

視点情報を用いた文書上の主観的高難易度単語の推定

大社 綾乃[†] 石丸 翔也[†] Olivier Augereau[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科 〒559-8531 大阪府堺市中区学園町 1-1

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

あらまし 本稿では、アイトラッカから取得した視点情報より、Deep Learning を用いた、文書上の主観的高難易度単語の推定手法を提案する。「主観的高難易度単語」とは、文書を読んだ際にユーザが難しいと感じた単語を表す。文書を読むだけで、主観的高難易度単語を推定することが出来れば、ユーザはどの単語を学習しなければならないかを簡単に知ることが出来る。文書を読んでいる際に、難しいと感じた単語に印をつける手間が省け、読むリズムが乱れないという利点もある。そのため、学習補助に大変有効であると言える。一般的に、難しい単語を目にすると、読む速度が遅くなったり、何度も読み返すといったような視点パターンがみられる傾向がある。しかし、難しい単語の意味を文脈から判断できる場合、読み飛ばし、視点パターンに特徴が見られないことがある。そこで、提案手法では視点情報に加え、単語の出現頻度情報を用いる。まず、各単語を中心に前後2～3単語を読んでいる際の視点パターンを抽出し、視点情報による入力データを得る。そして、入力データに単語の一般的な出現頻度情報を付加し、Deep Learning を用いて学習することで、より高精度な主観的高難易度単語の推定を目指す。提案手法の有効性を示すため、25名の視線データを記録し学習し、主観的高難易度単語の推定精度を検証した。検証の結果、推定精度は、再現率90.8%、適合率34.4%、F値49.2%であった。

キーワード 学習補助, 英語習熟度, 主観的高難易度単語推定, 視点情報, アイトラッカ, Deep Learning

1. まえがき

近年グローバル化が進み、英語が必要とされる場面が増えている。小学校における英語の必修化だけでなく、英語を公用語とする企業も増え、語学学習の必要性が高まっている。語学学習において、自分の学習成果を正しく知ることは大変重要な要素の1つである。自分の学習成果を知ることで、苦手な部分を把握したり、今後の学習計画を立てることができる。

学習成果を測る方法としては、定期的に試験を受けたり、日々の学習で難しいと感じた単語や文章に印をつけておくことなどが挙げられる。しかし、試験は、その時の体調や問題との相性に影響を受ける場合がある。加えて、出題箇所と関係のない部分については、理解できているかどうかを知ることが出来ないという欠点がある。また、日々の学習で難しいと感じた単語や文章の全てに1つ1つ印をつけておくことは、大変手間がかかり、容易ではない。より簡単に学習成果を知るため、文書上の難しいと感じた部分を自動で記録することができれば大変有用である。我々は、文書を読むだけで、文書上のユーザにとって難しいと感じた部分を自動記録するシステムの実現を目指す。

文書を読むという行為には目の動きが関係している。例えば、人は文書を読む際、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、同じ場所を注視し続けるといった傾向がある。そこで、文書を読んでいる際の視点情報を解析することで、文書上の難しいと感じた部分を推定することが考えられる。難しいと感じた部分には、段落、文章、単語などがあるが、本

稿では単語に着目する。ユーザにとって難しいと感じた単語を、ここでは「主観的高難易度単語」と呼ぶ。文書を読むだけで、主観的高難易度単語を知ることが出来れば、ユーザはどの単語を学習しなければならないかを簡単に知ることが出来る。文書を読んでいる際に、難しいと感じた単語に印をつける手間が省け、読むリズムが乱れないという利点もある。また、教師が生徒の難しいと感じている単語を簡単に知ることが出来れば、より効率的な指導が可能になると考えられる。

しかし、視点情報のみから、主観的高難易度単語を推定することは容易ではない。以前我々は、視点情報を用いて主観的高難易度単語の推定精度を検証した[1]。各単語の合計注視時間を特徴量とし、各単語について、一定時間よりも注視時間が長い時に、その単語を難しいと判定した。その結果、再現率36.2%、適合率15.6%、F値21.8%と十分な精度を得ることが出来なかった。この原因は大きく2つあると考えている。第一の原因は、注視時間が短くても主観的高難易度単語となる場合があることである。例えば、文書を読んでいる際に難しい単語を目にしても、文脈から判断可能であると読み飛ばしてしまい、視点に特徴的な動きが見られないことがある。第二の原因は、視点情報からどのような特徴を抽出すればよいかの自明ではない点である。

