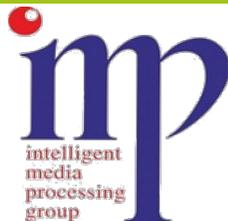




OSAKA  
PREFECTURE  
UNIVERSITY



# 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

大阪府立大学 工学域 情報工学課程  
知能メディア処理研究室

東村 理功

※大阪府立大学大学院工学研究科倫理委員会で承認済み

# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

## 1. 背景・目的

### ■背景～アプリで単語学習～

読んでいる英文の中から

自分が知らない英単語(未知単語)を

自動で記録できると便利



## 1. 背景・目的

### ■ 出現頻度

英単語の難しさを表す出現頻度

難しい単語はどの学習者にも未知単語である可能性が高い



易しい

難しい



speak  
go  
camera

choir  
stubby  
braille

## 1. 背景・目的

### ■ 出現頻度の問題点～客観的な指標～

易しい

難しい



speak  
go  
camera

choir  
stubby  
braille

出現頻度の低い単語を知っていることもあれば  
出現頻度の高い単語を知らない こともある

## 1. 背景・目的

### ■改善策

個人の読書活動データを利用することで  
学習者個人に応じた未知単語を推定



文書特徴量



読書活動

学習者固有の  
振る舞い

## 本研究の目的

---

スマートフォンで得られる  
読書中の活動データを用いて  
**未知単語を推定**

# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

## 2. 関連研究

### ■ 長沼らの研究 英文読解能力と読書時間

学習者の英文読解能力の指標として

読書時間を用いることの妥当性を確認



# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

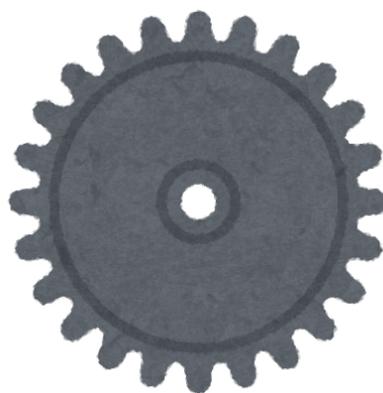
### 3. 提案手法

#### ■ 提案手法の流れ

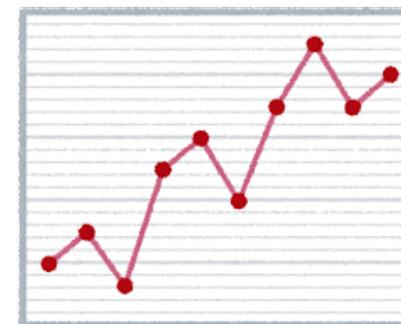
本研究の最終目標：未知単語の推定



データ取得



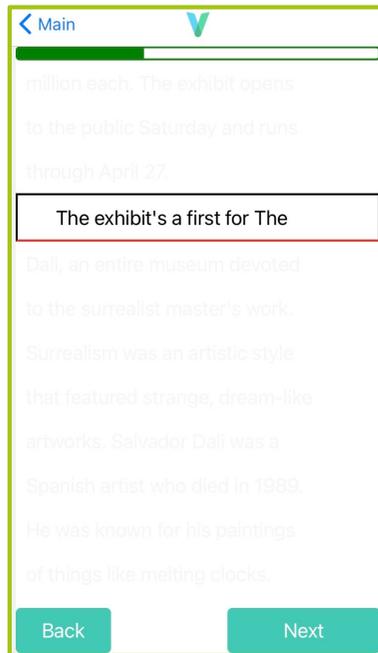
