

コーパスを用いた言語習得度の推定

坂田浩亮 新保仁 松本裕治
{kosuke-s, shimbo, matsu}@is.naist.jp

奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学講座

概要

言語教育において、学習者の言語習得度を知ることは教師にとって重要なことである。また、言語学習者の言語習得度を客観的に量る手段があれば、学習者は教師がいなくても自分の言語習得度を評価できるため、学習効率の向上につながると考えられる。本発表では、言語学習者の作文と習得度別コーパスとの類似度に基づいて、学習者の言語習得度を推定する手法を提案する。NICT JLE コーパスを用いて行った提案手法の評価実験と、その結果について紹介する。

Automatic estimation of English proficiency level using corpora

Kosuke Sakata Masashi Shimbo Yuji Matsumoto
{kosuke-s, shimbo, matsu}@is.naist.jp

Computational Linguistics Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology

Abstract

We propose methods for automatically evaluating second language learners' proficiency levels. These methods not only help teachers evaluate students' proficiency levels, but also accelerate students' learning, since they can evaluate their proficiency levels whenever they like without consulting their teachers. The proposed methods compute students' proficiency levels on the basis of the similarity between their free compositions and the corpus of compositions divided into skill levels.

1 はじめに

言語教育に従事する教師にとって、学習者(生徒)の言語習得度を知ることは重要なことである。また、学習者個々人の作文から言語習得度を客観的、自動的に量る手段があれば、学習者は教師がいなくても自分の作文能力を評価できることから、学習効率の向上につながると考えられる。

関連研究として、機械翻訳機(MT)の性能を評価するためによく使われる BLEU [3] という尺度がある。この評価結果は、人間が下す評価に近いと言われている。しかしながら、BLEU は MT の翻訳結果と人間の翻訳結果に使われている単語の一致度を測定して性能を評価するため、人間によ

て翻訳された回答例をあらかじめ準備しておかなければならないという問題がある。我々の先行研究では、MT の性能評価の指標に使われる BLEU を参考にして、MT の翻訳結果の代わりに言語学習者の作文を、人間の翻訳結果の代わりにコーパスを用いて類似度を測定し、学習者の言語習得度を推定する手法の提案、NICT JLE コーパス [1] を用いた提案手法の評価実験および実験結果について紹介した。本発表では、(i) 前回とアルゴリズムは変わらないが、学習者の作文との比較対象となる参照コーパスおよび訓練データの分割方法を変えた手法、そして (ii) 今回新しく、使われている単語と品詞の頻度から直接レベル推定関数を回帰する手法の比較を行う。

以下では、2節で関連研究について、3節で我々が提案する3つの手法の詳細について、4節で実験および実験結果について述べ、5節で論文を結ぶ。

2 関連研究

2.1 BLEU

機械翻訳機 (MT) の性能評価に用いられる。BLEU は、MT の翻訳結果と人間による翻訳例との類似度 (用いられている単語の一致度) 尺度の1つであり、これによって MT の翻訳結果の妥当性と流暢性を評価することができる。BLEU は、人間が下す評価に近い性能評価をするという結果が示されている [3]。BLEU スコアは、評価したい MT の翻訳結果と人間の翻訳結果に関して単語 n -gram の重なりに基づいて、次式によって計算される。スコアは 0 から 1 の値をとり、スコアが大きいほどよい翻訳であるとされる。

$$S_{BLEU} = BP \cdot \exp \left(\sum_{n=1}^N w_n \log P_n \right)$$

ただし、

$$BP = \begin{cases} 1 & \text{if } c > r \\ \exp \left(\frac{1-r}{c} \right) & \text{if } c \leq r \end{cases}$$

c は MT の翻訳結果の文長、 r は人間による回答例の文長であり、 P_n は *modified n-gram precisions* と呼ばれ、

$$P_n = \frac{\sum_{C \in \Gamma} \sum_{n\text{-gram} \in C} \text{Count}_{\text{clip}}(n\text{-gram})}{\sum_{C' \in \Gamma} \sum_{n\text{-gram}' \in C'} \text{Count}(n\text{-gram}')}$$

