

フィルタリング SQL におけるフィルタリングの合成と和積について

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

E-mail: †{rie,tuka,tsutomu,nishio}@ist.osaka-u.ac.jp

あらまし 筆者らはこれまで、フィルタリングのためのユーザ要求記述言語であるフィルタリング SQL を提案し、ユーザの嗜好やデータ管理ポリシーを自由に表現できるようにした。また、フィルタリングを関数として定義したフィルタリング関数により、フィルタリング SQL の意味を明確化した。フィルタリング SQL 記述の意味をフィルタリング関数で表現することにより、筆者らがこれまで構築してきたフィルタリングの数学的基盤を用いて、記述内容を実行するフィルタリングの性質を明らかにできる。さらに、ユーザが複雑な要求を記述する場合、複数の構文を組合せて用いるのが一般的である。そこで本研究では、合成フィルタリング関数と和積フィルタリング関数の関係を明らかにすることで、複数の構文を用いたフィルタリング SQL の意味を明確にする。合成フィルタリング関数と和積フィルタリング関数により、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できる。本研究では、それらの関係を示すことで、さまざまな記述内容を実行する手法の組合せ方に関する性質を明確にする。本研究により、複数の手法を組合せる際、手法の特性やフィルタリング環境を考慮して、より効率的なフィルタリング手法の組合せが実現できる。

キーワード 情報フィルタリング, フィルタリング SQL, フィルタリング関数, 合成, 和, 積

Composition, Union and Intersection of Filtering in FilteringSQL

Rie SAWAI[†], Masahiko TSUKAMOTO[†], Tsutomu TERADA[†], and Shojiro NISHIO[†]

[†] Graduate School of Information Science and Technology, Osaka University

E-mail: †{rie,tuka,tsutomu,nishio}@ist.osaka-u.ac.jp

Abstract In our previous works, we have proposed FilteringSQL, a user request description language for information filtering. FilteringSQL enables users to express their preference and policy of data management. Moreover, we have defined the semantics of FilteringSQL by using filtering function that represents information filtering as a function. By expressing the semantics of FilteringSQL using filtering functions, we can apply the mathematical foundation of information filtering that we have constructed, and can clarify the characteristics of filtering methods that carry out the request users described. Generally users describe their complex request using multiple sentences of FilteringSQL in practice. Therefore, in this paper, we clear the semantics of FilteringSQL combining multiple sentences by clarifying the relationship between composite filtering functions and union (or intersection) filtering functions, which can reveal the properties of filtering consisting of multiple methods. Furthermore, we reveal the properties on the way to combine the filtering methods that carry out each sentence described by users. Exploiting the results of this paper, we can achieve more efficient way of combining filtering methods considering the characteristics of each method and filtering environments.

Key words Information Filtering, FilteringSQL, Filtering Function, Composition, Union, Intersection

1. はじめに

近年、ブロードバンドネットワークの普及や、デジタル放送の開始により、さまざまな放送型サービスが提供されるようになった[4],[5],[16]。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

ことはコストの高い作業である。そのため、自動的に受信データの取捨選択をする情報フィルタリング技術に対する要求が高まっている[1]~[3],[6]。そこで筆者らは、フィルタリングに対するユーザ要求を記述するためのフィルタリング SQL を提案した[7]。フィルタリング SQL とは、データベースへの問合せ言語である SQL(Structured Query Language)をフィルタリングのために拡張した言語である。フィルタリング SQL により、

ユーザはコンテンツの嗜好やデータ管理ポリシーを自由に表現できる。また、筆者らは、フィルタリング関数 [8], [9] により、フィルタリング SQL の意味を明確化してきた [13], [15]。フィルタリング関数とは、フィルタリングを関数として定義したものである。筆者らは、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し、それらの性質間の関係を明らかにすることで、さまざまなフィルタリングの特性を明確にしてきた。したがって、フィルタリング SQL 記述の意味をフィルタリング関数で表現することにより、筆者らがこれまで構築したフィルタリングの数学的基盤を用いて、記述内容を実行するフィルタリングの性質を明らかにできる。

さらに、実際にユーザが複雑な要求を記述するには、複数の構文を組合せて用いるのが一般的であるため、合成フィルタリング関数 [11], [12]、和フィルタリング関数や積フィルタリング関数 [14] により、複数の構文を用いた記述の意味を明確化した [15]。合成フィルタリング関数により、ある簡単な手法で前処理を行ってから別の複雑な手法で精細な結果を計算するフィルタリングのように、複数の手法を連続的に組合せたフィルタリングを表現できる。一方、和フィルタリング関数と積フィルタリング関数では、複数の手法を並列に組合せ、それらの手法から得られたフィルタリング結果の和 (Union) や積 (Intersection) の演算を行うフィルタリングを表現できる。

