

異種領域ドキュメント群を対象にしたコンテキストの動的計量を伴う選択的情報配信機構の実現

高松 耕太[†] 倉林 修一[†]
佐々木 史織[†] 清木 康^{††}

本稿では、ユーザの指定したコンテキストに応じて、様々な分野のニュースを自動的に、かつ意味解釈を伴う情報選択によって配信する、意味的情報ルーティングエンジンを示す。ユーザの情報消費行動の多様性に対応するためには、意味の数学モデルなどの多様なコンテキストに応じた意味的な情報の評価を行うベクトル空間モデルを情報配信システムに導入することが肝要である。しかし、意味解釈の結果としての相関量の値は、コンテキストに応じてその値域を大きく変えるため、日々配信される動的な情報資源を対象にしたとき、十分に有意な相関量の閾値を自動的に判定することは容易ではない。本方式は、コンテキスト依存の閾値を自動設定可能な、高度に自律的な選択的情報配信機構を実現する。これは、二つの異なる評価基準によって、コンテキストに対して有意な相関量の閾値を自動的に設定する。さらに、複数の既存のベクトル空間群を、各分野間の共通概念を用いて統合的に利用可能とする問い合わせ解釈機構を示す。これにより、単一のニュースを、意味の解釈を伴い多角的に計量することが可能になる。本稿では、評価実験により、本方式による選択的情報配信の有効性を示す。

A Context-Based Pub/Sub System for Multi-Disciplinary News Documents

In this paper, we present a context-based Pub/Sub system which enables a publication system to notify a user of incoming news articles of various domains according to the user's context automatically. To respond to diversification in informational consumption behavior, it is important to incorporate a context-dependent vector space model into Pub/Sub systems. However, there is at least one problem that the correlation values as a result of semantic computation vary depending on each domain of target documents. Therefore, the system needs some other automatic measures to decide what document should be notified, when it treats dynamic information resources such as daily-published-news. We present a method to determine the context-dependent threshold automatically and adaptively by combining negative and positive sample documents. We also present an integrated query interpretation system with multiple vector spaces which reflect cross-cutting knowledge. We perform evaluation experiments and demonstrate the effectiveness our methods.

1. 概要

現在、一次情報としての最新のニュースをインターネットから電子的に取得することが可能である。特に、IT や経済など、高度に専門的なニュースはインターネット上の Web サイトが最初の情報発信源となることが一般的になってきている。しかし、Web は既存のテレビメディアのような情報集約型メディアと異なり、情報の巡回・分析・編集は、一般ユーザにはコストが高

く、現状では更新情報の概要を作成する専門のニュースサイトという人的資源が必要となる。ニュース情報の爆発的な増加に伴い、インターネットをニュース資源として利用することは徐々に困難になりつつある。そこで、このようなニュース情報資源をユーザの興味に応じて配信することが出来れば、ユーザの情報獲得を強力に支援することが出来る。新しい情報の発生を配信に短絡せず、ユーザの興味に応じた情報の選択的配信 (Selective Dissemination of Information, 以下 SDI と呼ぶ) を実現するシステムの概要を図 1 に示す。

本稿では、既存の異種の意味空間を統合することで、様々な分野のニュースを自動的に配信する、意味的情

[†] 慶應義塾大学大学院政策・メディア研究科
^{††} 慶應義塾大学環境情報学部

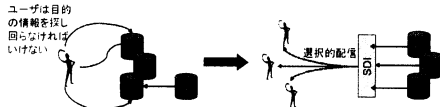


図1 選択的情報配信の概要

報ルーティングエンジン (*S*EMantic Routing Engine for News: *SEREN*) を示す。SEREN は、動的集合をターゲットに、コンテキスト依存の検索結果の妥当性を評価する能力を有する、高度に自律的な選択的情報配信機構の実現方式を示す。本方式は、動的に絶え間なく変化し続ける情報資源を配信する情報配信技術と、ユーザから発行されるコンテキストに応じた意味的な解釈を行う情報検索技術である意味的連想検索^{8)~10),20)}を統合することで、情報資源のダイナミズムとコンテキスト依存の情報獲得のダイナミズムの高次の統合を行う、*Context-Based Pub/Sub* システムの実現を目指す。本方式は主に二つの特徴を有する。