と定義される。ここで、 $\text{Count}(s)$ は MT の翻訳結果に含まれる単語 n -gram s の数であり、一方 $\text{Count}_{\text{clip}}(s)$ は回答例に含まれる s の最大数を t として、 $\min(\text{Count}(s), t)$ である。 Γ は MT の翻訳結果の集合である。

w_n は、正規化の定数であり、各 n -gram に対する重みと解釈できる。一般には、 $1/N$ の均一重みが使われる。

2.2 安田らの手法

安田ら [5] は受験者の英語コミュニケーション能力を単語 n -gram の一致度を用いて推定している。

この手法は以下のステップからなる。(1) 対訳コーパスから自動選択した日本語文を受験者に英語に訳してもらう。(2) 受験者の訳文と対訳コーパスの翻訳例との間の一致度を訳文中の単語 n -gram の重なり (BLEU など) や編集距離等を用いて機械的に測定する。コーパス内の翻訳例は複数個ある例もあり、一致度を求める際には上の BLEU や DP 等を用いる。(3) 予め様々な能力値と一致度の相関を学習しておき、未知の受験者の能力値を一致度から推定する。使われている能力値としては TOEIC スコアを採用している。

2.3 我々の先行研究

我々の先行研究 [6] ではレベルが既知のデータとの単語 n -gram に基づく類似度を素性に用いる手法を提案した。さらに、回帰分析を利用した手法を提案した。安田ら [5] の手法と異なる点は、安田らの手法は対象となる作文のテーマを指定しているが、我々の手法はテーマを指定しない自由作文を対象にレベルを推定する点である。

3 提案手法

BLEU を言語学習者のレベル推定に応用することを考えると、MT の翻訳結果の代わりに学習者の作文を、人間による翻訳例の代わりに教師によって訂正された学習者の作文を用いる方法である。安田らの手法はこれに該当する (学習者の作文と対訳コーパスの翻訳例を用いる手法)。しかし、この方法で推定する場合、人間によって訂正された作文の回答例をあらかじめ用意しておかなければならないという問題がある。レベルを推定したい学習者が多数の場合、各々の学習者の作文の回答例を用意することは非常に手間がかかり、さらに何を書くか予測できない作文の回答例をあらかじめ用意しておくことは困難である。現実的に考えると学習者の作文に対する回答例が不要な手法が望ましい。

我々が以下で提案するレベル推定手法では、作文の回答例の代わりに、多数のレベル付けされた作文 (以下、参照コーパス、あるいは単にコーパスと呼ぶ) を用いる。参照コーパス中の作文の内容はレベル推定の対象となる作文と必ずしも一致しない。したがって、BLEU で想定している状況 (同一内

容の翻訳どうしの比較)とは異なり、別内容の作文どうしを比較することになる。

以下では3種類の提案手法について紹介する。提案手法1は、 n -gramの類似度に重みを与えない手法である。一方、提案手法2では別途訓練データを用意し、 n -gramの重みを回帰分析を用いて最適化する。提案手法3は、単語と品詞の頻度の重みを直接回帰分析を用いて最適化する手法である。

我々は先行研究 [6] の実験において、類似度の大きさをレベルを推定する手法(提案手法1)と類似度を回帰分析の学習に用いてレベルを推定する手法(提案手法2)を比較した結果、回帰分析を利用する手法がより有効であることを示した。このため、後述の実験では提案手法1については評価しないが、便宜上、提案手法1についても解説する。

3.1 提案手法1

提案手法1は、学習者の作文と各レベル毎の参照コーパスとの n -gramに基づく類似度を計算し、最も類似度が大きくなった時に比較した参照コーパスのレベルを推定値とする手法である。学習者による作文コーパスをレベル別に M 個に分割したもの¹を用いて以下のステップを実行する。

1. 各レベルのコーパスから素性ベクトルを生成する

- レベル $1, 2, \dots, M$ の各コーパスから n -gram($n=1, 2, \dots, N$)を抽出し、頻度ベクトルを計算する。
- レベル m のコーパスの素性ベクトルを \mathbf{c}_m とすると、 \mathbf{c}_m は、各 n -gram 単語列の頻度を要素とするベクトル $\mathbf{c}_{m,n}$ を連結したベクトルで、 $\mathbf{c}_m = (\mathbf{c}_{m,1}, \mathbf{c}_{m,2}, \dots, \mathbf{c}_{m,N})$ と書ける。(図1を参照)

2. レベルを推定したい作文の素性ベクトル \mathbf{v} を以下のように計算する

- 作文の素性ベクトルを \mathbf{v} とすると、 $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$ であり、ここで \mathbf{v}_n は単語 n -gram 頻度ベクトルである。

