

深層学習を用いた医療機器に表示される数値の認識モデルの提案

Medical Numerical Record Recognition using Deep Learning

大槻 涼^{1*} 杉山 治² 松村 耕平³ 多田 昌裕⁴ 野間 春生³ 黒田 知宏²
 Ryo Otsuki¹, Osamu Sugiyama², Kohei Matsumura³, Masahiro Tada⁴,
 Haruo Noma³, Tomohiro Kuroda²

¹ 立命館大学大学院 ² 京都大学医学部附属病院 ³ 立命館大学 ⁴ 近畿大学

¹ Grad. Sch. of Ritsumeikan University, ² Kyoto University Hospital, ³Ritsumeikan University
⁴ Kindai University

Abstract: In this study, we propose a numerical record recognition method from the image on medical equipments. A medical numerical record recognition can be divided into two processes, a digit recognition on the medical equipments and a numerical record recognition concerning the meaning of each digit. We designed a digit recognition as a partial numeric image recognition and adopt Faster R-CNN model for the process. While, we designed a numerical record recognition as a digit sorting method, which rearranges out-of-order numeric strings into the templated numeric strings based on the recognized digits and their positions from the Faster R-CNN. In this paper, we focused on designing the latter process and report the results of the first recognition trial with date and step values on the pedometers.

1 はじめに

近年、医療現場への情報機器の導入が急速に進み、日常の診療業務において幅広く用いられている。しかし、現在導入されている医療機器は、情報の収集と確認を人がその機器に表示されている数値に対して、目視で行うことに頼っている。そのため、収集時の人的エラーや情報の確認ミスを原因とする医療インシデントが増加する傾向を招いている。医療インシデント削減のために医療現場におけるダブルチェックが重要であると言われており [1]、看護師同士のダブルチェックが行われているが、2人によるダブルチェックはお互いを信用しすぎ、ダブルチェックが疎かになってしまう場合があり、インシデントの削減にならず、より作業を多忙にし、別のインシデントを引き起こす原因になることがある [2]。

そこで本研究では、人と機械によるダブルチェックを提案する。提案するようなダブルチェックを行う上で、システムが機器からの情報を取得する方法は、NFCなどの近距離通信や、インターネット通信を用いて情報取得する方法や、機器に表示された情報を読み取る方法が考えられる。本研究では、人は目視で機器からの情報取得を行なっているという観点から、機器が通信



図 1: 数値が表示された医療機器画像の例 (数値は後から付け加えたものである)

機能を有していなくとも、数値の表示さえできていれば情報を取得できる技術の開発を目指す。これが実現できれば、医療機器を外部に出力するインターフェースがなくとも、人間向けの医療機器の情報表示部分から、画像処理によって数値を認識して自動的に看護師に提示できるようになり、人と機械のダブルチェックが可能となる。

システムが機器の画像から数値を取得するには3つの処理が必要となる。1つ目は画像の中から数値の位置を取得する処理、2つ目は個々の数値の値が何であるかを識別する処理、3つ目は数値が何を表しているのか数値の意味を取得する処理である。

画像から数値を認識する手法として、Ian et al. は深層学習のモデルの1つである Convolutional Neural Network (以下、CNN とする) を用いて、Google Street

*連絡先: (立命館大学大学院)
 (滋賀県草津市野路東 1 丁目 1-1)
 E-mail:rotsuki@mxdlab.net

View に写った数字の認識を行なっている。この手法では、道路や家屋の画像に含まれた数字が 0-9 の何であることを認識することに成功している [3]。

また、Pan et al. は店舗の看板画像をウィンドウごとに画像として切り出し CNN によって特徴量を抽出し、その特徴量を Recurrent Neural Network (以下、RNN とする) を用いて単語に復元している [4]。

しかし、既存手法では、数字認識は多く研究されているが、数値取得の 3 処理目に挙げた、数値の意味認識はあまり研究されておらず、画像内の数値表示部分を切り出した画像に対してだけ、数値の意味認識が可能であるなど限定的である。

本研究では、数値取得の 1、2 番目の処理を部分数字認識処理と呼び、3 番目の処理を数字意味理解処理と呼ぶ。本研究では特に、既存研究で深く研究されていない、数字意味理解処理に着目し、この処理を、意味を持った数値列への並べ替え問題と定義することで、医療機器に表示された数値情報の取得を目指す。