- 問い合わせコンテキストに適した閾値を自動的に設定し、コンテキスト依存の相関量を用いて、自動的に適合ドキュメントと非適合ドキュメントの分類を可能にする、意味的情報配信機構を有する。本稿では、コンテキストに適合するドキュメントと適合しないドキュメントの境界を指定する二つの閾値、**適合ボトム閾値**、**非適合トップ閾値**を自動的に設定する方式を示す。本稿では、これを、**二層閾値モデル**と呼ぶ。
- ジャンル別の複数の意味空間を用いて単一のニュースを多角的に計量する評価機構を示す。本方式では、分野別に構築された既存のベクトル空間のマトリクスを対象として、分野間の共通概念を用いて、意味の解釈を伴った複数のベクトル空間の統合的な利用方法を示す。本方式は、一つのサブスクリプション（購読問い合わせ）から、複数の意味空間への問い合わせへと変換する**購読モデレータ**と、一つのドキュメントを複数の専門分野意味空間へ写像する**発行モデレータ**から成る。これにより、多種多様な情報資源の関連性を動的に評価することが可能となり、動的集合を対象にしたコンテキスト依存の選択的情報配信を実現する。

意味の数学モデルは、検索語ベクトルに対応する動的なコンテキストに応じて、意味的に近い情報の検索を実現している。本方式と関連がある研究として、意味の数学モデルの検索結果を、RDBなどのパターンマッチング処理と組み合わせることで、意味モデルの検索結果をフィルタリングする方式²¹⁾が挙げられる。こ

の方式では、意味の数学モデルから得られた順序集合を、順序を考慮しないリレーショナルモデルに写像してしまうため、意味モデルによる検索精度を大きく損なう可能性があった。しかし、相関量の値はコンテキストに応じてその値域を大きく変えるため、コンテキストとの十分な相関を保証する閾値を自動的に設定することは容易ではない。日々配信される動的な情報資源を対象にしたとき、コンテキスト依存相関量を有する検索結果に対して自動的に閾値を設定する機能の実現が必要である。本方式は、二層閾値により、結果集合の順序を反映した情報選択を行うため、意味モデルの検索結果の精度を損なうことなく、フィルタリング処理を実現する。

関連する研究として、個別のニュースサイトの情報を配信することを目的とした、メタデータ配信技術^{11),14)}や連続的問い合わせ技術^{7),12),13)}が挙げられる。また、情報フィルタリングシステム¹⁸⁾は、ユーザのプロファイルに応じて、情報の選択を自動的に行うシステムである。文献¹⁹⁾では、ユーザのプロファイルと関係度の低い非適合プロファイルを、適合プロファイルと併用することで、高い情報フィルタリング能力を実現する方式を示している。Pub/Sub システム^{2)~6),11)}は、情報の提供者と消費者を関係が疎結合に保ったまま、メッセージを配信する、代表的なSDI機構である。これらのPub/Sub システム群は、メッセージとサブスクリプションとの関連性の計量において、統一的な評価基準を用いているため、大量のメッセージを処理する際に最適化を容易に行うことが可能であり、スケーラビリティに優れる。しかしながら、動的に絶え間なく変化するユーザのコンテキストに依存して、選択的情報配信を実現することは困難である。

2. 意味の数学モデルおよび意味的連想検索方式の概要

情報資源の意味的な関連性を計量する機構として、言葉と言葉の相関量を計量する意味の数学モデルが提案されている^{8)~10),20)}。意味の数学モデルを用いた意味的連想検索方式では、検索語ベクトルと検索対象データをメタデータ空間と呼ばれる直交空間上に写像する。以下、その概要を説明する。

