データベースシステム 128−45
（２００２．７．１８）

フィルタリング関数の合成順序について

澤井 里枝† 塚本 昌彦† 寺田 努‡ 西尾 章治郎†

†大阪大学大学院情報科学研究科マルチメディア工学専攻
{rie,tuka,nishio}@ist.osaka-u.ac.jp
‡大阪大学サイバーメディアセンターサイバーコミュニティ研究部門
tsutomu@cmc.osaka-u.ac.jp

筆者らはこれまで，情報フィルタリングの数学的基盤を構築するために，フィルタリングを関数として表すフィルタリング関数を定義し，さまざまなフィルタリングが満たす性質間の関係を明らかにしてきた．フィルタリングの数学的基盤を構築することにより，フィルタリングの定性的な評価や最適化，宣言的なフィルタリング言語の設計などが可能となる．本研究では，複数の手法を組合せたフィルタリングの性質を明らかにするため，フィルタリング関数の合成順序を入れ換えたときの性質を調べる．本研究により，複数の手法を組合せたフィルタリングにおいて，実行順序がフィルタリング結果に与える影響を定性的に示すことができる．

キーワード 情報フィルタリング，フィルタリング関数，合成関数，順序

On Composition Order of Filtering Functions

Rie SAWAI† Masahiko TSUKAMOTO† Tsutomu TERADA‡ Shojiro NISHIO†

†Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University
‡Cybercommunity Division, Cybermedia Center, Osaka University

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 relationship between properties of filtering functions. The constructed mathematical foundation makes it possible to qualitatively evaluate various filtering methods, to optimize processing methods in filtering, and to design a declarative language for describing the filtering policy. In this paper, to clarify the characteristics of filtering functions that combines two filtering methods, we investigate properties in case of changing the composition order of filtering functions. Exploiting the results of this paper, we can qualitatively indicate the effect of the execution order on the filtering results in filtering consisting of some filtering methods.

Keywords Information Filtering, Filtering Function, Composite Function, Order

1 はじめに

近年，ネットワークのブロードバンド化や，放送のデジタル化および多チャンネル化により，さまざまな放送型サービスが提供されるようになった [7, 8]．このような環境では，多様で膨大なデータを受信できるが，一般にユーザが必要とする情報はごく一部に限られているため，受信データから必要なデータを探し出すことは非常にコストの高い作業である．そこで，自動的に受信データの取捨選択をするフィルタリング機構や，フィルタリングのためのユーザ要求記述言語が多数提案されている [2, 3, 6, 9]．しかし，各フィルタリング機構は，キーワードマッチングや関連フィードバックなど，それぞれ独自の手法によってデータのフィルタリングを行っているにもかかわらず，それらの手法を定性的に表現する数学的基盤がなかった．そのため，フィルタリングの性質の定性的な評価や処理手法の最適化，宣言的なフィルタリング言語の設計などができなかった．そこで，筆者らはこれまでにフィルタリングを関数として表す**フィルタリング関数**を定義し，処理方法に関する基本的なフィルタリングの性質をフィルタリング関数が満たす制約条件として定性的に表現することを可能にした [11, 13]．さらに，筆者らは複数の手法を組合せたフィルタリングを表現するために，フィルタリング関数の合成関数の性質を明らかにした [10]．

しかし，筆者らがこれまでに構築した枠組みでは，ほとんどのフィルタリング関数の合成関数はもとのフィルタリング関数の性質を保てないことが明らかになっている [10]．すなわち，複数のフィルタリングを組合せると，一括処理から並列処理と

いった処理の等価変換がほとんどできない．

複数の手法を組合せたフィルタリングにおいて，処理の等価変換ができなくても，次のようにフィルタリングの実行順序を変えることで，より効率的な処理ができる．例えば，簡単な論理演算を行うフィルタリングと複雑な論理演算を行うフィルタリングを組合せる場合，先に簡単なフィルタリングを行い，後の複雑な計算に適用するデータを減らすことで，フィルタリング全体の計算コストを軽減できる [14]．また，受信するデータの内容や構造に応じて，蓄積条件にマッチするデータがより少ないフィルタリングを先に行うことで，初期の段階に必要なデータをより絞り込むことができる．このような処理の効率化を実現するには，処理の途中でフィルタリングの実行順序を変更しても，一貫したフィルタリング結果が得られるかどうかを明らかにする必要がある．

