

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-152864  
(P2019-152864A)

(43) 公開日 令和1年9月12日(2019.9.12)

(51) Int.Cl.	F I	テーマコード (参考)
G09B 19/00 (2006.01)	G09B 19/00	H 2C028
G09B 7/02 (2006.01)	G09B 7/02	
G09B 19/06 (2006.01)	G09B 19/06	

審査請求 未請求 請求項の数 10 O L (全 26 頁)

(21) 出願番号 特願2019-37662 (P2019-37662)  
 (22) 出願日 平成31年3月1日(2019.3.1)  
 (31) 優先権主張番号 特願2018-37535 (P2018-37535)  
 (32) 優先日 平成30年3月2日(2018.3.2)  
 (33) 優先権主張国・地域又は機関  
 日本国(JP)

(出願人による申告)平成28年度、国立研究開発法人  
 科学技術振興機構、戦略的創造研究推進事業、産業技術  
 力強化法第19条の適用を受ける特許出願

(71) 出願人 505127721  
 公立大学法人大阪府立大学  
 大阪府堺市中央区学園町1番1号  
 (74) 代理人 100065248  
 弁理士 野河 信太郎  
 (74) 代理人 100159385  
 弁理士 甲斐 伸二  
 (74) 代理人 100163407  
 弁理士 金子 裕輔  
 (74) 代理人 100166936  
 弁理士 稲本 潔  
 (72) 発明者 黄瀬 浩一  
 大阪府堺市中央区学園町1番1号 公立大学  
 法人大阪府立大学内

最終頁に続く

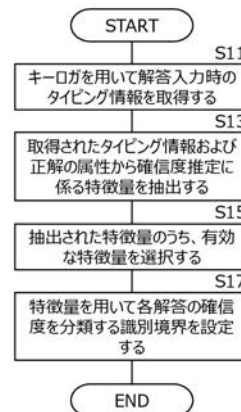
(54) 【発明の名称】 解答に対する確信判定方法および確信判定プログラム

(57) 【要約】

【課題】キーログを用いて取得されるタイピング情報または手書き文字入力に係る情報に基づいて、設問解答時の確信の有無を判定する。

【解決手段】設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定する方法であって、コンピュータが実行するステップとして、解答の入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備える確信判定方法。

【選択図】図1



**【特許請求の範囲】****【請求項 1】**

設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定する方法であって、

コンピュータが実行するステップとして、

解答の入力に係る履歴を文字入力情報として取得するステップと、

得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、

抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備える解答者の確信判定方法。

10

**【請求項 2】**

前記特徴量が、解答に要した時間、タイピングまたは手書きによる入力操作の時間間隔または消去の操作に係るものである請求項 1 に記載の確信度判定方法。

**【請求項 3】**

前記特徴量が、最後の文字のタイプ入力または最後のストロークの手書き入力から解答完了の操作までの時間間隔を含む請求項 1 または 2 に記載の確信判定方法。

**【請求項 4】**

前記特徴量が、設問が提示されてから最初の文字のタイプ入力または最初のストロークの手書き入力までの時間間隔を含む請求項 1 ~ 3 の何れか一つに記載の確信判定方法。

**【請求項 5】**

前記特徴量が、タイプ入力における文字入力間または手書き入力におけるストローク入力間のインターバルの最大値を含む請求項 1 ~ 4 の何れか一つに記載の確信判定方法。

20

**【請求項 6】**

前記特徴量が、設問が提示されてから解答完了の操作までに要した時間に基づく値、タイプ入力における文字入力間または手書き入力におけるストローク入力間のインターバルの平均、前記インターバルの標準偏差、前記インターバルの最小値、前記インターバルの中央値の少なくとも何れかを含む請求項 1 ~ 5 の何れか一つに記載の確信判定方法。

**【請求項 7】**

確信の有無の判定がサポートベクトルマシンの手法に基づき、選択された特徴量を用いて、確信の有無を判定する請求項 1 ~ 6 の何れか一つに記載の確信判定方法。

30

**【請求項 8】**

設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定するプログラムであって、

コンピュータに実行させるステップとして、

解答の入力に係る履歴を文字入力情報として取得するステップと、

得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、

抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備える解答者の確信判定プログラム。

**【請求項 9】**

コンピュータが、

予め定められた複数の設問のうち何れかを出題するステップと、

解答者から解答のタイピングまたは手書きによる文字入力を受付けて、請求項 1 ~ 7 の何れか一つの確信判定方法により解答に対する確信の有無を判定するステップと、

解答の正誤および確信の有無の判定を参照し、確信なしと判定された正答を再出題すべき設問として選択するステップとを備える出題選択方法。

40

**【請求項 10】**

前記コンピュータが、確信ありと判定された誤答を、確信なしと判定された誤答よりも短い間隔または高い頻度で再出題されるように決定するステップをさらに備える請求項 9 に記載の出題選択方法。

50

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、情報処理装置を用い、設問に対して解答者に文字を入力させ正解を完成させる形式の質疑応答において、解答に対する解答者の確信の有無を判定する手法に関する。

【背景技術】

【0002】

質問に対する解答を、情報処理装置に文字入力させ、学習や情報収集に用いることが近年多くなっている。

例えば、国際化の影響により、英語学習への関心が高まっているが、英語の習得には英単語の暗記、英文法の理解が必要である。その中でも、英単語の暗記は、英文法の理解に比べて、暗記の確認を反復して行う必要があるため、非常に手間のかかる地道な作業である。

地道な努力を要する英単語の暗記を、ゲーム感覚で効率的に行えるツールとしてスマートフォンやタブレットなどの情報処理装置で利用できる英単語アプリある。

一般的に英単語アプリでは、英単語を暗記し、英単語の定着度合いをテスト形式で確認し、その中から間違えた英単語を復習リストに登録して、復習リストを復習し、テストに再挑戦する学習方法を採用している。

【0003】

一方、情報処理装置にタイプ入力された内容および関連情報をデータとして記録するキーログが知られている。キーログは個人認証の自動入力ツール等として一般に用いられている。キーログは、アプリケーションの一つとして記録対象の情報処理装置に実装される形態が一般的である。

数は少ないものの、キーログから取得したタイピング情報を人の心理状態の推定に利用している研究がいくつか存在している。

【0004】

中村らは、e-learning教材において設問に解答する際のユーザに係る情報を取得してそのユーザの教材に対する主観的難易度を推定している（例えば、非特許文献1参照）。ただし、キー入力以外に顔画像、マウス操作からユーザに係る情報を取得しており、キー入力に関する特徴としては、キー入力の平均発生間隔のみを推定に使用している。

Eppらは、キーログから取得したタイピング情報から、ユーザの心理状態を推定する手法を提案している（例えば、非特許文献2参照）。ただし、対象としているのは、用意された文章をそのままタイプ入力するというタスクで、設問に解答するような形式はとっていない。

また、Allenらは、タイピングでエッセイを作成している際にキーログから取得した情報と文書情報から退屈度、集中度を推定する手法を提案している（例えば、非特許文献3参照）。退屈度の推定ではタイピングの停止時間とタイプした文字数が、集中度の推定ではそれらに加えタイピング速度とミスタイプの回数が有効であることが述べられている。

Ghoshらは、スマートフォンでタイピングを行った際にキーログから取得した情報を用いて、ユーザのストレス度合いなどを推定する手法を提案している（例えば、非特許文献4参照）。Ghoshらの手法においても、タイピング速度、タイプした文字数、そしてタイプミスの回数が有効であることが示されている。

【0005】

次に、設問解答時の確信を取り上げた研究について述べる。

浅井らは、タブレットで数学の設問に解答している際の筆圧や筆記速度、筆記間隔などの情報を用いて、つまづいた箇所を検出する手法を提案している（例えば、非特許文献5参照）。

三木らは、英単語の並べ替え設問に解答している際のマウスの軌跡情報や解答時間を用いて迷いの推定を行っている（例えば、非特許文献6参照）。

10

20

30

40

50

山田らは、多肢選択問題に解答している際の視点情報を用いて、解答している設問に対する確信度を推定する手法を提案している。山田らの手法では、確信度判定のために、視点情報から抽出した特徴量や解答時間を使用している（例えば、非特許文献7参照）。

しかし、キーログから取得したタイピング情報と設問解答時の確信についての関係性について触れている研究は確認されていない。

【0006】

情報処理装置への入力手段としては、タイプ入力の他に手書き入力も多用されている。手書き入力を人の心理状態の推定に利用する研究として、例えば以下のものが挙げられる。

Yuらは、異なる作業負荷の問題を手書きの入力操作で解いて、その作業負荷を見積もるように参加者に求めた。彼らは、特徴量として手書きの各ストロークの速度、圧力およびストローク長を抽出した（例えば、非特許文献8、9参照）。

Ugurluらは、宿題や試験問題を解くために書かれた文字を集めることによって、ユーザの感情の変化を分析した（例えば、非特許文献10参照）。

Kishiらは、書込み間隔特徴量だけを使用する弱点検出アルゴリズムを提案した。このアルゴリズムでは、しきい値は手動で設定され、各ユーザによって異なる（例えば、非特許文献11参照）。

AsaiとYamanaは、ストロークレベルの手書き分析に基づいて、ユーザのフラストレーションと忘れがちな項目を推定している（例えば、非特許文献12および13参照）。

【先行技術文献】

【非特許文献】

【0007】

【非特許文献1】中村光貴、荒見篤郎、上久保竜輝、福嶋政期、苗村健“英単語暗記システムのための綴りの類似性に関する基礎検討”PRMU（電子情報通信学会 パターン認識・メディア理解研究会）2017-116, Jan. 2018.