これまで、合成フィルタリング関数で表現される記述は、必ず複数の手法を連続的に組合せたフィルタリングで実行し、和積フィルタリング関数で表現される記述は、必ず複数の手法を並列に組合せたフィルタリングで実行することとしていた。しかし、状況によっては、手法の組合せ方を変更することで、効率的なフィルタリング処理が可能である。例えば、一方の手法によるデータの絞り込み効果が高い場合は、その手法で前処理を行い、もう一方の手法を連続的に組合せることで、2 段階目のフィルタリングが処理すべきデータ数を削減できる。逆に、複数の端末が用意できる場合は、手法を並列に組合せることで、フィルタリング処理を分散できる。このように、手法の組合せ方を変更することで、効率的なフィルタリングを実現できるが、組合せ方を変更しても一貫したフィルタリング結果が得られることを保証する必要がある。

そこで本研究では、合成フィルタリング関数と和積フィルタリング関数の関係を明らかにすることで、手法の組合せ方を変更することがフィルタリング結果に与える影響を明らかにする。フィルタリング SQL のさまざまな記述構文に対し、それを実行する手法の組合せ方に関する性質を明確化することで、複数の構文を用いたフィルタリング SQL の性質を明らかにする。本研究により、複数の手法を組合せる際、手法の特性やフィルタリング環境を考慮して、より効率的なフィルタリング手法の組合せ方法を実現できる。

以下、第 2 章でフィルタリング関数の概要、第 3 章でフィルタリング SQL の構文について述べる。第 4 章では、合成フィルタリング関数と和積フィルタリング関数の関係を明らかにする。第 5 章では、第 4 章の結果をもとに、いくつかの記述例の意味を示し、効率的なフィルタリング処理について考察する。

最後に第 6 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の概要について述べる。

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よぶ。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よぶ。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よぶ。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よぶ。

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 T とする。フィルタリング関数とは、任意の $T \subset T$ に対し^(注1)、以下の 2 つの条件を満たす 2^T 上の関数 f のことをいう [8], [9]。

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)

$$f(S \cup T) \supset f(S \cup f(T))$$

逐次等価性 (SE: Sequential Equivalence)

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

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)

$$f(S \cup T) \supset f(S) \cup f(T)$$

分配等価性 (DE: Distributed Equivalence)

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

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)

$$f(S \cup T) \supset f(f(S) \cup f(T))$$

並列等価性 (PE: Parallel Equivalence)

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

単調性 (M: Monotone)

$$S \subset T \text{ ならば } f(S) \subset f(T)$$

一貫性 (C: Consistency)

$$f(S) \supset f(S \cup T) \cap S$$

(注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。

ここで, S, T は \mathbf{T} の任意の部分集合とする.

逐次等価性は一括処理と逐次処理の結果が等価であることを意味する. 同様に, 分配等価性は一括処理と分配処理の結果が等価であり, 並列等価性は一括処理と並列処理の結果が等価であることを意味する. 文献 [8], [9] では, 逐次増加性, 分配増加性, 並列増加性, 一貫性が同値であり, 分配減少性と単調性, 逐次等価性と並列等価性が同値であることを明らかにした ($SI \Leftrightarrow DI \Leftrightarrow PI \Leftrightarrow C$, $DD \Leftrightarrow M$, $SE \Leftrightarrow PE$).

2.3 合成フィルタリング関数

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 [11] では, 合成関数がフィルタリング関数となるための条件を次のように示した.

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, $f: D_1 \rightarrow D_2$ において, D_1 を定義域とよぶ. 集合 D_2 の要素のうち, f の像になっているもの全体の集合 $Im(f) \triangleq \{f(X) | X \in D_1\}$ を f の値域とよび, フィルタリング結果として取り得る値の集合を意味する. 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. \square

2.4 和積フィルタリング関数

筆者らは文献 [14] において, 和フィルタリング関数と積フィルタリング関数を次のように定義した.

f, g をフィルタリング関数とする. 任意の $S \subset \mathbf{T}$ に対して $f^\vee g(S) \triangleq f(S) \cup g(S)$ と定義される $f^\vee g$ を f と g の和フィルタリング関数とよぶ. また, $f^\wedge g(S) \triangleq f(S) \cap g(S)$ と定義される $f^\wedge g$ を f と g の積フィルタリング関数とよぶ. 一般に $f^\vee g(S) = g^\vee f(S)$, $f^\wedge g(S) = g^\wedge f(S)$ が成り立つ.

フィルタリング関数の和積関数は必ずしもフィルタリング関数でない. フィルタリング関数 f, g に対して, f と g がフィルタリング和可能であるとは, $f^\vee g$ がフィルタリング関数であることをいう. また, f と g がフィルタリング積可能であるとは, $f^\wedge g$ がフィルタリング関数であることをいう. フィルタリング関数 f, g は減少性を満たすため, $f^\vee g, f^\wedge g$ も減少性を満たすことは自明である. ゆえに, $f^\vee g$ がフィルタリング和可能であることと任意の $X \in Im(f^\vee g)$ に対して $X = f(X) \cup g(X)$ が成立することは同値であり, $f^\wedge g$ がフィルタリング積可能であることと任意の $Y \in Im(f^\wedge g)$ に対して $Y = f(Y) \cap g(Y)$ が成立することは同値である.

