

# 文書への依存性を考慮した TOEIC スコア 推定法の実験的評価

星加 健介<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: †hoshika@m.cs.osakafu-u.ac.jp, ††augereau.o@gmail.com, †††kise@cs.osakafu-u.ac.jp

あらまし 英語能力を測定する方法に TOEIC がある。しかし、TOEIC は試験時間と結果発表まで時間がかかり、直ぐに結果を知ることができないという問題がある。そのため、現在、TOEIC の試験を受けることなく、英語能力を推定する手法が提案されている。本稿では、英語能力推定時に重要な 3 つの要素について考察する。1 つ目の要素は、学習データとテストデータのユーザが同じであるかである。2 つ目の要素は、学習データとテストデータの文書が同じであるかである。3 つ目の要素は、特徴選択が推定に使用されるユーザと文書に依存するかである。これら 3 つの要素によって、推定精度や実際のシステムとしての使用方法が異なる。本稿では、学習データとテストデータのユーザが異なる場合、文書と特徴選択に関する要素を変更する手法を提案する。実験の結果、学習データとテストデータでユーザと文書が異なる場合、ユーザが読んだ 4 文書のデータを推定に用いることで、平均絶対誤差 73.8 点を得た。

キーワード アイトラッキング, 視点情報, 文書情報, TOEIC

## 1. はじめに

近年、グローバル化により、日常的に英語を使用する機会が増え、英語学習の必要性が高まっている。一般的に英語能力は Reading, Listening, Speaking, Writing という指標から評価される。これらの英語能力を確認する方法に TOEIC [1] がある。TOEIC は国際コミュニケーションのための英語能力測定試験であり、進学や就職など様々な場面で使用される。しかし、TOEIC は試験時間が 2 時間であるため、受験者の負担が大きいという問題がある。また、結果発表まで 1 か月かかるため、直ぐに結果を知ることができないという問題もある。そのため、現在、TOEIC の試験を受けることなく、英語能力を推定する研究が多く行われている。手軽に自身の英語能力を確認することで、目標とする能力を身に着けるために効率的な学習計画を立てることができる。また、英語能力の推移を確認することで、日々の英語学習のモチベーション維持にも貢献できる。そこで、本研究では、手軽に英語能力を推定することを目指す。

英語能力を推定する際には、重要な 3 つの要素が存在する。3 つの要素の関係を表 1 に示す。1 つ目の要素は、学習データとテストデータのユーザが同じであるかである。学習データとテストデータのユーザが同じでなければならない手法をユーザ依存の手法と呼ぶ。ユーザ依存の手法は 1 人のユーザに対して、ユーザ非依存の手法は全てのユーザに対して適用可能である。2 つ目の要素は、学習データとテストデータの文書が同じであるかである。学習データとテストデータの文書が同じでなければならない手法を文書依存の手法と呼ぶ。文書依存の手法では、手法を構築する際に用いた文書を、ユーザが読むという条件のもとで、英語能力を推定する。文書非依存の方法は、ユー

ザが読む文書に制限のない手法である。3 つ目の要素は、ユーザ依存、または文書依存な場合に、特徴選択がユーザあるいは文書に依存するかである。これを特徴選択のユーザあるいは文書への依存性と呼ぶ。特徴選択がユーザあるいは文書に依存する場合には、ユーザあるいは文書が変更されると、特徴選択からやり直す必要があることを意味する。一方で依存しない場合には、特徴選択は適用し直す必要はなく、学習のみ再実行するだけで済む。これら 3 つの要素によって、英語能力の推定精度と実際のシステムの使用方法は変化する。

表 1 のユーザ依存、文書依存の組み合わせのうち、ユーザと文書の双方に依存した手法を構築することはできない。従って、これら以外の 3 つの組み合わせが研究の対象となる。ユーザ依存の手法は 1 人のユーザにのみ適用可能であるため、手法の一般性が低い。そのため従来研究では、英語能力を推定する手法がユーザ非依存の手法である場合が多い [2-6]。本研究でも、ユーザ非依存の推定を対象とする。ユーザ非依存の推定を行う従来研究の中には、文書依存だけではなく非依存の場合を考慮した研究も存在する [5,6]。この際、特徴選択は文書依存で行われていることが多い。

