LNCS 2973

YoonJoon Lee
Jianzhong Li
Kyu-Young Whang
Doheon Lee (Eds.)

Database Systems for Advanced Applications

9th International Conference, DASFAA 2004
Jeju Island, Korea, March 2004
Proceedings

Springer

Proc. of the 9th International Conference on Databese Systems for
Advanced Applications (DASFAA 2004), pp. 738--749 (Mar. 2004)

Union and Intersection of Filtering Functions for Information Filtering

Rie Sawai, Masahiko Tsukamoto, Tsutomu Terada, and Shojiro Nishio

Graduate School of Information Science and Technology, Osaka University
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Abstract. In our previous works, to establish mathematical foundation of information filtering, we defined the notion of filtering function that represents filtering as a function, and clarified the characteristics of filtering. The constructed mathematical foundation makes it possible to qualitatively evaluate various filtering methods, to optimize processing methods in filtering, or to design a declarative language for describing the filtering policy. Moreover, since current filtering methods consist of multiple methods, we have revealed the properties of composite filtering functions. However, we have not considered operations without composition. In this paper, we define filtering functions that carry out union and intersection of the filtering results, and clarify their properties. Results show that we can qualitatively represent the filtering combined by more diverse strategies and reveal their characteristics.

1 Introduction

In recent years, the number of broadcast services has increased due to the introduction of new satellite-based services and the digitization of broadcasts[9]. In this environment, not only is the amount of data being broadcast increasing, but so is the variety. However, users often only need small amounts of specific data, and it is very difficult for users to retrieve the information they are interested in from the large amount of broadcast data available. Therefore, various mechanisms that automatically filter data, and user-request description languages for filtering, have been proposed[1, 3, 4, 7, 8]. These filtering mechanisms filter data by different criteria such as keyword matching or relevance feedback. 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, nor to design a declarative language for filtering processes. In [10], we defined a *filtering function* that expresses filtering as a function, and this function made it possible to qualitatively represent several properties of filtering by satisfying relevant constraints. Moreover, since an actual filtering method generally consists of multiple methods, we clarified the properties of composite functions of filtering functions, which we call composite filtering functions[12].

Composite filtering functions can represent filtering combining multiple methods sequentially, such as a filtering that uses pre-processing by a simple method, and calculates the precise result by another complex method. However, there are many filterings composed of multiple methods in other ways, which cannot be represented by composite filtering functions. For example, consider a user request "I want both data items that include certain keywords and data items that belong to a particular genre." This type of request requires all results of filtering methods whose policies are different. In other words, the filtering method carries out union of different filtering results. On the other hand, there is a filtering to extract the data items which multiple filtering methods recommend in order to improve filtering precision. Such filtering carries out intersection of different filtering results. In this way, a composite filtering function, which has been dealt with in our previous works, cannot represent these filterings that operate union or intersection of multiple methods.

In this study, we introduce the concept of union and intersection into the framework of filtering functions. We define new filtering functions that carry out union and intersection of the filtering results, and clarify their properties. By introducing the concept of union and intersection into the framework of filtering functions, we can qualitatively express the filtering combined by more diverse strategies, not only composition. Exploiting the results of this paper, we can reveal the characteristics of filtering to combine multiple methods that satisfy the various properties.

This paper is organized as follows. Section 2 outlines the filtering function we defined in our previous works. Section 3 defines new filtering functions that carry out union and intersection of the filtering results, and clarifies their properties. Section 4 considers the filtering methods currently applied in practice through the results clarified in this paper. Finally, we conclude our paper in Section 5.

2 Preliminaries

In this section, we outline the filtering function described in [10], which is the foundation of this study.

2.1 Categorization of the Filtering Processes

We categorize in this study the filtering processes in the real world into several patterns by the number of filtering processes and receivers, as follows:

In a system using *sequential processing*, newly received data and previously filtered results, which have already been stored, are merged and filtered every time new data is received. On the contrary, in a system using *batch processing*, a receiver accumulates broadcast data and filters them out in bulk. In a system using *distributed processing*, the received data set is divided into multiple arbitrary data subsets, and each subset is filtered separately before the results are merged. Moreover, in a system using *parallel processing*, the merged filtering results of distributed processing are re-filtered.

