

情報検索における視線情報を用いた適合性推定

大沼 俊輔¹ 藤井 敦¹

概要：情報検索ではユーザの情報要求に適合する文書を発見するため、文書の適合性推定に関する研究が重要である。適合性推定について、ユーザの入力する検索質問に基づく研究やユーザの文書の閲覧行動に注目する研究がある。本研究はユーザの文書の閲覧行動に注目し、視線情報を用いて適合性を推定する手法を提案する。提案手法は、閲覧行動の個人差の度合いが文書の適合性により異なる点を利用する。本研究で扱う文書の適合性は、ユーザの望む情報を含む適合、含まない不適合、一部を含む部分適合の3種類である。実験協力者を雇い複数の文書に対して適合性判定をしてもらい、その際に記録した視線情報に基づいて提案手法の有効性を評価した。

1. 序論

情報化社会の発展にともないインターネットを介して多種多様で大量の情報が利用可能になった。なかでも自然言語テキストは主要な媒体の一つであり、大量の文書から情報要求に適合する文書を探し出す情報検索技術の重要性は急速に高まっている。本研究では情報要求に対する文書の適合性を推定する手法を提案する。

本研究で扱う文書の適合性は次の3種類とする。

- 適合:ユーザの望む情報を含んでいる
- 部分適合:ユーザの望む情報の一部を含んでいる
- 不適合:ユーザの望む情報を含まない

ユーザは検索システムが提示した文書を閲覧し求める情報を探す。このためユーザの閲覧行動は文書の適合性について示唆に富んでいると考えられ、閲覧行動から文書の適合性を推定する研究が盛んである。

閲覧行動の代表的な素性に文書の表示時間やマウスカーソルの挙動、スクロールの文量、視線情報がある。

Claypool ら [2] は文書の表示時間やスクロールの文量がユーザの文書への興味と相関があることを示した。一方で Guo ら [3] は表示時間は長くても文書は不適合であり得ると考えた。そこで文書を閲覧し始めた後のマウスカーソルの挙動に注目し適合性の推定、再ランキングについて性能を向上させた。

視線情報に注目する研究では、Gwizdka [4] はユーザの適合性判定時における認知的努力、つまりよく閲覧しようとする行動が部分適合、適合、不適合の順に強く観察されたと報告し、これに基づく適合性推定を行っている。しか

しこれは全体的な傾向であり、閲覧行動の程度は各ユーザにより異なる。そのため Buscher ら [1] は異なるユーザの視線情報の素性値をそのまま扱わず、ユーザごとに正規化して適合性の推定を行っている。