そこで本稿では，これまで定義してきたフィルタリング関数の合成順序を入れ換えることがフィルタリング結果に与える影響について議論する．本稿で明らかになる性質を用いることで，複数の手法を組合せたフィルタリングの実行順序を考慮した実装ができ，合成順序がフィルタリング結果に与える影響を定性的に評価できる．

以下，第 2 章でフィルタリング関数の概要を述べる．第 3 章では，筆者らがこれまで定義した性質を満たすフィルタリング関数について，合成順序を交換したフィルタリング結果の包含関係を明らかにする．第 4 章では，本稿で明らかになった結果をもとに，実際のフィルタリングシステムや関連研究を考察する．最後に第 5 章でまとめを行う．

2　フィルタリング関数

本章では，本稿の基礎となるフィルタリング関数とその合成関数について述べる．

2.1　フィルタリング処理の分類

あるフィルタリング手法が与えられたとき，実際の処理方法は以下に示すいくつかのパターンに分類できる．

データアイテムを受信する度に受信データと前回までのフィルタリング結果を合せてフィルタリングする処理方法を**逐次処理**と呼ぶ．それに対し，放送データを受信側にある程度ためておいてから一括してフィルタリングする処理方法を**一括処理**と呼ぶ．また，データ集合を 2 つ以上の任意の集合に分割して各々フィルタリングし，結果をマージしたものをフィルタリング結果とする処理方法を**分配処理**と呼ぶ．さらに，分配処理の結果を再びフィルタリングする処理方法を**並列処理**と呼ぶ．

2.2　フィルタリング関数の性質

データアイテムの集合を $\mathbf{T}$ とする．フィルタリング関数とは，任意の $T \subset \mathbf{T}$ に対し[1]，以下の 2 つの条件を満たす $2^{\mathbf{T}}$ 上の関数 f のことをいう [11, 13].

減少性 (D: Decreasing)
$f(T) \subset T$

ベキ等性 (ID: Idempotent)
$f(f(T)) = f(T)$

また，フィルタリング関数について以下のような性質が定義されている．

逐次増加性 (SI: Sequential Increasing)
$f(S \cup T) \subset f(S \cup f(T))$

逐次減少性 (SD: Sequential Decreasing)
$f(S \cup T) \supset f(S \cup f(T))$

分配増加性 (DI: Distributed Increasing)
$f(S \cup T) \subset f(S) \cup f(T)$

分配減少性 (DD: Distributed Decreasing)
$f(S \cup T) \supset f(S) \cup f(T)$

並列増加性 (PI: Parallel Increasing)
$f(S \cup T) \subset f(f(S) \cup f(T))$

並列減少性 (PD: Parallel Decreasing)
$f(S \cup T) \supset f(f(S) \cup f(T))$

単調性 (M: Monotone)
$S \subset T$ ならば　$f(S) \subset f(T)$

一貫性 (C: Consistency)
$f(S) \supset f(S \cup T) \cap S$

ここで，S, T は $\mathbf{T}$ の任意の部分集合とする．

これまでに筆者らはこれらの性質間に，図 1に示すような相互関係があることを明らかにした．図 1の矢印は包含関係を表し，包含関係がないものは “×” を付す．さらに，一つの枠内に列記した性質は同値であることを示し，異なる性質を囲った枠はそれらの性質を全て満たす性質を表す．例えば M を満たすフィルタリング関数は SD も満たし，PD かつ SI を満たすフィルタリング関数は SD を満たす．また，M と DD は同値である．ただし，PD を満たすフィルタリング関数が SD を満たすかどうか

