不均衡性を考慮した深層強化学習による読影レポート生成

西埜 徹[1] 桃木 陽平[2] 谷口 友紀[1] 田川 裕輝[1] 谷口 元樹[1] 大熊 智子[1] 中村 佳児[2]
[1] 富士ゼロックス株式会社
[2] 富士フイルム株式会社
{nishino.toru, taniguchi.tomoki, tagawa.yuki, motoki.taniguchi, ohkuma.tomoko}@fujixerox.co.jp
{yohei.momoki, keigo.nakamura}@fujifilm.com

1 はじめに

急速な高齢化や読影医不足の状況下で, X 線や CT 検査画像を観察し, 発見した異常所見をレポートに記述する読影作業の業務負荷は高まっている. 中でも読影レポート作成は時間を要する業務であり, 読影医は業務中に大量の読影レポートを作成する. 読影医の業務効率化のために, 読影レポート作成自動化は急務といえる.

我々は, 読影レポート中の所見文の自動生成に取り組んでいる. 多くの読影レポート生成の先行研究 [1] では, 検査画像を入力としてレポートを直接生成する End-to-End 型の手法が提案されている. 一方我々のシステムは, 入力した検査画像から所見ラベルを出力する画像認識システム及び, 所見ラベルからレポートを生成するテキスト生成システムの 2 段階の構成をとる. 画像認識とテキスト生成システムの分割により, 各モデルを独立して最適化でき, 人手による所見ラベルの修正が可能である利点がある.

本研究では, 所見ラベルから読影レポートを生成するテキスト生成システムに焦点を当てる. このシステムは入力データの内容を人間が解釈しやすいよう自然文として出力する Data-to-Text 生成技術の一種といえよう. しかし, 医療分野のテキスト生成システム実現への課題として, (1) 医療分野のレポート生成には適切性が重視され, 出力される生成レポートには, 入力ラベルに対応する記述を過不足なく含む必要がある点 (2) 読影現場において異常所見の出現頻度には大きくバラツキがあるため, 学習データの入力ラベルの出現頻度に著しい不均衡が存在する点が挙げられる. Seq2Seq など既存のテキスト生成手法では, 学習データ中に高頻度で出現する入力ラベルの情報ばかり学習が進み, 低頻度ラベルの情報の抜け落ちが生じやすい.

本研究では, 生成レポートから逆に所見ラベルを予測し適切性を推定する Reconstruction 機構を導入し, テキスト生成システムの強化学習を行う手法を提案する. 本研究の貢献は以下の 3 点である.

- Class-Balanced Loss [2] の導入により Reconstruction 機構の性能向上を実現し, 低頻度ラベルの抜け落ちを正しく検出可能にした.
- Reconstruction 機構により算出される適切性の推定値を報酬として強化学習を行い, 低頻度ラベルの情報が抜け落ちなく含まれるようテキスト生成システムを学習した.
- Reconstruction 機構と強化学習を用いて, 学習データ量の拡張無しでの追加学習を行った.

これらの手法の導入により, 生成レポートの情報抜け落ちを低減し, 適切性向上が確認された.

表 1 読影レポートの入出力例

正しいレポート
【入力所見ラベル】「結節あり」「吸収値 (充実型)」「胸膜陥入像あり」「辺縁不明瞭」
【読影レポート】 ＜肺の部位＞ に成分径 ＜大きさ＞ の境界不明瞭な充実型結節影が見られる。結節の辺縁部には胸膜陥入像を伴う。
誤りを含むレポート
【入力所見ラベル】「結節あり」「吸収値 (充実型)」「**空洞なし**」「**石灰化像なし**」「辺縁不明瞭」
【読影レポート】 ＜肺の部位＞ に成分径 ＜大きさ＞ の境界不明瞭な充実型結節影が見られる。結節の内部には**空洞がみられる**。

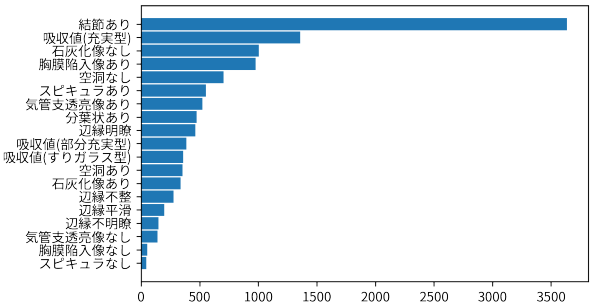

図 1 各入力所見ラベルの出現件数の一例

2 読影レポート生成タスクの特徴

医療分野におけるレポート生成には, (1) **適切性**の向上 (2) **不均衡**データへの対処の 2 点が課題となる.

