

英文問題解答時の視点情報を用いた英語能力推定法

藤好 宏樹[†] 吉村 和代[†] Kai Kunze^{††} 黄瀬 浩一[†]

[†] 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

^{††} 慶應義塾大学大学院メディアデザイン研究科

〒 223-8526 横浜市港北区日吉 4-1-1

E-mail: [†]{fujiyoshi,yoshimura}@m.cs.osakafu-u.ac.jp, ^{††}kai@kmd.keio.ac.jp, ^{†††}kise@cs.osakafu-u.ac.jp

あらまし 本稿では、アイトラッカで取得した英文問題解答時の視点情報を用いて、英語習熟度を推定する手法を提案する。一般に、英語習熟度を測るテストとして TOEIC が挙げられる。しかし、TOEIC には結果が出るまでに時間を要する、問題数が多く受験者に負担が大きいという課題がある。そこで、TOEIC よりも少ない問題数で手軽に英語習熟度を測る事ができれば有益であると考えられる。このような推定を可能にするための方法として、テストの問題正答数の他に、問題解答時の視点情報を利用することが考えられる。本手法では眼球運動の一種である fixation と saccade に関する特徴を利用する。英文問題 10 題を用いて被験者 9 人の英語習熟度を推定したところ、TOEIC スコアを平均絶対誤差 36.3 点で推定することができた。

キーワード アイトラッカ, 視点情報, 英語習熟度, TOEIC

1. はじめに

近年、グローバル化によって英語で会話する、英文書を読む、書くなど英語に触れる機会が増えている。そのため、英語の重要性が増し、英語教育の充実が必要となっている。英語教育を充実させるためには優秀な教員や良い教材を揃えなければならない。加えて、生徒の知識や技術を確認して、苦手な分野に応じて適切に指導することが必要である。そこで、教員は生徒にテストを受けさせることで英語習熟度を測る。英語習熟度を測るテストとして最も認知され、多くの人が受験しているものとして TOEIC が挙げられる。[1]。TOEIC は英語のリスニングとリーディングのスキルを測る世界共通のテストである。社会において TOEIC のスコアは重要視され、就職活動や進級、昇進などにも必要で、多くの人が受験する。

しかし、TOEIC には、主に二つの課題がある。第一の問題点として、受験者全員の得点を集計するまで結果が出ないため、受験から結果が出るまでに一ヶ月近くを要することが挙げられる。第二の問題点として、問題数が多いため、試験時間が 2 時間と長く、受験者に負担が大きいということが挙げられる。したがって、TOEIC の二つの課題を改善した方法で、英語習熟度を測ることが可能なら、非常に便利である。

TOEIC よりも少ない問題数で、手軽に英語習熟度を測る方法を実現するためには、問題の正解、不正解以外の情報を得る必要がある。本研究では、問題解答時の視点情報を用いて英語習熟度を測る。「目は口ほどにものを言う」という言葉があるように、人の視線にはその人の思考が反映されている。例えば、問題解答時の視点情報を解析することで、文章の読む速さ等が分かる。英語習熟度の高い者は単語や文法などの知識が豊富で

運用能力も高いため、読む速さも自然と速くなる。反対に、英語習熟度の低いものは、知識量が少ないために、文章中に知らない単語がより多くあり自然と読む早さが遅くなると考えられる。したがって、視点情報を用いることによって英語習熟度の高い者と低い者の間を分けることができると考える。

本稿では、TOEIC よりも問題数が少なく、手軽に英語習熟度を推定する手法を提案する。提案手法の特徴は、英文問題解答時の視点情報を取得し、視点情報から得られる特徴量と、問題の正答数を用いて英語習熟度を推定することである。問題の正答数は、英語習熟度を反映しているが、問題数が少ない場合には、英語習熟度を推定するのに十分な情報とはいえない。したがって、問題解答時の視点情報に関する特徴量とあわせて用いることによって、英語習熟度を推定するのに十分な情報を得ることを試みる。また、本手法では、英語習熟度を表すものとして TOEIC スコアを用いる。9 人の被験者が 10 題の英文問題を解答する際の視点情報と問題の正答数を用いて、TOEIC スコアを推定した結果、平均絶対誤差は 36.3 点が得られた。なお、本研究は大阪府立大学工学研究科の倫理委員会で承認済みであることを付記しておく。

