

視点特徴とテキスト特徴の分析による日本語テキストの難易度推定

楽 卓登[†] Olivier Augereau[†] C.Lima Sanchez[†] 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

E-mail: rou@m.cs.osakafu-u.ac.jp, charles-ls@hotmail.fr, augereau.o@gmail.com, kise@cs.osakafu-u.ac.jp

あらまし 本稿では、視点情報と文書情報を組み合わせた日本語文書の難易度推定手法を提案する。日本語文書の難易度を推定できれば、日本語を学習する外国人は自身の語学力に適した難易度の文書を用いることで、より効率的に学習することが可能になり、語学の学習サポートに有効である。従来手法では、様々な方法でテキストを解析することで文書の難易度の推定している。しかし、これらの特徴にはユーザが文書を読むときにどのように感じたかの情報が不足しているため、高い推定精度を得られず、使用できる文書も限定されている。そこで本研究では人が難しい文書を読むときに、読む速度が遅くなる、読み止まる、読み返すなどの視点情報に着目し、機械学習 (SVR) を用いることでより高精度な文書難易度推定手法を提案する。提案手法の有効性を示すため、18名の実験参加者に視点情報を記録し、日本語文書の難易度推定精度を検証した。検証の結果、難易度推定精度の平均絶対誤差は0.088であった。この結果は比較手法である文書情報のみを用いた手法と比べて、平均絶対誤差が13%改善していることから、提案手法の有効性を示すことができた。

キーワード アイトラッカ、視点情報、文書難易度推定、機械学習

1. はじめに

近年、社会のグローバル化により、留学や仕事で来日する外国人が増えている。このような人達は、今後日本での活動を円滑に進めるために、日本語を素早く習得する必要がある。そのため、日本語を第二言語として習得する必要がある人が増加傾向にある。日本語を習得するには、読む、聞く、話す、書くなどの練習が必要とされる。この中でも読む練習は、語彙力、文章力、知識力が養うことができ、聞く、話す、書くの成長につながるため、効率よく語学を学習するための重要な行為と考えられる。本研究はこの読む行為を対象として、日本語を学習する外国人をサポートできる方法について考える。

読む行為から語学を効率的に学習できるサポートシステムとして、Newsela^(注1)がある。Newselaは政治、科学、生物などの様々なジャンルから、注目度の高い時事ネタの文書を集め、専門家がその文書の文法や単語など入れ替えることで、複数の文書難易度を作成し、読者に提供している。読者は興味のある文書を自身語彙力に適した難易度で学習することができるため、知識を広める以外に、読解力の向上や語学学習へのモチベーションを高めることができる。このようなメリットがあるため、現在約80万の先生および770万の生徒に利用され、語学学習の新たなサポート方法として期待されている。しかし、現在は英語とスペイン語のみに対応しており、日本語化はされていない。そこで本研究は、Newselaの魅力である「文書を難易度別に提

供する」サービスを、日本語でも提供できる方法について考える。日本語文書を難易度別に提供する方法の一つの方法は、Newselaのように専門家が文書の文法や単語を変換し、作成することである。しかし、この方法は時間と労力がかかり、提供できる文書数が限られる。そこで、本研究は文書を一から作成するのではなく、既存の様々な文書を自動的に難易度別に分類することで、より多くの文書を提供する方法を考える。現状では多くの既存文書には難易度が明記されていないため、文書の難易度を推定する手法が必要と考えられる。

日本語の文書難易度を推定する手法は、これまで多く提案されている [1] [2] [3] [4]。これらの手法では、文書の平均文字数、語彙の頻出度や構文などを使用して、難易度の評価式を作成し、それに基づいて難易度を推定している。しかしこれらの手法には幾つかの問題点がある。第一の問題は、多くの手法の文書難易度の指標が、難しい語句や漢字の割合を用いており、それより高次の情報（言葉の言い回しや意味合いなど）を考慮していないことである。第二の問題として、語彙の頻出度を使う手法では使用するコーパスに大きく依存するため、適用できない文書が多い点が挙げられる。このように、文書情報だけでは難易度を判別することに限界があるため、文書情報以外の情報を利用する必要があると考えられる。

本研究では、文書を読む時の視点情報が文書難易度の推定に有効であると考えている。理由として、「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ことが挙げられる。例えば、人は難しい文書を読むときは、知らない単語や文法があるため自然と読む速度が遅