2.2 Properties of Filtering Functions

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

The following properties of a filtering function are defined:

M: monotone	$if \quad S \subset T \quad then \quad f(S) \subset f(T)$.
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	$f(S \cup T) \subset f(f(S) \cup f(T))$.
PD: parallel decreasing	$f(S \cup T) \supset f(f(S) \cup f(T))$.
PE: parallel equivalence	$f(S \cup T) = f(f(S) \cup f(T))$.
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	$f(S \cup T) = f(S \cup f(T))$.
C: consistency	$f(S) \supset f(S \cup T) \cap S$.

Here, S and T are arbitrary subsets of $\mathbf{T}$. Fig. 1 shows the relationship between these properties of the filtering function as proved in our previous works. The arrows in Fig. 1 represent the inclusion relation between the properties, while the arrows with an added "×" mean that there is no inclusion relation between them. The arrows between "M, DD" and "SD," for example, 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)). A rectangular frame including some ellipses represents the property that satisfies all properties within the frame. However, since the proposition that the filtering function that satisfies only the parallel decreasing property (PD) also satisfies the sequential decreasing property (SD), which we call assumption PD⇒SD, is not proved at this time, it is represented by a dotted line.

The sequential equivalence property (SE, which is equivalent to the property that satisfies both SI and SD) signifies that the filtering results of batch processing and sequential processing are equivalent. Similarly, the distributed equivalence property (DE (DI and DD)) and the parallel equivalence property (PE (PI and PD)) signify that the filtering results of batch processing and the corresponding processing are equivalent. From the relationship between the properties shown in Fig. 1, we know that if the filtering results of batch processing

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

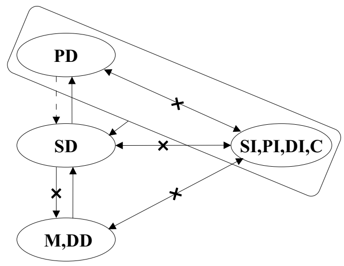

Fig. 1. The relationship between the properties of filtering function

and distributed processing are equivalent, then the filtering results of sequential processing and parallel processing are also equivalent (DE ⇒ SE, PE). Similarly, if the filtering results of batch processing and sequential (parallel) processing are equivalent, then the filtering result of parallel (sequential) processing is also equivalent (SE ⇔ PE).

3 Union and Intersection of Filtering Functions

In this section, we define the union function and the intersection function of filtering functions, and clarify their properties. First of all, we define filtering functions that carry out union and intersection of the filtering results as follows:

Let f and g be filtering functions. We define $f^{\vee}g(S) \stackrel{\triangle}{=} f(S) \cup g(S)$ for all $S \subset \mathbf{T}$, and we call this function $f^{\vee}g$ *union filtering function* of f and g. Similarly, we define $f^{\wedge}g(S) \stackrel{\triangle}{=} f(S) \cap g(S)$ for all $S \subset \mathbf{T}$, and we call this function $f^{\wedge}g$ *intersection filtering function* of f and g. Generally, the following equations are satisfied: $f^{\vee}g(S) = g^{\vee}f(S)$, $f^{\wedge}g(S) = g^{\wedge}f(S)$.

Here, we note that a union function and an intersection function of filtering functions are not necessarily always filtering functions. For filtering functions f and g, we say "*f and g are union-valid*" when the union function $f^{\vee}g$ is a filtering function. Additionally, we say "*f and g are intersection-valid*" when the intersection function $f^{\wedge}g$ is a filtering function. Here, when $f : D_1 \rightarrow D_2$, we designate $Im(f) \stackrel{\triangle}{=} \{f(X)|X \in D_1\}$ as the range of f[12]. It is clear that $f^{\vee}g$ and $f^{\wedge}g$ satisfy the decreasing property because filtering functions f and g satisfy the decreasing property; therefore, the fact that f and g are union-valid is equivalent to $X = f(X) \cup g(X)$ being satisfied for all $X \in Im(f^{\vee}g)$. Furthermore, the fact that f and g are intersection-valid is equivalent to that $Y = f(Y) \cap g(Y)$ is satisfied for all $Y \in Im(f^{\wedge}g)$. Moreover, we present the following theorems on union-validity and intersection-validity:

Theorem 1 *If filtering functions f and g satisfy the consistency property (equivalent to the distributed increasing, sequential increasing, or parallel increasing property), then f and g are union-valid and intersection-valid.* □

Theorem 2 *If filtering functions f and g satisfy the monotone property (equivalent to the distributed decreasing property), then f and g are union-valid.* □

3.1 The Properties of Union Filtering Functions

In this subsection, we clarify the properties of union filtering functions. We initially show the properties of union filtering function composed of filtering functions that satisfy the increasing or decreasing properties. Second, we present the properties of union filtering function composed of filtering functions that satisfy the equivalence properties.