2. 関連研究

本研究では、ユーザが英文問題を解答する際の視点情報を用いてユーザの英語習熟度を推定することを目標とする。関連研究として視点情報を用いて英語能力を識別、推定する研究について述べる。読書時の視点情報の特徴と文書特徴から理解度と言語習熟度を推定する手法を Gomez らが提案している [2]。Gomez 手法では、ユーザが英文書を読む際の視点情報を取得し、視点に関する特徴を抽出する。加えて、ユーザが読んだ文

書中にどのような単語を含むかという文書特徴を計算する。さらに、それぞれの文書特徴に対して注視時間を計算し、これを特徴量として抽出する。視点に関する特徴量と文書特徴と注視時間を組み合わせた特徴量を用いてユーザの文書に対する理解度、習熟度を推定する。結果として理解度の低いユーザと高いユーザを大まかに分けることができる。本研究では、理解度推定で有効である視点情報から得られる特徴量を用いて英語習熟度を推定する。

Gomez らと同様に、読書時の視点情報を用いてユーザの文書に対する理解度を推定する手法を吉村らが提案している [3]。吉村らの手法では、理解度は、ユーザが読んだ文書の内容をどの程度理解しているかという指標であり、文書に関する問題の正答数としている。ユーザが英文書を読む際の視点情報を取得して、得られた視点情報を理解度ごとに異なるパターンとして学習して理解度を推定する。しかし、この手法では十分な精度が得られないことが分かっている。

また、吉村らは英語能力として、理解度に加えて、習熟度を挙げている。文書に関する問題を解答させることによって正答数を理解度としている。習熟度とは英語全般に関する運用能力としており、習熟度を表すものとして、TOEIC スコアを用いている。吉村らは理解度推定とは異なる特徴量を用いて英語習熟度を推定する手法を提案している [4]。この手法では、吉村らは、文書を読む際の視点情報が、文書の難易度やユーザの理解度によって受ける影響を小さくすることを考える。そこで、吉村らは理解度推定とは異なり、複数文書分の視点情報から得られた特徴量を統合して英語習熟度を推定する。結果として、複数文書分の視点情報を用いることが英語習熟度推定に有効である。

本稿で提案する手法は吉村らの手法と異なり、文書を読んでいる時の視点情報に加えて英文問題を解答している時の視点情報を用いて英語習熟度を推定する。問題解答時の視点情報を含むことで、文書を読む時の特徴に加えて、問題の解き方の特徴が得られるので、より正確に英語習熟度を推定できると考える。また、吉村らの手法と同様に、複数問題分の視点情報から得られる特徴量を用いることで、視点情報が、ユーザと問題との相性から受ける影響を小さくする。

3. 提案手法

3.1 提案手法の流れ

英文問題解答時の視点情報を用いて、英語習熟度を推定する手法を提案する。提案手法の流れを以下の図 1 に示す。アイトラッカを用いて英文問題解答時の視点情報を取得する。次に得られた視点情報を文書画像検索を用いて文書画像上に変換し、特徴量を抽出する。最後に、線形回帰を用いて被験者の英語習熟度を推定する。また、本研究では英語習熟度を TOEIC スコアとして定義する。

3.2 視点情報の取得

まず、アイトラッカを使用して視点情報を取得する。アイトラッカは眼球運動を測定する装置のことで、装着型と非装着型がある。本節では、本手法で用いる装着型アイトラッカにつ