以上の状況の下、本稿では、新しい英語能力推定法を提案するとともに、表 1 に示す 3 つのタイプのいずれかではなく、全てについて、推定精度を評価する。すなわち、タイプ 1 (ユーザ非依存で文書依存、かつ特徴選択が文書依存の手法)、タイプ 2 (ユーザ非依存で文書依存、かつ特徴選択が文書非依存の手法)、ならびにタイプ 3 (ユーザ非依存で文書非依存の手法) の 3 通りについて評価する。実験の結果、タイプ 1 では、ユーザが読んだ 3 文書のデータを用いることで平均絶対誤差 10.8 点を得た。タイプ 2 では、ユーザが読んだ 4 文書のデータを用い

表 1 要素の関係

		ユーザ	
		依存	非依存
文書	依存	存在しない	タイプ 1: 特徴選択が文書依存 タイプ 2: 特徴選択が文書非依存
	非依存	本稿の対象外	タイプ 3

ること平均絶対誤差 70.8 点を得た。タイプ 3 では、ユーザが読んだ 4 文書のデータを用いることで平均絶対誤差 73.8 点を得た。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する。

## 2. 関連研究

本節では、英語能力に関する研究や英語能力推定に関する研究について述べる。

英語能力と語彙力に関する研究に鏗木らの研究がある [7]。鏗木らは、21 人の実験参加者が英訳した文から単語の出現率に関するコーパスを作成した。作成したコーパスは、JACET8000 [8] から作成したコーパスより実験参加者の英語能力に高い相関を示した。同様に安田らも 469 人の実験参加者が英訳した文からコーパスを作成し、英語能力との相関関係を求めた [9]。結果として、GTM [10] と実験参加者自身が英訳対象の文に感じる難易度、スペルミスの指摘回数との相関が高いことが確認できた。読解効率と英語能力の相関に関して、小谷らの研究がある [11]。小谷らは、設問の正答数と読解速度の積を読解効率と定義した。102 人の実験参加者に対して、英語能力を推定するテストを作成して、信頼性を調査した結果、読解効率と TOEIC の Reading section のスコアの相関係数は 0.73 であった。

Writing に関して、自由英作文から英語習熟度を推定する研究に小林らの研究がある [2]。小林らは TOEIC スコアを 4 クラスに分類し、総単語数や異語数など 21 個の特徴量とランダムフォレストを使用して英語能力を推定した。相関が高い特徴量は異語率と平均文長であり、推定精度は 58.23% であった。Listening に関して、英語リスニング中の脳血流量を使用して TOEIC スコアを推定する研究に亀山らの研究がある [4]。亀山らは脳血流量の分散や平均から主成分分析によって特徴量を抽出し、lasso 回帰を使用して、実験参加者の TOEIC スコアを推定した。実験の結果、最小の誤差は 89.93 点であった。英文読書時の視点情報を用いて、英語能力を推定する研究に吉村らの研究がある [3]。吉村らは TOEIC スコアを 3 クラスに分類して、90.9% の精度で識別することに成功している。吉村らの手法では、fixation の持続時間と saccade 時の眼球運動の速度が推定に有効であると示されている。

藤好らは視点情報と文書情報を用いて、英語能力を推定する手法を提案している [5,6]。吉村らと異なり、藤好らの手法では英文読書時に加えて問題解答時の視点情報を使用して、TOEIC スコア自体を推定する。藤好らはユーザ非依存で文書依存の場合、推定に 3 文書のデータを使用するとき、平均絶対誤差 32 点を得た。また、ユーザ非依存で文書非依存の場合、推定に 5 文書のデータを使用するとき、平均絶対誤差 96 点を得た。TOEIC