Filtering Functions that Satisfy the Increasing or Decreasing Properties. For the increasing and decreasing properties denoted in Section 2, the monotone (M), sequential increasing (SI), sequential decreasing (SD), and parallel decreasing (PD) properties are not equivalent to each other. In this subsection, for filtering functions that satisfy those four properties, we reveal the properties of the union filtering functions and introduce the following lemmas. We omit the proofs for the lemmas.

Lemma 1 *If filtering functions f and g satisfy* M, *then $f^{\vee}g$ satisfies* M. □

Lemma 2 *For filtering functions f and g, if f and g are union-valid, and f satisfies* M, *g satisfies* SI, *then $f^{\vee}g$ does not necessarily satisfy* M *or* SI. □

Lemma 3 *For filtering functions f and g, if f and g are union-valid, and f satisfies* M, *g satisfies* SD, *then $f^{\vee}g$ does not necessarily satisfy* M. □

Lemma 4 *For filtering functions f and g, if f and g are union-valid, and f satisfies* M, *g satisfies* SD, *then $f^{\vee}g$ satisfies* SD. □

Lemma 5 *For filtering functions f and g, if f and g are union-valid, and f satisfies* M, *g satisfies* PD, *then $f^{\vee}g$ does not necessarily satisfy* M. □

Lemma 6 *If filtering functions f and g satisfy* SI, *then $f^{\vee}g$ satisfies* SI. □

Lemma 7 *For filtering functions f and g, if f and g are union-valid, and f satisfies* SI, *g satisfies* SD, *then $f^{\vee}g$ does not necessarily satisfy* SI *or* SD. □

Lemma 8 *For filtering functions f and g, if f and g are union-valid, and f satisfies* SI, *g satisfies* PD, *then $f^{\vee}g$ does not necessarily satisfy* SI *or* PD. □

Lemma 9 *For filtering functions f and g, if f and g are union-valid, and f and g satisfy* SD, *then $f^{\vee}g$ satisfies* SD. □

It is not clarified at this time whether assumption PD⇒SD (the filtering function that satisfies the parallel decreasing property (PD) also satisfies the sequential decreasing property (SD)) is satisfied. However, if it is clarified whether assumption PD⇒SD is satisfied, then we can show the properties of some union filtering functions from the following lemmas.

Lemma 10 *For filtering functions f and g, assume that f and g are union-valid, f satisfies* M*, and g satisfies* PD*. If assumption* PD⇒SD *is satisfied, $f^{\vee}g$ satisfies* SD*. If assumption* PD⇒SD *is not satisfied, $f^{\vee}g$ does not necessarily satisfy* SD. □

Lemma 11 *For filtering functions f and g, assume that f and g are union-valid, f satisfies* SD*, and g satisfies* PD*. If assumption* PD⇒SD *is satisfied, $f^{\vee}g$ satisfies* SD*. If assumption* PD⇒SD *is not satisfied, $f^{\vee}g$ does not necessarily satisfy* SD. □

Lemma 12 *For filtering functions f and g, assume that f and g are union-valid, and they satisfy* PD*. If assumption* PD⇒SD *is satisfied, $f^{\vee}g$ satisfies* SD*. If assumption* PD⇒SD *is not satisfied, $f^{\vee}g$ does not necessarily satisfy* SD. □

Filtering Functions that Satisfy the Equivalence Properties.