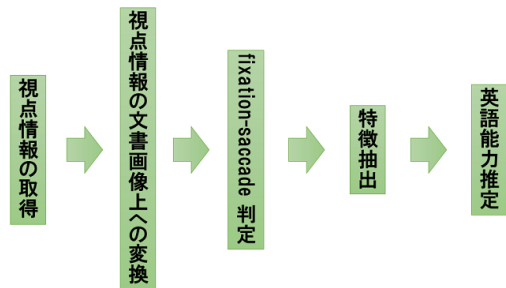


図 1 提案手法の流れ



図 2 装着型アイトラッカ

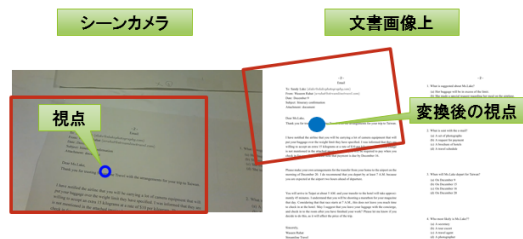


図 3 文書画像上に変換された視点

いて説明する。装着型アイトラッカはメガネ型のものが多く、ユーザのしている風景を記録するシーンカメラと眼球を測定するアイカメラによって構成されている。アイカメラによって眼球運動を測定して、シーンカメラ上のどこを見ているかを特定する。頭部に装着して使用するので、被験者の動きを制限しないという特徴がある。本手法では英文問題解答時の自然な振る舞いを記録するため、図 2 に示すようなウェアラブルな装着型アイトラッカを使用して被験者の視点情報を取得する。

3.3 視点情報の文書画像上への変換

次に、アイトラッカで取得する視点情報を文書画像上に変換する。問題解答時の被験者の振る舞いを分析するためには、被験者が文書や問題文をどのように見たかという情報が必要である。アイトラッカによって取得する視点情報はシーンカメラ上の視点情報である。そこで、アイトラッカによって取得した視点情報を文書画像検索手法の一つである、LLAH (Locally Likely Arrangement Hashing) [6] を用いて文書画像上に変換

する。

図3に文書画像上に変換後の視点情報を示す。まず、アイトラッカのシーンカメラ画像上の視点を中心に画像を切り出す。切り出した画像を検索質問画像として、LLAHによる文書画像検索を行う。LLAHは、検索質問画像の特徴点とそれに対応する登録文書画像の特徴点の関係から射影変換パラメータを推定できる。得られた射影変換パラメータを用いて、視点情報をシーンカメラ画像上から文書画像上に変換する。これによって、視点情報は、書画像上の座標に変換され、被験者が問題解答時にどこを見ていたかが分かる。

3.4 fixation-saccade 判定

眼球運動は、視点がある範囲内で一定時間留まる fixation、fixation 間の素早い目の動きである saccade に分類される。人は、fixation と saccade を繰り返すことによって文書を読んでいる。

まず、文書画像上に変換した視点情報に対し、fixation-saccade 判定を行う。判定方法には Buscher らの手法を用いる [5]。図4に Buscher らの手法の手順を示す。文書画像上に変換された視点情報を時間順に見ていく。まず、連続する4つの視点互いの視点間に関して距離 T_1 以内に存在するとき、これら4点を minimum fixation とする。そして、以後の視点が minimum fixation の4点との距離が全て距離 T_2 以内であるとき、large fixation とする。 T_1 と T_2 は $T_1 < T_2$ の関係である。連続して3点以上の視点 large fixation に含まれない場合、fixation 判定を終了する。large fixation と minimum fixation に含まれる全ての視点を1つの fixation とする。また、large fixation に含まれない視点が3点連続しない場合はノイズとして除去する。fixation の中心座標は fixation に含まれる視点の重心で、円の大きさは fixation に含まれる視点の数に依存する。以後、large fixation に含まれなかった3点から新たに fixation 判定を始める。fixation の中心間を saccade とする。この様にして fixation-saccade 判定を行う。

3.5 特徴抽出

本手法で用いる特徴量について説明する。

fixation-saccade 判定を行った視点情報から眼球運動の特性を利用して特徴量を抽出する。図5は文書上に表現した fixation と saccade の例である。円が fixation で、円をつなぐ線が saccade である。本手法で用いる英文問題は左側に文書、右側に問題文と分かれている。人は英文問題を解く際に、文書部分を読む、問題文部分を読む、問題の答えが含まれる文章を探す、等の行動をとる。英語習熟度によってこれらの行動に差が出ると思われる。

