

ユーザの閲覧意図にレスポンスな Web ページ最適化に関する検討

楠牟禮章[†] 牛尼 剛聡^{††}

[†]九州大学芸術工学部 〒815-8540 福岡県福岡市南区塩原 4-9-1

^{††}九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]1ds12181p@gmail.com, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし 近年、パーソナルコンピュータを始め、スマートフォンやタブレットなどの携帯型情報端末の普及により、Web サイトの閲覧は様々なデバイスで行われるようになってきた。そうした背景の下、デバイスの特性に対して動的に適合する Web ページデザイン（レスポンス Web デザイン）が主流になってきた。レスポンス Web デザインでは、閲覧するデバイスの特性に合わせて、ユーザが閲覧しやすいデザインで表示がなされる。しかし、同一の Web ページに対して、ユーザの閲覧意図は様々であり、閲覧意図に適したデザインが存在するにもかかわらず、ユーザの閲覧意図に基づいたレスポンスな Web デザインを提供する機構は存在しない。例えば、ユーザが、現在のページから他のページに移動したいと考えている場合、意図しているページに移動する為のアンカーテキストを探す必要がある。具体的には、電機メーカーのページを閲覧する状況で、ユーザがページ A でナビゲーションメニューの「会社概要」、「製品一覧」、「お問い合わせ」の順に並んでいる中から、「製品一覧」を選択した場合、ユーザは製品について興味があり比較している、という推定ができる。このとき、別のページ B のナビゲーションメニューでも「製品一覧」を優先的に左側に表示するような最適化を動的に行うことで、ユーザは目的の情報に移動する為のリンクを効率的に発見できることが期待できる。本研究では、このような、ユーザの意図を反映し、ユーザの意図に対してレスポンスな Web ページの最適化を行う機構を実現するための基礎的な検討を行う。

キーワード 閲覧意図、レスポンスデザイン、コンテンツ最適化、インターフェース

1. はじめに

現在、人々が情報を知りたいとき、Web ページを利用することが多い。Web ページは、書籍等の従来型のメディアに比べると、インターネットを利用できれば、容易かつ効率的に情報を取得できる。Web では、HTML(Hypertext Markup Language)を用いることで、人々は簡単に Web ページを制作し、公開できるため、Web 上には利用可能な膨大な情報が存在するようになった。そうした中で、膨大な情報の中から、ユーザが必要とする情報を効率的に取得することが困難になっている。ユーザが Web ページにアクセスする際は、何か特定の情報を得たい、という意図が存在することが多い。本研究では、ユーザが Web ページにアクセスした意図を、Web ページ閲覧の振り舞いから推定し、ユーザが求める情報を効率よく取得するための支援を行うことを目標とする。つまり、ユーザの意図に基づいて、ユーザにとって利用しやすく、かつ効率的な Web ページ閲覧ができるようにするための手法の開発を目指す。

一方、近年、パーソナルコンピュータ、スマートフォン、タブレット型 PC、TV 等、Web ページを閲覧するためのデバイスが多様化している。画面のサイズや解像度がデバイスによって異なることはもちろん、スマートフォンやタブレットなどのデバイスではタッチパネル、パーソナルコンピュータであればマウスやトラックパッドのようなインタフェースを用いるため、デバイスによって操作や表示の仕方が大きく異なっている。こうした状況の下で、全てのデバイスで同じように Web ページ

を表示すると、ユーザビリティが低下する状況が発生する可能性がある。そのような場合には、ユーザは求める情報を効率よく得ることができない。

このような問題に対処するために、近年、レスポンス Web デザインが注目されている。レスポンス Web デザインでは、表示するデバイスの幅に合わせてコンテンツ自体の幅や文字のサイズを動的に変更したり、種類に合わせて表示を変えたりなど、デバイスの特性にあわせた適切な表示を自動的に行う。

しかし、現在のレスポンス Web デザインは、あくまでもデバイスに合わせて表示を最適化するものであるため、いわば、デバイスに対するレスポンス性を提供しているだけに過ぎない。したがって、Web ページにアクセスしたユーザの意図が異なっていたとしても、同一の Web ページが提示される。ユーザは共通のページに対して様々な意図を持ってアクセスしていると考えられるが、ページ内のコンテンツはユーザの意図を考慮していない。すなわち、現在のレスポンス Web デザインではあくまでもデバイスに応じて最適化を行っているため、個人の閲覧意図に対してはレスポンスではなく、このような問題には対処できない。

上記の問題を解決するために、本研究では、ユーザの Web ページの閲覧履歴から閲覧意図を推定し、それに応じて Web ページを最適化することを目的とする。Web ページの最適化には様々な手法が考えられるが、本論文では、最も一般的で代表的な例として、リンクナビゲーション最適化手法の提案を行い、評価実験によって有用性を評価する。また、結果に基づいて本