メタデータ空間 *MDS* の設定としてベクトル表現された検索対象メディアデータをマッピングするための正規直交空間 (以下、メタデータ空間 *MDS*) を設定する。 m 個の基本データに関する基本的情報が n 個の特徴として表現された $m \times n$ のマトリクス M を用意する。このマトリクス M を固有値分解し、正規

直交空間 MDS を生成する。メディアデータ P には、メタデータとして t 個の基本データ (w_1, w_2, \dots, w_t) が付与されていることを前提としている。また各基本データ w_i は、ベクトル表現された特徴 (f_1, f_2, \dots, f_n) を持っている。各メディアデータは、メタデータとして付与されている t 個の基本データが合成され、ベクトル表現された後、メタデータ空間 MDS へ写像される。

コンテキスト c はメタデータ空間 MDS から固有部分空間への射影である。このとき $p(x; c)$ はコンテキスト c を与えられたときのメディアデータ x のノルムの値である。

3. 二層閾値による選択の情報配信方式

本節では、意味的情報ルーティングシステム SEREN の要素技術について示す。SEREN はコンテキストに応じ、配信の可否を決定する二層閾値モデル、ヒューリスティックデータベース、入力されたドキュメントからメタデータを生成するメタデータ生成モデルによって構成される。以下でそのモデルについて示す。

3.1 二層閾値モデル

二層閾値モデルは意味的連想検索の相関量に代表される順序性のある値を2つの閾値を用いて適合分類・混合分類・非適合分類に分類するモデルである。

意味的ルーティングエンジン SEREN は複数のドメイン依存ベクトル空間に対してコンテキストの応じた意味的連想検索を行い、相関量に応じて、ニュースを配信するシステムである。ドメイン依存ベクトル空間は、意味的連想検索の空間生成方式¹⁶⁾によって共通辞書による一般用語と各分野の専門辞書の専門用語から生成された空間である。一般用語と専門用語の対応付けが行なわれているため、この空間を適用することによって、ユーザは一般用語から高度な専門用語を引くことが可能となる。意味的連想検索は情報検索の技術であり、静的なコンテンツを検索の対象としている。対して、意味的ルーティングエンジン SEREN はニュースという動的なコンテンツを検索の対象としている。情報検索では相関量より全体における順位が重要なものに対して、意味的ルーティングエンジン SEREN においては対象が動的なために相関量の値が重要となる。

二層閾値 $T^{[12]}$ は、意味的に正しい順位とコンテキストとドキュメントの相関量の順位に相違がある場合に対して、確実にユーザのコンテキストに適合していると判断可能なドキュメント (適合分類ドキュメント) と確実にユーザのコンテキストに適合していないと判断可能なドキュメント (非適合分類ドキュメント) と

相関量では判別が出来ないドキュメント (混合ドキュメント) を分類する。二層閾値を図式化したものが図2である。

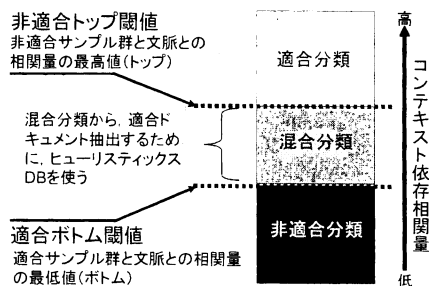


図2 閾値によるドキュメントの切り分け

相関量の値域はコンテキストと空間により大きく変わるため、意味的ルーティングエンジン SEREN の前提とする複数のドメイン依存ベクトル空間において、複数のコンテキストに応じた多角的な検索を行うためには、多数の閾値を設定しなければならない。したがって、閾値を自動的に設定することが必要である。

