文章変換技術に基づくデータ拡張を用いた問題横断型自動採点手法

Cross-prompt Automated Essay Scoring Based on Data Augmentation Using Text-Transfer Method

伊藤佑真 *1，宇都雅輝 *1
Yuma Ito*1, Masaki Uto*1
*1 電気通信大学
*1The University of Electro-Communications
Email: {ito_yuma, uto}@ai.lab.uec.ac.jp

あらまし： 従来の小論文自動採点の研究では，目的の小論文問題ごとに自動採点モデルを構築する問題固有型自動採点が一般に研究されてきた．他方で，近年では，目的の問題とは異なる問題に対する採点済み小論文データを用いて，汎用的な自動採点モデルを構築する問題横断型自動採点手法が注目を集めている．本研究では，高精度な問題横断型自動採点を目指し，文章の構造は維持したままドメインのみを変換する技術を応用した手法を提案する．

キーワード： 自動採点，深層学習，自然言語処理，データ拡張

1　はじめに

小論文試験の採点業務にかかる人的・経済的コストを削減する方法の一つとして，小論文自動採点（Automated Essay Scoring ; AES）が近年注目を集めている．従来の小論文自動採点の研究では，目的の小論文問題に対する採点済み小論文データを訓練データとして，深層学習モデルに代表される機械学習モデルを教師あり学習する問題固有型自動採点手法の研究が一般的であった．しかし，問題固有型自動採点において高精度を達成するためには，目的の問題に対する採点済み小論文データが大量に必要になる．

この問題を解決するために，近年では，目的の問題とは異なる問題に対する採点済み小論文データを用いて汎用的な自動採点モデルの構築を目指す問題横断型自動採点手法の研究が注目を集めている（e.g.,[1]）．問題横断型自動採点は，転移学習と呼ばれる機械学習タスクの一種とみなせる．転移学習手法の一つとして，目的のタスクに対応する教師データを手元にある別のタスクのためのデータから生成する「データ拡張」のアプローチが有効であると知られている．しかし，問題横断型自動採点の従来研究では，問題非依存の特徴量を構築してモデルを汎化させるアプローチが主流であり，データ拡張を用いた転移学習に基づく手法は見当たらない．そこで本研究では，データ拡張に基づく新たな問題横断型自動採点手法を提案し，その有効性を評価する．

2　提案手法

本研究では，データ拡張の手法として，ドメイン調整可能な文章変換技術 [2] を利用する．この手法は，目的外のドメイン（文章のジャンルやトピックを意味する）に関連する文章データから，個々のドメインに固有の単語をマスクし，マスクした単語に目的のドメインに固有の単語を埋め込むことで文章のドメイン変換を行う．提案手法では，一つ一つの小論文問題をドメインとみなし，目的の問題とは異なる問題に対する採点済み小論文データを目的の問題の特徴に合った文章に変換することでデータ拡張を行う．提案手法は，1）文章ドメイン変換を行うモデルの訓練，2）文章ドメイン変換モデルを用いたデータ拡張，3）拡張したデータを用いた自動採点モデルの構築の 3 つの手順で構成される．

手順 1）と 2）の概念図を図 1 に示す．ここからは，これらの手順の詳細を説明する．なお，以降では，I 個の問題集合を $\mathcal{I} = \{1, \ldots, I\}$ とし，自動採点したい問題を $p \in \mathcal{I}$ とする．さらに，目的の問題とは異なる問題 $i \in \mathcal{I} \backslash \{p\}$ に対する採点済み小論文データを $\mathcal{E}_i = \{(e_{ij}, y_{ij})\}_{j=1}^{J_i}$，目的の問題 p に対する小論文データを $\mathcal{E}_p = \{e_{pj}\}_{j=1}^{J_p}$ で表す．ここで，J_i は問題 i に対する小論文の数，e_{ij} は問題 i に対する j 番目の小論文，y_{ij} は小論文 e_{ij} の得点である．