スコアは約 35 点の誤差があると言われており、文書依存の場合、十分な推定精度であると言える。

## 3. 提案手法

本節では、まず TOEIC スコアの推定方法について、ユーザ依存、文書依存という観点から説明する。その後、提案手法の処理の流れを述べる。

### 3.1 TOEIC スコアの推定方法

まず、ユーザ依存や文書依存という観点から、TOEIC スコアの推定方法の難易度と一般性について説明する。

従来研究では、TOEIC スコアを推定するために、視点情報が用いられる [5,6]。視点情報はコンテンツやレイアウトなどの文書の差、興味などの心理状態、疲れなどの身体状態、言語能力などの様々な要素を反映している。英語能力推定法では、このうち言語能力に起因する差異を取り出し言語能力を推定するため、可能な限り他の要素が同じであることが望ましい。

文書依存の手法は、文書に起因する差異を生じにくくするため、文書非依存の手法より、推定精度が一般に高くなる。一方、文書非依存の手法は、視点情報が文書から受ける影響を取り除く必要があり、文書依存の手法より推定が難しい。

ユーザ依存の方法は、ユーザに起因する様々な差異を生じにくくするので、同様に、推定が容易になる。ただし、ユーザ依存で TOEIC スコアを推定することが意味を持つのは、時間的な変化を扱う場合のみであり、適用場面が限られてくる。注意すべき点は、ユーザ依存かつ文書依存の方法は意味を持たないことである。これは、同じ文書を用いて同じユーザの TOEIC スコアを推定することが意味をなさないためである。

以上を考えると、TOEIC スコアの推定方法は、ユーザ依存か否か、文書依存か否かによって、推定の容易さや適用性が変わってくると言える。最も適用範囲が広いが、推定が一番難しいのは、ユーザ非依存でかつ文書非依存な推定である。ユーザ非依存で文書依存の方法は、それに比べて推定が容易になるものの、一度、用いた文書は再度同じユーザの能力推定に用いることができないため、その点への注意が必要である。従って、文書依存の手法では、文書の追加が容易にできるかどうか、重要なポイントとなる。

特徴線選択が文書に依存しない場合には、新しい文書から同様に特徴を抽出し、それを追加して学習を行えば良い。一方で、特徴選択が文書に依存する場合には、すべてのデータについて、特徴選択から適用し直す必要があるため、手間が増える。従って、文書依存の方法の中では、特徴選択が文書に依存しないものは、依存する場合に比べて、適用性が高いものの、性能は低くなる可能性があると言える。

### 3.2 処理の流れ

提案手法の処理の流れ図 1 に示す。始めに、アイトラッカを用いて、視点情報を取得する。次に、取得した視点情報に対して fixation-saccade 判定を行い、特徴量を抽出する。その後、遺伝的アルゴリズムによって特徴量を選択し、SVR(Support Vector Regression) を使用して TOEIC スコアを推定する。以下、詳細を述べる。



図 1 処理の流れ

### 3.2.1 視点情報の取得

アイトラッカを用いて、ユーザの視点情報を取得する。アイトラッカは、内蔵されたカメラでユーザの眼球運動を計測して、視点情報を取得する機器である。本手法では、据置型アイトラッカを使用して、ディスプレイ上の英文問題を読む際の視点情報を取得する。

### 3.2.2 fixation-saccade 判定

眼球運動は、fixationとsaccade, blinkに分類することができる。fixationは視点がある範囲内で一定時間留まること、saccadeはfixation間の素早い動き、blinkは瞬きのことである。人はfixationとsaccadeを繰り返すことで、ものを見たり読んだりしている。本手法ではSMI社のBegazeによって、視点情報をfixationとsaccadeに分類する。図2は文書画像上にfixationとsaccadeを表現した例である。円がfixationを表し、円と円間の線がsaccadeを表す。

### 3.2.3 特徴量の抽出

fixation-saccade判定後の視点情報と文書情報を使用して、特徴量を抽出する。抽出する特徴量の一覧を表2に示す。特徴量は、視点情報に関する特徴量が50個、文書情報に関する特徴量が10個、視点情報と文書情報、正答数を組み合わせた特徴量が5個、その他の特徴量が3個である。

