

英語多肢選択問題解答時の視線に基づく確信度推定

山田 健斗[†] 大社 綾乃^{††} 藤好 宏樹^{††} 星加 健介^{††} OlivierAugereau^{††}

黄瀬 浩一^{†††}

[†] 大阪府立大学工学域

〒 599-8531 大阪府堺市中区学園町 1-1

^{††} 大阪府立大学大学院工学研究科

〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: [†]{kento,okoso,fujiyoshi,hoshika}@m.cs.osakafu-u.ac.jp, ^{††}augereau.o@gmail.com,

^{†††}kise@cs.osakafu-u.ac.jp

あらまし 英語学習において、復習は重要な学習プロセスである。十分な復習を行うためには、間違えた問題だけでなく、解答が正解であると確信をもてなかった問題も見落としとしてはならない。そこで本研究では、問題解答時の確信度を推定し、解答の正誤と合わせて解答者に提示するシステムの実現を目指す。このシステムの実現によって、見落としがなく、効率的で優先順位を考慮した復習が可能となる。人は読んである文書を難しいと感じると、読む速度が遅くなる、読み返しが起こるといった特徴が見られる。このため、確信度の推定には視点情報を利用する。本稿では、被験者 11 名に文法や語彙を問う英語の四択問題に解答してもらい、その際の問題に対する確信の有無について推定を行った。実験においては、被験者依存で学習を行い、推定に SVM を使用した。その結果、被験者 11 名の平均推定精度は 89.8%であった。

キーワード アイトラッカ、学習補助、SVM

1. はじめに

近年はグローバル化の影響から、英語学習の必要性が非常に高まっている。英語学習では、単語や文法を暗記し、テストを通して学習の成果を確認した後、復習するという流れが一般的である。復習することで、新たな知識を獲得し、曖昧な知識を補完できる。そのため、復習は非常に重要な学習プロセスであるといえる。

一般的に、人々は間違えた問題に対して復習する。しかし、正解した問題にも復習が必要な場合がある。具体的には、自身の解答が正解であると確信がもてずに解答し、偶然正解した場合である。このような問題は、正解したことに気を許してしまい、復習の機会が失われがちである。しかし、復習する際、このような問題を見落としとしては、十分な学習効果が得られない。すなわち、間違えた問題だけでなく、解答が正解であると確信をもてなかった問題に対しても復習を行う必要がある。さらには、復習する問題の中でも、優先順位をつけることが重要である。問題に間違える際には、その背景に知識の記憶違いが存在する。例として、確信をもって解答したにも関わらず間違えた問題では、正解に必要な知識を勘違い、もしくは記憶違いしていることが考えられる。この勘違いや記憶違いが深刻であるほど、より多くの問題に間違える原因となりうる。そのため、このような勘違いや記憶違いに対する復習を最優先で行う必要が

ある。これにより、確信を持って間違えたものを最優先とする点、確信なく正解したものを復習に含める点で、通常の復習とは大きく異なり、効率の良い復習に繋がる。以上より、解答時に確信をもって解答したかどうか分かれれば、見落としがなく、優先順位を考えた復習が期待でき、学習において大変有益である。

解答に対する確信度を考慮に入れた復習の方法として、解答時に確信をもてなかった問題に印を付け、後からその箇所を復習するというものがある。しかし、この方法には、印を付けながら解答するために問題に集中できず、学習の成果を確認するというテスト本来の役割を十分に果たせない問題がある。これに対して、問題解答時の確信度を自動で推定し、解答者にフィードバックすることができれば、問題解答時の振る舞いを妨げることなく復習に活用できるため、有益である。そこで、本研究では、上記の問題を解決するため、確信度を自動で推定し、解答の正誤と合わせて解答者に提示するシステムの実現を目指す。

確信度を推定する方法として、視点情報の利用に着目する。これまでの研究から、視点情報には言語能力や確信の程度に関する情報が含まれる事が分かっている [1] [2]。例えば、人は読んである文書が難しいと感じると、読む速度が遅くなったり、何度も読み返す傾向がある。