¹各レベルのコーパスには複数の作文が含まれるが、これらをまとめて一個として扱い n -gram を抽出する

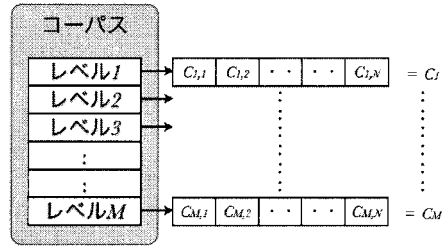


図1: コーパスの素性ベクトル

3. 作文の素性ベクトルに対して全てのコーパスとのコサインをとる

$$x_{m,n} = \cos \angle (\mathbf{v}_n, \mathbf{c}_{m,n})$$

これらを並べて \mathbf{v} のレベル別類似度ベクトル \mathbf{x} とする。すなわち

$$\mathbf{x} = (x_{1,1}, \dots, x_{m,n}).$$

(Algorithm 1 を参照)

4. レベル推定を行う

- $1 \sim N$ -gram のコサイン類似度の和をそのレベルのスコアとし、スコアが一番高いコーパスレベル y を推定値とする。すなわち、

$$y = \arg \max_m \sum_{n=1}^N x_{m,n}$$

Algorithm 1 ComputeSimilarityVector: レベル別類似度ベクトルを計算

入力: 作文の素性ベクトル $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_N)$
出力: \mathbf{v} のレベル別類似度ベクトル \mathbf{x}

```

for  $m \leftarrow 1 \dots M$  do { 各レベル  $m$  について }
  for  $n \leftarrow 1 \dots N$  do { 各  $n$ -gram について }
     $x_{m,n} \leftarrow \cos \angle (\mathbf{v}_n, \mathbf{c}_{m,n})$ 
  end for
end for
 $\mathbf{x} \leftarrow (x_{1,1}, \dots, x_{m,n})$ 
return  $\mathbf{x}$ 

```

3.2 提案手法2

3.2.1 手法

提案手法1では、コサイン類似度が最大のレベルをその作文レベルの推定値とした。これに対し

て提案手法 2 では、参照データと訓練データに分割し、参照データとの類似度を素性に用いた回帰によって各 n -gram の重みを最適化する。この際、各レベル $m = 1, \dots, M$ の n -gram コサイン類似度 $x_{m,n}$ に対して重み $w_{m,n}$ を定義し回帰式 $y = \mathbf{w} \cdot \mathbf{x} + b$ を求める。ここで、 $\mathbf{x} = (x_{1,1}, \dots, x_{m,n})$, $\mathbf{w} = (w_{1,1}, \dots, w_{M,N})$, b は定数項である。学習者レベルが既知の作文 K 個と、レベル別コーパスを用いて以下のステップを実行する。

1. 各レベルのコーパスから素性ベクトルを生成する

- レベル $1, 2, \dots, M$ の各コーパスから単語 n -gram ($n=1, 2, \dots, N$) を抽出し、頻度ベクトルを計算する。
- レベル m のコーパスの素性ベクトル \mathbf{c}_m を計算する。ここで、 $\mathbf{c}_{m,n}$ は提案手法 1 で用いたのと同じ単語 n -gram の頻度ベクトルとして、 $\mathbf{c}_m = (\mathbf{c}_{m,1}, \mathbf{c}_{m,2}, \dots, \mathbf{c}_{m,N})$ である。

2. レベルが既知の訓練データから素性ベクトルを生成する

- K 個の作文各々について n -gram ($n=1, 2, \dots, N$) の統計をとる。
- k 番目の作文の素性ベクトル $\mathbf{v}^{(k)}$ を計算する。ここで、 $\mathbf{v}^{(k)} = (\mathbf{v}_1^{(k)}, \mathbf{v}_2^{(k)}, \dots, \mathbf{v}_N^{(k)})$ であり、ここで $\mathbf{v}_n^{(k)}$ は k 番目の作文の単語 n -gram 頻度ベクトルである。図 2 を参照。

3. 訓練データの素性ベクトルに対して全てのコーパスとのコサイン類似度をとる

$$x_{m,n}^{(k)} = \cos \angle (\mathbf{v}_n^{(k)}, \mathbf{c}_{m,n})$$

これらを並べて各訓練データの類似度ベクトル $\mathbf{x}^{(k)} = (x_{1,1}, \dots, x_{M,N})$ を得る。(Algorithm 2 を参照)