(注1) : <https://newsela.com/>

くなる。一方、簡単な文書を読むときは既知の単語やわかりやすい表現が多くあるため、スラスラ読むことができる。このような明らかな違いがあることから、視点情報は難易度を推定する重要な手がかりとなり得る。これまでの様々な視点情報に関する先行研究から、読者の視点情報を分析することによって、読者の文書に対する理解度、読者の英語スキル、読んだ単語数、読んでいる文書の種類などが推定されている。本研究は、これらの研究と同様に視点情報に注目することで、文書難易度を推定するものと位置づけられる。

以上のことを踏まえて、我々は従来の文書情報に加えて、文書を読む時の視点情報を利用することで文書の難易度を推定する手法を提案する。本手法の特徴は、視点情報から、読者が主観的に感じる難易度の情報を得て、難易度の推定に用いる点にある。実験の結果、文書情報のみを使用するときと比べて、提案手法はより高い精度で難易度を推定できることが分かった。以降 2 節で関連研究について述べ、3 節で提案手法、4 節では提案手法の推定精度を検証する為の実験について説明する。そして、5 節で本稿のまとめについて述べる。

2. 関連研究

本節では、文書の難易度推定に関する先行研究と視点情報に関する先行研究について述べる。

2.1 文書の難易度推定に関する先行研究

本節では文書情報を用いた文書の難易度推定に関する研究について述べる。文書の難易度の評価基準は大まかに 2 種類に分かれる。一つは使用する単語の難しさを基準とするものと、もう一つは文書自体の構造の難しさを基準とするものである。

単語の難しさを基準とする研究例として以下の二つがある。近藤らは小学 1 年から高校 3 年までの 12 学年に、大学を加えた 13 区分で使用されている文書からコーパスを作成し、そのコーパスを元に文書の難易度を推定している [5]。しかしこの手法は使用した文書への依存性が高く、また教科書以外の文書には対応できないという問題がある。また、同一学年内で使用される文書の細かい難易度の差異を判別することもできない。川村らは、文書中に存在する単語に対して日本語能力試験の等級 (1~4) を出力するシステムを提案している [6]。このシステムは、文書内で各級がどの程度使用されているかの割合を出力することができるが、文書全体の難易度を推定する方法ではない。

文書の構造の難しさを基準とする研究例として建石らと柴崎らの研究がある [1] [3]。建石らの手法は日本語の文書を対象とし、文の平均の文字数、ひらがな、漢字、カタカナの相対頻度及び平均の長さ、読点と句点の数に対する割合などの特徴量を使用して、主成分分析により文書難易度の評価式を導出している。しかし、使用した特徴量が表層的なものに留まっており、高次の情報を考慮していないため、難易度を推定できないと考えられる。柴崎らは小学校の国語教科書 (6 学年) のコーパスを作成し、文の平均文字数と平均単語数、平均係り受け数など特徴量を変数とし、重回帰分析を行い、文書難易度を算出する手法を提案している。しかし、適用している文書は小学校の国語科の文書なので、それ以外の文書への適用については考慮されて

いない。

このように、現状では、文書情報として文書の表層的な情報しか利用しておらず、複雑な文脈や言葉の言い回しなどによる難しさを判別できないという欠点がある。

2.2 視点情報に関する先行研究

視点情報を用いた研究について述べる。文書と視点情報に関連する研究として、ユーザの読む行動を認識する研究がある。Kunze らはユーザの視点情報を記録して、読んだ単語数を推定する手法を提案している [7]。この手法は、一行当たりの平均単語数といった文書の情報とユーザの眼球運動の情報を組み合わせることにより、単語数を推定するものである。

永田らは文書を読む時のスピードを用いて文書の難易度を推定する手法を提案している [2]。文書を読むスピードとは、単位時間あたり読む文字数を示す。実験の結果、文書を読むスピードは個人性が強く、文書の難易度推定評価基準として利用することは適切ではないことが示されている。

文書を読む際、ユーザの読んだ単語数以外に、ユーザが難しく感じた単語や行などを推定する方法を大社らが提案している [8]。その結果、注視時間が難しいと感じた部分の識別に有効で、読む速度と注視回数は理解度に関連していることを示している。

