

ポスター6

画像情報・生体信号処理

2019年11月23日(土) 14:40 ～ 15:40 ポスター会場1 (国際展示場 展示ホール8)

[3-P1-3-04] Encoder-Decoder Modelによる最適 MRIプロトコル予測システムの開発

○谷川原 綾子^{1,2}、上杉 正人³、安渡 大輔⁴、タキンキン⁵、遠藤 晃⁶、藤田 勝久⁴（1. 北海道科学大学保健医療学部, 2. 北海道大学大学院保健科学研究院, 3. 北海道情報大学医療情報学部, 4. 北海道大学病院医療技術部, 5. 北海道大学病院放射線診断科, 6. 北海道大学病院医療情報企画部）

キーワード：Encoder-Decoder Model, Brain MRI examination, Natural Language Processing

【目的】MRI検査における最適撮影プロトコルは、放射線科医が患者の臨床所見を基に決定している。ディープラーニング技術は診療の質の向上に向けた最適撮影プロトコルの決定支援に貢献できる可能性がある。本研究では、患者の臨床所見に基づき Encoder-Decoder Modelを用いた最適な MRIプロトコルの予測システムの開発を行い、その精度評価を行った。

【方法】北海道大学病院から2010年1月から2017年12月の期間において6,831件の造影脳 MRI、10754件の非造影脳 MRIの検査申込書を収集した。これらのデータには、検査依頼医が記載した MRI検査目的、臨床所見と治療経過に加え、放射線科医より指示された MRIプロトコルが記載されている。なお、MRIプロトコルは MRIの画像の種類、撮像断面および厚さの組み合わせ(FLAIR sag 3mmなど)とする。解析方法は、依頼医からの記述を bigramで切り出し、Encoder-Decoder Modelに入力後、最適 MRIプロトコルを予測した。造影、非造影検査のデータセットをそれぞれ10等分に分割し、訓練9、テスト1の比率で交差検定を行った。評価は放射線科医の指示に対して予測された MRIプロトコルの完全一致率、過大評価率、過小評価率、部分一致率、完全不一致率を算出した。

【結果】造影検査における完全一致率、過大評価率、過小評価率、部分一致率、完全不一致率は、それぞれ13.0%、39.3%、2.3%、41.4%、4.0%であった。非造影検査では、それぞれ、4.4%、50.0%、1.8%、42.6%、1.4%であった。造影、非造影検査共に、撮像頻度の高い MRIプロトコルは予測精度が高い傾向が見られた。今後は、ハイパーパラメータを調整し、精度のさらに向上を目指したいと考えている。

Encoder-Decoder Model による最適 MRI プロトコル予測システムの開発

谷川原綾子^{*1,2}、上杉正人^{*3}、安渡大輔^{*4}、タキンキン^{*5}、遠藤晃^{*6}、藤田勝久^{*3}

*1 北海道科学大学保健医療学部、*2 北海道大学大学院保健科学研究科、*3 北海道情報大学医療情報学部
*4 北海道大学病院医療技術部、*5 北海道大学病院放射線診断科、*6 北海道大学病院医療情報企画部

Development of Automatic Prediction System for Optimal MRI Examination Protocol Using Encoder-Decoder Model

Ayako Yagahara^{*1,2}, Masahito Uesugi^{*3}, Daisuke Ando^{*4}, Khin Khin Tha^{*5}, Akira Endoh^{*6}, Katsuhisa Fujita^{*3}

*1 Hokkaido University of Science, *2 Faculty of Health Sciences, Hokkaido University,

*3 Hokkaido Information University, *4 Department of Radiological Technology, Hokkaido University Hospital,

*5 Department of Diagnostic Imaging, Hokkaido University Hospital,

*6 Department of Medical Informatics, Hokkaido University Hospital

The purpose of our study was to predict the optimal brain MRI protocols using encoder-decoder model. We collected 17,585 request forms, in Japanese language, for contrast-enhanced (CE) and non-CE MRI of the brain from Hokkaido University Hospital (from Jan 2010 to Dec 2017). These data included descriptions regarding the purpose of MRI, summary of clinical findings and treatment progresses, and the MRI sequence protocols ordered by radiologists. We then created bi-gram tokens in the description. Next, we applied the encoder-decoder model to predict the optimal CE and non-CE MRI protocols. The learning parameters were as follows: iteration, 50; embedding, 100 dimensions. We performed 10-fold cross validation using each of the ten ordered versions of our dataset. We used 'match,' 'overpredict,' 'underpredict,' 'partial match,' 'mismatch' as the performance index. For CE MRI, the results of 'match,' 'overpredict,' 'underpredict,' 'partial match' and 'mismatch' were 13.0%, 39.3%, 2.3%, 41.4% and 4.0%, respectively. The results for non-CE MRI were 4.4%, 50.0%, 1.8%, 42.6% and 1.4%, respectively. The prediction accuracy of MRI protocols, which were used routinely or frequently, tended to be higher. On the other hand, the accuracy of infrequent protocols was low. The percentage of successfully predicted protocols, which included 'match' and 'overpredict', was 52.3% for CE MRI and 54.4% for non-CE MRI.

