

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 16 日現在

機関番号：32663

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26770205

研究課題名(和文)機械学習によるスピーキングの基準特性抽出と習熟度推定

研究課題名(英文)Predicting the oral proficiency levels with machine learning techniques

研究代表者

小林 雄一郎(Kobayashi, Yuichiro)

東洋大学・社会学部・助教

研究者番号：00725666

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究は、自動採点の技術を用いて、第二言語としての英語学習者(外国語としての英語学習者を含む)の話し言葉の習熟度を自動評価するものである。本研究では、NICT JLE Corpusを主な分析データとし、Biber(1988)で用いられている言語項目に関する情報をデータに付与した。そして、分析に用いた手法は、他の手法よりも比較的高い分類精度が得られると報告されているランダムフォレスト法である。その結果、9段階の習熟度を61.28%の精度で正しく分類できた。これは、ベースラインの精度よりも37.63ポイント高いものである。また、分類に大きく寄与した言語項目は、前置詞、人称代名詞などであった。

研究成果の概要(英文)：The present study aims to automatically evaluate second language (L2) spoken English using automated scoring techniques. I used the NICT JLE Corpus, a corpus of 1,281 Japanese EFL learners, which is coded into nine oral proficiency levels, for the analysis. The nine levels were used as a criterion variable and linguistic features analyzed in Biber (1988) as explanatory variables. Random forests was employed to predict oral proficiency. As a result, 61.28% of L2 spoken productions were correctly classified. Compared to the baseline accuracy of the simplest possible algorithm of always choosing the most frequent level (37.63%), my random forests model improved prediction by 23.65 points. Predictors that can clearly discriminate oral proficiency levels were prepositions, first person pronouns, adverbs, and contractions in the order of strength. The results of this study can be applied to creating assessments that are more appropriate for scaling oral performances of EFL learners.

研究分野：コーパス言語学

キーワード：自動採点 学習者コーパス 自然言語処理 機械学習

1. 研究開始当初の背景

英語教育の分野では、数多くの英語力テストが存在し、中学校や高校のカリキュラムに組み込まれている場合もある。これらのテストの多くは、熟練した教師や採点者が学習者の作文や発話を評価するという形式を取っている。しかし、熟練した採点者を育成するには、かなりの時間が必要とされるであろう。また、いかに熟練した採点者たちが厳密な基準に基づいて評価を下したとしても、複数の採点者間の評価が一致しないこともある。そのような状況において、客観的な評価基準と統計モデルを用いて習熟度を推定する技術を開発することは、言語教育分野にとって非常に有意義なことである。しかしながら、現在の習熟度推定の研究の多くは、次のような2つの大きな問題を抱えている。

まず、客観的な評価基準に関して、習熟度を弁別するための統一的な基準が確立されていないという問題がある。例えば、近年注目を集めている CEFR (Common European Framework of Reference for Languages) における B1 レベルや A2 レベルなどは、1970年代の概念・機能シラバス全盛期に策定された基準に基づいており、文法項目に関する明確な評価基準を持たない(投野, 2010)。また、学習者コーパスから習熟段階別の言語的特徴を記述する試みもあるが、その殆どが非常に少数の言語項目(動詞共起構造、決定詞など)の記述にとどまっており(e.g. Hawkins & Filipović, 2012)、コーパスに基づく網羅的な研究はまだ行われていない。そこで、大規模なコーパスを用いて、習熟度を判定できる項目を明らかにすることができれば、学習者向けテストの評価と開発のために必要な基礎資料を提供することが可能になる。