確信度を推定する対象として様々なものが考えられる。本研究では、確信度推定の第一ステップとして、多肢選択問題を取

り扱うこととする。その理由は、多肢選択問題では問題文や複数の選択肢といった見るべき対象が多く存在するため、確信度に応じた視点情報が如実に現れると予想されることである。以上より、確信度を推定するために視点情報を利用し、推定の対象には多肢選択問題を採用する。視点情報はアイトラッカというデバイスによって容易に取得できる。さらに、問題解答時の妨げになりづらいという利点がある。

そこで、本稿では、解答者の視点情報を用いて、多肢選択問題に対する確信度を推定する手法を提案する。提案手法では、問題解答時の視点情報から特徴量を抽出し、SVMを用いて学習した後、識別する[3]。提案手法の有効性を検証するため実験を行い、推定精度を算出する。提案手法による推定精度は89.8%であった。本稿では、実験によって得られた結果について考察する。なお、本研究は、大阪府立大学工学研究科倫理委員会の承認を得ている事を付記しておく。

本稿は、第2節が関連研究、第3節が提案手法、第4節が実験、第5節がまとめという構成である。

2. 関連研究

視点情報が文書の読み方や言語習熟度、理解度と関係があることがこれまでの研究から分かっている。本節では、読書時の視点情報や問題解答時の確信度に関する関連研究について述べる。

視点情報と言語習熟度に関する研究について、藤好らの研究[4]や、吉村らの研究[5]がある。藤好らは、アイトラッカを用いて得られた視点情報とテストの問題の正答数から、ユーザの英語習熟度を推定する手法を提案している。この推定手法により、TOEICスコアを平均絶対誤差36.3点で推定することができる。また、吉村らは英文書を読む際の視点情報を利用して、その文書に対するユーザの理解度を推定する手法を提案している。この手法では、fixation(注視点)とsaccade(注視点間の素早い目の動き)の位置や方向、出現割合など8個の特徴量を用いている。しかし、十分な推定精度を得るに至っていない。他に、大社らの手法[6]に、視点情報を利用してユーザが難しいと感じた文書内の箇所を推定するものがある。難しいと感じた箇所を推定するために、読み返しの回数や読む速さなどの特徴量を用いているが、精度は十分でない。

問題解答時の視点の振る舞いに関する研究に、Tsaiらの研究[7]や、Ishimaruらの研究[8]がある。Tsaiらは、四択問題における問題と選択肢それぞれに対する時間のかけ方や、正答者と誤答者との視点の動きに違いがあることを示した。Tsaiらの研究により、問題で問われているものに関係が深い箇所ほど長時間注視され、誤答者は正答者と比べて問題文に読み戻す傾向が強く見られることが分かった。Ishimaruらは、物理に関するテキストを読み、問題を解く際の視点情報を、学習者の物理に対する習熟度で分けて調査した。この研究により、学習者の理解の深さと、テキストを読む際の視点情報や時間のかけ方には相関があることが明らかになった。

問題解答時の確信度に関する研究として、小島らの研究がある[2]。小島らは雑学や一般常識に関する四択問題を取り扱っ

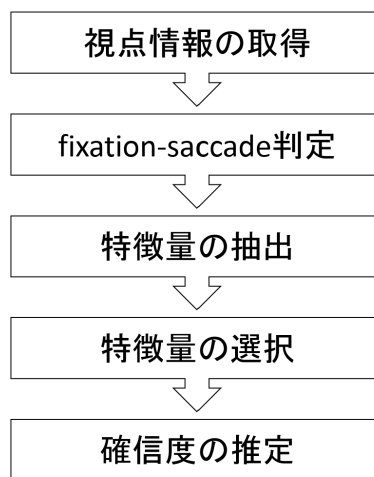


図1 提案手法の流れ

ており、解答者がどの程度確信をもって解答したのかを確信度と定義している。小島らは、解答者の視点情報から、確信度を推定する手法の考案を目的に、解答時の視点の動きについて調査した。これによって、確信度に応じて視点の動きに一定のパターンがあることを明らかにした。しかし、確信度を推定するには至っていない。