二層閾値 $T^{[12]}$ はニュースのコンテキストがユーザのコンテキストと適合しているか、ニュースのドメインが計量すべきドメイン依存ベクトル空間のドメインと適合しているか、という異なる2つの視点によって閾値を設定する。適合ボトム閾値とは、ニュースのコンテキストがユーザのコンテキストと適合しているかという視点によって設定される閾値であり、非適合トップ閾値とは、ニュースのドメインが計量すべきドメイン依存ベクトル空間のドメインと適合しているかという視点によって設定される閾値である。適合ボトム閾値を設定するため、コンテキストに適合するサンプル群を適合サンプル群として設定する。充分な数の適合サンプル群が用意されたとき、最も相関量の低い適合サンプルの相関量を今後、出現する適合ドキュメント群の最下限の値として適用する。これに基づき、適合サンプル群とユーザのコンテキストとの相関量を計量し、その中で最も低い値を持つ適合サンプルの相関量を適合ボトム閾値とする。適合サンプル群はコンテキスト c に依存するが空間 MDS には依存しない性質を持つ。

非適合トップ閾値を設定するため、ドメイン依存ベクトル空間とはドメインが異なるサンプル群を非適合サンプル群として設定する。充分な数の非適合サンプル群が用意されたとき、その中で最も相関量の高い非適合サンプルの相関量を今後、出現する非適合ドキュ

メント群の最上限の値として適用する。これに基づき、非適合サンプル群とユーザのコンテキストとの相関量を計量し、その中で最も高い値を持つ非適合サンプルの相関量を非適合トップ閾値とする。非適合サンプル群は空間 MDS に依存するがコンテキスト c には依存しない性質を持つ。

入力されるドキュメント集合を X としたとき、適合ドキュメント集合 $\mathcal{P}(X)$ 、非適合ドキュメント集合 $\mathcal{N}(X)$ 、混合ドキュメント集合 $\mathcal{C}(X)$ は以下のように定義される。

$$\mathcal{P}(X) := \bigcup_{n=0}^j \left(x \mid x \leftarrow \begin{cases} X_{[j]}, & \text{if } p(X; c) \geq T^{[p|n]} \\ \emptyset, & \text{otherwise} \end{cases} \right) \quad (1)$$

$$\mathcal{N}(X) := \bigcup_{n=0}^j \left(x \mid x \leftarrow \begin{cases} X_{[j]}, & \text{if } p(X; c) \leq T^{[p|n]} \\ \emptyset, & \text{otherwise} \end{cases} \right) \quad (2)$$

$$\mathcal{C}(X) :=$$

$$\bigcup_{n=0}^j \left(x \mid x \leftarrow \begin{cases} X_{[j]}, & \text{if } p(X; c) \leq T^P \wedge p(X; c) \geq T^n \\ X_{[j]}, & \text{if } p(X; c) \geq T^P \wedge p(X; c) \leq T^n \\ \emptyset, & \text{otherwise} \end{cases} \right) \quad (3)$$

$$\mathcal{P}(X) \cup \mathcal{N}(X) \cup \mathcal{C}(X) \equiv X \quad (4)$$

精度が良くないドメイン依存ベクトル空間上において、モデルを適用したとき、混合分類ドキュメント群 $\mathcal{C}(X)$ は多くなる可能性がある。このような場合、統計的な手法などを使って、混合分類が減るように閾値を変動させずに、意味のスペクトルを使った空間の学習方式¹⁷⁾を適用し、ベクトル空間自体の精度を上げることで、混合分類ドキュメント群 $\mathcal{C}(X)$ を減らすことが妥当である。

3.2 ヒューリスティックデータベースによる配信可否の決定モデル

ヒューリスティックデータベースは二層閾値モデルによっては、配信すべきか判断不能な混合分類ドキュメント群 $\mathcal{C}(X)$ に対して、配信可否を決定するモデルである。ヒューリスティックデータベースには、コンテキストに応じてユーザが設定した正解ワード群 WT_c が存在する。そのときヒューリスティックデータベースは混合ドキュメント $\mathcal{C}(X)$ 内にユーザが設定した正解ワード群 WT_c が存在すれば配信を行い、そうでなければ配信を行わない。これにより、相関量では判別が不可能であった混合ドキュメント群 $\mathcal{C}(X)$ の配信可否を決定する。