次に、習熟度を推定するための統計モデルが確立されていないという問題が挙げられる。言語教育分野では、語彙や文法項目のリストに基づいて習熟度を推定することが多い。しかし、このような can-do 方式では、数百・数千に及ぶ多数のチェック項目を効率的に取り扱うことが困難である。その一方、自然言語処理の分野において、習熟度推定や自動採点は、テキスト分類問題の一種として位置付けられ、回帰や分類のための機械学習の手法がその解決に用いられることが多い(e.g. Larkey & Croft, 2003)。そして、機械学習によるテキスト分類を行うにあたっては、どのような言語的特徴を特徴量に用いるかが重要となる(e.g. Grieve, 2007)。だが、多くの場合は、語数や文数といった単純な統計値(e.g. Shermis & Burstein, 2003)、品詞構成率(e.g. Supnithi・内元・齋賀・和泉・Sornlertlamvanich・井佐原, 2004)、n-gram(e.g. 坂田・新保・松本, 2007)、語彙指標(e.g. Crossley, Salisbury & McNamara, 2012)などが特徴量に用いられ、語彙、品詞、統語、談話など多角的な視点から学習者の言語を

評価する試みはあまり見られない

2. 研究の目的

本研究では、英語学習者の習熟度を自動推定すると同時に、その推定に寄与する言語項目を特定することを目的としている。具体的には、9つの異なる習熟度グループに属する日本人英語学習者の話し言葉を対象に、63項目にわたる多種多様な言語項目(語彙、品詞、統語、談話など)を特徴量とし、習熟度の推定と言語項目の特定を行う。

3. 研究の方法

(1) 分析データ

本研究は、NICT-JLE Corpus (和泉・内元・井佐原, 2004)を用いる。このコーパスは、ACTFL OPI に準拠した Speaking Standard Test (SST) を受験した日本人英語学習者 1,281 人の発話データを書き起したものである。

SST の受験者は、1枚の絵の描写、ロールプレイ、複数の絵を使った物語の作成といった、複数のタスクを15分間で行う。それぞれのタスクで同一の絵が用いられている訳ではなく、また、ロールプレイも同一ではない。そのため、タスクの違いによって、学習者の言語使用に差が出ることも予想されるが、本研究においては、全てのデータを分析対象とし、タスクによる言語使用の差は考慮しない。

NICT-JLE Corpus は、SST を受けた学習者の発話から構築されているため、専門の評価官が判定した9段階の習熟度情報が全ての学習データに付与されているという大きな利点を持つ。SST のレベルを ACTFL OPI のレベルと比較すると、日本人英語学習者を判定するための上級レベルは1つしかないが、中級レベルが5つに分けられているのが特徴的である。

本研究では、TreeTagger を用いて NICT JLE Corpus に Penn Treebank タグ (Marcus, Santorini & Marcinkiewicz, 1993) に基づく品詞情報を自動付与したデータを用いる。なお、このデータの各レベルから層化抽出したサンプルを手作業で調査した結果、90%以上の精度で品詞情報が正しく付与されていることが確認されている(阿部, 2012)。この品詞情報は、次節に挙げた言語項目を自動抽出する際に用いる(自動抽出のアルゴリズムに関しては、Biber, 1988 を参照)。

(2) 言語項目

本研究では、Biber (1988) が英語母語話者の話し言葉と書き言葉の分析に用いた 67 種類の言語項目のうち、以下の 60 項目を分析対象とする(なお、type-token ratio は、100 語単位で算出した standardized type-token ratio を使用する)。これらの言

語項目を用いることで、少数の言語的特徴のみを対象とすることが多かった先行研究と比べて、学習者による話し言葉を多角的に評価することが可能になる。

agentless passives, attributive adjectives, be as main verb, by passives, conjuncts, contractions, demonstrative pronouns, discourse particles, emphatics, hedges, independent clause coordination, mean length of words, other adverbial subordinators, past tense, perfect aspect, phrasal coordination, possibility modals, past particle postnominal clauses, predicative adjectives, predictive modals, total prepositional phrases, present tense, pro-verb do, split auxiliaries, split infinitives, stranded prepositions, total adverbs, synthetic negation, that adjective complements, that verb complements, that relatives in object position, that relatives in subject position, infinitives, other total nouns, WH-clauses, direct WH-questions, WH-relatives with front preposition, WH-relatives in object position, WH-relatives in subject position, amplifiers, analytic negation, causative adverbial subordinators, concessive adverbial subordinators, conditional adverbial subordinators, downtoners, existential there, first person pronouns, indefinite pronouns, pronoun it, necessity modals, nominalizations, place adverbials, private verbs, public verbs, second person pronouns, seem and appear, suasive verbs, third person pronouns, time adverbials, type-token ratio

