一般口演 | 知識工学

一般口演10
知識工学

2019年11月23日(土) 09:00 〜 11:00 C会場 (国際会議場 2階国際会議室)

[3-C-1-06] 機械学習を用いた画像診断レポートからの情報抽出

○杉本 賢人[1]、和田 聖哉[1]、山畑 飛鳥[1]、武田 理宏[1]、真鍋 史朗[1]、松村 泰志[1] （1. 大阪大学大学院　医学系研究科）

キーワード：Natural Language Processing, Information extraction, Radiology Information Systems

【背景と目的】画像診断レポートには放射線医が記述した診断における重要な情報が記述されており，この情報は，臨床研究や診断支援システムなど様々な分野での活用が期待されている．二次利用のためには，レポートから必要な情報を抽出して構造化する必要がある．しかし，抽出する情報の対象は自明ではなく，また，定義したとしても，フリーテキスト形式のレポートからの正確な情報抽出は難しい．我々は二次利用に向けて，まず，画像診断レポートの「情報モデル」を定義し，必要な構成要素を整理した．また，機械学習を用いて，レポートから各構成要素の対象を抽出した．

【方法】まず，複数の医師・放射線検査技師で議論し，画像診断レポートの所見に記述された文章から，重要な構成要素を網羅した「情報モデル」を定義した．次に，機械学習（双方向再帰ニューラルネット）を用いて，文章中の重要な構成要素を抽出するための分類器を構築した．本研究では，2016年から大阪大学医学部附属病院の画像診断レポートシステムに蓄積されている胸部単純 CT画像の所見（44,383件）を利用した．データセットから，訓練用に500件，評価用に150件を無作為に抽出し，分類器の学習及び評価を行った．

【結果】画像診断レポートの「情報モデル」として，「臓器や部位に関する表現・観察物を示す表現・臓器の異常所見を示す表現・肯定／否定などの表現・観察物の特徴を示す表現・観察物のサイズを示す表現・観察物の変化状態を示す表現」の合計7つのクラスを定義した．次に，学習した分類器を用いて，それらの抽出精度を評価した．各クラスの平均の F1値は，0.936であり，高い精度でレポートから，定義した構成要素が抽出可能であることを示している．

【結語】二次利用に向けて，画像診断レポートの「情報モデル」を定義した．また，その構成要素を機械学習を用いて高精度で抽出した．

©一般社団法人 日本医療情報学会

機械学習を用いた画像診断レポートからの情報抽出

杉本 賢人, 和田 聖哉, 山畑 飛鳥, 小西 正三, 武田 理宏, 真鍋 史朗, 松村 泰志
大阪大学大学院医学系研究科 医療情報学

Information Extraction from Radiology Reports with deep learning approaches

Sugimoto Kento, Wada Syoya, Yamahata Asuka, Konishi Syozo, Takeda Toshihiro, Manabe Shiro, Matsumura Yasushi
Department of Medical Informatics, Osaka University Graduate School of Medicine

Radiology reports have a lot of valuable information for diagnosis, which is expected to utilize for secondary use, such as clinical research, clinical decision support system. Radiology reports are written in a free-text form, which makes difficult to process. Natural Language Processing (NLP) is a key to address this problem. In addition, defining some necessary elements from radiology reports is needed to convert structured format. We defined an information model of a radiology report and extract important elements by machine learning. Our information model has seven classes of concept: Anatomical entity, Imaging observation, Clinical finding, Certainty descriptor, Characteristics descriptor, Size descriptor, Change descriptor. These cover the majority of clinically significant information contained in radiology reports. We built a neural network architecture to extract seven classes of concept. Average F-score is 0.90 (10-fold cross-validation), which shows our model can extract target elements accurately.

Keywords: Natural Language Processing, Radiology Report, Information Model, Machine Learning

1. 背景

画像診断レポートには，放射線医が記述した診断における重要な情報が記述されているおり，これらの情報は，臨床研究や診断支援システムなど様々な分野での活用が期待されている．二次利用のためには，画像診断レポートから必要な情報を抽出して構造化する必要があるが，フリーテキスト形式で記述されたレポートから情報を抽出するのは容易ではない．そのため，画像診断レポートに関する自然言語処理を用いた研究が多く行われてきた [1)].