視点情報と理解度に関する研究として、吉村らは視点情報を用いて、英語習熟度を推定する手法を提案している [9]。吉村らは、TOEIC スコアに基づく 3 段階の英語習熟度 (上級・中級・初級) を高い識別率で推定することに成功している。

Manuel らは、ユーザが日本語能力試験の問題を解答する際の視点情報のみを使用して、ユーザが 4 つの難易度中、どの難易度の問題を解いているかを推定している。実験した結果、難易度の推定精度は 35.4% となった。

このように、視点情報には人の感じている情報が含まれており、文書の複雑な文脈や言葉の言い回しなどによる難しさを抽出することが可能である。しかし、これらの情報以外に、文書のレイアウト、読む人のコンディションや文書の内容に対する興味度合いなどの情報も含まれているため、視点情報のみでは文書の難易度を判別することは容易ではないと考えられる。

3. 提案手法

先行研究で示すように、文書情報のみで文書難易度を推定する場合、文面からある程度の難易度を推定できるが、文脈や言葉の言い回しなどによる難しさを抽出することができない。一方、視点情報のみ使用した場合、文書情報で抽出できない部分を取得できるが、文書のレイアウトや内容、読む人のコンディションなどに大きく影響されるため、難易度の推定が困難だと考えられる。そこで、本研究では、視点情報と文書情報の両方を使用し、上記の問題点をそれぞれの長所によりに補うことで、より高精度な日本語文書の難易度推定手法を提案する。

3.1 処理の流れ

日本語の文書を読む際の視点情報と文書情報から、文書の難易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。まず、日本語文書を読む時の視点情報を記録する。次

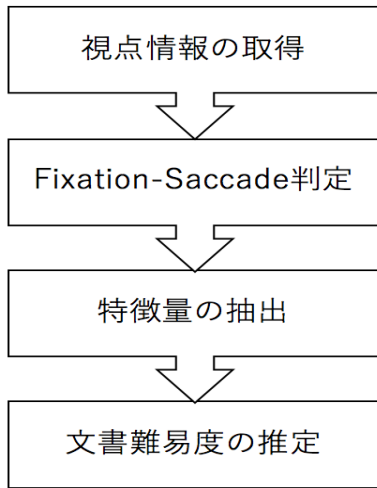


図 1: 提案手法の流れ

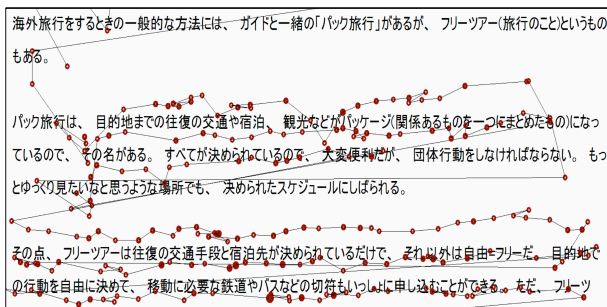


図 2: Fixation-saccade 判定後の視点情報

に、得られた視点情報に対して、Fixation-Saccade 判定を適用する。そして、視点や文書からそれぞれの特徴量を抽出する。最後に、SVR を用いて、日本語の文書難易度と特徴の関連性を学習し、未知な文書に対して難易度を推定する。

3.2 文書の難易度の定義

難易度の定義について説明する。文書難易度の決め方として様々な方法が存在する。その一つとして、文書の内容に関する問題を複数の人に解いてもらい、その正解率を使用する方法が考えられる。一般的に人は難しい文書を読むと、内容について理解ができず、問題を正しく解答できないと考えられる。この方法は、読者に直接難しさを聞く方法と比べて、文書を理解できたかをより客観的に判別し、難易度を推定することができる。また、難易度を学年や等級の区分で分ける方法と比べて、より細かい難易度を推定できることが期待できる。しかし、この方法は問題自身の難易度にも依存するため、使用する際、問題の設定に注意する必要がある。そこで、本研究では文書に関する適切な問題を使用し、複数の読者に解答してもらい、その正解率を文書の難易度として定義付ける。

3.3 視点情報の取得

視点情報の取得にはアイトラッカを用いる。アイトラッカとは、カメラやセンサを用いて眼球運動を測定する装置である。本研究では、据置型のアイトラッカを用いて、ユーザがディスプレイ上に表示された文書を読む際の視点情報を取得する。