適切性: 医療分野では情報の正確性が重視されるため, 入力した情報を過不足なく記述した, 適切性の高いレポートの生成が要請される. 読影レポート生成タスクでは, 入力した所見ラベルに対応する記述を含んだレポートを過不足なく出力する. 入力所見ラベルにない所見記述がレポート中に含まれるとき, もしくは入力所見ラベルに対応する記述がレポート中にない時, 生成レポートの適切性が低いと言える.

本研究では読影レポート生成タスクにおけるレポートの適切性を (1) 式で求まる適合率・再現率及び調和平均 F 値として定義した.

$$\text{Precision} = \frac{\text{所見ラベルのうち, 対応する記述がレポートに含まれる数}}{\text{生成レポートに含まれる所見記述の総数}}$$
$$\text{Recall} = \frac{\text{所見ラベルのうち, 対応する記述がレポートに含まれる数}}{\text{入力した所見ラベルの総数}} \quad (1)$$

例として, 読影レポート生成タスクにおける入力所見ラベルと出力読影レポートの組を表 1 に示す. 上の例では, 4 件の入力所見ラベルの情報がレポート内に過不足なく記載されており, 適切性が 1.0 と算出される. 下の例では, 入力所見ラベルのうち「空洞なし」「石灰化像なし」の情報がレポートに含まれておらず, 逆に「空洞

あり」所見ラベルに相当する記述が含まれるため, 適合率は 0.75, 再現率は 0.6 となり, 従って適切性は 0.67 と算出される.

不均衡性: 一般に医療分野の機械学習では, データの不均衡問題が課題となる. 読影現場においても, 各異常所見の出現頻度には極めて大きな偏りがある. 故に, 学習データ内のレポートに付与される所見ラベルの出現頻度も極めて不均衡となっている.

例として, 本タスクで使用する読影レポートデータセットの各所見ラベルの出現頻度を図 1 に示す.「胸膜陥入像あり」のラベルは学習データ中 977 件のレポートに付与されているが,「胸膜陥入像なし」のラベルは 52 件のレポートに付与されたに過ぎない. このデータセットを用いて従来手法のの Seq2Seq 等で学習した場合, 高頻度で出現する所見ラベルの情報ばかり学習が進み, 低頻度ラベルの情報の抜け落ちが生じ適切性が低下する問題が生じやすい.

本研究の課題は, 不均衡な分布の入力所見ラベルを持つ学習データを使用しても, 適切性の高いレポートが生成されるように学習する必要のある点である.

3 提案手法

読影レポート生成タスクは Data-to-Text タスクの一種で, 所見ラベル $x = \{x_1, x_2 ... x_n\}$ を入力に与え単語列 $y = \{y_1, y_2, ... y_n\}$ を出力するモデル $P(y|x)$ を学習するタスクと定義できる.

3.1 深層強化学習によるテキスト生成モデル学習

Data-to-Text タスクでは機械翻訳・要約等で広く採用される Attention 機構付き Seq2Seq モデルが広く用いられる [3]. 一般に Seq2Seq モデルの学習にはクロスエントロピー損失関数を使用し, 予測した単語列 $\hat{y} = \{\hat{y}_1 ... \hat{y}_t\}$ と正解となる単語列 $y = \{y_1 ... y_t\}$ が一致するよう学習する. しかしテキスト生成モデルでは, 正解文と異なる単語列だが正解文と同じ意味を持つ文章が生成される例が多い. このような文章では, クロスエントロピー損失関数は損失を過大に算出してしまう.

本研究では, テキスト生成モデルの学習に強化学習を導入し, 生成レポートの適切性が向上するよう学習を行う. 強化学習の利点は, 適切に報酬を設計すれば, 単語列が正解と一致せずとも報酬が最大化される文章が生成されるようにモデルを学習できる点にある. キャプション生成による読影レポート生成 [4] や読影レポート要約タスク [5] では ROUGE 等の文章間自動評価尺度と共に適切性を報酬として強化学習することで生成文章に含まれる事実の適切性向上が報告されている.