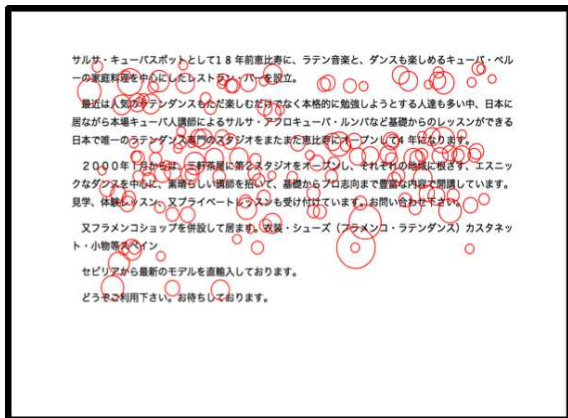
以下、本研究の着想について述べる。

Gwizdka の報告する閲覧行動の傾向は適合性判定の難度の傾向と見なせる。つまり部分適合、適合、不適合の順に適合性判定が難しいと考える。

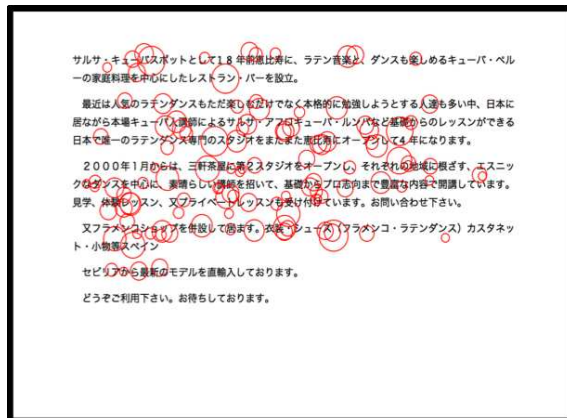
また適合性判定の難度と観察される視線情報の個人差に関係があると考えられる。判定の難しい文書ではユーザ全員が慎重に判定するため個人差は観察されない一方で、判定が簡単な文書については、依然慎重に判定するユーザと判定を簡素化するユーザの間に視線情報の個人差が観察されると考える。これより、文書の適合性と視線情報の個人差には相関がある可能性がある。具体的には部分適合文書では個人差が観察されにくく、不適合文書では観察されやすいと考える。

二人のユーザ A, B が部分適合文書と不適合文書を閲覧し適合性判定をした際の視線情報の実例を図1, 2に示す。この視線情報は3節で述べる評価実験で記録したものである。この図の赤い円の中心がユーザの注視した点であり、注視時間は円の大きさで表される。ユーザが注視している点を注視点と呼ぶ。注視点の様子を見ると、部分適合文書はいずれのユーザも同程度に閲覧している様子がわかる。一方で不適合文書についてはユーザ A とユーザ B の注視点の数が大きく異なり、ユーザ B はユーザ A と比べると閲覧していない様子がわかる。この例では、部分適合では個人差が小さく、不適合では個人差が大きくなることが予想できる。

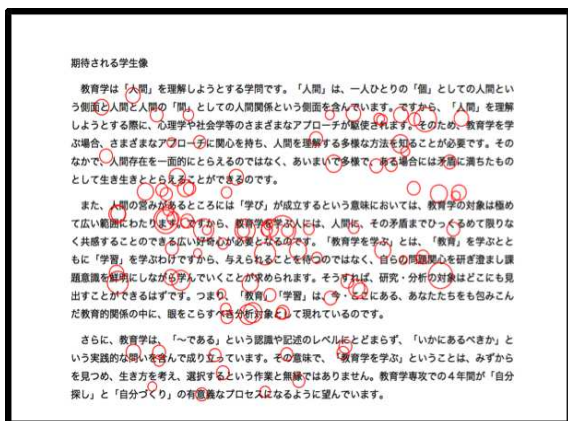
¹ 東京工業大学
Tokyo Institute of Technology



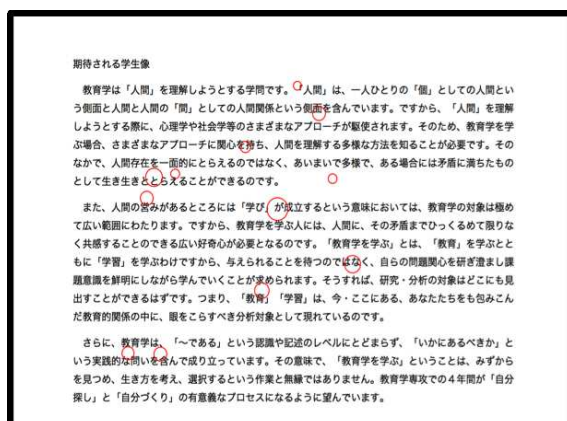
(a) 部分適合文書



(a) 部分適合文書



(b) 不適合文書



(b) 不適合文書

図 1 ユーザ A の視線情報の様子

図 2 ユーザ B の視線情報の様子

この着想に基づき、本研究では視線情報の個人差から文書の適合性を推定する手法を提案する。ユーザが適合性判定を行う際の視線情報を記録し、提案手法の評価実験を行い有効性を示す。

2. 提案手法

提案手法の全体像を図3に示す。以下、この図に沿って説明する。

提案手法は複数のユーザが同じ文書に対し適合性判定する際の視線情報に注目して、当該文書の適合性を推定する。文書の適合性は情報要求に対し定義されるので、比較されるユーザは同じ情報要求を認識している必要がある。

また、情報要求と文書について信頼性の高い適合性判定を正解として用意する。各ユーザが主観的に判定した適合性は本研究では扱わない。

対象の文書について複数のユーザが適合性判定を行う際の視線情報を得たら、注視点を抽出する。注視点の抽出方法は3.3小節で説明する。

人間の眼球運動には注視点を動かすためのサッケードがある[6]。しかしサッケード中はほとんど情報が認識されないため、本研究では視線情報としてディスプレイ上の注視点座標の時系列データを扱う。