表 1: 特徴量一覧

特徴量	詳細
視点に関する特徴量	Fixation の頻度
	Fixation 持続時間の平均, 最大値, 標準偏差 Saccade 各方向の頻度 前方, 後方
	Saccade 持続時間の平均, 最大値, 標準偏差
	Saccade 時の速度の平均, 最大値, 最小値, 標準偏差
	Saccade 移動距離の平均, 最大値, 標準偏差
文書に関する特徴量	漢字, ひらがな, カタカナ, ローマ字, 漢語, 和語 句読点, 動詞, 接続詞の使用割合
	語彙, 文の平均の長さ
	語彙 N1 級, N2 級, N3 級, N4 級, N5 級, 範囲外, その他の使用割合
言語能力に関する特徴量	日本語能力試験のスコア

3.4 Fixation-Saccade 判定

眼球運動は、Fixation と Saccade の二つに分類できる。Fixation は視点がある範囲内で一定時間留まる動きを示し、Saccade は Fixation 間の素早い目の動きを示す。図 2 で示すように、人は Fixation と Saccade を繰り返すことで、文書を読んでいる。本手法では、読書時の視点から Fixation と Saccade を抽出し、難易度推定の特徴量に用いる。Fixation の判定方法には Buscher らの手法を使用する [10]。

3.5 特徴量の抽出

本節では、文書難易度を推定するための特徴量について説明する。文書を読む際の人の視点情報は、その人の語学力や読み方により大きく影響を受けると考えられる。そのため、本手法では同一文書を読む複数のユーザの特徴量を平均することによって、ユーザによるばらつきを減少させることを考える。そして、どのレベルのユーザがどのような文書をどうやって読んだかを知るために、ユーザの言語能力に関する特徴量 1 個、文書情報に関する特徴量 18 個、視点情報に関する特徴量 16 個、合計 35 個の特徴量を文書ごとに抽出する。特徴量の一覧を表 1 に示す。

3.5.1 言語能力に関する特徴量

言語能力と読む能力には深い関連がある。言語能力の高い人は語彙処理能力が優れていて、読書速度も速いことが示されている [11]。これは、同一の文書を読む時でも、言語能力の違いによって視点情報が大きく異なることを意味する。そのため、ユーザの言語能力は文書を読むときの視点情報に深い関連があり、文書の難易度推定に考慮すべき特徴量だと考えられる。そこで、本手法はユーザの日本語能力を知るために、日本語能力試験の問題を解いてもらい、そのスコアをユーザの日本語能力を表す特徴として使用する。

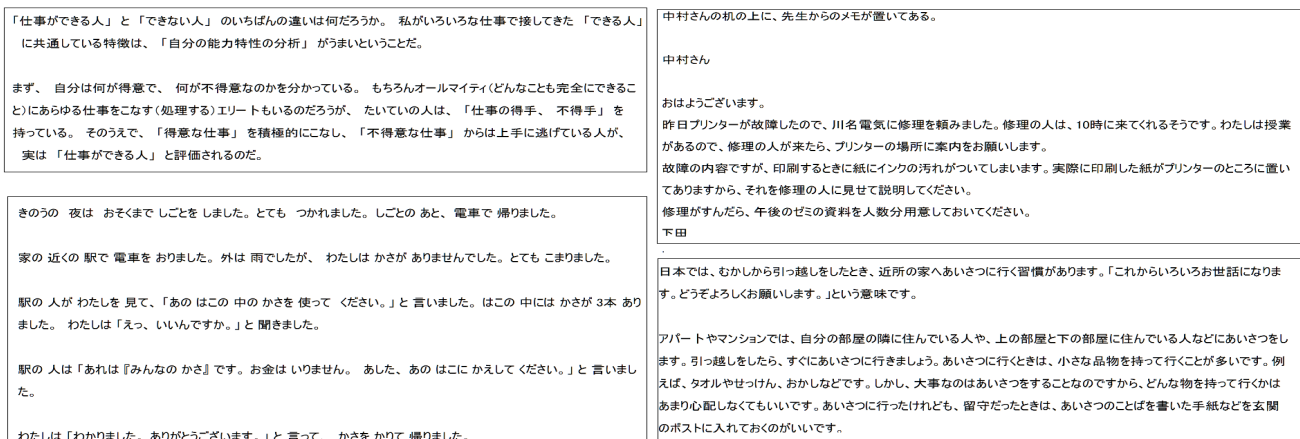


図 3: 実験で使用した文書の一例 (左上:N2 級 右上:N3 級 左下:N5 級 右下:N4 級)