Lemma 13 *If filtering functions f and g satisfy* DE*, then $f^{\vee}g$ satisfies* DE. □

Lemma 14 *For filtering functions f and g, if f satisfies* DE*, and g satisfies* SE*, then $f^{\vee}g$ does not necessarily satisfy* DE. □

Lemma 15 *For filtering functions f and g, if f satisfies* DE*, and g satisfies* SE*, then $f^{\vee}g$ satisfies* SE. □

Lemma 16 *If filtering functions f and g satisfy* SE*, then $f^{\vee}g$ satisfies* SE. □

We omit the lemmas on whether $f^{\vee}g$ satisfies the properties other than those satisfied by the original functions f and g. Table 1 shows the properties of union filtering functions for all filtering function combinations that satisfy the increasing or decreasing properties, and Table 2 presents those for all filtering function combinations that satisfy the equivalence properties as proved by the above lemmas. In these tables, each element represents the property of union filtering function $f^{\vee}g$ when f and g respectively satisfy the properties in the columns and rows, and they are union-valid. Additionally, "¬" means that the union filtering function does not necessarily satisfy the property added to it.

The property in parentheses represents that it is not yet clarified whether $f^{\vee}g$ satisfies the property. However, we revealed that this property is deeply associated with the assumption PD⇒SD. "(SD)" represents that if PD⇒SD is satisfied, $f^{\vee}g$ also satisfies the sequential decreasing property, and that if PD⇒SD is not satisfied, $f^{\vee}g$ does not necessarily satisfy the sequential decreasing property. Moreover, it has not proved at this time whether $f^{\vee}g$ satisfies the parallel decreasing property in the following cases: f satisfies the monotone property, and g satisfies the parallel decreasing property; f satisfies the sequential decreasing property (or the parallel decreasing property), and g satisfies the parallel decreasing property. However, if assumption PD⇒SD is satisfied, then since PD and SD are equivalent (from Fig. 1), it is clarified that those union filtering functions satisfy the parallel decreasing property. In this way, the properties of some

Table 1. The properties of union filtering functions $f^{\vee}g$ for f, g that satisfy the increasing or decreasing properties

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

Table 2. The properties of union filtering functions $f^{\vee}g$ for f, g that satisfy the equivalence properties

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

union filtering functions depend on whether assumption PD⇒SD is satisfied; however, for the filtering functions whose properties have not been clarified, we must pay attention to the actual use of these filtering methods.

Table 1 clarifies that the union filtering function $f^{\vee}g$ satisfies some of the properties only when both f and g satisfy the monotone property (or sequential increasing, sequential decreasing property), or f satisfies the monotone property and g satisfies the sequential decreasing property. Additionally, from Table 2, for all filtering function combinations that satisfy the equivalence properties, the union filtering functions certainly satisfy the sequential equivalence and parallel equivalence properties.

3.2 The Properties of Intersection Filtering Functions

In this subsection, we clarify the properties of intersection filtering functions. First, as with Subsection 3.1, we show the properties of intersection filtering function composed of filtering functions that satisfy the increasing or decreasing properties (M, SI, SD, and PD, which are not equivalent to each other). Second, we present the properties of intersection filtering function composed of filtering functions that satisfy the equivalence properties.

Filtering Functions that Satisfy the Increasing or Decreasing Properties.

Lemma 17 *For filtering functions f and g, if f and g are intersection-valid, and they satisfy* M*, then $f^{\wedge}g$ satisfies* M. □

Lemma 18 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* M*, g satisfies* SI*, then $f \wedge g$ does not necessarily satisfy* M *or* SI*.* □

Lemma 19 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* M*, g satisfies* SD*, then $f \wedge g$ does not necessarily satisfy* M *or* SD*.* □

Lemma 20 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* M*, g satisfies* PD*, then $f \wedge g$ does not necessarily satisfy* M *or* PD*.* □

Lemma 21 *If filtering functions f and g satisfy* SI*, then $f \wedge g$ satisfies* SI*.* □

Lemma 22 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* SI*, g satisfies* SD*, then $f \wedge g$ does not necessarily satisfy* SI *or* SD*.* □

Lemma 23 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* SI*, g satisfies* PD*, then $f \wedge g$ does not necessarily satisfy* SI *or* PD*.* □

Lemma 24 *For filtering functions f and g, if f and g are intersection-valid, and they satisfy* SD*, then $f \wedge g$ does not necessarily satisfy* SD*.* □

Lemma 25 *For filtering functions f and g, if f and g are intersection-valid, and f satisfies* SD*, g satisfies* PD*, then $f \wedge g$ does not necessarily satisfy* SD *or* PD*.* □

Lemma 26 *For filtering functions f and g, if f and g are intersection-valid, and they satisfy* PD*, then $f \wedge g$ does not necessarily satisfy* PD*.* □

Filtering Functions that Satisfy the Equivalence Properties.