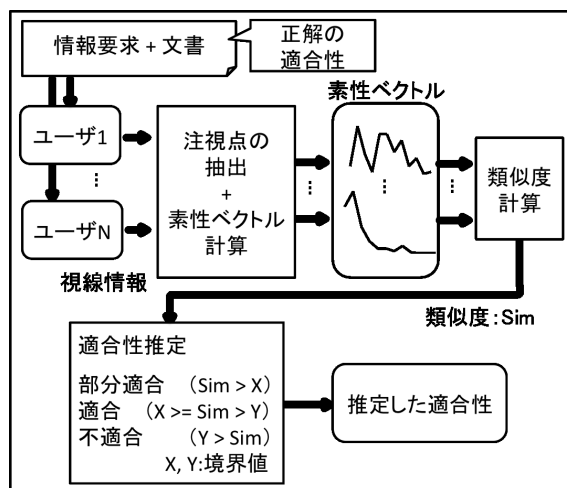


図 3 提案手法の全体像

次に視線情報について、閲読行動を表現する素性ベクトルを計算する。素性ベクトルは次の3種類を扱う。

- 移動方向:各方向への注視点の移動の発生頻度。10°ごとに方向を分けて移動回数を数える
 - 移動距離:各距離の注視点の移動の発生頻度。移動距離を100pixelごとに階級に分けて移動回数を数える
 - 注目箇所:文書の各行にある注視点の数
- 2ユーザの素性ベクトルについての個人差は、コサイン

類似度を用いて表す。類似度と個人差は極性が反転するので、個人差が大きいとき類似度は小さくなる。N ユーザについては、N 人中から 2 人を選ぶ組み合わせについてそれぞれコサイン類似度を計算し、平均類似度を計算する。

計算された類似度が大きければ当該文書は部分適合、小さければ不適合、中間だったら適合と推定する。この推定結果が正解の適合性と一致すれば推定は成功である。適合性の境界値となる類似度の値を二つ設定すれば 3 クラスの適合性の推定が可能であり、一つ設定すれば 2 クラスの適合性の推定が可能である。

本手法の展望として、検索システムへの応用を考える。

視線情報を用いるため、文書を読んだ際の視線情報の記録が利用可能であるとする。複数のユーザが似た情報要求について検索を行い、同じ文書を読んだ際の記録がある場合、本手法により当該文書の適合性を推定できる。これにより将来、似た情報要求について検索されたとき適合性の高い文書を提示することが可能である。このような応用方法ではユーザの情報要求の推定技術と組み合わせる必要がある。

別の展望としてテストコレクション構築支援を考える。情報検索に関わる研究において、大量の文書について信頼性の高い適合性の情報が必要である場合がある。そのためテストコレクションの構築では、各文書について情報要求と照らし合わせて人手で適合性を判定することになる。これについて視線情報から適合性を推定することによって、判定者の負担軽減、判定結果の信頼性の向上などが望める。

提案手法の評価実験では、提案手法の前提となる適合性と視線情報の類似度の関係について調査し、次に提案手法の有効性を検証する。

3. 評価実験

3.1 評価実験で用いた文書

評価実験で用いた文書は、NTCIR-3 WEB[5] のテストコレクションから選択した。このテストコレクションは Web 検索に関する実験評価用コレクションである。

このコレクションは以下の内容から構成される。

- 検索対象の文書
- 情報要求を記述した検索課題
- 各検索課題に関する適合性判定の結果

評価実験ではこの検索課題のうち 3 件を情報要求として設定し、各検索課題について 4 文書の計 12 文書を利用する。

検索課題は図 4 の形式で記述されている。この情報のうち、DESC タグで示される情報要求の記述、NARR タグで示される検索の背景・目的、語の定義、適合性判定基準の補足をユーザに提示する。