テキスト生成モデルの強化学習手法には REINFORCE [6] が広く採用されている. REINFORCE ではサンプリングにより生成された文章 $\hat{y}^s$ に対し報酬 $r(\hat{y}^s)$ を算出し, (2) 式に従い損失関数の勾配を近似する. b は任意の関数で定義されるベースラインである.

$$\nabla_\theta L_\theta \approx -\nabla_\theta \log P_\theta(\hat{y}^s)(r(\hat{y}^s) - r_b) \qquad (2)$$

本研究では, REINFORCE を改良して安定的に学習可能とした Self-Critical Sequence Training (SCST) [7] を採用する. SCST での損失関数は (3) 式に表される. SCST では貪欲法で生成された 文章 $\hat{y}^g$ に対する報酬 $r(\hat{y}^g)$ をベースライン r_b として定義している.

$$\nabla_\theta L_\theta \approx -\nabla_\theta \log P_\theta(\hat{y}^s)(r(\hat{y}^s) - r(\hat{y}^g_t)) \qquad (3)$$

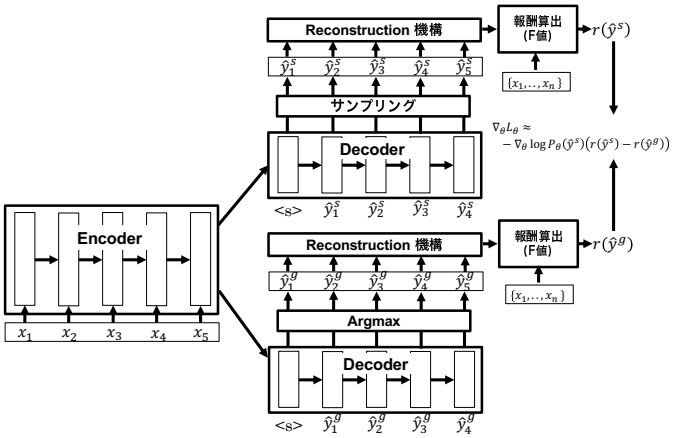

図 2 Reconstruction 機構を用いた深層強化学習

3.2 Reconstruction 機構による適切性の推定

適切性の高いテキスト生成モデルの強化学習には, 適切性を推定する関数を予め設計し, そのスコアを報酬として与える必要がある. Data-to-Text タスクでは生成文章から逆に入力データを復元する Reconstruction 機構を導入し, 生成モデルと同時に学習する手法の有効性が報告されている [8] . Reconstruction 機構で復元された予測入力ラベルが実際の入力ラベルと近いほど, 生成レポートには入力所見ラベルの情報が正しく含まれ, 適切性が高いレポートと言える. 本研究では, Reconstruction 機構により得られた予測入力ラベルの入力ラベルに対する F 値を生成レポートの適切性の推定値と定義し, 報酬として利用する. Reconstruction 機構を用いた強化学習モデルを図 2 に示す.

Reconstruction 機構は読影レポートの所見文から逆に所見ラベルを復元するよう学習される. この復元処理はマルチラベルの文書分類タスクとして扱えるため, 本研究では Reconstruction 機構として Hierarchical Attention Network (HAN) [9] を採用した. HAN はテキスト生成モデル学習に用いたデータセットを使用して学習する. しかし, 2 章で触れたように当該データセットには所見ラベルの著しい不均衡性があり, 分類精度に影響を及ぼす. 本研究では不均衡性のあるデータセットに対して効率的に学習を行う目的で, HAN 学習時の損失関数として Class-Balanced Loss (CBL) [2] を導入する. Class-Balanced Loss は (4) 式に示す重みを各クラス毎に与える損失関数である.

$$w_i = (1 - \beta)/(1 - \beta^{n_i}) \qquad (4)$$

β はハイパーパラメータ, n_i はサンプル数であり, 低頻度ラベルに高い重み付けが与えられる.

3.3 疑似データ生成による追加学習

低頻度ラベルを入力とした場合の学習を促進するため, 本研究では作成した疑似データによる追加学習を行う. 機械翻訳では, 逆翻訳を用いて単言語コーパスから擬似的対訳コーパスを得るデータ拡張を行い, 対訳コーパスの増加無しに翻訳精度が向上している [10] . Data-to-Text タスクでは入力データと出力テキストのデータ拡張の難易度に差があり, 入力データは比較的容易にデータ拡張できるが, 出力テキストのデータ拡張コストは高い. ゆえに入力-出力ペアのデータ拡張は高コストであり, 困難である.