視点情報はユーザ特有の情報であり、英語能力の差を表現することに有効である[3, 5, 6]。文書情報を用いることで、視点情報が文書から受ける影響を取り除くことができる可能性がある[6]。表2に示されるCLI[12]は文書の読みやすさの指標であり、 $C_{LI}$ とすると式(1)で表される。

$$C_{LI} = 5.89 \times A_c - \frac{30}{A_w} - 15.8 \quad (1)$$

式(1)の $A_c$ は文書内の1単語あたりの平均文字数、 $A_w$ は文書内の1文あたりの平均単語数である。組み合わせた特徴量はJacksonらやShizukaら、小谷らの研究で示されている英語能力を表す指標を参考にして作成する[11, 13, 14]。表2の設問の正答数とは、一つの文書に対して、ユーザが正解した設問の数である。表2の問題の正答率 $R$ は、

$$R = \frac{C_n}{U_n \times Q_n} \quad (2)$$

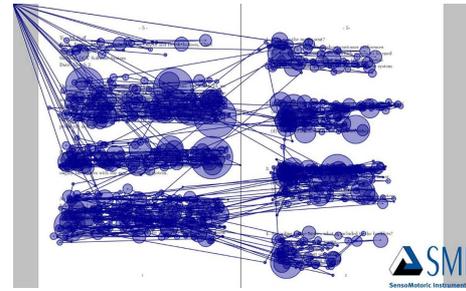


図 2 fixation-saccade 判定後の視点情報



図 3 問題文と設問の区切り

で表される。ここで、 $U_n$ はユーザ数、 $Q_n$ はある文書の設問数、 $C_n$ は全ユーザがその文書の設問を正解した数である。

本手法では、図3のように文書を問題文と設問に分割して、問題文と設問に関する特徴量を新たに追加する。図3のような文書に関して、ユーザが問題文を読み、設問に解答する際、ユーザ独自の解答方法が存在する。例えば、問題文と設問にかける時間の割合や問題文と設問の往復回数はユーザによって異なる。このようなユーザ独自の解答方法は英語能力の差を表す可能性がある。

表2に示す68個の特徴量は文書毎に作成する。視点情報は文書の差、心理状態、身体状態、言語能力などの要素を反映している。文書毎に視点情報を取得するとき、文書の差や心理状態、身体状態は異なり、特徴量にばらつきが生じる。特徴量にばらつきが生じている際、ユーザ本来の言語能力と異なる特徴を学習する可能性があり、好ましくない。そこで、ばらつきを軽減するために、学習と推定に使用する複数文書の特徴量を合算する。特徴量の合算方法について説明する。例えば、 $m$ 文書を使用する場合、式(3)によって合算後の特徴量を作成する。

$$F = \left( \sum_{i=1}^m f_1^{d_i}, \sum_{i=1}^m f_2^{d_i}, \dots, \sum_{i=1}^m f_{n_f}^{d_i} \right) \quad (3)$$

式(3)の $F$ は合算後の特徴量、 $m$ は文書の数である。また、 $f_{n_f}^{d_i}$ は文書 $d_i$ の $n_f$ 番目の特徴量を表す。

### 3.2.4 特徴量の選択

遺伝的アルゴリズムによって、推定に使用する特徴量を選択する。遺伝的アルゴリズムはHollandの単純なアルゴリズムを使用する[15]。本手法で、個体の選択方法はエリート保存戦略とトーナメント戦略であり、交差方法は一様交差である。遺伝的アルゴリズムで用いられる適応度は式(4)から得られる平均