コレクションでは各検索課題について文書が高適合、適合、部分適合、不適合の 4 種の適合性で判定されている。

```
<NUM>0008</NUM>
<TITLE CASE="b">サルサ, 学ぶ, 方法</TITLE>

<DESC>サルサを踊れるようになる方法が知りたい</DESC>
<NARR><BACK>最近はやっているサルサという踊りを学ぶためにどうすればよいか具体的な方法が知りたい。例えば教室に通うという場合には、その場所や授業形態など、具体的な内容を必要とする。</BACK><RELE>具体的な方法の表記のない、流行であることのみを扱った文書は不適合とする。</RELE></NARR>

<CONC>サルサ, 習う, 方法, 場所, カリキュラム</CONC>
<RDOC>NW011992774, NW011992731, NW011992734</RDOC>
<USER>大学院修士1年, 女性, 検索歴2.5年</USER>
</TOPIC>
```

図 4 検索課題の形式 [5]

しかし高適合の文書は検索課題によっては存在しないため、本研究では高適合な文書と適度な文書をあわせて「適合」であると扱った。

手法の評価のため一つの検索課題について少なくとも適合・部分適合・不適合の 3 件の文書が必要である。さらに情報要求に関連がありそうなキーワードの存在によりユーザの閲覧行動が変わる可能性を考慮し、不適合文書においてキーワードを含むもの、含まないものの 2 通りを用意した。適合文書、部分適合文書はキーワードを含むものを抽出している。これは検索システムのアルゴリズムとして単純なキーワードの一致を前提としたためである。キーワードは、コレクションの検索課題にあらかじめ設定されている擬似的なクエリの情報(図 4 の TITLE タグ)を採用した。

そのため用意した文書は各検索課題について次の 4 件である。

- 適合 : Relevant, R
- 部分適合 : Partially Relevant, P
- 不適合 : Non Relevant, N
- 不適合(キーワードを含まない) : Non Keywords, K

以下、文書の適合性を R, P, N, K で表す。今回の実験では、コレクションで判定されている適合性を文書の適合性の正解として評価を行った。

各検索課題について 4 文書を選択した基準は、スクロールの必要なく全文表示できること、表示領域の半分以上の表示になること、表示内容が単語・数字の羅列ではなく文章になっていることとした。しかし最後の条件については文書数を確保するため、文書の行数の三分の一行までは含んでよいとした。この条件を満たす文書を各検索課題で識別番号が若い順に 4 件選択した。また文書を 4 件選択できる検索課題について識別番号が若い順に 3 件選択した。

3.2 視線情報の記録

選択した文書について、提案手法の評価に用いる視線情

報を記録した。視線情報の記録は2回行い、それぞれ男性9名、計18名に協力してもらった。協力者は通信情報工学、計算工学の専攻に所属する大学院生である。実験は最大1時間を予定し、謝礼は各人に1000円支払った。以下、協力者のことをユーザと呼ぶ。

視線情報の記録にはTobii T60 (以下、Tobii) を用いた。Tobii は赤外線を照射し、その反射光を用いて視線の向きを測り、ディスプレイ上の視点の座標を計算する。そのため、ユーザは通常のディスプレイの使用時と同様の環境で文書を読める。ディスプレイとユーザの距離は60cmとした。表1にTobiiの仕様を示す。

表1 Tobii T60の仕様

項目	仕様
ディスプレイ	17inch TFT
解像度	1280 × 1024
サンプリングレート	60 Hz
正確度	0.5°

視線情報の記録にあたりユーザに課したタスクを説明する。協力者には文書の適合性を、適合・部分適合・不適合のいずれかで判定し、理由とともに判定結果を回答してもらった。また、適合性判定が済んだ時点で回答するように指示し、文書の閲読を開始してからユーザが回答を行うまでの視線情報を計測した。適合性の判定時は、他に見た文書とは独立に判定するよう指示した。

表示される文書の文字の大きさは誤差を考慮し24pixelとした。改行時は行間として18pixelが挿入される。つまり1行あたり42pixelの高さになる。また、段落形成時にはさらに18pixelの空白が挿入される。表示領域については、画面端において計測の精度が落ちるため上下左右の100pixelは文書を表示しない領域とした。