4. 回帰分析による重み付けを行う

- 訓練データ $\{(\mathbf{x}^{(k)}, y^{(k)})\}_{k=1}^K$ の近似式

$$y = \mathbf{w} \cdot \mathbf{x} + b \quad (1)$$

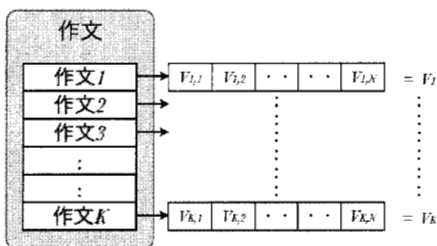


図 2: 作文の素性ベクトル

Algorithm 2 全訓練データに対してレベル別類似度ベクトルを計算

```

入力: 作文の集合  $\{\mathbf{v}^{(k)}\}_{k=1}^K$ 
出力: レベル別類似度ベクトル集合  $\{\mathbf{x}^{(k)}\}_{k=1}^K$ 

for  $k \leftarrow 1 \dots K$  do { 各作文  $k$  について }
   $\mathbf{x}^{(k)} \leftarrow \text{ComputeSimilarityVector}(\mathbf{v}^{(k)})$ 
end for
return  $\{\mathbf{x}^{(k)}\}_{k=1}^K$ 

```

を回帰によって求める。ここで、

$$\mathbf{w} = (w_{1,1}, w_{1,2}, \dots, w_{M,N})$$

はレベル別類似度ベクトルの各要素に対する重みである。

5. レベル推定を行う

- 求めた回帰式 (2) に作文の素性ベクトル \mathbf{x} を代入し、 y を計算する。 y を四捨五入した値を推定レベルとする。

3.2.2 参照データと訓練データ

提案手法 2 では、回帰学習モデルを作成するために参照データと訓練データを必要とする。訓練データと参照データとの類似度を素性に用いるためである。実際にレベル推定に用いる場合を考えると、テストに用いるデータが参照コーパスに含まれることはほとんどない。これと同じ状況を作るため、先行研究 [6] では、コーパスを参照データと訓練データに二等分し互いに背反である状態で実験を行った。しかしながら、参照コーパスは文単位ではなく集合としてベクトルが作られる。また、類似度尺度にコサイン尺度を使っているため、訓練データ自身が参照コーパスに含まれていたとしても、参照コーパスが十分大きければオーバーフィッティング等の悪影響は避けられると推測できる。本実験ではこの仮説を確認するため、全体を参照データと訓練データに全く同じデータ (コーパス全体) を使った場合についても実験を行う。

3.3 提案手法 3

提案手法 2 では、参照コーパスとのコサイン類似度を素性に用いて作文レベルの推定をした。提案手法 3 では、提案手法 2 と同様に別途訓練データを用意するが、参照用コーパスは用いず、各作

文に出現する単語²を素性に用いて直接回帰によってレベルを推定する。具体的には、学習者のレベルが既知である作文 K 個を用いて以下のステップを実行する。

1. レベルが既知の訓練データから素性ベクトルを生成する
 - K 個の作文各々について各単語の出現回数を数え頻度ベクトル $\mathbf{v}^{(k)}$ を計算する。
2. 回帰分析による重み付けを行う
 - 訓練データ $\{(\mathbf{v}^{(k)}, y^{(k)})\}_{k=1}^K$ の近似式
$$\mathbf{y} = \mathbf{w} \cdot \mathbf{v} + b \quad (2)$$
を回帰によって求める。
3. レベル推定を行う
 - 求めた回帰式 (2) に作文の素性ベクトル \mathbf{v} を代入し、 y を計算する。 y を四捨五入した値を推定レベルとする。

3.4 素性

BLEU では MT の翻訳結果に出てくる単語のみを用いて人間による回答例との類似度を測定しているが、使われる単語は書かれている文の話題に依存しやすいため偏りが生じる可能性がある。人間による回答例の代わりに使われるコーパスは翻訳の場合とは異なり評価したい文の内容と必ずしも一致しないため、単語のみの素性は適切ではないこともあり得る。その場合には、単語の代わりに品詞、または単語と品詞の両方を素性として使うことが可能である。実際、4 章の実験では単語とともに品詞も用いている。

3.5 類似度尺度

今回、上記の提案手法の説明 (および後述の実験) においてはコサイン類似度を用いたが、レベル別類似度ベクトルの要素として任意の類似度尺度を用いることが可能である。BLEU で用いられている modified n-gram precisions などコサインのかわりに用いる、あるいは併用することも可能である。

