{T})$, then $f(A) = A \cap X$, $g(A) = A \cap Y$ for all $A \subset \mathbf{T}$. For both sides in (5), the following equations are respectively derived:

$$\begin{aligned} &f(g(S \cup T)) \\ &\quad = \quad f((S \cup T) \cap Y) \\ &\quad = \quad f((S \cap Y) \cup (T \cap Y)) \\ &\quad = \quad ((S \cap Y) \cup (T \cap Y)) \cap X \\ &\quad = \quad (S \cap Y \cap X) \cup (T \cap Y \cap X), \end{aligned} \tag{6}$$

$$\begin{aligned} &f(g(S)) \cup f(g(T)) \\ &\quad = \quad f(S \cap Y) \cup f(T \cap Y) \\ &\quad = \quad (S \cap Y \cap X) \cup (T \cap Y \cap X). \end{aligned} \tag{7}$$

Therefore, from (6) and (7),

$$f(g(S \cup T)) = f(g(S)) \cup f(g(T))$$

is deduced. □

Lemma 18. For filtering functions f and g, if f satisfies the distributed equivalence property, and g satisfies the sequential equivalence property (equivalent to the parallel equivalence property), then $f \circ g$ does not necessarily satisfy the distributed equivalence, sequential equivalence, or parallel equivalence property. □

Lemma 19. For filtering functions f and g, if f satisfies the sequential equivalence property (equivalent to the parallel equivalence property), and g satisfies the distributed equivalence property, then $f \circ g$ satisfies the sequential equivalence and parallel equivalence properties.

Proof. Since the sequential equivalence property is equivalent to the parallel equivalence property from Figure 2, we prove that

$$f(g(S \cup T)) = f(g(S \cup f(g(T)))) \quad (8)$$

for f that satisfies the sequential equivalence property, and g that satisfies the distributed equivalence property. In [13], it is proved that

$$\forall S, \forall T, g(S \cup T) = g(S) \cup g(T) \iff \exists X, \forall S, g(S) = S \cap X.$$

Hence, if we assume that $X = g(\mathbf{T})$, then $g(A) = A \cap X$ for all $A \subset \mathbf{T}$. For both sides in (8), the following equations are respectively derived:

$$\begin{aligned} & f(g(S \cup T)) \\ & \quad = f((S \cup T) \cap X) \\ & \quad = f((S \cap X) \cup (T \cap X)) \\ & \quad = f((S \cap X) \cup f(T \cap X)), \quad (9) \\ & f(g(S \cup f(g(T)))) \\ & \quad = f((S \cup f(T \cap X)) \cap X) \\ & \quad = f((S \cap X) \cup (f(T \cap X) \cap X)). \quad (10) \end{aligned}$$

Here, we introduce the following lemma:

Lemma 20. For all sets T and X, $f(T \cap X) = f(T \cap X) \cap X$ is satisfied.

Proof.

i) It is trivial that $f(T \cap X) \supset f(T \cap X) \cap X$.

ii) We prove that $f(T \cap X) \subset f(T \cap X) \cap X$. From the decreasing property, for all $x \in f(T \cap X)$,

$$x \in f(T \cap X) \subset T \cap X \subset X$$

$$\therefore x \in f(T \cap X) \cap X.$$

From i) and ii), this lemma is proved. □

Therefore, from (10),

$$\begin{aligned} & f(g(S \cup f(g(T)))) \\ & \quad = f((S \cap X) \cup f(T \cap X)) \quad (11) \end{aligned}$$

is formed, and from (9) and (11), we deduce

$$f(g(S \cup T)) = f(g(S \cup f(g(T)))).$$

□

Lemma 21. For filtering functions f and g, if f satisfies the sequential equivalence property (equivalent to the parallel equivalence property), and g satisfies the distributed equivalence property, then $f \circ g$ does not necessarily satisfy the distributed equivalence property. □

$f \setminus g$	DE	SE, PE
DE	DE	¬DE, ¬SE, ¬PE
SE, PE	SE, PE, ¬DE	¬SE, ¬PE

Lemma 22. For filtering functions f and g, if f and g satisfy the sequential equivalence property (equivalent to the parallel equivalence property), then $f \circ g$ does not necessarily satisfy the sequential equivalence or parallel equivalence property. □

Table 3 shows the properties of composite filtering functions for all filtering function combinations that satisfy the equivalence properties as proved by the above lemmas. From Table 3, we know that if the filtering function g that satisfies the distributed equivalence property is precedently applied to the data set, then the composite filtering function $f \circ g$ maintains the property satisfied by the subsequently applied function f. This is true even if the filtering function f satisfies any one of the distributed equivalence, sequential equivalence, and parallel equivalence properties.