各文書の適合性の流れを説明する。1件の検索課題について、最初に検索課題の内容を表示して確認してもらった。検索課題の内容の理解を促すため検索課題の表示の後に課題内容についての質問を表示し、確認してもらった後に文書の適合性判定に移ってもらった。文書の適合性判定が済んだら音声による回答を行ってもらい、準備ができた時点で次の文書に移ってもらった。これを4文書分繰り返した。

以上の流れを3件の検索課題について繰り返してもらった。表示内容の切り替えはユーザのキー入力により行った。

また、検索課題の内容はいつでも確認できるようにした。文書の表示中に検索課題の確認が行われた場合、確認中の視線情報は解析から除いた。

文書の提示順について説明する。最初の9名の視線情報の記録では、すべてのユーザに同じ順で文書を提示した。同じ文書における視線情報を比較するにあたり、その文書を見るまでの実験上の条件を同じにするためである。提示順は乱数で表2のように決定した。

表2 ユーザに提示した文書の適合性の順番

文書の表示順	1	2	3	4
検索課題1	R	K	P	N
検索課題2	K	P	N	R
検索課題3	N	P	K	R

次の9名についてはユーザごとに提示順を変更した。これは、例えば最初の文書では適合性判定のタスクの不慣れさ故に皆慎重になるというような順番の影響を考慮したためである。評価実験の結果はおよそ同じ傾向であったため、本論文では提示順を固定した9名の視線情報について結果を示す。

3.3 データの処理

注視点の抽出について説明する。Tobiiが計測し記録する情報はユーザが視ているディスプレイ上の座標である。この座標について、100ms以上、半径16pixelの円内に連続する視点が取まる場合、連続する視点をまとめて注視点として抽出した。注視点の座標は円内にある視点の重心とした。

注目箇所の計算では、注視点のある行の判別が必要である。これは注視点座標が各行を中心とする42pixelに取まるかどうかで判定した。

次に記録した視線情報に施すエラー処理について説明する。ある時刻の視線情報が計測出来ない場合、エラーが出力される。このエラーが長時間にわたる場合、そのエラー前後の注視点を連続するものとは扱わないことにした。基準として500msを設定した。

計測時にエラーでなくとも、注視点抽出処理をするとウィンドウ外に注視点ができることがある。この場合は当該注視点を除外し、その前後の注視点は連続性が無いものとして扱う。

以上のエラー処理を施した上で、次の条件を満たすものを不適当な視線情報として解析より除いた。

- 適合性判定タスクにおいて指示に違反しているもの
- 視線情報のうちエラー率が20%以上のもの(500ms以上のエラーは除く)
- 注視点が4点未満のもの

このエラー率、注視点数の基準はGwizdkaと同じ条件である。これによりではR文書で4件、P文書で3件、N文書で4件、K文書で3件の視線情報を除外した。

次に注視点の連続性の扱いについて説明する。注視点の移動方向など、抽出した注視点について連続性を考慮する素性がある。しかし、次の場合は上記した通り、その前後の注視点到連続性がないとみなす。

- 検索課題の確認による文書の閲読の中止
- 長時間のエラーの発生
- ウィンドウ外の注視点の除外

連続性がないとされる箇所がある場合、視線情報はその箇所にて複数の視線情報に分割されると考える。この分割された視線情報については個別に素性値を計算し、全体で素性値を合計した。

以上で評価実験の手続きの説明を終える。次に視線情報の類似度と適合性の関係についての解析結果と提案手法の評価結果を述べる。

3.4 視線情報の類似度と適合性の関係の解析結果

記録した視線情報を文書の適合性ごとに解析し、類似度と適合性の関係について調査した結果を説明する。

各素性について各検索課題・各文書について9ユーザから考えられるペア36組の類似度の平均を示す。ただしペアの一方の視線情報がエラー処理で除かれる場合は類似度を考えない。また各文書における類似度の群間の分布の差についてKruskal-Wallis検定を行った。信頼区間は95%とする。

表3に注視点の移動方向についての平均類似度を示す。

表3 注視点移動方向について各検索課題、各適合性の文書ごとの類似度の平均

平均類似度	R	P	N	K
検索課題1	0.84	0.93	0.65	0.74
検索課題2	0.92	0.90	0.90	0.82
検索課題3	0.84	0.92	0.80	0.83