2.1　文章ドメイン変換モデルの訓練

文章変換は，問題に固有の語彙やフレーズをマスキングして，目的の問題に固有な語彙やフレーズに置き換えることで行われる．マスキング操作では，初めに閾値 τ，変換元の問題 s，変換先の問題 t を設定する．次に，問題 s に関する各小論文 e_{sj} 中の各 n グラム $w^{(n)}$（$n \in \{1, 2, 3\}$ のみを対象）に対し，個々の n グラム $w^{(n)}$ が問題 t と比較してどの程度問題 s に関連しているかを表す式（1）のマスキングスコア $\mathrm{m}(w^{(n)}, \mathcal{E}_s, \mathcal{E}_t)$ を計算する．このスコアが閾値 τ を上回る場合，対象の n グラムを問題 s に固有のものと判断して，マスクする．

$$\mathrm{m}(w^{(n)}, \mathcal{E}_s, \mathcal{E}_t) = \rho(w^{(n)}, \mathcal{E}_s) - \rho(w^{(n)}, \mathcal{E}_t). \quad (1)$$

ここで，

$$\rho(w^{(n)}, \mathcal{E}_i) = P(\mathcal{E}_i|w^{(n)}) \left(1 - \frac{H(w^{(n)})}{\log I}\right), \quad (2)$$

$$H(w^{(n)}) = -\sum_{k=1}^{I} P(\mathcal{E}_k|w^{(n)}) \log P(\mathcal{E}_k|w^{(n)}), \quad (3)$$

$$P(\mathcal{E}_i|w^{(n)}) = \frac{P(w^{(n)}|\mathcal{E}_i)}{\sum_{k=1}^{I} P(w^{(n)}|\mathcal{E}_k)}, \quad (4)$$

$$P(w^{(n)}|\mathcal{E}_i) = \frac{n_{w^{(n)}|i} + \alpha}{J_i}. \quad (5)$$

表 1　実験結果

手法	問題 1	問題 2	問題 3	問題 4	問題 5	問題 6	問題 7	問題 8	平均
提案（$\tau = 0.08$）	0.545	0.421	0.624	0.635	0.690	0.593	**0.626**	0.357	0.561
提案（$\tau = 0.2$）	0.547	0.439	**0.628**	0.641	**0.701**	0.606	0.570	0.374	0.563
提案（$\tau = 0.5$）	0.597	0.511	0.580	**0.646**	0.568	**0.665**	0.558	**0.411**	0.567
提案（$\tau = 0.8$）	**0.605**	0.520	0.591	0.601	0.589	0.653	0.610	0.398	**0.571**
既存	0.483	**0.530**	0.563	0.610	0.547	0.590	0.516	0.400	0.530

ただし，$n_{w^{(n)}|i}$ は問題 i における $w^{(n)}$ を含む小論文の個数を表す．また，α はハイパーパラメータである．

文章変換は，上記のマスキング手法でマスクした箇所を任意の問題に特徴的な語彙やフレーズに変換する生成言語モデルを訓練することで行われる．ここで生成言語モデルには，*T5 (Text-to-Text Transfer Transformer)* を用いる．T5 の訓練手順は次のとおりである．まず，各問題 $i \in \mathcal{I}$ に対する小論文データ $\{e_{ij}\}_{j=1}^{J_i}$ から，問題に固有の単語をマスクした小論文データ $\{\mathrm{M}'(e_{ij})\}_{j=1}^{J_i}$ を，マスキング手法を用いて作成する．ただし，ここではマスキングスコアを $\mathrm{m}'(w^{(n)}, \mathcal{E}_i) = \max_{l \in \mathcal{I}\backslash\{i\}} \mathrm{m}(w^{(n)}, \mathcal{E}_i, \mathcal{E}_l)$ とする．そして，マスクした小論文と問題を識別するラベルを入力として，元の文章 e_{ij} を復元するように T5 を訓練する．