特徴選択



識別

### 3. 提案手法

#### ■ 読書活動データを 1 window ごとに取得



← **1 window** : スマホ画面上の 1 行に相当

- ・ 文書特徴量
- ・ 読書時間
- ・ スマホの3軸加速度(2Hz)
- ・ 読み返し回数      を取得

データ取得



特徴選択



識別

### 3. 提案手法

#### ■ 推定に用いる特徴量

一例

The exhibit's a first for The



- exhibit
- first

データ取得



特徴選択



識別

No	推定に用いる特徴量
1-2	単語数, 文字数
3-4	平均単語長, 平均文章長
5	Automated Readability Index
6	単語の出現頻度
7-9	出現頻度の {最小値, 最大値, 平均値}
10-11	機能語の {数, 占める割合}
12-13	7文字以上の単語の {数, 占める割合}
14	総読書時間
15-16	{1単語, 1文字} あたりの読書時間
17-19	スマートフォンの {x, y, z} 軸加速度の平均値
20-22	スマートフォンの {x, y, z} 軸加速度の最大値
23-25	スマートフォンの {x, y, z} 軸加速度の最小値
26-28	スマートフォンの {x, y, z} 軸加速度の分散
29	読み返し回数

※ 冠詞などのstop wordsや固有名詞は除く

### 3. 提案手法

#### ■ 特徴選択

Sequential Backward Floating Selection

- ◆ スコアリング : PR曲線のAUC
- ◆ 検証法 : Leave-One-Document-Out交差検証



### 3. 提案手法

#### ■ 識別

SVMで未知単語であるか否かを識別

- ◆ カーネル：RBF
- ◆ 2値分類 (未知単語であるか否か)
- ◆ ユーザ依存で推定



# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

## 4. 実験

### ■ 実験の概要

提案手法を検証するためデータを収集

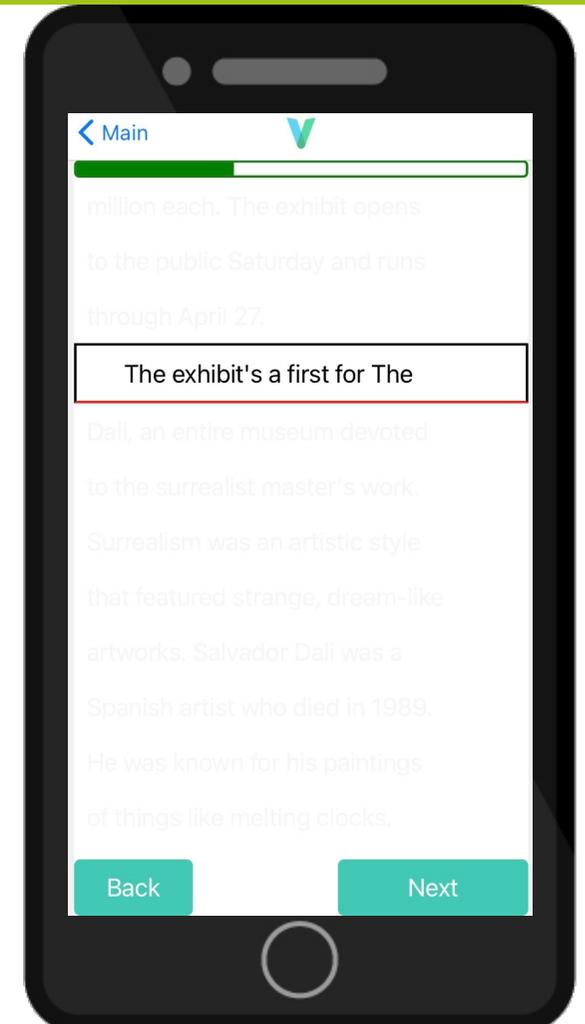
参加者 : 22名 (男性11名, 女性11名)

年齢 : 22.3 ± 1.5 歳

報酬 : 4000円 (4時間分)

文書数 : 9文書 (Newselaデータセット)