Keywords: Encoder-Decoder Model, Brain MRI Examination, Natural Language Processing

1 緒論

MRI 検査において、最適な MRI 撮影プロトコルは、放射線科医が患者の臨床所見を基に決定している。この撮影プロトコルの決定には、知識と経験を要する。それは、T1 強調像や拡散強調像など多種のコントラストの画像が存在し、スライス厚や撮影断面の方向も疾患によって適切に選択しなければならない。加えて、1つの MRI 画像を収集するのに数分かかり、決められた検査時間内で撮像可能なプロトコルを選定する必要もあるためである。ルーチンプロトコルを設定している医療機関もあるが、検査依頼書から患者や疾患に合わせた適切なプロトコルを撮像することができれば、更なる画像診断の質の向上や最良な治療方針の選択に貢献できる。

近年、Deep learning の技術向上が進み、医療応用の研究も盛んに行われている。この Deep learning 技術を MRI 検査における最適撮影プロトコルの決定支援にも適用できる可能性がある。特に、言語処理における Deep learning 技術である Encoder-Decoder Model¹⁾は主に翻訳で用いられているが、翻訳したい文章を検査依頼書の記述、翻訳文章を MRI プロトコルとすると、MRI プロトコルが自動で予測出来るのではないかと考えた。そこで、本研究では、患者の臨床所見に基づき Encoder-Decoder Model を用いた最適な MRI プロトコルの予測システムの開発を行い、その精度評価を行った。

2 Encoder-Decoder Model

Encoder-Decoder Model とは、機械翻訳に応用されている手法で、翻訳元の単語列($x_1, x_2, \dots, x_m$)を入力とし、翻訳先の単語列($y_1, y_2, \dots, y_n$)を出力としてマッピングする手法である。具体的には、Encoder 部に入力された単語 x_t は Embed_x にて分散表現に変換され、LSTM(Long Short Term Memory)ブロックに入力される。LSTM ブロックでは、その時点までの文脈情報を次の LSTM ブロックに渡す。Encoder 部にて<eos>が読み込まれると Decoder 部にて翻訳された単語 y_1 が出力される。この y_1 が次の入力となる。Decoder 部において、入力された単語 y_t は Embed_y により分散表現に変換され、LSTM ブロックに入力される。その後、LSTM の出力が線形変換され、softmax 関数により翻訳言語の全単語に対する出力確率値が求められ、その最大の単語が出力される(図 1)。

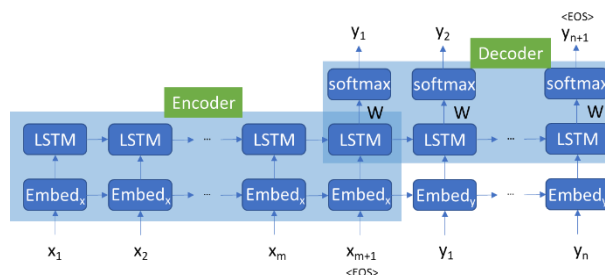


図 1 Encoder-Decoder Model の概要

3 方法

北海道大学病院において 2010 年 1 月から 2017 年 12 月に実施された 6,831 件の造影脳 MRI (CE)、10754 件の非造影脳 MRI (non-CE) の検査申込書を収集した。これらのデータには、検査依頼医が記載した MRI 検査目的、主な臨床所見と治療経過に加え、放射線科医より指示された複数の MRI プロトコルが記載されている。ここで、MRI プロトコルは MRI の画像の種類、撮像方向を主とし、必要に応じて FOV(Field of view)、マトリックスサイズおよびスライス厚さ等を組み合わせたもの(例: FLAIR sag 3mm)とした。FOV、マトリックスサイズ、スライス厚について、下垂体や小脳橋角部(C-P)では、部位ごとに独自のパラメータが決定されているため、プロトコルに部位名を付与することでそれを示した。データセットの生成について、検査目的等の記述を1文ずつ抽出し、bigram で文字列を切り出した後、放射線科医が指示した複数の MRI プロトコルを付与させた。Chainer により実装され、Attention が組み込まれた Encoder-Decoder Model にデータセットを入力し、学習を行った。学習は CE と non-CE を分けて実施した。学習パラメータは、埋め込みベクトル数 100、学習回数は 50 回とした。