3. 提案手法

3.1 提案手法の流れ

図1に提案手法の流れを示す。まず、ディスプレイ上に英語の問題を表示し、解答する際の視点情報をアイトラッカで取得する。次に、得られた視点情報に対して、fixation-saccade判定を行う。fixation-saccade判定後の視点情報を利用して、特徴量を抽出する。抽出した特徴量から、確信度の推定に有効なものを選択し、SVM(Support Vector Machine)を用いて確信度を推定する。ここで、確信度としては最も基本的なもの、すなわち、確信の有無を用いる。これらの手法の内、特徴量の選択と確信度の推定は、被験者ごとに行う。その理由は、問題解答時の視点情報は、被験者の癖や問題に対する戦略などに左右されると予想されることである。以下に詳細を述べる。

3.2 視点情報の取得

アイトラッカを用いてユーザの眼球運動を測定し、視点情報を取得する。アイトラッカとは、赤外線を照射することでユーザの瞳孔の位置を捉え、眼球運動を測定する装置である。本手法では、Tobii EyeXという図2のようなディスプレイ装着型のものを用いる。これにより、ディスプレイ上に表示された問題を解答している際の視点情報を取得する。

3.3 fixation-saccade判定

眼球運動はfixation, saccade, blinkの3種類に分類される。fixationは視点が一定時間留まることを表し、saccadeはfixation間の素早い視点の移動を表す。blinkは、瞬きのことである。人はfixationとsaccadeを繰り返すことで物を見たり文章を読んだりしている。本手法ではBuscherらの手法[9]を用いてfixation-saccade判定を行う。



図 2 Tobii EyeX

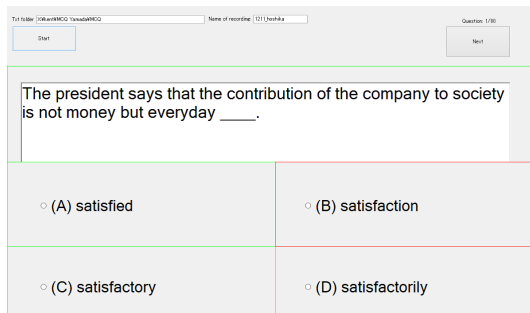


図 3 矩形の設定

3.4 特徴量の抽出

fixation-saccade 判定後の視点情報から、特徴量を抽出する。使用する特徴量を表 1 に示す。特徴量は、fixation 関連のものが 14 個、saccade 関連のものが 14 個、問題の解答時間を表すものが 1 個、合計 29 個である。一般的に、ユーザが問題を解答する際の振る舞いとして、悩んだ選択肢に視点が集まり、そうでない選択肢には視点が集まらない傾向がある。そこで、fixation と選択肢との対応を取るために、各選択肢を囲う矩形を図 3 のように設定する。この矩形に含まれる fixation は、その選択肢、あるいは問題文に対するものとして判定する。アイトラッカの精度の都合上、視点が対象の文字列から多少ずれることを考慮し、矩形の大きさに余裕を持たせてある。

3.5 特徴量の選択

表 1 に示した 29 個の特徴量は、確信度の推定を行う上で有効であるか自明ではないため、特徴量を選択する必要がある。そこで、特徴選択手法として遺伝的アルゴリズムを用いる。遺伝的アルゴリズムにおける適応度を、確信度の推定精度とする。

3.6 確信度の推定

選択した特徴量を用い、SVM によって解答者の設問ごとの確信度を推定する。確信度は表 2 に示す 2 クラスである。

4. 実験

本節では、提案手法を用いた実験について述べ、この実験によって推定された確信度の精度について検証を行う。

4.1 実験条件・実験手順

実験の様子を図 4 に示す。被験者に図 5 の上のような問題に解答してもらい、その際の視点情報を記録した。被験者は日

表 1 特徴量

No.	特徴量名	説明
1	fixation_choices	_count
2		_ratio
3	fixation_sentence	_count
4		_ration
5	fixation_duration_choices	_total
6		_average
7		_maximum
8		_minimum
9	fixation_duration_sentence	_total
10		_average
11		_maximum
12		_minimum
13	fixation_variance	_x
14		_y
15	saccade_length	_total
16		_average
17	saccade_count	_total
18		_choices
19		_sentence
20		_c-s
21	saccade_duration	_total
22		_average
23		_maximum
24		_minimum
25	saccade_velocity	_total
26		_average
27		_maximum
28		_minimum
29	readtime[s]	解答時間

