2J3-OS-16b-1

退院時要約自動分類器の構築

Construction of Classifiers for Discharge Summaries

津本周作 *1 Shusaku Tsumoto　平野章二 *1 Shoji Hirano　岩田春子 *2 Haruko Iwata　木村知広 *3 Tomohirno Kimura

*1島根大学医学部医療情報学
Department of Medical Informatics, Faculty of Medicine, Shimane University

*2島根大学医学部附属病院入退院管理センター
Center for Bed Control, Shimane University Hospital

*3東芝メディカルシステムズ
Toshiba Medical Systems

This paper proposes a method for construction of classifiers for discharge summaries. First, morphological analysis is applied to a set of summaries and a term matrix is generated. Second, correspond analysis is applied to the classification labels and the term matrix and generates two dimensional coordinates. By measuring the distance between catergories and the assigned points, ranking of key words will be generated. Then, keywords are selected as attributes according to the rank, and training example for classifiers will be generated. Finally learning methods are applied to the training examples. Experemental validation shows that random forest achieved the best perfomrance and the second best was the deep learner with a small difference , but decision tree methods with many keywords performed only a little worse than neural network or deep learning methods.

1. はじめに

診療録の電子化が進み，診療記録，診断書等膨大な電子的テキストが蓄積されるようになった。これらをリソースとしてテキストマイニングを用いて，有用な知識を取り出し，診療・研究・病院管理に役立てることが可能となってきている。診療文書には医療における用語の標準化がなされていないこともあり，多種多様な用語が混在し，解析しにくいことはよく知られているが，退院時要約は各症例の診断・治療経過を時系列的にコンパクトにまとめて記載していることから，一般の診療記録に比べ，キーワードの冗長性が少なく，横断的解析に適していると考えられる [三浦 10, 外池昌嗣 大熊智子 増市博 大江和彦, 鈴木隆弘 横井英人 井宮 淳 里村 洋一 04]。

一方，テキストマイニングによる分類学習では，テキストからデータセットを作成し，それらを用いて，決定木・SVM 等の手法を用いる方法が提案されてきた [Sebastiani 02]。ただし，従来の手法では，キーワードの選択は，キーワードの頻度のみに依存しているため，キーワードとターゲットとなる概念との対応付けが弱いため，分類精度がそれほど上がらないことが指摘されている。医学的なテキスト分析のコンテクストでは，MeSH[MeSH] 等の医療オントロジーの助けを得て，キーワードの絞り込みを容易にしてから，テキストマイニングするという方法 [Srinivasan 01] も提案されている。本研究では，このようなキーワードによる知識をいっさい仮定することなく，分類に必要なキーワードを変数選択することを目的とし，テキストマイニングに形態素解析を適用した後，対応分析による DPC 毎のキーワードの選定を行い，選ばれたキーワードに機械学習で用いられる手法を適用し，キーワードによる DPC コーディングを行う分類器を構築した。

分類器構築後，反復交差検証法 [Kim 09] を用いて，実際の退院時要約を用いて，その分類の正答率を比較した。キーワード数増加により，深層学習，ニューラルネットによる正答率が上昇したが，200 程度のキーワード数でほぼ平坦になった。一方，決定木については，単調に正答率が上昇することがわかった。

連絡先: 津本周作, 島根大学医学部医療情報学講座, 出雲市塩冶町 89-1, 0853-20-2171, 0853-20-2170, tsumoto@med.shimane-u.ac.jp

2. 方法

2.1 プロセス

マイニングのプロセスは Fig. 1 の通りである。まず，退院時要約を抽出後，形態素解析を行い，キーワードに関する分割表を作成する。次いで，対応分析 (2 次元) を行って，各キーワードと DPC について布置座標を与える。これらの布置座標について，DPC とキーワード間のユークリッド距離を計算し，DPC 毎に距離の値によって，キーワードのランク付けを行う。ランク付けされたキーワードを用いて，文書内のキーワードの有無についての表形式のデータ集合を生成する。生成されたデータを用いて，分類学習，ここでは決定木，SVM, BNN あるいは深層学習等の手法を用いて，分類器を構築する。