Lemma 27 *If filtering functions f and g satisfy* DE*, then $f \wedge g$ satisfies* DE*.* □

Lemma 28 *For filtering functions f and g, if f satisfies* DE*, and g satisfies* SE*, then $f \wedge g$ does not necessarily satisfy* DE *or* SE*.* □

Lemma 29 *If filtering functions f and g satisfy* SE*, then $f \wedge g$ does not necessarily satisfy* SE*.* □

We omit the lemmas on whether $f \wedge g$ satisfies the properties other than those satisfied by the original functions f and g. Table 3 shows the properties of intersection filtering functions for all filtering function combinations that satisfy the increasing or decreasing properties, and Table 4 presents those for all filtering function combinations that satisfy the equivalence properties as proved by the above lemmas.

Table 3. The properties of intersection filtering functions $f \wedge g$ for $f,\ g$ that satisfy the increasing or decreasing properties

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

Table 4. The properties of intersection filtering functions $f \wedge g$ for $f,\ g$ that satisfy the equivalence properties

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

Tables 3 and 4 clarify that the intersection filtering function $f \wedge g$ satisfies the properties satisfied by the original functions f and g only when both f and g satisfy the monotone property (or sequential increasing, distributed equivalence property). On the other hand, if filtering functions f and g satisfy the properties other than those properties, then the intersection filtering function $f \wedge g$ does not necessarily satisfy the properties dealt with in this paper.

4 Observations

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

4.1 Application of Union Filtering Functions

Fab[2] is a filtering system that has the characteristics of both a contents-based filtering method and a collaborative filtering method for web pages. In Fab, multiple collection agents collect web pages, after which a selection agent extracts necessary data from the collected data according to the user's preference. Each collection agent considers the keywords included in each data item, and collects the data associated with a particular topic. Thus, a collection agent uses the filtering method that specifies whether each data item is to be stored. This type of filtering method satisfies the distributed equivalence property[11]. Therefore, the collecting process by the collection agents is represented by a union function of the filtering functions that satisfies the distributed equivalence property.

Consequently, from Table 2, it is assured that the filtering results of batch processing, distributed processing, sequential processing, and parallel processing are equivalent.

On the other hand, a selection agent extracts the data that the user has not browsed, and selects the data from various web sites evenly. Hence, since the selection agent does not necessarily satisfy the properties dealt with in this paper, it is impossible to interchange the processing methods during the filtering process while maintaining equivalent filtering results.

Here, if interchanging a part of (or all of) the filtering methods of collection agents with filtering methods that satisfy the sequential equivalence property, then the collecting process performed by collection agents can be represented by the union function of the filtering function that satisfies the distributed equivalence property and the filtering function that satisfies the sequential equivalence property (or by the union function of the filtering functions that satisfy the sequential equivalence property). Examples of filtering that satisfy the sequential equivalence property include a ranking method and a filtering method that degrades the evaluation value of multiple data items when they are together. The ranking method is a filtering method that arranges the received data in order of importance according to the user's preferences, and extracts a particular quantity of top-ranked data. The filtering method, which degrades the evaluation value of multiple data items when they are together, is a method that considers the correlation between the contents of data items. For data items broadcast daily, such as weather forecasts and program guides, this filtering method degrades the evaluation value of an old data item when its update data item is received. If collection agents include those filtering methods, then the collection agent process satisfies the sequential equivalence and parallel equivalence properties from Table 2. Consequently, it is assured that the filtering results of batch, sequential, and parallel processing are equivalent.

Using the above characteristics, we can reduce the processing cost of filtering by changing the processing method according to environments and properties the filtering satisfies[10]. In the filtering that satisfies the parallel equivalence or sequential equivalence property, batch processing can reduce the server load when the network bandwidth is large enough. On the contrary, when the network bandwidth narrows, even if the processing method is replaced by a parallel processing method that can decentralize the network load by downloading data from multiple sites in parallel, it is certain that the filtering results will still be equivalent. Moreover, when the computational capacity of the receivers is low in the filtering that satisfies the distributed equivalence or parallel equivalence property, a high level of throughput is achieved by equipping some receivers or asking the other unoccupied ones, as in [5]. Particularly when the filtering satisfies the distributed equivalence property, we can select a distributed processing method that is more efficient than one that uses parallel processing because the number of processing actions becomes smaller. Furthermore, if filtering satisfies the sequential equivalence or parallel equivalence property, when we want the filtering results immediately, it is possible to select sequential processing.

