

地域ポータルサイトにおける 情報配信支援システムの構築

小林拓也[†] 岡本東^{††} 堀川三好^{††} 菅原光政^{††}

ポータルサイトの運用に関する研究は、情報収集の自動化や情報のまとめ方の観点から数多くなされている。しかしながら、特定分野の情報を地域住民に対して配信することが多い地域ポータルサイトを対象とした場合は、運営者の負荷を考慮しながら特定分野の情報のみを収集・配信する方法について解決しなければならない。

本研究では、地域ポータルサイトを対象に、運営者の負荷軽減と継続的な情報収集・配信を行うために、情報を自動収集・分類を支援する仕組み、また、利用者の目的に関連する情報配信を行う仕組みを提案する。特に、利用者の目的に関連する情報配信の仕組みについて、数値実験を行い検証した結果を報告する。

Construction of Information Dissemination Supporting System for Regional Portal Site

Takuya Kobayashi[†] Azuma Okamoto^{††}
Mitsuyoshi Horikawa^{††} and Mitsumasa Sugawara^{††}

A number of researches have been made on automation of the information gathering and how to summarize information in a portal site. In case of the regional portal site which disseminates information on a specific field, the search difficulties occur by the difference in the method of collecting the specific information.

This research's purposes are automatic collection of information, construction of a directory based on each user's purpose and dissemination of information on related purpose, in order to reduce operator's burden and to have continuous information gathering. This paper reports evaluation on the proposal of each structure and the technique of grasping the purpose of users using numerical simulation.

1. はじめに

地域ポータルサイトは、地域情報に特化したサイトであり、地域住民や観光に訪れる人などを対象に、情報の配信を行なっている。地域ポータルサイトの運営者は、地域特性や特定分野に応じた情報収集・分類を行い、それらを利用者の目的に基づいて配信する必要がある。情報収集・分類は、利用者の目的を満たすために継続的に行うことが重要である。しかし、情報の取捨選択、加工や整理などの作業が多く、頻繁に行うことが難しい。情報配信は、利用者ごとに目的やニーズを考慮した検索機能を構築することが重要である。これらの作業は、継続的に行う必要があるが、運営主体が比較的小規模な場合、運営者の負荷を考慮する必要がある。

このことから、地域ポータルサイトの運営を行うために必要な項目を以下にまとめる。

- 地域特性や特定分野に応じた情報収集・分類
- 個々の利用者の目的に合わせた情報配信
- 継続運営を行うため運営者の負荷の考慮した仕組み

本研究では上記の3項目を支援するための方法として、地域特性や特定分野に応じた情報を自動収集・分類、利用者ごとの目的に基づいたディレクトリの構築、および、目的に関連する情報の提示を行う情報システムを提案する。情報収集・分類機能では外部サイトから地域特性や特定分野に応じた情報を自動収集する。これは、ポータルサイト内のコンテンツをもとに、収集する情報の特徴を抽出することによって実現する。収集したコンテンツは内容に基づいて、カテゴリに自動分類する。また、目的に基づいた情報提示は、利用者ごとにプロフィールを作成し、情報推薦や、利用者ごとのディレクトリの構築を行う。

岩手県の子育て情報を配信する地域ポータルサイト（以下、子育てポータルサイトと呼ぶ）の運用を目指し、提案システムを岩手県地域子育て支援拠点施設 CMS¹⁾ (Content Management System)（以下、子育て CMS と呼ぶ）へ導入する。そのため、本稿では、提案システムの各機能について提案を行い、特に利用者の検索目的を把握し、それに基づいた情報推薦を行う手法について、子育てポータルサイトを想定した数値実験によって検証した結果を報告する。

2. 関連研究

一般に、ポータルサイトでは情報の収集や分類をほぼ人手で行っている。その際に

[†] 岩手県立大学大学院ソフトウェア情報学研究科

Graduate School of Software and Information Science, Iwate Prefectural University

^{††} 岩手県立大学ソフトウェア情報学部

Faculty of Software and Information Science Iwate Prefectural University

運営者はどのような情報を収集するか、どのようなまとめ方を行えば良いかについて考慮する必要がある。以下へ、地域ポータルサイトの情報の収集と分類の仕方、扱い方に関する研究を示す。

情報の収集について、大槻²⁾は地域情報ウェブディレクトリの自動編集を行うシステムの提案を行っている。このシステムは属性、地域名、都道府県名、.jp というドメインを基に地域に関係する URL を自動で収集し、システム内で用意したカテゴリに分類を行うことで運営者の情報収集・ディレクトリ構築を行う際の負荷軽減を目的としている。

