データベースシステム　122－9
(2000. 7. 26)

新鮮度・流行度に基づく蓄積型受信端末の廃棄制御とコンテンツの再構造化

馬　強[†]　田中克己[†]

蓄積メディアの大容量化，低価格化に伴って，クライアント側でのコンテンツの蓄積が可能となり，ユーザが多様な方法で配信コンテンツを再利用することが可能となる．しかし，日々絶えずに配信してくるコンテンツと比べて，蓄積メディアの容量が一定であり，蓄積の限度があるので，ユーザの再利用目的に応じて蓄積コンテンツの選択，廃棄制御および構造化が重要である．放送型情報配信システムでは，配信される情報は時間的に追加もしくは更新され，キーワードなどが予測困難であるので，従来のキーワードベースの手法では，新しい情報や話題ニュースを獲得・蓄積困難である場合がある．本論文では，配信コンテンツの時系列性に基づいてニュース記事の時系列的特徴量（新鮮度・流行度）を定義し，それに基づく蓄積型受信端末でのコンテンツの蓄積・廃棄制御手法を提案する．さらに，大量な蓄積コンテンツを効率よく呈示するためのコンテンツの再構造化手法を提案する．

Disposing and Restructing of Stored Time-Series Articles Based on Freshness and Popularity

MA QIANG[†] and KATSUMI TANAKA[†]

With the storage device is becoming increasing high capacity and low price, it becomes possible that archiving the contents at the user side. There is a new issue that how to manage the stored information at the client side. In this paper, we propose a new method to dispose and restructure the stored time-series articles based on the time-series features(*freshness* and *popularity*).

1. はじめに

放送型情報配信システム[1]では，ユーザが情報を明示的に指定して取り出すのではなく，ユーザによってあらかじめ設定されたプロファイルにしたがって情報を自動的に配信する．ユーザが受動的に情報を受信できるので，大量の情報にアクセスできると共に，情報を獲得するためのユーザの負担が減少される利点がある．

また，蓄積メディアの大容量化によって，TVなど受信端末でも，受信コンテンツなどを大量に蓄積可能となっている．ユーザが配信情報を蓄積して様々な方法で再利用することが可能である．

しかし，受信コンテンツが日々絶えずに増加しているに対して，蓄積メディアの容量が一定であり，蓄積できるコンテンツの量は限度があるので，ユーザの再利用方法に応じた蓄積コンテンツの選択および廃棄制御が必要となる．

大量の配信情報からユーザがほしい情報を見つけて蓄積するために，キーワードベースのユーザプロファイルを用いた情報フィルタリングがよく利用される．しかし，これらの手法は，ニュースなどキーワードが未知な情報を獲得困難である場合があるので，放送型ニュース配信システムのフィルタリング機構に更なる工夫が必要である．

我々は，ニュース記事の時間的に連続して配信されるという特徴を配慮して，配信情報の時系列的特徴量を定義し，ユーザプロファイルと併用したフィルタリング手法を提案している[2),3)]．この手法では，受信コンテンツの時系列的特徴を捕え，過去の配信記事との類似・非類似を計算し，より重要，より新鮮な情報を選択することが可能である．

しかし，この手法は，リアルタイム処理を対象としたもので，コンテンツの蓄積を想定していない．蓄積コンテンツは選択された時点では，情報の価値が高いが，蓄積メディアのなかでの価値は必ずしも高いわけではない．よって，蓄積されたコンテンツの特徴量を再計算するなどさらなる工夫が必要であると考えられる．

FIFOなど従来の廃棄制御手法では，配信コンテンツの連続性と関連性を考慮せず，コンテンツの配信時

† 神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University

間のみでコンテンツの廃棄を行っているので，再利用を目的とする配信コンテンツの蓄積には不十分である．

- FIFO では，記事の内容を考慮せず時間のみで廃棄処理を行うので，蓄積された情報は重複である場合がある．
- 蓄積されるコンテンツはある期間中の配信記事の全集合であり，コンテンツの構成は固定である．ユーザが自分の再利用方法に応じてコンテンツの構成をカスタマイズするのは困難である場合がある．