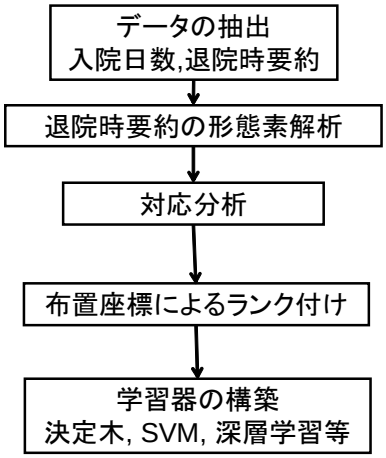

図 1: 分類器構築プロセス

表 1: 入院患者数 10 位までの内訳 (2015 年度)

No	DPC 名称	件数	文字数
1	白内障、水晶体の疾患 手術手術あり 重症度等片眼	445	270.12 ± 295.12
2	白内障、水晶体の疾患 手術手術あり 重症度等両眼	152	287.87 ± 161.05
3	２型糖尿病（糖尿病性ケトアシドーシスを除く。）	145	6888.51 ± 1549.08
4	肺の悪性腫瘍 手術手術あり 処置等２なし	131	4535.57 ± 979.36
5	子宮頸・体部の悪性腫瘍 手術手術なし 処置等２４あり 副傷病名なし	121	1508.72 ± 1023.73
6	肺の悪性腫瘍 手術手術なし 処置等１あり 処置等２なし 副傷病名なし	120	2506.34 ± 1132.42
7	子宮の良性腫瘍 手術腹腔鏡下腟式子宮全摘術等	111	2038.57 ± 910.91
8	肺の悪性腫瘍 手術手術なし 処置等１なし 処置等２４あり	110	3505.10 ± 1121.77
9	妊娠期間短縮、低出産体重に関連する障害（出生時体重 2500g 以上）手術手術なし 処置等２なし 副傷病名なし	110	1182.35 ± 646.16
10	肘、膝の外傷（スポーツ障害等を含む。） 手術腱縫合術等	99	1867.22 ± 639.75

