

視線情報からの未知語検出における個人適応と 単語親密度の影響の調査

Effects of Personalization and Word Familiarity on Eye Gaze based Unknown Words Detection

平岡 類 田中 宏季 サクティ サクリアニ 中村 哲
Rui Hiraoka Hiroki Tanaka Sakriani Sakti Satoshi Nakamura

奈良先端科学技術大学院大学 情報科学研究科

Nara Institute of Science and Technology, Graduate School of Information Science

We present initial work towards processing eye gaze while reading sentences. Various factors interfere understanding when a user reads sentences. For example, the appearance of unknown words which are not included in the user's vocabulary might be a factor for understanding especially in non-native language. We previously proposed a method to automatically detect unknown words based on eye gaze. However, the performance of detection is not sufficient for practical usage. To improve detection performance, this paper examines effects of (1) personalized classifier to each participant and (2) using word familiarity corpus instead of word frequency corpus. The experimental result demonstrated that personalization improves classification performance. In addition, word familiarity corpus was not effective for classification performance while it is correlated with known/unknown words label.

1. はじめに

人間の社会的な相互行為に関する非言語情報, 社会的信号を計測・解析し, 利用する取り組みが広く行われている [12][10]. 社会的信号処理を利用する取り組みの一つとして, 本稿では文章読解中のユーザの未知単語の推定技術を提案する. 未知語は文章理解の妨げになることが知られており, 特に第二外国語においては, 未知語の割合が増えることで, 母語に比べて影響が大きくなると考えられる. 未知語を書き手や読み手にフィードバックすることが出来れば, 自動で辞書を引いて注釈を付与したり, 読み手の言語習熟度レベルを把握することが可能となるため有用である. 未知語や難解な表現を処理している際には, 眼球運動に一定の反応が出ることが広く研究されており [11], ユーザの未知語を自動で検出し解説を付与するシステムも作成されている. しかしながら, 自己ペースでの眼球運動からの未知語検出において十分な精度が得られていない [5][4]. この理由の一つとして, 読み手による眼球運動の違いが存在するにも関わらず, いずれの手法も読み手の特性を考慮していないことが考えられる. そこで本稿では, 眼球運動計測を用いた未知語検出の精度向上のため (1) 個人に適応した識別器の作成, (2) 単語親密度コーパスを用いることを検証した. 識別器を個人に適応させることで F 値において精度改善が確認された. 一方で単語親密度は未知語と既知語間で有意に差が出る特徴量となったが, 識別精度の改善は確認されなかった.

2. 関連研究

眼球運動と文章読解に関する研究として, Rayner の研究がある. Rayner は文脈から推定しづらい単語や頻度の低い単語に対して, 固視の時間が通常よりも長くなることを示している [11]. Frenck-Mes らは注視の時間長さが他動詞の場合は名詞の方が長くなり, 自動詞の場合は動詞のほうが長くなることを示している [3]. また文章の理解度と眼球運動に関する研究に Gomez らの手法がある [8]. Gomez らは眼球運動計測を用いて, ユーザの英語習熟度と文章の理解度を推定する手法を提案

している. 同様の研究として, Kunze らはウェアラブルデバイス着用時の英文読解時の視線の動きから, ユーザの英語力を推定する手法を提案している [7]. 単語単位での理解度推定に関する研究に, Hyrskykari の研究がある [5]. Hyrskykari は, 未知語視認時に固視や注視の時間長さなどの眼球運動に反応が出ることを示している. 加えて眼球運動に関する特徴量と言語特徴量である語彙頻度に基づいた線形な未知語検出法を提案し, 視線からの辞書引きシステムを開発している. 以前我々は (1) 識別器に Support Vector Machine (SVM) を使用, (2) 新たな特徴量の使用, (3) 特徴量の選択法を提案することで, Hyrskykari の手法と比較し検出精度の改善効果が得られることを示した [4]. 本稿では (1) 識別器を個人に適応させること, (2) 特徴量に単語親密度コーパスを使用することで, SVM を用いた未知語検出を改善する手法を提案する.