そこで本研究では、以下の2つの特徴を伴う新しい推定手法を提案する。第一に、視点情報だけでなく、単語の一般的な出現頻度を特徴量として付加する。第二に、Deep Learning を導入し、学習により特徴量を自動で選定する。具体的には以下の

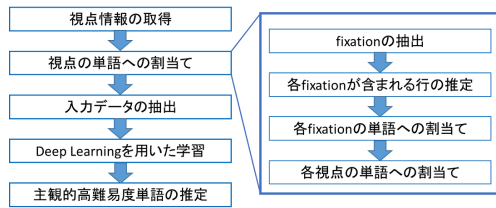


図 1 処理の流れ

通りである。提案手法では、ユーザが英語の文書を読む際の視点情報を記録し、主観的高難易度単語を読む際の視点パターンに単語の一般的な出現頻度情報を付加した入力データを Deep Learning を用いて学習し、主観的高難易度単語を推定する。提案手法では、単語の出現頻度を用いることで、視点情報では不十分であった特徴量を補う。単語は、一般的に出現頻度が低いほど難易度が高いと考えられるためである。

本稿では、提案手法の有効性を検証するため、単語の出現頻度情報のみで推定する手法、視点情報のみで Deep Learning を用いて推定する手法と比較実験をおこなう。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

視点情報を用いた読書補助システムとして、Cheng らのアノテーションシステムがある [2]。Cheng らは、教師が論文を読む際の視点情報から得た、段落ごとの読む速さや読み返しの頻度などの情報を、文書画像上にアノテーションを付与している。生徒がこのアノテーションを参考に論文を読むことで、より理解が深まっているという結果が得られている。また、Biedert らは、視点情報を用いて、文書の書き方が読者にとって分かりやすいものかどうかを評価する手法を提案している [3]。Biedert らの手法では、読者の視点情報から、文書上のどの部分が意味が分かりにくいかを推定し、書き手にフィードバックすることを目的としている。文書中の部分的な難しさについて着目した研究として、Kunze らの研究がある [4]。Kunze らは、視点情報を用いてユーザが難しいと感じた単語の推定を提案しているが、定量的な評価には至っていない。また、Godfroid らの研究から、文書を読む際、知らない単語に対する注視時間より、知っている単語に対する注視時間の方が長くなる傾向にあることが分かっている [5]。

以前我々は、ユーザの段落ごとの視点情報を文書画像上に表示することで、ユーザの英語スキルの差を簡単に知ることが出来るシステムを提案した [6]。また、我々は、文書中の難しいと感じた部分を推定するために有効な特徴量を段落ごと、時間による segment ごと、単語ごとの 3 つについて調査した [7]。その結果をふまえ、視点情報を用いて文書上の難しいと感じた段落および単語を推定し、文書画像上にアノテーションとして表示するシステムを提案した [1]。しかし、難しいと感じた単語の推定精度は十分ではなかった。

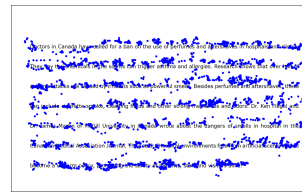


図 2 画像上の視点情報の例

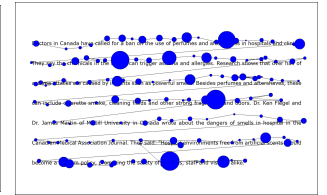


図 3 画像上の fixation の例

3. 提案手法

3.1 処理の流れ

図 1 に処理の流れを示す。まず、アイトラッカを用いてユーザの眼球運動を記録し、視点情報を取得する。各視点がある単語を読んでいるのかという情報を得るため、各視点を単語に割り当てる。このプロセスはさらに 4 つに分類できる。詳細は 3.3 で述べる。次に、Deep Learning に入力するため、各単語を中心に前後 2～3 単語を読む視点パターンを取得し、入力データを抽出する。また、予め文書情報を抽出しておき、単語の一般的な出現頻度を入力データに付加する。その後、Deep Learning を用いて学習し、難しい単語を読んでいる際の視点パターンを推定する。その結果を用いて、主観的高難易度単語を推定する。