そこで、図5の様に文書部分、問題文部分、全体の3つの領域内に分ける、3つの領域それぞれにおいて表1の fixation、saccade に関する特徴量を抽出する。

加えて、視点の移動に関する特徴量を用いる。英文問題を解答する際の振る舞いとして図6において線で示される様な文書部分と問題文部分の間を往來する視点移動がある。この視点移

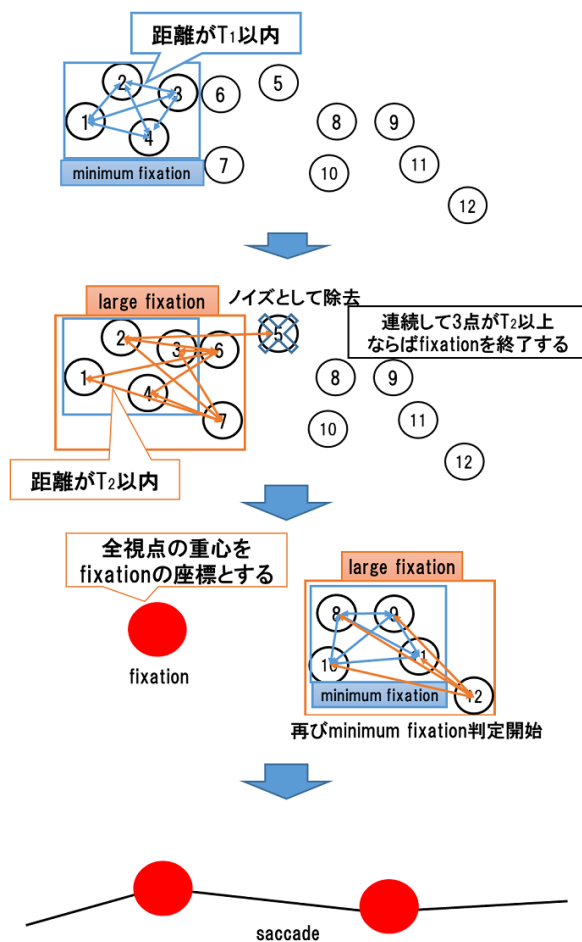


図4 fixation-saccade 判定

動は、問題文を読み答えが分からない場合、文書部分から答えを探す時に発生する。英語習熟度の高い人は、読む回数が少なくても文書の内容を理解できる。しかし、英語習熟度の低い人は、何度も読まなければ難しい文書は理解できない。したがって、文書部分と問題文部分の間を視点が行き来する回数が多いほど、英語習熟度が低いと考え、この視点移動の回数を特徴量とする。

以上の視点情報に関する特徴量とユーザの解答する問題の正答数の特徴量として用いる。問題の正答数は、英語習熟度を反映している。ただし、問題数が少ない場合には、習熟度を推定するのに十分な情報とはいえない。したがって、他の特徴量とあわせて用いる。

次に、特徴抽出について説明する。得られる特徴量が、問題の難易度やユーザの理解度等の、被験者と問題の相性に影響を受けると考える。そこで、複数題分の視点情報から得た特徴量を合計することで、問題との相性による影響を小さくする。本手法では、問題1題ごとに特徴量を抽出し、特徴量ごとに複数の問題から得られる特徴量を合計する。

表 1 提案手法で用いる特徴量

特徴量	詳細
fixation に関する特徴量	fixation 持続時間の平均
	fixation の総数
	fixation 持続時間の和
	fixation 持続時間の割合
	saccade の平均の長さ
saccade に関する特徴量	saccade の平均の長さ
	saccade 時間の平均
	saccade 時間の和
	saccade の速さの和
	saccade の距離の総和
その他の特徴量	文書、問題間の視点移動の回数
	問題の正答数