2 提案システムの設計

提案するシステムは、医療機器の画像から数値を情報として取得するシステムである。提案するシステムの処理の流れを図 2 に示す。提案システムではまず、入力された画像から、図 2 に示す、部分数字認識処理において個々の数値の値と位置座標を出力する。部分数字認識処理は Faster R-CNN[5] を用いて設計される。処理としては、入力された画像から Region Proposal Network によって数値の場所を特定し、その場所の Feature Map を用いて CNN で値が 0-9 の何であるか識別する。

次に、部分数字認識処理より出力された数値と位置座標のセットを入力とし、図 2 に示す、数字意味理解処理で数値を目的の情報へと変換する。数字意味理解処理は Long-Short Term Memory (以下、LSTM とする) を用いた Encoder-Decoder モデル [6] と数値列を目的の情報に変換する変換器によって設計される。処理としては、入力された数値を位置座標を元に、数値が何を示しているかを認識し、目的の数値列に並べ替える。決められた数値列から目的の情報に変換する変換器に、出力された数値列を入れることで、目的の情報が出力される。

本論文では、部分数字認識処理から数値の値と位置座標が正しく得られたと仮定して、数字意味理解処理の設計を示し、特定の医療機器に特化した提案システムを作成し、検証実験を行う。

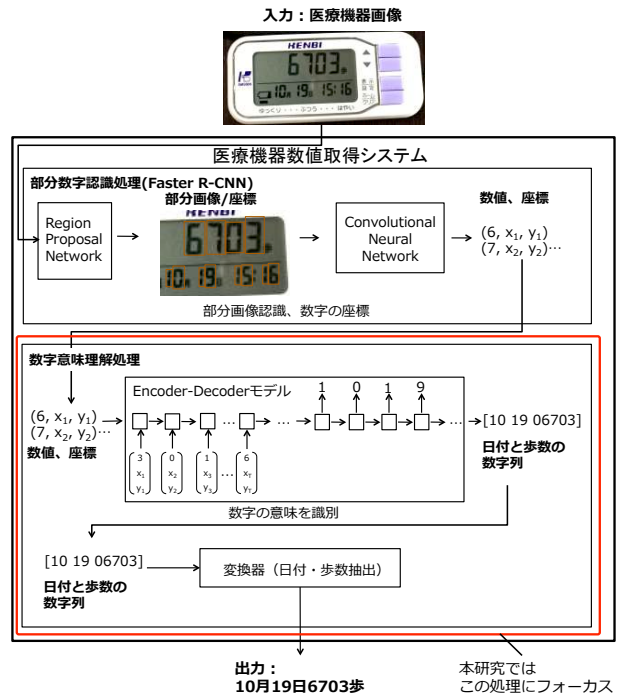


図 2: 提案システムの処理の流れ

2.1 数字意味理解処理の設計

本研究では、数字意味理解処理に Encoder-Decoder モデルを用いる。Encoder-Decoder モデルは一般的に機械翻訳や、対話モデルを構築する際に用いられており、入力側、出力側に RNN の拡張として作られた、時系列データに対するモデルの 1 つである LSTM を用いし、中間ノードで繋ぎ合わせたモデルとなっている。

提案システムの数字意味理解処理では、数値と数値の画像内位置座標を入力とし、指定した並びに並べ替えられた数値列を出力とする。

入力として用いる数値は医療機器のディスプレイに表示される数値であるため、0-9 までのデジタル数字となる。各デジタル数字に対して、デジタル数字を 10 次元のカテゴリカルなデータに変換し、デジタル数字の中心座標 (x, y) を加えた 12 次元のベクトルを作成する。この 12 次元のベクトルを表示された数字分だけ列とした、12 次元 × 画像内のデジタル数字の数の行列が入力データの型となる。また、Encoder-Decoder モデルでは、入力データの次元数が一定である必要があるため、画像ごとにデジタル数字の数が異なる場合、表示されたデジタル数字の桁数や数に応じて、足りない列数を補間する必要がある。

出力は、目的の情報を数値列に置き換えたものとなる。例えば、読み取りを行う対象の機器が月日、時間、薬剤の残量の情報を表示しているのであれば、月日 4 桁、時間 4 桁、薬剤の残量 3 桁、のような 11 桁の数値列となる。出力に関しても、薬剤の残量など桁数が不