表 2 クラスの分類

クラス	確信度
0	解答に確信をもてた
1	解答に確信をもてなかった

本人大学生および大学院生 11 名（男性：9 名、女性：2 名）である。被験者の TOEIC スコアは、最高値が 940 点、最低値が 450 点、平均は 709 点であった。本実験ではアイトラッカとして Tobii EyeX を用いた。使用した問題、およびアンケートのフォーマットを図 5 に示す。問題には、TOEIC の Part5 を想定した四択穴埋め問題 80 問を使用した。アンケートの内容は、「その問題に自信をもって解答できたかどうか」であり、このアンケートの結果をもとに、確信度を表 2 に示す 2 つのクラスに分類した。

実験手順は、まず被験者に対してキャリブレーションを行った。キャリブレーションとは、アイトラッカが取得する視点情報を被験者ごとに補正することである。その後、被験者は問題に解答し、1 問解くごとにその問題に関するアンケートに回答する。この手順を 80 問解答し終えるまで繰り返した。

次に、得られた特徴量から、確信度の推定に有効なものを選択した。選択には遺伝的アルゴリズムを用いた。詳細には、始

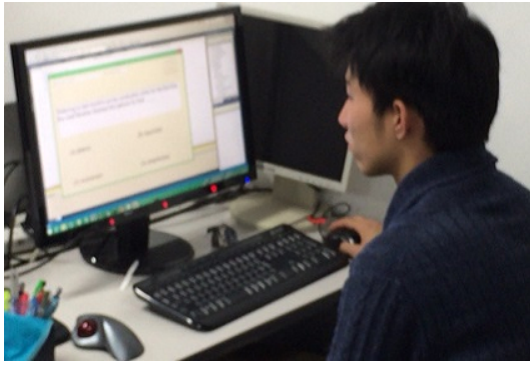


図 4 実験の様子

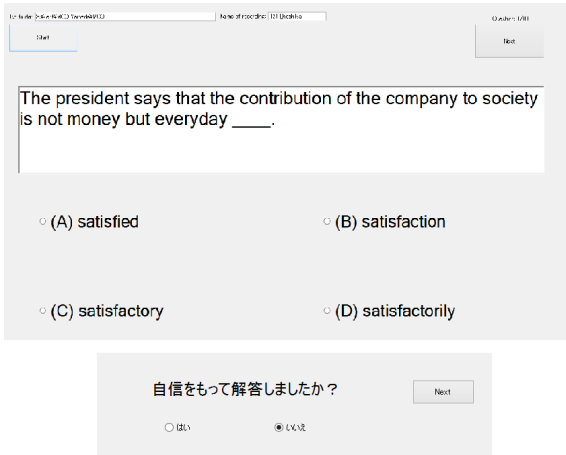


図 5 問題とアンケートのフォーマット

めに 200 個の親世代をランダムに生成した。各個体は、29 次元のバイナリデータで表現され、符号の 1 つ 1 つが各特徴量を表現している。これらを交叉させることによって 200 個の次世代個体を生成し、各個体の適応度を求めた。これを、300 世代繰り返した。今回は、交叉させる個体の選択方法としてルーレット方式、交叉方法として一様交叉を採用した。また、変異確率を 3% とし、適応度が高い個体を次世代に残すエリート保存戦略を採用した。

その後、SVM を用いて確信度を推定し、その精度を求めた。SVM を使用する際、被験者 1 名が解答した問題の内、1 問分のデータをテストデータ、残りの 79 問分のデータを学習データとすることで被験者依存の学習を行い、Leave-One-Document-Out Cross-Validation を用いて推定した。これを被験者 11 名に対して行った。

