

CEFR 準拠 CDS 自動分類システムの精度改善を目的とした 推定改良手法の提案

On Improving Precision of CEFR-Compliant CDS Classification System

里見 陸^{*1}, 宮崎 佳典^{*1}, 谷 誠司^{*2}, 安 志英^{*3}, 元 裕璟^{*4}
Riku SATOMI^{*1}, Yoshinori MIYAZAKI^{*1}, Seiji TANI^{*2}, Jiyoung AN^{*3}, Yukyoung WON^{*4}

^{*1} 静岡大学

^{*1}Shizuoka University

^{*2} 常葉大学

^{*2}Tokoha University

^{*3} 群山大

^{*3}Kunsan University

^{*4} 高麗大

^{*4}Korea University

Email: satomi.riku.21@shizuoka.ac.jp

あらまし：外国語の能力について世界の各言語で共通した参照枠である CEFR 内には、対象言語を用いて何ができるかを表す CDS が定義されている。本研究は機械学習による日本語文章の CDS 自動分類を行う既存システムに対してその精度改善を目的とする。習得レベルである文章の CEFR レベルを推定したのうち、同一レベル内の文章を対象とした CDS 推定を適用した結果、適合率、再現率の調和平均値が有意に向上したことが観測された。

キーワード：CEFR, CDS, 分類システム, 日本語, 機械学習

1. はじめに

CEFR とはヨーロッパ言語共通参照枠の略で、欧州の言語教育・学習の場で共有する枠組みとして、2001 年に欧州評議会が発表した⁽¹⁾。欧州をはじめとする世界の各言語で実際に利用されており、現在、英語やヨーロッパ各国の言語を含め、多くの言語に対応している。CEFR には「読む、書く、話す、聞く」といった技能項目があるが、本研究では「読む」能力を対象とする。CEFR レベルは 難易度毎に A1, A2, B1, B2, C1, 最高レベルの C2 に分けられているが、2017 年には CEFR を補完するものとして⁽²⁾が公開され、初学習者向けに、A1 レベルよりもさらに初級段階のレベルとして PreA1 レベルが追加された。

各 CEFR レベルは「その言語を使って何ができるか」という形で言語力を表す Can-Do Statement (以下 CDS) を複数用いて構成されており、PreA1 から C2 レベルまでの CDS は合計 38 個ある。ただし、CEFR レベルや CDS の中にはその言語を母語話者とする人にとっても難易度の高いものも存在し、本研究では B2 レベルに属する CDS の内の 1 つと、C1, C2 レベルに属する CDS を除いた計 34 個の CDS を研究対象とすることとしている。CDS は、CEFR レベルとは別に、それぞれが表す文章を読むのに必要な能力の種類によって 5 つの「下位尺度」のいずれかに属する。下位尺度が表す能力としては、「情報や要

点を読み取る能力」、「手紙やメールを読む能力」等といったものがある。そのため、CDS は CEFR レベルと下位尺度の組み合わせで表現でき、文章の難易度と種類を内包した指標であると言える。

2. 本研究の目的

日本語教育へ CEFR を活用することに関して、日本語の CEFR 準拠テキストコーパスが作成されている例が管見の限りほとんど見受けられないことから、網羅的に研究されている事例は現時点において少ないと考えられる。これに対し、我々研究グループは日本語文章の CDS を自動推定するシステムの作成に取り組んでいる。特に本発表では、先行研究におけるシステムの推定手法に対して新たなアプローチを適用することでシステム推定精度の改善を目的としている。

3. 先行研究の推定手法

先行研究⁽³⁾では、PreA1 レベルの文章とそれ以外の文章には難易度や特徴に大きな差があると考えられたため、まず文章が PreA1 レベルかそれ以外の A1-B2 レベルの文章かを 2 値分類してからそれらに属する CDS のみを対象として CDS を推定するという手法を用いた。最新の先行研究⁽⁴⁾においても同様の手法を用いている (図 1)。

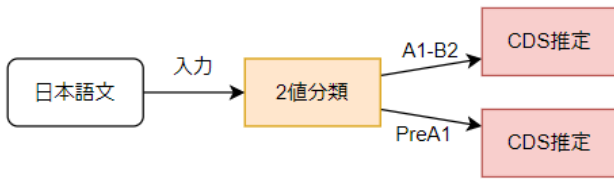


図1 先行研究の推定手法

この際、2 値分類と、A1-B2 と推定された文章に対する CDS 推定には同じ特徴量が用いられているが、PreA1 と推定された文章に対する CDS 推定にはそれらとは異なった特徴量が用いられている。