また、情報の分類について、田村³⁾は地域情報を「動的情報、静的情報」と「生活関連情報、観光情報、行政情報」のマトリックスで分類し、収集、配信する情報を定義した。近藤⁴⁾は配信者の特徴を基に、地域住民が持つ情報、地域の公的な情報、地域企業が持つ情報、地域そのものが持つ情報の4つに分類している。これらは、利用者の検索目的や配信される情報の不変的な特徴から分類を行っており、この分類によって情報を提示するだけでは、利用者ごとに異なる目的や季節や時間の変化に対応し難い。

これらに対し提案するシステムは、情報収集については、地域ポータルサイト内で配信している情報を基に収集する情報をフィルタリングすることで、特定目的の情報を収集する。情報分類については、利用者の検索目的によってディレクトリを再構築する事によって、利用者ごとに異なる目的や季節や時間の変化に対応する。

3. 地域ポータルサイトの運用に対する要求定義

田村³⁾は地域ポータルサイトの定義を、「地域への入り口のサイト」「地域情報に特化したサイト」としている。本研究の対象とする地域ポータルサイトは、範囲を限定し、「ある特定分野の地域情報に特化したサイト」とする。子育てポータルサイトをあてはめると利用者は、子育てを行っている親、子育ての予定がある親などであり、その中でも食事に関すること、病気に関することなど様々な検索目的を持っていることが想定できる。

本章では、運営者の情報収集・分類と、利用者への情報提示について子育てポータルサイトをもとに要求定義を述べる。

3.1 運営者の情報収集・分類

運営者の負荷軽減を行う一つの方法として、運営者を複数用意し、一人あたりの負荷を軽減する方法が考えられる。現在運用を行なっている子育て CMS はこの方法を採用しており、各子育て拠点施設の管理者が運営者となる。管理者は自身の施設のおたよりなどの情報を収集している。子育て CMS については4章で詳しく説明する。表1に子育て CMS の現状を例とした時の収集者と収集される情報をまとめる。運営

表1 情報収集・配信する情報と形態の関係

	収集・配信する情報	子育てに関わる情報	各施設のおたより	各施設の情報	各施設への口コミ	外部サイトのURL	外部サイトの情報
収集者							
運営者		○	×	○	×	×	×
各施設の管理者		×	○	○	×	×	×
利用者		×	×	×	○	×	×

者は子育てに関わる全般的な情報、子育て支援拠点施設に関わる情報を収集しているが、現時点で収集する範囲はあまり広くない。子育てポータルサイトとして運用するためには、運営者は子育て支援拠点施設に関わる情報以外にも幅広く収集する必要がある。

以上のことから、複数の運営者から情報収集する仕組みは維持し、さらに、ポータルサイト運営者に対して、情報収集する際の負荷軽減、幅広く継続的な情報収集を行うため、現在収集が行われていない、Web上に存在する情報の収集を行う。また、継続的に子育てに関する情報を配信している外部サイトのURLの収集を行う。これら収集された情報は、自動的に情報の内容によって判断され、ポータルサイト内に存在するカテゴリへ分類される。

3.2 利用者への情報提示

一般に、検索方法はキーワード検索とディレクトリ検索が存在する。検索したいキーワードが決まっている場合には、キーワード検索を用いる事で、目的に近い情報が比較的簡単に取得できる。検索したいキーワードがわからない場合は、検索目的と関連した内容の情報がまとまったディレクトリ検索から情報を取得する方法が用いられる。しかし、こういった情報が関連すると考えるかは、利用者ごとに違うため、すべての利用者の要求を満たす様なディレクトリを構築するのは困難である。また、多くの分類方法を網羅しようとする、検索が難しいディレクトリとなってしまう。

このことから、検索しやすいディレクトリを実現するためには、利用者ごとの目的に合わせた情報提示をする必要がある。そのため、利用者ごとの目的を把握し、それを基に利用者ごとの小規模なディレクトリを構築することでカテゴリ数と階層を減らし、また、情報推薦を行うことで、目的のカテゴリにたどり着く際のステップ数の減少が図れる。

4. 岩手県の子育て情報を配信する地域ポータルサイト

3章で述べた要求定義を子育てポータルサイトにあてはめる。ここでは情報収集・分類の機能と利用者への情報提示の機能について述べる。また、複数の限定的な権限を持った運営者から収集する仕組みとして、子育て CMS の概要について述べる。

4.1 岩手県子育て支援拠点施設 CMS の概要

子育て CMS は地域子育て支援拠点施設の情報を一元管理するシステムである。地域子育て支援拠点は、子どもや親たちの交流を深める場所の提供、子育てに関する育児講座、子育てに関する不安や悩みなどを相談する育児相談の3つの業務を行っている。地域子育て支援拠点で配布されていた紙媒体のおたよりは、各施設から送られた原稿を複製・編集し、再び各施設へ配布していた。システムの目的として場所・距離・時間に制約があったおたよりを電子化することで、制約を気にすること無く、おたよりを得ることが出来る。このようなおたよりを中心とした、テキストや画像などの情報を統合的に管理する子育て CMS を構築している。子育て CMS の概要を図1に示す。現在子育て CMS は複数の拠点施設の情報と集約した情報のみの発信を行っている。