3.2 視点情報の取得

アイトラッカを用いて視点情報を取得する。アイトラッカとは、眼球運動を測定する装置である。本システムでは、据置型のアイトラッカを用いて、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。得られた視点を文書画像上に表示したものを図 2 に示す。青い点が視点を表している。

3.3 視点の単語への割当て

図 1 の右側のフローチャートに、視点の単語への割当ての手順を示す。まず、視点情報を、注視を意味する fixation と、注視間の素早い目の動きを意味する saccade に分ける。得られた fixation 情報を用いて改行を判定し、各 fixation がどの行に属するかを調べる。次に、各 fixation がどの単語に割り当てられるかを調べ、最後に、各 fixation に含まれる視点を単語に割り当てる。このように、各視点を直接単語に割り当てるのではなく、fixation を用いて単語に割り当てている理由は以下の通りである。視点の推定誤差などの影響により、各視点は最近傍の単語を見ているとは限らない。ところが、fixation として複数の視点を統合すると、このような誤差の影響が軽減される。一方、入力データとしては、より詳細な情報を抽出可能な個々の視点をを用いる。すなわち、fixation を単語に割り当てたあと、fixation に含まれる視点を単語に割り当てて、入力データを抽出する。以下、詳細について述べる。

3.3.1 fixation の抽出

アイトラッカから取得した視点情報から、fixation を抽出する。fixation は視点がある範囲内に一定時間停留すること、saccade は fixation 間の素早い目の動きのことである。人は、fixation と saccade を繰り返すことで、ものを見たり、文書を読んでいる。提案手法では、Buscher らの手法 [8] を用いて、視

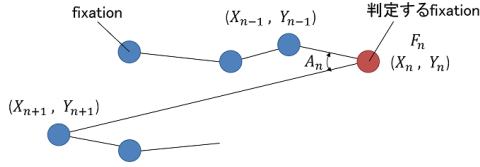


図4 改行判定の際の fixation の様子

点情報から fixation を抽出する。fixation を抽出することで、単語を読んでいる視点とそうでない視点に分けることができる。fixation を文書画像上に表示したものを図3に示す。図中の青い円は fixation を表し、円の大きさは fixation の長さを表している。

3.3.2 各 fixation が含まれる行の推定

各 fixation がどの単語にあたるかをより正確に推定するため、各 fixation が含まれる行を推定する。行情報を得るため、抽出した fixation 情報から改行を判定する。改行の際、視点は、読んでいた行の右端から次の行の左端へ向かう動きをする。この特性を利用し、横方向の閾値を設定することで改行を判定する。また、各行を読んでいる最中に別の行を読み返す際も改行とみなし、縦方向の閾値も設定する。図4のように、ある fixation を F_n とし、その座標を (X_n, Y_n) とする。 T_x , T_y を閾値とする、

$$X_{n+1} - X_n > T_x \quad (1)$$

または、

$$|Y_{n+1} - Y_n| > T_y \quad (2)$$

を満たすとき、 F_n と F_{n+1} の間を改行とみなす。

次に改行判定されるごとの fixation 群の y 座標の平均値を求める。そして、予め抽出しておいた文書情報を利用して、最も近い行に割り当てる。

3.3.3 各 fixation の単語への割当て

予め抽出しておいた文書情報を利用して、単語領域を抽出する。ある単語 W_m を包含する最小の矩形について、左端の x 座標を X_{wl} 、右端の x 座標を X_{wr} とする。また、視点の誤差を考慮し、単語 W_m の両側に w の幅をもたせる。ある fixation $F_n(X_n, Y_n)$ について、3.3.2で割り当てられた行中の単語 W_m に対し、

$$X_{wl} - w \leq X_n \leq X_{wr} + w \quad (3)$$

が成り立つとき、 $F_n(X_n, Y_n)$ を単語 W_m に割り当てる。なお、隣り合う2つの単語の両方において式(3)が成り立つとき、その両方の単語に割り当てる。