【非特許文献2】C. Epp, M. Lippold, and R. L. Mandryk "Identifying Emotional States using Keystroke Dynamics," CHI, May. 2011.

【非特許文献3】L. Allen, C. Mills, M. Jacovina, S. Crossley, S. D' Mello, and D. McNamara "Investigating Boredom and Engagement during Writing Using Multiple Sources of Information: The Essay, The Writer, and Keystrokes," Learning Analytics and Knowledge, Apr. 2016, pp.114-123.

【非特許文献4】S. Ghosh, N. Ganguly, B. Mitra, and P. De "Evaluating Effectiveness of Smartphone Typing as an Indicator of User Emotion," ACII, 2017.

【非特許文献5】浅井洋樹、野澤明里、苑田翔呉、山名早人“オンライン手書きデータを用いた学習者のつまずき検出”、DEIM（電子情報通信学会 データ工学と情報マネジメントに関するフォーラム）、2012.

【非特許文献6】三木良介、宮崎佳典、厨子光政、法月健“英単語並べ替え問題における迷い抽出機能の実装”、情報処理学会第76回全国大会、2014.

【非特許文献7】山田健斗、大社綾乃、藤好宏樹、星加健介、Olivier Augereau、黄瀬浩一“英語多肢選択問題解答時の視線に基づく確信度推定”、PRMU2016-192, Feb. 2017.

【非特許文献8】Kun Yu, Julien Epps, and Fang Chen. 2011. Cognitive load evaluation of handwriting using stroke-level features. In Proceedings of the 16th international conference on Intelligent user interfaces. ACM, 423-426.

【非特許文献9】Kun Yu, Julien Epps, and Fang Chen. 2013. Mental workload classification via online writing features. In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 1110-1114.

【非特許文献10】Bora Ugur, Rembiye Kandemir, Aydin Carus, and Ercan Abay. 2016. An Expert System for Determining the Emotional Change on a Critical Event Using Handwriting Features. TEM Journal 5, 4 (2016), 480.

【非特許文献11】Kazuya Kishi and Motoki Miura. 2018. Detecting Learners aAZ We

10

20

30

40

50

ak Points Utilizing Time Intervals of Pen Strokes. International Journal of Learning Technologies and Learning Environments 1, 1 (2018), 61-77.

【非特許文献12】Hiroki Asai and Hayato Yamana. 2013. Detecting student frustration based on handwriting behavior. In Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology. ACM, 77-78.

【非特許文献13】Hiroki Asai and Hayato Yamana. 2015. Detecting Learner's To-Be-Forgotten Items using Online Handwritten Data. In Proceedings of the 15th New Zealand Conference on Human-Computer Interaction. ACM, 17-20.

【発明の概要】

10

【発明が解決しようとする課題】

【0008】

例えば、前述した英単語アプリで、間違えた英単語だけを復習するだけでは、まだ不十分といえる。なぜなら、正解した英単語にも復習が必要な場合があるからである。具体的には、確信が持てずに正解した英単語を復習せず放置していると、再びテストを行った際には間違えてしまうおそれがある。そのため、確信が持てずに正解した英単語も復習リストに加える必要がある。また、正解であると確信をもって解答したにも関わらず間違えてしまった場合にも、英単語の意味や綴りを勘違いしている可能性がある。この場合についても、注意して復習に取り組みなければ再び間違えてしまう。したがって、確信をもって解答したにも関わらず間違えてしまった英単語については、復習リストでも注意を促す必要がある。

20

【0009】

効率的な復習のためには、単なる正誤だけでなく解答時の確信の有無に係る情報を取得し、プランを策定する必要がある。

この発明は、以上のような事情を考慮してなされたものであって、キーログを用いて取得されるタイピングに関する情報（以下、この明細書でタイピング情報という）または手書き文字入力に係る情報（以下、前記タイピング情報とまとめて文字入力情報ともいう）に基づいて、設問解答時の確信の有無を判定する手法を提供するものである。

【課題を解決するための手段】

【0010】

30

この発明は、設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定する方法であって、コンピュータが実行するステップとして、解答の入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備える解答者の確信判定方法を提供する。

また、異なる観点からこの発明は、設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定するプログラムであって、コンピュータに実行させるステップとして、解答の入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備える解答者の確信判定プログラムを提供する。

40

さらに、異なる観点からこの発明は、コンピュータが、予め定められた複数の設問のうち何れかを出题するステップと、解答者から解答のタイピングまたは手書きによる文字入力を受けて、前述の確信判定方法により解答に対する確信の有無を判定するステップと、解答の正誤および確信の有無の判定を参照し、確信なしと判定された正答を再出题すべき設問として選択するステップとを備える出题選択方法を提供する。

【発明の効果】

50

## 【 0 0 1 1 】

この発明による確信判定方法は、解答のタイピングまたは手書きによる文字入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するので、文字入力情報に基づいて、設問解答時の確信の有無を判定できる。

この発明による確信判定プログラムについても同様のことがいえる。

さらに、この発明による出題選択方法は、解答の正誤および前述の確信判定方法によって得られた確信の有無を参照して、再出題すべき設問を選択するので、解答者が確信を持ってない正答や勘違いしている誤答を再出題できる。

10

## 【 図面の簡単な説明 】

## 【 0 0 1 2 】

【 図 1 】 第 1 の実施形態において、解答者の確信の有無を判定する処理の流れを示すフローチャートである。

【 図 2 】 第 1 の実施形態において、特徴量に関連するインターバルの定義を示す説明図である。

【 図 3 】 第 1 の実施形態において、質疑応答に用いる英単語アプリの設問および解答入力画面の例を示す説明図である。

【 図 4 】 第 1 の実施形態における実験で取得したデータセットについて、ユーザ依存の方式の識別器作成に用いるデータと、その識別器の判定精度の評価に用いるデータを示す説明図である。

20

【 図 5 】 第 1 の実施形態における実験で取得したデータセットについて、ユーザ非依存の方式の識別器作成に用いるデータと、その識別器の判定精度の評価に用いるデータを示す説明図である。

【 図 6 】 第 1 の実施形態における実験において、ユーザ依存の方式における確信の有無の判定精度を示すグラフである。

【 図 7 】 第 1 の実施形態における実験において、ユーザ非依存の方式における確信の有無の判定精度を示すグラフである。

【 図 8 】 第 2 の実施形態において、実験に用いた語彙テストの設問画面の例を示す説明図である。

30

【 図 9 】 第 2 の実施形態において、機器別期の性能を表す再現率-適合率曲線 ( recall-precision 曲線 ) を示すグラフである。

【 図 1 0 】 第 2 の実施形態において、ユーザ非依存の方法における解答時の確信度と各特徴量との相関を示すグラフである。

【 図 1 1 】 第 2 の実施形態において、学習サンプルの数と判定精度との関係を示すグラフである。

【 図 1 2 】 第 2 の実施形態において、各分類の典型例を示す説明図である。

## 【 発明を実施するための形態 】

## 【 0 0 1 3 】

以下、図面を用いてこの発明をさらに詳述する。なお、以下の説明は、すべての点で例示であって、この発明を限定するものと解されるべきではない。

40

第 1 の実施形態では、解答がタイピング入力される場合について述べる。タイピング情報を取得するためにキーログを用いる。キーログは、入力したキーやキーを入力した時間などを記録することができるソフトウェアである。キーログは、スマートフォン等の情報処理装置に搭載されたハードウェアにアプリをインストールするだけで実現可能であり、さらに、設問解答時のユーザ操作の妨げにならないという利点がある。

## 【 0 0 1 4 】

第 1 の実施形態では、英単語アプリにおける設問解答時の確信を判定し、解答の正誤と合わせて提示することで、復習をより効果的に進めることのできる英単語学習システムについて述べる。ただし、英単語アプリは、この発明の一態様に過ぎない。所定文字数をタ

50

イピング入力することにより設問に解答するものであれば、この発明の適用範囲といえる。

また、第1の実施形態では、解答時の確信判定に用いる、解答者の挙動の特徴を表す情報として、ユーザのタイピング情報のみを収集している。他の手段を用いて解答者の挙動を検出することはしていない。

タイピング情報には人の心理状態に関する情報が含まれる。解答に確信が持てない場合、迷いが生じるため、タイピング速度が遅くなり、タイプする文字数も多くなる。したがって、適切なタイピング情報を取得して分析すれば、解答時の確信の有無を判定できる。

【0015】

第2の実施形態では、解答が手書き入力される場合について述べる。

10

手書き入力を扱う理由は2つある。(1)手書きログは、タブレットとスタイラスペンのみを必要とするに過ぎない。ユーザは追加の機器を着用する必要がない。その結果、通常の学習環境とほぼ同じデータを記録することができる。(2)多くの生徒にとって、手書きで書くのが最も身近な方法である。したがって、このシステムは小学校から大学までのすべての学生にとって使いやすいものである。

手書きの入力操作は質問にも依存するため、特徴量は質問に依存しないようにする必要があり。ストロークレベルの手書き特徴量は、質問に依存しないように慎重に選択すべきである。