[1] 本稿では $A \subset B$ は A が B の部分集合である (A = B の場合を含む) ことを意味するものとする．

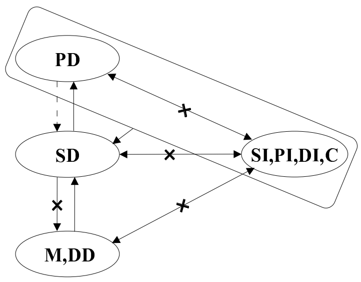

図 1: 性質間の関係

は現在のところまだ明らかとなっていないため点線で示している．

ここで，逐次増加性かつ逐次減少性を満たす性質を**逐次等価性** (SE: Sequential Equivalence) と呼び，逐次等価性を満たすフィルタリングは一括処理と逐次処理の結果が等価となることを意味する．同様に，分配増加性かつ分配減少性を満たす性質を**分配等価性** (DE: Distributed Equivalence) と呼び，分配等価性を満たすフィルタリングは一括処理と分配処理の結果が等価となることを意味する．さらに，並列増加性かつ並列減少性を満たす性質を**並列等価性** (PE: Parallel Equivalence) と呼び，並列等価性を満たすフィルタリングは一括処理と並列処理の結果が等価となることを意味する．これらの性質を用いると，図 1の性質間の関係より，ある性質を満たすシステムが他の性質を満たすかどうかが判断でき，環境に応じてより効率的な処理方法に変換できる．

2.3　フィルタリング関数の合成

フィルタリング関数の合成関数は必ずしもフィルタリング関数になるとは限らない．そこで文献 [10] では，合成関数がフィルタリング関数となるための次のような条件を示した．

フィルタリング関数 f, g に対して，**f が g にフィルタリング合成可能**であるとは，合成関数 $f \circ g$ がフィルタリング関数であることをいう．また，$f : D_1 \rightarrow D_2$ のとき $Im(f) \triangleq \{f(X) | X \in D_1\}$ を f の値域とよぶ．**f が g 不変**であるとは，任意の $X \in Im(f \circ g)$ に対して $f(X) = g(X)$ が成立することをいう．このとき以下の定理が成り立つ．

［定理 1］ フィルタリング関数 f, g に対して，f が g にフィルタリング合成可能であることと f が g 不変であることは同値である． □

3　合成関数の包含関係

本章では，前章で述べた各性質を満たすフィルタリング関数について，合成順序の交換がフィルタリング結果に与える影響を明らかにする．以下，3.1節では増加性または減少性を満たすフィルタリング関数について，3.2節では等価性を満たすフィルタリング関数について述べる．

3.1　増加性または減少性を満たす関数

2.2節に示した増加性または減少性を満たすフィルタリング関数に関して，合成順序を交換したフィルタリング結果の包含関係について以下の補題が成り立つ．

［補題 1］ フィルタリング関数 f, g が単調性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるならば，任意の $S \subset \mathbf{T}$ に対して $f(g(S)) = g(f(S))$ である．

≪ 証明 ≫

i) 任意の $S \subset \mathbf{T}$ に対して

$$f(g(S)) \subset g(f(S)) \tag{1}$$

であることを示す．

$$f(g(S)) \not\subset g(f(S)) \tag{2}$$

と仮定すると，ある $x \in f(g(S))$ に対して，

$$x \notin g(f(S)) \tag{3}$$

となる．f は g にフィルタリング合成可能なので，定理 1より，

$$f(g(S)) = f(f(g(S))) = g(f(g(S))) \tag{4}$$

が成立する．(4) と f のベキ等性より，

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

が導き出される．したがって，$x \in f(g(S))$ より，

$$x \in g(f(f(g(S)))) \tag{6}$$

が成り立つ．

一方，f と g は減少性を満たすので，

$$f(g(S)) \subset S \tag{7}$$

が成立し，さらに f と g は単調性を満たすことから，

$$
\begin{aligned}
f(f(g(S))) &\subset f(S) \\
g(f(f(g(S)))) &\subset g(f(S)) \qquad (8)
\end{aligned}
$$

となる．ここで，(3)，(6) より，(8) が矛盾する．ゆえに，(1) が成立する．

ii) 任意の $S \subset \mathbf{T}$ に対して

$$f(g(S)) \supset g(f(S)) \qquad (9)$$

であることを示す．
(省略．g が f にフィルタリング合成可能であることから，i) と同様に証明できる．)

i) と ii) から，題意は示された． □

［補題 2］ フィルタリング関数 f, g について，f が単調性，g が逐次増加性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるならば，任意の $S \subset \mathbf{T}$ に対して $f(g(S)) \subset g(f(S))$ を満たす．