5. Observations

In this section, we address some filtering methods currently applied in practice, and discuss properties of those methods by applying the notion of composite filtering functions.

The filtering methods that decide whether the data should be stored per data item (for example, the filtering by keyword matching or the expiration date of the data) do not consider the correlation between the data filtered together. Thus, those filtering methods satisfy the distributed equivalence property (equivalent to the monotone and consistency properties)[12]. Therefore, from the results shown in 4.2, the filtering results of batch processing, distributed processing, sequential processing and parallel processing are equivalent since the combined method of those filtering methods satisfies the distributed equivalence property.

On the other hand, some filtering methods change the evaluation value of the data according to the correlation between the data filtered together. These types of filtering methods can be divided into the following two methods. The first type of filtering methods upgrade the evaluation value of the data when particular data are filtered together. Those methods consider the correlation between the data of serialized broadcast, or segmentalized data, etc. The second type of filtering methods downgrade the evaluation value of the data when particular data are filtered together. Those methods downgrade, for example, the evaluation value of the existing data when their updated data are received, such as data on weather forecasts or TV programs. The first methods satisfy the monotone, sequential decreasing, and parallel decreasing properties, but not the consistency property. The second methods satisfy the consistency, sequential decreasing, and parallel decreasing properties, but not the monotone property[12].

From the results of 4.1, the composite functions of the filtering functions satisfy the monotone property only if the filtering functions satisfy the monotone property. Otherwise, whichever combination of properties the filtering functions satisfy, the composite filtering functions may not necessarily maintain the properties satisfied by the original functions. Consequently, if the methods that upgrade the evaluation value of the data when particular data are filtered together are combined, then the combined method satisfies the monotone property. Moreover, since it also satisfies the sequential decreasing and parallel decreasing properties shown in Figure 1, it has the same characteristics as the original filtering methods. On the contrary, if the methods that downgrade the evaluation value of the data when particular data are filtered together are combined, then the combined method does not necessarily satisfy the properties satisfied by the original functions. Therefore, the combined method's characteristics differ from those of the original methods. Similarly, in the case of the combining the method that upgrades the evaluation value of the data with the method that downgrades the value when particular data are filtered together, the characteristics of original methods do not take over.

Bell and Moffat[3] propose a filtering method based on tf-idf. They intend to decrease the calculation time and memory capacity needed, and to improve the throughput and scalability by incorporating various information retrieval and filtering methods. Their method satisfies the distributed equivalence property, for it does not consider the correlation between the data filtered together, and stores the data if its evaluation value, which is calculated per data item, exceeds a certain threshold. In the same way, SIFT[15] and WebMate[5] use tf-idf and thresholds. SIFT is applied to USENET news, and WebMate recommends the Web pages. Moreover, Callan[4] verifies the effect of threshold value on precision and recall, and proposes an algorithm to automatically decide the appropriate threshold value.

When combining those filtering methods, a low-cost method can be used by appropriately combining strategies whose threshold values or vectors expressions are different. For example, the processing cost can be reduced by using pre-processing by the filtering with low threshold value before precise filtering with high-dimensional vector operations. Here, since the above filtering methods satisfy the distributed equivalence property, all combinations of these filtering methods satisfy the distributed equivalence property. Therefore, under those methods, the characteristics of the original methods remain. Moreover, since the filtering results of batch processing, distributed processing, sequential processing, and parallel processing are equivalent, the processing cost can be reduced further by changing the processing method according to the environment, such as the number and disposal capacity of the receivers, or network bandwidth.

ProfBuilder[14] filters data by ranking method according to the user's profile if the user selects the content-based filtering option. This filtering satisfies the consistency and sequential decreasing properties (equivalent to the sequential equivalence property) while the user's profile is not updated. Therefore, from the results given in this paper, a combined filtering method of this type does not necessarily satisfy the sequential equivalence property. In such filtering, there is no assurance that the filtering result of batch processing is equivalent to that of distributed processing, sequential processing, or parallel processing. In other words, it is impossible to change a processing method in the process of filtering while attempting to maintain the equivalence of filtering results. Consequently, we must sufficiently examine the filtering environment during implementation, and decide the most appropriate processing method. However, if we combine ProfBuilder after the filtering that satisfies the distributed equivalence property, then it satisfies the sequential equivalence and parallel equivalence properties. Thus, the filtering results of batch processing, sequential processing, and parallel processing are equivalent. In such filtering, it is possible to keep the ranking up to date by sequential processing, and the processing cost of the receivers and the necessary network bandwidth can be reduced by parallel processing.