²実際には品詞も併用する。

4 実験

提案手法の有効性および素性ベクトルの有効性を検証する。

4.1 使用コーパス

本実験で使用するデータとして、NICT JLE コーパスを用いる。NICT JLE コーパスは、英語話者ではなく日本人英語学習者の会話を書き起こした言語データベースである。このコーパスの特徴として、英語話者とのインタビュー形式で学習者が質問に答えており、話し言葉のため、フィラーや言い直し等もそのまま記載されている³。また、最大の特徴は SST (Standard Speaking Test) により発話能力を 9 レベルに分けてあり、このレベルは文法、語法、発音、流暢性を客観的に評価している点である。

4.2 実験設定

NICT JLE コーパスに対して 2 通りのレベル分けを行って実験をする。ひとつは、NICT JLE コーパスにもともと与えられた 9 段階のレベルに分ける。もうひとつは、NICT JLE コーパスの 9 レベルを 3 等分し、3 段階に分けて⁴実験に用いる。素性としては、単語 n -gram だけではなく、品詞 n -gram 及び単語 n -gram と品詞の n -gram を組み合わせたものも用いる。これは単語 n -gram はコーパスの会話によって使われる単語に偏りが生じてしまうためである (3.4 節を参照)。なお、本実験では単語および品詞、いずれについても $N = 5$ 。すなわち 1~5-gram を素性に使う。提案手法 2, 3 の回帰分析には、Support Vector Regression [4]⁵ を用いる。

4.3 カーネル

本実験では、回帰学習に多項式カーネルおよび RBF カーネルを適用し、パラメータを変えながら実験を行った。各データセットで最も良かった精

³本実験では前処理によってフィラー等は取り除いた。

⁴初級 (レベル 1~3)、中級 (レベル 4~6) および上級 (レベル 7~9) の 3 段階。

⁵実験には SVMlight[2] を利用した。

表 1: NICT JLE コーパスの語彙リスト ([1] より)

レベル	全体	1	2	3	4	5	6	7	8	9
被験者数	1,201	3	35	222	482	236	130	58	25	10
平均発話数	1,115	338	475	790	1,060	1,298	1,412	1,505	1,715	1,632
文の平均長	7.65	3.09	4.04	5.9	7.44	8.44	9	9.14	9.25	9.42

度をそのデータセットにおける結果とし、全データセットの平均を実験結果とする。

4.4 比較する手法

提案手法 1 と提案手法 2 の比較は、先行研究 [6] で示したので省略し、本論文では提案手法 2 と提案手法 3 を比較する。また、提案手法 2 については、参照コーパスと訓練データの量を変えて以下の 2 種類の状態で実験を行う。

提案手法 2-A コーパスを参照コーパスと訓練データに 2 等分し、互いに背反であるデータ ([6] と同じ)。

提案手法 2-B コーパス全体を参照コーパスと訓練データに使用し、参照コーパスと訓練データが等しいデータ。

4.5 テストデータと訓練データ

提案手法の評価を行うためにテストデータを用いる。このテストデータは、NICT JLE コーパスから各レベルの分布を維持した上でランダムに 30 個の作文を取り出したものを使う。この際、レベルに偏りができないように初級、中級、上級の 3 段階にコーパスを分けた後、各々の段階から 10 人分ずつ取り出す。ここで、各々の段階に含まれる 3 つのレベルの分布は、各段階ごとのレベルの分布を模倣している。従って、本実験で用いるテストデータの各レベルの分布は以下の表 2 のようになる。提案手法 2 では回帰素性に参照コーパスとの類似度を回帰学習に用いるため別途参照データを用意する。この参照データとの類似度を用いて回帰学習したモデルにテストデータを用いて評価を行う。NICT JLE コーパスからテストデータを引いた残りを訓練データ、および参照データとして用いる。提案手法 2-A ではコーパスを 2 等分し、参照データと訓練データは背反であり、提案手法

2-B ではコーパス全体を参照データと訓練データに用い、お互いが等しいデータである。また、提案手法 3 については、参照データを用いないため、NICT JLE コーパスからテストデータを引いた残りを単に訓練データとして用いる。本実験に用いるテストデータ、参照データおよび訓練データをまとめて単に実験データセットと呼ぶ。本実験では各提案手法に対して、これらを 3 セット用意して評価を行う。

