

# 問題解答時の視点情報を用いた文書非依存な英語能力推定法

藤好 宏樹<sup>†</sup> OlivierAugereau<sup>†</sup> 黄瀬 浩一<sup>†</sup>

<sup>†</sup> 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: †fujiyoshi@m.cs.osakafu-u.ac.jp, ††augereau.o@gmail.com, †††kise@cs.osakafu-u.ac.jp

あらまし 英語能力を測るテストとして TOEIC がある。TOEIC は、問題数が多く受験者に負担が大きいという課題があるため、TOEIC よりも少ない問題数で英語能力を測る事を目標に、我々は、問題解答時の視点情報を用いた英語能力推定法を提案した。しかし、この手法は他のユーザによって解答時の視点情報が記録された問題でしか英語能力を推定出来ない。そこで、本稿では、推定に使用する文書に依存しない英語能力の推定法を提案する。視点情報は、解答する問題の難易度や文書の長さに影響を受ける。本手法では、従来法に加えて解答する問題の文書特徴を用いることによって、視点情報が文書から受ける影響を取り除くことを考える。結果として、21 人の被験者が英文問題を解答する際の視点情報と問題の正答数、文書特徴を用いて TOEIC スコアを推定した結果、5 文書を用いた際の平均絶対誤差は 96 点となった。

キーワード アイトラッカ、視点情報、英語能力、TOEIC

## 1. ま え が き

近年、グローバル化によって英語の重要性が増し、英語を学習する人が増えている。英語学習は成果が表れるまでに時間を要することから、目標を設定することが重要である。「いつまでに」「どの程度のスキルが欲しいのか」を明確にし、自分の現在の能力と目標の差を把握することによって、具体的な学習計画を立てることができ、着実な学習に繋がる。現在、英語能力を測る指標として最も認知されているのは TOEIC テストである。[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り、TOEIC のスコアは、就職活動や進級、昇進などに利用されている。しかし、TOEIC は、試験時間が 2 時間と長く、受験者に負担が大きいといった課題がある。そこで、本稿では、TOEIC よりも少ない問題数で、手軽に英語能力を測ることを目指す。

TOEIC スコアを手軽に測る方法として、TOEIC の問題集やアプリを利用する方法がある [2] [3]。これらの方法は TOEIC 形式の模擬テストの正答数から TOEIC スコアを測る。問題集は TOEIC と同様に解答に 2 時間かかるが、アプリは所要時間 10 分ほどで結果を取得できる。しかし、TOEIC スコアの推定精度が悪いことが問題であり、問題の解答時間とスコアの精度はトレードオフの関係にある。

そこで、この問題を解決するために、我々は問題解答時の視点情報を用いて TOEIC スコアの推定を目指す。これまでの研究により、視点情報は言語能力と高い関係があることが分かっている [11]。例えば、英語能力の高い者は単語や文法などの知識が豊富で読む速さが速い、読み返しが少ないといった傾向がある。したがって、問題解答時の視点情報を解析することで、英語能力の高いものと低いものを分けることを考える。

以前我々は英文問題解答時の視点情報を用いて TOEIC スコアを推定する手法を提案した [5]。問題の正答数に加えて視点情報を用いることによって、TOEIC よりも少ない長文問題 2 題を解答する際の視点情報を利用して、平均絶対誤差 33 点で推定出来た。しかし、この手法には、他のユーザによって解答時の視点情報が記録された問題でしか英語能力を推定出来ないという課題がある。この原因は視点情報が文書の長さや問題の難しさに影響を受けているためと考える。

そこで、本稿では、推定に使用する文書のコンテンツを考慮して、TOEIC スコアの推定を試みる。具体的には、従来法に加えて、解答する問題の文書特徴を用いることによって、視点情報が文書から受ける影響を除く。結果として、21 人の被験者が英文問題を解答する際の視点情報と問題の正答数、文書特徴を用いて TOEIC スコアを推定した結果、7 文書を用いた際の平均絶対誤差は 91 点となった。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

本節では関連研究として、視点情報を用いた英語能力の推定に関する研究について述べる。その後、本手法で用いる文書特徴について、視点情報との関係を調査した研究について述べ、本手法での視点特徴と文書特徴の利用方法を考える。