3. フィルタリング SQL

本章では, フィルタリング SQL の基本的な構文とその意味について述べる [7], [15]. まず, フィルタリング SQL の基本構文を以下に示す.

```
EXTRACT <属性>
FROM <リソース>
WHERE <プリファレンス>
<付加条件>
```

EXTRACT 句では, 抽出されたデータのうち蓄積する属性を指定する. FROM 句にデータ放送名などのデータリソースを指定し, WHERE 句にユーザの嗜好を表現するプリファレンスを記述する. また, 必要に応じてデータの蓄積に関するオプションとなる条件を付加できる.

次に, 複数の構文を組合せた記述を示す. ここで, Σ をフィルタリング SQL 記述の全体, F をフィルタリング関数とすると, セマンティクス関数とは, 関数 $S: \Sigma \rightarrow F$ である.

WHERE 句に複数の構文 σ_1, σ_2 を AND 演算子でつなげた記述 σ を以下に示す. ただし, 以下, 本章では WHERE 句の記述に焦点を絞った議論を行うため, 蓄積する属性やデータリソースの記述は省略する.

```
EXTRACT
WHERE  $\sigma_1$ 
AND  $\sigma_2$ 
```

このような記述 σ の意味を, 任意の $X \subset \mathbf{T}$ に対して

$$S(\sigma)(X) \triangleq S(\sigma_1) \circ S(\sigma_2)(X) \quad (1)$$

あるいは

$$S(\sigma)(X) \triangleq S(\sigma_2) \circ S(\sigma_1)(X) \quad (2)$$

と定義する. ここで, $S(\sigma_1)$ と $S(\sigma_2)$ の実行順序は, それぞれが満たす性質に応じて変更できる場合がある. 例えば, $S(\sigma_1)$ と $S(\sigma_2)$ が単調性あるいは分配等価性を満たすならば, 合成は可換である [12]. $S(\sigma_1)$ と $S(\sigma_2)$ が可換でないとき, (1) と (2) が一致しないため, あらかじめ実行順序を決定しておく必要がある.

一方, 以下のような記述 σ に対して, σ の意味を

$$S(\sigma)(X) \triangleq S(\sigma_1) \circ S(\sigma_2)(X) \quad (3)$$

と定義する.

```
EXTRACT
FROM ( EXTRACT
      FROM
      WHERE  $\sigma_2$ )
WHERE  $\sigma_1$ 
```

このように, 記述が入れ子の形になっているものについては, $S(\sigma_1)$ と $S(\sigma_2)$ の実行順序を一意に定義する.

また, 以下の記述は, 複数のフィルタリング SQL 文を演算子で結び付けたものである. 下記のように, フィルタリング SQL 文を UNION 演算子で結び付けた記述の意味を

$$S(\sigma)(X) \triangleq S(\sigma_1)^\vee S(\sigma_2)(X) \quad (4)$$

と定義する．

EXTRACT	
WHERE	σ_1
UNION	
EXTRACT	
WHERE	σ_2

同様に，下記のように，INTERSECT 演算子を用いた記述の意味を

$$S(\sigma)(X) \triangleq S(\sigma_1) \wedge S(\sigma_2)(X) \quad (5)$$

とする．

EXTRACT	
WHERE	σ_1
INTERSECT	
EXTRACT	
WHERE	σ_2

ここで，

$$S(\sigma_1) \vee S(\sigma_2)(X) = S(\sigma_2) \vee S(\sigma_1)(X) \quad (6)$$

$$S(\sigma_1) \wedge S(\sigma_2)(X) = S(\sigma_2) \wedge S(\sigma_1)(X) \quad (7)$$

が成立することは容易に確かめられる．

本稿では，上記の合成フィルタリング関数で表現される記述と，和積フィルタリング関数で表現される記述の関係を明らかにする．

4. 合成フィルタリング関数と和積フィルタリング関数の関係

本章では，これまでに筆者らが定義した合成フィルタリング関数と和積フィルタリング関数の関係を明らかにする．まず，フィルタリング関数 f, g に対して，合成フィルタリング関数 $f \circ g$ と和積フィルタリング関数 $f \vee g$ の包含関係を明らかにするために，以下の補題を示す．

[補題 1] フィルタリング関数 f, g について， f が g にフィルタリング合成可能であり， f と g がフィルタリング和可能であるならば，任意の $S \subset T$ に対して $f \circ g(S) \subset f \vee g(S)$ が成り立つ．

《証明》ある $S \subset T$ に対して $f(g(S)) \not\subset f(S) \cup g(S)$ と仮定すると，ある $x \in T$ に対して，

$$x \in f(g(S)) \quad (8)$$

$$x \notin f(S) \cup g(S) \quad (9)$$

が成立する．したがって， f の減少性と (8) より

$$x \in g(S) \quad (10)$$

となり，また (9) より

$$x \notin f(S) \text{ かつ } x \notin g(S) \quad (11)$$

表 1 反例 1

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

が導き出される．ここで，(10) と (11) が矛盾するため，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 g にフィルタリング合成可能， f と g がフィルタリング和可能であり， f が単調性あるいは逐次増加性，逐次減少性，並列減少性を満たし， g が単調性あるいは逐次増加性，逐次減少性，並列減少性を満たすとき，ある $S \subset T$ に対して $f \circ g(S) \supset f \vee g(S)$ を満たさない f, g が存在する．