Kedzie ら [11] は Data-to-Text タスクにおいて生成文章をパースして入力ラベルを得るデータ拡張を実施し, 生成文の適切性を向上させている. データ拡張が容易な入力データからテキスト生成モデルを用いて出力テキストを生成することで低コストで入力-出力ペアのデータ拡張が実現できる. しかし, 本研究で扱う読影データセットでは, 読影レポート中に出現するフレーズの頻度にも著しい不均衡性がある. これらの手法と同様にテキスト生成モデルによりデータ拡張を行っても, 読影レポート中に頻出するフレーズの量が増えるばかりでデータセットの不均衡性を増大させる結果になる.

本研究では, 所見ラベルと読影レポートのペアをデータ拡張をせずに, 所見ラベルのみをデータ拡張して追加学習する手法を提案する. データ拡張及び追加学習の手続きを以下に示す.

(1) **所見ラベルのサンプリング** 学習データ中の所見ラベル列 $x = \{x_1, ...x_n\}$ に対して, 新たに所見ラベル x_0^s を 1 件追加した疑似所見ラベル列 $x^s = \{x_1, ..., x_0^s, ..., x_n\}$ を作成する. 追加する所見ラベルは x に含まれない所見ラベルの集合からランダムサンプリングにて抽出する. それゆえ, 学習データでは出現頻度の低い所見ラベルが疑似所見ラベル列には比較的高頻度で含まれ, 不均衡性が緩和される.

(2) **レポート生成及び所見ラベルの推定** 疑似所見ラベル列 x^s からテキスト生成モデルでレポート $\hat{y}^s$ を生成し, Reconstruction 機構で $\hat{y}^s$ に対応する予測所見ラベル $\hat{x}^s$ を推定する.

(3) **強化学習** 疑似所見ラベル x^s 及び予測所見ラベル $\hat{x}^s$ から F 値を算出し, これを報酬として (3) 式に従いテキスト生成モデルを学習する.

この追加学習により, 読影レポートのデータ数を増やすことなく, 入力所見ラベルのみのデータ拡張から生成レポートの適切性を改善可能となる. 疑似所見ラベルは不均衡性が緩和するため, より低頻度ラベルに対する学習を促進すると期待できる.

3.4 強化学習の損失関数

適切性の推定値のみを報酬とした強化学習だけでは流暢性の高い文章を生成するようにモデルを学習できない. ゆえにテキスト生成モデルの学習時には強化学習の損失関数に加え, クロスエントロピー損失関数も合わせて学習することで適切性の高く, かつ流暢なテキスト生成モデルの学習を行う.

本研究では, テキスト生成モデルは SCST の損失関数 L_{rl}, 追加学習の損失関数 L_{aug} 及びクロスエントロピー損失関数 L_{xent} を用いた (5) 式に示される損失関数により学習する.

$$L_{all} = \lambda_{rl}(L_{rl} + \lambda_{aug}L_{aug}) + (1 - \lambda_{rl})L_{xent} \quad (5)$$

MIXER [12] 同様に, ハイパーパラメータ $\lambda_{rl} \in [0, 1]$ は学習が進むにつれ強化学習の損失関数を重視するよう動的に変更する. n_p エポック目における $\lambda_{rl}(n_p)$ は全学習エポック数 n_{all} を用い, $\lambda_{rl}(n_p) = \lambda_{const}n_p/n_{all}$ 式に従い決定される.

表 2 読影レポートデータセットの統計量

	レポート件数	平均ラベル数	平均単語長
学習データ	3,637	4.71	27.5
開発データ	418	9.46	52.7
評価データ	399	9.49	51.4

表 3 ハイパーパラメータの一覧

共通			
語彙数	339	Optimizer	Adam
所見ラベル数	57	バッチサイズ	32
Dropout rate	0.5	勾配クリッピング閾値	2.0
Reconstruction 機構			
学習エポック数	100	学習率	0.0001
1 文書の最大文数	8	最大文長	40
隠れ層ユニット数	256	単語埋込層次元	300
CBL β	0.999		
テキスト生成モデル			
学習エポック数	100	学習率	0.0005
ラベル埋込層次元	32	単語埋込層次元	64
最大入力長	20	ビームサーチ幅	5
最大出力長	100	隠れ層ユニット数	128

表 4 Reconstruction 機構の分類精度評価

	Precision	Recall	F 値
CNN	90.7	76.2	82.9
RNN	90.2	73.9	81.2
HAN	91.1	83.4	87.1
HAN + CBL	**91.4**	**87.3**	**88.6**

4 実験

4.1 実験設定

実験データ: 本研究では, 病院から提供された肺結節に関する読影レポートのうち, 所見文の部分をデータセットとして使用する. Data-to-Text タスクのデータセットとして利用するために, 肺癌取扱い規約 [13] を参考に定義した計 57 種の所見ラベルを人手作業で所見文に付与した. E2E NLG Challenge [3] 等の Data-to-Text タスクに倣い, 表記の多様性が高い肺の部位に関するフレーズ及び数値はトークンに置換している. 読影レポートと付与した所見ラベルの例を表 1 に示す.