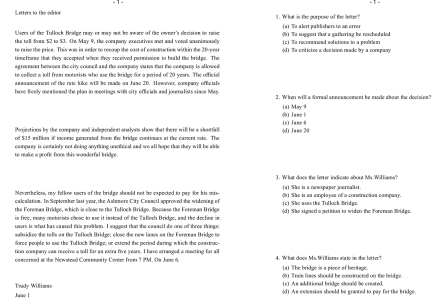


図 7 実験に使用した文書

データの誤差を算出する。その後、全ての特徴を含むモデルから特徴を1つずつ取り除いた35通りのモデル作成する。35通りのモデルで線形回帰を用いて英語習熟度を推定し、推定結果と正解データとの誤差を算出する。この結果の中から、最も誤差が小さくなり、かつ全ての特徴を含むモデルよりも誤差が小さい結果が得られたモデルを選択する。選択するモデルに対しても同様の処理を行い、1つずつ特徴を取り除く。誤差が最も小さくなるまで繰り返すことによって、有効な特徴を選択する。

4. 実験

今回の実験では、提案手法による英語習熟度推定の精度を検証する。被験者は9名で被験者ごとにデータセットを作成し、Leave one subject out によって評価した。

4.1 実験条件

被験者に TOEIC の英文問題を解答させ、その際の視点情報を記録した。被験者は9人で TOEIC スコアの最高値が930点、最低値が435である。使用した問題は TOEIC のテスト対策の問題集から選んだ英文問題10題である。英文問題は図7の様に左側に文書、右側に小問4つという構成でフォーマットは統一した。また、アイトラッカには SMI 社の Eye Tracking Glasses 2.0 (ETG2.0) を使用した。

実験の流れを示す。アイトラッカを図8の様に装着した後、キャリブレーションを行った。その後、被験者に英文問題を1題解答させ、視点情報を記録した。この手順を10題分繰り返した。被験者の疲労を考慮して5題解答したところでアイトラッカを外し休憩を挟んだ。また、英文問題解答時の自然な動作や視点情報を記録するため、解答中は被験者の動作に制限を設けなかった。

4.2 実験結果

まず、特徴量を選定した結果を表2に示す。これは、10題分の視点情報と問題の正答数を用いた英語習熟度推定の結果が最も良くなる様に選定した結果である。相関係数は被験者の TOEIC スコアと特徴量との相関係数である。相関係数の絶対値が大きいほど有効な特徴量で、絶対値が低いほど TOEIC スコアの推定に適さない特徴量と考える。

表2にから、相関係数の絶対値が大きい順に「問題の正答数」、「全体の平均 fixation 持続時間」、「全体の saccade の距離

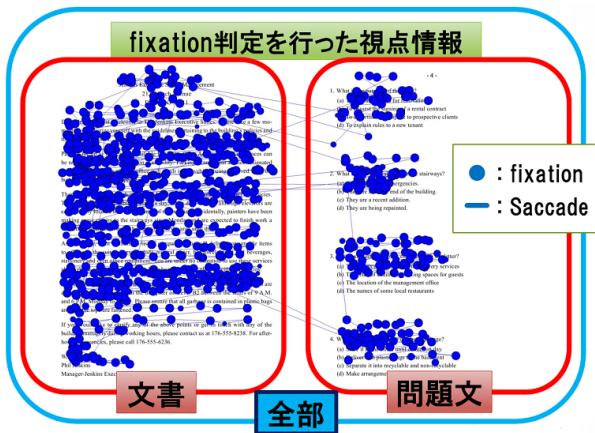


図 5 文書画像上に変換された視点情報

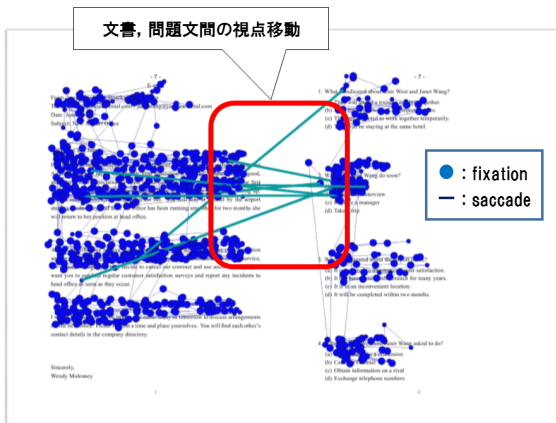


図 6 文書、問題文間の視点移動

3.6 英語習熟度推定

得られた特徴に対し、線形回帰を用いて学習、英語習熟度を推定する。その後、全ての特徴から backward stepwise selection を用いて英語習熟度の推定に有効な特徴を選定する。backward stepwise selection とは、全ての特徴を含むモデルから特徴を1つずつ取り除いて、有効な特徴を選択する手法である。以下に、具体的な方法を示す。まず、35個全ての特徴を含むモデルで、線形回帰を用いて英語習熟度を推定し、推定結果と正解