《証明》 $T = \{a, b\}$ とする．表 1 に示すフィルタリング関数 f, g は，単調性，逐次増加性，逐次減少性，並列減少性を満たすが， $S = \{a, b\}$ のとき $f \circ g(S) \supset f \vee g(S)$ を満たさない． □

[補題 3] フィルタリング関数 f, g について， f が g にフィルタリング合成可能， f と g がフィルタリング和可能であり， f が分配等価性あるいは逐次等価性を満たし， g が分配等価性あるいは逐次等価性を満たすとき，ある $S \subset T$ に対して $f \circ g(S) \supset f \vee g(S)$ を満たさない f, g が存在する．

《証明》省略 (補題 2 と同様に証明できる)． □

次に，フィルタリング関数 f, g に対して，合成フィルタリング関数 $f \circ g$ と積フィルタリング関数 $f \wedge g$ の包含関係を明らかにするために，以下の補題を示す．

[補題 4] フィルタリング関数 f, g について， f が g にフィルタリング合成可能， f と g がフィルタリング積可能であり， f が単調性を満たすならば，任意の $S \subset T$ に対して $f \circ g(S) \subset f \wedge g(S)$ が成り立つ．

《証明》ある $S \subset T$ に対して $f(g(S)) \not\subset f(S) \cap g(S)$ と仮定すると，ある $x \in T$ に対して，

$$x \in f(g(S)) \quad (12)$$

$$x \notin f(S) \cap g(S) \quad (13)$$

が成立する．したがって，(13) より

$$x \notin f(S) \quad (14)$$

または

$$x \notin g(S) \quad (15)$$

となる．一方， f の減少性と (12) より

$$x \in g(S) \quad (16)$$

であるため，(15) は成り立たず，(14) のみ成り立つ．

ここで， g の減少性， f の単調性より

$$S \supset g(S) \quad (\because g : D)$$

$$f(S) \supset f(g(S)) \quad (\because f : M) \quad (17)$$

となるが、これは (12), (14) に反する。したがって、題意が示された。 □

[補題 5] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が単調性を満たすならば、任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge} g(S)$ が成り立つ。

◀ 証明 ▶ ある $S \subset T$ に対して $f(g(S)) \not\subset f(S) \cap g(S)$ と仮定すると、ある $x \in T$ に対して、

$$x \in f(S) \cap g(S) \quad (18)$$

$$x \notin f(g(S)) \quad (19)$$

が成立する。また、 f と g がフィルタリング積可能であることから、

$$f^{\wedge} g(S) = f^{\wedge} g(f^{\wedge} g(S))$$

$$f(S) \cap g(S) = f(f(S) \cap g(S)) \cap g(f(S) \cap g(S)) \quad (20)$$

が成り立つ。したがって、(18), (20) より、

$$x \in f(f(S) \cap g(S)) \cap g(f(S) \cap g(S)) \quad (21)$$

となるため、

$$x \in f(f(S) \cap g(S)) \quad (22)$$

$$x \in g(f(S) \cap g(S)) \quad (23)$$

が導き出される。ここで、

$$f(S) \cap g(S) \subset g(S) \quad (24)$$

が成り立つことは容易に確かめられる。(24) より、 f が単調性を満たすことから

$$f(f(S) \cap g(S)) \subset f(g(S)) \quad (25)$$

が導き出されるが、(25) は (19), (22) に矛盾する。 □

[補題 6] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が逐次増加性を満たすならば、任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge} g(S)$ が成り立つ。

◀ 証明 ▶ 文献 [8], [9] より、逐次増加性と一貫性は同値である。したがって、 f が一貫性を満たすことから、

$$\begin{aligned} f(g(S)) \supset f(g(S) \cup S) \cap g(S) \\ = f(S) \cap g(S) \quad (\because g : D) \end{aligned} \quad (26)$$

が導き出される。 □

[補題 7] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が逐次増加性あるいは逐次減少性、並列減少性を満たし、 g が単調性あるいは逐次増加性、逐次減少性、並列減少性を満たすとき、ある $S \subset T$ に対して $f \circ g(S) \subset f^{\vee} g(S)$ を満たさない f, g が存在する。

◀ 証明 ▶ $T = \{a, b\}$ とする。表 2 に示すフィルタリング関数 f は逐次増加性、逐次減少性、並列減少性を満たし、 g は単調性、逐次増加性、逐次減少性、並列減少性を満たすが、 $S = \{a, b\}$ のとき $f \circ g(S) \subset f^{\wedge} g(S)$ を満たさない。 □