3.3 メタデータ生成モデル

メタデータ生成モデルは意味的ルーティングエンジ

ン SEREN において、ドキュメントを各ドメイン依存空間ごとにドメインに依存した専門性の高いメタデータを自動的に生成するモデルである。

入力されたドキュメント D からその中の含まれる単語 w_1, w_2, \dots, w_n を抽出する。複数のドメイン依存空間 MDS^j の m 個の専門用語を $f_{r-1}^j, f_{r-2}^j, \dots, f_{r-m}^j$ とするとき、空間 MDS^j におけるドキュメント群 D につけられるメタデータ語群 M^j は各空間の専門用語のみで構成される。したがって、ある専門ドキュメントはドメインに合致した空間においては空間 MDS^j の m 個の専門用語との共通部分が多いため、メタデータ語群が多めに付く。逆にドメインが異なる空間ではドキュメントにメタデータ語群がほとんどつかない。メタデータが付かない場合は相関量が 0 になるため、専門ドキュメントが関係の薄いドメイン依存空間において計量され、相関量より配信することになるケースは少なく、専門ドキュメントはドメインの一致する専門空間において配信の可否を決定されやすいシステムを実現している。

4. 実現方式

4.1 意味的情報ルーティングシステム SEREN

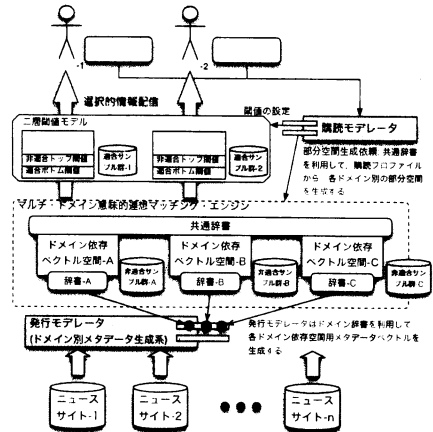


図3 意味的情報ルーティングシステムの概要

意味的情報ルーティングシステム SEREN の概要を図3に示す。本システムはマルチ・ドメイン意味的連想検索エンジン、発行モデレータ、購読モデレータの3つの部分で構成される。

以下に意味的情報ルーティングシステム SEREN がニュースサイトからのドキュメント配信時にユーザへの通知の可否を判定する手順を示す。Step1 が発行モ

デレータ, Step2 がマルチ・ドメイン意味的連想検索エンジン, Step3 が購読モデレータに対応する。

Step1: ドキュメントに存在する単語群と, ドメイン依存ベクトル空間の専門用語群の集合和をとり, 各ドメイン依存ベクトル空間に対応したドキュメントのメタデータ語群を生成する。

Step2: ドキュメントからメタデータ語群をそれぞれのドメイン依存ベクトル空間へ写像し, ユーザが定義したドメインに依存しない共通の辞書によるコンテキストによってドメイン依存ベクトル空間の部分空間を選択し, 各部分空間におけるメタデータ語群のノルムをそれぞれ計量する。

Step3: 各部分空間で計量されたノルムを部分空間ごとに設定されている二層閾値と比較する。すべてのドメイン依存ベクトル空間において, 非適合分類に分類された場合は, そのドキュメントを配信しない。一つでも, 適合分類に分類された場合は, ドキュメントを配信する。混合分類に分類された場合, ヒューリスティックデータベースにより, 配信の可否を決定する。

本研究では SEREN の実現にあたり, 異種の専門空間の一例として, 情報通信分野の空間 (以下 IT 空間) と国際関係分野の空間 (International Relation 空間, 以下 IR 空間) を用いた。IT 空間生成にはオンラインの専門用語集, IR 空間生成には当該分野で汎用的な専門辞書を用い, 共通概念を記述する部分的マトリクス生成 (一般語による専門用語定義) にはロングマン英英辞書を用いた。

なお, 各専門空間は, IT 空間が 1937 次元, IR 空間が 2449 次元であり, 内ロングマン辞書による共通概念部分のみのベクトル空間は 1761 次元となっている。