自然言語処理による画像診断レポートからの情報抽出は，辞書ベースでの手法が古くから用いられているが，この手法の場合，網羅的な表現の抽出が難しいことが知られている [2)]．特に，医学表現の場合，表現の多様性が大きな問題となり，同じ概念であっても，書き手によって異なる用語が使われるため，レポート中の表現の多様性を網羅するのは容易ではない [3) 4)]．一方，機械学習を用いた抽出手法も多く研究されている．機械学習による抽出手法の場合，単語だけではなく，レポートの文脈情報も考慮して抽出を行うことができるため，辞書ベースと比較して幅広い表現の抽出が期待できる．また，深層学習を利用した手法は，従来の特徴エンジニアリングベースの手法と比べて，高い抽出精度を記録できることが知られている [5)].

我々は先行研究にて，胸部単純X線画像レポートから，部位と所見に関する用語を，深層学習を用いて高い精度で抽出できることを示した [6)]．しかし，この手法は，二次利用先を「胸部単純 X 線画像からの病変検出システム」を想定しており，対象の病変を「所見」としてまとめ，病変に関する属性情報を対象としていなかった．しかし，幅広い二次利用先を想定した場合，この手法は好ましくない．たとえば，肺がんに関するレポートを分析したい場合，結節の有無だけではなく，その大きさや性状などの属性情報が必要となる．また，過去のレポートとの変化も重要な情報となるかもしれない．このような二次利用を考える場合，先行研究で定義した部位と所見に関する情報だけでは，不十分であると考えた．

2. 目的

本研究では，幅広い二次利用先を想定して，レポートから抽出するべき情報について定義する．また，定義した情報を正確に抽出するためのモデルを構築し，その精度を評価する．本研究では，胸部 CT レポートを対象として，実験を行う．CT 画像は単純X線画像と比べて，病変の位置や大きさ，その性状などをより詳細に検査できるため，レポートも病変に関する記述がより詳細になることが多く，これらの詳細な情報を含めた精度評価をすることが望ましいと考えた．

3. 方法

まず，我々が定義した情報モデルについて説明する．次に，機械学習の教師データを作成するためのアノテーション作業について説明する．最後に，構築した深層学習モデルについて述べる．

3.1 情報モデルの定義

情報モデルとは，コンピューター上で情報を整理するための基本的な枠組みである．情報モデルを定義することで，それに基づいてテキストから抽出するべき情報を整理することが可能となる．我々は，先行研究にて提案されている画像診断レポートの情報モデル [7)-9)]に加えて，RadLex[10)]の分類を参考にして，情報モデルの定義を検討した．RadLex は，RSNA によって作成された放射線領域に関する用語集であり，多様な用語が階層構造で整理されている．我々は，「臓器や部位に関する表現（Anatomical entity）・観察物を示す表現（Imaging observation）・臓器の異常所見を示す表現（Clinical finding）・肯定／否定などの表現（Certainty descriptor）・観察物の特徴を示す表現（Characteristics descriptor）・観察物や臓器のサイズを示す表現（Size descriptor）・観察物や異常状態の変化状態を示す表現（Change descriptor）」の 7 つの概念から情報モデルを定義した（以降，これらの概念を「エンティティ」と呼ぶ）．胸部 CT レポートの所見を定義した情報モデルで整理した例を表 1 に示す．所見内にエンティティ候補が存在しない場合，そのエンティティは空となる．また，所見に複数のエンティティ候補が存在する場合，複数の要素を含めるようにした．

表 1 情報モデルによりレポートを整理した例

①右肺 S6 に径 15mm の炎症後変化を疑う
不整な結節影を認める.