本論文では，配信コンテンツの時系列性を考慮して，蓄積型受信端末におけるコンテンツの蓄積，廃棄制御手法を提案する．この手法では，過去に蓄積された記事との類似・非類似に基づいて蓄積する新しい配信記事を選択する．同時に，より新しく蓄積された記事との類似・非類似に基づいて廃棄記事を選択する．ユーザは自分の再利用目的に応じて，蓄積・廃棄の基準を選択して蓄積コンテンツの構成をコントロールすることが可能である．

さらに，本論文では，蓄積されたコンテンツの構成に応じて，ユーザのコンテンツの再利用を支援するためのコンテンツの再構成手法を提案する．

本論文で提案しているコンテンツの蓄積・廃棄制御および再構造化手法は次のような特徴がある：

- **時系列的特徴量に基づくコンテンツの蓄積・廃棄を行う．**
  放送型情報配信システムでは，新しい配信情報のキーワードが未知であるため，キーワードのみでは獲得困難な情報がある．本論文で提案している手法は，蓄積記事の類似・非類似を計算することによって，より新しい情報や話題ニュースを獲得・蓄積可能である．また，時間のみではなく，記事の時系列的特徴量に基づいて廃棄記事を選択しているので，より幅広く情報を蓄積することが可能である．
- **コンテンツの蓄積，廃棄の基準が選択可能であり，再利用の目的に応じて蓄積コンテンツの構成をコントロールすることが可能である．**
  FIFO などコンテンツの廃棄手法は，単に時間を基準にコンテンツを廃棄しているので，ユーザのカスタマイズが困難である場合がある．その上，蓄積されたコンテンツは短い期間中に限られ，単一的なコンテンツの構成になってしまう可能性がある．本論文で提案している手法では，ユーザが蓄積・廃棄の基準を自由に設定できるので，いろんな構成でコンテンツを蓄積することが可能である．例えば，あらゆるイベントのオーバービュー，メジャーイベント，スクープ記事，結果記事などの構成でコンテンツを蓄積することが可能である．
- **蓄積コンテンツの構成に応じて再構造化を行う．**
  本論文で提案している手法では，ユーザは自分のニーズに応じて蓄積コンテンツの構成を選択可能である．このような異なる構成の蓄積コンテンツの再利用を支援するために，コンテンツの構成に応じて蓄積コンテンツを再構造化することが可能である．

以下，本論文の構成を示す：2 章では，本論文で提案している蓄積型受信システムについて述べる．3 章では，時系列文書の時系列的特徴量: 新鮮度と流行度を述べる．4 章では，蓄積型受信端末における時系列文書の蓄積と廃棄制御および再構造化について述べる．5 章では，関連研究について述べる．6 章では，まとめと今後の課題について述べる．

2. 蓄積型受信システム

図 1 で示しているように，本研究で想定している蓄積型受信端末は，蓄積フィルタ，蓄積メディア，再構造コントローラ，廃棄フィルタから構成される．蓄積フィルタは，蓄積すべきコンテンツを選択する．再構造コントローラは，蓄積されたコンテンツの再構造化を行う．廃棄フィルタは必要に応じてコンテンツの廃棄を行う．

蓄　　積

蓄積フィルタは，過去に蓄積された記事との類似・非類似に基づいて記事を選択して蓄積する．

ユーザは自分が重視する情報のタイプに基づいて，類似と非類似のいずれかを記事の選択基準にすることが可能である．例えば，新しい情報を重視するなら非類似を，話題を重視するなら類似を蓄積フィルタの選択基準にすることが可能である．

廃　　棄

廃棄フィルタは，蓄積フィルタと同様に，蓄積メディアでの記事の類似・非類似に基づいて廃棄記事を選択する．ただし，記事の類似・非類似の比較対象は，蓄積フィルタではその記事より以前に蓄積された記事であり，廃棄フィルタではその記事より新しく蓄積されている記事である．

蓄積フィルタと廃棄フィルタの相互作用があり，ユーザは自分の再利用目的に応じて，蓄積フィルタと廃棄フィルタを組み合わせていろんな構成でコンテンツを蓄積することが可能である．