4.2 実験結果・考察

各被験者が解答した問題 80 問に対して選択された特徴量を用いて確信度を推定した。また、比較実験として、特徴量に解答時間 (表 1 の No.29) のみを用いて確信度を推定した。推定により得られた結果を図 6 に示す。棒グラフ上の横線は被験者ごとの事前確率、すなわち被験者が確信を持つと回答した割合を表している。選択された特徴量によって被験者 11 名の確信度を推定した精度は、最高値は 92.5%、最低値は 85.0%、平均は 89.8% であった。一方、特徴量に解答時間のみを使用して

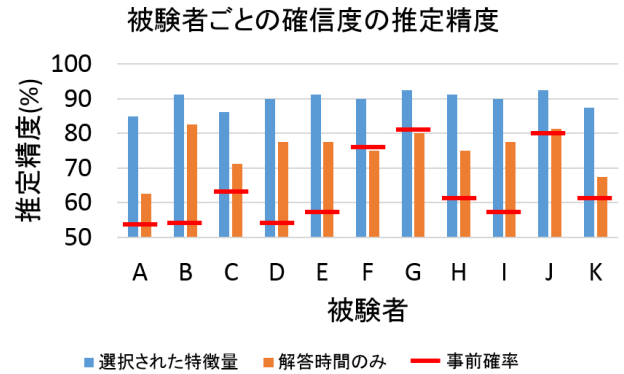


図 6 確信度の推定精度

表 3 符号検定の結果。**は有意水準 1%、*は 5% で帰無仮説 (提案手法と比較手法に差がないとする仮説) が棄却されることを示す。

比較手法	全体	A	B	C	D	E	F	G	H	I	J	K
事前確率	**	*	**		**	**	**			**	**	
解答時間のみ	**											

確信度を推定した場合、推定精度の最高値は 81.3%、最低値は 62.5%、平均は 75.2% であった。提案手法と比較手法で符号検定 [10] を行った結果を表 3 に示す。これより、被験者ごとの検証では、全てのケースについて帰無仮説が棄却できたわけではないが、全体を通してみると、比較手法の両者に対して、提案手法の優位性が示された。

推定精度が高かった被験者は G と J で、推定精度が低かった被験者は A であった。図 7 と図 8 は、それぞれ被験者 A が解答に確信をもった際の視点情報と、解答に確信をもてなかった際の線点情報を表しており、円が fixation、円の大きさが fixation の持続時間、円の中の線が saccade を表している。被験者が確信をもって解答した際の特徴として、図 7 のように各選択肢に対する fixation の回数が少なく、短い fixation 持続時間で選択肢を絞り解答している。一方で、被験者が解答に確信をもてなかった際の特徴として、図 8 のように fixation の回数が多く、それぞれの fixation 持続時間が長くなる傾向が見られた。fixation 数の増加に伴って、画面全体を通して saccade の回数も増加していることが分かる。なお、特徴量として fixation 持続時間のみを使用して確信度を推定したとき、全ての被験者において、視点情報を用いた際の推定精度を下回ったため、確信度の推定に視点情報が有効であることが確かめられた。

次に、被験者 A の確信度の推定精度が他の被験者のものと比べて低かった理由について考える。図 9 と図 10 は、それぞれ被験者 A の確信度の推定に失敗した問題における視点情報を表している。図 7 と図 9 はどちらも被験者 A が確信をもって解答した際の視点情報だが、図 7 の問題では推定に成功し、図 9 の問題では推定に失敗している。一方、図 8 と図 10 はどちらも被験者 A が確信をもてず解答した際の視点情報だが、図 8 の問題では推定に成功し、図 10 の問題では推定に失敗している。図 7 と図 10 から、被験者 A に関しては、fixation、saccade の数が少なく、fixation 持続時間が全体的に短い場合、確信をもつ