また、本研究では、上記の 60 項目に加えて、mean length of sentences, tokens, types の 3 項目も分析に用いる。これらの項目は、英語学習者の習熟度推定に大きく寄与すると報告されている項目である (e.g. 小林・金丸, 2012a, 2012b)。

(3) 実験手法

近年、多くのパターン認識や機械学習の技術がテキスト分類に応用されている (e.g. Sebastiani, 2002)。その中で、本研究の習熟度推定に用いる手法は、ランダムフォレスト (Breiman, 2001) である。ランダムフォレストとは、決定木のアンサンブル学習である。決定木は、非線形判別分析の 1 つとして位置付けられ、特徴量の値を何らかの基準で分岐させ、分類モデルを構築する (Breiman, Friedman, Stone & Olshen, 1984)。分岐の過程は、木構造で図示することができ、IF-THEN のような簡単なルールで表すことも可能である。また、アンサンブル学習とは、

必ずしも精度の高くない複数の分類器の結果を組み合わせ、精度を向上させる機械学習の手法である (Breiman, 1996)。

ランダムフォレストでは、まず、与えられたデータセットから、 n 組のブートストラップサンプルを作成する。次に、各々のブートストラップサンプルデータを用いて、未剪定の最大の決定木を生成する (但し、分岐のノードは、ランダムサンプリングされた特徴量のうち最善のものを使用する)。そして、全ての結果を多数決で統合し、新しい分類器を構築する (Hastie, Tibshirani & Friedman, 2009)。

ランダムフォレストの長所としては、精度が高いこと、非常に多くの特徴量を効率的に扱うことができること、それぞれの特徴量が予測に寄与する度合いが分かること、などが挙げられる (金, 2007)。

そして、ランダムフォレストは、文学作品の著者推定 (金・村上, 2007) や英語科学論文の質判定 (小林・田中・富浦, 2011) などにも用いられており、言語分析において非常に有効であることが確認されている。

4. 研究成果

ランダムフォレストによる習熟度推定を行うにあたって、ランダムサンプリングする特徴量の数は、全特徴量数の正の平方根を取り、生成する木の数は 500 とした。また、推定の精度は、OOB (out-of-bag) による交差妥当化によって検証した。OOB とは、ブートストラップサンプルの 3 分の 2 で分類モデルを作成し、残りの 3 分の 1 を用いて評価を行った結果の誤判別率である。

その結果、1,281 人のうち 785 人分の発話データが正しく推定されており、全体の精度 (exact agreement) が 61.28%であった。また、レベル 4 をレベル 3 やレベル 5 と推定した例のように、1 つまでの誤差を正答とみなした場合の精度 (adjacent agreement) は、93.91%である。そして、本実験の推定結果と SST の評価官による評価結果の積率相関係数は、0.867 である。

そして、推定に寄与した言語項目は、総語数と異語数に加えて、前置詞句、1 人称代名詞、副詞、否定、短縮形、不定詞、時制などの頻度であった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 8 件)

Yuichiro Kobayashi & Mariko Abe (2016). A corpus-based approach to the register awareness of Asian learners of English. Journal of Pan-Pacific Association of Applied Linguistics,

20(2), 1-17. (査読有)

Yuichiro Kobayashi & Mariko Abe (2016). Automated scoring of L2 spoken English with random forests. *Journal of Pan-Pacific Association of Applied Linguistics*, 20(1), 55-73. (査読有)