岩手県内 86 ケ所の地域子育て支援拠点を対象に作成したシステムの運用を 2009 年 1 月から行っており、2011 年 6 月から岩手県の幅広い子育て情報の配信をコンセプトとした、岩手県保険福祉部児童家庭課による「岩手県子育て情報・応援ポータルサイトいわて子育てiランド」内にシステムを移行した。トップページを図2に示す。

おたよりの登録件数は 2009 年:198 件、2010 年:231 件、2011 年 (6/30 現在):96 件であり、イベントの登録件数は 2009 年:860 件、2010 年:1164 件、2011 年 (6/30 現在):413 件となっている。システムの閲覧総数は 19530 回となっている。また、おたよりを毎月配信している施設は 26 ケ所、イベントを毎月登録している施設は 34 ケ所となっている。このことから利用している施設に偏りがあると言えるが、各施設の情報の配信と閲覧は行われており、拠点施設の情報配信としては、十分利用されていると考えられる。子育て CMS を用いることで、岩手県の地域子育て支援拠点の情報収集が行える。

4.2 情報収集・分類の機能

情報収集を行うために、複数の限定的な権限を持った運営者から収集する仕組みとポータルサイト運営者を対象とした負荷を考慮し継続的に収集する仕組みを構築する。前者は子育て CMS を用いることで実現し、後者は、外部サイトから情報を収集する仕組みを構築することで実現する。

子育てポータルサイトの外部から情報収集を行うため、外部サイトの URL を収集する必要がある。外部サイトの URL にアクセスし、地域に関する情報を収集する。収集した情報を子育てポータルサイト内に存在するコンテンツを基にフィルタリングし、子育てに関する情報のみに限定する。また、継続的に子育てに関する情報を配信している外部サイトの URL を同様にフィルタリングし収集する。地域に関する情報の取得は大概ら²⁾の手法を用いる。さらに、

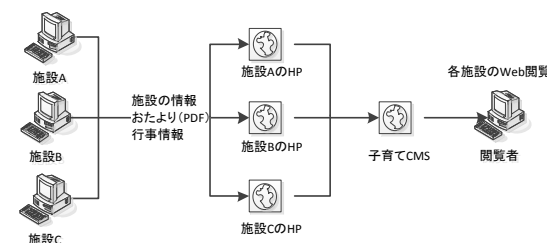


図1 子育て CMS



図2 システムのトップページ

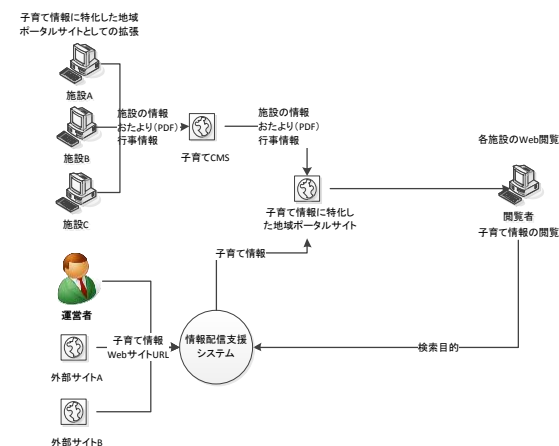


図3 子育て CMS と提案システムの関係

- 外部サイトの URL だけでなく、文章などのコンテンツ収集
 - ある特定分野の情報のみ収集
- を実現するために、大槻らの手法に以下の 2 つの手法を追加する。
- 外部サイトの URL とそのページの内容（文書）2 つについて収集
 - フィルター用いてページの内容によって収集を行うか否か判別

まず、収集した地域サイトにアクセスし、ページを収集する。このページに含まれる文書に対して形態素解析を行い、単語を取り出す。それぞれの単語の重みを TFIDF 法によって求め、閾値を超えた単語を特徴語とし、ページごとの特徴語ベクトルを作成する。次に、これらのページがポータルサイトに関連する情報かどうかを以下の方法で判別する。子育て CMS から収集したおたよりの特徴語ベクトルを作成する。作成したおたよりの特徴語ベクトルとページの特徴語ベクトルの類似度を算出し、閾値未満のページを破棄する。最後に、残ったページを予め用意したカテゴリに自動分類する。これも同様に、カテゴリ毎の特徴語ベクトルを用意し、ページの特徴語ベクトルとすべてのカテゴリの特徴語ベクトルの類似度を算出し、最も値が大きいカテゴリにページを分類する。