エンティティ	要素例
Anatomical entity	右肺 S6
Imaging observation	結節影
Clinical finding	炎症後変化
Certainty descriptor	認める
Characteristics descriptor	不整な
Size descriptor	径 15mm
Change descriptor	–

②右胸膜に石灰化を伴う限局性肥厚あり.
前回と著変なし

エンティティ	要素例
Anatomical entity	右胸膜
Imaging observation	–
Clinical finding	肥厚
Certainty descriptor	あり
Characteristics descriptor	石灰化，限局性
Size descriptor	–
Change descriptor	前回と著変なし

定義したエンティティについて，Anatomical entity，Imaging observation，Clinical finding は，RadLex の最上位階層の概念を利用した．また，Certainty descriptor，Size descriptor については，他の概念の属性表現を広く定義した Radlex descriptor の下位分類の概念を利用した．一方，Characteristics descriptor に関しては，Radlex descriptor の下位分類の形態上の属性，密度に関する属性など分類をまとめて新しく定義した．Change descriptor についても同様に Radlex descriptor の状態変化に関するいくつかの分類をまとめた概念として新しく定義した．これらは，詳細に分類し過ぎると，訓練データを作成するための人手でのアノテーションが困難になり，機械学習による抽出精度が低下することを懸念したためである．

3.2 対象データ

本研究では，2010 年から大阪大学医学部附属病院の画像診断レポートシステムに蓄積されている胸部 CT 画像のレポート（118,078 件）を利用した．尚、本研究は大阪大学医学部附属病院の観察研究倫理審査委員会の承認（承認番号 17166）を得て実施した．

3.2 前処理

レポート中の文章に対して，クレンジング（ヘッダー フッターのレポートの内容に関係の無い情報の除去、半角・全角の統一など）処理を行い，データを整形した後、文章を1文ごとに分けた．

次に，整形された文を形態素単位に分割した．本研究では、形態素解析ソフトウェアとして MeCab[11]を利用した．

3.3 アノテーション

システムに蓄積されたレポートから，訓練及び評価用のデータセットとして無作為に 250 件を抽出して，アノテーション作業を実施した．ここで，アノテーション作業とは，形態素単位に分割されたレポートについて，1 つ以上の形態素の範囲に適切なエンティティを付与する作業を指す．アノテーション作業は，3 名の医療従事者（医師 2 名，放射線技師 1 名）により実施された．アノテーションツールとしては，TALEN[12]を利用した．

表 2 TALEN の出力例

入力	出力
両側	B-Anatomical entity
肺	I-Anatomical entity
尖部	I-Anatomical entity
に	O
結節	B- Imaging observation
影	I- Imaging observation
を	O
認める	B- Certainty descriptor

まず，作業前にエンティティ別にどのような要素を対象とするか，その範囲をどのように定義するか等を定義したガイドラインを作成して，アノテーターに配布した．その後，サンプルデータを用いて，各アノテーターのタスクの理解度を高めた．

本番作業では，1 つのレポートを各アノテーターに配布して，2 人以上が一致した内容を Gold Standard として採用した．全員不一致の場合，ガイドラインが不十分であったと考え，ガイドラインに関連する内容を追加した．その後，アノテーターに結果の修正を依頼し，一致度を高めるように努めた．アノテーターの結果の修正後，全員不一致になる箇所については，別の医療従事者（医師）に意見を求め，その判断を Gold Standard として採用した．

3.4 データセット

TALEN の出力結果は，アノテーションした 1 つの範囲を IOB2 スキーマ [13]により同定する．IOB2 スキーマでは，B はアノテーションした先頭の形態素，I はアノテーションの先頭以外の形態素，O はアノテーションされなかった形態素に付与される．表 2 に「両側肺尖部に結節影を認める」という文を入力として，適切にアノテーションされた TALEN の出力例を示す．教師データとして，1 文の入力系列とそのラベル情報となる TALEN の出力系列のペア 1,668 文を教師データとして利用した．

3.5 系列ラベリング

系列ラベリングでは，系列情報を機械学習の入力として，入力の各系列に対応するラベル情報を予測するタスクである．このタスクは，深層学習を用いることで，高精度な予測モデルが構築できることが知られており [5]，我々も深層学習を用いてモデルを構築した．

3.6 Bi-LSTM-CRF

本研究では，系列ラベリングの共通タスクである CoNLL[14]でも高い精度を記録した Bi-LSTM-CRF モデル [15] [16]を Pytorch[17]を用いて実装した．以下，モデルの構成要素について述べる．