吉村らは英文読書時の視点情報を用いて英語能力を推定する手法を提案している [4]。英語能力とは、英語全般の運用能力としており、TOEIC スコアを用いて評価している。TOEIC スコアの得点によって分けられた 3 クラスの英語能力を 90.9% の精度で識別することに成功している。本手法では、英語能力の識別に有効だった、吉村らの特徴量を用いる。また、我々は従来法として英文読書時の視点情報に加えて問題解答時の視点情

報を用いて TOEIC スコアを推定する手法を提案している [5]. 英文問題 2 題の視点情報を用いて TOEIC のスコアを推定し、結果として平均絶対誤差 33 点を得た. 読書時と問題解答時の視点情報を用いて推定を行うことによって吉村らの手法より精度の高い結果が得られることが分かった.

視点特徴と文書特徴を組み合わせて英語能力を推定した例として, Gomez らの手法がある [6]. Gomez らが, 読書時の視点情報の特徴に加えて文書特徴から理解度と言語習熟度を推定する手法を提案している. Gomez らはユーザの英文読書時の視点に関する特徴と, ユーザが読んだ文書中の単語を名詞や動詞といった品詞に分け, 品詞ごとに注視時間を算出した特徴量を用いて, ユーザの文書に対する理解度, 習熟度を推定している. 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たが, 品詞ごとの注視時間に差は見られなかった.

また, 英語能力を表す指標として, 「問題の正答数」×「読みの速さ」という指標がある [7]. Jackson ら, Shizuka らは文書を読んだ際の速さと, その後に解答する問題の正答率をかけた値を用いて読み手言語能力を評価した [8] [9]. 問題の正答数のみを用いて学生の英語の評価した場合「問題の正答数×読みの速さ」を用いて英語能力を評価した結果, 「問題の正答数×読みの速さ」を用いたほうが, 学生の実際の英語能力に近い指標となることが分かっている.

次に, 視点情報と文書特徴に関する研究について述べる. Rayner らは読書時の注視時間が注視している単語の単語頻出度やその単語の持つ意味の数, 前の単語との意味の関係性といった言語特徴に影響を受けると報告している [10]. また, Rayner らは文書の難易度が読書時の眼球運動に与える影響も調査している [11]. 問題を解答する際の人の視点情報は英語能力の影響を受けると考えられる. 例えば, 文書の難易度が増すと, 英語能力の低い読者ほど, 注意深く文書を読むため注視時間が増す, 視点の移動距離が短くなる, 読み返しが多くなるといった影響があることが分かっている.

以上から, 本稿では従来法と同様に, 問題解答時の視点情報と問題の正答数に加えて, 単語の頻出度や文書の難易度といった文書特徴を加えることによって, コンテンツを考慮した TOEIC スコアの推定を試みる. また, 「読みの速さ」は, 「文書の単語数÷読書時間」で算出することが出来る. この様に, 視点情報や問題の正答数, 文書特徴を複合して得られる特徴量を新たに考案することも目標とする.

### 3. 提案手法

#### 3.1 処理の流れ

英文問題解答時の視点情報を用いて, TOEIC スコア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. 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する. 次に得られた視点情報から眼球運動の特性を利用した特徴を, 解答した問題から文書特徴を抽出する. 最後に, 回帰を用いて被験者の TOEIC スコアを推定する.

#### 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



図 1 処理の流れ

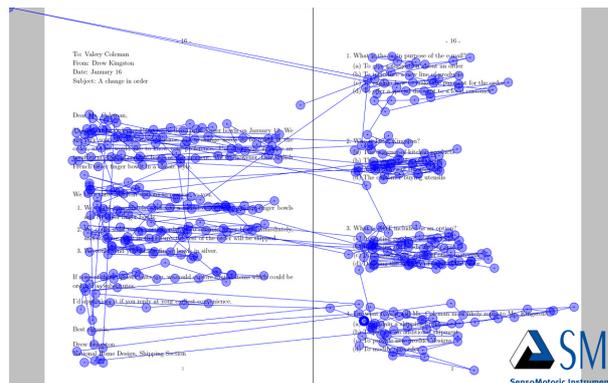


図 2 fixation-saccade 判定後の視点情報

ラッカは眼球運動を測定する装置である. 付属のアイカメラで眼球運動を測定して, ディスプレイ上に表示した英文問題を解答する際の視点情報をディスプレイ上の座標として取得する.