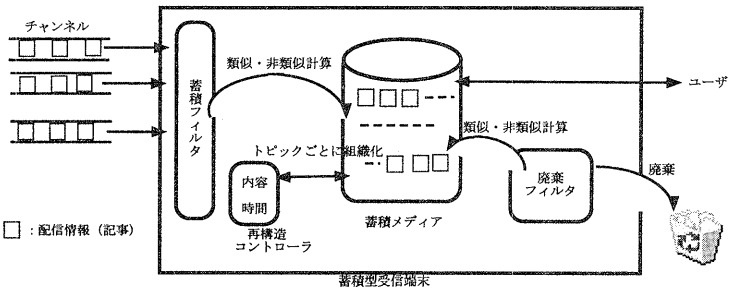

図 1　蓄積型受信システム

再構造化

ユーザの再利用目的を反映するコンテンツの構成に応じて，再構造コントローラは，コンテンツの類似度に基づいて蓄積記事をトピックごとに組織化し，配信時間と類似度に基づいてトピックをサブトピックにわける．さらに，トピックあるいはサブトピックの代表記事を選択する．

3. 時系列的特徴量

時間の経過と共に配信されるニュース記事の価値は，その記事自身の内容だけでなく，過去の配信履歴，配信頻度など要素にも依存する．本論文では，ニュース記事の時系列的特徴量として，配信履歴に基づく新鮮度と流行度を定義する．

3.1 新鮮度

新鮮度の計算とは，一定の時空間において，オブジェクト (記事) を既存のオブジェクト集合と比べて，そのオブジェクトの新しさに対する評価を行うことである．

図 2 に示すように，ある遡及範囲（記事の集合)Ω に対する記事 a の新鮮度を，(1)Ω 内の類似記事の数に基づく新鮮度 ($fresh_{num}(a)$)，(2)Ω 内の類似記事との内容距離に基づく新鮮度 ($fresh_{cd}(a,\omega)$)，(3) 記事の密度に基づく新鮮度 ($fresh_{de}(a)$)，(4) 類似記事との時間距離に基づく新鮮度 ($fresh_{td}(a,\omega)$) によって定義する．これらの新鮮度は，ユーザの選択によって，独立に用いることができる．次に，これらの新鮮度を混合して，それぞれに重みを付けた統合新鮮度 $fresh_{\Omega}(a)$ を次式のように定義する．

$$fresh_{\Omega}(a) = \alpha * fresh_{num}(a) \quad (1)$$
$$+\beta * fresh_{cd}(a,\omega) \quad (2)$$
$$+\gamma * fresh_{de}(a) \quad (3)$$
$$+\sigma * fresh_{td}(a,\omega) \quad (4)$$

ただし，ω は Ω における a の類似記事の集合である．$\alpha,\beta,\gamma,\sigma$ は重み付け定数である．

ここで，記事 a の新鮮度を計算するための遡及範囲 Ω の記事数を n とする．Ω における a の類似記事の集合 ω の記事数を m とする．

(1) 類似記事の数による新鮮度

過去の配信記事の集合 Ω の中で，a と類似している記事の数が少なければ，その記事の内容が新しく，新鮮度が高いと考えられる．そこで，類似記事数に基づく新鮮度を次のようにする．

$$fresh_{num}(a) = \frac{1}{log_2(2+m)} \quad (5)$$

(2) 内容距離による新鮮度

各記事 a は，k 次元特徴ベクトル $v(a) = (w_1, w_2, ..., w_k)$ を持つとする．ただし，w_i はキーワード k_i の重みである．

記事 a と b の違いを表す内容距離 $dis(a,b)$ を次のように定義する：

$$dis(a,b) = |v(a) - v(b)| \quad (6)$$

つまり，$v(a), v(b)$ の差を記事 a, b の内容距離とする．

記事 a が前報記事（過去の類似記事）と比べてどの程度新しい情報を追加しているかを，記事 a と前報記事との内容距離で表すことができる．前報の記事とは異なる情報が多く追加される場合（内容距離が大きい場合），a の新鮮度が高いと考えられる．つまり，記事 a の類似記事集合 $\omega(\omega \subseteq \Omega)$ 内の記事との平均内容距離が大きいほど a の新鮮度が高いと考えられる．よって，記事 a の類似記事集合 ω に対する新鮮度を

$$fresh_{cd}(a,\omega) = \log(\frac{1}{m}\sum_{i=1,b_i\in\omega}^{m} dis(a,b_i)) \quad (7)$$