3.3.4 各視点の単語への割当て

各 fixation に含まれる視点を、その fixation に割り当てられた単語に割り当てる。また、fixation に含まれない視点はどの単語にも属さないとする。

3.4 入力データの抽出

Deep Learning を用いて学習するための入力データを抽出する。文書を読んでいる際、難しい単語を目にすると、読む速

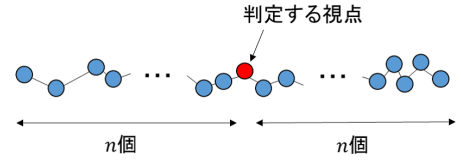


図5 視点パターンの例

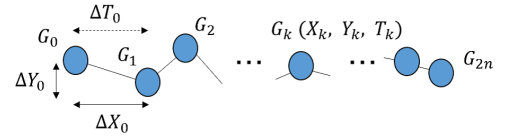


図6 視点パターンの特徴量

度が遅くなったり、読み返したり、または、読み飛ばしたりといったように、読むリズムが乱れる傾向がある。そこで、そのような読むリズムを入力データとして抽出するため、各単語を読んでいる際の前後2~3単語を読む視点群を視点パターンとして取得する。視点パターンの例を図5に示す。赤と青の円は視点を表す。特に赤い円は、ある視点パターンの中心の視点を表す。ある視点を中心とし、前後 n 個、合計 $2n + 1$ 個の視点群を視点パターンとして、中心の視点が難しい単語を読んでいる時のものかどうかを推定する。この視点パターンを全視点において抽出する。ただし、読み始めの n 点および読み終わりの n 点の合計 $2n$ 点については抽出しない。この視点パターンにおいて、図6のような特徴量を得る。視点パターンに含まれる $2n + 1$ 個の視点群を $G_0 \dots G_{2n}$ とする。また、 $G_k (0 \leq k \leq 2n)$ における座標および視点の取得時刻を (X_k, Y_k, T_k) とする。

$$\Delta X_0 = X_1 - X_0 \quad (4)$$

$$\Delta Y_0 = Y_1 - Y_0 \quad (5)$$

$$\Delta T_0 = T_1 - T_0 \quad (6)$$

とすると、視点パターンから抽出した入力データは、

$$(\Delta X_0, \Delta Y_0, \Delta T_0, \dots, \Delta X_{2n}, \Delta Y_{2n}, \Delta T_{2n}) \quad (7)$$

の $6n$ 次元となる。

次に、単語の出現頻度情報を抽出し、入力データに付加する。これは、同じ視点パターンであっても、難しいと感じている場合とそうでない場合があるためである。例えば、視点パターンが読み飛ばしを表しているとき、難しい単語を読み飛ばしている場合と、ストップワードのような簡単な単語を読み飛ばしている場合が考えられる。この問題を解決するため、「出現頻度が低いほど難易度の高い単語である可能性が高い」という仮定を導入する。すなわち、読んでいる単語の出現頻度情報を利用することで、視点情報では不十分な情報を補う。約10万語の単語の一般的な出現頻度のデータセット[9]から、全単語の出現頻度について対数を取り、0以上1以下の値を持つように正規化する。視点パターンの中心点である G_n が割り当てられた単語の出現頻度を、その視点パターンの入力データに付加する。なお、出現頻度のデータセットに含まれない単語は、出現頻度