手法の問題点や具体的なシステムの実装について検討する。

2. 関連研究

Web ページ閲覧の支援として、重要だと判断された単語をハイライトして表示することは一般的な手法である。この手法は Google Quick Scrol [1] という Google chrome の拡張機能でも用いられている手法である。この手法では入力した検索キーワードを閲覧意図としており、そのキーワードが出現する位置に自動スクロールし、ハイライト表示することで Web ページ閲覧の効率を高めた機能だといえる。しかし、この機能ではキーワードが正しいことが前提となっており、曖昧なキーワードでしか検索できなかった場合、不十分な結果が得られてしまう可能性が高い。

また、横尾ら [2] は、検索キーワードの類語や同義語をページ内から探し出し、表示する手法を提案している。これは検索キーワードに対して類語、同義語をオンライン辞書から取得し、いずれかの単語が含まれる箇所を重要箇所とするものである。この手法では、類語、同義語も考慮に入れているため、曖昧なキーワードやコンテキストにあった箇所が探し出せるところにメリットがある。本研究では、Web ページの閲覧履歴からユーザーにとって重要だと思われる情報を取得することによって、ユーザーの Web ページ閲覧を支援することを目的にしている点で、この研究とは異なる。

ユーザーの目的である Web ページ上の情報の発見支援を行うシステムとして、PopoutPrism [3] や ScoutView [4] が挙げられる。

PopoutPrism では、Web ページをページ全体のサムネイル画像と通常のページと分割した上で、キーワードを強調して表示することでユーザーが求める情報が記載されている箇所の発見を支援する。更に、サムネイル画像をダブルクリックすることでその箇所までスクロールするため、閲覧行動の支援も行う。

ScoutView は、PopoutPrism の機能を拡張したもので、マウスカーソルをサムネイルに合わせるとその箇所が拡大される。ユーザーはページ内でキーワード検索を行うと、スクロールすることなく、どの箇所に目的の情報が書かれているかを確認出来る。

本研究では、これら研究と異なり、ユーザーは通常の閲覧行動を繰り返すだけでユーザーの意図を推定し、目的の情報を発見する支援を行うことを目指す。

閲覧行動からユーザーの意図を推定する手法として、顔ら [5] はユーザーのスマートフォンでのオンラインショッピングサイトに対する振る舞いから購買意図を読み取り、最適化する研究を行っている。閲覧時間やスワイプ速度からユーザーが興味のある商品を推定し、未読部分をユーザーにとってより興味のある商品の順に並び替えるというものである。この研究ではオンラインショッピングサイトに対して意図の推定と最適化を目指している点と本研究では操作の振る舞いについては考慮しないという 2 点が、本研究とは異なる。

本手法では閲覧履歴から意図推定を行うが、木本ら [6] はブラウザ操作履歴に基づいた Web サイト改善ツールを提案してい

る。このツールでは、ユーザーがブラウジングした際の操作ログを記録し、行動に合わせた最適化を行う。例えば、多くのユーザーが検索するキーワードにはあらかじめ注釈をつけたり、よく印刷されるページには印刷ボタンをつけることで閲覧行動の支援を行う。これにより、ユーザーの Web ページ閲覧の効率は良くなると考えられる。しかし、本研究とは履歴を用いた Web ページ閲覧支援という点では類似するが、本研究は個人単位で閲覧意図を推定し、閲覧意図をもとに Web ページを最適化することを指すため、本研究とは異なる。

以上を踏まえ、本研究ではユーザーはこれまでとは違う操作をせず、ただ複数の Web ページを閲覧をしているだけでユーザーの意図を推定し、目的の Web ページまでアクセスの支援ができるようになるような手法を提案する。

3. アプローチ

本研究では、対象とする状況として、ユーザーが複数のサイトから同一テーマのものを比較するという状況を想定する。ユーザーの閲覧履歴を基に何を閲覧しようとしているか（閲覧意図）を推定し、ユーザーの意図に適したページ構成に最適化する。しかし、ページ内のコンテンツのレイアウトを大幅に変更してしまうとユーザーの混乱を招いてしまうため、本研究では、最適化を同レベルの要素間での並び替えによって行う。HTML で作成された Web ページは入れ子構造をしているため、並列する要素は同じ階層に記述される。そのため、並び替え自体は容易に行うことができ、レイアウトを大きく変えることもない。今回は同一階層に記述され、ページの閲覧を行う中で非常に重要な役割を果たすであろうと考えられるグローバルナビゲーションのリンク配置の並び替えを行う。

ユーザーは図 1 のように、Web ページをブラウジングをする際、左上から右上に向かって閲覧した後、また左側に視点を戻し、下に向かって進むというのが基本的な行動 [7] であるため、ユーザーの閲覧意図に近いコンテンツのリンクを目につきやすい左上に配置し直すことによって効率的なブラウジングが可能になると考える。

4. 提案手法

提案手法の概略を説明する。提案手法では以下の流れで処理を行う。

1. ユーザーの閲覧履歴から閲覧意図を推定する
2. Web ページの構造分析を行う
3. 閲覧意図に応じて最適化を行い、ユーザーに提示する

ユーザーの Web ページの閲覧履歴をもとに閲覧意図ベクトルを作成し、次に閲覧する可能性があるページとの相関を計算し、相関が高いページのアンカーリンクを優先的に目に入る位置に再配置する。