と定義する．

(3) 類似記事の密度による新鮮度

a の類似記事の Ω における密度は $d = m/n$ であ

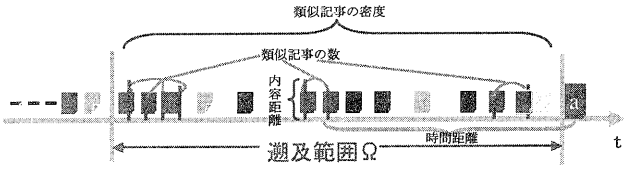

図 2　新鮮度

る．d を a の予想出現確率と考えると，a の情報量は $log_2\frac{1}{d}$ となる．情報量が大きいと，新鮮度が高いと考えられるので，記事 a の類似記事密度に基づく新鮮度を

$$fresh_{de}(a) = log_2\frac{n}{m} \tag{8}$$

と定義する．

(4) 時間距離による新鮮度

過去の類似記事との時間距離も記事 a の新鮮度に影響を与える．類似記事と時間的に離れれば離れるほど a の新鮮度が高くなると考えられる．直観的には，例えば，図 **3** では，$(a.1)$ の a の新鮮度は $(a.2)$ の a より大きい，$(b.2)$ の a の新鮮度は $(b.1)$ の a より小さい．

a と類似記事集合 ω の平均時間距離が大きければ大きいほど a の新鮮度が大きくなると考え，記事 a の類似記事集合 ω との時間距離に基づく新鮮度を次式のように定義する．

$$fresh_{td}(a,\omega) = \log(\frac{1}{m}\sum_{i=1,b_i\in\omega}^{m}(t(a)-t(b_i))) \tag{9}$$

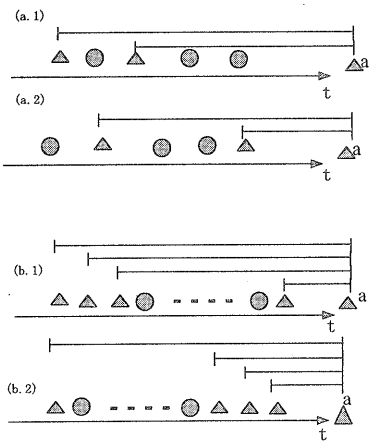

図 3　時間距離の例

ただし，$t(a)$ は記事 a の配信時間を表わす．

3.2　流　行　度

最近の配信記事の中で，新しい配信記事と類似しているものが多数存在する場合，この記事はあるトピックの続報であり，現在話題になっているのであると考えられる．このような記事は，新鮮度が低いが，流行度が高く，ニュース価値が高いと考えられる．新鮮度と流行度の計算では，共に過去の配信記事との比較を行っているので，流行度と新鮮度は基本的にお互いに依存する．ただ，遡及範囲の選択方式などが異なると依存度が変化する場合があり，本論文では，新鮮度と流行度を別個のものとして扱っている．

本論文では，記事 a の流行度 $pop(a)$ は類似記事の密度と時間距離を用いて定義する (図 **4**)．つまり，最近の配信記事の中で，a との類似記事が多ければ多いほど，a の流行度が高くなると考え，次式のように定義する．

$$pop(a) = e^{\lambda_1 k} + e^{-\lambda_2 t_d} \tag{10}$$

ただし，$\lambda_1(>0), \lambda_2(>0)$ は重みづけ定数である．k は類似記事の密度 m/n，t_d は，

$$t_d = \frac{1}{m}\sum_{i=1,b_i\in\omega}^{m}(t(a)-t(b_i)) \tag{11}$$

である．

4.　コンテンツの蓄積，廃棄制御と再構造化

メディアの大容量化と低価格化に伴って，ユーザ側では，配信コンテンツの蓄積が可能となる．しかし，配信コンテンツが絶えずに増加するが，メディアの容量が固定であるので，蓄積の限度がある．従って，メディアにつねにユーザにとって価値が高い情報を保つことが重要であり，ユーザの再利用目的に応じてコンテンツの蓄積，廃棄および再構造化を行うことが重要である．