表 2 特徴量一覧

特徴量	詳細
fixation に 関する特徴量	fixation の数, 頻度
	fixation 中の視点の分散値の 合計, 平均, 最大値, 最小値
	fixation 持続時間の 合計, 平均, 最大値, 最小値
	問題文における fixation の数, 比率, 持続時間の合計, 最大値, 最小値, 平均
	設問における fixation の数, 比率, 持続時間の合計, 最大値, 最小値, 平均
saccade に 関する特徴量	saccade の $x$ 軸距離の 平均, 分散, 最大値
	saccade の平均潜在時間, 数, 頻度
	saccade 持続時間の 合計, 平均, 最大値, 最小値
	saccade 時の眼球回転角度の 合計, 平均, 最大値, 最小値
	saccade 時の角速度の 合計, 平均, 最大値, 最小値
	視点の総移動距離
	問題文, 設問, 問題文と設問間の saccade の数
blink に 関する特徴量	blink の数, 頻度
	blink 持続時間の 合計, 平均, 最大値, 最小値
	単語, 文, 行, 文字の数
文書に 関する特徴量	単語頻出度の平均, 分散
	CLI(文書の読みやすさ)
	単語, 文の平均の長さ
	平均単語水準
組み合わせた 特徴量	(正答数 $\times$ 単語数) / (解答時間 $\times$ 問題の正答率)
	(正答数 $\times$ 単語数) / (fixation 持続時間の合計 $\times$ 問題の正答率)
	正答数 $\times$ saccade 時の角速度の平均
	単語数 / fixation 持続時間の合計
	単語数 / saccade 時の角速度の合計
その他の 特徴量	解答時間
	設問の正答数
	設問の正答数 / 問題の正答率

絶対誤差である.

$$S_a = \frac{\sum_{i=1}^{u_n} |S_i - S_e|}{u_n} \quad (4)$$

式 (4) の  $S_a$  は平均絶対誤差,  $u_n$  はユーザの数,  $S_i$  はユーザ本来の TOEIC スコア,  $S_e$  は本手法によって推定された値である. 平均絶対誤差を小さくするように, 遺伝的アルゴリズムで世代を更新する.

### 3.2.5 英語能力推定

遺伝的アルゴリズムにより選択された特徴量のみを使用して, SVR(Support Vector Regression) によって, TOEIC スコアを推定する.

## 4. 実験

本節では, 提案手法の推定精度を検証した結果と考察について述べる. なお, 実験で使用するデータセットは藤好らのデータセット [6] と同一である.

### 4.1 実験条件

本稿での実験参加者は日本人大学生 21 人である. 実験参加者の TOEIC スコアの範囲は最低 390 点から最高 945 点である. 実験に用いた文書の形式は TOEIC の長文読解問題であり, 図 3 のように左側に問題文, 右側に設問という構成で統一している. 各文書の設問は 4 問である. 各実験参加者が 15 文書を読み, 解答する際の視点情報を据置型アイトラッカを使用して取得した. アイトラッカは SMI 社の iViewXTM RED 250 であり, 眼球運動の解析には SMI 社の Begaze を使用した.

手法の各種パラメータは以下のように設定した. 特徴選択に用いた遺伝的アルゴリズムについては, 初期個体数は 200 個, 世代更新回数は 100 回, 突然変異率は遺伝子長  $L = 68$  より,  $\frac{1}{68}$  と設定した. SVR については, 線形カーネルを用いてコストパラメータを  $C = 0.001$  とした.

次に, 実験結果の評価方法について述べる. 交差検証は Leave-One-User-Out Cross-Validation を用いる. これは, 1 人のユーザをテストデータとして, 残りの 20 人を学習データとする手法である.

文書依存な場合と文書非依存な場合における文書の選択方法について述べる. 文書依存の場合は, テストデータと学習データで共通の  $n_1 (n_1 = 1, 2, \dots, 15)$  文書を選択する. 文書非依存の場合は, テストデータとして, 全 15 文書から,  $n_2 (n_2 = 1, 2, \dots, 7)$  文書を選択する. 次に, 学習データとして, テストデータで選択された文書と異なる  $n_2$  文書を選択する. 選択する文書の組み合わせが 100 通り以上である場合, ランダムに 100 通り選択する.