≪ 証明 ≫ ある $S \subset \mathbf{T}$ に対して，

$$f(g(S)) \not\subset g(f(S)) \qquad (10)$$

と仮定すると，ある $x \in f(g(S))$ に対して

$$
\begin{aligned}
x &\in g(S) \qquad (11) \\
x &\notin g(f(S)) \qquad (12)
\end{aligned}
$$

が成立する．このとき $f(S)$ について次のような場合分けをする．

i) $x \in f(S)$ のとき
逐次増加性と一貫性は同値なので [11, 13]，g が一貫性を満たすことから，

$$
\begin{aligned}
g(f(S)) &\supset g(f(S) \cup S) \cap f(S) \\
&= g(S) \cap f(S) \\
&\ni x \quad (\because (11)) \qquad (13)
\end{aligned}
$$

となり，(13) は (12) と矛盾する．

ii) $x \notin f(S)$ のとき
g は減少性，f は単調性を満たすので，

$$
\begin{aligned}
g(S) &\subset S \\
f(g(S)) &\subset f(S) \qquad (14)
\end{aligned}
$$

が成り立つ．$x \in f(g(S))$，$x \notin f(S)$ なので，(14) は矛盾する．

表 1: 反例 1

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

表 2: 反例 2

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

i)，ii) より，任意の $S \subset \mathbf{T}$ に対して，$f(g(S)) \subset g(f(S))$ となる． □

［補題 3］ フィルタリング関数 f, g について，f が単調性，g が逐次増加性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \supset g(f(S))$ を満たさない．

≪ 証明 ≫ $\mathbf{T} = \{a, b\}$ とする．表 1に示すフィルタリング関数 f は単調性，g は逐次増加性を満たすが，$S = \{a, b\}$ のとき $f(g(S)) \supset g(f(S))$ を満たさない． □

［補題 4］ フィルタリング関数 f, g について，f が単調性，g が逐次減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ $\mathbf{T} = \{a, b, c\}$ とする．表 2に示すフィルタリング関数 f は単調性，g は逐次減少性を満たすが，$S = \{a, b, c\}$ のとき $f(g(S)) \subset g(f(S))$ も $f(g(S)) \supset g(f(S))$ も満たさない． □

［補題 5］ フィルタリング関数 f, g について，f が単調性，g が並列減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ず

表 3: 反例 3

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

しも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 4と同様に証明できる)． □

［補題 6］ フィルタリング関数 f, g が逐次増加性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ $\mathbf{T} = \{a, b\}$ とする．表 3に示す f, g は逐次増加性を満たすが，$S = \{a, b\}$ のとき $f(g(S)) \subset g(f(S))$ も $f(g(S)) \supset g(f(S))$ も満たさない． □

［補題 7］ フィルタリング関数 f, g について，f が逐次増加性，g が逐次減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

［補題 8］ フィルタリング関数 f, g について，f が逐次増加性，g が並列減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

［補題 9］ フィルタリング関数 f, g が逐次減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

［補題 10］ フィルタリング関数 f, g について，f が逐次減少性，g が並列減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

表 4: 増加性または減少性を満たす f, g に対する $f \circ g$ と $g \circ f$ の包含関係

$f \setminus g$	M	SI	SD	PD
M	$=$	$\subset, \neg\supset$	$\neg\subset, \neg\supset$	$\neg\subset, \neg\supset$
SI	–	$\neg\subset, \neg\supset$	$\neg\subset, \neg\supset$	$\neg\subset, \neg\supset$
SD	–	–	$\neg\subset, \neg\supset$	$\neg\subset, \neg\supset$
PD	–	–	–	$\neg\subset, \neg\supset$

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

［補題 11］ フィルタリング関数 f, g が並列減少性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

以上の補題から，増加性または減少性を満たすフィルタリング関数の合成関数について，合成順序を交換したフィルタリング結果の包含関係を表 4に示す．表 4中の "=" は，フィルタリング関数 f, g がそれぞれの性質を満たすときに，$f \circ g = g \circ f$ であることを示し，"⊂" は $f \circ g \subset g \circ f$ であることを示す．また，"¬" を付した関係は，必ずしもその包含関係が成立しないことを表し，"–" は重複項目があるために省略することを示す．