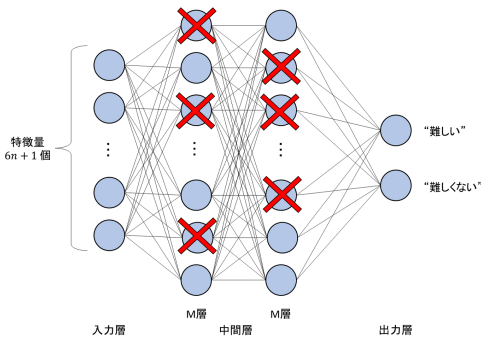


図7 提案手法の多層ニューラルネットワーク



図8 実験の様子

0 とする。

以上より、1つの視点パターンから合計 $6n + 1$ 次元の入力データを抽出する。なお、本提案手法では $n = 25$ とする。

3.5 Deep Learning を用いた学習

3.4 で抽出した視点パターンを Deep Learning を用いて学習し、主観的高難易度単語を推定する。Deep Learning とは、多層ニューラルネットワークを用いた学習である。提案手法で用いた多層ニューラルネットワークを図7に示す。提案手法で用いる多層ニューラルネットワークは、入力層 $6n + 1$ 個、中間層 M 個と M 個の2層、出力層2個の合計4層からなる。また出力層は、視点パターンが難しい単語を読んでいる時のものか、そうでないかを表している。過学習を防ぐため、各訓練データに対し中間層の素子をランダムに50%の確率で無視 (Dropout) しながら学習を進める。入力層には、3.4 で抽出した入力データを入力する。この際、難しい単語を読んでいる時の視点パターンとラベル付け (ラベル1) された入力データ数と、そうでない視点パターン視点パターンとしてラベル付け (ラベル0) された入力データ数に大きな差があると、学習が十分にできないことがわかっている。そこで、ラベル1の入力データ数と、同じ数のラベル0の入力データをランダムに抽出する。また、出力層では、ラベル1と推定される確率が閾値 T_p 以上であるとき、主観的高難易度単語を読んでいる際の視点パターンとして出力する。なお、提案手法では $M = 1500$ とし、活性化関数にはランプ関数、Optimizer には Adagrad を用いる。学習の際、バッチサイズ100のバッチ処理をおこなう。

3.6 主観的高難易度単語の推定

3.5 で、各視点パターンが主観的高難易度単語を読んでいる時のものかどうかを推定した結果から、主観的高難易度単語を推定する。各単語において、割り当てられた視点パターンのうち半数以上の視点パターンがラベル1 (難しい単語を読んでいる時の視点パターン) である場合、その単語を主観的高難易度単語として出力する。

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。提案手法と2つの比較手法との比較実験をおこなう。まず、入力データである各視点パターンが、難しい単語を読んでいる時のものかどうかについて、Deep Learning による推定精度の検

証をおこなう。ここでは、提案手法と視点情報のみを用いた手法について比較する。次に、主観的高難易度単語の推定精度について、提案手法と2つの比較手法を用いて検証する。

4.1 実験概要

大学生および大学院生の被験者34名 (男性:22名, 女性:12名) に英語の文書を読んでもらい、その際の視点情報をアイトラッカを用いて記録した。英語の文書は第二外国語として英語を勉強する人を対象としたニュースサイト [10] から抜粋した。文書のフォーマットは、より正確に視点を単語に割り当てるため、行間を広くとっている。文書は学習データ用に20文書、テストデータ用に1文書用意した。学習データ用の文書には、1~2行に1単語の割合で、造語に置き換えている。これは、英語能力の差に関わらず、全ての被験者から一定数以上の主観的高難易度単語のデータを取得するためである。この際、造語は英単語に見えるような単語にし、品詞が変化しないよう、接尾語を残した。アイトラッカは、ディスプレイに装着可能な Tobii EyeX を用いた。

実験の様子を図8に示す。被験者はディスプレイの前に座り、アイトラッカのキャリブレーションをおこなう。次に、被験者は学習データ用として10文書または15文書を選ぶ。被験者は、ディスプレイ上に表示された学習データ用の文書、および、テストデータ用の文書を1文書読み、その際の視点情報を記録する。その際、被験者には、読み返しを含め自由に読んでもらい、「意味を理解しながら一通り読むように」と指示した。文書を1文書読むごとに、文書の要約と難易度 (1:とても簡単, 2:簡単, 3:普通, 4:難しい, 5:とても難しい) を口頭で答えてもらう。これは、被験者が乱雑に読むのを防ぐためであり、これらの情報を学習には用いない。また、単語の頻出度情報を用いて、難しい可能性のある単語をリストアップしたものを被験者に渡す。この中には、予め置き換えておいた造語も含める。被験者は、その中で難しいと思った単語にマークをつける。これを学習のための正解データとして用いる。ただし、被験者が造語を難しい単語としてマークをしていなくても、難しい単語としてラベル付けをする。この際、難しい可能性のある単語をリストアップしたものをを用いるのは、文脈から意味が理解できた主観的高難易度単語についても、難しい単語としてラベル付けするためである。また、被験者の負担を減らすためでもある。