また, 開発データ・評価データが不均衡データの場合, 適切な評価結果が得られない. 適切な評価のために, 開発・評価データは不均衡性が減じるようにサンプルを選択して作成している. 学習・開発・評価データ間では同一入力所見ラベルの組み合わせを持つデータは含まれない. データセットの統計量を表 2 に示す.

実験条件: テキスト生成モデルには Attention 機構付き Seq2Seq モデルを使用する. モデルの定義及びハイパーパラメータを表 3 に示す. ハイパーパラメータは開発データにおける探索により決定した.

4.2 実験結果

Reconstruction 機構単体の性能評価: まず Reconstruction 機構が正しくレポートの適切性を推定可能か検証するために, Reconstruction 機構単体の分類精度を評価した. 比較対象として CNN 及び RNN を検討した. 表 4 に各モデルの分類精度を示す. (1) CNN や RNN に対し HAN が本タスクで適切 (2) HAN の学習に Class-Balanced Loss を適用することで不均衡学習データを用いる分類タスクで性能向上する の 2 点がわかる. この結果より, Reconstruction 機構として HAN + Class-Balanced Loss を採用した.

テキスト生成モデルの自動評価: テキスト生成モデルの簡便な評価指標として, Rouge-L・BLEU-2 及び学習した Reconstruction 機構により得られる適切性の推定値の F 値で評価を行った. 比較対象として (1) ベー

表 5　テキスト生成モデルの自動評価

	Rouge-L	BLEU-2	Reconstruction
Seq2Seq	65.8	66.4	72.2
Seq2Seq + DA	63.9	65.4	71.5
RL(BLEU)	66.0	65.8	74.8
RL(R)	**66.8**	65.9	75.2
RL(R) + DA	65.6	**67.0**	**76.7**

表 6　テキスト生成モデルの人手評価

	適切性			流暢性
	Precision	Recall	F 値	2 段階
Seq2Seq	88.4	76.2	81.2	**1.0**
RL(R) + DA	**91.1**	**76.8**	**82.7**	0.98

スラインの Seq2Seq モデル (2) テキスト生成モデルを用いて入力-出力ペアのデータ拡張を実施する追加学習手法 (Seq2Seq + DA) (3) BLEU-2 を報酬とする強化学習 (RL(BLEU)) (4) Reconstruction 機構の適切性推定値を報酬とした強化学習 (RL(R)) (5) RL(R) に加え, 3.3 節の追加学習を適用した強化学習 (RL (R) + DA) を比較した.

結果を表 5 に示す. Rouge-L での評価では RL(BLEU) が優れている一方, BLEU-2 及び Reconstruction 機構を用いた評価では Reconstruction 機構の適切性推定値を報酬として学習した提案手法の RL(R) + DA が最も精度が高い. この結果から, BLEU 等の評価尺度よりも, 適切性の推定値を報酬とした Reconstruction 機構を用いた強化学習が有効と言える. よって, 適切性の高いレポートを生成するには, 適切性と相関の高い報酬設計が必要と言えよう. また, 入力-出力ペアをデータ拡張した Seq2Seq に対し, RL(R) + DA の適用により Reconstruction 評価尺度が向上しており, 本研究の追加学習手法が有効と言える.

テキスト生成モデルの人手評価: 生成レポートのより正しい質の評価のために 50 件のレポートに対し人手評価を実施した. 評価者にはレポートに含まれる所見の適切性・ レポートの流暢性の 2 点で評価を依頼した. 適切性の評価には入力所見ラベルと生成レポート中の所見の合致・相違件数のカウントを依頼し, (1) 式を元に適合率・再現率及び F 値 を算出した. また, 流暢性は 2 段階評価を実施した. BLEU-2 と人手評価の相関係数は 0.278 と低く, 同様に Reconstruction 機構を用いた評価と人手評価の相関係数も 0.767 であり正確な評価は得られない. 人手評価の実施でより正しい評価結果が得られると期待される.