以上のような大まかな処理の流れについて、以下で詳しく説明する。

5. 閲覧履歴を用いた意図推定

本研究では、ユーザーが Web ページを閲覧した履歴を利用し

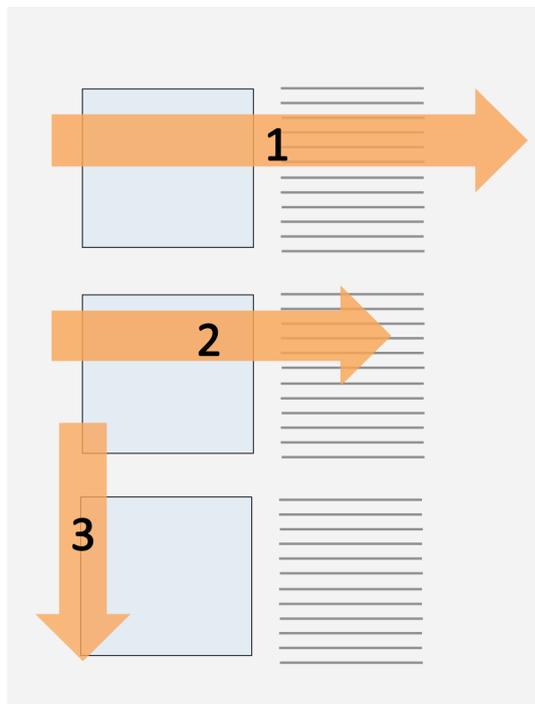


図1 ユーザの Web ページ閲覧時の視線の動かし方

て、ユーザがどのような意図を持って Web ページを閲覧しているのか推定する。本研究では、意図推定を行うための手法を2つ提案し、実験を通して比較、考察を行う。

5.1 手法 1

まず、ユーザの閲覧意図の推定し閲覧意図ベクトルを作成するアルゴリズムについて説明する。

意図の推定には直前に閲覧した Web ページの内容を使用する。これは、本研究で想定している状況が、ユーザが複数のサイトから同一テーマのものを比較するという状況であるため、直前に閲覧したページの方がより意図が反映されたページを閲覧していると考えられるためである。

閲覧履歴の各ページから HTML を取得し、形態素解析を行う。本研究ではテキストデータのみを使用し、扱う品詞を名詞のみとする。

ページ毎に抽出した名詞に対して、を TF-IDF 法 [8] を用いて重み付けする (1)。さらに求めた TF-IDF 値を単語毎に合計し、ページ数で割ることでユーザーが閲覧したページに含まれる単語の重みの重心を求める (2)。この重心をユーザの閲覧意図を表す特徴ベクトルとする。

$$tfidf(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}} \cdot \log \frac{N}{df(t)} + 1 \quad (1)$$

$$V_i = \frac{\sum tfidf_i}{N} \quad (2)$$

ここで、 $n_{t,d}$ は特定の単語が 1 ページに出現した単語数、 $\sum_{s \in d} n_{s,d}$ は 1 ページの総単語数、 N はユーザが閲覧した総ページ数、 df は対象となる単語が出現したページ総数をそれぞれ表す。

5.1.1 例

実際に、閲覧履歴を利用し、意図を抽出した例を示す。著者の一名が情報系の大学院について閲覧し、閲覧意図を表す特徴ベクトルを構成した例を示す。スタートは各大学のトップページからとした。ブラウザから閲覧したページの履歴を取得し、5. のアルゴリズムに基づいて計算した。総ページ数は 41 ページで名詞は 2299 語抽出できた。

表 2 に抽出された TF-IDF 値が大きい名詞上位 23 件とその TF-IDF 値を示す。

表 1 閲覧意図を表す特徴ベクトルの上位 23 件	
抽出された単語 (名詞)	TF-IDF 値 (小数点 3 桁以下四捨五入)
芸術	0.32
ここ	0.27
研究	0.23
情報	0.17
特	0.15
メニュー	0.15
学府	0.12
専攻	0.13
論	0.12
記事	0.12
工	0.11
選択	0.11
大学院	0.09
科	0.09
工学	0.09
工学部	0.09
デザイン	0.08
学科	0.07
教育	0.07
学	0.07
科学	0.07
学部	0.07
設計	0.07

以上の結果より、抽出した TFIDF 値が大きい名詞群の中に「研究」や、「学府」、「専攻」といった相関があると考えられる名詞が数件確認できた。よってこのアルゴリズムで作成した閲覧意図ベクトルはある程度の妥当性があると判断できる。