表 2 反例 2

S	$f(S)$	$g(S)$	$f \circ g(S)$	$f^{\vee} g(S)$	$f^{\wedge} g(S)$
ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
$\{a\}$	$\{a\}$	$\{a\}$	$\{a\}$	$\{a\}$	$\{a\}$
$\{b\}$	$\{b\}$	ϕ	ϕ	$\{b\}$	ϕ
$\{a, b\}$	$\{b\}$	$\{a\}$	$\{a\}$	$\{a, b\}$	ϕ

[補題 8] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が分配等価性を満たすならば、任意の $S \subset T$ に対して $f \circ g(S) = f^{\wedge} g(S)$ が成り立つ。

◀ 証明 ▶ 文献 [10] より、

$$\exists X, \forall S, f(S) = S \cap X$$

$$\iff \forall S, \forall T, f(S \cup T) = f(S) \cup f(T)$$

が成立するため、 $X = f(T)$ とすると、任意の $S \subset T$ に対して $f(S) = S \cap X$ とおける。したがって、

$$f(g(S)) = g(S) \cap X \quad (27)$$

$$\begin{aligned} f(S) \cap g(S) &= (S \cap X) \cap g(S) \\ &= (S \cap g(S)) \cap X \\ &= g(S) \cap X \quad (\because g : D) \end{aligned} \quad (28)$$

が成り立つ。ゆえに、(27), (28) より、 $f \circ g(S) = f^{\wedge} g(S)$ が導き出された。 □

[補題 9] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が逐次等価性を満たすならば、任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge} g(S)$ が成り立つ。

◀ 証明 ▶ 2.2 節で示した定義より、フィルタリング関数が逐次等価性を満たすならば、逐次増加性も満たす。したがって、補題 6 より、任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge} g(S)$ が成り立つ。 □

[補題 10] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が逐次等価性、 g が分配等価性あるいは逐次等価性を満たすとき、ある $S \subset T$ に対して $f \circ g(S) \subset f^{\wedge} g(S)$ を満たさない f, g が存在する。

◀ 証明 ▶ $T = \{a, b\}$ とする。表 2 に示すフィルタリング関数 f は逐次等価性を満たし、 g は分配等価性、逐次等価性を満たすが、 $S = \{a, b\}$ のとき $f \circ g(S) \subset f^{\wedge} g(S)$ を満たさない。 □

ここで、フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、 f が逐次減少性あるいは並列減少性、 g が単調性あるいは逐次増加性、逐次減少性、並列減少性を満たすとき、任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge} g(S)$ が成り立つかどうかは、現在のところ明らかになっていない。しかし、 f, g がある条件を満たすとき、以下の補題により包含関係が成立することがいえる。

[補題 11] フィルタリング関数 f, g について、 f が g にフィルタリング合成可能、 f と g がフィルタリング積可能であり、

任意の $S \subset T$ に対して $f \circ g(S) \subset f^{\wedge}g(S)$, かつ f が逐次減少性あるいは並列減少性を満たすならば, 任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge}g(S)$ が成り立つ.

《証明》ある $S \subset T$ に対して $f(g(S)) \not\subset f(S) \cap g(S)$ と仮定すると, ある $x \in T$ に対して,

$$x \in f(S) \cap g(S) \quad (29)$$

$$x \notin f(g(S)) \quad (30)$$

が成立する. また, f と g がフィルタリング積可能であることから,

$$f^{\wedge}g(S) = f^{\wedge}g(f^{\wedge}g(S))$$

$$f(S) \cap g(S) = f(f(S) \cap g(S)) \cap g(f(S) \cap g(S)) \quad (31)$$

が成り立つ. したがって, (29), (31) より,

$$x \in f(f(S) \cap g(S)) \cap g(f(S) \cap g(S)) \quad (32)$$

となるため,

$$x \in f(f(S) \cap g(S)) \quad (33)$$

$$x \in g(f(S) \cap g(S)) \quad (34)$$

が導き出される. ここで, f が逐次減少性を満たす場合も並列減少性を満たす場合も, 次のように同じ式が導き出される.

i) f が逐次減少性を満たすとき

$$\begin{aligned} & f((f(S) \cap g(S)) \cup g(S)) \\ & \supset f((f(S) \cap g(S)) \cup f(g(S))) \end{aligned} \quad (35)$$

となる.

ii) f が並列減少性を満たすとき

$$\begin{aligned} & f((f(S) \cap g(S)) \cup g(S)) \\ & \supset f(f(f(S) \cap g(S)) \cup f(g(S))) \end{aligned} \quad (36)$$

となる. ここで, (31) より, 任意の $y \in f(S) \cap g(S)$ に対して

$$y \in f(f(S) \cap g(S)) \quad (37)$$

かつ

$$y \in g(f(S) \cap g(S)) \quad (38)$$

が成り立つため, (37) より

$$f(S) \cap g(S) \subset f(f(S) \cap g(S)). \quad (39)$$

また, f の減少性より

$$f(S) \cap g(S) \supset f(f(S) \cap g(S)). \quad (40)$$

したがって, (39), (40) より,

$$f(S) \cap g(S) = f(f(S) \cap g(S)) \quad (41)$$

が成立するため, (36), (41) より, (35) が導き出される.