図 3: 左: 訓練に用いた歩数計画像 右: アノテーションされた歩数計画像

定に変わる部分に対しては、目的の数値列の桁数が不定になってしまうため、20 という値を 020 と 3 桁に直す、といった処理によって補う必要がある。入力の際に 0-9 までのデジタル数字を 10 次元のカテゴリカルなデータに変換しているため、実際に Encoder-Decoder モデルから出力される値としては、10 次元のカテゴリカルなデータが目的の桁数並べられた配列となる。

このような入出力データを用いて、数字意味理解処理の Encoder-Decoder モデルを訓練することによって、数値表示部分のみを切り出すなどの、画像内における数値の座標ではなく、数値同士の位置情報の意味を学習することができると考えられる。よって、追加のプリプロセッシングを行わずとも、数値の画像内サイズに関係なく、数値の相互の位置関係を元に、目的の数値列への並べ替えを行えると期待される。

提案する Encoder-Decoder モデルでは、訓練の際に正解データとして与える変換後の数値列を変更することによって、月日の情報は必要であるが、時間の情報は必要ない、薬剤の容量がかなり多く、5 桁から 1 桁まで変動する、といった機器の情報取得にも対応することが可能となる。



図 4: 歩数計画像に対して部分数字認識処理を行なった結果

3 歩数計を用いた数字意味理解システムの設計

数値がディスプレイに表示されている情報機器の例として、歩数計を用いて、提案モデルの設計を行なった。

図 3 に訓練に用いた歩数計の画像と、手動でアノテーションを行なった際の歩数計の画像を示す。入力データとしてはカメラで撮影した歩数計画像を用い、手動のアノテーションにより、歩数計画像内全ての数値の値と、数値ごとの左上と右下の (x, y) 座標を保持した訓練データを作成した。また、図 3 に示すように、訓練データの背景は不定であり、歩数計の回転、画像内サイズも様々である。

歩数計に特化した際の提案システムは歩数計画像を入力として受け取り、月日と歩数を出力する。

提案システムの部分数字認識処理では、入力された歩数計画像に対して、Faster R-CNN 内部の RPN によって、歩数計に表示された月、日、時間、歩数の数値部分が認識され、各数値部分に対して CNN によりクラス分類を行い、数値の値と画像内の位置座標を出力する。

提案システムにおいて、部分数字認識処理を 10 枚の歩数計画像に対してのみ訓練を行い、訓練に使った画像を用いた、クローズドなテストを行なったところ、図 4 に示すような結果が訓練で用いた 10 枚全ての画像において得られた。クローズドなデータに対してではあるが、ほぼ全ての表示されている数値が座標、値ともに正しく認識されていることが分かる。

数字意味理解処理には 2.1 節にも示したように、Encoder-Decoder モデルを用いた。歩数計に特化した提案システムの数字意味理解処理の構造を図 5 に示す。数字意味理解処理では、入力として数値と数値の画像内位置