5.2 手法 2

手法 1 では、閲覧履歴から各ページ毎の単語について TF-IDF 値を計算し、それをもとに閲覧意図ベクトルを作成した。しかし、1つのサイト内でしか用いられない単語があると、その単語の値が大きくなりやすいということが考えられる。例えば、ユーザが複数の電機メーカーのページを閲覧する時、ユーザにとって各メーカー名は既にそのメーカーの Web ページを閲覧しているため、重要ではない。しかし、サイト内では頻繁に用いられ、サイト外では基本的には記述されない。これは TF-IDF 値を用いる上で不適切な結果を生じさせる可能性がある。そこで、サイト単位での単語を考慮し、対応付けを行うことを考える。

手法 1 では、ベクトルの定義を TF-IDF 値をもとに行った。

しかし、実験で算出された閲覧意図ベクトルの上位にメーカー名が含まれていた。本実験においては3つのメーカーのサイトのみを閲覧する状況設定であったため、ユーザの閲覧意図としては不適切だと考えられる。そこで、サイト内でよくつかわれている単語の重みを小さくし、かつ異なるサイトに共通に利用されている単語については重み付けを行う手法を提案する。

TF-IDF 法では、TF (Term Frequency) と、文書集合における単語を含む文書の出現頻度を表す DF (Document Frequency) を利用する。本研究では、サイト内でよく用いられている単語の重みを小さくするために、DF の代わりに PF (Page Frequency) を定義する。これは「サイト内のページに対する出現頻度」を意味する。さらに、複数サイトをまたいでよく出てくる単語はユーザの興味に近いと考えられるため、SF (Site Frequency) を定義する。SF は、その単語がいくつのサイトに含まれているかを指し示す値である。これらの値を用いて単語に対する閲覧意図ベクトルの値を以下の式で定義する。P はサイト内での総ページ数を示している。この演算を全ての単語毎に行い、ユーザの閲覧意図ベクトルを生成する。これを従来の TF-IDF 法にならって TF-IPF-SF 法 (3) とする。ここで、 $tf(w, p)$ は単語 w のページ p における出現頻度を表し、 $pf(w)$ は単語 w を含むページ頻度を表し、 $sf(w)$ は単語 w を含むサイト頻度を表す。

$$tfipfsf(w, p) = tf(w, p) \cdot \log \frac{|P|}{pf(w)} \cdot sf(w) \quad (3)$$

6. Web ページの構造分析

本研究では、Web ページに含まれるリンクの順序を最適化することを考える。そこでページ内の構造分析を行い、どの箇所について並び替えるのか、ユーザの興味を表す閲覧意図ベクトルと比較するベクトルをどのように定義するか、などの分析が必要である。Web ページの構造的な特徴を理解する必要がある。

6.1 最適化の対象

本論文では、今回はグローバルナビゲーションのリンクの順番を最適化することを目標とする。このとき、グローバルナビゲーションのリンクの順番を並び替えることでレイアウトが大きく崩れてしまう可能性がある。Web ページには決められた幅に大小様々なコンテンツが配置されているため、どのコンテンツでも好きに並び替えることができるわけではない。例えば、図2のように幅が800pxのページに350pxと450pxのコンテンツA、コンテンツBが並んでいて、下に同じように幅600pxのコンテンツC、200pxのコンテンツDが並んでいたとする。仮に、コンテンツBとコンテンツCを並び替えることになった場合、コンテンツBは最上段に出来た隙間に入ることができないため、その下に配置される。また、コンテンツDも同じ理由でコンテンツBの下に配置されることになり、下のレイアウトを大幅に崩してしまう。

この問題は、HTMLの構造上の特徴を利用して解決する。図2のようなコンテンツは1つ1つを個別の要素として記述さ

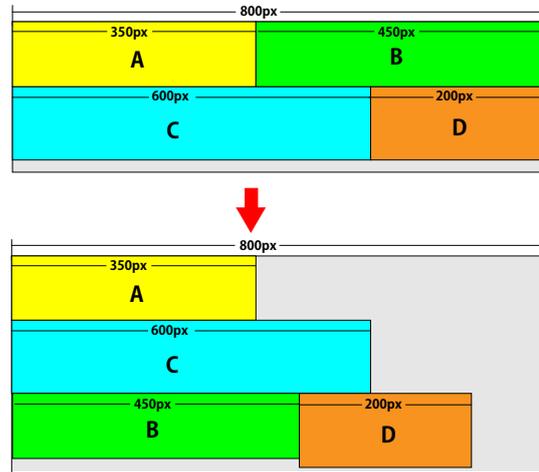


図2 再配置をした例

れるため、配置される順序を安易に変えることは適切でない。Web ページを構成する HTML は入れ子構造で記述されており、並列する要素は同階層に記述されることが一般的である。今回は `<nav>` というタグに注目し、そのタグの中のコンテンツに対して並び替えを行う。このタグの中の要素は、順序のないリストの要素が記述されることが一般的である。そのため、コンテンツの並び替えによるレイアウトの崩れは起きにくい。また、各コンテンツはタグの中に並列して記述されているため、並び替えは容易に可能だと考えられる。