図 3 に子育て CMS と提案する機能の関係を示す。2 つのシステムはそれぞれの対象に対して情報収集を行い、運営者を対象とした機能は子育て CMS で収集されたコンテンツを利用することで、ある特定目的の情報を収集することが可能となる。

4.3 利用者への情報提示の機能

利用者のプロフィールを作成し、それに基づいてポータルサイト内の情報推薦、利用者ごとのディレクトリの構築を行う。プロフィールはページごとの特徴語ベクトルと同様に特徴語の重みを持ったベクトルで表現する。作成方法については 5 章で説明する。情報推薦は、利用者が選択したカテゴリ内のコンテンツからプロフィールに類似したものを提示する機能である。ディレクトリの再構築は、プロフィールとカテゴリの類似度を求め、ある一定以上の類似度を持つカテゴリを利用者が興味ありと判断する。それらのカテゴリでディレクトリを構築する。例を図 4 に示す。図の(a)が全体のディレクトリ構造であり、(b)がプロフィールに基づいて再構築したものである。手順は、プロフィールを基に利用者が興味ありと判断したカテゴリを、利用者の目につきやすくするため、上位階層に移動させる。例として、3 階層目に興味ありのカテゴリが存在する場合、2 階層目の親カテゴリを飛ばして 3 階層目のカテゴリにアクセス出来るようにディレクトリを変化させる。この時 2 階層目のカテゴリは類似したカテゴリに集約し、ある一階層のカテゴリ数を減少されることによって、興味ありのカテゴリにアクセスしやすくなる。

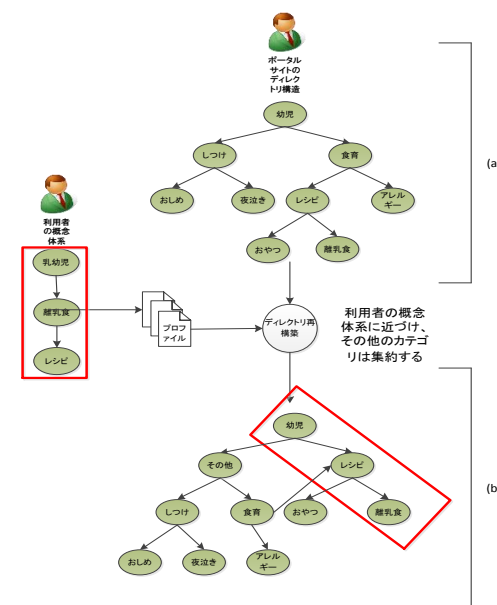


図 4 ディレクトリの再構築

5. 利用者の検索目的の把握

ポータルサイト内の情報推薦を行うため、利用者の検索目的はどのような傾向があるか調査し、それを数値として表す方法である。プロフィールの作成方法について提案を行う。利用者の検索目的は子育て CMS を対象に、検索キーワードを調査することによって明らかにする。

5.1 地域子育て支援拠点 CMS の運用データの分析

調査方法は、子育て CMS に利用者が訪れた際の検索キーワードとコンテンツの内容について情報の特性毎に分類を行った。対象期間は 2010/04 から 2011/04 である。検索キーワードは月平均約 100 種類が表れる。これらの一部を表 2 に示す。検索キーワードと検索キーワードによって検索されるコンテンツの特徴を表すために、属性を「地域性」、「季節性」、「嗜好性」を用意した。地域性とは検索キーワードまたは、コンテンツの内容に地名、方角、位置などの単語が表れたものとしている。季節性とは、節分やひな祭りといった毎年行われるイベントなど世の中の流れを示すものである。嗜好性とは、地域性以外で利用者ごとに変わる検索目的を示すものである。この結果

の考察として、検索キーワードの属性に地域性が含まれていなくても、検索されるコンテンツの属性には多くの地域性が含まれているものが存在することがわかる。

5.2 利用者ごとのプロフィール作成

プロフィールは利用者の興味があるキーワードや位置情報などが要素となっており、それらにどれだけ興味があるかなどを数値化したものである。プロフィールの作成方法は一般に明示的手法と暗黙的手法が存在する。2つの手法についてまとめたものを表3に示す。

明示的手法は、利用者が直接、興味のあるキーワードや文章を選択する手法である。システムや運営者の設問に対して利用者が直接回答するためプロフィールの精度は高いが、ニーズや嗜好が変わるたびに回答しなければならないため、利用者の負荷が大きい。