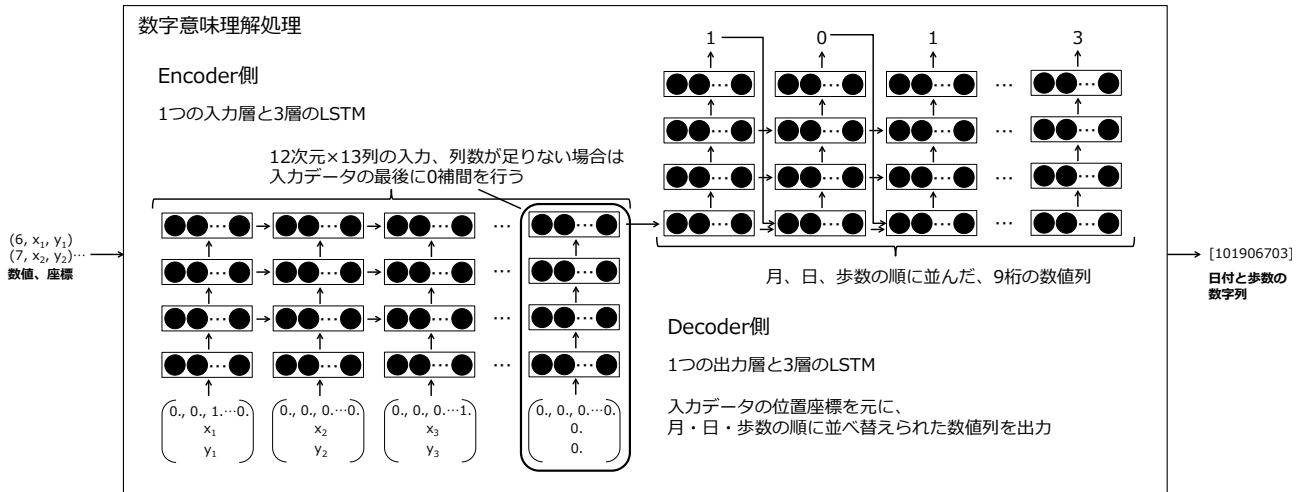


図 5: 歩数計に特化した提案システムの数字意味理解処理の構造

座標を受け取り、月、日、歩数の順に並べ替えられた数値列を出力する。Encoder-Decoder モデルの構造としては、Encoder 側に 1 つの入力層と 3 層の LSTM、Decoder に 1 つの出力層と 3 層の LSTM を持つモデルを設計した。LSTM は長期的な依存を学習することが可能であるため、このような Encoder-Decoder モデルを作成することで、Encoder 側では、入力データの学習に、直前だけではなく数ステップ前に入力された数値と位置座標を考慮に入れた学習を行うことができ、また Decoder 側では、過去の出力データを元に次の出力を決めることになるため、並び順に関連性がある、意味を持った数値列を出力するモデルとなることが期待できる。

3.1 数字意味理解処理部の訓練

提案システムにおける数字意味理解処理では、日付と歩数を各桁をおのおの別の 0-9 までの数値データとして扱う。すなわち、数字意味理解処理の入力としては、月の単位 2 次元、日の単位 2 次元、歩数 5 次元、時間(時/分)4 次元のデジタル数字がランダムな順番で得られる。訓練時の入力データは、得られたデジタル数字それぞれに対して、数値の画像内における中心座標 2 次元 (x, y) とカテゴリカルなベクタ 10 次元 (0-9 までの数字であるかの確率) の、12 次元 × 13 列の行列を用いた。しかし、歩数計に表示される日付部分では、2 桁の月日に関しては、10 月 10 日といった形で表示されるが、1 桁の月日に関しては、01 月 01 日ではなく、1 月 1 日のように表示されてしまうため、入力として与える月日の数値の次元数が変わってしまう。また、歩数に関しても、5 桁の歩数は 10000 歩のように表示されるが、4 桁以下の歩数は 1234 歩のように桁数

が 5 桁以下になってしまい、入力の次元数が変わってしまう。そこで提案システムを訓練する際には、歩数や日付の違いによって、入力データの次元が変動することを避けるために、列数が足りなかった場合は 0 を 12 個並べた配列を足りない列数分追加することで補間した。この際、追加した列は、歩数が 4 桁の場合に歩数を表す入力データの手前に 1 列追加するようには行わず、画像内数値全てにおいて 12 次元 × 数値数の行列を作成した後に、13 列に満たない列数を後ろに付け足すことを行なった。

訓練時の正解データとしては、歩数計画像ごとに月の単位 2 次元、日の単位 2 次元、歩数 5 次元それぞれに対して、カテゴリカルなベクタ 10 次元 (0-9 までの数字であるかの) の 10 次元 × 9 列を用いた。正解データにおいては、1 月 1 日を 0101 と表現し、歩数が 5 桁以下である場合には 01234 のようにし、必ず月 2 桁、日 2 桁、歩数 5 桁の順番に並んだ数値列を正解データとして与えた。

データセットとしてアノテーションされたデータは 150 枚あり、その 100 枚を訓練データとして用い、学習回数 10000 回で訓練を行なった。

4 数字意味理解処理部の検証

まず、学習済みのモデルに対して、訓練に用いた歩数計画像でクローズドなテストを行なった。結果を表 1 に示す。また、出力と正解の欄に下線が引かれているものが、出力と正解が一致しているデータである。表 1 に示すように、訓練に用いたデータに対しては、数値、順番が不定な入力データからでも、月、日、歩数の順の数値列に正しく並べ替えられていることが分かる。また、クローズドなテストを行なった際の認識率を図