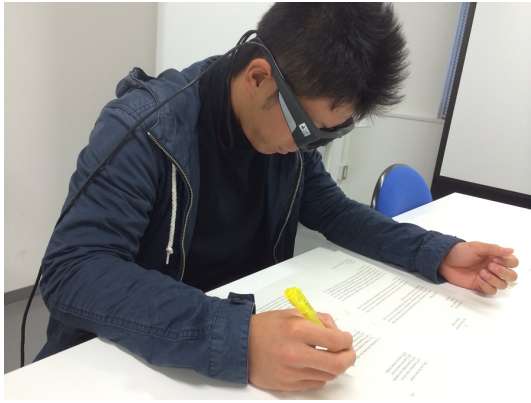


図 8 実験の様子

表 2 特徴量と TOEIC スコアの相関係数

特徴量	相関係数
問題の正答数	0.74
文書部分の fixation の総数	0.45
文書部分の平均 saccade 長	-0.38
文書部分の saccade の速さの平均	0.11
文書部分の fixation の割合	0.55
文書部分の saccade 時間の和	0.59
文書部分の saccade の速さの和	-0.54
問題文部分の平均 saccade 長	-0.22
問題文部分の文書部分の fixation の割合	0.58
問題文部分の fixation 時間の和	0.50
問題文部分の saccade 時間の和	0.70
問題文部分の saccade の速さの和	0.45
全体の平均 fixation 持続時間	0.66
全体の fixation の割合	0.54
全体の平均 saccade 時間	0.034
全体の saccade 時間の和	0.56
全体の saccade の速さの和	0.59
全体の saccade の距離の総和	0.66
全体の平均 saccade 時間の逆数	0.59
段落移動数	0.54

の総和」であることが分かる。中でも問題の正答数が最も相関係数の絶対値が大きい 0.74 である。これは、実験に使用した英文問題 10 題分の正答数という少ない数ではあるが、問題の正答数は英語習熟度に大きく関連することが分かる。次に、「全体の平均 fixation 持続時間」は解答中の全ての fixation の平均持続時間であり、相関係数 -0.66 となっている。これは相関係数が負となっているため、英語習熟度が低い被験者ほど、解答中に文書や問題をじっくり読んでいたということである。「全体の saccade の距離の総和」は解答中の視点の総移動距離で、相関係数が -0.66 となっている。これも相関係数が負であるため、英語習熟度が低い被験者ほど解答中に答えを探索するため、多くの文書を読んでいたということである。これらの特徴量は、英語習熟度による、被験者の解答中の振る舞いの差を良く表していると考えられる。

反対に、相関係数の絶対値が低い特徴量は、低い順に「全体の平均 saccade 時間」、「文書部分の saccade の速さの平均」、

表 3 TOEIC スコアの推定結果

被験者	A	B	C	D	E	F	G	H	I	平均
絶対誤差	87.0	86.9	1.52	28.1	42.9	24.1	9.38	24.5	21.9	36.3.

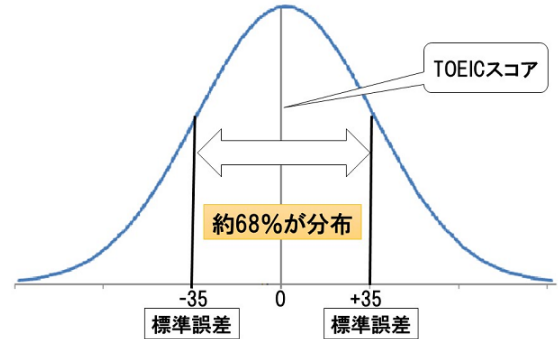


図 9 TOEIC スコアの分布

「問題文部分の平均 saccade 長」である。「全体の平均 saccade 時間」などは、相関係数が -0.03 であり、ほとんど英語習熟度とは関係ないことが分かる。この様に、10 文書を用いて英語習熟度を推定する際に用いた特徴量には、相関係数の低いが含まれていた。これは、今回の実験における推定結果に偶然、悪い影響を与えなかったためであり、問題や被験者などの実験条件をかえた場合は、推定結果に悪影響を及ぼすと考えられる。