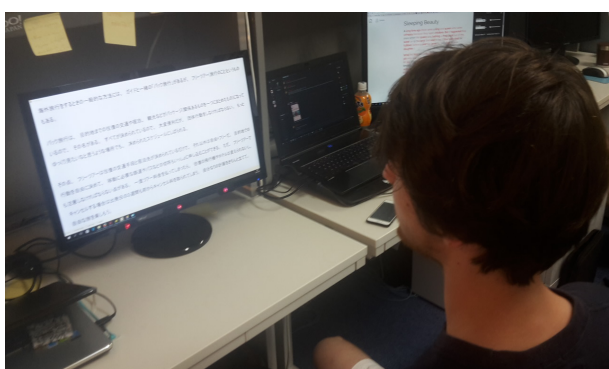


図 4: 実験の風景

3.5.2 文書情報に関する特徴量

日本語文書の難易度を決定する要素には語彙、文法、構文などがある。本研究では、文書の難易度に深く関わる語彙と構文の複雑さを考慮し、文書の難易度に関する特徴量を抽出する。文書は日常で使用頻度が低い語彙の多く含むほど、難易度が高くなる。そこで、文書の難易度を推定するために、本手法は読解学習支援システムであるリーディング・チュウ太・語彙チェッカー^(注2)(以下、語彙チェッカーと省略する)を用いて、文書中に使用される語彙の日常での出現頻度を求める。語彙チェッカーは「現代日本語書き言葉均衡コーパス (BCCWJ)」および「日本語教科書コーパス」に準拠して、文書に出現する頻度から語彙を5つのレベルに分類したものを使用している。本手法では、文書で使用された語彙の5つのレベルの割合を用いることで、文書の難易度を表す。また、構文の複雑さを求めるために、Mecab^(注3)を用いて構文を解析する。Mecabとは、形態素解析エンジンであり、入力された文書の形態素を解析することができる。提案手法では、Mecabを利用することで、文書中に漢語、和語、動詞、接続詞などの割合を求めて、文書の複雑さ

(注2) : <http://language.tiu.ac.jp/>

(注3) : <http://taku910.github.io/mecab/>

を表す特徴として使用する。

3.5.3 視点情報に関する特徴量

Buscher らの手法を用いて分類した Fixation と Saccade から、表 1 で示す各特長量を抽出する。本手法では、ユーザがどれぐらい注視しているか求めるために、Fixation の特徴量として、Fixation の頻度、Fixation 持続時間の平均、最大値、標準偏差などを用いる。そして、読み飛ばし、読み返しなどを求めるために、Saccade の特徴量として、Saccade の各方向への頻度、速度、持続時間及び移動距離などを用いる。

3.6 日本語文書の難易度推定

日本語文書の難易度推定には、サポートベクトル回帰 (SVR) を使用する。抽出した視点特徴量と文書特徴量をサポートベクトル回帰で回帰分析を行い、特徴量と難易度の関連性を学習することで、文書難易度の推定モデルを作成する。また、推定に有効な特徴が自明ではないため、Backward stepwise selection を用いて文書難易度推定に有効な特徴を選定する。Backward stepwise selection とは、使用する特徴量の一つずつ減らしていく山登り型の選定手法で、減らした前後の推定精度を比較することで、減らした特徴量が有効であるかを検証する方法である。推定に不要な特徴量を取り除くことで、推定精度の高い特徴量の組み合わせを探索する。

4. 実験

本節では、提案手法の有効性を検証する実験について述べる。本実験では提案手法を用いて、難易度が既知な文書に対してどれくらい正しく推定できるかを求める。そして、文書情報のみ、視点情報のみ使用した場合とその両方の情報を用いた場合の推定精度を比較する。実験では図 3 のような N2 級~N5 級の日本語能力試験で用いられた文書を利用する。日本語能力試験は、日本語を外国語とする人の日本語能力を認定する試験であり、世界 58 カ国・地域で幅広く実施されているため、日本語を外国語として学習している人の文書として適切であると考えられる。実験で使用する N2 級は幅広い場面での日本語をある程度理解できるレベルで、N5 級は基本的な日本語をある程度理解できるレベルとなる。以下、実験の詳細を述べる。