暗黙的手法は、利用者が閲覧した情報に興味があると仮定し、閲覧した情報を基にプロフィールを作成する方法や、協調フィルタリングによってプロフィール作成する方法である。利用者の行動やWeb上のネットワークから嗜好やニーズを予測するため、ニーズの変化に対応しやすく利用者の負荷が小さい。しかし、明示的手法に比べプロフィールの精度が低いといった欠点がある。また、ある程度行動などのデータを蓄積する必要があるため、システムの利用開始後すぐに利用できない。プロフィールの精度を向上させるための閲覧したページ以外の情報を用いた研究が多くなされている。Morita⁶⁾は利用者のWebページの閲覧時間と興味の度合いの相関関係を明確化し、閲覧時間によってプロフィールに反映する大きさを決定する手法を提案している。大野⁷⁾はWebページを閲覧する際の視線に注目し、視線の停留時間から利用者の興味がある部分を抽出、それを基にプロフィールを作成する手法を提案している。しかし、ある程度行動を蓄積する必要があるといった問題は考慮されていない。

本研究では、明示的手法と暗黙的手法の複合手法によってプロフィールの作成を行う。これによって、利用者の負荷が大きい、データの蓄積が必要といった互いの問題点の解決を図る。プロフィール作成の概要を図5に示す。明示的手法は、直接利用者が回答するという特徴からプロフィールの作成時で使い、作成したプロフィールを利用者の長期間のニーズと見なす。ここで回答するのは、ポータルサイトに存在するカテゴリであるが、全てのカテゴリから回答を求めることは利用者の負荷に繋がる。そこで、カテゴリの特徴語ベクトルを基にクラスタリングを行い、選択数を減少させる。このことによって、プロフィールの精度は低くなるが利用者の負荷は軽減される。暗黙的手法はニーズの変化に対応しやすいことから、プロフィールの変更・修正時に用いる。これは、利用者が閲覧したコンテンツを基とするため、短期間のニーズと見なす。

表2 検索キーワードと検索される情報の属性の一部

		検索キーワードの属性			検索される情報の属性		
	検索キーワード	地域性	季節性	嗜好性	地域性	季節性	嗜好性
1月	おたより タイトル	○	×	○	○	×	○
	(施設名) (イベント)	○	×	×	○	×	×
	節分	×	○	×	○	○	×
	節分 製作	×	○	×	○	○	×
	鬼のお面	×	○	×	○	○	×
	睡眠 ポイント	×	×	○	×	×	○
	幼児食セミナー	×	×	○	○	×	○
2月	ひな祭り 製作	×	○	×	×	○	×
	支援センター 一覧	○	×	×	○	×	×
	(施設名)	○	×	×	○	×	×
	(施設名) (イベント)	○	×	×	○	×	×
	ヨガ	×	×	○	×	×	○
	おひなさま 幼児食	×	○	×	○	○	×
	製作 おひなさま 遊び	×	○	×	○	○	×
3月	ゲーム	×	×	○	○	×	○
	除菌 ねずみ除去剤	×	×	○	×	×	○
	除菌 洗剤	×	×	○	×	×	○
	除菌 ミニマト	×	×	○	×	×	○
	おたより タイトル	○	×	○	○	×	○
	岩手の子育て しやすや	×	×	○	○	×	○

表3 プロファイルの作成手法の特性

	方法	ニーズの変化	利用者の負担	プロフィールの精度 (利用者のニーズを表せているか)	研究数
明示的手法	利用者がニーズを直接回答	対応し難い	大きい	高い	少ない ^[5]
暗黙的手法	利用者の行動(ページの閲覧等)からニーズを予測	対応し易い	小さい	低い	多い

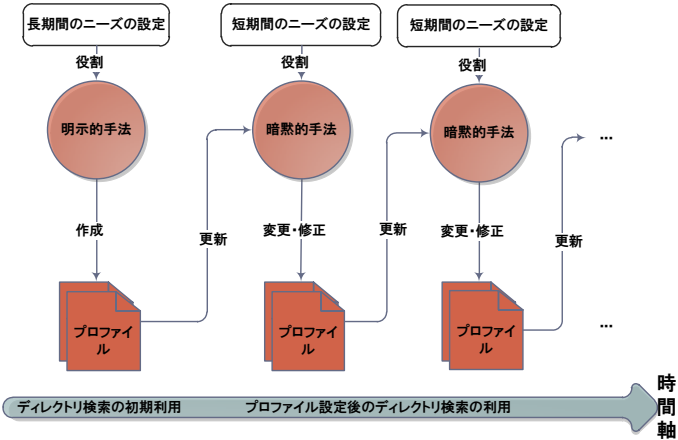


図5 明示的手法と暗黙的手法の複合手法

6. 提案手法の評価

提案したプロファイルの作成手法がどのくらいの頻度で利用する利用者に有効なのか検証を行う。検証を行うために利用者の満足度を指標にする。満足度は利用者の負荷度と、利用者の検索目的に対する検索の達成度からなるものとする。負荷度とは利用者がプロファイルを作成する際に発生し、負荷度は利用頻度が増加すると一回あたりの負荷が減少すると仮定する。利用者の負荷度は明示的手法を用いた場合のみ反映する。