特徴選択については以下の通りである. 文書依存の推定方法でかつ特徴選択が文書非依存の場合, 推定に用いる文書数ごとに, 全データを用いて特徴選択を適用して, 選択された特徴を学習に用いる. 一方, 文書依存の特徴選択を行う方法では, 学習に用いる文書の集合ごとに, 適切な特徴量の特徴選択法によって求め, それを学習に用いる. 従って, 一般には, 学習に用いる文書数が同じであっても, 異なる特徴量が学習に用いられることになる.

比較手法として用いる藤好らの手法と本手法の主な相違点を表 3 に示す. 本手法は, 藤好らの手法に比べて, 用いる特徴量の数が多い. また, 特徴選択法や学習法が異なるものとなっている.

### 4.2 実験結果・考察

本節では, 表 1 で示した 3 つのタイプについて, 実験結果と考察を述べる. タイプ 1 は, ユーザ非依存で文書依存(特徴選択も文書依存)の手法, タイプ 2 は, ユーザ非依存で文書依存(特徴選択は文書非依存)の手法, タイプ 3 は, ユーザ非依存で文書非依存の手法である.

表 3 藤好らの手法との相違点

	文書依存の手法			文書非依存の手法	
	タイプ 1	タイプ 2	藤好らの手法	タイプ 3	藤好らの手法
特徴量数	68		34	68	42
特徴選択法	遺伝的アルゴリズム		backward stepwise selection	遺伝的アルゴリズム	forward stepwise selection
特徴選択の文書への依存	依存	非依存	依存	非依存	非依存
学習法	SVR		重回帰	SVR	ridge 回帰

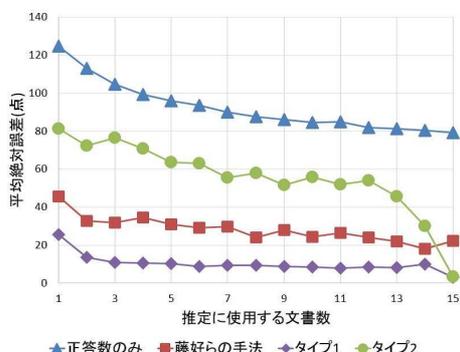


図 4 ユーザ非依存で文書依存の実験結果

表 4 タイプ 1 で選択された特徴量

特徴量	割合 (%)
設問の正答数	83.8
saccade 時の眼球回転角の最小値	55.7
単語数 / saccade の角速度の合計	54.8
単語の平均の長さ	52.2
saccade の $x$ 軸移動距離の分散	51.9

#### 4.2.1 タイプ 1 の実験結果

タイプ 1 の実験結果を図 4 に示す。図 4 に、タイプ 1 の結果に加えて、比較手法として、正答数のみを使用して推定した結果と藤好らの手法による結果も示す。タイプ 1 の平均絶対誤差は藤好らの手法の半分以下である。このことから、提案手法で用いた特徴量や特徴選択法、学習法が効果を挙げていることが分かる。

タイプ 1 で、選択された回数が上位である主な特徴量を表 4 に示す。表 4 より、saccade に関する特徴量が多く選択されている。saccade はユーザの視点の動きを意味しており、英語能力の差が視点の動きに関して顕著に表れたと考えられる。特に、saccade の  $x$  軸移動距離とは、視点の左右の動きを意味している。これは本実験で、ユーザが文書を読む際、問題文や設問を繰り返し読む回数に関係している。本実験のユーザの視点を確認したところ、TOEIC スコアが高くなるにつれて、繰り返し読む回数が減少する傾向があることを確認できた。

一方、藤好らの手法で選択された回数が上位である主な特徴量を表 5 に示す。表 5 では、fixation に関する特徴量が多く選択されているが、表 4 では選択されていない。一方、表 4 では saccade の  $x$  軸移動距離に関する特徴量が選択されているが、表 5 では選択されていない。タイプ 1 は藤好らの手法より推定精度が良いことから、saccade の  $x$  軸移動距離は推定精度の向上に有効であると言える。