#### 3.3 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, 上記の眼球運動を繰り返すことによって文書を読んでいる. 本手法では英文読書時の視点を fixation, saccade に分類することによってユーザの眼球運動を解析する. fixation, saccade の判定には SMI 社の Begaze を用いた. 図 2 は文書画像上に表現した問題解答時の fixation と, saccade の例である. 円が fixation で, 円をつなぐ線が saccade を表す.

#### 3.4 特徴抽出

本節では, TOEIC スコア推定のための特徴抽出について説明する. 本手法では, 英語能力の推定に, 文書や問題の難易度から受ける影響を考慮するため, 視点情報から得られる特徴と文書や問題の難易度に関する指標を特徴として用いる. また, 視点情報が文書や問題の難易度から受ける影響を考慮するために, 文書や問題の難易度を測定する指標を特徴とする. 表 1 に全ての特徴を示す. まず, 視点情報から得られる特徴として, fixation に関する特徴 saccade に関する特徴を使用する. 次に, 文書の難易度に関する特徴を説明する. 難易度を算出するために用いられる指標として, 1 文当たりの平均単語数や 1 単語当たりの平均文字数, 単語頻出度の平均, 分散がある.

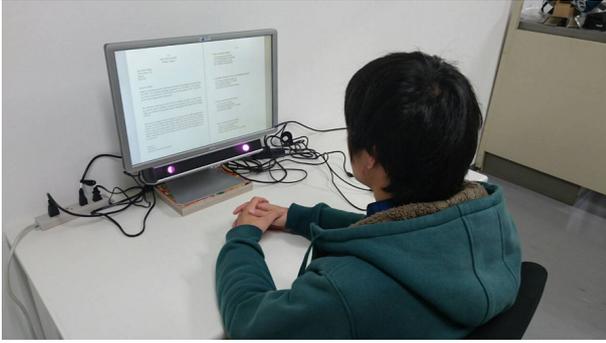


図3 実験風景

また、これらの指標を用いて算出される文書の読みやすさである、readability も特徴として利用する。問題の難易度を考慮するための特徴としては、問題の正答率、高得点者の正答率から低得点者の正答率を引くことで得られる項目弁別力といった値を特徴量とする。

また、英語能力を表す指標として、Shizuka らが検証した「読みの速さ×正確性」を特徴として用いる [9]。一般的に読みの速さは、文書の単語数/読書時間で算出される。本手法では、読書時間の指標として、解答時間や fixation 持続時間の合計といった視点から得られる類似特徴量を利用する。

以上の 42 個の特徴を使用する。

次に、特徴量の算出方法を説明する。問題を解答する際の人の眼球運動は、英語能力の影響を受けると考えられる。しかし、文書ごとに難易度や長さ、問題が異なるため、英語能力の影響にばらつきが生じる。吉村らは、複数文書分の眼球運動から得た特徴量を統合することで、文書によるばらつきを小さくした。本手法でも同様に、1 文書ごとに算出した特徴を、特徴ごとに複数文書分合計する。

### 3.5 TOEIC スコアの推定

抽出した特徴量を最大値 1、最小値 0 で正規化し、リッジ回帰を用いて TOEIC スコアの推定を行う。リッジ回帰は、最小二乗法に正則化項を加えた回帰分析で、通常の重回帰分析に比べ過学習を防ぐ効果がある。また、SVR などの回帰式に比べ、特徴の数が多くてもモデルの計算が速く、TOEIC スコアと特徴量との関係性を明確に示す利点があるため、リッジ回帰を利用する。また、特徴選択手法には、forward stepwise selection を用いる。forward stepwise selection とは、モデルに特徴を一つずつ加えていく選定方法で、特徴を加える前後の推定精度を比較することで、加える特徴の有効性を検証する手法である。どの特徴を加えても推定精度が向上しなくなるまで続けることで、推定精度の高い特徴の組み合わせを検証する。