最適化を行うの箇所はグローバルナビゲーションの中の並列するコンテンツ (アンカーリンク) 間で行うが、ページ内でグローバルナビゲーションを見つける必要がある。本手法では `<nav>` のタグを用いる。このタグは、ナビゲーション部分をマークアップするためのタグで、ページ構造をより簡潔に解釈できるようになるためのものである。現状では HTML5 は互換性の関係上の問題であまり普及しているとは言えないが、今後は普及してくることが予想される。本研究ではこのタグ内のコンテンツに注目し、並び替えを行う。

7. Web ページ最適化

並び替えを行う対象となるページ内のグローバルナビゲーションに含まれるリンク先のページに対して、閲覧意図ベクトルを作成した時と同じように、リンク先のページの名詞毎に TF-IDF 値を計算し、ページの特徴量を抽出する。各ページの特徴量を特徴ベクトルとして定義する。このベクトルと 3.1 で求めたユーザの閲覧意図のベクトルを、コサイン類似度 (4) を用いて類似度を計算する。

閲覧意図ベクトルと各ページの特徴ベクトルが類似していると判断される場合は、類似していると判断されたページにリンクするアンカーリンクを優先的に目のつきやすい左側に並び替える。例えば、ユーザが電機製品の購入するためにページの閲覧をしていた場合を考える。ページのナビゲーション部分に「個人のお客様」、「法人のお客様」、「会社情報」、「製品情報」、「採用情報」、「お問い合わせ」の6つのリンクがあり、この順序で並んでいる場合、ユーザにとって興味があると考えられる

ページは「製品情報」のページである。よって図3のように製品情報を左に配置する。

$$\cos(q, d) = \frac{\sum_{i=1}^{|V|} q_i \cdot d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \cdot \sqrt{\sum_{i=1}^{|V|} d_i^2}} \quad (4)$$



図3 グローバルナビゲーションの最適化

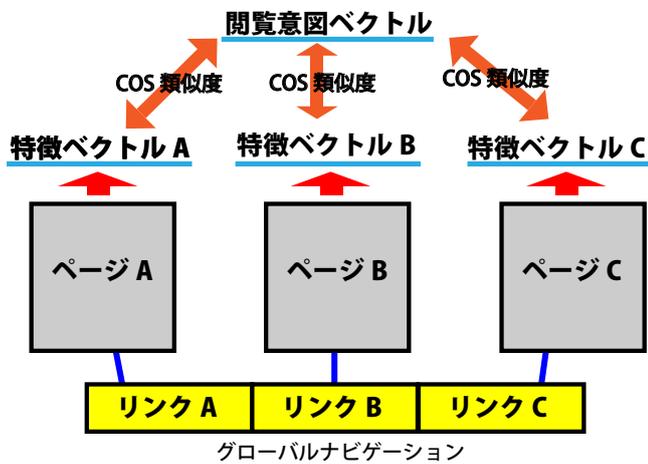


図4 最適化のイメージ

8. 実験

提案手法を利用して、ユーザの閲覧履歴から目的のページが推定が可能かどうかの実験を行った。

8.1 実験の概要、手順

被験者6名に、「現在興味があるが複数のメーカーが存在するために選択に悩んでいるもの」というテーマを設定し、それぞれの興味があるものについてWebページをブラウジングしてもらった。時間による興味の減衰が考えられるため、閲覧時間は5分間とした。比較するメーカーを全部で3社とし、その3社のWebページから意図に基づいて自由に閲覧してもらった。

リアルタイムに解析するシステムは未実装であるため、被験者の閲覧が終了したら、手動でブラウザの閲覧履歴からURLを取得し、手法1と手法2に基づいて閲覧意図を計算し、ユーザの閲覧意図ベクトルを定義した。