表 1: 訓練に用いた画像に対して数字意味理解処理を行なった出力の一部

	入力された数値	数値の位置座標	出力			正解		
			月	日	歩数	月	日	歩数
0	[4, 3, 2, 3, 1, 9, 4, 1, 2, 2, 2, 0, 1]	[[0.33, 0.64], [0.36, 0.63]...	10	22	13234	10	22	13234
1	[2, 4, 6, 1, 1, 0, 2, 2, 1, 1, 5, 7]	[[0.39, 0.43], [0.45, 0.43]...	10	22	02461	10	22	02461
2	[1, 5, 2, 0, 1, 0, 2, 3, 8, 3, 3]	[[0.51, 0.45], [0.51, 0.47]...	10	23	01520	10	23	01520
3	[4, 8, 0, 4, 1, 0, 2, 8, 1, 4, 0, 7]	[[0.53, 0.50], [0.56, 0.49]...	10	28	04804	10	28	04804
4	[1, 5, 5, 5, 2, 2, 4, 1, 8, 2, 0, 1]	[[0.46, 0.45], [0.51, 0.45]...	10	28	05551	10	28	05551

表 2: 訓練に用いなかった画像に対して数字意味理解処理を行なった出力の一部

	入力された数値	数値の位置座標	出力			正解		
			月	日	歩数	月	日	歩数
0	[7, 3, 2, 1, 6, 9, 1, 0, 1]	[[0.43, 0.62], [0.45, 0.66]...	10	19	00035	10	19	00037
1	[9, 4, 1, 2, 2, 2, 0, 1, 4, 3, 2, 3, 1]	[[0.36, 0.51], [0.38, 0.51]...	10	23	15121	10	22	13234
2	[1, 6, 8, 5, 1, 0, 2, 8, 1, 4, 2, 6]	[[0.53, 0.42], [0.57, 0.45]...	10	23	06482	10	28	05861
3	[2, 4, 4, 1, 1, 0, 2, 6, 1, 0, 2, 3]	[[0.46, 0.40], [0.50, 0.39]...	10	25	01592	10	26	02441
4	[8, 5, 3, 5, 1, 0, 2, 1, 1, 4, 4, 2]	[[0.52, 0.53], [0.52, 0.59]...	10	20	03550	10	21	05358

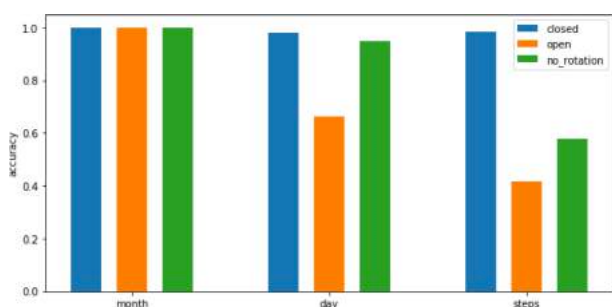


図 6: 認識率の比較

6 に青色の棒グラフで示す。本論文における認識率は、数値列の月 2 桁、日 2 桁、歩数 5 桁の値が正解とどれだけ一致しているかを示したものであり、例えば歩数同士の比較であれば、12345 を 1 桁ずつ '1', '2', '3', '4', '5' と分け、1 桁ずつ一致しているかを確認した。図 6 に示されたように、訓練に用いた画像に対しては、どの項目に対しても 100% に近い認識ができていたことが分かる。このことより、提案する数字意味理解処理のモデルは、数値と位置座標を元に意味を理解し、意味を持った数値列に落とし込むという数値の並び替え問題を扱えることが分かった。

次に、訓練に用いなかった 50 枚の歩数計画像に対して、オープンなテストを行なった。オープンなテストの出力の一部を表 2 に示す。出力と正解の欄に下線が引かれているものが、出力と正解が一致しているデータである。また、正解率のグラフを図 6 にオレンジ色の棒グラフで示す。表 2 と図 6 に示されたように、数値列の最初の 2 桁である月を表す 10 は全ての画像に

して認識できている。日を表す 3-4 桁目は、10 の位は高い精度で認識できている。一方、歩数部に関しては、認識率が低く、桁数が間違っているデータ数が、1 回あたり平均して 7.4 個であった。

また、テストの出力とテスト画像を照らし合わせてみると、画像内の歩数計部分がかなり小さい時、画像内の歩数計が 90 度や 270 度回転している時にも、文字列への並べ替えがうまく行っていないことが分かった。

歩数計画像の回転による認識率の変化を検証するため、歩数計画像を正面から撮影していた 69 枚の画像のみを用いて学習回数 10000 回で提案システムを訓練し、10 枚のテストデータを用いて追加検証を行なった。追加検証の結果を図 6 に緑色の棒グラフで示す。図 6 に示すように、オープンテストに比べて、月、日、歩数どの数値に対しても認識率が向上していることが分かる。