表 4より，単調性を満たすフィルタリング関数同士の場合のみ合成は可換となる．また，逐次増加性を満たすフィルタリングの後に単調性を満たすフィルタリングを行うフィルタリング結果のみ，合成順序を交換したフィルタリング結果に含まれることが明らかになった．それ以外の組合せでは，合成順序を交換しても包含関係は必ずしも成立しない．したがって，そのような場合，処理の途中で実行順序を変更すると，変更前に蓄積されるべきデータを変更後にも蓄積することがこれらの性質だけからでは保証できない．

3.2 等価性を満たす関数

本節では，2.2節に示した等価性を満たすフィルタリング関数，および**セレクション関数**，**ランキング関数**について，合成順序を交換したフィルタリング結果の包含関係を示す．

セレクション関数とランキング関数は文献 [12]

において次のように定義した．ある $X \subset \mathbf{T}$ に対する X のセレクション関数 B_X とは，任意の $S \subset \mathbf{T}$ に対して $B_X(S) \triangleq S \cap X$ と定義される関数である．また，ある全順序 $R = (\mathbf{T}, <)$ に対する n ランキング関数 f とは，任意の $X \subset \mathbf{T}$ に対し，ある $a \in \mathbf{T}$ について $f(X) \triangleq \{x \in \mathbf{T} | x < a\} \cap X$ と定義される度数 n の関数である．関数 f が度数 n であるとは，任意の $X \subset \mathbf{T}$ に対して

$$\begin{cases} |f(X)| = n \ (X \text{ が無限集合，あるいは} \\ \qquad\qquad X \text{ が有限集合であり } |X| \geq n \text{ のとき}) \\ f(X) = X \ (X \text{ が有限集合であり } |X| < n \text{ のとき}) \end{cases}$$

が成立することである．

セレクションとは，各データの取捨選択が潜在的に決まっている手法である．例えば，特定のキーワードを含むデータを蓄積するキーワードマッチングや，データの内容から評価値を計算し，評価値とユーザが設定した閾値とを比較することで蓄積するデータを決定する手法などはセレクションである．一方，ランキングとは，ユーザの嗜好に応じて受信データを重要な順序に並べ，その上位から特定の数のデータを選択する手法である．

セレクション関数とランキング関数に関して，以下の定理が成立する [12].

[定理 2] フィルタリング関数 f がセレクション関数であることと，f が分配等価性を満たすことは同値である． □

[定理 3] 関数 f がある全順序 $(\mathbf{T}, <)$ に対して n ランキング関数であることと，f が度数 n で逐次等価性を満たすことは同値である． □

まず，等価性を満たすフィルタリング関数について以下の補題を示す.

[補題 12] フィルタリング関数 f, g が分配等価性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるならば，任意の $S \subset \mathbf{T}$ に対して $f(g(S)) = g(f(S))$ である．

≪ 証明 ≫ 文献 [12]，定理 2より，f をある $X \subset \mathbf{T}$ のセレクション関数とし，g をある $Y \subset \mathbf{T}$ のセレクション関数とすると，

$$\begin{aligned} f(S) &= S \cap X \\ g(S) &= S \cap Y \end{aligned} \tag{15}$$

とおける．したがって，

$$\begin{aligned} f(g(S)) &= f(S \cap Y) \\ &= (S \cap Y) \cap X \\ &= (S \cap X) \cap Y \\ &= g(S \cap X) \\ &= g(f(S)) \end{aligned} \tag{16}$$

が成立する． □

[補題 13] フィルタリング関数 f, g について，f が分配等価性，g が逐次等価性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるならば，任意の $S \subset \mathbf{T}$ に対して $f(g(S)) \subset g(f(S))$ である．

≪ 証明 ≫ f が単調性，g が逐次増加性を満たすので，補題 2より成立する． □

[補題 14] フィルタリング関数 f, g について，f が分配等価性，g が逐次等価性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \supset g(f(S))$ を満たさない．

≪ 証明 ≫ 省略 (補題 3と同様に証明できる)． □

[補題 15] フィルタリング関数 f, g が逐次等価性を満たし，f が g にフィルタリング合成可能，かつ g が f にフィルタリング合成可能であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

次に，セレクション関数，ランキング関数，等価性を満たすフィルタリング関数に関して，合成順序を交換したフィルタリング結果の包含関係について以下の補題を示す.