4.1 実験条件

実験参加者 18 名に各難易度の文書を 5 文書読んでもらい、計 360(4 難易度 × 5 文書 × 18 名) 個の視点情報をアイトラッカを用いて記録した。各文書を読み終えた後に、その文書に関する問題を一つ解答した。そして、各文書問題に対して参加者全員の解答から正解率を求め、文書の難易度として使用した。使用した問題は日本語能力試験から抜粋したものであり、問題作成者が文書の難易度に合わせて慎重に設定したものと考えられる。次に、実験参加者が確信をもてず、偶然選んだ答えが正解する場合があるため、文書に対する理解度についてアンケートを行い、全く理解できなかった文書を実験から除外した。提案手法の評価方法として、Leave one document out cross validation を使用した。この手法は 1 文書をテストデータとして、残りの全ての文書を学習データとする手法である。文書難易度の推定精度を評価する方法としては、式 1 で示す平均絶対誤差 MAE (Mean Absolute Error) を使用する。式 1 の E は平均絶対誤差、 N は文書数、 X_i と Y_i はそれぞれ推定値と真値を示す。実験において、参加者は全員日本語を外国語とする人で、国籍はフランス、ドイツ、ベトナムなど様々である。アイトラッカは Tobii 社の Tobii EyeX^(注4) を用いた。実験の風景を図 4 に示す。

$$E = \frac{\sum_{i=1}^N |X_i - Y_i|}{N} \quad (1)$$

4.2 実験結果・考察

まず実験で使用した各文書の難易度を難易度順にソートしたものを図 5 に示す。文書の平均難易度は 0.750、標準偏差は 0.210 である。各特徴量を用いた時の推定平均絶対誤差を表 2 に示す。表 2 で示すように、視点情報のみを使用して文書難易度を推定する場合、推定平均絶対誤差が 0.151 となり、基準値(難易度の平均値との平均絶対誤差)である 0.160 より誤差が小さく、従来の文書情報のみを使用する場合よりも推定誤差が大きかったことが見られた。しかし、文書情報に視点情報を付加することで推定平均絶対誤差が 0.088 となり、従来手法と比べて推定誤差が 13% 改善し、より高い推定精度が得られた。

表 2: 各特徴量を用いた時の文書難易度の推定誤差

特徴量	平均絶対誤差
文書情報	0.101
視点情報	0.151
文書情報+視点情報	0.088
基準値	0.160

Backward stepwise selection で選択された特徴量について考察する。推定で使用された特徴量を表 3 にまとめる。視点情報では、後方への Saccade の頻度、Fixation 持続時間の標準偏差、Saccade 時の最大速度が選択された。これは、文書を読む際に、ユーザが難しいと感じた文書では、読み返しや部分的に集中して読むことが多く、簡単な文書では読む速度が速く、視点の移動幅が多いからだと考えられる。次に、文書特徴量と