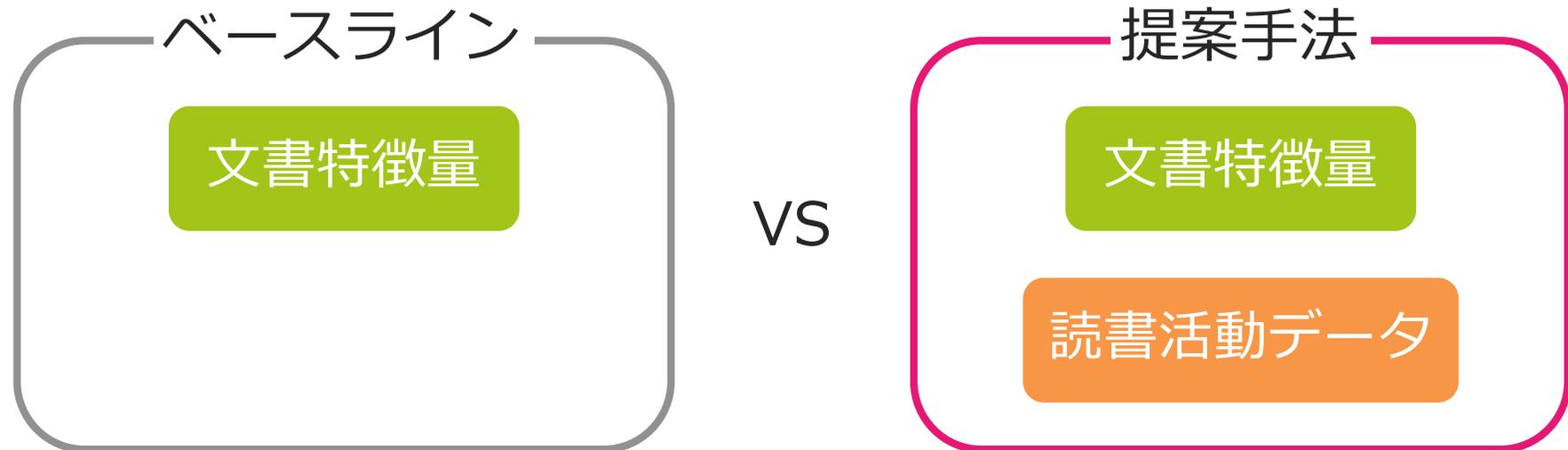
1文書を読む毎に  
未知単語 (意味を知らない単語) を記録



## 4. 実験

### ■ 評価方法

- ◆ 評価指標は、**実験参加者ごと**に求めたPR曲線のAUC



## 4. 実験

### ■ 提案手法の検証

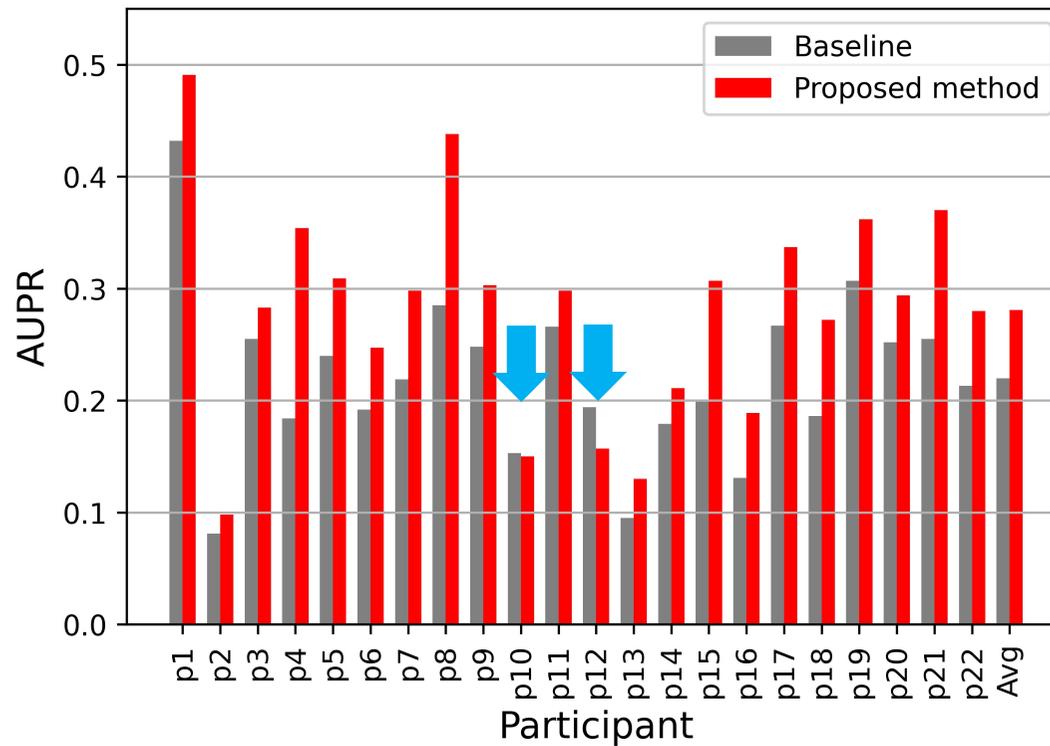
実験参加者ごとに

Leave-One-Document-Out交差検証



## 4. 実験

### ■ 結果と考察 (AUPRの比較)

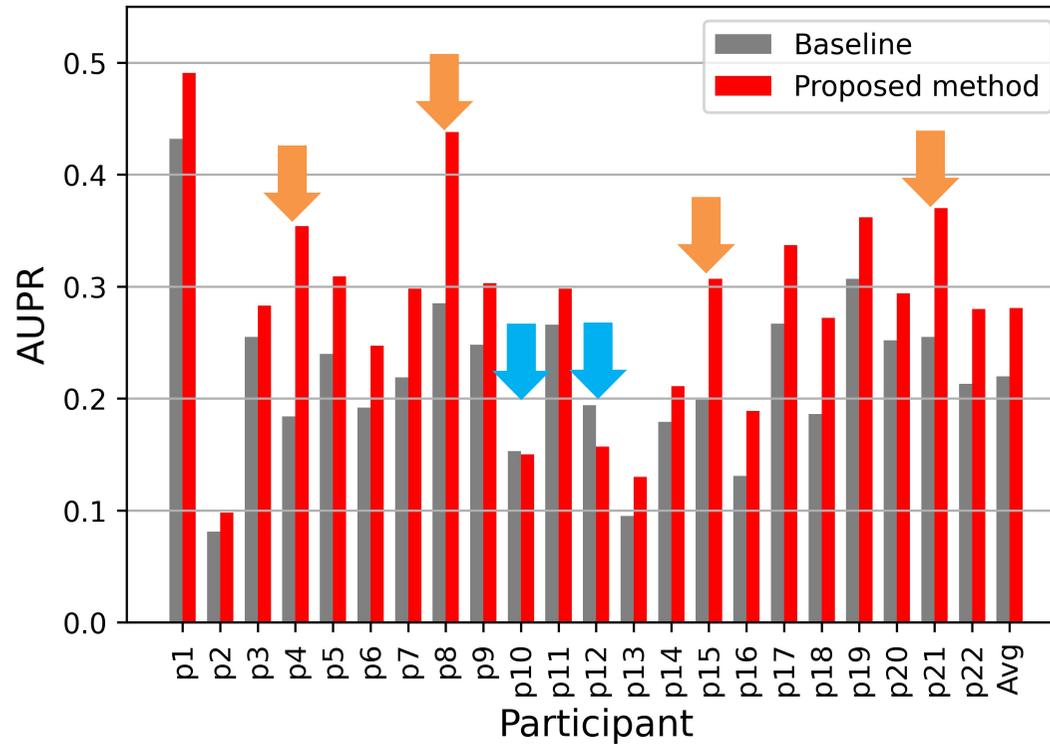


- 22名の内20名がベースラインを上回った