[補題 16] 関数 f, g について，f が逐次等価性を満たすフィルタリング関数，g がランキング関数であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

[補題 17] 関数 f, g がランキング関数であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \subset g(f(S))$ または $f(g(S)) \supset g(f(S))$ とは限らない．

≪ 証明 ≫ 省略 (補題 6と同様に証明できる)． □

さらに，以下の補題は文献 [12] で示した.

[補題 18] 関数 f, g について，f がセレクション関数，g がランキング関数ならば，任意の $S \subset \mathbf{T}$ に対して $f(g(S)) \subset g(f(S))$ である． □

[補題 19] 関数 f, g について，f がセレクション関数，g がランキング関数であるとき，任意の $S \subset \mathbf{T}$ に対して，必ずしも $f(g(S)) \supset g(f(S))$ を満たさない． □

表 5: 等価性を満たす f, g に対する $f \circ g$ と $g \circ f$ の包含関係

f \ g	分配等価性 (セレクション)	逐次・並列等価性	ランキング
分配等価性 (セレクション)	$=$	$\subset, \neg\supset$	$\subset, \neg\supset$
逐次・並列等価性	$-$	$\neg\subset, \neg\supset$	$\neg\subset, \neg\supset$
ランキング	$-$	$-$	$\neg\subset, \neg\supset$

定理 2より，セレクション関数と分配等価性を満たすフィルタリング関数は等価なので，セレクション関数と逐次等価性を満たすフィルタリング関数，分配等価性を満たすフィルタリング関数とランキング関数，およびセレクション関数同士を合成した場合の補題は省略する．

以上の補題から，等価性を満たすフィルタリング関数の合成関数について，合成順序を交換したフィルタリング結果の包含関係を表 5に示す．定理 2より，セレクション関数と分配等価性を満たすフィルタリング関数は同じ項目に記す．表 5より，分配等価性を満たすフィルタリング関数同士の場合のみ，合成は可換であることが示された．また，逐次等価性，またはそれと等価である並列等価性を満たすフィルタリングと，セレクションによるフィルタリングとの合成において，合成順序を交換したフィルタリング結果の包含関係は，度数に依存しないことが明らかになった．

4 考察

本章では，実際に用いられているいくつかのフィルタリング手法を取り上げ，本稿で示した性質から，各手法で実現できる処理方法について述べる．

表 4に示す結果より，ほとんどのフィルタリング関数の合成関数は，合成順序を変更すると必ずしも等しいフィルタリング結果が得られないことが明らかになった．これにより，一般に処理中に実行順序を変更すると，一貫したフィルタリング結果を保証されないことが示された．

セレクションによるフィルタリングとして，キーワードマッチングを用いたものに，XML データの構造を利用する XFilter[1] や NiagaraCQ[4] などがある．また，各データにそのコンテンツから評価値を与え，評価値と閾値を比較することで蓄積するデータを決定するものに，SIFT[16] や距離の公理を用いた手法 [5] などがある．これらのフィルタリングを組合せたフィルタリングでは，表 5に示す結果より，どちらの処理を先に行っても等価な結果が得られる．ゆえに，環境に応じて，より計算コストが低く，条件にマッチするデータがより少ないセレクションを先に行うことで，初期の段階で必要なデータを絞り込み，システム全体の計算コストを軽減できる．

例えば XFilter と SIFT を組合せる場合，受信する XML データの多くが同じデータ構造をしているとき，XFilter で先に処理してもあまりデータを絞り込めない．したがって，先に SIFT で閾値以上の評価値をもつデータに絞り込むことで次の処理コストを低くできる．逆に，データ中に多数のキーワードを含むとき，データとユーザのプロファイルをベクトル表現する SIFT では，ベクトルのサイズが大きくなるため，それらのベクトルの積を計算する SIFT の計算コストが高くなってしまう．そこで，XML のデータ構造を利用する XFilter が先にデータを絞り込むことで全体の計算コストを低くできる．

ランキングによるフィルタリングとして，コンテンツによるフィルタリングをオプションとして選択した場合の ProfBuilder[15] などがある．ランキングによるフィルタリング同士を組合せる場合，実行順序にかかわらず，フィルタリング結果に包含関係が必ずしも成立しないことが明らかになった．したがって，フィルタリング処理の開始前に最適な実行順序を決定しておく必要がある．