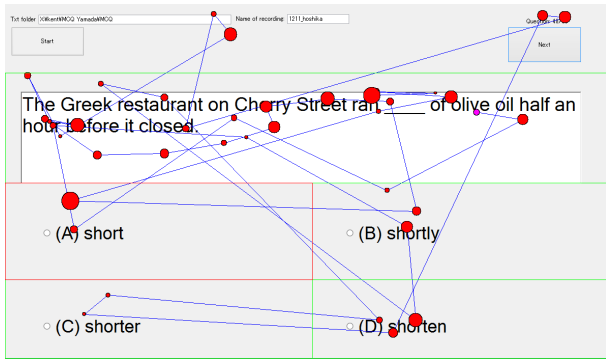


図 7 被験者 A が確信をもった際の視点情報

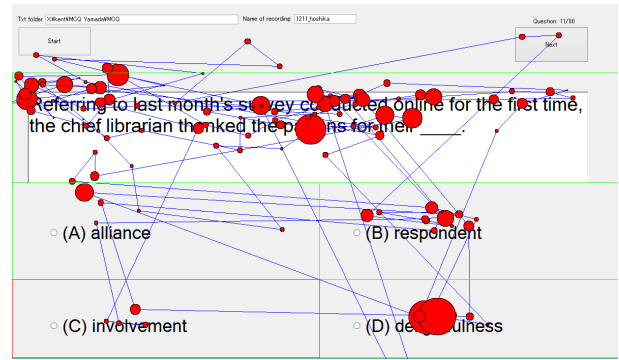


図 9 確信をもって解答して推定に失敗した際の視点情報

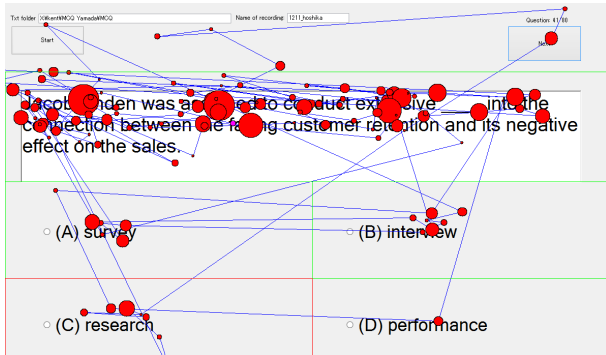


図 8 被験者 A が確信をもてなかった際の視点情報

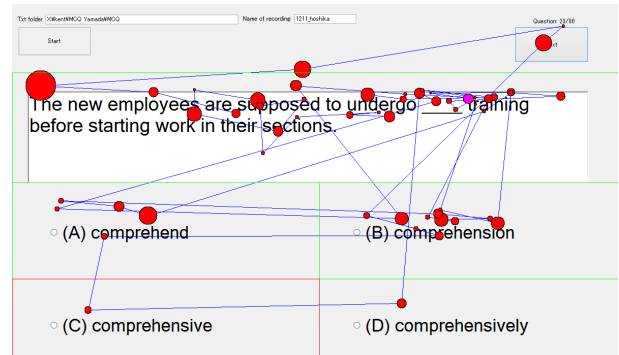


図 10 確信をもてず解答して推定に失敗した際の視点情報

て解答された問題であると SVM に推定されたと考えられる。図 9 において、被験者 A は問題文や選択肢を注意深く眺め熟考し、選択肢を 1 つに絞っている可能性がある。このとき、他の選択肢を注意深く精査した上で解答しており、多くの fixation や saccade が残るため、確信なしと推定される。しかし、全ての選択肢を時間をかけて見た上で解答したという点を考えれば、復習を行う余地がある箇所であるといえるため、復習を補助するシステムとしては容認できる範囲である。また、図 10 のように被験者が確信をもてず解答した場合については、問題や選択肢を流し見し、自分の知識では解けないと判断した際の少ない fixation や saccade によって確信ありと誤って推定されている。被験者の問題解答時の多様な特徴が確信ありと確信なしのどちらにも見られるため、正しく推定されていないと考えられる。これに対して、被験者が解答を諦めて適当に解答した場合と通常通り解答した場合を比較し、変化がないか調べることを今後の課題とする。その他、視点情報以外を取得するため、新たなデバイスを追加することが考えられる。