推定精度の算出は、CEFR の知識を有する、外国にて日本語教育の経験がある 10 名から収集した CDS 情報付きの 555 例文を用いた 10 分割交差検定によって行う。教師データを用いて 2 値分類のモデル、並びに A1-B2 と推定された文章の CDS 推定のモデル、PreA1 と推定された文章の CDS 推定のモデルを作成し、それにテストデータを入力して、本来のテストデータに付与された CDS と推定された CDS を比較して精度を算出する。

テストデータの正解 CDS と推定された CDS から適合率と再現率を算出する。さらに、適合率と再現率の調和平均である F 値を算出し、これを最終的な精度として考えることとする。すると、先行研究⁽⁴⁾の推定手法における各値は表 1 のようになった。

表 1 先行研究の推定手法の精度

適合率	再現率	F 値
49.6%	71.2%	56.5%

4. 本研究の提案手法

先行研究の手法では、2 値分類と A1-B2 における CDS 推定に同じ特徴量を用いており、これは難易度を表す CEFR レベルと難易度と能力の種類を内包した CDS の推定に同じ特徴量を用いていることとなり、適切ではないと考えられる。また、先行研究の手法では PreA1 とそれ以外では難易度や特徴に大きな差があると考えられたため 2 値分類をする手法を取っていたが、A1 から B2 レベルに関しても、各レベル間には難易度や特徴には差があると考えられる。そこで、文章の CEFR レベルを推定した後、その CEFR レベルに属する CDS のみを対象とした推定をするという手法を提案する (図 2)。また、本発表では先行研究に対して推定手法を変更することによる精度の変化を得ることを目的とするため、CEFR レベルや CDS の推定には先行研究の手法における 2 値分類等と同じ特徴量を用いている。

その結果、10 分割交差検定による精度は表 2 のようになり、提案手法の精度は先行研究に比べて適合

率は上がって再現率は下がり、F 値は上昇した。

表 2 提案手法の精度

適合率	再現率	F 値
73.1%	56.4%	63.6%

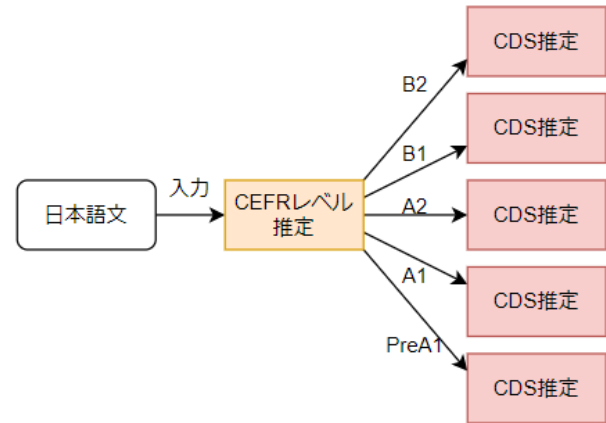


図2 本研究の提案手法

5. まとめと今後の展望

本研究では、現在開発を行っている CEFR 準拠 CDS 自動分類システムに対し、先行研究の推定手法に対して新たなアルゴリズムを適用することで精度の改善を行うことができた。

また、本研究の提案手法では、先行研究の手法と精度を比較するため、CEFR レベルや下位尺度の推定、そして CEFR レベル毎の CDS の推定全てにおいて、参照する特徴量を同じものとしているが、やはり文章の難易度を表す CEFR レベルや種類を表す下位尺度の推定と、それらを合わせた指標である CDS の推定に同じ特徴量を用いることには疑問が残る。また、それぞれの推定により適した特徴量を適用することによってそれぞれの推定の精度を向上させることができると考えられるため、それらの推定により適した特徴量を考察し、適用することは今後の展望として挙げられる。

参考文献

- (1) Council of Europe: Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Cambridge University Press (2001)
- (2) Council of Europe: Common European Framework of Reference for Languages: Learning, Teaching, Assessment Companion Volume with New Descriptors (2017)
- (3) VUONG HONG DUC “CEFR Companion Volume に対応した日本語例文自動分類支援 Web アプリケーションの改良” (2021)
- (4) CAO HOAI GIANG, 宮崎 佳典, 谷 誠司, 安 志英, 元 裕環, CEFR に対応した日本語例文自動分類システムの BERT 適用による精度改善の試み, 韓国日本學會 第 102 回国際學術大會 (2021)