## 4. 実験

本章では、提案手法による TOEIC スコアの推定精度検証のための実験について述べる。

### 4.1 実験条件

被験者が TOEIC の英文問題を解答する際の視点情報を記

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	一秒当たりの fixation の回数
	fixation 中の視点の分散値の合計, 平均, 最大値,
	fixation 持続時間の平均, 最大値
saccade に関する特徴量	saccade の x 軸距離の合計
	saccade の回数
	saccade 時間の平均, 最大値
	一秒当たりの saccade の回数
	saccade 持続時間の平均, 最大値
	saccade 時の眼球回転角度の平均, 最大値
	saccade 時の角速度の合計, 平均, 最大値
文書に関する特徴量	単語の数
	文の数
	単語頻出度の平均, 分散
	readability
	問題の正答率
	問題の項目弁別力
速さ × 正確さに関する特徴量	単語数/(解答時間 × 問題の正答率)
	正答数 × 単語数/(fixation 持続時間の合計 × 問題の正答率)
	saccade の X 軸距離/単語
	単語数/fixation 持続時間の合計
	単語数/saccade 時の眼球角速度の合計
その他の特徴量	解答時間
	問題の正答数
	問題の正答数/問題の正答率

録した。被験者は 21 人で TOEIC スコアの最高値が 945 点、最低値が 390 点で、平均は 681 点である。実験に使用した問題は TOEIC の対策問題集から選んだ英文問題 15 題である。英文問題は図 2 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、アイトラッカは SMI 社の iViewXTM RED250 を使用した。眼球運動の解析分析には同じく SMI 社の Begaze を使用した。

実験の流れを示す。図 3 の様に被験者はアイトラッカが眼球を認識する位置で頭を固定し、実際の視点の位置とアイトラッカの推定する視点を一致させるキャリブレーションを行った。次に、英文問題を 1 題解答し、問題を解答する際の視点情報を記録した。頭を動かすと視点の推定が正しく行えないため、被験者には頭を出来るだけ動かさないように指示した。また、解答の際に頭が動かないように、被験者には問題の解答が分かった時点で、口頭で実験者に解答を伝える様に指示した。

この手順を 15 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休憩を挟んだ。また、英文問題解答時の自

表 2 選択された特徴量

特徴	選択率
問題の正答数/問題の正答率	1
問題の正答率	0.99
単語頻出度の分散	0.87
解答時間	0.83
正答数 × 単語数/(fixation 持続時間の合計 × 問題の正答率)	0.76
正答数 × 単語数/(解答時間 × 問題の正答率)	0.76
問題の項目弁別力	0.69
文書の行数	0.68
問題の正答数/解答時間	0.65

然な動作や視点情報を記録するため、解答中は頭を動かさないこと以外は被験者の動作に制限を設けなかった。

次に、評価方法について説明する。本実験では、被験者非依存かつ文書非依存の精度検証を行う。被験者非依存の学習を行うために、被験者 1 人をテストデータとして、残りの被験者 20 人を学習データとして分割する。次に、文書非依存の学習を行うための文書の分割方法を説明する。本手法では、3.4 節で説明した通り、一つの文書から得られる特徴ベクトルを複数文書を統合することで、文書特徴から受ける視点情報の影響を小さくすることを考える。そのため、データセットを作成する際は、テストと学習に用いる文書の数を等しく、文書の組み合わせの重複を無くす必要がある。特徴ベクトルを合算する文書の数を  $n$  とすると、まず、テストに用いる  $n$  文書を選択する。次に、全 15 文書の内、テストデータと重複を許さない  $n$  文書の組み合わせをランダムに 20 通り選択し、学習データとする。

以上から、テストデータに 1 人の被験者と文書の組み合わせを 1 通り選択し、残りの被験者 20 と文書の組み合わせ 20 通りからなる 400 のデータを学習に用いる。また、本手法では、特徴を選択する際に、特徴を加える基準として、学習データ内で leave one subject out cross validation による推定精度を行い、最も推定精度が向上する特徴をモデルに加える。

#### 4.2 実験 1: TOEIC スコアの推定精度検証

推定に用いる文書数を 1 から 7 まで変化させた場合の推定精度を図 4 示す。本稿で提案する手法は、問題の正答数と視点情報、文書特徴に関する特徴を用いて TOEIC スコアを推定する。