その後、同業他社であるメーカーのWebページのグローバルナビゲーションに含まれるリンク先のページの特徴ベクトルを計算し、類似度を求めた。

8.2 実験結果

被験者6名のそれぞれの閲覧意図とその結果を記す。

被験者AはAV機器、パソコン・スマートフォン・タブレットの製品情報の閲覧に興味があった(図5)。

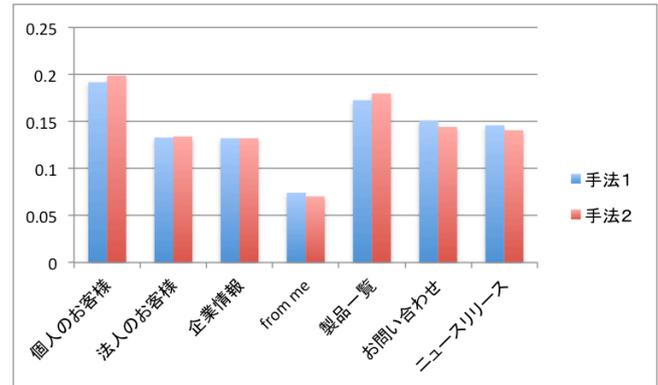


図5 被験者Aの意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

三菱電機 <http://www.mitsubishielectric.co.jp/>

被験者Bはどんなパソコンがあるか、性能や特徴は何か、アクセサリや使用できるアプリケーションはあるか、価格に興味があった(図6)。

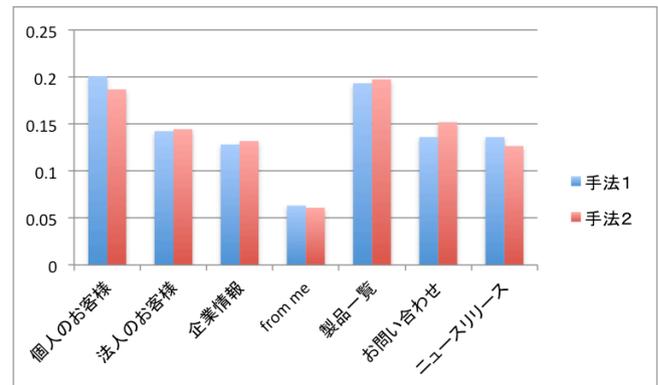


図6 被験者Bの意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

三菱電機 <http://www.mitsubishielectric.co.jp/>

被験者Eはソフトテニスのラケットの購入を検討している(図9)。

被験者Fは10inch程度のタブレット2in1ノートパソコンの比較(金額、スペック、レビュー)を行った(図10)。

8.3 実験の考察

本実験では、被験者の閲覧意図が製品の購入の検討や、スペック比較であるため、基本的に製品情報に関わるグローバルナビゲーションのリンク先のページに類似する内容が記述されているものと考えられる。