4.1　蓄積と廃棄制御

4.1.1　蓄　　積

蓄積フィルタは，新しく配信してきた記事 a を，蓄

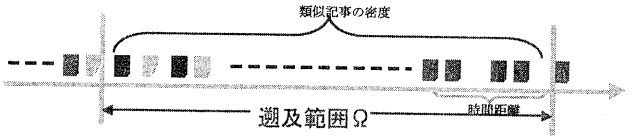

図 4　流行度

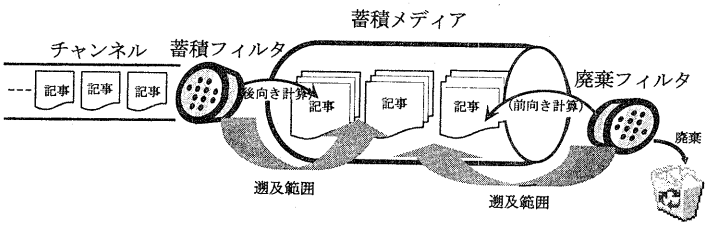

図 5　蓄積と廃棄制御モデル

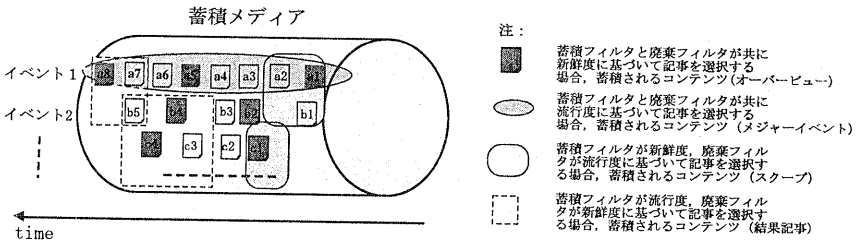

図 6　フィルタの組合せと蓄積コンテンツの構成

積された記事，つまり過去の配信記事との類似/非類似を計算して，a は内容が新しいか，あるいは最近の話題であるかを評価する (図 5)．つまり，a より古い記事と比べて，a の新鮮度・流行度を計算し，a を蓄積するかどうかを判断する．

蓄積フィルタでは，流行度重視と新鮮度重視の二つのオプションがある．ユーザが自分のニーズに応じてどちらかを選択することが可能である．たとえば，話題情報を重視するのであれば，蓄積フィルタは過去の配信記事と類似度が高い記事，つまり流行度が高い記事を選択して蓄積する．一方，新しい情報重視であれば，蓄積フィルタは過去の記事との非類似が高い記事，つまり新鮮度が高い記事を選択して蓄積する．

4.1.2 廃　　棄

廃棄フィルタは，蓄積メディアから古い順に記事を廃棄候補 a_c とし，a_c より新しく蓄積された記事との類似・非類似に基づいて a_c の新鮮度・流行度を計算して，a_c を廃棄するかどうかを判断する (図 5)．つまり，新しい配信情報と比べて計算された新鮮度・流行度に基づいて廃棄記事を選択する．

蓄積フィルタと同様に，ユーザは新鮮度を重視するかあるいは流行度を重視するかを選択可能である．新しい情報を重視するのであれば，類似記事が少ない，つまり新鮮度が高い記事が保留される．新鮮度低い記事が廃棄される．話題情報重視であれば，類似記事が多い，つまり流行度が高い記事が保留される．流行度が低い記事が廃棄される．

4.1.3 蓄積と廃棄制御の相互作用

蓄積フィルタは記事 a の新鮮度・流行度を，a より先に蓄積された記事と比較して計算し，記事を蓄積するかどうかを決定する．一方，廃棄フィルタは a より新しい記事と比べて計算された新鮮度・流行度に基づ

いて，a は廃棄すべき記事であるかどうかを判断する．二つのフィルタは，共に新鮮度・流行度を計算しているが，遡及範囲が時間的に逆であるので，二つのフィルタは相互作用し，蓄積されるコンテンツの構成に影響を与える (図 6)．

(a) 新鮮度と新鮮度

蓄積フィルタが新鮮度が高い記事を選択して蓄積し，廃棄フィルタが新鮮度が低い記事を選択して廃棄処理を行う場合，廃棄フィルタと蓄積フィルタが共に，新鮮度の高い記事を蓄積フィルタに保存する働きがあるので，蓄積期間中に発生したあらゆるイベントの記事は少なくとも 1 つ以上が蓄積メディアに保存されることが予想される．