推定精度が最も高かった被験者 G, J について考える。被験者 G, J は TOEIC スコアがそれぞれ 930 点, 940 点と英語の習熟度が非常に高い。そのため、全体の 8 割前後の問題に確信をもって解答している。また、この 2 名が問題を解く際の視点情報は、図 11 や図 12 のように fixation の個数や持続時間、saccade の回数に明確な違いが見られたため、高い精度で推定できたと考えられる。また、被験者 J には図 13 のような視点が見られた。この問題においては、選択肢 B 以外の選択肢を注視していないにもかかわらず、正解である選択肢 D を解答して

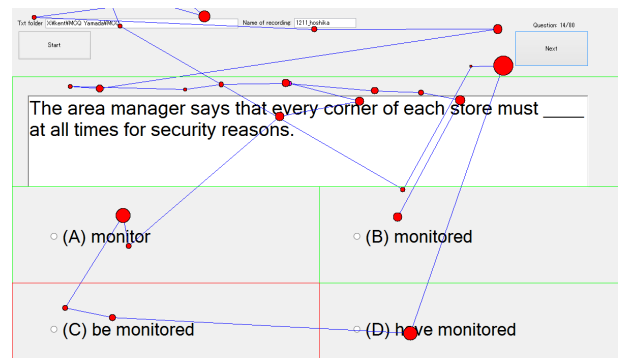


図 11 被験者 J が確信をもった際の視点情報

いる。このとき、問題文に目を通した段階で解答が予想できている。このことから、人は必ずしも問題中における目当ての情報を注視したり読んだりしているとは限らないといえる。一方で、藤好らの研究 [4] から、視点情報と英語習熟度に相関があることが分かっているため、TOEIC スコアなどの英語習熟度の指標となるもので被験者を分類し、それぞれの習熟度に応じて使用する特徴量に重み付けすることが考えられる。

29 個の特徴量から特徴選択を行った結果、全ての被験者に採用された特徴量や、採用されなかった特徴量は存在しなかった。最も採用回数が多かった特徴量を表 4 に示す。採用回数が多かった特徴量として、saccade 距離の合計、問題文内の saccade の回数、選択肢-問題間の saccade の回数、解答時間の 4 つで、11 名中 6 名に採用された。これら 4 つの内、saccade 距離の合

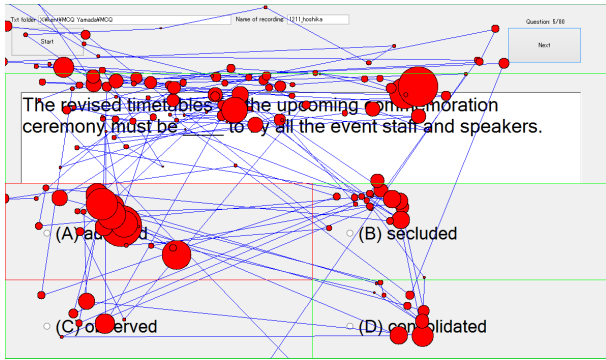


図 12 被験者 J が確信をもてなかった際の視点情報

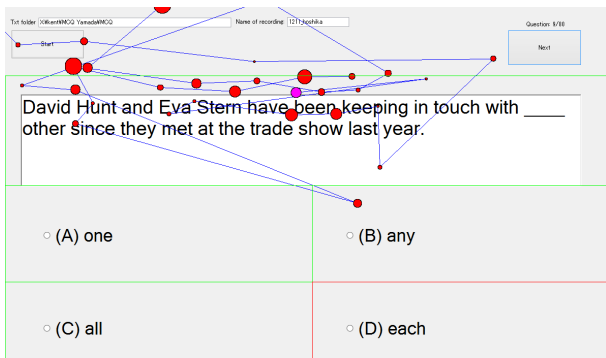


図 13 視点をほとんど残さなかった例

表 4 特徴量の採用回数

採用回数	特徴量の番号
6 回	16, 23, 24, 29
5 回	3, 6, 7, 8, 12, 15, 21
4 回	4, 9, 13, 14, 17, 18, 20, 25, 27, 28
3 回	1, 5, 19
2 回	2, 10, 11
1 回	22, 26