2.2　文章ドメイン変換によるデータ拡張

上記で訓練した T5 を用いて，目的の問題とは異なる問題に対する小論文データのドメインを変換することで，目的の問題に対応した擬似的な採点済み小論文データを作成する．具体的には，まず目的の問題とは異なる各問題 $i \in \mathcal{I}\backslash\{p\}$ に対するマスクした小論文データ $\{\mathrm{M}(e_{ij})\}_{j=1}^{J_i}$ と目的の問題の識別ラベルを 2.1 節で訓練した T5 に入力し，マスク箇所を目的の問題に固有の語彙で補完した小論文データ $\{e'_{ij}\}_{j=1}^{J_i}$ を生成する．この変換された小論文を元の得点データと統合することで，擬似的な採点済み小論文データ $\{(e'_{ij}, y_{ij})\}_{j=1}^{J_i}$ を得る．

2.3　自動採点モデルの構築

以上の手順で拡張されたデータ $\{(e'_{ij}, y_{ij})\}_{j=1}^{J_i}$ を，目的の問題とは異なるすべての問題 $i \in \mathcal{I}\backslash\{p\}$ に対して作成し，得られたデータセットを用いて自動採点モデルを訓練する．自動採点モデルには，様々な最先端の自動採点研究でベースラインとして利用されている [3] *BERT*（*Bidirectional Encoder Representations from Transformers*）と呼ばれる深層学習モデルを用いたモデルを使用する．

3　評価実験

実データを用いて提案手法の有効性を評価した．実験には，8 つの小論文問題で構成されている ASAP（Automated Student Assessment Prize）データセットを使用した．ここでは，個々の問題ごとに提案手法で自動採点モデルを構築して性能を評価した．具体的には，問題 i に対する自動採点モデルを構築する際には，その他の問題 $\mathcal{I}\backslash\{i\}$ に対する採点済み小論文データを提案手法で変換したデータの 80% を訓練データ，20% をアーリーストッピングのための検証データとし，訓練されたモデルの性能を問題 i に対するデータセットで評価した．自動採点モデルの訓練は 30 エポックずつ行った．本実験では，この手続をそれぞれの問題 $i \in \mathcal{I}$ に対して行った．また，提案手法の性能を比較するために，訓練データを変換せずに元々の文章のまま使用して自動採点モデルを訓練する「従来手法」についても同様に実験を行った．評価指標には，予測スコアと教師スコアの一致度を測るための指標である QWK（Quadratic Weighted Kappa）を用いた．なお，提案手法におけるハイパーパラメータ α は 1 グラム，2 グラム，3 グラムでそれぞれ 1，5，7 と設定した．この実験をマスキングスコアの閾値 $\tau = 0.08, 0.2, 0.5, 0.8$ についてそれぞれ行った．

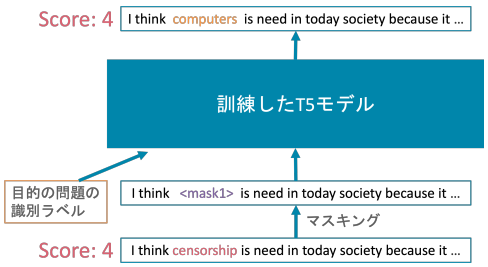

図 1　文章ドメイン変換によるデータ拡張の概念図

結果を表 1 に示す．表 1 から，問題 2 を除く全ての場合で提案手法が従来手法より高い性能を示したことがわかる．また，提案手法の中では，$\tau = 0.8$ のときに平均的な性能が最も高くなっていることが読み取れる．このことから，提案手法は問題横断型自動採点の精度改善に有効であるといえる．

4　まとめ

本研究では，文章変換技術に基づくデータ拡張を用いた問題横断型自動採点手法を提案し，その有効性を示した．今後は，ハイパーパラメータの最適化の方法論について検討していきたい．

参考文献

[1] Cancan Jin, Ben He, Kai Hui, and Le Sun. TDNN: a two-stage deep neural network for prompt-independent automated essay scoring. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 1088–1097, 2018.

[2] Nitay Calderon, Eyal Ben-David, Amir Feder, and Roi Reichart. DoCoGen: Domain counterfactual generation for low resource domain adaptation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 7727–7746, 2022.

[3] Masaki Uto. A review of deep-neural automated essay scoring models. *Behaviormetrika*, Vol. 48, No. 2, pp. 459–484, 2021.