表 5 藤好らの手法における選択回数が上位の特徴量

文書依存な手法	文書非依存な手法
解答時間	設問の正答数 / 問題の正答率
fixation 持続時間の合計	設問の正答数
設問の正答数	fixation 持続時間の合計
fixation の数	単語頻出度の分散
saccade の数	解答時間
saccade 持続時間の合計	行数

表 6 タイプ 2 で選択された特徴量

特徴量	割合 (%)
設問の正答数	100
設問の正答数 / 問題の正答率	93.3
saccade の $x$ 軸移動距離の平均	80
saccade の $x$ 軸移動距離の分散	
saccade 時の眼球回転角度の最小値	
文の数	

#### 4.2.2 タイプ 2 の実験結果

タイプ 2 の実験結果を同様に図 4 に示す。特徴選択を文書に依存せずに行うことで、TOEIC スコアの推定精度が大幅に低下していることがわかる。このことから、文書に依存した特徴選択は、文書に起因する変動を吸収する上で有効であることがわかる。ただし、特徴選択を文書に依存せずに行う場合であっても、正答数のみを用いた TOEIC スコアの推定よりは推定精度が高いことから、視線情報を用いることは推定に有効であることがわかる。

なお、文書に依存した特徴選択を用いた結果のうち、15 文書を用いた場合の結果が、文書非依存の特徴選択を行う結果と同じになっている。これは、15 文書を用いて学習する場合、作成できる文書の集合が 15 文書すべてを含むもの 1 通りになるため、文書非依存の場合と同じデータを使って特徴選択を行うことになる。これが同じ結果になる理由である。

次に、タイプ 2 で、選択された回数が上位である主な特徴量を表 6 に示す。表 6 でも、表 4 と同様に saccade に関する特徴量が多く選択されている。タイプ 1 とタイプ 2 はユーザ非依存で文書依存の手法である。ユーザが同じ文書を読む際、英語能力の差が視点の動きに表れている可能性がある。

#### 4.2.3 タイプ 3 の実験結果

タイプ 3 の実験結果を図 5 に示す。比較手法として、正答数のみを使用して推定した結果と藤好らの手法による結果を示している。図 5 より、推定に使用する文書数に関わらず、タイプ 3 は藤好らの手法より推定精度が良いことがわかる。ただし、提案手法の優位性は、図 4 に示した文書依存の場合に比べて、

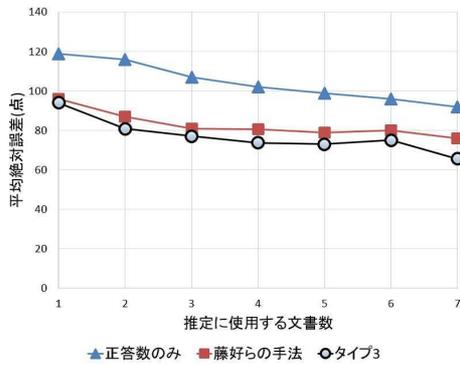


図5 ユーザ非依存で文書非依存の実験結果

表7 タイプ3で選択された特徴量

特徴量	割合 (%)
設問の正答数	100
設問の正答数 / 問題の正答率	
fixation の数	85.7
単語数	71.4
CLI(文書の読みやすさ)	
問題文における fixation 持続時間の平均	57.1
saccade の x 軸移動距離の平均	
saccade の x 軸移動距離の分散	
文の平均の長さ	
行数	
単語頻出度の分散	

低くなっていることも事実である。これはユーザ非依存で文書非依存の手法の難しさを示している。さらなる推定精度の向上のためには、ユーザが異なる文書を読む際、英語能力の差を表す工夫が必要である。

タイプ3で、選択された回数が上位の特徴量を表7に示す。表7では、視点情報の特徴量に加えて、単語数やCLIなどの文書情報の特徴量が多く選択されている。ユーザが異なる文書を読む際、単語数やCLIなどの文書情報の特徴量が推定に有効であると考えられる。また、表7では、問題文における fixation 持続時間の平均も選択されている。この特徴量は、ユーザがTOEICの長文読解問題を解答する際の戦略が推定に有効であると考え、本手法で新たに追加したものである。