次に、10 題分の視点情報から得た、表 2 の特徴量を用いて TOEIC スコアを推定した結果を表 3 に示す。これは、線形回帰によって推定されたスコアと被験者の持つ TOEIC スコアの絶対誤差を示す、推定値との平均絶対誤差が 36.3、標準偏差が 46.5 という結果が得られた。

一般的に、能力に変化がない限り、TOEIC は何度受験しても、スコアは一定に保たれるとされている。しかし、実際は TOEIC スコアは被験者と問題の相性などが原因で、本来の英語習熟度とは誤差がある。図 9 の様に、ユーザの実際の英語習熟度から ± 35 点の範囲内に TOEIC スコアが含まれる確率が 68% と言われる [7]。本手法は TOEIC と同程度の精度で英語習熟度を推定することを目標としている。そこで、推定したスコアと被験者本来 TOEIC スコアとの差が 35 点以内であれば、推定成功とする。これにより、結果として、9 人中 6 人が推定成功したと考える。

以上から、視点情報から英文問題解答時の振る舞いから英語習熟度の差を表すことができる程度できるが、特徴量の選択に工夫が必要であることがわかった。

また、この結果は 10 題分の視点情報を用いて推定した結果である。英文問題 10 題は TOEIC を解くよりも問題数が少ないが、量は少なくない。そこで、本実験において選定した特徴量が、より少ない問題数で英語習熟度を推定することが可能かを検証した。

10 題から問題数を減らしていく場合、10 題から 9 題の問題

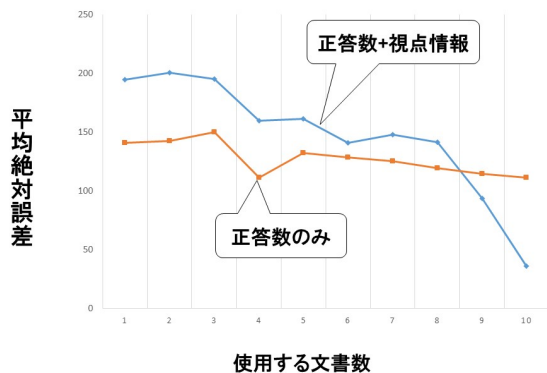


図 10 問題数ごとの推定精度

を選ぶ組み合わせは 10 通りで、全てのデータセットに対して英語習熟度を推定することが可能である。しかし、10 題から 2 題以上 8 題以下の問題を選ぶ組み合わせは非常に多くなる。そのため、全てのデータセットに対して英語習熟度を推定することは現実的ではない。そこで、2 題分以上 8 題分以下の特徴量を用いて英語習熟度を推定する場合、ランダムに 10 通りに組み合わせを選び、10 個のデータセットとする。なお、10 題から 9 題、または 1 題を選ぶ組み合わせについては 10 通りであるため、それらを 10 個のデータセットとする。10 個のデータセットに対して、10 題のときに選定した表 2 の特徴量を算出し、線形回帰を用いて英語習熟度を計算し、全ての平均絶対誤差を求めた。

求めた平均絶対誤差を図 10 に示す。これは、横軸が、推定に用いる英文問題の数である。縦軸は、線形回帰を用いて推定した結果と本来のスコアとの平均絶対誤差である。

図の赤いグラフが特徴量として、問題の正答数と視点情報を用いた場合の結果である。緑のグラフは正答数のみの特徴量として用いた場合の結果である。二つのグラフを比較すると、9 題のときまでは、視点情報を用いた結果の方が優れていることが分かる。しかし、8 題からは、問題の正答数に加えて視点情報を用いて推定した結果の方が悪化している。8 題以下の時は正答数のみを用いた結果と同様に徐々に推定結果が悪くなっている。