表3 増加性または減少性を満たすフィルタリング関数 f, g に対する $f \circ g$ と $f^{\vee}g$ の関係

$f \setminus g$	M	SI	SD	PD
M	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$
SI	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$
SD	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$
PD	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$	C, $\neg \supset$

表4 等価性を満たすフィルタリング関数 f, g に対する $f \circ g$ と $f^{\vee}g$ の関係

$f \setminus g$	DE	SE, PE
DE	C, $\neg \supset$	C, $\neg \supset$
SE, PE	C, $\neg \supset$	C, $\neg \supset$

すなわち, i), ii) より, f が逐次減少性あるいは並列減少性を満たすならば, (35) が成立する. ゆえに, $g(S) \supset f(S) \cap g(S)$ が自明であることと, 条件 $f(S) \cap g(S) \supset f(g(S))$ が与えられていることから,

$$f(g(S)) \supset f(f(S) \cap g(S)) \quad (42)$$

が導き出される. (42) は (30), (33) に矛盾する. \square

以上の補題より, 表3, 表4に, 合成フィルタリング関数と和フィルタリング関数の関係を示す. 図中の“ \subset ”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, 任意の $S \subset T$ に対して $f \circ g(S) \subset f^{\vee}g(S)$ が成立することを表し, “ $\neg \supset$ ”は, ある $S \subset T$ に対して $f \circ g(S) \not\subset f^{\vee}g(S)$ となることを示す. 同様に, 表5, 表6に, 合成フィルタリング関数と積フィルタリング関数の関係を示す. 図中の“ $=$ ”は, 任意の $S \subset T$ に対して $f \circ g(S) = f^{\wedge}g(S)$ であることを示す. さらに, “ \supset ”は, 任意の $S \subset T$ に対して $f \circ g(S) \supset f^{\wedge}g(S)$ が成立するかどうか直接的には明らかになっていないが, 任意の $S \subset T$ に対して $f \circ g(S) \subset f^{\wedge}g(S)$ という条件を満たすときは成立することを表す. このとき, f, g が満たす条件を考慮すると, 任意の $S \subset T$ に対して $f \circ g(S) = f^{\wedge}g(S)$ が成立することがわかる.

表3, 表4より, 本稿で扱うフィルタリング関数はどのように組合せても, 合成フィルタリング関数と和フィルタリング関数が同値でないことが明らかになった. 一方, 表5, 表6より, フィルタリング関数 f が単調性あるいは分配等価性を満たすならば, フィルタリング関数 g が本稿で扱ういずれの性質を満たす場合も, 合成フィルタリング関数と積フィルタリング関数が同値となることが示された.

5. 記述例

本章では, 複数の構文を用いた記述例を示し, 前章までに明らかにしたフィルタリング関数の性質から, 記述内容を実行するフィルタリングの処理方法について考察する.

図1の記述例は, 「“A_Broadcast”により放送されたデータの中から, コンテンツのジャンルがニュースであり, かつコンテンツに対して放送側があらかじめ設定した有効期限 (“MET-ALIMIT”) の切れていないものが欲しい」というユーザの要求を表す. 文献[13]より, “GENRE = news”を実行するフィ

表5 増加性または減少性を満たすフィルタリング関数 f, g に対する $f \circ g$ と $f \wedge g$ の関係

$f \setminus g$	M	SI	SD	PD
M	=	=	=	=
SI	$\supset, \neg C$	$\supset, \neg C$	$\supset, \neg C$	$\supset, \neg C$
SD	$\neg C, (\supset)$	$\neg C, (\supset)$	$\neg C, (\supset)$	$\neg C, (\supset)$
PD	$\neg C, (\supset)$	$\neg C, (\supset)$	$\neg C, (\supset)$	$\neg C, (\supset)$

表6 等価性を満たすフィルタリング関数 f, g に対する $f \circ g$ と $f \wedge g$ の関係

$f \setminus g$	DE	SE, PE
DE	=	=
SE, PE	$\supset, \neg C$	$\supset, \neg C$

ルタリング (f とする) も, “TILL METALIMIT” を実行するフィルタリング (g とする) も分配等価性を満たすことが既に明らかになっている。すなわち, この記述の意味は, 分配等価性を満たすフィルタリング関数どうしの合成関数により表現できる。したがって, 表6より, この記述の意味は同じ2つのフィルタリング関数の積関数でも表現できるため, 図1の記述の意味と図2の記述の意味が一致する。ゆえに, 図1と, 図2のいずれの記述に対しても, 一貫したフィルタリング結果が得られるため, ユーザはどちらか記述しやすい表現を用いればよい。