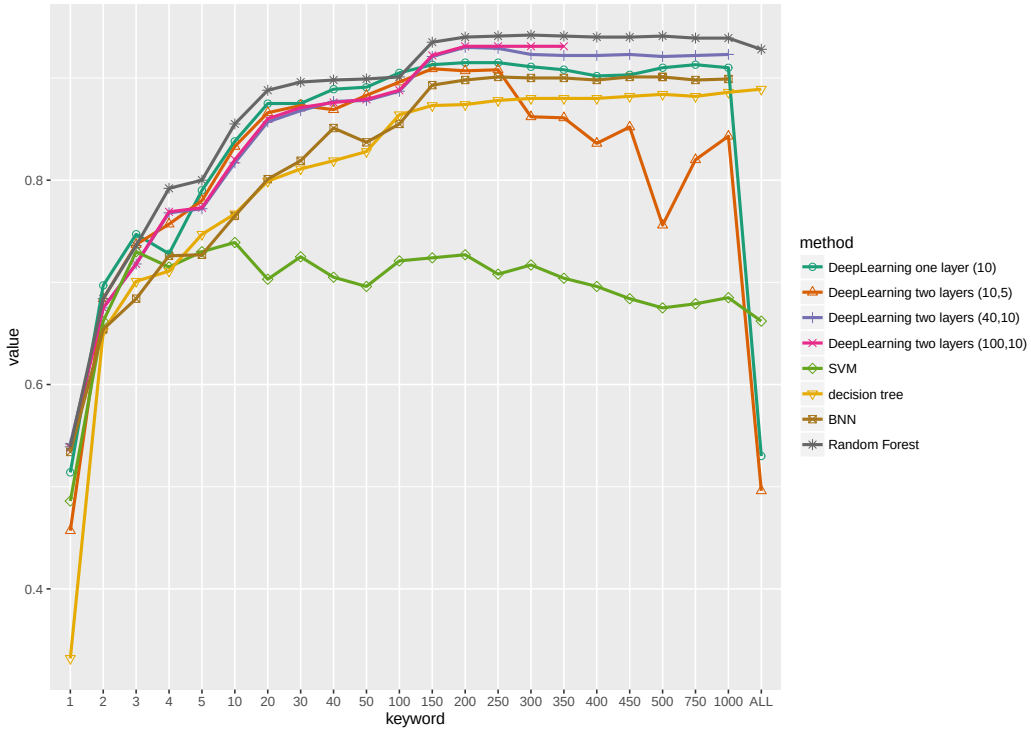

図 2: 実験結果

2.2 実験

島根大学医学部附属病院に 2015 年度に蓄積された退院時要約のうち，DPC コードで上位 10 までのものを抽出した (表 1)。

このデータを RMeCab[石田 16] を用いて，形態素解析を行った後，対応分析 (R3.3.2) を行い，各 DPC について，キーワードのランク付けを行った。各 DPC 間で重複したキーワードについては削除した。分類器構築のプラットフォームとして，R3.3.2 を用い，決定木には rpart[Therneau 15], Random forest には randomForest[Liaw 02], SVM には kernlab[Karatzoglou 04], BNN には nnet[Venables 02], 深層学習については，darch[Drees 13] を用い，Darch のパラメーターとしては，中間層 10 および中間層 (10,5), (40,10), (100,10)，反復学習回数 100 とした。今回はどのパッケージについても，中間層を 2 層にした場合以外は，すべて Default のパラメーター設定を利用した。

次に，構築した分類器の性能評価については，データ集合をランダムに 2 分割し，片方を訓練標本，もう一方をテスト標本として正答率を算出することを 100 回繰り返し，平均正答率を算出した (repeated 2-fold cross validation[Kim 09])。選択するキーワード数は 1 位のみから 1000 位までそれぞれについて，性能を評価した。対応分析，分類器構築，性能評価については，HP Proliant ML110 Gen9 (Xeon E5-2640 v3.2 2.6GHz 8Core, 64GB メモリ) を用いた。

3. 結果

Fig. 2 に各手法のキーワード数別による性能評価のプロットを示す。キーワード数が 4 まででは，すべての学習方法では正答率が 70%程度で拮抗するが，5 以上では，SVM はむしろ低下，他手法は単調に増加，キーワード数 200 付近でほぼ平坦となった。

キーワード数 5 以上では，Random Forest の正答率が高く，他の手法よりも分類精度が高かった。BNN は 5〜100 のキーワード数では Darch に劣っていたが，100 以上では正答

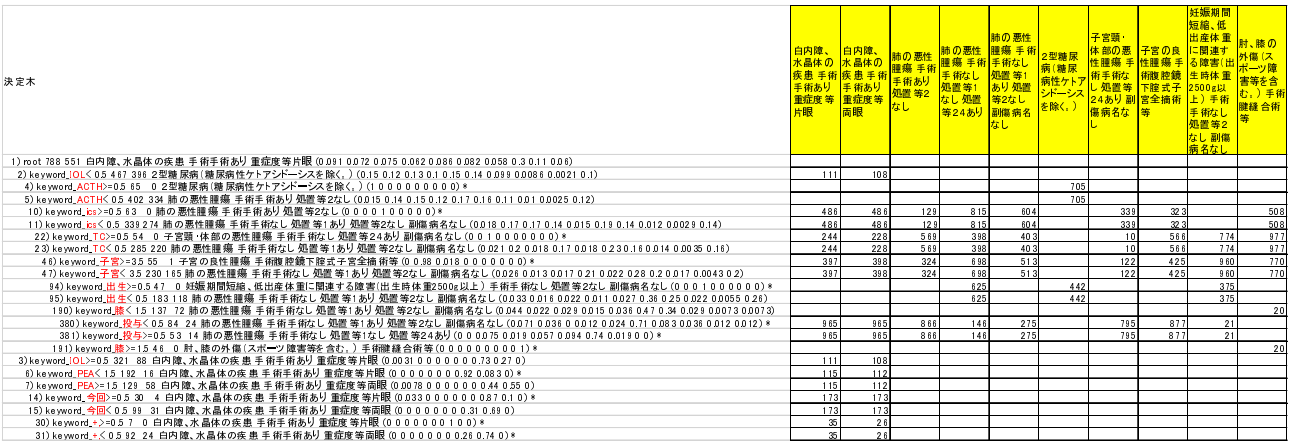

図 3: キーワードすべてで生成された決定木

率が接近していく。Darch2 層 (10,5) は 1 層に比べて性能が劣っていた。2 層 (40,10),(100,10) については，BNN の正答率より数パーセントの差を保っていたが，(100,10) については，Random forest の正答率に接近した。この実験条件下では，中間層が増加した方が精度は上昇し，中間層のニューロン数を増加させることで，Random forest の正答率を越える可能性があるかどうかどうか検証を進める必要がある。