3.6.1 Bidirectional LSTM

再帰ニューラルネット(RNN)は，音声や文章などの系列データを扱うニューラルネットであり，可変長の入力列の各要素の並びの特徴を取り扱うことができる．しかし，RNN は長い入力系列の学習において，勾配爆発や勾配消失が問題となるため、うまく学習が進まないことが知られている [18]．それを克服するため LSTM（Long Short Term Memory）と呼ばれるモデ

ルが考案された [19]. LSTM は, RNN の各ユニットに memory cell, input gate, output gate と呼ばれる制御構造を追加し, これにより長期の情報を保存することを可能としている. 初期の LSTM に加えて, memory cell の情報をリセットするために forget gate を加えたバージョン [20]が現在の標準的な LSTM として広く利用されており, 本研究でも forget gate を加えた LSTM を利用した. その後, LSTM を応用したモデルとして, Bidirectional LSTM[21]（以降, Bi-LSTM と表記）が提案された. 通常の LSTM は, 系列の各要素を前向き（先頭から末尾に）に伝播させる1方向だけであるが, 双方向の場合, 前向きと後ろ向き（末尾から先頭）の双方向の伝播を組み合わせることで, 入力系列全体の情報を考慮できるようになる.

3.6.2 Conditional Random Fields

Conditional Random Fields[22]（以降, CRF と表記）は, 確率モデルを系列データに適用したものであり, 深層学習以前の系列ラベリングのモデルとして広く利用されていた. CRF の特徴は, 出力のラベルは独立しているのではなく, 前のラベルに依存していると仮定し, ラベルの出力時に前のラベル情報を考慮できる点である. Bi-LSTM の出力層に CRF を組み合わせることで、ラベルの依存情報を考慮することが可能となる.

3.6.3 Word Embedding

単語などの離散的な要素を数学的に計算可能なベクトル空間にマッピングすることを「Embedding（埋め込み）」と言う. 本研究の系列は形態素であるが, 一般的に用いられる Word Embedding という表現を使用する. Word Embedding の初期値は, ランダムに決めるのではなく, 事前学習の結果を適用することで, 目的のタスクの精度が向上することが知られている [23]. 我々も事前学習の手法として, CBoW(Continuous Bag-of-Words)[24]を用いて事前学習した単語ベクトルをモデルの入力に用いた. 尚, 事前学習は, 本研究の対象データとした胸部CT画像のレポートから, 訓練及び評価用として抽出したデータセットを除いたレポートを利用した.

4. 結果

4.1 精度評価

アノテーション済のレポート 250 件（1,668 文）データセットを 10-fold 交差検証を用いて評価を行った, 最終的に評価した値は, 10-fold の平均値により計算した. 評価指標としては, Precision, Recall 及びその調和平均である F1 値を用いた.

実験では, 訓練データへの過学習を抑えるため, drop out を採用した, また, 勾配爆発を抑えるため, 勾配クリッピング（上限値 5.0）を設定した. 最適化のアルゴリズムとしては, RMSProp[25]を用いた. その他, 設定したモデルのハイパーパラメータを表 3 に示す.

表 3 モデルのハイパーパラメータ

ハイパーパラメータ名称	値
Batch size	32
Epoch	10
Word Embedding	50
dropout rate	0.5
LSTM unit	128

事前学習した Word Embedding の効果を検証するため, 事前学習の有無についても評価を行った. 表 4 は事前学習の有無別に全体の精度を比較した結果である.

表 4 事前学習の有無別の精度

事前学習	P	R	F
無し	0.84	0.83	0.84
有り	0.90	0.91	0.90

* P: Precision, R:Recall. F: F1

事前学習無しの場合, F1 値は 0.84 に対して, 事前学習有りの場合, F1 値は 0.90 であった. この実験より, 事前学習が精度の向上に貢献していることが分かる.

次に, 事前学習ありでのエンティティ別の精度を表 5 に示す.

表 5 エンティティ別の精度

エンティティ	P	R	F
Anatomical entity	0.89	0.95	0.92
Imaging observation	0.86	0.91	0.89
Clinical finding	0.88	0.88	0.88
Certainty descriptor	0.96	0.97	0.97
Characteristics descriptor	0.76	0.78	0.77
Size descriptor	0.90	0.93	0.92
Change descriptor	0.95	0.90	0.92
Total (avg)	0.90	0.91	0.90

Characteristics descriptor を除くエンティティの F1 値は, 0.9 前後かそれ以上であり, 概ねレポート中の対象となるエンティティを高精度で抽出できており, その予測精度も正確であることが分かる. 一方, Characteristics descriptor の F1 値は 0.77 と他のエンティティと比較して, 0.1 程度低い結果となった.