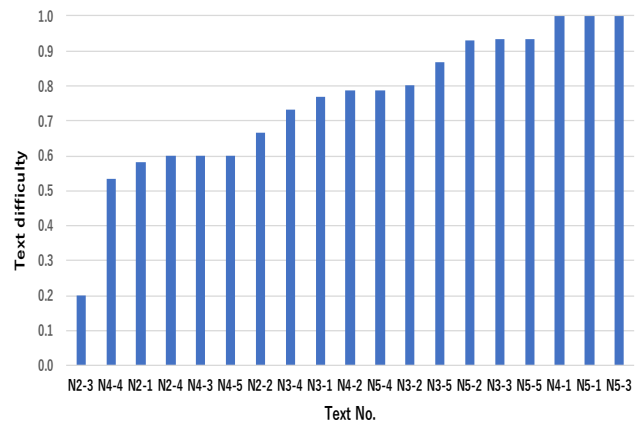


図 5: 実験で使用した各文書の難易度

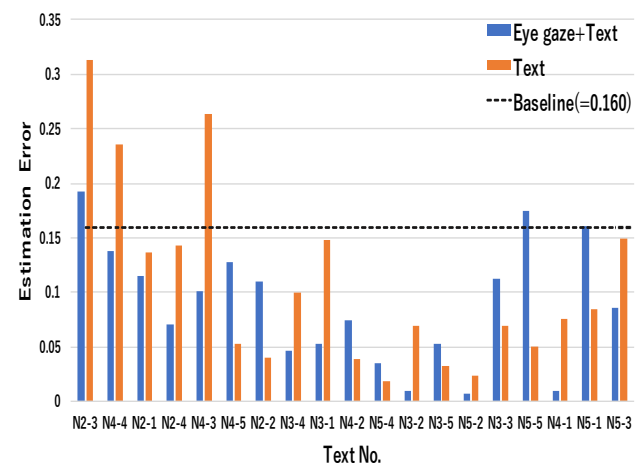


図 6: 実験で使用した各文書各文書の推定誤差

表 3: 選択された特徴量

特徴量	詳細
文書特徴量	語彙 4 級, 語彙 5 級, ひらがな, カタカナ, 句読点, 接続詞, 動詞の割合
視点特徴量	Saccade 後方への頻度 Fixation 持続時間の標準偏差 Saccade の最大速度

して、文書で使用された語彙 4 級、語彙 5 級、ひらがな、カタカナ、句読点、接続詞、動詞の割合が選択された。これは、実験で使用した文書では、N2 級では漢字やカタカナが多く含まれており、ひらがなの使用率が N5 級など語学初心者が読む文書と比べて少ないため、難易度の推定に有効的だと考えられる。

次に、どのような文書が推定困難かを調べるため、各文書の推定結果を解析した。文書情報のみの場合と文書情報、視点情報の両方を使った時の、各文書の推定誤差を図 6 で示す。図 5 と比較すると、全体的に難易度が高い文書に対して、提案手法の視点情報と文書情報を組み合わせた手法が、従来の文書情報のみを使用した手法より推定誤差が小さいことが見られた。一方、難易度が低い文書に対しては、従来手法の方が推定誤差が

(注4) : <http://www.tobiipro.com/ja/>

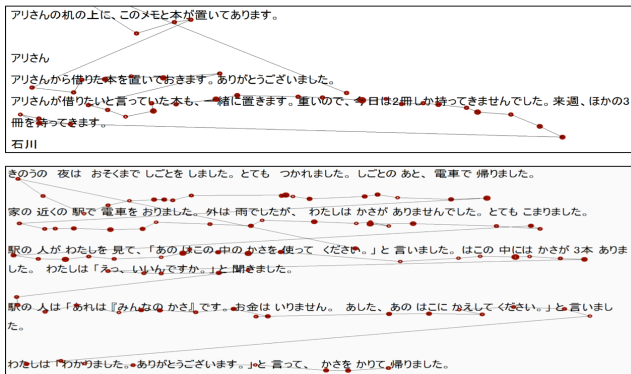


図 7: N4 級, N5 級の文書を読んでいる時の視点情報

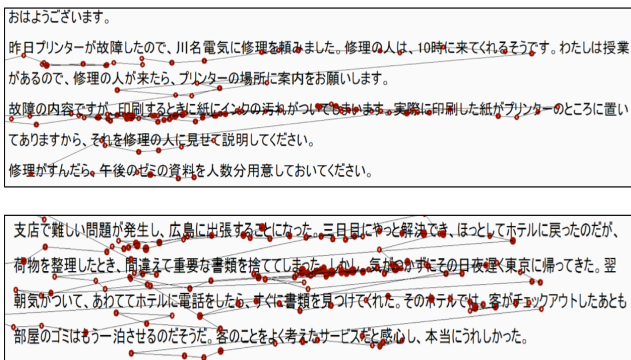


図 8: 文書 N3-3, N3-4 を読んでいる時の視点情報

小さいことが見られた。次に、難易度が一番高い文書 N2-3 の推定誤差が最も大きいことが見られた。この原因として、図 5 からわかる様に、SVR で学習する際、N2-3 の様な難易度が高いサンプルが不足しているため、正しく学習できなかったからだと考えられる。