計以外の 3 つが頻繁に選択されたことで、関連研究 [2] [7] [8] で得られた知見を検証する結果となった。しかし、多くとも被験者全体の半分でしか採用されなかったことから、解答時の確信度を表現する特徴量には個人差があるといえる。最も採用回数が少なかった特徴量は、選択肢間の saccade の回数と saccade 時の視点の速度の平均の 2 つで、どちらも 11 名中 1 名のみに採用された。選択肢間の saccade の回数については、選択肢を絞り切れないために確信をもてなかった場合に有効であると想定していたが、ほとんど採用されることはなかった。以上のことから、被験者の性格や問題を解く際の戦略、癖など、被験者それぞれに依存するところが大きいため、有効な特徴量は被験者によって異なると思われる。

5. まとめ

本稿では、優先度に応じた復習の重要性に着目し、優先度を考慮に入れた復習を補助するシステムの実現のため、学習者の問題解答時の確信度を判定する手法を提案した。本手法では、問題解答時の視点情報を取得した。視点情報には個人差がある

ため、被験者ごとに遺伝的アルゴリズムを用いて特徴選択を行った。確信度の推定には SVM を用いた。被験者 11 名の平均推定精度は 89.8% であった。実験によって有効な特徴量として、saccade 距離の合計、問題文内の saccade の回数、選択肢-問題間の saccade の回数、解答時間の 4 つが得られたが、これらの採用率は約 5 割に留まっており、問題解答時の振る舞いは、被験者の性格や問題を解く際の戦略などから影響を受けるため、有効な特徴量は被験者によって異なることが分かった。また、英語習熟度が高いほど、確信度に応じた視点情報に違いが出やすい傾向があることが明らかとなった。

今後の課題として、解答に対する確信度が異なるにも関わらず、視点情報に違いが見られない場合について、より詳細に調査することや、新たに視点情報以外を取得するデバイスを選定することが挙げられる。

謝辞 本研究の一部は、JST CREST, 科研費基盤研究 (A) (25240028), ならびに科研費挑戦的萌芽研究 (15K12172) の補助による。

文 献

- [1] K. Rayner, "Eye movements in reading and information processing: 20 years of research.," *Psychological bulletin*, vol.124, no.3, p.372, 1998.
- [2] 小島一晃, 村松慶一, 松居辰則, "視線を用いた多肢選択問題の回答プロセスと確信度の分析手法の実験的考案," *人工知能学会全国大会論文集*, vol.26, pp.1-4, 2012.
- [3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol.12, no.Oct, pp.2825-2830, 2011.
- [4] 藤好宏樹, 吉村和代, K. Kunze, 黄瀬浩一, "英文問題解答時の視点情報をを用いた英語能力推定法," *電子情報通信学会技術研究報告 IEICE technical report: 信学技報*, vol.115, no.24, pp.49-54, PRMU2015-10, 2015.
- [5] 吉村和代, 川市仁史, 黄瀬浩一, "アイトラックで取得した視点情報と文書理解度の関係," *電子情報通信学会技術研究報告 IEICE technical report: 信学技報*, vol.112, no.495, pp.261-266, PRMU2012-224, 2013.
- [6] 大社綾乃, K. Kunze, O. Augereau, "学習補助のための視点情報に基づく文書アノテーション," *電子情報通信学会技術研究報告 IEICE technical report: 信学技報*, vol.115, no.23, pp.161-166, PRMU2015-30, 2015.
- [7] M.-J. Tsai, H.-T. Hou, M.-L. Lai, W.-Y. Liu, and F.-Y. Yang, "Visual attention for solving multiple-choice science problem: An eye-tracking analysis," *Computers & Education*, vol.58, no.1, pp.375-385, 2012.
- [8] S. Ishimaru, S.S. Bukhari, C. Heisel, J. Kuhn, and A. Dengel, "Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics," *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp.1041-1045 2016.
- [9] G. Buscher, A. Dengel, and L. van Elst, "Eye movements as implicit relevance feedback," *CHI'08 extended abstracts on Human factors in computing systems*, pp.2991-2996 2008.
- [10] 高木英行, "使える! 統計検定・機械学習: I: 2 群間の有意差検定," *システム/制御/情報: システム制御情報学会誌*, vol.58, no.8, pp.345-351, 2014.