したがって，蓄積されるコンテンツは蓄積期間中のあらゆるイベントのオーバービューであると考えられる．

(b) 流行度と流行度

蓄積フィルタが流行度が高い記事を選択して蓄積し，廃棄フィルタが流行度が低い記事を選択して廃棄処理を行う場合，廃棄フィルタと蓄積フィルタは共に，流行度が高い記事を保存する働きがある．類似記事が少ない記事が徐々に廃棄されるので，蓄積されたコンテンツは期間中に支配的なイベント，つまりメジャーイベントの記事集合となることが予測される．

(c) 新鮮度と流行度

蓄積フィルタが新鮮度が高い記事を選択して蓄積し，廃棄フィルタが流行度が低い記事を選択して廃棄処理を行う場合，蓄積されるコンテンツは蓄積期間中のあらゆるイベントのスクープ記事（第一報）の集合である可能性が高い．

廃棄フィルタでの流行度の計算は，より新しく蓄積された記事を対象にしているので，最初に蓄積された記事は流行度が高く計算され，保留される可能性が高い．一方，新しく配信してきた記事が廃棄される可能性が高い．したがって，イベントの最初記事，スクープ記事が保存される可能性が高いと予測される．

(d) 流行度と新鮮度

蓄積フィルタが流行度が高い記事を選択して蓄積し，廃棄フィルタが新鮮度が低い記事を選択して廃棄処理を行う場合，蓄積されるコンテンツは蓄積期間中のあらゆるイベントの結果 (まとめ) 記事の集合となることが予測される．

過去に蓄積された記事との類似度が高い記事，つまり流行度が高い記事が蓄積フィルタによって選択され蓄積メディアに保存されるが，廃棄フィルタは，自分より新しく蓄積された記事と比べて，新鮮度が低い記事をメディアから削除する．この場合，イベントの最初記事は新鮮度が低く，廃棄される可能性が高い．一方，イベントの結果記事が新鮮度が高く，保留される可能性が高い．つまり，イベントの結果記事が保留される可能性が高いと考えられる．

4.2 蓄積コンテンツの再構造化

大量に蓄積された配信情報を，効率よく再利用するためには，コンテンツの再構造化が重要である．

本論文では，蓄積コンテンツをトピックごとに組織する：まず，蓄積された記事を記事間の類似度に基づいてトピックに分類する．そして，類似度と配信時間に基づいてトピック内の記事をサブトピックに分ける．さらに，各々のサブトピックには，最も特徴ある記事をそのサブトピックの代表記事とする (図 7)．

記事 a があるトピックに属するすることは，a の内容がトピックの記事内容に近いを意味する．サブトピックに属するのは，そのサブトピックの記事内容に近い，しかも時間的に近いのを意味する．

ここでは，記事の類似度によって記事をトピックに分け，さらに，類似度と配信時間に基づいてトピックの記事をサブトピックに分ける．

類 似 度

候補となるトピックを t_c とする．記事 a の tf・idf ベクトルと t_c の tf・idf ベクトル (t_c に属するすべての記事の平均 tf・idf ベクトル) のコサイン相関値を，a と t_c の類似度 $sim(a, t_c)$ とする．$sim(a, t_c)$ が閾値 θ_1 より大きければ，a は t_c のメンバーとなる．

時 間 相 関

t_c のサブトピック t_{sub} の i 番目記事の配信時間 t_i と i の相関関数 $te(i)$ に基づいて，ラスト記事 (m 番目の記事，m は t_{sub} のサイズである．) の次の記事 ($m+1$ 番目の記事) の予測配信時間 t_{m+1} を計算する．記事 a の実際配信時間 t_a と t_{m+1} との比率は，a の t_{sub} との時間相関 ($tr(a, t_{sub})$) である．

$$tr(a, t_{sub}) = \frac{t_a}{t_{m+1}} \tag{12}$$

相関関数 $te(i)$ は実際のデータに依存するが，一番シンプルなのは，t_{sub} 内で隣接している二つの記事 (a_i, a_{i+1}) の時間間隔の平均 ti_{avg} を用いた線形関数である：