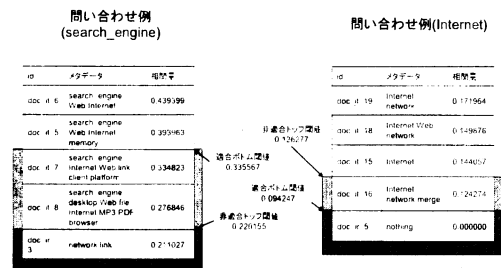


図 4 閾値によるドキュメントの切り分けの例

5. 実験

ここでは二層閾値モデルによる動的集合をターゲット

に, コンテキスト依存の検索結果の妥当性を評価する能力を有する, 自律的な選択的情報配信機構の実現可能性を検証する。

5.1 実験目的

本実験では本方式である適合ボトム閾値, 非適合トップ閾値の値の自動設定の信頼性, 有効性の検証を行うことを目的とする。

5.2 実験方法

5.2.1 実験環境

ロイター通信社のオンラインニュース記事 53 件を対象ドキュメントに設定し, 4.1. において実現した IT 空間を用いて実験を行った。情報通信の分野を対象とした意味空間 M_{IT} をオンラインの IT の専門用語集¹⁵⁾ から意味的連想検索の空間生成方式¹⁶⁾ によって生成する。この方式によって 156 語の専門語と 362 語の一般語によって構成された統合空間を実験に使用する。

5.2.2 問い合わせとドキュメントの設定

問い合わせとして情報通信 (IT) の専門用語 5 件 (desktop, DVD, Internet, MP3, search_engine) を設定した。また, それぞれの問い合わせに対し, コンテキスト上正しいと思われるドキュメントを 10 件ずつ設定した。そのうち各問い合わせにつき 5 件を適合ボトム閾値設定のための適合サンプルドキュメント群として設定し, 残りの 5 件を閾値精度検証のための正解ドキュメント群として設定した。

IT 関連の問い合わせに対してコンテキスト上関係が薄いと思われる国際関係 (IR) 関係ドキュメントを用意し, そのうち 15 件を非適合トップ閾値設定のための非適合サンプルドキュメント群として設定し, 残りの 13 件を閾値精度検証ための不正解ドキュメント群として設定した。

5.2.3 ドキュメントのメタデータ生成

用意されたすべてのドキュメントに対して, 本方式の 3.2. に従い, メタデータ生成を行なう。具体的には以下のステップを実行する。

Step1: シソーラスや専門家が作成した対訳表を用いて IT 専門用語 (複合語を含む) を特定し, 活用や語尾の変化などを統一する。

Step2: IT 空間を構成する専門語のみを抽出し, これをメタデータとする。

このプロセスにおいて, 不正解と設定された IR 関連ドキュメントには IR 関連の専門用語のみがメタデータとして付与されるため, IT 空間において相関量が極端に低くなると期待される。

5.2.4 閾値の設定

Step1: 5.2.2 で設定した問い合わせ 5 件と各適合サンプルドキュメント 5 件との相関量を量り、最も低い相関を持つドキュメントの相関量を適合ボトム閾値の値とする。

Step2: 問い合わせ 5 件と非適合サンプルドキュメント 15 件との相関量を量り、最も高い相関量を持つドキュメントの相関量を非適合トップ閾値の値とする。

以上のプロセスを経て、精度検証の為に用意された正解ドキュメント群および不正解ドキュメント群が適合分類・非適合分類・混合分類のいずれに該当するかを計測し、比較を行なった。