人手評価結果を表 6 に示す. 提案手法は Seq2Seq に対して適切性が改善され, 流暢性では Seq2Seq と遜色ない評価が得られている. この結果より, 提案手法によって流暢性を犠牲にすることなく生成レポートの適切性を改善できたと言える.

5　考察

エラー分析: 表 7 に, 読影レポート 50 件における Reconstruction 機構のエラー分析結果を示す. 「○○あり」と「○○なし」のラベルを取り違える否定表現のエラーが最も多い. ラベル間には ある 2 つのラベルは同時に出現しない排他関係, あるラベルは他の所見ラベルの詳細である階層関係にあるペアが多い. しかし, これらの関係は Reconstruction 機構では考慮されておらずエラーに繋がっている. また, 追加学習での入力所見ラベルのデータ拡張時にもこれらの関係を考慮していないため, 排他関係のラベルが同時に含まれる等臨床的に存在しない所見ラベルの組み合わせが生成されている可能性がある. 以上より, 臨床的に適切な所見ラベルの組み合わせを生成する制約が必要と推察できる.

表 7　Reconstruction 機構のエラー分析

要因	件数
否定表現	22
所見の度合い	20
詳細な所見	8
その他	4
ラベル推定ミス総計	54

表 8　生成されたレポートの例

【入力所見ラベル】「結節あり」「吸収値 (充実型)」「**胸膜陥入像なし**」(中略)「辺縁明瞭」
【Seq2Seq の出力】＜肺の部位＞に＜大きさ＞の充実型結節を認めます。境界明瞭で (中略) **胸膜陥入像を伴っています**。
【提案手法の出力】＜肺の部位＞に＜大きさ＞の充実型結節を認めます。境界は明瞭ですが (中略) **胸膜の陥入像は認めません**。

生成例: 表 8 に生成レポートの例を示す. Seq2Seq では入力ラベル「胸膜陥入像なし」に矛盾する, 胸膜陥入像が存在すると記述された出力が得られた. 一方提案手法では「胸膜陥入像なし」に対応する記述 (**太字部**) が正しく記載されている. 「胸膜陥入像なし」は学習データでは出現頻度が低いラベルであり, 提案手法では低頻度入力所見ラベルに対して対応する記述を適切に含んだレポートが生成可能となったと言えよう.

6　おわりに

本研究では, 適切性の推定値を報酬とした強化学習及び追加学習手法を提案し, 入力所見ラベルの不均衡性の高い読影レポート生成タスクにおいて効果を示した. 読影レポート生成に限らず多くの Data-to-Text タスクの学習には不均衡性が課題であると予想され, 不均衡性なデータセットから効果的にテキスト生成モデルを学習できる本手法は広く有用な手法と期待できる.

今後の課題は (1) 所見ラベル間の関係を考慮した Reconstruction 機構の導入とデータ拡張手法の導入 (2) 本手法は学習が不安定のため, 安定的に学習可能となる手法の探索 の 2 点が挙げられる.

参考文献

[1] Baoyu Jing, Pengtao Xie, and Eric Xing. On the automatic generation of medical imaging reports. In *ACL 2018*.

[2] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR 2019*.

[3] Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. Findings of the E2E NLG Challenge. In *INLG 2018*.

[4] Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, and Marzyeh Ghassemi. Clinically accurate chest x-ray report generation. *arXiv:1904.02633*, 2019.

[5] Yuhao Zhang, Derek Merck, Emily Bao Tsai, Christopher D Manning, and Curtis P Langlotz. Optimizing the factual correctness of a summary: A study of summarizing radiology reports. *arXiv:1911.02541*, 2019.

[6] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 1992.

[7] Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. Self-critical sequence training for image captioning. In *CVPR*, 2017.

[8] Sam Wiseman, Stuart Shieber, and Alexander Rush. Challenges in data-to-document generation. In *EMNLP*, 2017.

[9] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *NAACL*, 2016.

[10] Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Unsupervised machine translation using monolingual corpora only. *ICLR*, 2018.

[11] Kedzie Chris and Kathleen McKeown. A good sample is hard to find: Noise injection sampling and self-training for neural language generation models. In *INLG*, 2019.

[12] Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *ICLR*, 2016.

[13] 特定非営利活動法人 日本肺癌学会. **肺癌取扱い規約 第 *8* 版**. 金原出版株式会社.