4.6 実験結果と考察

3 つの提案手法の実験結果を図 3 および図 4 に示す。実験結果は 3 セットの実験データセットの平均値である。

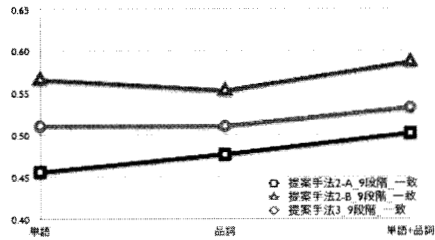


図 3: 実験結果 1 (9 段階)

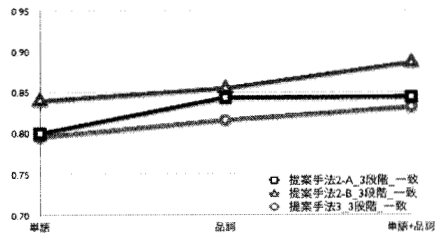


図 4: 実験結果 2 (3 段階)

図 3 は提案手法 2-A, 2-B および提案手法 3 による 9 段階に分割したコーパスレベルを推定したと

表 2: テストデータの各レベルの分布

レベル	1	2	3	4	5	6	7	8	9	合計
分布	1	2	7	5	3	2	6	3	1	30

きの精度を示している。一方図 4 は 3 段階に分割したコーパスレベルを推定したときの精度を示している。2 つの実験結果より、提案手法 2-A より提案手法 3 の方が良い精度であったが、コーパス全体を参照コーパスと訓練データに用いること (提案手法 2-B) により提案手法 3 よりもよい精度が得られた。これにより、提案手法 2-A はコーパスの半分しか参照コーパス、および訓練データに当てられなかったことによるデータ不足が原因と考えられる。一方、提案手法 3 と提案手法 2-B は同じ量の訓練データを使用しているが、類似度を素性にしている提案手法 2-B の方が良い精度が得られている。このことより、単純に単語を素性ベクトルに用いるよりも、類似度を用いた方が有効であることがわかる。また、図 4 より、提案手法 2-B による 3 レベル推定ではほぼ 9 割正しくレベル推定することができた。以上 2 つの実験結果より、単語と品詞に基づく類似度を素性ベクトルにした回帰分析による重みの最適化は有効な手法であることが確認できた。

5 おわりに

学習者が書いた作文から言語習得度を推定する手法を提案した。提案手法では個々の作文内容に依存した回答例を必要とせず、レベル付けされたコーパスを用いる。提案手法では、類似度、単語の頻度および回帰分析を利用した手法によりレベルを推定した。提案手法 1 は各レベルに対して 1 ~ 5-gram のコサイン類似度を素性に使う手法。提案手法 2 は各レベルの作文毎に使われている単語または品詞、および単語と品詞の頻度を素性に使い、回帰を行う手法。そして提案手法 3 は、単語の頻度を直接回帰学習に用いる手法をそれぞれ提案した。今回の実験では、実験データに NICT JLE コーパスを採用したところ、9 段階の推定では約 58%、3 段階の推定では約 88% の精度で習得度を推定することができた。

NICT JLE コーパスのようにレベル付けされたコーパスはほとんどない上にサンプル数が少ない

ため、回帰分析を行うための訓練データにサンプルを多くとることができない。さらに、中級レベルの学習者が多く、初級および上級レベルの学習者が極端に少ない問題もある。今後は、レベル付けされていないコーパスを使って学習者のレベルを推定する手法を考えたい。また、レベルを推定した際に学習者にとって本当に必要な「なぜそのレベルに推定されたのか」という情報をフィードバックする手法についても考えなければならない。

参考文献

- [1] 和泉絵美, 内元清貴, 井佐原均. 日本人 1200 人の英語スピーキングコーパス. アルク, 2004.
- [2] Thorsten Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*. MIT Press, 1999.
- [3] Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th ACL*, pp. 311-318, 2002.
- [4] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, second edition, 1998.
- [5] 安田圭志, 隅田英一郎, 山本誠一, 柳田益造, 前川喜久雄, 菅谷史昭. 英語コミュニケーション能力の自動測定技術の提案. 情報処理学会研究報告. 自然言語処理研究会報告, pp. 65-70. 情報処理学会, 2003.
- [6] 坂田浩亮, 新保仁, 松本裕治. コーパスを用いた言語習得度の推定. 言語処理学会第 13 回年次大会論文集, pp. 793-796. 言語処理学会, 2007.