【0016】

(第1の実施形態)

20

この実施形態では、キーログから取得したタイピング情報を用いて、英単語タイピング設問解答時の確信を判定する手法を述べる。

この実施形態における処理の流れを図1に示す。まず、キーログからタイピング情報を取得し(ステップS11)、タイピング情報から特徴量を抽出する(ステップS13)。次に、抽出した特徴量から有効なものを選択する(ステップS13)。サポートベクトルマシン(SVM、Support Vector Machine)の手法に基づいて、選択された特徴量を用いた機械学習を行って識別器(入力されたデータの判定(クラス分類)を行う主体、分類器ともいう)を作成し、作成された識別器を用いて認識を行う(ステップS17)。SVMは、機械学習によるパターン認識の一手法として知られている。なお、SVMによる識別器は、入力されたデータを2つのクラスに分類するので、この明細書では認識と呼ばず判定と呼ぶこととする。

30

【0017】

タイピング情報の取得

スマートフォンアプリを用いてタイピング情報を取得する。取得されるタイピング情報は、具体的には、タップした箇所を中心座標といった位置に関する情報や、図2のように、情報処理装置であるタブレット端末の画面をタップしてから指を離すタップ終了までの時間、指を離してから次にタップするまでの時間といった時間に関する情報である。それぞれの設問への解答についてタイピング情報を取得する。

次に、取得されたタイピング情報に含まれる座標から、どのキーをタップしたのかを特定し、キーに対応する文字情報を含んだタイピング情報に変換する。キーが配置されていない座標については、どのキーにも対応していないので、タイピング情報には含めないように設定する。

40

【0018】

特徴量の抽出

取得されたタイピング情報から、設問ごとの特徴量を抽出する。抽出する特徴量を表1に示す。

【表 1】

No	特徴量
1	解答時間 [ $\mu$ s]
2	インターバルの平均 [ $\mu$ s]
3	インターバルの標準偏差
4	インターバルの最大値 [ $\mu$ s]
5	インターバルの最小値 [ $\mu$ s]
6	インターバルの中央値 [ $\mu$ s]
7	最初のインターバル [ $\mu$ s]
8	最終のインターバル [ $\mu$ s]
9	タイプした文字数
10	デリートキーをタイプした回数
11	正解単語の文字数
12	正解単語の使用頻度
13	編集距離

10

20

## 【0019】

表 1 に示す特徴量は、上から順に、タイピングの時間に関連するものが 8 個 (No.1-8)、タイピングのキー情報に関連するものが 2 個 (No.9,10) である。それらのタイピング情報に基づく特徴量に加えて、正解の単語に関するものが 2 個 (No.11,12)、タイピングのキー情報と正解の単語の何れにも関連するものが 1 個 (No.13) の合計 13 個である。

「解答時間」は、前の設問に対する解答完了の操作であるエンターキーのタップから、次の設問に対する解答完了の操作であるエンターキーのタップまでの時間間隔である。「インターバル」は、図 2 に示す通り、あるキーをタップして文字入力 (タイプ) してから次のキーをタップして文字入力するまでの時間間隔である。「最初のインターバル」は、設問が提示されてからキーをタップして最初の文字を入力するまでにかかった時間間隔である。「最終のインターバル」は、解答の最後の文字をタイプして文字入力してから、エンターキーをタップして解答を完了するまでの時間間隔である。

30

## 【0020】

表 1 の No.9、「タイプした文字数」は、設問が提示されてからエンターキーをタップして解答を完了するまでにタイプした文字数であって、No.10の「デリートキーをタイプした回数」、No.11の「正解単語の文字数」と関連する。タイプ後にデリートキーで削除された文字の数も「タイプした文字数」に含まれる。

表 1 の No.12、「正解単語の使用頻度」は、その単語が一般の文書で使用される頻度である。コーパスに基づいて、英単語を使用頻度順に表示する WORDCOUNT という Web サイトのデータを参照し、正解単語に係る順位を用いた。

40

表 1 の No.13、「編集距離」は、タイプした文字列と正解単語がどれ程異なるのかを示す指標であって、最初に考えた人物の名前にちなんでレーベンシュタイン距離とも呼ばれる。具体的には、文字の挿入、削除、置換といった操作の数のことである。

## 【0021】

## 有効な特徴量の選択

各設問への解答について抽出された特徴量について、有効な特徴量の選択を行う。

この実施形態ではすべての特徴量の組合せについて探索を行う。13 個の特徴量があるので、すべての特徴量の組合せは  $(2^13 - 1)$  通りある。特徴量の各組合せに対して SVM を用いて識別境界を設定し、識別境界の設定に用いられた特徴量を得る。識別境界に

50



係る特徴量の数が最も少なく、かつ判定精度が最も高い特徴量の組合せを選択する。

【 0 0 2 2 】

識別境界の設定

各設問に対する解答について選択された特徴量を用いて、SVMを適用して識別境界の設定を行い、各設問への解答を表2に示すように、以下の2つのクラスに分類する。解答が正解であるという確信が持てたクラスと、解答が正解であるという確信が持てなかったクラスである。両者の何れかに各設問への解答を分類する。即ち、確信の有無を判定する。

【表2】

ラベル	クラス
0	解答が正解であるという確信が持てた
1	解答が正解であるという確信が持てなかった

10

なお、この実施形態ではSVMを用いて確信の有無に係るクラス分類を行うが、SVM以外の手法を用いてクラス分類を行ってもよい。

【 0 0 2 3 】

実験例

この実施形態が確信判定に有効であるかを検証するための実験を行った。実験の手順について述べ、実験によって得られた確信判定の精度について検証を行う。

1) 実験条件

実験参加者の日本人学生12名に図3に示す英単語アプリ画面のような形式の設問に解答するよう指示し、その際のタイピング情報を記録した。設問としては、通信教育用の英単語アプリとして知られている株式会社リクルートマーケティングパートナーズ社の「スタディサプリ 英単語」のライティング問題の中から12セクション(各セクションは10個の設問からなる)、合計120問を使用した。このとき単語の難易度には偏りが出ないように選択した。

20

【 0 0 2 4 】

2) 実験手順

実験手順について述べる。まず、Auto Touchというキーロガのアプリがインストールされ起動されたスマートフォンを用いて、1セクション10問の設問に解答するよう各実験参加者に指示した。解答が終わった後に、各設問への解答について確信が持てたかどうかのアンケートを各実験参加者に対して行った。この手順を12セクションすべてに対して行った。なお、解答するセクションの順番は実験参加者ごとにランダムになるように設定した。各実験参加者が各解答についてアンケートで確信の有無について答えた内容が正解であるとする。

30

【 0 0 2 5 】

3) 評価方法

評価方法について述べる。この実施形態では、2つのやり方について検証を行った。

第1はユーザ依存での検証である。ここで、ユーザ依存とはSVMの機械学習に際して、特定のユーザのデータのみを用いて機械学習を行って識別器を作成し、作成した識別器を用いて同一ユーザに対して確信判定を行うやり方をいう。ユーザ依存で学習された識別器を他のユーザに用いることはできない。ユーザ依存での学習を行うことで、そのユーザが解答する際に確信がある場合あるいは確信がない場合にどのような振舞いを行うのかを学習することができる。その振舞いの差が識別できれば、確信判定がより容易になる。また、すべての実験参加者について、各特徴量の選択回数を計算することで、ユーザの個人差に依らず、どの特徴量が判定に寄与しているのかを求めることができる。

40

評価方法については、図4に示すように、各実験参加者について、1セクション分のデータ10問をテストデータとし、残りの11セクションを学習データとして分割する。そのうえで、第1から第12のすべてのセクションについて、各セクションをテストデータとした12通りの場合について評価を行うLeave-one-section-out Cross-validationの手

50

法を適用した。

【 0 0 2 6 】

第 2 はユーザ非依存での検証である。ここで、ユーザ非依存とは、SVM の機械学習に際して 1 以上のある実験参加者のデータを用いて識別器を作成し、機械学習にデータを使用した実験参加者と異なる実験参加者のデータをテストデータとして識別器に判定を行わせるやり方である。ユーザ非依存で確信の有無の判定が可能であれば、まだ一度も設問に解答したことのない新しいユーザについて確信の有無を判定する場合にも、本実験の実験参加者のデータを用いて作成した識別器を用いることができる。ユーザ非依存での判定精度が高ければ、新たなユーザについて確信判定を行う場合に改めて機械学習を行わずとも、学習済の識別器を適用できる。したがって、ユーザ非依存の方式はユーザ依存の方式と比較して実用上より簡便な手法といえる。

10

ユーザ非依存の方式の評価方法については、図 5 のように、12 名の実験参加者 A ~ L のうち 1 名分のデータ 120 問をテストデータとし、残り 11 名の実験参加者のデータを学習データとする。そのうえで、第 1 から第 12 のすべてのユーザについて、各ユーザをテストデータとして 12 通りの場合について評価を行う Leave-one-user-out Cross-validation の手法を適用した。

【 0 0 2 7 】

ユーザ依存と、ユーザ非依存の 2 つの方式を適用した実験結果を、正誤情報のみに基づく判定精度と比較する。ここで、正誤情報のみに基づく判定精度とは、正解した設問に係る解答は、「解答が正解であるという確信が持てた」クラスに分類し、間違えた設問に係る解答は、「解答が正解であるという確信が持てなかった」クラスに分類した場合の判定精度である。