利用者の検索目的の達成度は、利用者の目的に合った情報の探しやすさの事である。そのため利用者の検索目的にあったコンテンツが上位に表示される事で増加する適合率で表す。この適合率は情報推薦・プロファイルの精度に依存する。提案するプロファイルの作成方法は、クラスタを選択する明示的手法と、閲覧したコンテンツから利用

者の検索目的を得る暗黙的手法の複合手法である。暗黙的手法は、利用者の検索目的が変化せず、閲覧したコンテンツが利用者の検索目的に近いものだとすれば、プロファイルは利用頻度の増加に合わせて精度が向上すると考えられる。そこで利用者の検索目的に対する検索の達成度について

- 利用者の検索目的が変化しない
- 利用者は検索目的に近いコンテンツを閲覧する

という仮定のもと、利用頻度に対してどのように精度が上がっていくのか、また、提案手法でどの程度の利用頻度において有効かを明らかにするため、実験を行った。

6.1 実験

仮定に基づいてコンテンツを閲覧するたびに上昇していく満足度を数値として表すための事前実験を行う。利用者の検索目的を想定し、コンテンツを閲覧するたびにプロファイルを作成し、その際の適合率の上昇の仕方を観察する。利用者を2人想定し、検索目的を表4の様に設定した。

6.1.1 実験1

まず、検索目的1の暗黙的手法のみで作成したプロファイルを用いた場合の適合率を図6に表す。このグラフは縦軸に適合率を示し、横軸に利用頻度を示す。利用頻度はコンテンツを閲覧するたびに増加していく。適合率は、利用頻度が増加していくにつれ増加することがわかる。また、対数関数の様な増加の仕方をしていることがわかる。これは、杉山ら⁸⁾の実験結果でも同様な結果を示しており実験結果は妥当では無いかと考えられる。

この結果から、適合率は利用頻度に対して対数関数的に増加していくのでは無いかと考え、対数曲線を求めた。結果を図7に示す。父親1の決定係数は0.684であり、父親2の決定係数は0.875という精度であった。

表4 利用者の検索目的

対象者	子供	検索目的1(嗜好性)	検索目的2(地域性)
子育てを行なっている父親1	小学校低学年の子供が一人 2歳の子供が一人	子供の夏休みの自由研究について調べたい	盛岡地区
子育てを行なっている父親2	3ヶ月の子供が一人 1歳半の子供が一人	1歳半の子供の離乳食を作りたい	北上地区

6.1.2 実験2

同様に、提案手法で作成したプロファイルを用いた場合の適合率と対数曲線を図8に示す。適合率が暗黙的手法より低くなっているのは、クラスタに含まれる特徴語にノイズがあり、適合率が上がりにくくなったためだと考えられる。

6.1.3 満足度の定式化

実験1, 2から得られた結果を基に利用者の満足度をモデル化し、提案手法の有効性を検証する。満足度は以下の式で定義する。

$$S = (1 - \gamma) \left(1 - \frac{\beta}{FoU}\right) + \gamma(a \ln(FoU) + b) \quad (1)$$

S: 満足度

FoU: 利用頻度

$\left(1 - \frac{\beta}{FoU}\right)$: 利用者の負荷度の逆数

$a \ln(FoU) + b$: 利用頻度によって対数関数的に変化する利用者の検索目的の達成度

$$\begin{cases} a \ln(FoU) + b & (0 \leq a \ln(FoU) + b \leq 1) \\ 1 & (a \ln(FoU) + b > 1) \end{cases}$$

β : 明示的手法の有無[0,1]

γ : 負担度、達成度の割合を決めるパラメータ

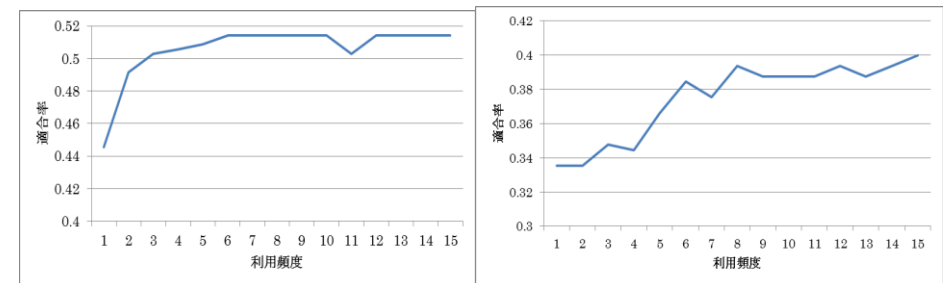


図6 暗黙的手法のみの利用頻度と適合率の関係 (左: 父親1, 右: 父親2)

