

スマートフォンを用いた 読書活動データからの未知単語推定

ESTIMATION OF UNKNOWN WORDS FROM LEARNER'S BEHAVIOR USING SMARTPHONE

東村理功¹
Riku Higashimura

岩田基²
Motoi Iwata

黄瀬浩一²
Koichi Kise

大阪府立大学 工学域¹
College of Engineering, Osaka Prefecture University

大阪府立大学 工学研究科²
Graduate School of Engineering, Osaka Prefecture University

1 はじめに

語彙の習得は語学学習の基本である。新たな語彙を習得するためには、学習者個人が意味を知らない単語(以下、未知単語)を何らかの方法で語学学習の対象にしなければならない。そこで、文書中の未知単語を推定する指標の一つとして、単語の難しさを表す出現頻度を用いる方法がある。ここで、出現頻度は全ての学習者に共通する客観的な指標である。そのため、出現頻度の高い簡単な単語が未知単語である場合、学習の必要性が高いにも関わらず、出現頻度のみに基づいた学習では学習漏れの可能性がある。

本研究ではスマートフォンのセンサデータで得られる読書活動データを用いて、学習者個人の英文読解能力 [1] を加味することで未知単語推定の精度向上を目指す。実験の結果、実験参加者 14 名のうち 12 名に対して、読書活動データが未知単語推定に寄与することを確認した。

2 提案手法

提案手法では、客観的な指標である文書特微量とスマートフォンのセンサデータで得られる読書活動データを用いて未知単語であるか否かを識別する。

まず、スマートフォンの画面上で文書中の数単語を 1 行ずつ表示するモバイルアプリケーションを実装し、実験に用いる。この 1 行ごとに文書特微量・読書時間・スマートフォンの 3 軸加速度 (2Hz)・読み返し回数を取得する。文書は Newsela データセット¹を使用する。

次に、取得した特微量を 1 行単位から 1 単語単位に抽出する。この際、冠詞などの stop words や固有名詞を除く。また、抽出した単語の出現頻度を、推定に用いる文書特微量として追加する。

そして、得られた特微量の中から有効な特微量を Sequential Backward Floating Selection によって選択する。選択された特微量をもとにサポートベクターマシン (SVM) を用いて未知単語であるか否かを識別する。

3 実験

提案手法の精度を検証するため、実験参加者 14 名(男性 6 名、女性 8 名)を対象に実験を行った。

3.1 文書特微量と読書活動データの取得

端末は全て実験参加者各自のスマートフォンを利用し、9 文書読んでもらった。実験参加者が 1 つの文書を読んでいる間、スマートフォンの画面上で 1 行ごとに文書特微量・読書時間・スマートフォンの 3 軸加速度 (2Hz)・

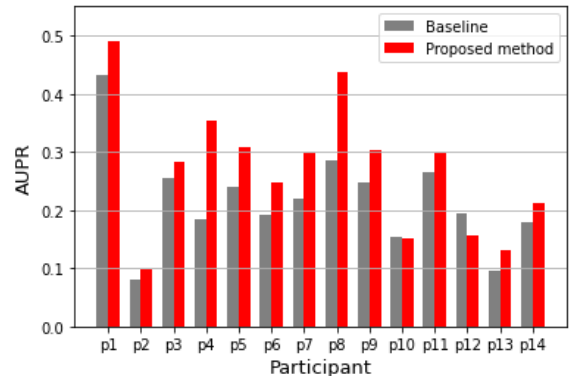


図 1: 提案手法とベースラインの AUPR

読み返し回数を取得した。そして、1 つの文書を読み終えるごとに、未知単語を選択してもらった。

3.2 評価方法

ユーザ依存で Leave-One-Document-Out 交差検証を行い、PR 曲線の AUC である AUPR を評価指標として計算した。全ての学習者に対して共通する客観的な指標である文書特微量のみで学習した AUPR をベースラインとし、提案手法の AUPR と比較した。

3.3 結果

提案手法およびベースラインの AUPR を比較した結果を図 1 に示す。図 1 より、実験参加者 14 名のうち 12 名に対して、提案手法がベースラインを上回る結果となった。また、この 12 名に対して選択された特微量の内、読書活動データの数は平均で 4.4 個だった。一方、ベースラインが提案手法を上回る結果となった p10 と p12 の 2 名に対して選択された特微量の内、読書活動データの数はそれぞれ 2 個と 1 個だった。このことから、今回の実験で取得した読書活動データが未知単語推定に寄与しない実験参加者も存在することが分かった。

謝辞

本研究の一部は JST CREST (JPMJCR16E1), JST Trilateral AI Research (JPMJCR20G3), JSPS 科研費基盤 (B) (20H04213), JSPS 国際共同研究強化 (B) (20KK0235), 阪大 Society5.0 グランドチャレンジの補助による。

参考文献

[1] 長沼君主, 和田朋子. 速読力とテキストの読みやすさによる英語リーディング能力の測定. 日本言語テスト学会研究紀要, Vol. 5, pp. 34-45, 2002.

¹<https://newsela.com/>