まず、特徴選択の結果について述べる。推定に用いる文書の数が 1 の時に、選択された特徴量を表 2 に示す。全 42 の特徴の内、9 個の特徴が過半数の訓練モデルに対して選択される結果となった。中でも「問題の正答数/問題の正答率」、「問題の正答率」は、ほぼ全てのモデルで選択されており、推定に用いる文書数が少ない時に重要な要素であると考えられる。選択された特徴の中には、視点に関する特徴は、ほとんど含まれていなかった。唯一多くのモデルで選択された視点に関する特徴は、他の特徴と組み合わせた「正答数 × 単語数/(fixation 持続時間の合計 × 問題の正答率)」のみだった。このことから、1 文書から得られる視点特徴は、文書の難易度の影響を大きく受けており、視点特徴単体では TOEIC スコアの推定に不向きであ

表 3 1 文書を用いて推定した結果

推定に使用した特徴	平均絶対誤差
正答数, 文書, 視点特徴	110.8
正答数のみ	125.8
正答数, 文書特徴	125.0
正答数, 視点特徴	127.7
TOEIC スコアの平均値	137.2

ると考える。また、選択された文書特徴として「文書の行数」や「単語の頻出度の分散」があった。「文書の行数」は実験に用いられた文書では、文の大きな長さを表しており、文書の長さが視点に影響を与えていると考えられる。「単語頻出度の分散」は頻出度の低い単語が多く含まれる場合、視点に影響を与えることを示していると考えられる。表 2 の特徴量を用いて TOEIC スコアを推定した結果、推定された TOEIC スコアと正解値との絶対平均誤差 110 点を得た。

本手法では、問題の正答数に加え、視点特徴と文書特徴を用いて TOEIC スコアを推定している。そこで、比較手法として、「問題の正答数のみ」、「正答数と視点特徴」、「正答数と文書特徴」を用いて TOEIC スコアを推定した結果と被験者の TOEIC スコアの平均値との平均絶対誤差を表 3 に示す。表 3 より、最も推定精度が高いのは、全種類の特徴を用いる提案手法であることがわかる。このことから、表 2 で選択された視点情報と文書特徴を組み合わせた「正答数 × 単語数/(fixation 持続時間の合計 × 問題の正答率)」が精度向上に貢献していると考えられる。

先行研究から、推定に用いる文書数を増やした場合、推定精度が向上することがわかっている [5]。そこで、推定に用いる文書数を 1 から 7 まで増やした場合の結果について述べる。テストに用いる文書の組み合わせをランダムに 20 通り選び、その平均誤差を求めた。推定した平均絶対誤差を図 4 に示す。比較手法として、「問題の正答数のみ」、「正答数と視点特徴」、「正答数と文書特徴」を用いて TOEIC スコアを推定した結果も図 4 に示している。図 4 より、全ての手法において、使用する文書数が最も少ない 1 文書の時に比べて、文書数を増やすにつれて精度は向上している。提案手法で最も推定結果が良いのは文書数が 7 の時で、推定結果が 91.3 点となった。

推定に用いる文書数を増やした場合の比較手法と比べると、推定に用いる文書数が 4 の時、「問題の正答数のみ」と「正答数と文書特徴」を用いて推定した場合の方が精度がよい。この原因は、提案手法の特徴選択にあると考える。表 4 に、推定に用いる文書数別に選択された数が多い上位 5 つの特徴を示す。表 4 より、推定に用いる文書数が 1 の時に選択された特徴は、選択率が高い。対して、文書数を増やした場合に選択される特徴は、選択率が低い。文書数が増えると問題の正答数を見るだけで精度よく TOEIC の点数が推定できるため、その他の特徴の選択が推定モデルの過適合を引き起こしていると考えられる。