20

【 0 0 2 8 】

4) 実験結果とその分析

12 名の実験参加者の事前確率を表 3 に示す。ここで、事前確率とは、アンケートにおいて各実験参加者が設問への解答について確信が持てたと答えた割合、即ち、各実験参加者の解答のうちで「確信あり」と判定されるべきものの割合である。

【表 3】

実験参加者	事前確率 [%]
A	65.8
B	63.3
C	43.3
D	32.5
E	38.3
F	56.7
G	67
H	50
I	41.7
J	65
K	35.8
L	35.8
平均	49.6

30

40

【 0 0 2 9 】

事前確率の平均値が 50 % 前後であることから、2 つのクラス（「確信あり」と判定さ

50

れるべき解答のグループと「確信なし」と判定されるべき解答のグループ)のデータ数には偏りが無いことが確認された。

この実施形態によるユーザ依存の方式の判定精度と正誤情報のみに基づく判定精度を比較した結果を図6に示す。図6で、ユーザ依存の方式の判定精度の値は、すべての解答のうちで、ユーザ依存の方式で確信の有無を判定した結果と実験参加者がアンケートに回答した確信の有無が一致した割合を示す。正誤情報のみに基づく判定精度は、すべての解答のうちで、正解の解答について実験参加者がアンケートに確信ありと答えた割合と不正解の解答について実験参加者がアンケートに確信なしと答えた割合を足したものの割合を示す。

平均の判定精度は、正誤情報のみに基づく場合は86.7%、この実施形態によるユーザ依存の方式は91.2%であった。

#### 【0030】

実験参加者全体に見られた傾向としては、正解した設問については確信が持てたと解答し、間違えた設問については確信が持てなかったと解答したケースが多かった。そのため、正誤情報のみに基づく場合でもほとんどの実験参加者で高い判定精度が得られた。一方、すべての実験参加者について、この実施形態による判定精度が、正誤情報のみに基づく判定精度を上回っている。

実験参加者ごとで見られた特徴として、実験参加者F, G, H, Kについては、ユーザ依存の方式による判定精度も、正誤情報のみに基づく判定精度も、ともに高い値を示す(誤判定が少ない)ことが確認された。これらの実験参加者では、少し考えて分からない設問はスキップする傾向が認められた。またその場合には、スキップしていない設問については確信あり、スキップした設問については確信なしとするケースが多いことも確認できた。一方で、実験参加者A, B, E, J, Lについては、ユーザ依存の方式の判定精度が正誤情報のみに基づく判定精度を大きく上回った。これらの実験参加者では、分からない設問は時間をかけて思い出そうとする傾向が確認された。またその場合には、解答時間はもちろん、タイピング速度も遅くなることも確認できた。

#### 【0031】

実験参加者Dの判定精度が、ユーザ依存の方式、正誤情報のみに基づく場合の両方で低い(誤判定が多い)理由について述べる。実験参加者Dは、ある1つのセクションにおいて、確信が持てずに正解した設問と、確信を持ちつつ間違えた設問の数が他のセクションと比べて極端に多かったということが確認された。Leave-one-section-out Cross-validationを行った場合、これらのパターンの学習データが少なかったために誤認識が起きやすくなったと考えられる。

ユーザ依存の方式における各特徴量の選択回数を表4に示す。ここで、特徴量の選択回数は、以下のことを意味する。ユーザ依存の評価方法で述べたように12名の実験参加者のそれぞれについて(即ち12通りの場合について)SVMの機械学習を行って識別器を作成する。それぞれの識別器において識別境界の設定に用いられた特徴量を、その識別器において選択された特徴量として+1カウントする。例えば、12通りのすべてについて選択された特徴量があるとすれば、その選択回数は12であり、何れの識別器においても選択されなかった特徴量があるとすれば、その選択回数は0(ゼロ)である。

#### 【0032】

10

20

30

40

【表 4】

選択回数	特徴量
11	編集距離
6	解答時間 正解単語の文字数
5	インターバルの平均 最終のインターバル タイプした文字数
4	インターバルの標準偏差 インターバルの最大値
3	インターバルの最小値 最初のインターバル 正解単語の使用頻度
2	インターバルの中央値
1	デリートキーをタイプした回数

10

20

## 【0033】

実験参加者は12人なので、選択回数の最大値は12回である。最も多く選択された特徴量は編集距離であった。次に多く選択されたのは解答時間、正解単語の文字数であった。3番目に多く選択されたのはインターバルの平均、最終のインターバル、タイプした文字数であった。編集距離は解答の正誤情報も含んでいるため、選択回数が多くなったと考える。正解単語の文字数は単語の難易度に関連する情報であり、文字数が多い単語ほど暗記するのが難しいという傾向がある。解答時間、インターバルの平均、最終のインターバル、タイプした文字数は、解答に確信が持てない場合には、値が大きくなるため、選択回数が多くなったと考えられる。

30

## 【0034】

この実施形態による確信判定手法の有意性を検証するために、実験結果と事前確率のみに基づく場合の判定精度、実験結果と正誤情報のみに基づく場合の判定精度について、それぞれ符号検定を行った。その結果を表5に示す。

前述のように、事前確率とは、解答のうちで「確信あり」と判定されるべきものの割合、即ち「確信あり」が正解である確率である。

表5の中央の欄は、実験結果の判定精度が事前確率のみに基づく判定精度に対して有意な差を持つことを符号検定により検証した結果を示している。ここで、事前確率のみに基づく判定精度は以下のようにして判定を行った場合に判定が当たる確率である。事前確率が50%以上の場合は、すべての設問に対して「確信あり」と判定するものとする。一方、事前確率が50%未満の場合はすべての設問に対して「確信なし」と判定するものとする。そうした場合の判定精度は結局、事前確率が50%以上の場合は事前確率に等しく、事前確率が50%未満の場合は(100 - 事前確率)に等しいパーセンテージになる。

40

また、表5の右欄は、実験結果の判定精度が正誤情報のみに基づく判定精度に対して有意な差を持つことを符号検定により検証した結果を示している。

## 【0035】

なお、表5においては、符号検定で多用される5%および1%の基準の何れかを満たすか、あるいは満たさないかを示している。\*は有意水準5%の基準を満たすことを示しており、さらに、\*\*は有意水準1%の基準を満たすことを示している。空白は、いずれの有意水準についてもその基準を満たさないことを示している。有意水準1%の基準を満た

50

すとは、この実施形態による確信判定手法が、事前確率のみに基づく場合あるいは正誤情報のみに基づく場合の判定手法と変わらない性能である（正しく判定する確率が同等）にもかかわらず、両者に差があると誤って判断する確率が1%未満であることを示す。

【0036】

【表5】

実験参加者	事前確率	正誤情報
A	**	
B	**	**
C	**	
D	*	
E	**	
F	**	
G	**	
H	**	
I	*	
J	**	
K	*	
L	**	
全体	**	**

10

20

【0037】

表5に示すように、実験結果と事前確率との間には十分な差がみられた。しかし、ユーザ依存の方式による判定精度と正誤情報のみに基づく場合の判定精度との間には全体としては十分な差があるものの、各実験参加者については実験参加者B以外での十分な差が認められなかった。このことから、12セクション分のデータでは有意差を検出するには不十分なのではないかと推測する。したがって、各実験参加者に解答してもらうセクション数を増やすことで、有意性が得られるようになるものと推測する。

30

ユーザ非依存の方式の判定精度の実験結果を図7に示す。この実施形態の平均判定精度は89.8%であった。表6に示す通り、これはタイピング情報を用いなかった他の4つの手法を上回る精度であった。他の4つの手法のうち第1の手法は、正誤情報のみに基づく確信判定である。第2の手法は、正誤情報に解答時間の特徴量（表1に示す特徴量のNo.1）を加えた確信判定である。第3の手法は、正誤情報に正解単語の情報の特徴量（表1に示す特徴量のNo.11および12）を加えた確信判定である。第4の手法は、正誤情報に解答時間および正解単語の情報の特徴量を加えた確信判定である。前記第2および第4の手法における解答時間の取得だけならば、あえてキーログを用いなくてもよい。例えば、英単語アプリの側で解答時間を測定できる。

40

【0038】

【表 6】

手法	推定精度
提案手法	<b>89.8%</b>
正誤情報	86.7%
正誤情報+解答時間	86.6%
正誤情報+正解単語の情報	86.5%
正誤情報+解答時間+正解単語の情報	86.7%

10

ユーザ非依存の方式に選択された特徴量は、表 7 のとおりである。特徴量として、解答時間、インターバルの平均、インターバルの最大値、インターバルの中央値、最初のインターバル、最終のインターバル、正解の単語の文字数、正解の単語の使用頻度および編集距離が選択された。

【 0 0 3 9】

【表 7】

特徴量
解答時間
インターバルの平均
インターバルの最大値
インターバルの中央値
最初のインターバル
最終のインターバル
正解単語の文字数
正解単語の使用頻度
編集距離

20

30

【 0 0 4 0】

一般的に、ユーザ非依存の方式では、ユーザ固有の振舞いは反映されないもので、判定精度が下がり、誤判定が多くなる傾向がある。本実験でも、多くのユーザに対してそのような傾向がみられた。