提案システムの汎用性を検証するために行なったオープンなテストの結果は、月の 2 桁に関しては、訓練データに用いた歩数計画像が全て 10 月に撮影されたものであり、日付部分が全て 10 月であるためだと考えられる。また、日にち部分と歩数部分に関しては、日にちに比べ、歩数は 1 桁-5 桁までの数字となり、自由度が高いため、日にちよりも認識率が下がったものと考えられる。

クローズドなテストとオープンなテストの結果を比較すると、訓練データが少ないために、特に自由度が高い歩数部分の認識率が下がっているため、訓練データを増やすことで認識率の向上が見込める。

5 考察

本研究では、Faster R-CNN で認識された歩数計画像内の数値とその位置座標を入力とし表示された数値を認識する Deep Encoder-Decoder モデルの実装に取り組んだ。順不同の数字列から意味のある数字列への変換には、並び替えの他にも、その位置情報と数字の出現確率などの関係の考慮や、取得する必要のない数字の扱いなど、モデルベースの手法で問題を解こうとすると、複数の手法を組み合わせていくことになり、ハイパーパラメータの調整が困難であることが予想される。このような処理をできるだけ廃して、Encoder-Decoder モデル単体で問題を解いた場合、どれだけの性能が出るかを検証することも実験の重要な目的の一つであった。

提案モデルは当初、1) 数値の順番の違い、2) 数値の桁数の違い、3) 不要な数字の存在という3つの項目をモデルによって吸収することを意図していたが、実際には、用いた訓練データの制約によって、4) 縮尺の違い、5) 角度の違い、6) 不十分な訓練データ、というより複雑な処理を必要とするエンコーダを学習することとなり、検証結果より、特に歩数において低い認識率になることが分かった。

4章の追加検証では、吸収すべき項目のうち、5) の角度の違いを無くし、他の5つの項目にフォーカスを絞ることで、認識率の改善が見られた。残り2つの項目に関しても、縮尺の違いに関しては、画像の縦幅、横幅を縮尺を認識できるデータとして入力データに追加することで、不十分な訓練データに関しては、それぞれの値域を考慮した Data Augmentation を行うことで認識率を改善させることがそれぞれ可能であると考えられる。今後は、このような処理を追加し、1) 数値の順番の違い、2) 数値の桁数の違い、3) 不要な数字の存在の吸収に再度、学習対象を絞ることによって、提案するモデルの過学習を抑制し、オープンテストにおける認識率の向上を試みる。

6 まとめと展望

本研究では、医療現場において人と機械によるダブルチェックを提案し、ダブルチェックをサポートするために、機器に表示された数値を取得するシステムのプロトタイプ的设计と、数字意味理解処理部の実装を行った。本論文では、数値がディスプレイに表示されている情報機器の例として、歩数計を挙げ、歩数計画像に特化した提案システムを作ることによって検証を行った。歩数計に特化した提案システムの数字意味理解処理において、クローズドなテストを行なったところ、期待していた結果が得られ、提案システムによる画像からの数値取得は可能であることが示唆された。

今後の展望としては、訓練データを擬似的に増やし、提案システムの汎用性の向上を目指す。また、システムの設計部分では、現在、部分数字認識処理と数字意味理解処理、数値列の変換器が独立しているため、これら3つの処理を1つに統合した数値認識ネットワークモデルとすることで、ダブルチェックを行うためのシステムへの導入や、アプリケーションとしての提供がより容易になると考える。

謝辞

本研究の一部は国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」および、厚生労働省科学研究費補助金 (14528949) の支援によって行われた。

参考文献

- [1] 日本看護協会: 医療安全推進のための標準テキスト, 東京都: 日本看護協会 (2013)
- [2] 医真会八尾総合病院 医療安全管理室: 医療安全ニュース 7期 8号 (2016)
- [3] Ian et al: Multi-digit number recognition from street view imagery using deep convolutional neural networks, *arXiv preprint arXiv:1312.6082* (2013)
- [4] Pan et al: Reading Scene Text in Deep Convolutional Sequences., *AAAI*, Vol.16, pp.3501-3508 (2016)
- [5] Ren et al: Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, pp.91-99 (2015)
- [6] Cho et al: Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078* (2014)
- [7] Sepp et al: Long short-term memory, *Neural computation*, Vol.9, No.8, pp.1735-1780 (1997)
- [8] 渋谷ら: 地域高齢者の健康づくりにおける1日平均歩数の有用性について, *日本老年医学会雑誌*, Vol.44, No.6, pp.726-733 (2007)