検索課題1, 3については $P > R > K > N$ という傾向である。しかし検索課題2については $R > P = N > K$ となっており、 P, R の関係、 N, K の関係は逆になっている。Kruskal-Wallis検定の結果、すべての検索課題で群間において分布の差が認められた。

表4に注視点の移動距離について平均類似度を示す。

表4 注視点移動距離について各検索課題、各適合性の文書ごとの類似度の平均

平均類似度	R	P	N	K
検索課題1	0.84	0.94	0.83	0.81
検索課題2	0.93	0.91	0.86	0.78
検索課題3	0.85	0.93	0.85	0.84

検索課題1, 3では $P > R > N > K$ という傾向がある。しかし検索課題2は P, R の関係は逆になっている。検定結果からすべての検索課題で群間に分布の差が認められた。

表5に注目箇所について平均類似度を示す。

表5 注目箇所について各検索課題、各適合性の文書ごとの類似度の平均

平均類似度	R	P	N	K
検索課題1	0.71	0.80	0.60	0.65
検索課題2	0.68	0.81	0.77	0.74
検索課題3	0.52	0.75	0.75	0.71

検索課題1については $P > R > K > N$ という傾向がある。検索課題2, 3については $P > N > K > R$ という傾向である。検定結果からすべての検索課題で群間に分布の差が認められた。

以上より、全体として部分適合の類似度が高い結果である。また、注視点の移動方向、移動距離の類似度では検索課題1, 3において $P > R > N, K$ という傾向であり、Gwizdkaの結果から類推した仮定と矛盾しない。ただし、検索課題2のR文書、P文書の関係が全体の傾向の反対の関係が見られた。これについては3.6小節で原因について考察する。またN文書、K文書については共通する傾向は見いだせなかった。

3.5 類似度に基づく適合性推定手法の評価

次に提案手法の評価結果を説明する。

提案手法による推定の種類は次の2種類とした。

- 3クラス:K文書を除く3種の文書について、3種の適合性から適合性推定する
- 2クラス:用意した4種の文書について、適合性を適合・不適合の2種から適合性推定する

3クラスの推定においてK文書を除いたのは、文書数の偏りをなくすためである。すべての文書を用いると文書数の比が適合:部分適合:不適合=1:1:2になってしまうため、一方の不適合文書のみ扱った。K文書以外の文書はキーワードを含むという条件は同じであるため、不適合文書としてN文書を採用した。また2クラスの推定では、R, Pを適合、N, Kを不適合と扱う。

視線情報の類似度は最小で2人のユーザがいれば定義できるので適合性推定はユーザのペア(ユーザペア)を単位として行う。

あるユーザペアの視線情報の類似度について、12文書の適合性の推定を行い正解率を計算する。このとき適合性の境目となる類似度の値を当該ペアのユーザを含まないユーザペアの類似度から学習する。

具体的には、当該ユーザを含まないユーザペアの類似度データにおいて最も誤分類が少なくなる境界値を求め、その境界値を当該ユーザペアによる12文書の類似度に対して適用して正解率を計算する。

ユーザペアは9名のユーザについて36組考えられるので、全体の正解率は36個の正解率のマクロ平均で計算する。

提案手法による適合性推定を行った結果の正解率を示す。正解率の検定は無作為に適合性を選択した場合正解率は2クラスの推定で0.5、3クラスの推定で $1/3=0.333\dots$ であるため、それぞれの36組の正解率と0.5、0.333\dotsの一樣分布の間で符号検定を行った。信頼区間は95%である。*は p 値 < 0.05 を意味する。

結果を表6に示す。

素性の中では移動方向、移動距離、注目箇所の順に正解

表 6 提案手法による適合性推定の正解率

	3クラス	2クラス
移動方向	*0.55	*0.68
移動距離	*0.50	*0.64
注目箇所	*0.42	0.54

率が高い。また注目箇所による2クラス推定を除き無作為に適合性を選択する手法より正解率が高く、有意差があった。そのため、視線情報の類似度に基づく適合性推定は有効である。

3.6 考察

検索課題1, 3においてはP文書の類似度 > R文書の類似度となる傾向があった。しかし検索課題2においてはこの関係は逆転している傾向があった。この現象について考察する。