#### 4.3 実験 2: 最良の特徴選択をした場合の精度検証

実験 1 では、推定モデルの過適合により、汎化性能が悪くなり、推定精度が下がった。そこで、4.2 で述べた手法のように

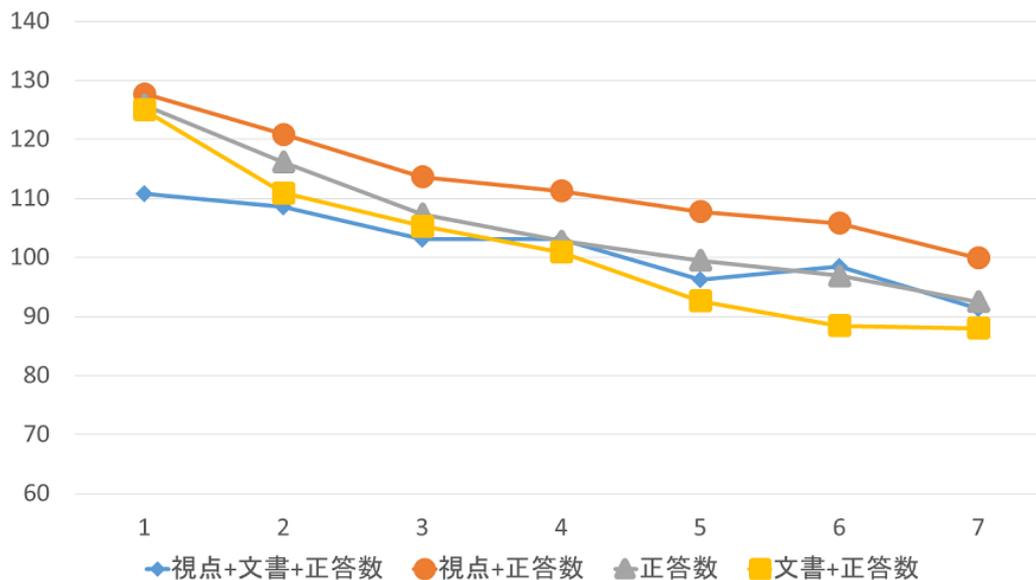


図 4 文書数別の推定精度

表 4 選択された特徴の順位

文書数	1 位 (選択率)	2 位 (選択率)	3 位 (選択率)	4 位 (選択率)	5 位 (選択率)
1	正答数/正答率 (1)	正答率 (0.99)	単語頻度度の分散 (0.87)	解答時間 (0.83)	(正答数 × 単語数)/ (fixation 時間 × 正答率) (0.76)
2	正答数/正答率 (1)	正答率 (0.92)	単語頻度度の分散 (0.68)	解答時間 (0.61)	問題の項目弁別力 (0.56)
3	正答数/正答率 (0.99)	正答率 (0.55)	fixation 時間の合計 (0.5)	頻度度の分散 (0.48)	行数 (0.42)
4	正答数/正答率 (0.99)	fixation 時間の合計 (0.73)	正答率 (0.37)	行数 (0.35)	頻度度の分散 (0.30)
5	正答数/正答率 (0.94)	fixation 時間の合計 (0.61)	正答率 (0.33)	解答時間 (0.26)	行数 (0.24)
6	正答数/正答率 (0.86)	fixation 時間の合計 (0.56)	正答率 (0.4)	正答数 (0.3)	単語頻度度の分散 (0.25)
7	正答数/正答率 (0.61)	fixation 時間の合計 (0.52)	正答率 (0.45)	正答数 (0.38)	単語数 (0.32)