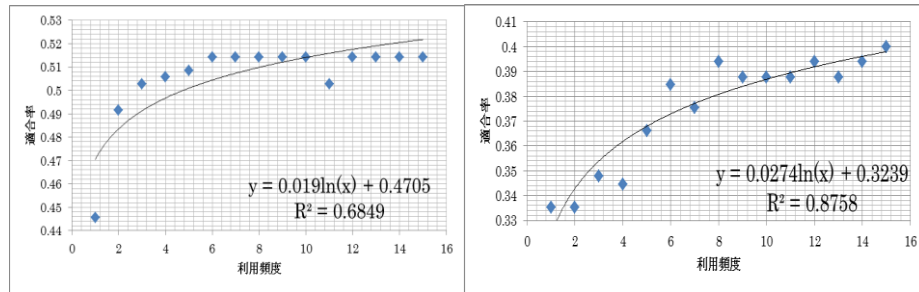


図7 暗黙的手法のみの対数曲線 (左: 父親1, 右: 父親2)

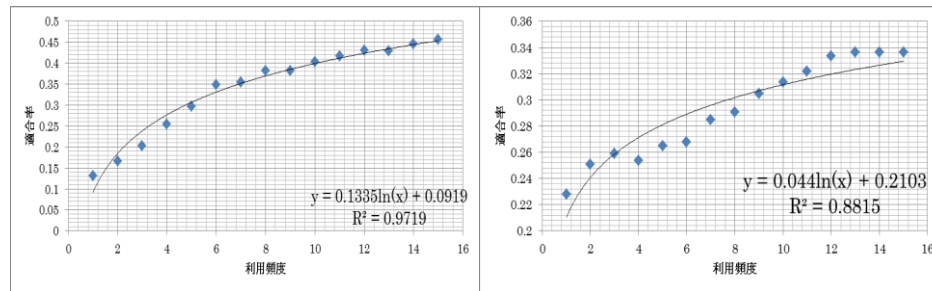


図8 提案手法の対数曲線 (左: 父親1, 右: 父親2)

表5 提案手法が有効である対象範囲

	クラスタの選択	満足度
提案手法が有効	検索目的に近い	暗黙的手法のみを超える
提案手法が有効でない	検索目的に遠い	暗黙的手法のみのを超えない

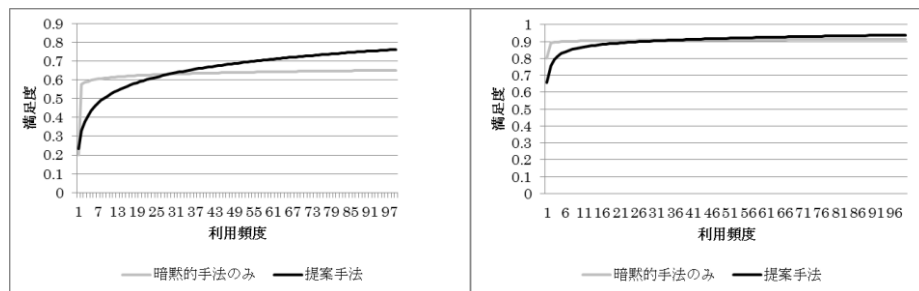


図9 父親1の暗黙的手法と提案手法の満足度の比較 (左: $\gamma=0.8$, 右: $\gamma=0.2$)

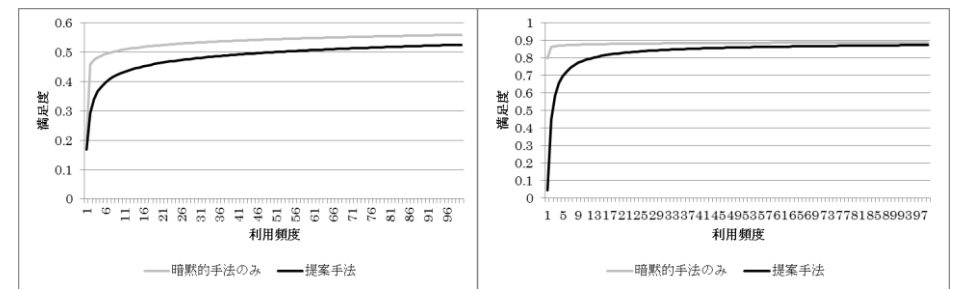


図10 父親2の暗黙的手法と提案手法の満足度の比較 (左: $\gamma=0.8$, 右: $\gamma=0.2$)
利用者の負荷度は、仮定から明示的手法がある場合のみとし、利用頻度に反比例する。これは利用頻度が増加すれば、プロフィールを作成する一度目の負荷が小さくなっていくだろうという仮定に基づく。検索目的の達成度は実験1, 2の結果から対数関数で表す。この時、利用者の負荷度の逆数と検索目的の達成度は0から1の間に存在し、 γ によって、どちらを重視するかを決定する。検索目的を重視する場合には γ の値を高く、負荷度を重視する場合は低くする。実験結果を図9に示す γ は0.8と0.2を用いて行った。ここで用いた式は以下の様になる。

$$\text{暗黙的手法のみ } (\gamma=0.8) : S = (1-0.8)(1-\frac{0}{FoU}) + 0.8(0.0019\ln(FoU) + 0.4705) \quad (2)$$

$$\text{提案手法 } (\gamma=0.8) : S = (1-0.8)(1-\frac{1}{FoU}) + 0.8(0.1335\ln(FoU) + 0.0919) \quad (3)$$