特に、確信がない場合の振舞いは学習が難しく、ユーザ依存の方式と比べて誤判定が起きやすくなることが分かった。なぜなら、確信がある場合は、頭に浮かんでいる文字列をそのままタイプしていくだけで、その振舞いには大きな個人差はないが、確信がない場合の振舞いは人それぞれである。具体的には、キーを押さずに頭の中でじっくり考えて解答する人、キーを押して自身のタイプした文字列を見ながら解答を探す人、あるいはスキップする人など、様々であったからである。

40

【 0 0 4 1】

その中でも、実験参加者 A , B , C については、ユーザ依存の方式と同等の判定精度を得ることができた。この理由については、A , B , C においてユーザ依存の方式で選択された特徴量の組合せが似通っていることから、この 3 者のタイピング作業における振舞いが似ていたということが推測される。

また、実験参加者 G についても、ユーザ依存の方式と同等の判定精度を得ることができた。これは、A , B , C の場合とは異なり、この実験参加者は確信が持てない問題はほとんどスキップしていたので、得られたタイピングに関する情報は確信が持てた問題に限られていた。確信が持てた問題はユーザ非依存の方式でも判定が容易であったので、判定が

50

上手くいったと考えられる。

#### 【0042】

実験参加者Eについて、ユーザ依存の方式と比較して判定精度が著しく下がった理由について考察する。この実験参加者については、デリートキーをタイプした回数が確信判定に寄与していることが確認されたが、ユーザ非依存の方式にはこの特徴量は選択されなかった。したがって、ユーザ非依存で学習を行った際に、ユーザ依存で学習を行った場合に比べて確信がない場合の振舞いを学習しきれていないことが確認された。このことから、この実験参加者は確信がない場合の振舞いが特殊であるということが推測される。Leave-one-user-out Cross-validationを行った場合、確信がない場合の振舞いが他の実験参加者の振舞いと合致しなかったため、誤判定が起きやすくなったと考える。

10

このように、ユーザ非依存の方式は、ユーザ依存の方式に程高い判定精度が得られないが、それでも上述したように、タイピング情報を用いなかった他の4つの手法を上回る判定精度が得られた(表6参照)。

#### 【0043】

##### 実験のまとめ

以上のように、この実施形態では、確信を考慮した英単語学習システムを実現するために、キーロガを用いたタイピング情報の取得という簡便な構成で、解答者の挙動に係る特徴量を取得し、設問解答時の確信の有無を前記特徴量に基づいて判定する手法について述べた。キーロガから取得したタイピング情報から解答者の挙動に係る特徴量を抽出し、正解の属性に係る特徴量と併せて、SVMを用いて確信の有無を判定(即ち、確信ありと確信なしの2クラスに分類)した。

20

実験の結果、この実施形態の平均判定精度は、各実験参加者のデータセットについて特徴量の選択および判定を行う方式(ユーザ依存の方式)には、91.2%、データセット全体について特徴量の選択および判定を行う方式(ユーザ非依存の方式)では89.8%であった。

#### 【0044】

この実施形態によれば、スマートフォン等の情報処理装置にキーロガのアプリをインストールするだけで簡便にユーザの挙動の特徴を取得でき、膨大なデータ量の機械学習を行わずとも精度よく確信判定を行えるという利点がある。

この実施形態では、英単語タイピングに関する設問への解答を例に挙げてこの発明による確信判定方法を述べた。これは一例であって、設問に対する解答をタイピング入力させるものであれば、言語を問わず、あるいは単語に限らず、この発明による手法が適用できる可能性がある。

30

#### 【0045】

また、この実施形態で得られた実験結果は編集距離に大きく依存した結果となった。英作文などでは、解答の多様性から、明確な解答が提示されていない場合が多いため、編集距離や正誤情報を用いずに確信の有無を判定する手法を考案する必要がある。具体的には、まず、タップしてから指を離すまでの時間、指を離してから次にタップするまでの時間を加えることが考えられる。それに加えてキーをタップした際にディスプレイにかかった圧力なども特徴量として導入することが考えられる。確信がある場合には、キーをしっかりとタップすることが推測されるからである。ただし、Auto Touchには現状、そのような機能は実装されていないため、より高機能のキーロガを用意する必要がある。

40

#### 【0046】

##### 確信判定に基づく復習プランの生成

この実施形態で述べたユーザ依存の方式あるいはユーザ非依存の方式による識別器を用いて確信判定を行った結果に基づいて、質疑応答の復習プランを生成する処理について述べる。

例えば、前述した英単語アプリを実行するスマートフォンのCPU(Central Processing Unit)は、ユーザの解答の正誤と、解答時の確信判定の結果を履歴としてメモリーに格納しておく。

50

そして、ユーザが少なくとも一度解答した設問を、将来再び出題すべきか否かを判断する。即ち、復習すべき英単語を選択する。

【0047】

ユーザが間違えた英単語は、復習すべき英単語に分類し、復習リストに登録する。ただし、それらについては、確信判定を参照することなく判断できる。

この実施形態では、確信なしと判定されたが正解した英単語についても、復習リストに登録する。

さらに、確信ありと判定されたが間違えた英単語は、確信なしと判定されて間違えた英単語よりも復習の出題頻度が高くなるように復習リストに登録する。

前記CPUは、実行中の英単語アプリの次の出題の機会に、あるいはその英単語アプリが次回起動された機会に、前述の復習リストを参照し、復習リストに登録された英単語を再度出題する。

確信ありと判定されて正解した英単語であっても、間隔をおいて繰り返し出題してもよいが、復習リストに登録された英単語はそれより短い間隔で、あるいはそれよりも高い頻度で再出題されるように前記CPUは出題のプランを策定する。

【0048】

なお、確信なしと判定されたが正解した英単語については、再出題の際にユーザが正確に記憶していない可能性があることを知らせるようにしてもよい。

確信ありと判定されたが間違えた英単語についても、再出題の際にユーザが勘違いしている可能性があることを知らせるようにしてもよい。

スマートフォンを用いた英単語アプリは単なる一例であって、上述の出題選択手法は、情報処理装置を用いて解答をタイピング入力させる質疑応答において、再出題プランを決定する場合に拡張して適用可能である。

【0049】

(第2の実施形態)

以下、設問解答時の確信度判定に係る第2の実施形態を構成する、(1)データの記録、(2)特徴量の抽出、(3)特徴量の選択、(4)分類の4つの処理について簡単に述べ、さらに、それらの処理に基づいて行った実験例について述べる。

記録

タブレットとスタイラスペンを用いて、ストロークレベルの手書きの記録を行う。タブレットは、入力操作(押し、放し、移動)、タイムスタンプ、x-y座標、デバイスの種類、筆圧、および各イベントの消去の有無を記録する。

【0050】

特徴量の抽出

手書き入力操作は、確信度だけでなく質問やユーザの書き込みの全体の長さによっても変わることが想定している。これらの要因を取り除くために、質問や書き込み長さに依存しないように特徴量を慎重に選択する。

最後に、提案手法は応答時間、書き込み間隔、書き込み速度、書き込み圧力、消去率について5つのカテゴリ(解答時間、インターバル即ち時間間隔、速度、筆圧、消去)の特徴を抽出する。抽出された各特徴量の詳細を表8に示す。

表8に示すように、解答時間に係る特徴量は、設問が提示されてから解答完了の操作までの時間を手書き入力の総ストローク長で正規化した解答時間(f1)である。

【0051】

各ストローク間の時間間隔(インターバル)に係る特徴量は、その平均(f2)、標準偏差(f3)、最大値(f4)、最小値(f5)、中央値(f6)を含む。さらに、設問が提示されてから最初の文字の最初のストロークを入力し始めるまでの時間(f7)を含む。さらにまた、最後の文字の最後のストロークを入力してから解答完了の操作までの時間(f8)を含む。また、前記f7に係る特徴量を解答時間で除したもの(f9)、および前記f8に係る特徴量を解答時間で除したもの(f10)を含む。さらにまた、各ストローク間の時間間隔の合計を解答時間で除したもの(f11)を含む。



## 【 0 0 5 2 】

速度に係る特徴量は、各ストロークを書き込む速度の平均（f12）、標準偏差（f13）、最大値（f14）、最小値（f15）、中央値（f16）を含む。

筆圧に係る特徴量は、各ストロークを書き込む筆圧の平均（f17）、標準偏差（f18）、最大値（f19）、最小値（f20）、中央値（f21）を含む。

消去に係る特徴量は、各設問への解答中に行ったデリート操作の回数を、その設問に対するストロークの数で除したもの（f22）である。

【表 8】

カテゴリ	No	特徴量
解答時間 インターバル	f1	解答時間／総ストローク長
	f2	インターバルの平均
	f3	インターバルの標準偏差
	f4	インターバルの最大値
	f5	インターバルの最小値
	f6	インターバルの中央値
	f7	最初のインターバル
	f8	最終のインターバル
	f9	最初のインターバル／解答時間
	f10	最終のインターバル／解答時間
	f11	インターバルの合計／解答時間
速度	f12	速度の平均
	f13	速度の標準偏差
	f14	速度の最大値
	f15	速度の最小値
	f16	速度の中央値
	筆圧	f17
f18		筆圧の標準偏差
f19		筆圧の最大値
f20		筆圧の最小値
f21		筆圧の中央値
消去	f22	デリートの回数／ストロークの数