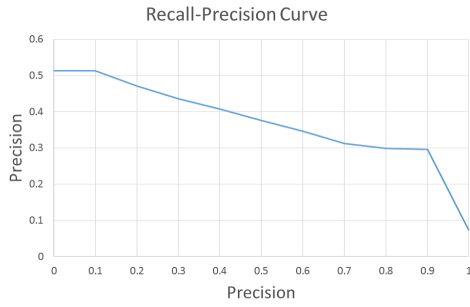


図 9 出力層の閾値 T_p を変化させた時の適合率と再現率の関係

表 1 難しい単語を読んでいる際の視点パターンの推定精度の比較

手法	再現率 (%)	適合率 (%)	F 値 (%)
提案手法	94.1	30.2	45.7
視点情報のみの手法	54.0	14.4	22.8

4.2 比較手法

提案手法の有効性を検証するため、単語の出現頻度情報のみで推定する手法、視点情報のみで Deep Learning を用いて推定する手法と比較実験をおこなう。

4.2.1 比較手法 1：単語の出現頻度情報のみを用いた推定

一般的に、出現頻度が低い単語ほど難しい単語だと考えられる。そこで、視点情報を用いず、単語の出現頻度から、主観的高難易度単語を推定する。まず、提案手法と同様に、約 10 万語の単語の出現頻度のデータセット [9] から、全単語の出現頻度を正規化したものを特徴量とする。出現頻度が閾値よりも小さい時、その単語を難しい単語として推定する。学習データに対し、F 値が最大になる時の閾値を求め、テストデータの主観的高難易度単語を推定する。なお、学習は被験者依存でおこなう。

4.2.2 比較手法 2：視点情報のみを用いた推定

提案手法と同様に視点パターンから視点情報による特徴量を取得する。その際、単語の出現頻度情報は付加しない。その後、提案手法と同様に Deep Learning を用いて学習し、各視点パターンが難しい単語を読んでいる時のものかを推定する。その結果を用いて、主観的高難易度単語を推定する。

4.3 実験 1：難しい単語を読んでいる際の視点パターンの推定精度検証

入力データである各視点パターンが、難しい単語を読んでいる時のものかどうかについて、Deep Learning による推定精度の検証をおこなった。被験者 34 名のうち、9 名がキャリアレーションに失敗したため、推定精度検証に含めなかった。キャリアレーションに失敗した原因として、眼鏡をかけていたことが考えられる。また、アイトラッカは瞳孔を取得する必要があり、目が細く睫毛が長い被験者の瞳孔を取得出来なかった可能性が考えられる。以上より、推定精度検証に用いた被験者は、25 名（男性：17 名、女性：8 名）であった。そのうち、9 名が裸眼、11 名がコンタクトレンズ着用者、5 名が眼鏡着用者であった。学習用データ用の文書を 15 文書読んだ被験者は 22 名、10 文書読んだ被験者は 3 名であった。また、そのうち 2 名において 1 文書ずつキャリアレーションに失敗していたため、学習用デー

表 2 主観的高難易度単語の推定精度の比較

手法	再現率 (%)	適合率 (%)	F 値 (%)
提案手法	90.8	24.0	37.5
単語の出現頻度情報のみの手法	59.3	17.4	26.7
視点情報のみの手法	46.9	15.9	21.8

Doctors in Canada have called for a ban on the use of perfumes and aftershave in hospitals and clinics. They say the chemicals in the scents can trigger asthma and allergies. Research shows that over half of asthma attacks are caused by irritants such as powerful smells. Besides perfumes and aftershave, these can include cigarette smoke, cleaning fluids and other strong fragrances and odors. Dr. Ken Fiegel and Dr. James Martin of McGill University in Canada wrote about the dangers of smells in hospital in the Canadian Medical Association Journal. They said: "Hospital environments free from artificial scents should become a uniform policy, promoting the safety of patients, staff and visitors alike."