3. SVM を用いた眼球運動計測に基づく未知語検出法

以下に SVM による眼球運動計測に基づく未知語検出法を述べる [4]. 一般に読解時の眼球運動の主な特徴は saccade (跳躍), fixation (固視) の 2 つとされている. fixation は定点を見るために眼球の動きを固定する運動のことである. 注視は対象物に連続で向け続けられた fixation の集合を表す. saccade は fixation 間に発生する跳躍運動であり, この間は認知活動がほとんど行われない. したがって, 得られた眼球運動データに対し, 固視なのか跳躍なのかを判断し, 主に認知活動を反映する fixation に対して特徴量を抽出する. またそれがどの単語に与えられたのかを判定し, 単語ごとに特徴量を生成する. 得られた特徴量の中で有効なものを選び出し, 識別器を作成する. 図 1 に各手順での処理の流れを示す. また以下に各処理の詳細を示す.

3.1 fixation と単語のタグ付け部

単語 w_x ごとの fixation の判定に対しては, 単語 w_x の始点座標 s_x と終点座標 e_x を取得し, fixation f_k がどの単語に与えられたかは以下の式で定義する.

連絡先: 連絡先: 平岡 類, 奈良先端科学技術大学院大学 情報科学研究科, hiraoka.rui.hj9@is.naist.jp



図 1: SVM を用いた眼球運動計測に基づく未知語検出法の流れ

$$w(k) = \arg \min_x (D(s_x, f_k), D(e_x, f_k)) \quad (1)$$

ここで $w(k)$ は k 番目の fixation が何番目の単語に属するかを定義する関数である。 $w(k) = x$ のとき k 番目の fixation は x 番目の単語に属することを指す。 $D(s_x, f_k)$ は fixation f_k と x 番目の単語 w_x の始点 s_x , $D(e_x, f_k)$ は終点 e_x のユークリッド距離を表す関数である。このいずれか一方の距離が最も小さくなる w_x に f_k は属する。

3.2 特徴量抽出部

この節で fixation を元にした特徴量の定義を示す。定義にしたがって、fixation と単語がタグ付けされたデータから特徴量を抽出する。 w_x 上での L 個の注視の集合を $G(w_x) = \{g_1(w_x), g_2(w_x), \dots, g_L(w_x)\}$ と定義する。

初期注視時間

単語 w_x 上の最初の注視時間 $g_{\text{first}}(w_x)$.

最大注視時間

単語 w_x 上に与えられた注視時間の中で最大のもの。

総注視時間

単語 w_x 上の注視時間の総和 $g_{\text{total}}(w_x)$.

fixation の回数

単語 w_x 上の fixation の回数 $N(w_x)$.

逆行回数

単語 w_x 上での視線の逆行回数 $Reg(w_x)$. 一度でも単語 $w_i, i > x$ を満たす w_i に視線が落ちたとき、別の単語から単語 w_x に fixation が遷移した回数で定義する。

fixation の平均時間

単語 w_x 上の fixation の平均時間 $f_{\text{mean}}(w_x)$.

瞳孔直径の変化量

単語 w_x 上の瞳孔直径の変化量 $P_{\text{variation}}(w_x)$ [8].

単語希少度

文章には a や the など、主観的な未知語となりにくいストップワードが大量に存在する。これらを排除するための事前情報として単語の希少度を考慮する。希少度は以下のように定義される [5]. 視覚刺激となる文章の全ての単語に対し、構文解析器を用いて、単語の原型を抽出する。各原型の単語について、コーパスにおける出現頻度の順位(降順)を抽出する。この順位を語彙の希少度とするが、ここで単語出現頻度総数の寄与率を考慮する。非母語を英語に設定しているため、単語頻度の抽出には、BNC(British National Corpus)[1]を使用する。BNCにおいて、上位100位の単語は、寄与率が50%を超え、上位6000位までで85%を超える。頻度の順位を正規化するため、上位100位以上の単語はすべて100とし、下位6000位以下のものについては、すべて6000とする。

単語長

1600x900の画面上での単語 w_x の物理的長さ。

3.3 特徴量選択部

いくつかの特徴量が識別に有効でない可能性を考慮し、識別に有効な特徴量の組み合わせだけを用いる。特徴量が増えるにつれて特徴量の組み合わせ数は指数乗に増えていくため、特徴量を評価し、評価の低いものを組み合わせ候補から除き、計算量を削減する。特徴量の評価として、未知語と既知語の間で、平均について非等分散 t 検定を行い、有意に差が出るものを特徴量として用いる。さらに未知語ラベルを1、既知語ラベルを0とした場合に、これらのラベルと各特徴量の相関係数を計算し、相関の高いものを特徴量として用いる。