これは、問題数が少なくなることによって、得られる特徴量に、被験者と問題との相性や問題の構成等が大きく影響を与えているためと考えられる。更に、先述の通り、選定された特徴量には、TOEIC スコアとの相関係数が低いものも含まれているため、これらが悪影響を及ぼしたと考える。

以上から、提案法は推定に使用する問題数を減らした場合は、結果が顕著に悪くなった。しかし、英文問題 10 題ほどユーザに解いてもらえば英語習熟度を推定できると考える。

5. まとめ

本稿では、TOEIC よりも問題数が少なく、手軽に英語習熟度を推定する手法を提案した。提案手法の特徴は、英文問題解

答時の視点情報を取得し、視点情報から得られる特徴量と、問題の正答数を用いて英語習熟度を推定することである。問題の正答数は、英語習熟度を反映しているが、問題数が少ない場合には、英語習熟度を推定するのに十分な情報とはいえない。したがって、問題解答時の視点情報に関する特徴量と組み合わせることで、高い精度で英語習熟度を推定することを目指した。また、本手法では、英語習熟度を表すものとして TOEIC スコアを用いた。

提案手法では、まず、アイトラッカを用いて英文問題解答時の視点情報を取得した。次に得られた視点情報を文書画像上に変換し、眼球運動の特性を利用した特徴量を抽出した。視点情報に関する特徴量と、問題の正答数を特徴量として用いて線形回帰によって英語習熟度を推定した。その後、全ての特徴から backward stepwise selection を用いて英語習熟度の推定に有効な特徴を選定した。

特徴量を選定した結果、20 の特徴量が選択された。中でも「問題の正答数」、「問題文部分の平均 fixation 持続時間」、「問題文部分の saccade の距離の総和」が英語習熟度推定に有効であることがわかった。10 題の英文問題を解答する際の視点情報と問題の正答数を用いて、英語習熟度を推定した結果、平均絶対誤差は 36.3 が得られた。英文問題 10 題ほどの視点情報と問題の正答数を用いることで英語習熟度を推定できることが分かった。

今後の課題としては、精度向上のために、使用する問題数が少ない場合でも推定が可能な特徴を調査することが挙げられる。また、被験者により個人差があることから、多くの被験者のデータを収集し、分析することも課題の 1 つである。

謝 辞

謝辞 謝辞 本研究の一部は、JST CREST および日本学術振興会 科学研究費補助金 基盤研究 (A)(25240028) ならびに若手研究 (B)(26730095) の補助による。

文 献

- [1] <http://www.toeic.or.jp/>
- [2] Martinez-Gomez, Pascual, Akiko Aizawa, "Recognition of understanding level and language skill using measurements of reading behavior." Proceedings of the 19th international conference on Intelligent User Interfaces, ACM, 2014, pp. 95-104.
- [3] 吉村 和代, 川市 仁史, Kai Kunze, 黄瀬 浩一, "アイトラッカ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, vol. 112, no. 495, PRMU2012-224, pp. 261-266, Mar. 2013
- [4] 吉村 和代, Kai Kunze, 黄瀬 浩一, "読書時の眼球運動を利用した英語習熟度推定法," 電子情報通信学会技術研究報告, vol 114, no. 454, PRMU2014-123, pp. 63-68, Feb. 2015
- [5] G. Buscher, A. Dengel, "Gaze-based filtering of relevant document segments," In Workshop on Web Search Result Summarization and Presentation. Workshop on Web Search Result Summarization and Presentation (WSSP-2009), located at in conjunction with WWW, Vol. 9, pp. 20-24, 2009
- [6] 竹田 一貴, 黄瀬 浩一, 岩村 雅一, "1 億ページのデータベースを対象とした大規模文書画像," 電子情報通信学会議出研究報告, vol. 112, no. 441, PRMU2012-161, pp. 131-136, Feb. 2013
- [7] <http://eng.alc.co.jp/newsbiz/hinata/2008/08/toeic70.html>