4.2 Application of Intersection Filtering Functions

Foltz et al.[6] have shown that the precision of filtering result extracted by multiple methods is higher than that of filtering result extracted by a single method. Filtering methods that use this theorem are ones that carry out intersection of filtering results. Tables 3 and 4 show that only when all filtering functions satisfy the sequential increasing property (or monotone, distributed equivalence property), their intersection function satisfies the property that was satisfied by the original functions. Especially, if filtering functions satisfy the distributed equivalence property, their intersection function satisfies the equivalence properties (DE, SE, PE). Therefore, since it is assured that the filtering results of batch, distributed, sequential, and parallel processing are equivalent, we are able to replace the processing method with a more efficient one in accordance with the environment. However, in the other combinations of filtering functions, the intersection functions do not necessarily satisfy the equivalence properties dealt with in this paper; consequently, since there is no guarantee that the filtering results are consistent if the processing method is changed, we must sufficiently examine the filtering environment during implementation, and decide what is the most appropriate processing method to adopt.

5 Conclusions and Future Work

By introducing the concept of union and intersection into the framework of filtering functions, in this paper we established the mathematical foundation to qualitatively represent filtering that carries out union and intersection of the filtering results, and clarified their properties. Moreover, 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 in accordance with the environment by applying the mathematical foundation established in this paper to filtering methods currently used in practice.

Our future works include the following points:

- The properties of $M^{\vee}PD$, $SD^{\vee}PD$, and $PD^{\vee}PD$.
  In this paper, we clarified the properties of various combinations of filtering methods currently used in practice. However, Table 1 indicates that it has not been determined whether the above three union filtering functions satisfy the sequential decreasing and parallel decreasing properties described in this paper.
- Adding constraints to the union filtering function and intersection filtering function
  Union filtering functions and intersection filtering functions denoted in this paper do not necessarily satisfy the properties dealt with in this work. However, by placing specific constraints on each filtering function, the union function and intersection function may satisfy some properties addressed in this paper.

- Fusing to composite filtering function
  In Section 4, we considered collection agents and selection agents of Fab system separately. However, the total Fab system can be represented by one composite filtering function. In this way, there are filtering methods that employ not only union filtering functions and intersection filtering functions, but also composite filtering functions. Therefore, we will clarify the characteristics of filtering that are represented by multiple operations, such as filtering that carries out union and intersection of composite filtering functions' results.

Acknowledgements

This research was supported in part by "The 21st Century Center of Excellence Program," Special Coordination Funds for promoting Science and Technology, and Grant-in-Aid for Scientific Research (B) numbered 15300033 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. VLDB2000*, pp. 53–64 (2000).
2. M. Balabanovic and Y. Shoham: "Fab: Content-based, collaborative recommendation," *Communications of the ACM*, vol. 40, no. 3, pp. 66–72 (1997).
3. 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).
4. T. A. H. Bell and A. Moffat: "The design of a high performance information filtering system," in *Proc. SIGIR1996*, pp. 12–20 (1996).
5. T. F. Bowen, G. Gopal, G. Herman, T. Hickey, K. C. Lee, W. H. Mansfield, J. Raitz, and A. Weinrib: "The datacycle architecture," *Communications of the ACM*, vol. 35, no. 12, pp. 71–81 (1992).
6. P. W. Foltz and S. T. Dumais: "Personalized information delivery: an analysis of information filtering methods," *Communications of the ACM*, vol. 35, no. 12, pp. 51–60 (1992).
7. S. Loeb and D. Terry: "Information filtering," *Communications of the ACM*, vol. 35, no. 12, pp. 26–28 (1992).
8. 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).
9. Satellite Magazine: http://www.satemaga.co.jp.
10. R. Sawai, M. Tsukamoto, Y. H. Loh, T. Terada, and S. Nishio: "Functional properties of information filtering," in *Proc. VLDB2001*, pp. 511–520 (2001).
11. 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).
12. R. Sawai, M. Tsukamoto, T. Terada, and S. Nishio: "Composition of filtering functions," in *Proc. DASFAA2003*, pp. 293–300 (2003).