被験者Eについて、「PRODUCT」のページに対する類似度

被験者 C はスマホの購入を検討している (図 7)。

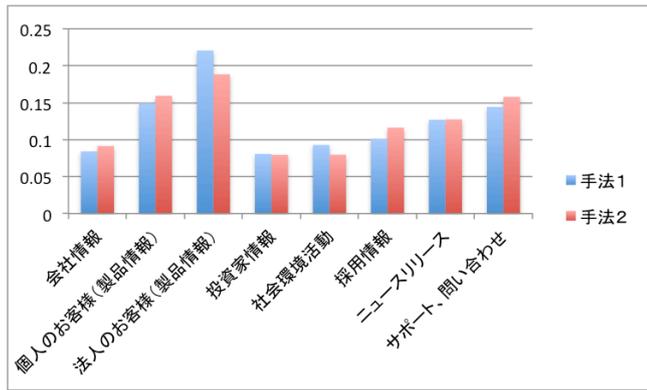


図 7 被験者 C の意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

シャープ株式会社 <http://www.sharp.co.jp/>

被験者 D はノートパソコンの購入を検討している (図 8)。

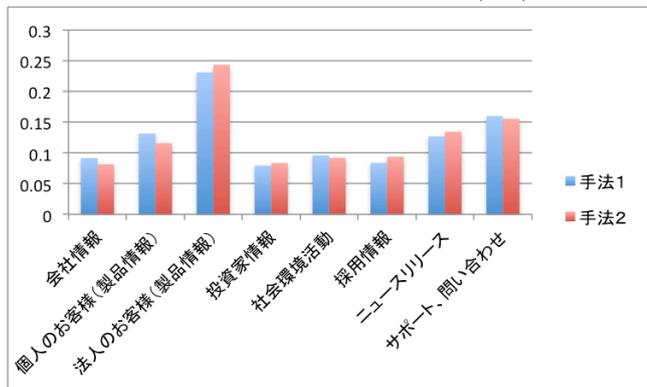


図 8 被験者 D の意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

シャープ株式会社 <http://www.sharp.co.jp/>

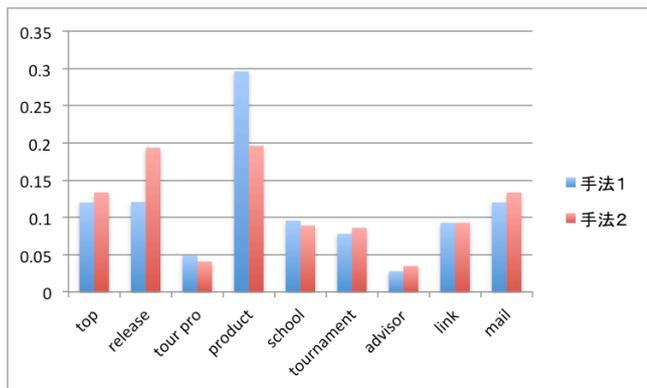


図 9 被験者 E の意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

DUNLOP TENNIS NAVI
<http://www.srixon-tennis.com/index1.html>

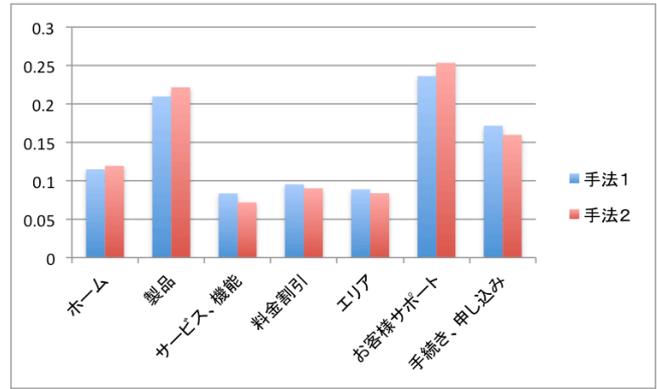


図 10 被験者 F の意図に基づいたグローバルナビゲーションのリンク先のページに対する類似度

NTT ドコモ ホーム <https://www.nttdocomo.co.jp/>

はかなり高くなっている。しかし、全体を通して見てみると、製品紹介がされていると思われるページの類似度はある程度高いが、他のページの類似度と大きな差はない。すなわち、ユーザにとって興味のあるページの推定ができていない。また、手法 1 と手法 2 を比較してもほとんど効果は得られなかった。

今回は閲覧意図ベクトルを全名詞を対象にして TF-IDF 値を計算して単語ごとの TF-IDF 値の平均値を閲覧意図ベクトルとし、類似度の計算を行ったが、ベクトル内の値が小さいものも計算に使用していたことも原因として考えられる。ベクトル内で値が小さいが、意図とはかけ離れた多くの単語に対しても類似度を考慮してしまったため、結果として、類似度が高い単語が均等になってしまったと考えられる。ある一定の閾値を設け、値が大きいものを類似度の計算に用いる必要があると考えられる。

そこで、被験者 A の閲覧履歴から TF-IDF 法で取得した閲覧意図ベクトルと、TF-IPF-SF 法で取得した閲覧意図ベクトルの上位 40 件を表 2 に示す。閲覧意図ベクトルの次元の総数は、閲覧したページに含まれる全ての単語であるため、全ての単語を比較対象として類似度計算をしてしまうと、多くの値の小さいベクトル値によって、類似度に差が現れない可能性があるため、閾値を設けベクトル値の大きい単語を利用する。

表 2 を見ると、ベクトルに含まれる単語には、ほとんど違いがないことがわかる。上位の単語は類似度に与える影響が大きいため、手法 1 と手法 2 では大きな差が得られなかったと考えられる。

また、閲覧意図ベクトルの中の単語を確認したところ、TF-IPF-SF 法では社名などのサイト特有の単語の排除には成功している。

次に、閾値を設けた場合の類似度の変化を検証するため、通常の手法 2 とグローバルナビゲーションに含まれるリンク先ページの特徴ベクトルの上位 20 件と閲覧意図ベクトル上位 5 件の類似度を計算した。その結果を図 11 に示す。

結論として、TF-IDF 値と TF-IPF-SF 値では類似度にほとんど差は生まれなかった。原因として、サイト内のみでよく用

表 2 閲覧意図を表すベクトルの上位 40 件

閲覧意図ベクトル (TF-IDF 法)	閲覧意図ベクトル (TF-IPF-SF 法)
シャープ	製品
製品	情報
情報	商品
商品	スマート
フォン	日
携帯	こちら
電話	サイト
ソニー	月
ライン	サービス
スマート	タブレット
品	年
アップ	公式
フッタ	設定
こちら	お知らせ
機	ページ
サイト	対応
サービス	紹介
年	キャンペーン
月	快適
日	こと
タブレット	入力
公式	購入
対応	無料
機種	ディスプレイ
ページ	環境
設定	コンテンツ
紹介	時
お知らせ	画面
こと	特長
会員	利用
パフォーマンス	カメラ
時	モデル
キャンペーン	名
特長	液晶
ディスプレイ	家電
コンテンツ	ガイド
利用	一覧
カメラ	用
購入	必要
販売	サポート

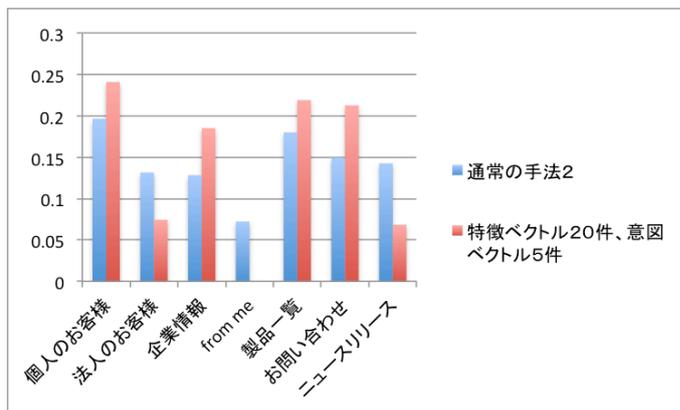


図 11 2種の閲覧意図ベクトルの比較

11 を実例を示す。「個人のお客様」、「製品一覧」、「お問い合わせ」を見ると、すべて類似度としてみると同じくらいユーザの意図と近いと判断できる。実際にページを見ると、「個人のお客様」のページには個人用の家電や電子機器の紹介、「製品一覧」のページには個人向け、法人向けと分けて製品の紹介をするコンテンツが並べられており、製品を購入する、という意図がある場合に有効なページだと判断できる。被験者 A の閲覧意図は製品情報の閲覧であるため、この2つのページはユーザの意図に近いと考察できる。一方、類似度は、「お問い合わせ」のページが高いが、商品の一覧が表示されており、その製品について個別に対応、サポートの案内をするページである。今回は製品情報の閲覧をする、という意図なのでユーザの意図とは異なると考えられる。