4.2 誤り例の評価

機械学習による精度は, 全体の F1 値が 0.90 を達成した. しかし, 実験からいくつかの抽出誤りや見逃しのケースも明らかになった. 誤り例の 1 つとして, 1 つのエンティティの境界を正しく予測できなかった例がいくつか見られた. 以下に例文を示す.

A) <u>左側第 7、8 肋骨</u>に骨硬化像を認めます

この文の場合,「左側第 7、8 肋骨」を 1 つの Anatomical entity として抽出するのが正しい. しかし, モデルの予測結果では,「左側第 7」と「8 肋骨」を別の Anatomical entity として分類していた. これは, 訓練データにおいて, 句点（、）によって 1 つの Anatomical entity が分割される例が多かったことに起因していると考えられる（例えば、「右肺 S3、左肺 S10」のような文の場合, Gold Standard では,「右肺 S3」と「左肺 S10」を別々の Anatomical entity アノテーションしていた例が多かった）. また, Characteristics descriptor が正しく判断できない例もいくつか見られた. 以下に例文を示す.

B) 食道に明らかな<u>腫瘤状</u>肥厚を指摘できません

C) 両側肺に<u>空洞</u>を伴う腫瘤影を認めます

「腫瘤状, 空洞」などは, 後ろの「肥厚, 腫瘤影」などの特徴を表した修飾表現なので, Characteristics descriptor と分類されるのが望ましい. しかし, モデルは, これらを誤って Imaging observation と予測した例が見られた. これも, 先ほどの例と同様に,「腫瘤状, 空洞」などの表現が訓練データ中に Imaging observation として多く出現したことが原因と考えられる.

5. 考察

本研究では，画像診断レポートの内容から，情報モデルを定義した．我々が定義した情報モデルは，いくつかの先行研究や RadLex の概念を参考にしており，画像診断レポートの重要な要素の多くを網羅していると考えている．一方で，どの程度汎用的に利用できるかについては，今後，検証が必要であり，より汎用的な二次利用に向けて情報モデルを今後も検討していきたいと考える．

定義した情報モデルの各エンティティの要素については，機械学習により抽出することを試みた．抽出精度は，全体の Precision が 0.90，Recall が 0.91，F1 値が 0.90 であり，高い精度で目的の表現を抽出できたと言える．しかし，エンティティ別に見ると， Characteristics descriptor は，他のエンティティと比べて，精度が低い結果となった．当初，Characteristics descriptor で獲得する表現として，結節の性状を表す「不整な，石灰化した」などの簡単な表現のみを抽出することを目的としていた．しかし，アノテーション作業を始めてみると，「水と同程度の濃度を呈する（腫瘤）」や「不均一に造影される（結節影）」など非常に多様な表現で表される場合があることが分かった．その結果，Characteristics descriptor は，医療従事者によるアノテーション作業でも最も意見が割れたエンティティとなった．更に，Characteristics descriptor は文中に必ず出現するエンティティではないため，その出現頻度は低くなる．多様な表現が低頻度でしか出現しなかったことが原因となり，精度が低い結果となったと考えられる．また，「4.2 誤り例の評価」で述べたように，1 つの単語が文脈によって異なるエンティティになることがある．これを正確に分類するためには，多様な表現を含んだ訓練データとそれを考慮できる表現力のあるモデルの構築が必要であると考える．

医療自然言語処理のアノテーション作業は，医学知識が求められる場合が多いため，医療従事者の協力が不可欠である．しかし，医療従事者に大量のアノテーション作業を依頼することは，時間とコストの観点から現実的に困難である．我々の先行研究 [6]では，3,009 文（2,000 件）の教師データを用意して学習を行ったのに対して，本研究では，その半分程度の 1,668 文（250 件）教師データで全体の F1 値が 0.90 を達成した．レポートの内容が単純 X 線と CT で異なるので，単純な比較だけでの判断は難しいが，より少数のアノテーション作業でも，事前学習を利用した深層学習により，一定の精度が出せるということを示した．しかし，本研究では，教師データの量の妥当性に関する検証まで行えていないため，教師データを増やすことで，更なる精度向上が期待できる．今後，教師データの量の妥当性を検証するとともに，Active Learning などの手法を利用したデータ選択を行うことで，モデルの精度とアノテーターの労力の両方を考慮した効率的な手法を検討していきたい．