5.3 実験結果

各問い合わせごとに設定された適合ボトム閾値、非適合トップ閾値、分類されたドキュメント群の内訳を図 5 に示す。

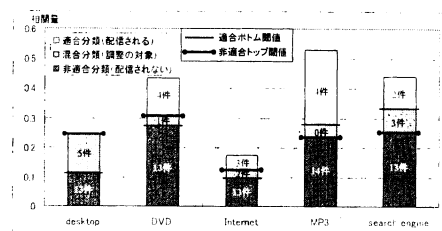


図 5 各問い合わせごとの相関量の分布

問い合わせ用の正解および不正解ドキュメント群が適合分類・非適合分類・混合分類に分けられた結果を、表 1 に示す。

	適合分類	混合分類	非適合分類
正解セット	13	11	1
不正解セット	0	0	65
全体	13	11	66

	正解率	識別率
正解セット	96%	52%
不正解セット	100%	100%
全体	99%	87%

二層閾値モデルを検証するために以下のように正解率と識別率を定義した。

● 正解率

正解ドキュメント群の場合:
 $(\text{適合分類} + \text{混合分類}) / \text{総数} * 100(\%)$

不正解ドキュメント群の場合:
 $(\text{非適合分類} + \text{混合分類}) / \text{総数} * 100(\%)$

正解率は二層閾値モデルの信頼性を量る尺度である。この値が高いとき、誤判断が少なく、信頼性があると言える。

● 識別率

正解ドキュメント群の場合:
 $\text{適合分類} / \text{総数} * 100(\%)$

不正解ドキュメント群の場合:
 $\text{非適合分類} / \text{総数} * 100(\%)$

識別率は二層閾値モデルの有効性を量る尺度である。この値が高いとき、二層閾値による判断能力が高く、有効であると言える。

5.4 考察

図 5 は、コンテキストごとに設定された二層閾値が、各コンテキストに応じた相関量の値域の変化に対応していることを示している。

図 5 の問い合わせ「Internet」の例のように相関量の値域が非常に小さい場合は、それに応じて閾値が低めに狭く設定されており、「DVD」の例のように相関量の値域が大きい場合は、それに応じて閾値が高めに広く設定される結果になっている。問い合わせ「desktop」の例では適合分類ドキュメントが存在しない。これは非適合トップ閾値が非常に高い値を持っていたためにその閾値を超える正解ドキュメントが存在しなかったためである。「desktop」の例のような IT 分野に限定されない一般的な言葉のコンテキストにおいては、非適合トップ閾値が正解ドキュメントに比べて高い値を持つことがあり、相関量では一概に判断できない場合が存在する。この場合は新しいコンテキストワードの追加や、意味のスペクトル¹⁷⁾によりベクトルを修正することで、コンテキストをより明確にする必要がある。

本実験による二層閾値の正解率は表 2 に表されているとおり、全体では 99% であり、識別率は全体で 87% である。不正解ドキュメントにおける正解率、識別率は共に 100% と非常に高い。これはメタデータ生成時に IT 専門用語のみのメタデータが設定されるため、不正解ドキュメントとされる IR 関連ドキュメントにはメタデータが付与されないためである。一方、正解ドキュメントは正解率は 96%、識別率は 52% であり、信頼度はかなり高いが識別率は半分程度の結果になっている。この結果は、本方式は全体として誤判断が少なく、不正解ドキュメントをフィルタリングに

成功していることを示している。

6. 結 論

本稿では、ユーザの指定したコンテキストに応じて、様々な分野のニュースを自動的に、かつ意味解釈を伴う情報選択によって配信する、意味的情報ルーティングエンジン SEREN を示した。本方式は、コンテキストに関係の強い情報と関係の弱い情報の境界線を示す二つの閾値を自動的に設定することで、それぞれのコンテキストに適した情報配信基準を自動的に設定する、高度に自律的な選択的情報配信機構を実現し、評価実験により、本方式による選択的情報配信の有効性を示した。

今後の研究の課題として、二層閾値モデルでは直接判定することが不能であった混合分類を空間の精度を向上させることによって、その数を減らし、本方式の問題の識別率の工場を目指す。