3.4 識別器の作成

識別器の作成部では選択した特徴量に対して、複数の特徴量を考慮するためSVMを用いる。また非線形の識別を考慮し、RBFカーネル関数を用いて識別を行う。

4. 提案手法

本稿では未知語検出精度向上のため、識別器の個人適応を行うことで識別器の作成部を改善し、単語親密度を考慮することで特徴量抽出部分を改善することを提案する。

4.1 個人適応

被験者ごとに眼球運動に違いがあることが、[5][9]によって報告されており、読み手の個性を考慮しての学習を行うことで精度が改善する可能性が考えられる。そこで学習部において、(1)特徴量増幅法、(2)個別の学習器の作成により、個人適応法を提案する。

4.1.1 特徴量増幅法

学習器は単一とし、特徴量増幅法による転移学習によって個人適応を行う [2]. 眼球運動特徴量についてのみ、被験者ごとの特徴ベクトルを作成する。 $\mathbf{X} \in \mathbf{R}^m$ を入力とし、 \mathbf{Y} を出力とする。被験者 $i \in n$ のデータを \mathbf{D}^i とする。入力次元を被験者の人数分増やし、 $\mathbf{X}' \in \mathbf{R}^{m(n+1)}$ に拡張する。このとき $\mathbf{X} \rightarrow \mathbf{X}'$ への写像関数を $\Phi^1 \dots \Phi^n$ とし、入力ベクトル $\mathbf{x} \in \mathbf{X}$ に対して、

$$\begin{aligned} \Phi^1(\mathbf{x}) &= \langle \mathbf{x}, \mathbf{x}, \mathbf{0}, \dots, \mathbf{0} \rangle, \mathbf{x} \in \mathbf{D}^1 \\ \Phi^i(\mathbf{x}) &= \langle \mathbf{x}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{x}, \dots, \mathbf{0} \rangle, \mathbf{x} \in \mathbf{D}^i \\ \Phi^n(\mathbf{x}) &= \langle \mathbf{x}, \mathbf{0}, \mathbf{0}, \dots, \mathbf{x} \rangle, \mathbf{x} \in \mathbf{D}^n \end{aligned}$$

と定義する。新たな入力 \mathbf{X}' は被験者 i のデータに対しては、元のデータに加え、 i のデータと $m-1$ 個の 0 ベクトルが追加される。これによって、追加したベクトルは i の眼球運動データだけが増やされることで、 i のデータに対してのみ重みを付けたものとなる。

4.1.2 個別識別器の作成

入力データがどの個人から得られたものなのかは既知であるとし、訓練データから個人ごとの識別器を作成する。各テストデータに対して該当個人の識別器のみを用いることによって、個人適応を行う。

4.2 単語親密度の考慮

先行研究で考慮された特徴量に加え、単語親密度 [13] を特徴量として抽出する。これを単語希少度と同様に、ストップワードを排除するためのフィルタとして特徴量に加える。単語親密度は BNC の高頻度語をベースに作成された、一般的な日本人大学生にとっての英単語の親密度を表す指標であり、値が高いほど親密度が馴染みのある語であることを示す。単語の親密度は単語の頻度に比べて、日本人大学生が「どの程度知っているか」を反映していると考えられるため、未知語の識別に有効であると考えられる。

単語親密度

単語 w_x の親密度。日本人大学生において 1-7 段階でどれほど単語 w_x に馴染みがあるのかを調査したコーパスより抽出する。コーパスに含まれない単語については全て 1 とした。