## 5. まとめ

本稿では、TOEICスコアを推定する手法として、ユーザ非依存の各種手法について実験的に評価した。従来の藤好らの手法でも設定されていたユーザ非依存でかつ文書非依存の推定方法(タイプ3)、ユーザ非依存で文書依存(特徴選択も文書依存)(タイプ1)の推定方法については、提案手法が、既存手法である藤好らの手法を上回る性能を持つことがわかった。また、ユーザ非依存で文書依存(ただし特徴選択は文書非依存)(タイプ2)の手法を評価することで、特徴選択を文書に依存して行うことの効果を検証した。

今後の課題として、特徴量を再考することが挙げられる。提案手法では、英語能力の差を表すことができる有効な特徴量を

人手で定めていた。大量の学習データが利用可能であれば、これを深層学習によって置き換えることが考えられる。

謝辞 本研究の一部は、JST CREST (Grant No. JP-MJCR16E1), 日本学術振興会科学研究費補助金挑戦的萌芽研究(15K12172), ならびに大阪府立大学キーププロジェクトの補助による。

## 文 献

- [1] “TOEIC,” <http://www.toEIC.or.jp>.
- [2] 小林雄一郎, “ランダムフォレストを用いた英語習熟度の自動推定,” 言語処理学会第18回年次大会発表論文集, 2012, pp.979–982, 2012.
- [3] K. Yoshimura, K. Kunze, and K. Kise, “The eye as the window of the language ability: Estimation of english skills by analyzing eye movement while reading documents,” Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR2015), pp.251–255, Aug. 2015.
- [4] 亀山拓朗, 錦木崇史, 栗原陽介他, “英語リスニング中の脳血流量を用いた正則付回帰モデルによる言語能力推定法,” 第78回全国大会講演論文集, vol.2016, no.1, pp.685–686, 2016.
- [5] 藤好宏樹, 石丸翔也, O. Augereau, 黄瀬浩一, “視点情報を用いた英語習熟度推定法の実験的検討,” 電子情報通信学会技術研究報告, PRMU2015-195, 第115巻, pp.185–190, March 2016.
- [6] 藤好宏樹, O. Augereau, 黄瀬浩一, “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,” 電子情報通信学会技術研究報告, 第116巻, pp.7–12, Feb. 2017.
- [7] 鏑木元, 安田圭志, 山本博史, 匂坂芳典, “統計的語彙情報に基づく日本人学習者の英語習熟度の分析,” 言語処理学会第16回年次大会, pp.539–542, 2010.
- [8] “JACET8000,” <http://www.jacet.org/>.
- [9] 安田圭志, 喜多村圭祐, 山本誠一, 柳田益造, “多重タグ付き英語学習者コーパスの開発と英語能力自動測定への応用,” 自然言語処理, vol.16, no.4, pp.447–463, 2009.
- [10] J. TURIAN, “Evaluation of machine translation and its evaluation,” Proceedings of MT Summit IX New Orleans, LA., 2003, pp.386–393, 2003.
- [11] 小谷克則, 吉見毅彦, 井佐原均他, “読解効率に基づく英文読解能力測定テストの開発とその信頼性・妥当性の検証,” 情報処理学会論文誌, vol.52, no.4, pp.1843–1851, 2011.
- [12] M. Coleman and T.L. Liao, “A computer readability formula designed for machine scoring,” Journal of Applied Psychology, vol.60, no.2, p.283, 1975.
- [13] M.D. Jackson and J.L. McClelland, “Processing determinants of reading speed,” Journal of Experimental Psychology: General, vol.108, no.2, p.151, 1979.
- [14] T. Shizuka, “Attaining 40-item-equivalent reliability and validity with only 4 items: Effects of incorporating confidence and speed measures in a computer-based reading test,” 2000.
- [15] J.H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence, MIT Press, Cambridge, MA, USA, 1992.