この結果から、暗黙的手法のみでは、1回目以降は満足度が停滞している。提案手法は、一回目は負荷度の逆数と検索目的の達成度が低く、満足度が低い、10回以降から増加していき、30回目前後で暗黙的手法の満足度を超えている。特に検索目的を重視した場合に満足度の差が大きくなる。これは、利用頻度が増えることで負荷を感じなくなったことと、明示的手法で選択したクラスタが、閲覧しただけでは取得できない特徴語を持っており、相乗効果によって達成度が増加したためだと考えられる。

しかし、父親2では暗黙的手法のみの方が提案手法より良い結果となり、利用頻度が1000回を超えても提案手法が暗黙的手法のみの満足度を超えなかった。結果を図10に示す。ここで用いた式は以下の様になる。

$$\text{暗黙的手法のみ } (\gamma=0.8) : S = (1-0.8)(1-\frac{0}{FoU}) + 0.8(0.0274\ln(FoU) + 0.3239) \quad (4)$$

$$\text{提案手法 } (\gamma=0.8) : S = (1-0.8)(1-\frac{1}{FoU}) + 0.8(0.044\ln(FoU) + 0.2103) \quad (5)$$

暗黙的手法のみ提案手法それぞれ、満足度が停滞している結果となった。これは、選択したクラスタに離乳食や食育と言った特徴語が無かったために、暗黙的手法で高

まる精度の妨げになったのだらうと考えられる。実験結果から検索目的に近いクラスタを選択したかどうかで、満足度が変わってくると推測される。現段階での実験結果では、検索目的に近いクラスタが選択できた場合、利用頻度が多い利用者にとって、提案手法は有効となることが考えられる。

7. おわりに

本研究では、運営者の負荷軽減と継続的な情報収集を行うために、情報を自動収集・分類し、利用者ごとの目的に基づいたディレクトリの構築、また、目的に関連する情報の提示を行う情報配信支援システムの提案を行った。

特に、利用者の検索目的を把握し、情報推薦を行うための仕組みについて子育て CMS を基に利用者の検索目的を調査し、利用者の負荷軽減を目的としたプロフィールとして表す方法の提案を行った。提案した手法の有効性を示すため、暗黙的手法のみのプロフィールを作成した場合との比較を数値実験によって行った結果、利用頻度が多い利用者にとって有効だと推測した。今後の課題として、ディレクトリの再構築のロジックの有効性とシステム構築、また、情報収集の自動収集・分類のシステム構築を行い、岩手県の子育て情報を配信する地域ポータルサイトとして運用し、システムの有効性を確かめていく。

参考文献

- 1) 下堀智史, 岡本東, 堀川三好, 菅原光政: 地域子育て支援拠点の利用満足度向上を目的とした地域子育て支援拠点 CMS の構築, 平成 20 年度日本経営工学会東北支部卒業論文・修士論文発表会抄録集, p.8(2009)
- 2) 大槻洋輔, 佐藤理史: 地域情報ウェブディレクトリの自動編集, 情報処理学会論文誌, Vol.42, No9, pp.2310-2318 (2001)
- 3) 田村信之: 地域ポータルサイトについて, 情報の科学と技術, 55 巻 2 号, pp.76-80(2005)
- 4) 近藤真由, 後藤昌人, 服部哲, 安田孝美, 横井茂樹: 地域ポータルサイトにおける CMS の実践的活用と今後の課題, 情報処理学会研究報告, 2008-IS-103 (5) (2008)
- 5) 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 巻 3 号, pp.365-372 (2004)
- 6) Morita, M. and Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, in Proc. of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp. 272-281 (1994)
- 7) 大野 健彦: IMPACT: 視線情報の再利用に基づくブラウジング支援法, in Proc. of the 8th Workshop on Interactive Systems and Software (WISS'2000), pp.137-146 (2000)
- 8) 杉山一成, 波多野賢治, 吉川正俊, 植村俊亮: ユーザからの負荷なく構築したプロフィールに基づく適応的 Web 情報検索, 電子情報通信学会論文誌, Vol.J87-D-1No.11, pp975-990 (2004)