5. 評価実験

本節では、提案手法が未知語検出精度に与える影響を実験的に検証する。実験は 20 代の大学卒業レベルの教育を受けた 10 名 (女性 3 名, 男性 7 名) のユーザーを対象に行った。全員が日本語を母語としており、正常または正常に矯正された視力を有していた。また全員が TOEIC で 475 点から 790 点のスコアを有していた。ディスプレイ上に視覚刺激となる英語文章を被験者に読んでもらい、その間の眼球運動を計測した。計測には、非装着型のイトラッカ Tobii pro X2-30 を使用した。刺激は PDF 形式で 31 インチのスクリーンに verdana font の 14pt で表示した。先行研究 [5] では行間を広げることで、計測誤差の影響を少なくしていたため、本研究でも同様に行間を広げた。文章読解のモチベーションを持続させるため、文章刺激終了後に内容についての要約課題を被験者に課した。読解時間の制約は特に設けず、被験者はマウス操作により、任意のタイミングで画面を切り替えるようにした。計測中はマウスカーソルが画面上に表示されないようにした。要約課題が終了次第、該当する刺激文章を被験者に再度見せ、未知語のアノテーションを行った。未知語のアノテーション基準は、ユーザーが注釈や説明を必要であったかどうかとし、各単語に印をつけてもらった。未知語は被験者一人あたり平均 64 語 (標準偏差 28.5 語) となり、9490 語の内、640 の単語が未知語とアノテーションされた。

識別器の作成部における個人適応の検証は、Leave-one-subject-out (LOSO) によって、個人性を一切考慮しない評価と、10-fold cross validation によって個人のデータを含む評価、特徴量増幅法、個別の学習器の作成によって個人適応を行った場合を比較する。また単語親密度の特徴量としての評価は、まず特徴量選択部で述べた特徴量の評価法に基づいて行う。また識別に単語親密度を必ず含んだ上での最適な特徴量の組み合わせを探索し、最も精度が良かったものを実際に用いて評価す

る。また単語の親密度と単語の希少度はストップワードのフィルタ処理として用いるため、同時に使用しないものとする。

6. 実験結果

本節で、個人適応と単語親密度の使用による実験結果を示す。

6.1 個人適応の検証

表 1: 個人適応の検証結果。

	LOSO	10-fold-cv	特徴量増幅法	個別識別器
適合率	0.502	0.507	0.493	0.539
再現率	0.608	0.616	0.650	0.612
偽陽性	0.038	0.043	0.058	0.035
F 値	0.550	0.556	0.561	0.573

表 3 より、未知語検出において提案した個人適応法は、F 値において精度向上に有効であることが確認された。個人適応を行った精度向上の有意性を確かめるため、Leave-one-subject-out と特徴量増幅法、個別識別器の作成を行った場合と Paired bootstrap 検定 [6] を行った結果、いずれの提案手法も Leave-one-subject-out と比較して有意差が確認された ($p < 0.05$)。しかしながら 10-fold cross validation との比較においては、特徴量増幅法は精度は改善したものの、有意差は見られなかった ($p > 0.1$)。一方で、個別識別器の作成と 10-fold cross validation との比較においては有意差が確認された ($p < 0.01$)。このことから識別器を個人適応することで、精度が改善する可能性が示唆された。

6.2 単語親密度の考慮

まず単語親密度が特徴量として有効なのかを検証するため、未知語と既知語の間で単語親密度の平均について非等分散 t 検定を行った。さらに未知語ラベルを 1、既知語ラベルを 0 とした場合に、これらのラベルと各特徴量の相関係数を計算し、識別に用いる特徴量を選別した。表 2 に特徴量の検証結果を示す。

表 2: 特徴量の評価結果。相関係数でソート済。太字は特徴量として識別に使用したものを表す。 t 検定の p 値は指数のみを記載してある。

特徴量	相関係数	p -value
単語親密度	0.484	e-273
単語希少度	0.462	e-233
単語長	0.339	e-171
最大注視時間	0.318	e-53
初期注視時間	0.265	e-15
総注視時間	0.252	e-45
fixation の回数	0.208	e-40
fixation 平均時間 :	0.153	e-31
瞳孔径変化量	0.136	e-19
逆行回数	0.109	e-16

評価の結果、いずれの特徴量も未知語と既知語間で平均に有意な差が表れており、提案した単語親密度が最も未知語既知語ラベルとの相関が高くなった。また未知語ラベルとの無相関検定

を行ったところ、いずれも相関は棄却されなかった ($p < 0.01$). このことから単語親密度が、主観的な未知語とある程度関連していることが示唆される。

以下に単語親密度を考慮した識別結果を示す。識別結果の評価は 10-fold cross validation で行った。単語親密度を考慮し

表 3: 単語親密度を考慮した識別結果.