10

20

30

40

この実施形態において、1ストロークとは、ユーザがペンで画面に触れてからペンが画面から外れるまでの一連の入力操作をいう。書き込み間隔は、各ストローク間の待ち時間を表す。書き込み速度は各ストロークの平均速度で計算される。

## 【 0 0 5 3 】

## 特徴量の選択

特徴量の選択方法として前方段階的選択（forward stepwise selection）を用いる。データセットは5つの部分に分割され、特徴を追加することで交差検証（cross-validation）を行い、平均の判定精度を計算する。次に、すべての平均判定精度の中で最も高い判定精度を記録した特徴セットを選択する。このステップはトレーニング中にのみ行った。予め選択された特徴量は未知のサンプルを分類するために使用される。

50

## 【 0 0 5 4 】

## 分類

各サンプルを2つのクラス、すなわち高確信度または低確信度に分類するために、RBFカーネルおよびハイパーパラメータ ( $C=1.0$ ,  $\gamma=0.045$ ) を有するサポートベクターマシン (SVM) を適用する。ハイパーパラメータは、学習アルゴリズムの挙動を制御するパラメータとして知られているものである。

## 【 0 0 5 5 】

## 実験例

この実施形態に係る確信度判定法を評価するために、以下の実験を行った。

## (参加者と課題)

10

参加者はドイツで勉強している11人の大学生であった。参加者のうち、3人は日本からの男性、7人は中国からの男性、2人は中国からの女性であった。なお、他の国からの学生を採用しなかったのは、彼らの母国語で設問等のフォーマットを準備することができなかったためである。

設問は、単語の訳語を書くもののみである。この種の質問はすぐに解答できるため、参加者ごとに多数のサンプルを収集できる。参加者は、日本語または中国語の単語を見せられる。そして、その単語の英訳を図8のような指定された空白に小文字で埋めるように求められる。また、解答の文字がわかりやすいように、なるべく分けて記入してもらうように求めた。答えがわからない場合は、その問題をスキップできる。問題がスキップされた場合、その問題に係る手書きの入力操作は記録できない。

20

## 【 0 0 5 6 】

参加者は1時間でできるだけ多くの問題を解答し続ける。これらの単語は、国際コミュニケーション英語力テスト (TOEIC) で頻繁に使用される語彙の集まりから選択されている。各質問の文字数は3~12文字である。

各質問に解答した後、参加者は答えに対する確信度を報告する。彼らには次の3つの選択肢が与えられる：(1) Unconfident：綴りと意味の両方に対する確信度が低い、(2) Confident with meaning：綴りに対する確信度が低い、意味に対する確信度がある、(3) Confident with both meaning and spelling：綴りと意味の両方に対する確信度が高い。

30

我々は、確信度判定のグラウンドトゥールスレベル (ground truth label)、即ちここでは各選択肢に紐付ける確信度として、(1)と(2)の両方を低確信度とし、(3)を高確信度として扱った。綴りに対する確信度((1)と(2)の間の分類作業)を判定しなかったのは、(2)とラベル付けされた学習モデルのサンプルの数が他に比べて少なすぎて選択肢を細分化するのは適当でないと考えたためである。

## 【 0 0 5 7 】

## (検証)

選択された特徴を用いて、我々は2つの異なる検証方法を使用してこの実施形態に係る確信度判定法の性能を比較した。2つの異なる検証方法とは、ユーザ依存の方法とユーザ非依存の方法である。

ユーザ依存の方法では、あるユーザから収集されたサンプル(記録)のデータの一部をトレーニングに使用し、残りのサンプルを性能テストに使用する。

40

我々はユーザ依存の方法について、特徴量を選択したうえでそれぞれのユーザから収集されたデータを5つの質問のサブセットに分割し、k分割交差検証を行った。

ユーザ非依存の方法については、あるユーザから収集されたデータを性能テストに使用し、他のすべてのユーザから収集されたデータを学習に使用する。

我々はユーザ非依存の方法について、特徴量を選択したうえで参加者の1人を除外する相互検証を実行した。

## 【 0 0 5 8 】

学習データには、テストデータの同じ質問がいくつか含まれている。我々の方法が質問に依存しないことを証明するために、我々はテストデータと重複するそれらの質問を学習

50

データから取り除いた。この検証方法を使用することで、コールドスタート問題 (Xuan N hat Lam, Thuc Vu, Trong Duc Le, and Anh Duc Duong. 2008. Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, 208-211. 参照) を解決できる。

通常、各サンプルを高確信度または低確信度に分類する分類器の性能は、ユーザ依存の方法による分類器の方がユーザ非依存の方法による分類器よりも優れている。ユーザ依存の方法の方法では対象のユーザ自身の振る舞いを学習することができるからである。ユーザ依存の方法の欠点は、システムに学習をさせるために、ユーザが幾つかのデータにラベル付けをする必要があること、即ち、あるユーザがシステムを使用するためには、そのユーザをシステムに学習させる処理をユーザがまず行わなければならない点である。

【 0 0 5 9 】

ユーザ依存の方法とユーザ非依存の方法のどちらについても、性能を比較するために、質問の数で重み付けしたすべての参加者間の平均判定精度を計算した。

データセットは不均衡である (クラスに属するサンプル数に偏りがある) ので、balanced baggingアルゴリズムを使ったアンダーサンプリングを適用した (Shohei Hido, Hisashi Kashima, and Yutaka Takahashi. 2009. Roughly balanced bagging for imbalanced data. Statistical Analysis and Data Mining: The ASA Data Science Journal 2, 5-6 (2009), 412-426. 参照)。アンダーサンプリングの重要性を評価するために、アンダーサンプリングと非アンダーサンプリングの両方の性能を以下に述べる。

【 0 0 6 0 】

結果と考察  
(性能)

確信度判定の各クラスの正解率について、平均および標準偏差ならびに曲線下面積 (AUC) を表 9 に示す。正解率は、高確信度の解答を高確信度であると正しく判定したもの (True positiveのクラス) および低確信度の解答を低確信度であると正しく判定したもの (False negativeのクラス) の、全ての判定に占める割合である。全ての判定には、上記True positiveおよびFalse negativeのクラスの他に、False positiveのクラスおよびTrue negativeのクラスを含む。False positiveは、低確信度の解答を高確信度であると誤判定したものである。False negativeは、高確信度の解答を低確信度であると誤判定したものである。

【表 9】

学習	サンプリング	分類 (正解率)	検出 (AUC)	
			高確信度	低確信度
ベースライン		0.66 ± 0.10	0.66	0.34
ユーザ依存	不均衡	<b>0.80 ± 0.06</b>	0.90	0.64
	均衡	0.79 ± 0.04	<b>0.91</b>	<b>0.73</b>
ユーザ非依存	不均衡	0.74 ± 0.07	0.87	0.62
	均衡	0.74 ± 0.03	0.88	0.65

ベースラインは、多数派クラスの比率として定義される。ユーザ依存型の不均衡データセットは、すべての手法の中で最高の性能を達成した。しかし、そのデータセットは一部の参加者にとってはあまりにも不均衡である。そのため、図 9 に各手法と各クラスの再現率-適合率 曲線 (recall-precision 曲線) も示している。

図 9 に係る適合率 (precision) は、あるクラス (例えばTrue positive) に着目した場合に、そのクラスに分類されたサンプル k のうちで、正しく分類されたサンプル j の割合 (j / k) である。全てのサンプルが正しく分類されると j は k と等しくなり適合率は 1 になる。誤分類が生じると、j < k となり、誤分類が多くなる程 k が大きくなって適合率 (判定の精度) は小さくなる。全てのサンプルが誤分類されると j = 0 となり、適合率は

0 (ゼロ)になる。

また、再現率 (recall) は、あるクラス (例えば True positive のクラス) に着目した場合に、そのクラスに分類されるべきサンプル n のうち、正しく分類されたサンプル m の割合 (m / n) である。全てのサンプルが正しく分類されると m は n と等しくなり再現率は 1 になる。誤分類が生じると、m < n となり、誤分類が多くなる程 m、そして再現率は小さくなり、全てのサンプルが誤分類されると再現率は 0 (ゼロ)になる。

再現率-適合率 曲線およびその AUC は、機械学習の分類タスクにおける性能を示すためによく用いられる。テストサンプルを Positive である (関連が強い) と予想される順にランキングした場合に、実際に Positive なサンプルをどれだけ上位に集められたか、即ちランキングの正確さを表す指標として AUC が用いられる。

10

ユーザ依存およびユーザ非依存の 2 つの手法を比較すると、高確信度検出の AUC はそれほど変化しない。ただし、均衡のとれたトレーニングデータセットを使用すると、確信度の低い検出の AUC が向上する。均衡の取れたトレーニングデータセットを使用したほうがよいと思われる。

ユーザ非依存の方法では、性能はユーザに依存する方法と比較して劇的には変化しない。データセットは基本的に均衡が取れているので、均衡の取れたトレーニングと不均衡のトレーニングの間にそれほど大きな違いはない。

【 0 0 6 1 】

( 選択された特徴量 )