評価について、CE、non-CE のデータセットをそれぞれ 10 等分に分割し、訓練データ 9、テストデータ 1 の比率で交差検定を行った。テストデータは、学習モデルに検査目的等の記述を1文ずつ入力し、MRI プロトコルを推測させた。1検査において複数文の記述がある場合、各文で出力された全ての MRI プロトコル統合し、これを該当検査における予測結果とした。評価は放射線科医の指示と予測された MRI プロトコルと比較し、完全一致率、過大評価率、過小評価率、部分一致率、完全不一致率を CE と non-CE について、それぞれ算出した。本研究を実施するにあたり、北海道大学病院倫理委員会の承認を得た。

4 結果

1 検査当たりの検査目的等の記述の文字数は CE で平均 132.0±89.7 字、non-CE で 94.7±69.2 字であった。全部で 160 の MRI プロトコルが存在し、正解データの 1 検査当たりの平均プロトコル数は CE で 5.7±1.6、non-CE で 4.2±1.2 であった。またプロトコルの組み合わせのパターンは CE で 1461、non-CE で 952 であった。

CE における完全一致率、過大評価率、過小評価率、部分一致率、完全不一致率は、それぞれ 13.0%、39.3%、2.3%、41.4%、4.0%であった。Non-CE MRI では、それぞれ、4.4%、50.0%、1.8%、42.6%、1.4%であった。表 1 に結果例を示す。

5 考察

Encoder-Decoder Model における、完全一致率は CE、non-CE とともに 13%、4.4%であった。完全一致となった MRI プロトコルの組み合わせの特徴として、そのプロトコルの実施頻度が高い傾向にあった。また、頻度の低いプロトコルの組み合わせでも、個々のプロトコルの全体の撮影頻度が高い場合も存在した。このことから、ルーチン検査やどの疾患でも撮影される MRI プロトコルが含まれている検査では一致となる可能性が高くなると考える。完全一致率が高い値を得られなかったが、放射線科医の指示を全て網羅することを考えた場合、予測精度は完全一致率と過大評価率を合わせて CE は 52.3%、non-CE は 54.4%となり、半数程度は診療に有用な予測が可能であると考ええる。

本研究の限界として、MRI の撮影プロトコルは標準化され

ていないため、同じ疾患でも放射線科医師の MRI プロトコル選択や診療科医師の要求によってばらつきが見受けられた。これが、完全一致を低下させ、過大評価や過小評価、部分一致率の増加に影響している可能性もあると考えられた。

表 1 CE における結果の例

判定	検査目的等の記述	予測	正解
完全一致	2009 年 12 月両側胸水貯留にて精査入院。2010 年 1 月 VATS 肺生検にて MPM と診断。3 月 17 日ケモ目的に入院。開始前メタスクリーニングをお願いします。	T1Ax GdAx GdFLAIRAx FsGdCor	T1Ax GdAx GdFLAIRAx FsGdCor
	14、5 年前からの顔貌の変化:3 年前からサンドスタチンにて治療中:成長ホルモン等は徐々に低下 腫瘍も徐々に縮小	下垂体 T1sag 下垂体 T1Cor 下垂体 T2Cor 下垂体 Gdsag 下垂体 GdCor	下垂体 T1Cor 下垂体 T2Cor 下垂体 Gdsag 下垂体 GdCor
	左踵悪性黒色腫にて入院中の患者。PET/CT にて鼠径・外腸骨領域リン節転移の可能性あり。脳転移検索必要。	T1Ax、 GdAx GdFLAIRAx FsGdCor	T1Ax T1cor GdAx GdFLAIRAx FsGdCor
	上記に対し MTX 化学療法を 3 コース及び 30Gy/15Fr 施行後。前回は 3T で撮像しており、今回も 3T での撮像(横断像、矢状、冠状断)を御願います。7/26~30 日の空いてる際に撮像を御願います。	T1Ax T2Ax FLAIRAx FLAIR_cor GdAx GdCor	T1Ax T2Ax DWI(1000) GdAx GdFLAIRAx FsGdCor Gd_Sag
完全不一致	4 月 15 日発症、聴力低下あり。左聴神経鞘腫。天使病院耳鼻咽喉科にて指摘。4 月 21 日、当院脳神経外科へ紹介受診。同日当科紹介となりました。治療のため、MRI での精査をお願いします。Fusion で使用します。T1、T2、Gd-T1、FLAIR Axi。可能であれば、CISS の追加もお願いします。	下垂体 T1Cor 下垂体 T2Cor 下垂体 Gdsag 下垂体 GdCor	T2Ax GdCISS C-PT1Ax C-PT2Ax C-PFLAIRAx C-PT1Cor C-P_GdAx

参考文献

1) K. Cho, B. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1724-1734.