6. 結語

本研究では，画像診断レポートの構造化に向けて，レポートの内容から情報モデルを定義し，各エンティティの構成要素を機械学習により抽出することを試みた．

本研究では，情報モデルの定義及びその抽出手法の開発に留まっており，二次利用を考える場合，各エンティティをどのように構造化するか，どのような形式でデータを保存するか等を検討する必要がある．また，抽出した用語の表記ゆれや略語などの正規化なども検討する必要がある．

今後は，構築したモデルの改善，二次利用に向けた構造化や正規化などに取り組みたいと考える．

参考文献

1) Pons E, Braun LMM, Hunink MGM, Kors JA. Natural Language Processing in Radiology: A Systematic Review. Radiology. 2016;279(2).
2) Cornegruta S, Bakewell R, Withey S, Montana G. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. In: Proceedings ofthe Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI). 2016.
3) Sobel JL, Pearson ML, Gross K, Desmond KA, Harrison ER, Rubenstein LV, Rogers WH, Kahn KL. Information content and clarity of radiologists' reports for chest radiography. Acad Radiol. 1996;3(9).
4) Reiner B, Siegel E. Radiology reporting: Returning to our image-centric roots. Am J Roentgenol. 2006;187(5).
5) Yadav V, Bethard S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018.
6) Kento S, Syoya W, Yoshie S, Asuka Y, Toshihiro T, Shiro M, Yasushi M. 画像診断レポートからの構造化データの抽出．第38回医療情報学連合大会論文集. 2018;718–20.
7) Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. Artif Intell Med. 2016;66.
8) Taira RK, Soderland SG, Jakobovits RM. Automatic structuring of radiology free-text reports. Radiographics. 2001;21(1).
9) Langlotz CP, Meining& L. Enhancing the Expressiveness and Usability of Structured Image Reporting Systems. In: Proceedings of the AMIA symposium. 2000.
10) Langlotz CP. RadLex: A New Method for Indexing Online Educational Materials. RadioGraphics. 2006;26(6).
11) Taku K, Kaoru Yamamoto, Yuji Matsumoto. Conditional Random Fieldsを用いた日本語形態素解析．情報処理学会研究報告自然言語処理（NL）. 2004;(47).
12) Mayhew S, Roth D. TALEN: Tool for Annotation of Low-resource ENtities. In: ACL System Demonstrations. 2018.
13) Sang EFTK, Veenstra J. Representing Text Chunks. In: Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics. 1999.
14) Tjong EF, Sang K, De Meulder F. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003. 2003.
15) Huang Z, Research B, Xu W, Baidu KY. Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv. 2015;1508.01991.
16) Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. In: Proceedings of NAACL-HLT 2016. 2016.
17) Adam Paszke, Sam Gross, Soumith Chintala GC. PyTorch. https://github.com/pytorch/pytorch
18) Bengio Y, Simard P, Frasconi P. Learning Long-Term Dependencies with Gradient Descent is Difficult. IEEE Trans Neural Networks. 1994;5(2).
19) Hochreiter S, Urgen Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997;9(8).
20) Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Comput. 2000;12(10).
21) Graves A, Schmidhuber J. Framewise phoneme classification

with bidirectional LSTM and other neural network architectures. Neural Networks. 2005;18(5–6).
22) Lafferty J, Mccallum A, Pereira FCN, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceeding ICML. 2001;
23) Collobert R, Weston J, Com J, Karlen M, Kavukcuoglu K, Kuksa P. Natural Language Processing (Almost) from Scratch. Vol. 12, Journal of Machine Learning Research. 2011.
24) Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv. 2013;1301.3781.
25) Dauphin YN, de Vries H, Bengio Y. Equilibrated adaptive learning rates for non-convex optimization.