Yuichiro Kobayashi (2016). Investigating metadiscourse markers in Asian Englishes: A corpus-based approach. *Language in Focus: International Journal of Studies in Applied Linguistics and ELT*, 2(1), 19-35. (査読有)

Yuichiro Kobayashi (2016). Heat map with hierarchical clustering: Multivariate visualization method for corpus-based language studies. *NINJAL Research Papers*, 11, 25-36. (査読有)

Kiyomi Chujo, Yuichiro Kobayashi, Atsushi Mizumoto, & Kathryn Oghigian (2016). Exploring the effectiveness of combined web-based corpus tools for beginner EFL DDL. *Linguistics and Literature Studies*, 4(4), 262-274. (査読有)

Yuichiro Kobayashi (2014). Computer-aided error analysis of L2 spoken English: A data mining approach. *Proceedings of the Conference on Language and Technology 2014*, 127-134. (査読有)

Yasutake Ishii, Satoru Uchida, Kyoko Hayashi & Yuichiro Kobayashi (2014). An analysis of three collocations dictionaries for learners of English. *Lexicon*, 44, 24-53. (査読有)

Yuichiro Kobayashi & Mariko Abe (2014). A machine learning approach to the effects of writing task prompts. *Learner Corpus Studies in Asia and the World*, 2, 163-175. (査読有)

[学会発表](計 12 件)

Yuichiro Kobayashi & Akira Murakami (2nd March, 2017). Variation across L2 speech and writing: A multidimensional study. *Language in Focus 2017*, Bafla (Cyprus)

Yuichiro Kobayashi & Akira Murakami (23rd October, 2016). Contrastive

analysis of L2 speech and writing: A multi-dimensional approach. *Asia Pacific Corpus Linguistics Conference 2016*, Beijing (China)

Mariko Abe & Yuichiro Kobayashi (10th March, 2016). Grammatical, morphological, and lexical errors between careful and vernacular styles. *Language in Focus 2016*, Istanbul (Turkey)

Yuichiro Kobayashi (5th December, 2015). Comparing native and non-native English through significance tests and effects sizes. *Pan-Pacific Association of Applied Linguistics (PAAL) 2015*, Seoul (Korea)

Yuichiro Kobayashi (19th November, 2015). Investigating metadiscourse markers in Asian Englishes: A corpus-based approach. *Symposium on Second Language Writing 2015*, Auckland (New Zealand)

Masumi Narita, Mariko Abe, & Yuichiro Kobayashi (21st July, 2015). Effects of a writing prompt on L2 learners' essays. *Corpus Linguistics 2015*, Lancaster (UK)

Yuichiro Kobayashi (6th March, 2015). A corpus-based contrastive analysis of L2 metadiscourse. *Language in Focus 2015*, Kapadokya (Turkey)

Yuichiro Kobayashi (14th November, 2014). Computer-aided error analysis of L2 spoken English: A data mining approach. *5th Conference on Language and Technology*, Karachi (Pakistan)

Yuichiro Kobayashi & Mariko Abe (27th September, 2014). The similarity and difference between human scoring and automated scoring. *Applied Linguistics Association of Korea (ALAK) 2014*, Seoul (Korea)

Yuichiro Kobayashi (17th August, 2014). Exploring error patterns in L2 spoken English with data mining techniques. *Pan-Pacific Association of Applied Linguistics (PAAL) 2014*, Tokyo (Japan)

Yuichiro Kobayashi & Mariko Abe (22nd July, 2014). How do L1 and proficiency level influence L2 written production? *Teaching and Language Corpora (TaLC)*

11, Lancaster (UK)

Yuichiro Kobayashi & Mariko Abe (31st May, 2014). A machine learning approach to the effects of writing task prompts. Learner Corpus Studies in Asia and the World (LCSAW) 2014, Kobe (Japan)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

小林 雄一郎 (KOBAYASHI, Yuichiro)

東洋大学・社会学部・助教

研究者番号：00725666

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()