セレクションによるフィルタリングとランキングによるフィルタリングを組合せる場合，処理の途中で実行順序を交換すると一貫したフィルタリング結果が得られなくなる．しかし，セレクションによるフィルタリングの後にランキングによるフィルタリングを行う手法の結果は，ランキングによるフィルタリングの後にセレクションによるフィル

タリングを行う手法の結果を包含することが明らかになった．したがって，実行順序を交換し，セレクションによるフィルタリングの後にランキングによるフィルタリングを行う方法へと変更しても，変更前に蓄積すべきデータも必ず残るため，フィルタリング結果を利用して処理する場合などに必要な処理が実行されることが保証できる．

5 おわりに

本稿では，さまざまな性質を満たすフィルタリング関数について，合成順序を交換したフィルタリング結果の包含関係を明らかにした．本稿によって，フィルタリング手法を組合せる場合，実行順序を定性的に考慮した実装が可能となる．また，本稿で示した体系を実際に複数の手法を組合せたフィルタリングに適用することで，環境に応じてより効率的な実行順序へと動的に変更できることを述べた．

今後の課題を以下に示す．

- 合成順序交換の条件
  本稿で論じた合成関数は必ずしも可換とならないことが明らかになった．しかし，合成順序を交換するとき，特定の条件を追加することで可換となる可能性がある．
- ベキ等性を満たさない合成関数
  本稿で構築する枠組みにおいて，フィルタリング関数はベキ等性を満たすものとしているために，扱える関数の範囲が狭められている．したがって，ベキ等性を満たさず減少性のみを満たす関数についても考察する必要がある．

謝辞

本研究は，科学技術振興調整費任期付研究者支援「情報フィルタリングの数学的基盤の確立」，および文部科学省振興調整費「モバイル環境向 P2P 型情報共有基盤の確立」の研究助成によるものである．ここに記して謝意を表す．

参考文献

[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. Chen, D. J. DeWitt, F. Tian, and Y. Wang: “NiagaraCQ: a scalable continuous query system for internet databases,” in *Proc. ACM SIGMOD2000*, pp. 379–390 (2000).

[5] カンギョウビ, 大和田俊和, 浅田一繁, 飯沢篤志, 古瀬一隆: “情報放送システムにおける距離の近似を利用したフィルタリング方式,” 電子情報通信学会第 11 回データ工学ワークショップ (DEWS2000) 論文集 (CD-ROM) (2000).

[6] 森田昌宏: “情報フィルタリングに関する研究動向,” JAIST Research Report, IS-RR-93-9I, 北陸先端科学技術大学院大学情報科学研究科 (1993).

[7] 西正, 野村敦子: “多チャンネル放送の衝撃,” 中央経済社 (1997).

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

[9] 澤井里枝, 寺田努, 塚本昌彦, 西尾章治郎: “フィルタリング SQL: フィルタリングのためのユーザ要求記述言語,” 電子情報通信学会第 11 回データ工学ワークショップ (DEWS2000) 論文集 (CD-ROM) (2000).

[10] 澤井里枝, 塚本昌彦, 寺田努, Loh Yin Huei, 西尾章治郎: “フィルタリング関数の合成とその性質について,” 情報処理学会研究報告 (データベースシステム研究会 2001-DBS-125(2)), Vol. 2001, No. 71, pp. 61–68 (2001).

[11] 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).

[12] 澤井里枝, 塚本昌彦, 寺田努, Loh Yin Huei, 西尾章治郎: “フィルタリング関数におけるセレクションとランキングの性質について,” 情報処理学会データベースと Web 情報システムに関するシンポジウム (DBWeb2001) 論文集, pp. 277–284 (2001).

[13] 澤井里枝, 塚本昌彦, 寺田努, Loh Yin Huei, 西尾章治郎: “情報フィルタリングの関数的性質について,” 電子情報通信学会論文誌 D-1 (2002, 掲載予定).

[14] 上原邦昭, 田中克己, 田島敬史: “マルチメディア・コンテンツの内容記述・検索モデル,” 日本学術振興会 平成 10 年度 未来開拓学術研究推進事業 マルチメディア・コンテンツの高次処理の研究 成果報告書, pp. 27–46 (1999).

[15] 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).

[16] 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).