また, この要求を実行するフィルタリングでは, f, g の処理を連続的に組合せても並列に組合せても, 等価なフィルタリング結果が得られることを保証できるため, 以下のように処理の効率化が図れる。例えば, 複数の処理端末を用意できる場合, 一つの端末で f と g の2つのフィルタリング処理を連続的に実行するよりも, f のフィルタリング処理と g のフィルタリング処理を別端末に分けて処理することで, 処理を分散できる。しかし, このとき, f のフィルタリング処理も g のフィルタリング処理も, 全ての受信データをフィルタリングする必要があるため, 一方の処理端末の計算能力が低い場合, 大量の受信データをフィルタリングすることは困難である。そこで, 例えば g のフィルタリング処理の後に f のフィルタリング処理を連続的に組合せ ($f \circ g$), 比較的計算能力の高い端末で前処理 g を行うことにより, 処理能力が低い端末が処理すべきデータを絞り込める。一般に, 処理すべきデータが少ないほど計算コストは低い。ゆえに, f と g のフィルタリング処理を並列に組合せるときよりも, 処理能力の低い端末が行うフィルタリング処理 f のコストを軽減できる。 f の処理コストは, g におけるデータの絞り込み効果が高いほど軽減できるため, 放送されるデータの特徴や g の性質に応じて変動する。ここで, 表6より, f と g のフィルタリング処理の順序を交換したもの ($g \circ f$) も, それらを並列に組合せたフィルタリングと同値であるため, 上と同様の処理変換をしても, 一貫したフィルタリング結果が保証できる。

さらに, 図2の “INTERSECT” を “UNION” に書き換えた記述について考える。この記述の意味は, f と g の和フィルタリング関数で表現できる。したがって, 表4より, この記述の意味と図1の記述の意味は一致しない。しかし, f, g の合成関数 $f \circ g$ と和関数 $f \vee g$ の間に包含関係が成立するため, 両手法

```
EXTRACT *
FROM A_Broadcast
WHERE GENRE = news
AND TILL METALIMIT
```

図1 ユーザ要求の記述例1

```
EXTRACT *
FROM A_Broadcast
WHERE GENRE = news

INTERSECT

EXTRACT *
FROM A_Broadcast
WHERE TILL METALIMIT
```

図2 ユーザ要求の記述例2

を並列に組合せるフィルタリング ($f \vee g$) から連続的に組合せるフィルタリング ($f \circ g$ あるいは $g \circ f$) へと変更すると, 変更前に蓄積されていたデータのうち, より重要度の高いデータのみ絞り込んで蓄積できる。逆に, 両手法を連続的に組合せるフィルタリング ($f \circ g$ あるいは $g \circ f$) から並列に組合せるフィルタリング ($f \vee g$) へと変更しても, 変更前に蓄積されていたデータが, 変更後も蓄積されることが保証される。ゆえに, 別のアプリケーションなどで利用していたデータを, 処理方法の変更後も継続して利用できる。ここで, f の結果と g の結果の積を計算するフィルタリング ($f \wedge g$) は, 各フィルタリング処理 f, g の結果を比較し, それぞれの共通データを見つける必要があるため, f の結果と g の結果の和を計算するフィルタリング ($f \vee g$) よりも計算コストが高いといえる。したがって, 図1のユーザ要求を実行するために f, g の手法を並列に組合せる場合, 本来蓄積すべきデータの一貫性のみが保証されていればよいならば, f と g の和フィルタリング関数 ($f \vee g$) で実行することは有効である。

次に, 図3の記述例は, 「“A_Broadcast” により放送されたデータの中から, 最近放送された500個以内のデータと, ここ1時間以内に放送されたデータが欲しい」というユーザの要求を表す。文献[13], [15]より, “best(500, Broadcast_Time, DESC)” を実行するフィルタリング (f とする) は逐次等価性を満たし, “PERIOD 1hour” を実行するフィルタリング (g とする) は分配等価性を満たすことが既に明らかになっている。すなわち, この記述の意味は, 逐次等価性を満たすフィルタリング関数と分配等価性を満たすフィルタリング関数の和関数により表現できる。ゆえに, 表4より, この記述は同じ2つのフィルタリング関数の合成関数で表現できないため, 図3の記述の意味と図4の記述の意味は一致しない。

しかし, f, g の合成関数 $f \circ g$ あるいは $g \circ f$ と, 和関数 $f \vee g$ の間に包含関係が成立するため, 図2の “INTERSECT”

```

EXTRACT *
FROM A_Broadcast
WHERE best(500, Broadcast_Time, DESC)

UNION

EXTRACT *
FROM A_Broadcast
WHERE PERIOD 1hour

```

図3 ユーザ要求の記述例3

```

EXTRACT *
FROM A_Broadcast
WHERE best(500, Broadcast_Time, DESC)
AND PERIOD 1hour