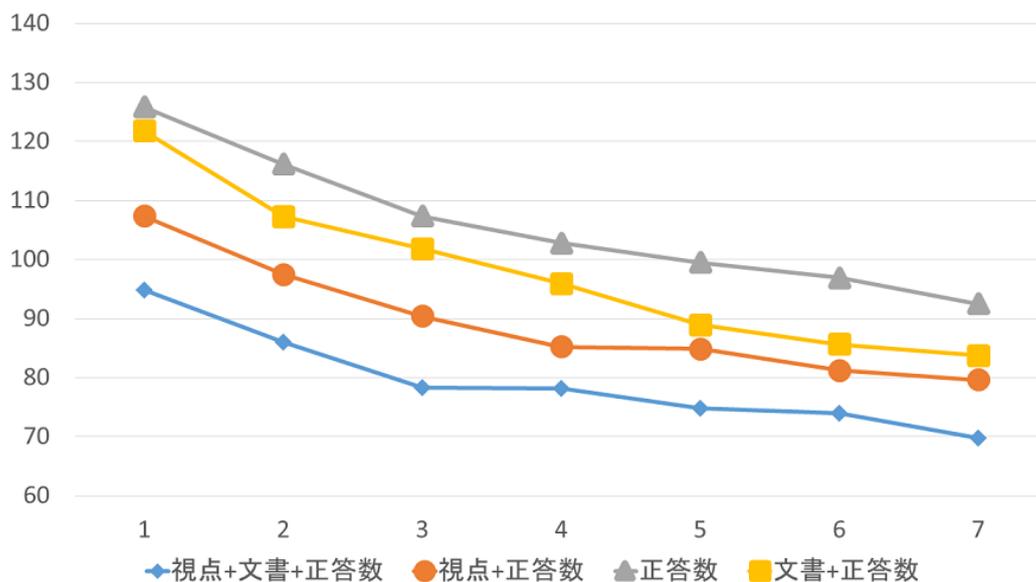


図 5 推定結果が最も良くなる時の推定精度

特徴選択を被験者ごとに行うのではなく、全ての被験者に共通した特徴を選択することで、過適合を防ぐことを試みる。具体的な手順は以下の通りである、まず、実験 1 と同様の方法で

テストデータと学習データを分割する。次に、学習データから回帰式を作成し、テストデータを入力し、推定結果を算出する。この時、全ての被験者の推定結果が最も良くなる様に共通の特

徴を選択する。

結果を図5に示す。全ての特徴を用いて TOEIC スコアを推定した場合が、推定に用いる文書数が増え変化しても最も良い精度が得られた。文書数が7の時、推定精度が最も良く、正解ラベルとの平均絶対誤差は69.78点だった。

## 5. ま と め

本稿では、21人の被験者が15題の英文問題を解答する際の視点情報と問題の正答数、文書特徴を用いて TOEIC スコアを推定した。最も推定結果が良くなるように特徴量を選定した結果、7文書を用いた時、TOEIC スコアの推定値と正解値の平均絶対誤差91点を得られた。問題の正答数に視点特徴と文書特徴を用いて TOEIC スコアを推定することによって、文書に依存しない TOEIC スコアの推定が行える可能性を示した。

今後の課題としては、精度向上のために、「問題の正答数×読みの速さ」といった複合した特徴を新たに考案すること。視点特徴単体では文書に依存しない TOEIC スコアの推定効果が無かったため、SVRを用いて、文書特徴と組み合わせを計算することを目標とする。

## 謝 辞

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] 『TOEIC』, <http://www.toEIC.or.jp>
- [2] 『TOEIC 公式問題集』, <http://square.toEIC.or.jp/kyouzai/store/pr/08/index.html>
- [3] 『TOEICRTTEST 実力判定「アブトレ」』, <https://itunes.apple.com/jp/app/toEIC-test-shi-li-pan-ding/id1048914472?mt=8>
- [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] 藤好 宏樹, 石丸 翔也, Olivier Augereau, 黄瀬浩一, “視点情報を用いた英語能力推定法の実験的検討,” 電子情報通信学会技術研究報告, vol. 115, no. 517, PRMU2015-195, pp. 185-195, March. 2016
- [6] 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.
- [7] 門田 修平, 野呂 忠司, 『英語リーディングの認知メカニズム』(くろしお出版, 2001)
- [8] Jackson Mark D, James L. McClelland. “Processing determinants of reading speed,” Journal of Experimental Psychology: General, 108.2 (1979): 151.
- [9] Shizuka T, “Attaining 40-item-equivalent reliability and validity with only 4 items: Effects of incorporating confidence and speed measures in a computer-based reading test,” A Paper Presented at Kanto-Koshin-Etsu English Language Education Society 24th Annual Conference, 2000.
- [10] Keith Rayner, Monica S. Castelano, “Eye movements during reading, scene perception, visual search, and while looking at print advertisements,” Visual marketing: From attention to action, pp. 9-42, 2008
- [11] Keith Rayner, “Eye movements in reading and information

processing: 20 years of research,” Psychological bulletin, vol. 124, no. 3, pp. 372-422, Nov. 1998