$$te(i) \quad = t_1 + i * ti_{avg} \tag{13}$$

$$ti_{avg} = \frac{1}{m-1} \sum_{i=1}^{m-1} (t_{i+1} - t_i) \tag{14}$$

ただし，t_i は t_{sub} の i 番目の記事の配信時間である．

また，a と t_{sub} の類似度を $sim(a, t_{sub})$ とする．

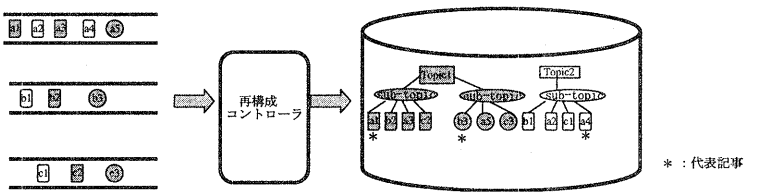

図 7　コンテンツの再構造化

$sim(a, t_{sub}) \geq \theta_2$, $tr(a, t_{sub}) \leq \theta_3$ であれば，a は t_{sub} の $m+1$ 番記事となる．ただし，θ_2,θ_3 は閾値である ($\theta_2 \geq \theta_1$)．

コンテンツの構成に応じた再構造化

ユーザが自分のニーズに応じて，蓄積フィルタと廃棄フィルタの選択基準を選択し，蓄積コンテンツの構成をコントロールすることが可能である．したがって，コンテンツの構成はユーザの特定の再利用の目的を反映していると考えられる．このようなユーザの異なるコンテンツの再利用を支援するために，ユーザの目的を反映するコンテンツの構成に応じてトピック (サブトピック) 分けと代表記事を選択することが重要であると考えられる．

(1) オーバービュー

蓄積フィルタと廃棄フィルタの選択基準が共に新鮮度に設定された場合，蓄積されたコンテンツはあらゆるイベントのオーバービューとなる可能性が高い．ユーザはその期間で発生したあらゆるイベントの概要情報を獲得するのが目的であると考えられる．

この場合，蓄積された記事の間の類似度が低いと考えられるので，閾値 θ_1, θ_2 をより小さく，θ_3 をより大きく設定する．また，代表記事はそのトピック (サブトピック) で内容が最も詳しい記事とする．つまり，トピック（サブトピック）内の他の記事との類似度の平均が一番高い記事は代表記事である．

(2) メジャーイベント

ユーザは，ある期間中の支配的なイベント (メジャーイベント) の情報を獲得する目的である場合，蓄積フィルタと廃棄フィルタの選択基準を共に流行度に設定する．

この場合，蓄積された記事の間の類似度が高いと予測され，イベントの流れに応じてサブトピックの切り分けが重要であると考えられるので，閾値 θ_1, θ_2 をより大きく，θ_3 をより小さく設定する．また，代表記事は各々のサブトピックに存在する．

(3) スクープ記事

ユーザは蓄積フィルタを新鮮度計算，廃棄フィルタを流行度計算に設定して，イベントのスクープ記事を獲得する場合，蓄積されたコンテンツはスクープ記事の集合である可能性が高い．

この場合，蓄積記事の間の類似度が低いと考えられるので，θ_1, θ_2 をより小さく，θ_3 をより大きく設定する．また，代表記事はそれぞれのトピックの最初記事とする．

(4) 結果記事

蓄積フィルタが流行度計算，廃棄フィルタが新鮮度計算に設定され，イベントの結果記事を獲得する場合，蓄積されたコンテンツは結果記事の集合である可能性が高い．

この場合，蓄積記事の間の類似度が低いと考えられる．この場合，閾値 θ_1, θ_2 をより小さく，θ_3 をより大きく設定する．代表記事はそれぞれのトピックの最も新しく蓄積された記事とする．

5. 関連研究

EntryPoint[4] は，ニュース型情報配信システム Pointcast[5][6] の新しいバージョンである．XML の対応などの新しい機能の追加やユーザインターフェースの改善などが行われている．EntryPoint では，一定期間 (3 日間位) の配信記事の蓄積を行っているが，ユーザの再利用を考慮していなく，基本的には FIFO を利用している．

Massachusetts 大学の Allan[7]，Carnegie Mellon University の Yang ら [8] は TDT(Topic Detection and Tracking) に関する研究を行っている．彼らは，既存の記事またはオンラインニュースから新しい話題 (topic) の検知と追跡について研究を行っている．彼らは記事のクラスタリングやクエリーなど手法を用いて，話題の切り分けを行っているが，本論文では，新しく配信された記事を，過去の配信記事との類似・非類似を計算してその記事のニュース価値（新鮮度・流行度など) を評価し，ユーザの再利用の目的に基づいてコンテンツの再構造化を行っている．