- 2標本t検定で有意差があった

※ 求められた p値 =  $4.54 \times 10^{-6}$

## 4. 実験

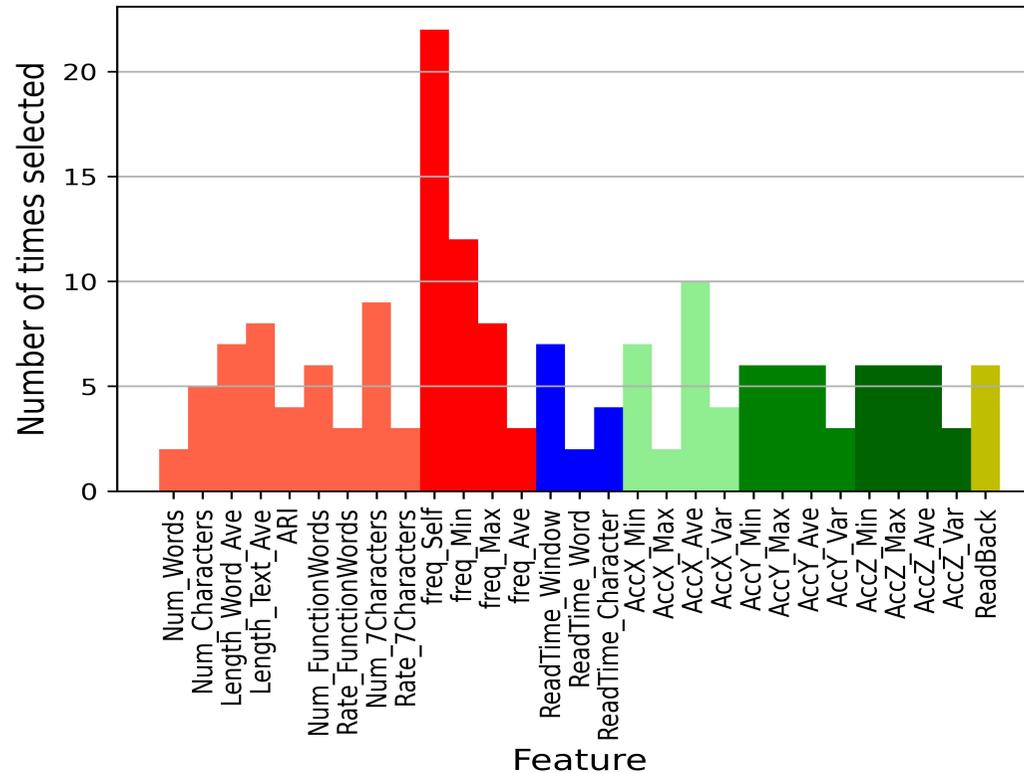
### ■ 結果と考察 (AUPRの比較)



- **効果的であった人** : 英文を読むのが**苦手**な傾向
- **効果的でなかった人** : 英文を読むのが**得意**な傾向

## 4. 実験

### ■ 結果と考察 (特徴量の比較)



- ・ **出現頻度**は多くの人に有効な特徴量
- ・ **スマホの3軸加速度(2Hz)**は**読書時間**と同程度に有効

# 目次

1. 背景・目的
2. 関連研究
3. 提案手法
4. 実験
5. まとめ・今後の課題

## 5. まとめ・今後の予定

### ■まとめ

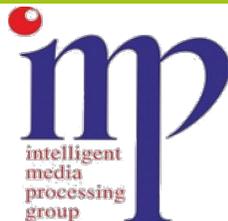
- ◆ 読書活動データを加味することで  
学習者個人に応じた未知単語推定を試みた

### ■今後の課題

- ◆ スマホで取得できる読書活動データをさらに追加



OSAKA  
PREFECTURE  
UNIVERSITY



# 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

大阪府立大学 工学域 情報工学課程  
知能メディア処理研究室

東村 理功

※大阪府立大学大学院工学研究科倫理委員会で承認済み