AIS (Active Information Store)[10] filters broadcast data by two steps. AIS precedently filters the received data by keyword matching, and subsequently decides the data to be stored by the method that downgrades the evaluation value of the data when particular data are filtered together. The precedent processing satisfies the distributed equivalence property because it filters per data item, and the subsequent processing satisfies the sequential equivalence property. Therefore, AIS can be regarded as a combined method of the filtering that satisfies the distributed equivalence property and the filtering that satisfies the sequential equivalence property. Thus, since AIS satisfies the sequential equivalence and parallel equivalence properties from Table 3, the filtering results of batch processing, sequential processing, and parallel processing are equivalent. Consequently, the load on receivers can be reduced if the system filters the data after accumulating a certain amount. Parallel processing can also be done if there are many channels which broadcast data.

6. Conclusions and Future Work

In this paper, we have shown a condition for composite filtering functions to act as filtering functions and revealed characteristics of composite functions of filtering functions that satisfy various properties. By introducing the concept of composition into the framework of filtering functions, we can qualitatively represent complex combined methods of filtering. Additionally, we classified the filtering methods currently used in practice according to their properties, and discussed the processing methods that can be replaced while preserving the equivalence of filtering results. We can achieve more efficient filtering processes according to the environment by applying the mathematical foundation established in this paper to filtering methods currently used in practice.

Our future work includes the following:

- Adding constraints to the composite filtering function
  Most composite filtering functions denoted in this paper do not necessarily hold properties satisfied by the original functions. Therefore, we must know when we may combine some methods currently used in practice. However, by placing specific constraints on each filtering function during composition, the properties of the original functions may be maintained after composition. We will define such constraints.

- Defining new properties
  To present any filtering methods, it is necessary to introduce filtering functions that satisfy new properties that are not given in this paper. In particular, we plan to define new properties that have the same feature as the monotone property: the property of a filtering function that holds after combination. Moreover, by discovering characteristics of the new property, we clarify contributing factors in cases where the original properties do not hold.

Acknowledgements

This research was supported in part by "The 21st Century Center of Excellence Program" and Special Coordination Funds for promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] M. Altinel and M. J. Franklin: "Efficient filtering of XML documents for selective dissemination of information," in *Proc. 26th International Conference on Very Large Data Bases (VLDB2000)*, pp. 53–64 (2000).

[2] N. J. Belkin and W. B. Croft: "Information filtering and information retrieval: two sides of the same coin?," *Communications of the ACM*, vol. 35, no. 12, pp. 29–38 (1992).

[3] T. A. H. Bell and A. Moffat: "The design of a high performance information filtering system," in *Proc. SIGIR '96*, pp. 12–20 (1996).

[4] J. Callan: "Learning while filtering documents," in *Proc. SIGIR '98*, pp. 224–231 (1998).

[5] L. Chen and K. Sycara: "WebMate: a personal agent for browsing and searching," in *Proc. Second International Conference on Autonomous Agents*, pp. 132–139 (1998).

[6] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry: "Using collaborative filtering to weave an information TAPESTRY," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70 (1992).

[7] S. Loeb: "Architecting personalized delivery of multimedia information," *Communications of the ACM*, vol. 35, no. 12, pp. 39–48 (1992).

[8] S. Loeb and D. Terry: "Information filtering," *Communications of the ACM*, vol. 35, no. 12, pp. 26–28 (1992).

[9] J. Mostafa, S. Mukhopadhyay, W. Lam, and M. Palakal: "A multilevel approach to intelligent information filtering: model, system, and evaluation," *ACM Transactions on Information Systems*, vol. 15, no. 4, pp. 368–399 (1997).

[10] S. Sanguantrakul, T. Terada, M. Tsukamoto, S. Nishio, K. Miura, S. Matsuura, and T. Imanaka: "A user customized selection and categorization for broadcast data," in *Proc. 1999 IEEE International Workshops on Multimedia Network Systems(MMNS)*, pp. 596–601 (1999).

[11] Satellite Magazine: http://www.satemaga.co.jp/.

[12] R. Sawai, M. Tsukamoto, Y. H. Loh, T. Terada, and S. Nishio: "Functional properties of information filtering," in *Proc. 27th International Conference on Very Large Data Bases (VLDB2001)*, pp. 511–520 (2001).

[13] R. Sawai, M. Tsukamoto, T. Terada, and S. Nishio: "On Selection and Ranking in Filtering Function," IPSJ Transactions on Databases, vol.43, no. SIG12(TOD16), pp. 80–91, in Japanese (2002).

[14] A. M. A. Wasfi: "Collecting user access patterns for building user profiles and collaborative filtering," in *Proc. 1999 International Conference on Intelligent User Interfaces*, pp. 57–64 (1999).

[15] T. W. Yan and H. Garcia-Molina: "The SIFT information dissemination system," *ACM Transactions on Database Systems*, vol. 24, no. 4, pp. 529–565 (1999).