```

図4 ユーザ要求の記述例4

を“UNION”に書き換えた場合と同様、処理端末の計算能力や、一貫性保証の必要性などを考慮して処理変換ができる。ここで、図3の“UNION”を“INTERSECT”に書き換えた記述の意味は積フィルタリング関数 $f \wedge g$ で表現できるが、合成関数との包含関係は f, g の合成順序 $(f \circ g, g \circ f)$ に応じて異なる。まず、 f, g の合成関数 $f \circ g$ と積関数 $f \wedge g$ は、図3の記述を書き換える前と逆の包含関係となる。すなわち、両手法を並列に組合せるフィルタリング ($f \wedge g$) から連続的に組合せるフィルタリング ($f \circ g$) へと変更したとき、変更前に蓄積されていたデータが、変更後も継続して蓄積されることが保証される。一方、 f, g の合成関数 $g \circ f$ と積関数 $f \wedge g$ は同値であるため、 f の後に g を連続的に組合せるならば、 f, g を並列に組合せる場合と等価なフィルタリング結果が得られる。

6. おわりに

本稿では、複数の構文を組合せたフィルタリングSQLに対して、それを実行するフィルタリング手法の組合せ方に関する性質を明らかにした。合成フィルタリング関数と和積フィルタリング関数の関係を示すことで、手法の組合せ方を変更することがフィルタリング結果に与える影響が明らかになった。本稿により、複数の手法を連続的に組合せたフィルタリングと並列に組合せたフィルタリングの間で組合せ方を変更しても、一貫したフィルタリングが保証されるかどうかを明らかにし、状況に応じて組合せ方法を変換できる。

今後の課題として、フィルタリング関数 f が逐次減少性あるいは並列減少性を満たし、フィルタリング関数 g が本稿で扱う増加性あるいは減少性を満たすとき、任意の $S \subset T$ に対して $f \circ g(S) \supset f \wedge g(S)$ が成立するかどうかを明らかにすることが挙げられる。また、本稿では、蓄積したいデータの条件を記述するための構文を扱っているが、ユーザのさらに複雑な要求を記述できるよう、“更新データが到着したら、古いデータはい

らない”といったように、蓄積したくないデータの条件を指定する構文も定義し、その記述内容を実行するフィルタリングの性質を明らかにする予定である。

謝 辞

本研究は、文部科学省振興調整費「情報フィルタリングの数学的基盤の確立」、「モバイル環境向 P2P 型情報共有基盤の確立」、および文部科学省 21 世紀 COE プログラム「ネットワーク共生環境を築く情報技術の創出」、科学研究費補助金（基盤研究 (B)(2)）「大規模な仮想空間システムを構築する放送型サイバースペースに関する研究」（プロジェクト番号:15300033）の研究助成によるものである。ここに記して謝意を表す。

文 献

- [1] 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).
- [2] T. A. H. Bell and A. Moffat: “The design of a high performance information filtering system,” in *Proc. of 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR1996)*, pp. 12–20 (1996).
- [3] 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).
- [4] 西正, 野村敦子: “多チャンネル放送の衝撃,” 中央経済社 (1997).
- [5] Satellite Magazine: <http://www.satemaga.co.jp>.
- [6] 澤井里枝: ミニサーベイ “ブロードバンド時代における情報フィルタリングの動向,” 電子情報通信学会第 13 回データ工学ワークショップ (DEWS2002) 論文集 (2002).
- [7] 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).
- [8] R. Sawai, M. Tsukamoto, Y. H. Loh, T. Terada, and S. Nishio: “Functional properties of information filtering,” in *Proc. of 27th International Conference on Very Large Data Bases (VLDB2001)*, pp. 511–520 (2001).
- [9] 澤井里枝, 塚本昌彦, 寺田努, Loh Yin Hwei, 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-I, vol. J85-D-I, no. 10, pp. 939–950 (2002).
- [10] 澤井里枝, 塚本昌彦, 寺田努, 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について,” 情報処理学会論文誌: データベース, vol.43, no. SIG12 (TOD16), pp. 80–91 (2002).
- [11] 澤井里枝, 塚本昌彦, 寺田努, 西尾章治郎: “合成フィルタリング関数の性質について,” 情報処理学会論文誌: データベース, vol. 44, no. SIG3 (TOD17), pp. 43–53 (2003).
- [12] 澤井里枝, 塚本昌彦, 寺田努, 西尾章治郎: “情報フィルタリングの実行順序に関する関数的性質について,” 情報処理学会論文誌: データベース, vol. 44, no. SIG3 (TOD17), pp. 54–64 (2003).
- [13] 澤井里枝, 塚本昌彦, 寺田努, 西尾章治郎: “情報フィルタリングのためのユーザ要求記述言語 FilteringSQL 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 2003-DBS-131 (I)), pp. 451–458 (2003).
- [14] 澤井里枝, 塚本昌彦, 寺田努, 西尾章治郎: “フィルタリング関数の和積について,” 情報処理学会論文誌: データベース, vol. 44, no. SIG12 (TOD19), pp. 86–97 (2003).
- [15] 澤井里枝, 塚本昌彦, 寺田努, 西尾章治郎: “フィルタリングSQLにおけるランキングについて,” 情報処理学会研究報告 (データベースシステム研究会 2004-DBS-132・放送コンピューティング研究グループ 2004-BCCgr-7 合同研究会), pp. 72–79 (2004).
- [16] 総務省 情報通信統計データベース: <http://www.johotsusintokei.soumu.go.jp/>.