宗像ら [9] は，周期的に発生するデータ系列から，

データの鮮度と同期度を基づいてデータの組み合わせを選択する手法を提案している．彼らは，データの鮮度を，得られたデータの中で一番古いデータの時刻から現在の時刻までの経過時間として定義している．データの同期度は，得られたデータの中で，一番新しいデータの時刻と一番古いデータの時刻の差として定義されている．宗像らの鮮度は，固定的な周期ごとに発生するデータを対象としているが，ニュースの場合，記事は周期的に更新されるとは限らない．

カンらは，デジタル放送システムでは情報フィルタリングはリアルタイム性が要求される点に着目し，距離の近似計算によるフィルタリング手法を提案している[10]．カンらの手法では，基本的にはユーザプロファイルを利用しているので，ニュースなどキーワード未知の情報を獲得困難な場合がある．また，ユーザ側でコンテンツを蓄積して再利用するのを考慮していない点は本研究と異なる．

6. おわりに

蓄積メディアの大容量化，低価格化に伴い，クライアント側でのコンテンツの蓄積が可能となっている．ユーザが様々な方法で配信コンテンツを再利用することが考えられる．しかし，日々絶えずに配信してくるコンテンツと比べて，蓄積メディアの容量が一定であり，蓄積の限度があるので，コンテンツの再利用の目的に応じて蓄積コンテンツの選択，廃棄および再構造化を行うことが重要である．

本論文では，配信コンテンツの時系列性に基づいてニュース記事の時系列的特徴量（新鮮度・流行度）を定義し，それに基づく蓄積型受信端末でのコンテンツの蓄積・廃棄制御手法を提案している．ユーザが自分の再利用目的に応じて，蓄積フィルタと廃棄フィルタを調整して蓄積コンテンツの構成をコントロール可能である．

さらに，本論文では，ユーザのコンテンツの再利用を支援するために，蓄積コンテンツの構成に応じた蓄積コンテンツの再構造化手法を提案している．

今後，シミュレーションを行い，適合率と再現率の二つの評価基準を用いて提案している手法の効果を評価する予定である．また，蓄積されたコンテンツの再構成および呈示方式について研究を行う予定である．

謝辞 本研究の一部は，文部省科学研究費「分散型ハイパーメディアからの構造発見とアクセス管理」（課題番号は「12680416」）の援助を受けています．また，本研究の一部は，日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」（プロジェクト番号 JSPS-RFTF97P00501）によっています．ここに記して謝意を表すものとします．

参考文献

1) 角谷和俊, 宮部義幸: 放送型情報配信のためのモデルとシステム, 情報処理学会論文誌：データベース Vol.40 No.SIG 8(TOD4), pp. 141–157 (1999).
2) 馬強, 角谷和俊, 田中克己: 放送型情報配信システムのための時系列性を考慮した情報フィルタリング, 情報処理学会論文誌：データベース (TOD7, to appear) (2000).
3) Ma, Q., Kondo, H., Sumiya, K. and Tanaka, K.: Virtual TV Channel: Filtering, Merging and Presenting Internet Broadcasting Channels, *Proceedings of ACM Digital Library Workshop On Organizing Web Space(WOWS)* (1999).
4) EntryPoint: http://www.entrypoint.com (2000).
5) PointCast: http://www.pointcast.com (1999).
6) Ramakrishnan, S. and Dayal, V.: The PointCast Network, *Proc. of ACM SIGMOD '98*, p. 520 (1998).
7) James Allan, Ron Papka, V. L.: On-line New Event Dectection and Tracking, *Proceedings of SIGIR'98*, pp. 37–45 (1998).
8) Yiming Yang, Tom Pierce, J. C.: A Study on Retrospective and On-Line Event Detection, *Proceedings of SIGIR'98*, pp. 28–36 (1998).
9) 宗像浩一, 吉川正俊, 植村俊亮: 鮮度と同期度に基づく周期データの選択方式, アドバンスト・データベース・シンポジウム'99(ADBS'99), pp.141–150 (1999).
10) カンギョウヒ, 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ワークショップ論文集 (DEWS'2000) (2000).