ここで、三菱電機のサイトのグローバルナビゲーションに含まれるページの特徴語上位 20 件をまとめた表 3 と閲覧意図ベクトルをまとめた 2 を見ると、どちらにも上位に製品という単語が並んでいる。今回の提案手法では単語レベルでの類似度を計算するため、いずれのページもユーザの意図に近いことになる。確かにどのページも製品について記述されたページであるが、役割が異なるため、ユーザの意図と合うページとそうでないページの対応付けが必要である。

いられる単語は閲覧意図ベクトルから排除することはできたが、それでも上位 40 件のベクトルのほとんど違いがなく、ページ間の類似度の差として現れなかったと考えられる。また、図 11 より、閾値を設けることで類似度に差が生まれることがわかる。閾値を適切に設けることで、閲覧意図の推定の精度を上げることが期待できる。しかし、二分化されてしまい、複数のページがユーザの閲覧意図に類似する、という結果になった。本手法では単語の出現頻度しか考慮しておらず、単語のつながりや流れを一切考慮していないため、ページの中に単語が並んでいるだけで重要な Web ページであるかを判断している。これも類似度が似通ってしまった原因だと考えられる。実際に図

表 3 三菱電機のサイトのグローバルナビゲーションに含まれるページの特徴語上位 20 件

個人のお客様	法人のお客様	企業情報	from me	製品一覧	お問い合わせ	ニュースリリース
機	支社	情報	三菱電機	製品	製品	年
製品	地域	採用	エッセンス	向け	ダウンロード	以下
家電	お客様	取組	テクノロジー	機	問い合わせ	日
たち	三菱電機	環境	五	個人	電話	月
終了	ページ	三菱電機	大陸	お客様	修理	名古屋
情報	エネルギー	企業	届け	法人	質問	局
三菱電機	別	家	広告	順	取扱	株式会社
機器	通信	投資	歴史	一覧	書	兼松
カー	利用	ページ	うま	音	カタログ	ニュース
キッチン	法人	会社	みなさま	三菱電機	三菱電機システムサービス	開発
器	製品	概要	やじ	エネルギー	説明	通信
生産	事例	研究	わたし	カー	購入	お知らせ
お客様	システム	経営	キー	セキュリティー	株式会社	三菱電機
個人	サービス	開発	プロジェクト	器	登録	マイクロ
サイト	事業	グループ	ユーザー	サイト	情報	加熱
エアコン	ビル	事業	南極	ページ	サポート	市
コンテンツ	公共	インターンシップ	好奇	検索	フォーム	波
サポート	分野	決算	心	エアコン	品	発売
テレビ	有効	報告	本質	システム	三菱電機ライブネットワーク	発表
カーナビ	産業	グローバル	理解	こちら	動画	省エネ

9. ま と め

本研究では、ユーザーの閲覧意図を推定し、それを基に目的の情報に類似するページのアンカーリンクをユーザーの目につきやすいように並び替えることによってページの最適化を図る手法について基礎的な検討を行った。

今回は、閲覧したページのみを利用して意図推定を行ったが、今後は時間経過による興味の減衰や、ページの階層を考慮した意図推定のシステムの設計と評価を行っていく予定である。

文 献

- [1] "Google Quick Scroll" <https://chrome.google.com/webstore/detail/google-quick-scroll/okanipcmceoemlbjnmbdibhgpbllgc?hl=ja>
- [2] 横尾俊一、吉浦紀晃、" 検索された Web ページにおける検索語に基づく重要箇所の表示"、情報処理学会研究報告 Vol.2013-IOT-20 No.4 2013/3/14
- [3] Bongwon Suh, Allison Woodruff, Ruth Rosenholtz, and Alyssa Glass. "Popout prism: Adding perceptual principles to overview+detail document interfaces." . Proc. CHI 2002,pp.251-258.
- [4] 井桁正人、寺田実、丸山一貴、" Scout View : Web ページにおけるナビゲーションインターフェース"、情報処理学会研究報告 Vol.2009-HCI-133 No.8 2009/5/16
- [5] 顔洪、牛尼剛聡、" スマートフォンでの効率的な商品選別を目的としたユーザの振る舞いに基づく閲覧リスト最適化手法"、情報処理学会論文誌 データベース Vol.8 No.4 1-15(Dec. 2015)
- [6] 木本亮司、市村哲、" ブラウザ操作履歴に基づいた Web サイト改善ツール"、情報処理学会研究報告 Vol.2011-GN-79 No.23 2011/3/17
- [7] "F-Shaped Pattern For Reading Web Content" <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>
- [8] 奥村学著 コロナ社 : 自然言語処理の基礎 pp.117-119