提案手法

Doctors in Canada have called for a ban on the use of perfumes and aftershave in hospitals and clinics. They say the chemicals in the scents can trigger asthma and allergies. Research shows that over half of asthma attacks are caused by irritants such as powerful smells. Besides perfumes and aftershave, these can include cigarette smoke, cleaning fluids and other strong fragrances and odors. Dr. Ken Fiegel and Dr. James Martin of McGill University in Canada wrote about the dangers of smells in hospital in the Canadian Medical Association Journal. They said: "Hospital environments free from artificial scents should become a uniform policy, promoting the safety of patients, staff and visitors alike."

単語の出現頻度のみを用いた手法

Doctors in Canada have called for a ban on the use of perfumes and aftershave in hospitals and clinics. They say the chemicals in the scents can trigger asthma and allergies. Research shows that over half of asthma attacks are caused by irritants such as powerful smells. Besides perfumes and aftershave, these can include cigarette smoke, cleaning fluids and other strong fragrances and odors. Dr. Ken Fiegel and Dr. James Martin of McGill University in Canada wrote about the dangers of smells in hospital in the Canadian Medical Association Journal. They said: "Hospital environments free from artificial scents should become a uniform policy, promoting the safety of patients, staff and visitors alike."

視点情報のみを用いた手法

図 10 主観的高難易度単語の推定結果例

タは全部で 358 文書であった。

全 358 文書の学習用データを用いて視点パターンを Deep Learning で学習した。学習に用いた入力データ数は 217500 であった。テストデータには、全被験者 25 名のテストデータ用の文書を読んだ際の視点パターンを用いた。全部で 25 文書、入力データ数は 86233 であった。Deep Learning を用いて 200 回学習した際の推定精度について、出力層の閾値 T_p を変化させた時のグラフを図 9 に示す。F 値が最大になるとき、再現率 94.1%、適合率 30.2%、F 値 45.7% であった。また、視点情報のみを用いた手法との精度の比較を表 1 に示す。視点情報のみを用いた手法の場合、F 値が最大になるとき、再現率 54.0%、適合率 14.4%、F 値 22.8% であった。この結果から、視点情報のみでは、入力データとして不十分であったことがわかる。

4.4 実験 2：主観的高難易度単語の推定精度検証

提案手法により、主観的高難易度単語の推定精度検証をおこなった。提案手法の有効性を検証するため、単語の出現頻度情報のみを用いた推定手法および視点情報のみを用いた推定手法と比較した。推定精度を表 2 に示す。提案手法の精度が最も高く、再現率 90.8%、適合率 24.0%、F 値 37.5% であった。

また、図 10 に各手法により得られた主観的高難易度単語の

推定結果を示す。上図が提案手法、左下図が単語の出現頻度情報のみを用いた手法、右下図が視点情報のみを用いた手法である。各図の水色の矩形は主観的高難易度単語を推定された単語を表し、橙色に塗りつぶされた矩形は正解データを表している。

推定精度の比較により、単語の出現頻度情報のみを用いた手法や視点情報のみを用いた手法より、視点情報と単語の出現頻度情報の両方を用いた提案手法の方が精度が高いことが分かった。図 10 の視点情報のみを用いた手法により得られた推定結果を見ると、“a”、“in”、“on”などといったストップワードが含まれている。これは、ある単語を読み飛ばしているような視点パターンの場合、それが難しい単語なのか、ストップワードのような意味をなさないような単語なのか、視点情報のみでは判断できないからだと考えられる。また、図 10 の単語の出現頻度情報のみを用いた手法により得られた推定結果を見ると、主観的高難易度単語の一部を正しく検出できていない。一般的に出現頻度が低い単語を難しい単語として推定したが、頻度の高い単語であっても、被験者にとって難しく感じる単語も多くあり、単語の出現頻度情報のみでは不十分だったと考えられる。以上より、視点情報や単語の出現頻度情報単体のみでは不十分であり、両方を組み合わせた提案手法が有効であるということが言える。