図 1 0 は、ユーザ非依存の方法における確信度と各特徴量との間の相関関係を示す。図 1 0 に示すように、最も強い負の相関を示す特徴量は f4 のインターバルの最大値である。それに続くのは、f1 の解答時間、f2 のインターバルの平均、f17 の筆圧の平均、f18 の筆圧の標準偏差、f20 の筆圧の最小値である。さらに、f3 のインターバルの標準偏差、f19 の筆圧の最大値、f6 のインターバルの中央値が続く。一方、f7 の最初のインターバルはほとんど相関を示さない。表 1 0 は、ユーザ非依存の方法において、確信度との相関があるものとして選択された特徴量を示している。

20

【表 1 0】

学習	サンプリング	分類 (正解率)	検出 (AUC)	
			高確信度	低確信度
ベースライン		0.66 ± 0.10	0.66	0.34
ユーザ依存	不均衡	<b>0.80 ± 0.06</b>	0.90	0.64
	均衡	0.79 ± 0.04	<b>0.91</b>	<b>0.73</b>
ユーザ非依存	不均衡	0.74 ± 0.07	0.87	0.62
	均衡	0.74 ± 0.03	0.88	0.65

30

数字「1」は選択された特徴を表し、「0」は選択されていない特徴を表す。書き込み間隔と確信度の間には弱い負の相関があることがわかった。言い換えれば、確信度が低いと書き込み間隔が長くなる。消去回数も同じ傾向である。それらの特徴量は確信度判定に有効である。スピードとプレッシャーは、確信度とあまり相関がない。しかし、それらのいくつかは選択されている。それらは、間隔または消去回数と強い相関関係がないので、間隔または消去回数と組み合わせると効果的に機能する可能性があると考えられる。

40

【 0 0 6 2 】

( 学習サンプル数 )

ユーザ非依存の方法では、適切な推定に必要なサンプル数を明確にするために、学習サンプルの数を調整することを試みた。結果を図 1 1 に示す。システムの性能は、サンプル数が増えるにつれて高くなる。その影響は私達が予想したより小さい。アンダーサンプリングと非アンダーサンプリングのどちらの方法でも正解率は 200 ~ 400 で安定しているので、200 ~ 400 で十分である。

【 0 0 6 3 】

50

## (誤分類)

この項目では、誤分類の典型的な例を紹介する。図12では、各例で時系列の書き込みストロークをプロットした。我々の方法では、「正」は高い確信度を意味し、「負」は低い確信度を意味する。横軸は応答時間で正規化されている。各図の下部に表示されている文字列は、参加者が書いた実際の単語である。ストローク数は文字数と等しくないことに注意されたい。

図12で、(a)はTrue positive、即ち正解は高確信度(True)で判定(予想)結果も高確信度(positive)の例である。(b)はFalse positive、即ち、正解は低確信度(False)であるにもかかわらず判定(予想)結果が高確信度の例である。(c)はFalse negative、即ち正解は低確信度(False)で判定(予想)結果も低確信度(negative)の例である。(d)はTrue negative、即ち正解は高確信度(True)であるにもかかわらず判定(予想)結果が低確信度(negative)の例である。

10

図12(b)に示す例が擬陽性に誤分類された理由は以下の通りである。ユーザは答えに対して低い確信度を感じたが、書き込み間隔が図5(a)の高い確信度パターンと同様になったので、識別器は予測に失敗した。

図12(d)のTrue negativeへの誤判定の例では、ユーザは問題を解く前にしばらく考えていたか、少し休憩を取ったために解答を始めるまでの間隔が長くなりすぎた。したがって、識別器はその質問を確信度が低いと予測した。

この種の誤分類は、我々の方法が書き込みの振る舞い、特に書き込み間隔に大きく依存しているために引き起こされた。

20

【0064】

## 結論

この実施形態では、ストロークレベルの手書き分析に基づく自己信頼判定法について述べた。この実施形態における確信度判定の判定精度は、ユーザ依存の方法で80%であり、ユーザ非依存の方法で74%である。

この実施形態に係る確信度判定方法は質問に依存しない特徴量のみを使用している点を強調したい。上述の実験結果から、この実施形態による確信度判定は、(1)書き込み間隔および消去回数との間に弱い確信度関係があること、(2)特徴量の組み合わせもまた、予測に重要であることが見出される。

## 確信判定に基づく復習プランの生成

30

この実施形態で述べた方式による識別器を用いて確信判定を行った結果に基づいて、第1の実施形態と同様に質疑応答の復習プランを生成することができる。

【0065】

以上に述べたように、

(i)この発明による確信判定方法は、設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定する方法であって、コンピュータが実行するステップとして、解答の入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備えることを特徴とする。

40

【0066】

ここで、解答者が入力する文字は、数字、記号を含んでもよい。

この発明において、設問は、解答の文字入力を受付ける情報処理装置を用いて提示されてもよいが、そうでなくてもよい。他の情報処理装置や、人によって提示されてもよい。その場合、解答を受付ける情報処理装置は設問が解答者に提示されたタイミングを取得できるものとする。

【0067】

タイピング情報を取得するステップその具体的な手段として、キーログが挙げられる。

この発明による確信判定方法は、典型的にはスマートフォンやタブレット端末等の情報

50

処理装置にアプリ（アプリケーション）としてインストールされ、情報処理装置のコンピュータにより実行される。処理は、情報処理装置が有するメモリー、表示デバイス、入力デバイス等のハードウェア資源を用いて実行される。

【0068】

さらに、この発明の好ましい態様について説明する。

(ii) 前記特徴量が、解答に要した時間、タイピングまたは手書きによる入力操作の時間間隔または消去の操作に係るものであってもよい。

これらの特徴量は、解答者の確信の有無と相関があるとので、好適な特徴量を選択することによって解答者の確信の有無を判定できる。

【0069】

(iii) 前記特徴量が、最後の文字のタイプ入力または最後のストロークの手書き入力から解答完了の操作までの時間間隔を含んでいてもよい。

解答に確信が持てない場合、解答者は最後の文字またはストロークを入力した後も迷いが残り、最後の文字またはストロークの入力から解答完了の操作までの時間間隔が大きくなることが考えられる。このようにすれば、解答者の迷いを特徴量として取得できる可能性がある。

【0070】

(iv) 前記特徴量が、設問が提示されてから最初の文字のタイプ入力または最初のストロークの手書き入力までの時間間隔を含んでいてもよい。

解答に確信が持てない場合、設問が提示されてから最初の文字またはストロークを入力するまでの時間間隔が大きくなることが考えられる。即ち、設問を見て戸惑いが生じた度合いに、最初の文字またはストロークを入力するまでの時間間隔が大きくなることが考えられる。このようにすれば、解答者の戸惑いを特徴量として取得できる可能性がある。

【0071】

(v) 前記特徴量が、タイプ入力における文字入力間または手書き入力におけるストローク入力間のインターバルの最大値を含んでいてもよい。

解答に確信が持てない場合、解答の途中で文字またはストロークを入力するまでの時間間隔が大きくなることが考えられる。インターバルの最大値は、どれだけ長いあいだ迷ったかに関連することが考えられる。このようにすれば、解答者の戸惑いを特徴量として取得できる可能性がある。

【0072】

(vi) 前記特徴量が、設問が提示されてから解答完了の操作までに要した時間に基づく値、タイプ入力における文字入力間または手書き入力におけるストローク入力間のインターバルの平均、前記インターバルの標準偏差、前記インターバルの最小値、前記インターバルの中央値の少なくとも何れかを含んでいてもよい。

解答に迷うと、設問が提示されてから解答者が解答完了の操作を行うまでの時間間隔が大きくなることが考えられる。また、文字を入力する時間間隔の平均、標準偏差、最小値あるいは中央値が大きくなることが考えられる。時間間隔の標準偏差は、各文字の入力の時間間隔にどの程度ばらつきがあるかを示し、どの程度迷いながら文字を入力したかに関連することが考えられる。このようにすれば、解答者の迷いを特徴量として取得できる可能性がある。

【0073】

前記特徴量が、タイプした文字数を含んでいてもよい。

タイプした文字数が多いと、解答が複雑であるかまたは修正した文字数が多いことが考えられる。このようにすれば、解答者の迷いを特徴量として取得できる可能性がある。

【0074】

設問が単語を構成する文字を入力させるものである場合、前記特徴量が、編集距離、正解単語の文字数、正解単語が文書に出現する使用頻度の少なくとも何れかを含んでいてもよい。

解答単語と正解単語の不一致の度合い、正解単語の複雑さ、正解単語を目にする機会の

10

20

30

40

50

希少さは、正解の難しさの度合いに関連すると考えられる。このようにすれば、確信のなさを正解の難しさに関連する特徴量として取得できる可能性がある。

【 0 0 7 5 】

(vii) 確信の有無の判定がサポートベクトルマシンの手法に基づき、選択された特徴量を用いて、確信の有無を判定してもよい。

このようにすれば、簡易な学習で優れた認識性能が得られることで知られるサポートベクトルマシンの手法に基づいて、タイピング情報から抽出された特徴量を用いて確信度を判定する識別器を作成できる。

【 0 0 7 6 】

(viii) また、この発明による確信判定プログラムは、設問に対して解答者にタイピングまたは手書きにより文字を入力させて解答を得る質疑応答において、解答に対する解答者の確信の有無を判定するプログラムであって、コンピュータに実行させるステップとして、解答の入力に係る履歴を文字入力情報として取得するステップと、得られた文字入力情報から解答者の挙動の特徴を表す情報として予め定められた複数種類の特徴量を抽出するステップと、抽出された特徴量と正解の属性に係る特徴量とに基づいて、解答者の確信の有無を判定するステップと、を備えることを特徴とする。