ユーザによる適合性判定では、検索課題2のR文書について適合と答えたユーザは3人であり、部分適合と答えたユーザは9人であった。この検索課題と文書について内容を確認したところ、適合とするための条件を充たしていないと判断できる内容だった。また適合でないと判断した15人中13人が、条件が充たされないことを判定理由に挙げていた。以上を踏まえると、この文書の適合性は適合でないとしたほうがよい可能性がある。

本研究では文書の適合性としてNTCIR-3 WEBの評価用コレクションの適合性を信頼しそのまま用いた。情報検索システムの評価用のデータであるので、コレクション全体としては信頼性は高いはずである。しかし上記したような文書とその適合性について考えると、個々の文書については与えられる適合性判定を完全には信頼できない可能性がある。

また、素性の注目箇所があまり有効でなかった理由として視線計測の縦方向の誤差を考える。他の素性は注視点間の相対的な情報に注目しているところ、注目箇所は注視点の絶対的な座標より計算される。そのため縦方向に常に同じ程度の誤差、例えば下方向に1行分の誤差が発生すると、素性の対応する要素がずれるので同じ行同士の注視点数の比較ができなくなることが考えられる。

実際に空白部分を視ていると計算されるような視線情報があった。これは視た箇所から誤差によりずれた注視点計算されたのかもしれない。しかし、実は誤差はなく空白部分を見ていたという可能性は排除できないため、一様の補正をかけることで対応するのは難しい。

この点に対する改善は、ほかの素性と同様に注視点間の相対的な関係により注目箇所を表現することが考えられる。

4. 結論

本研究では情報検索における複数ユーザの視線情報の個

人差に基づく文書の適合性を推定する手法を提案した。

手法の評価のため実験協力者を募り文書に対し適合性判定を行ってもらい、その際の視線情報を記録した。閲覧行動を表す視線情報の素性ベクトルとして注視点の移動方向、移動距離、注目箇所を扱った。また視線情報の個人差を素性ベクトルのコサイン類似度で表現し、類似度と適合性の間に関係があることを分析により示した。移動方向、移動距離について特に観察された傾向として部分適合文書、次いで適合文書、最後に不適合文書の順に類似度が高い結果であった。

視線情報の類似度に基づく適合性推定では、無作為に適合性を選択する手法よりも正解率の高い推定が可能であることを示した。視線情報を表す素性として移動方向、移動距離が優れていた。

本研究において残された課題を述べる。

本研究の実験協力者は全員男性で、工学系の大学院生である。一般的なユーザについて手法の有効性を検証するには、女性についてほぼ同数のデータが必要であろう。また年齢や受けた教育などについて幅広い性質を確保すべきである。

今回の実験では検索課題と文書の件数が少ない。そのためそれぞれの件数を増やして手法の検討を行うことが必要である。

参考文献

- [1] Buscher, G., Dengel, A., Biedert, R. and Elst, L. V.: Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond, *ACM Transactions on Interactive Intelligent Systems*, Vol. 1, No. 2 (2012).
- [2] Claypool, M., Le, P., Wased, M. and Brown, D.: Implicit interest indicators, *IUI '01 Proceedings of the 6th international conference on Intelligent user interfaces*, pp. 33-40 (2001).
- [3] Guo, Q. and Agichtein, E.: Beyond Dwell Time: Estimating Document Relevance from Cursor Movements and other Post-click Searcher Behavior, *WWW '12 Proceedings of the 21st international conference on World Wide Web*, pp. 569-578 (2012).
- [4] Gwizdka, J.: Characterizing Relevance with Eye-tracking Measures, *IIX '14 Proceedings of the 5th Information Interaction in Context Symposium*, pp. 58-67 (2014).
- [5] 江口浩二, 大山敬三, 石田栄美, 神門典子, 栗山和子: NTCIR-3 WEB: Web 検索のための評価ワークショップ, *NII journal*, Vol. 6, pp. 31-56 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/80015963005/>) (2003).
- [6] 大野健彦: 視線から何がわかるか 視線測定に基づく高次認知処理の解明, *認知科学*, Vol. 9, No. 4, pp. 565-579 (2002).