参 考 文 献

- 1) Dan Brickley and R.V. Guha. Rdf vocabulary description language 1.0: Rdf schema. <http://www.w3.org/tr/rdf-schema/>, 2004.
- 2) Antonio Carzaniga, David S. Rosenblum, and Alexander L. Wolf. Design and evaluation of a wide-area event notification service. *ACM Transactions on Computer Systems*, 19(3):332–383, 2001.
- 3) Mariano Cilia, M. Antollini, Christof Bornhövd, and Alejandro P. Buchmann. Dealing with heterogeneous data in pub/sub systems: The concept-based approach. In *Third International Workshop on Distributed Event-Based Systems (DEBS 2004)*, 2004.
- 4) Mariano Cilia, Christof Bornhövd, and Alejandro P. Buchmann. Cream: An infrastructure for distributed, heterogeneous event-based applications. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 482–502. Springer-Verlag Heidelberg, 2003.
- 5) Y. Diao, S. Rizvi, and M. Franklin. Towards an internet-scale xml dissemination service. In *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, 2004.
- 6) Patrick Th. Eugster, Pascal A. Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. The many faces of publish/subscribe. *ACM Computing Surveys*, 35(2):114–131, 2003.
- 7) Bugra Gedik and Ling Liu. Peerccq: A decentralized and self-configuring peer-to-peer information monitoring system. In *ICDCS '03: Proceedings of the 23rd International Conference on Distributed Computing Systems*, page 490. Washington, DC, USA, 2003. IEEE Computer Society.
- 8) T. Kitagawa and Y. Kiyoki. A mathematical model of meaning and its application to multidatabase systems. *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pages 130–135, April 1993.
- 9) Y. Kiyoki, T. Kitagawa, and T. Hayama. A metadata system for semantic image search by a mathematical model of meaning. *ACM SIGMOD Record*, 23(4):34–41, 1994.
- 10) Y. Kiyoki, T. Kitagawa, and Y. Hitomi. A fundamental framework for realizing semantic interoperability in a multidatabase environment. *International Journal of Integrated Computer-Aided Engineering*, 2(1):3–20, Jan 1995. Special Issue on Multidatabase and Interoperable Systems.
- 11) Manolis Koubarakis, Christos Tryfonopoulos, Stratos Idreos, and Yannis Drougas. Selective information dissemination in p2p networks: problems and solutions. *SIGMOD Record*, 32(3):71–76, 2003.
- 12) Ling Liu, Calton Pu, and Wei Tang. Continual queries for internet scale event-driven information delivery. *IEEE Transactions on Knowledge and Data Engineering*, 11(4):610–628, 1999.
- 13) Ling Liu, Wei Tang, David Buttler, and Calton Pu. Information monitoring on the web: A scalable solution. *World Wide Web*, 5(4):263–304, 2002.
- 14) Frank Manola and Eric Miller. Rdf primer. <http://www.w3.org/tr/rdf-primer/>, 2004.
- 15) Fact Monster. Computer glossary, 2005.
- 16) Shiori Sasaki and Yasushi Kiyoki. Space creation and evaluation method using specialized and general knowledge for semantic associative search. In *IEEE International Symposium on Applications and the Internet (SAINT 2005) - Proceedings of the International Workshop on Cyberspace Technologies and Societies (IWCTS 2005)*. IEEE Computer Society Press, Feb 2005.
- 17) 大橋 英博 and 清木 康. 意味的連想検索方式における意味表現ベクトルを対象とした学習機構の実現式. In *情報処理学会研究会報告*, number 72, pages 499–504, 2004.
- 18) 土方 嘉徳. 情報推薦・情報フィルタリングのためのユーザプロファイリング技術. *人工知能学会誌*, 19(19):365–372, 2004.

- 19) 帆足 啓一郎, 松本 一則, 井ノ上 直己, and 橋本 和夫. 非適合プロフィールを利用した文書フィルタリング手法. *情報処理学会論文誌*, 42(3):507-517, 2001.
- 20) 清木 康, 金子 昌史, and 北川 高嗣. 意味の数学モデルによる画像データベース探索方式とその学習機構. *電子情報通信学会論文誌*, j79-D-2(4):509-519, April 1996.
- 21) 吉田 尚史, 清木 康, and 北川 高嗣. 意味的連想検索機能を持つメディア情報検索システムの実現方式. *情報処理学会論文誌*, 39(4):911-922, 1998. 1998.