	単語親密度無	単語親密度有
適合率	0.507	0.416
再現率	0.616	0.717
偽陽性	0.043	0.073
F 値	0.556	0.526

た結果、再現率は高くなったものの、適合率が下がったため、全体として識別精度は低くなった。このことから、単語親密度は単語希少度と比べて、識別に有効な特徴ではないことが示唆される。

7. まとめと今後の課題

本稿では従来手法を改善するための手法を提案し、その効果を検証した。

個人適応の実験においては、いずれの提案手法でも従来手法と比較し、精度改善が確認された。特徴量増幅法は Leave-one-subject-out での学習と比べて有意な改善となったが、ランダムな 10-fold cross validation での評価と比較では有意差は見られなかった。一方で個人ごとに識別器を作成した場合は、10-fold cross validation での評価と比べて有意に精度が改善した。これらによって個人適応した識別器を作成することで、精度改善効果があることが示された。

単語親密度は、特徴量評価において、未知語/既知語ラベルと最も高い相関があり、平均にも有意な差があったが、ストップワードのフィルタとしては単語希少度のほうが有効であった。原因として、単語親密度は最頻語の 3000 語のみを考慮しているのに対し、単語希少度は最頻語 6000 語を考慮している。したがって、単語親密度コーパスに含まれていない語に対して適切に学習が行われなかった可能性がある。

今後の課題として本研究では、実験設定を単純にするために未知語の出現に対する言語処理を対象としたが、文としては違和感があるようなものや複雑で統語処理が困難なもの、曖昧性が高いもの、またユーザーが興味によって注意を向ける可能性もある。本研究においてはこれらの要因を考慮していないため、さらなる精度改善のためには、これらのような未知語以外の要因を考慮する必要があると考えられる。未知語以外の要因を異なるものとして検出できれば、未知語視認に加え、誤り知覚など幅広い社会的信号を考慮したシステムを作成することが可能となる。

謝辞

本研究は、JSPS 科研費 26540117 および 16K16172 の助成を受けて行われたものである。

参考文献

[1] The british national corpus, version 3 (bnc xml edition). *Distributed by Oxford University Computing Services on behalf of the BNC Consortium.*, 2007.

[2] H. Daumé III. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815*, 2009.

[3] C. Frenck-Mestre and J. Pynte. Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology A*, 50(1):119–148, 1997.

[4] R. Hiraoka, H. Tanaka, S. Sakti, G. Neubig, and S. Nakamura. Personalized unknown word detection in non-native language reading using eye gaze. In *18th ACM International Conference on Multimodal Interaction (ICMI)*, Tokyo, Japan, November 2016.

[5] A. Hyrskykari. *Eyes in attentive interfaces: Experiences from creating iDict, a gaze-aware reading aid*. Tampereen yliopisto, 2006.

[6] P. Koehn. Statistical significance tests for machine translation evaluation. In *EMNLP*, pages 388–395. Citeseer, 2004.

[7] K. Kunze, H. Kawaichi, K. Yoshimura, and K. Kise. Towards inferring language expertise using eye tracking. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 217–222. ACM, 2013.

[8] P. Martínez-Gómez and A. Aizawa. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 95–104. ACM, 2014.

[9] H. Miyata, Y. Minagawa-Kawai, S. Watanabe, T. Sasaki, and K. Ueda. Reading speed, comprehension and eye movements while reading Japanese novels: Evidence from untrained readers and cases of speed-reading trainees. *PLoS ONE*, 7(5), 2012.

[10] C. Pelachaud, V. Carofiglio, B. De Carolis, F. de Rosis, and I. Poggi. Embodied contextual agent in information delivering application. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*, pages 758–765. ACM, 2002.

[11] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[12] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.

[13] H. Yokokawa. Nihonjin eigogakushushano eitan-goshimitsudo onseipan: Kyouiku kenyuunotameno dainigengogakudeetabeesu [english word familiarity of japanese learners of english, audio edition: Second language research database for education and research]. *Japan: Kuroshio*, 2009.