また、図 6 より、N4 級と N5 級の文書を推定する際、多くの文書は文書情報のみを特徴量とした方がより高い精度が得られていることを確認できる。原因として、N4 級と N5 級の文書は、図 7 のように、その他の文書と比べて文書の内容が短く、また初心者に文書を読みやすくするように、その他の文書と比べて改行が多く含まれているため、視点情報を利用すると、文書が難しいといった誤った判断を得られたと考えられる。一方、図 8 で示すような漢字が多く含まれているが、文書としてそこまで難しくもないものに対しては、文書情報だけでは正しく難易度を推定することは困難となる。そのような場合、視点情報を加えて利用することでより難易度を正しく推定できる。

それ以外にも、多くの N2 級の文書の推定難易度が、実際の難易度より大きく上回る事が見られた。原因として挙げられるのは、今回ユーザが文書を理解できているかを調べるために使用した問題が、文書が難しいにもかかわらず問題自身の難易度が低いため、参加者が文書をあまり理解できなくても、正しく解答できたためだと考えられる。

5. まとめ

本稿では、日本語を学習する外国人を支援するために、視点

情報と文書情報を組み合わせた、文書の難易度推定手法を提案し、それに基づく実験について述べた。提案手法では、文書を読む時の視点情報と文書情報から特徴量を抽出し、サポートベクトル回帰を用いて学習モデルを作成することで、文書の難易度を推定するものである。提案手法の有効性を検証するため、18 名の実験参加者に 20 文書を読んでもらい、その視点情報を記録し、実験を行った。その結果、日本語文書の難易度推定精度は平均絶対誤差が 0.088 となり、従来の文書情報のみで推定する手法と比べて、視点情報を加えることでより高い推定精度が得られ、提案手法の有効性を示すことができた。

今後の課題としては、精度向上のための特徴量および使用する学習モデルの再考がある。また、本実験で用いた文書は日本語能力試験から抜粋したものであったが、今後様々な分野や対象についても検証する必要がある。

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] 建石由佳, 小野芳彦, 山田尚勇他, “日本文の読みやすさの評価式,” 情報処理学会研究報告ヒューマンコンピュータインタラクション (HCI), vol.1988, no.25 (1988-HI-018), pp.1-8, 1988.
- [2] 永田亮, 井口達也, 榊井文人, 河合敦夫, “リーディングスピードに基づいた文章の読み易さについて,” 電子情報通信学会技術研究報告. TL, 思考と言語, vol.102, no.491, pp.13-18, 2002.
- [3] 柴崎秀子, 沢井康孝, “国語教科書コーパスを応用した日本語リーダビリティ構築のための基礎研究 (言語理解のためのコーパスからの知識獲得),” 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, vol.107, no.246, pp.19-24, 2007.
- [4] S. Sato, S. Matsuyoshi, and Y. Kondoh, “Automatic assessment of japanese text readability based on a textbook corpus,” LREC, 2008.
- [5] 近藤陽介, 松吉俊, 佐藤理史, “教科書コーパスを用いた日本語テキストの難易度推定,” 言語処理学会第 14 回年次大会発表論文集, vol.14, pp.1113-1116, 2008.
- [6] 川村よし子, “語彙チェッカーを用いた読解テキストの分析,” 早稲田大学日本語研究教育センター講座日本語教育, 第 34 分冊, pp.1-22, 1998.
- [7] K. Kunze, H. Kawaichi, K. Yoshimura, and K. Kise, “The wordometer—estimating the number of words read using document image retrieval and mobile eye tracking,” Document Analysis and Recognition (ICDAR), 2013 12th International Conference on IEEE, pp.25-29 2013.
- [8] 大社綾乃, “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 (信号処理),” 電子情報通信学会技術研究報告= IEICE technical report: 信学技報, vol.115, no.22, pp.161-166, 2015.
- [9] 吉村和代, 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 (テーマセッション, 実世界指向, 産業),” 電子情報通信学会技術研究報告. CNR, クラウドネットワークロボット, vol.114, no.455, pp.63-68, 2015.
- [10] G. Buscher, A. Dengel, and L. vanElst, “Eye movements as implicit relevance feedback,” CHI’08 extended abstracts on Human factors in computing systems ACM, pp.2991-2996 2008.
- [11] 山科美和子, 釣井千恵, “第 2 言語の語彙処理能力とリーディング力: 単語認知と読書量・読解速度との関連を探る (岡田章子教授退任記念号),” 英米評論, pp.237-260, 2010.