提案手法では高い再現率を得ることができたが、適合率は 24.0% に留まった。推定された主観的高難易度単語の中には、“Canada”、“Dr”、“Martin”などといった固有名詞や敬称が誤った結果として出力されることが多くあったためである。これらの単語は、出現頻度のデータセット [9] に含まれなかったため、入力データ抽出時に頻度 0 (出現頻度が最も低い) としたことが原因であると考えられる。そこで、推定結果から固有名詞などを除く必要がある。提案手法により、主観的高難易度単語と推定された単語から、出現頻度のデータセットに含まれない単語を除いたところ、再現率 90.8%、適合率 34.4%、F 値 49.2% となった。適合率が 34.4% に留まった原因として、以下の 2 点が考えられる。まず、主観的高難易度単語の前後にある単語が誤って出力される点である。これは、難しい単語を読むリズムが、前後の単語の視点パターンにまで影響を及ぼしてしまったためだと考えられる。誤検出を防ぐために、入力データや入力する視点パターンの長さの再考が必要である。また、被験者が、理解していない単語を見落としている可能性が考えられる。そこで、出力された主観的高難易度単語に対し、被験者に再度、その単語が難しい単語かどうかを質問する必要があると考えられる。

5. まとめ

本稿では、アイトラッカから取得した視点情報より、Deep Learning を用いた、文書上の主観的高難易度単語の推定手法を提案した。提案手法では、各単語を中心に前後 2～3 単語を読んでいる際の視点パターンと単語の出現頻度情報を入力データとし、Deep Learning を用いて、主観的高難易度単語を推定する。推定精度は、再現率 90.8%、適合率 34.4%、F 値 49.2% であった。

提案手法の推定精度は、我々が実現を目指す、「文書を読むだけで、文書上のユーザにとって難しいと感じた部分を自動記録するシステム」に十分な精度といえる。推定された主観的高難易度単語を用いて、ユーザは個人の理解度に応じた単語テストを簡単に受けることが可能になり、英語学習に有用である。また、その単語テストの結果を入力データとして利用することで、システムがより高精度になることが見込める。

今後の課題としては、精度向上のための入力データの再考が必要である。提案手法では単語の出現頻度情報を用いたが、単語の文字数を入力データに付加したり、抽出した視点パターンの長さを変化させて比較することが考えられる。また、提案手法を用いたシステムのユーザビリティを調査する必要もある。本実験では、テストデータとして 25 名の被験者に対し 1 文書ずつしか検証していない。そこで、テストデータの文書を増やす必要がある。加えて、テストデータには学習データに入っていない被験者を入れた検証も必要である。

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) ならびに挑戦的萌芽研究 (15K12172) の補助による。

文 献

- [1] 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. 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.
- [2] Shiwei Cheng, Zhiqiang Sun, Lingyun Sun, Kirsten Yee, and Anind K Dey. Gaze-based annotations for reading comprehension. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1569–1572. ACM, 2015.
- [3] Ralf Biedert, Andreas Dengel, Mostafa Elshamy, and Georg Buscher. Towards robust gaze-based objective quality measures for text. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 201–204, 2012.
- [4] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 217–222, 2013.
- [5] Aline Godfroid, Frank Boers, and Alex Housen. An eye for words. *Studies in Second Language Acquisition*, Vol. 35, No. 03, pp. 483–517, 2013.
- [6] Ayano Okoso, Kai Kunze, and Koichi Kise. Implicit gaze based annotations to support second language learning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 143–146, 2014.
- [7] Ayano Okoso, Takumi Toyama, Kai Kunze, Joachim Folz, Marcus Liwicki, and Koichi Kise. Towards extraction of subjective reading incomprehension: Analysis of eye gaze features. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1325–1330. ACM, 2015.
- [8] Georg Buscher, Andreas Dengel, and Ludger van Elst. Eye movements as implicit relevance feedback. In *CHI'08 extended abstracts on Human factors in computing systems*, pp. 2991–2996, 2008.
- [9] Mark Davies. Word frequency data. <http://www.wordfrequency.info/>.
- [10] Sean Banville. Free esl materials.com. http://www.freeeslmaterials.com/sean_banville_lessons.html.