10

【 0 0 7 7 】

(ix) さらに、この発明による出題選択方法は、コンピュータが、予め定められた複数の設問のうち何れかを出題するステップと、解答者から解答のタイピングまたは手書きによる文字入力を受付けて、上述の確信判定方法により解答に対する確信の有無を判定する

20

ステップと、解答の正誤および確信の有無の判定を参照し、確信なしと判定された正答を再出題すべき設問として選択するステップとを備えることを特徴とする。

この出題選択方法によれば、確信なしと判定されたが正解した設問は、再出題されて解答者が再確認できるようになる。

【 0 0 7 8 】

(x) 前記コンピュータが、確信ありと判定された誤答を、確信なしと判定された誤答よりも短い間隔または高い頻度で再出題されるように決定するステップをさらに備えてもよい。

このようにすれば、確信ありと判定されたが間違えた設問は、確信なしと判定されて間違えた設問よりも文字間隔または高い頻度で再出題されて解答者が再確認できるようになる。

30

【 0 0 7 9 】

この発明の好ましい態様には、上述した複数の態様のうちの何れかを組み合わせたものも含まれる。

前述した実施の形態の他にも、この発明について種々の変形例があり得る。それらの変形例は、この発明の範囲に属しないと解されるべきものではない。この発明には、請求の範囲と均等の意味および前記範囲内でのすべての変形とが含まれるべきである。

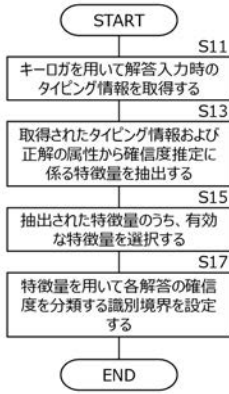
【 符号の説明 】

【 0 0 8 0 】

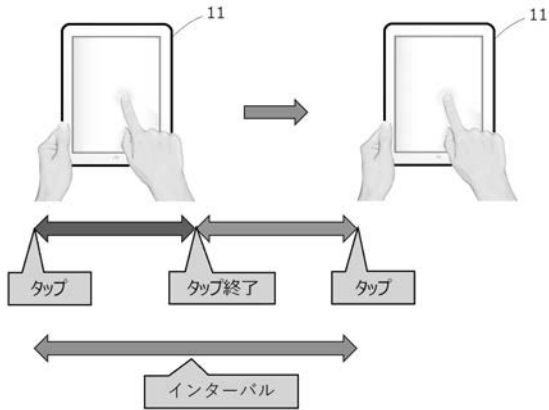
1 1 : 情報処理装置、 1 3 : 英単語アプリ

40

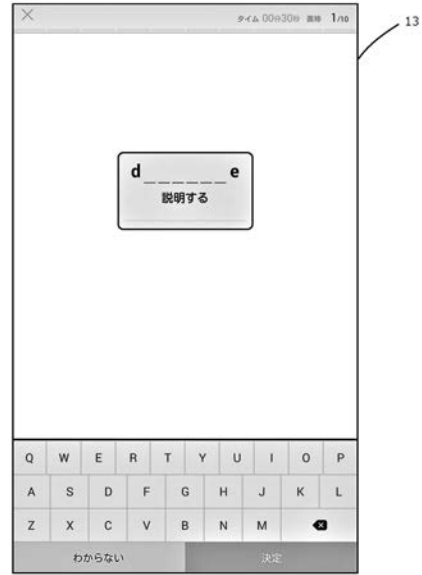
【 図 1 】



【 図 2 】



【 図 3 】

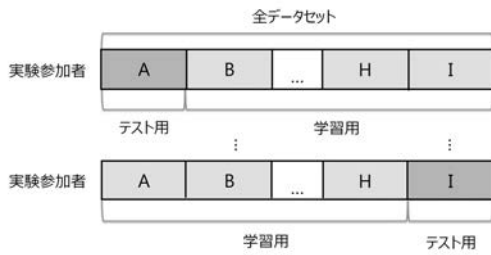


【 図 4 】

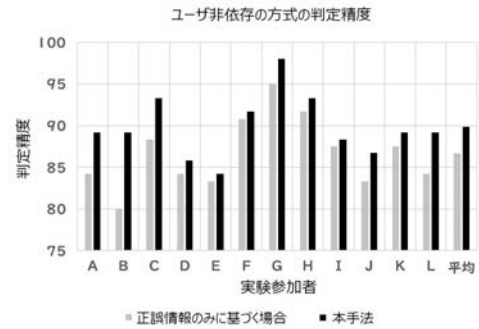
実験参加者 1 名分のデータセット

問題	1~10	11~20	...	101~110	111~120
	テスト用			学習用	
	:			:	
問題	1~10	11~20	...	101~110	111~120
	学習用			テスト用	

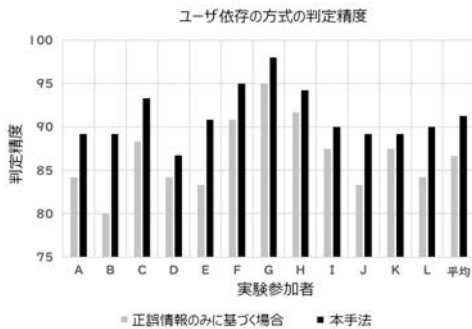
【 図 5 】



【 図 7 】



【 図 6 】



【 図 8 】

HOME

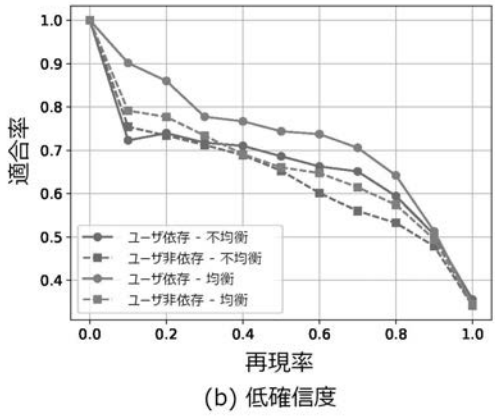
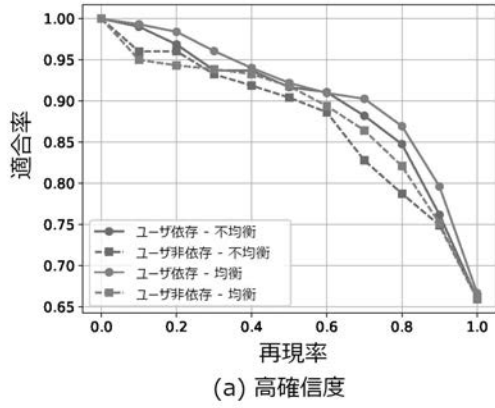
Q1.  
以下の意味を持つ英単語を書きなさい。  
～を足す

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

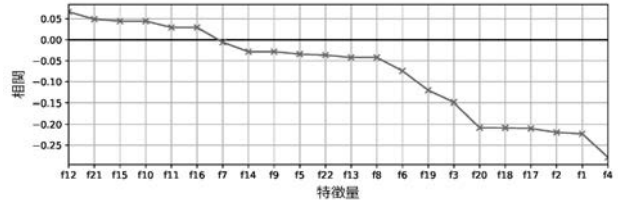
Unconfident	Confident with meaning	Confident with both meaning and spelling
-------------	------------------------	--



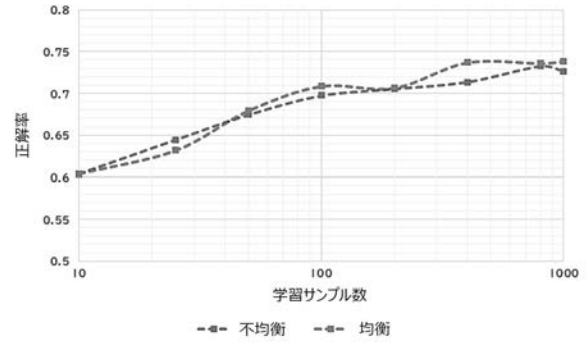
【 図 9 】



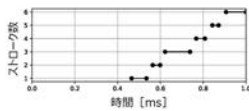
【 図 1 0 】



【 図 1 1 】

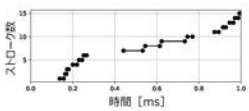


【 図 1 2 】



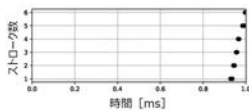
fair

(a) True positive



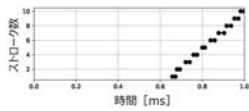
eternal

(b) False positive



fix

(c) False negative



procedure

(d) True negative

---

フロントページの続き

- (72)発明者 オジュロ オリビエ  
大阪府堺市中央区学園町1番1号 公立大学法人大阪府立大学内
- (72)発明者 丸市 賢功  
大阪府堺市中央区学園町1番1号 公立大学法人大阪府立大学内
- (72)発明者 石丸 翔也  
大阪府堺市中央区学園町1番1号 公立大学法人大阪府立大学内
- Fターム(参考) 2C028 AA03 BA02 BB04 BC01