一方，興味深いのは，決定木であり，正答率は単調に増加している。評価実験は抄録提出時には完了していないが，SVM, Darch1 層, 2 層，決定木については参考のため，すべてのキーワードを用いた分類器の性能も見た。決定木は正答率が上昇にしているにも関わらず，SVM, Darch については性能が劣化した。

4. 考察

4.1 分類精度

実験結果から意外であったのは，決定木の正答率が単調に増加すること，および Random Forest の正答率の高さである。決定木が 11 のキーワードを選択したことから改めて，Fig 2 を観察すれば，5 変量選択後，SVM(キーワードの線型結合) とネットワーク型結合との性能の差が顕著になってくることがわかる。では，これらの選択されたキーワードはどのような性質を持っているか，Fig. 3 に，すべてのキーワードを使用して作成した決定木を左に対応するキーワードの DPC それぞれにおける順位を示した。選択されたすべてのキーワードは順位が高く選択されたわけではないことがわかるが，これは，それぞれキーワードによって得られる評価値の差があまり大きくなく，対応分析による評価が接近しているためと考えられる。今後，これらのキーワードの性質を検討する必要があるが，選ばれたキーワードを見る限り，非常に巧妙な選択をしている印象を受ける。このような選択は，退院時要約の記述が疾患間で大きく異なるためのものと考えられるので，今後，類似症例での判別の場合，同様のキーワード選択が可能かどうかは検証する必要がある。

キーワード数による正答率の変化については，SVM, 深層学習ともにキーワード数 200 がピークで，SVM の正答率が 72.7%, 深層学習が 91.5%となり，20%近い性能の差が得られた。これは，SVM がキーワードの線型結合によるのに対し，深層学習では，非線形結合が主となることに起因すると考えられ，キーワードの結合の性質，すなわち線型とは言えない関係性がその精度を考える上で重要であることを暗示している。これは，BNN の正答率がやはり，200 程度においてピークを迎えることによっても支持されていると考えられる。

表 2: 中間層ニューロン増による正答率変化

キーワード数	中間層	平均正答率	実行時間
250	10, 10	0.917	4 時間 25 分
250	20,10	0.925	7 時間 15 分
250	70,10	0.930	16 時間 14 分
250	140, 10	0.930	33 時間 13 分
250	150, 10	0.931	35 時間 23 分

中間層が 1 層で，ここまで性能が高いということは，比較的，構造としては簡単な構造が分類の精度に関わる可能性がある。高次の単語間の関係を考える上では，2 層以上の展開が必要となると考えられるが，単純に 2 層にするだけではなく，中間層のニューロン層を適切に選定しなければならない。現在，中間層を 2 層，ニューロン数を増やした場合の正答率について評価を重ねている。

表 2 にその結果の一部を示した。キーワード数を 250 と固定し，同様の交差検証法で正答率を評価している。

以上のごとく，高次の関係性を想定することで性能が上がる可能性があるが，その性能向上は数パーセント程度にとどまっている。このあたり，中間層の設定に試行錯誤しているが，中間層を適切に調整すれば，Random forest と同等の性能あるいはそれ以上が期待できるのかもしれない。直観的には，Deep Learner が決定木あるいは random forest のような構造を内包していることによるのかもしれない。データの潜在構造を考える上で興味深いが，今後の研究課題としたい。

4.2 実行速度

今後，誤判別例の精査およびキーワード結合の性質の吟味を行うことによって，BNN, 深層学習で得られた学習器の性質を検討する必要があると思われる。一方，BNN, 深層学習で問題になるのは，その収束速度であろう。

Fig. 4 に 100 回試行の実行速度の比較を示す。ピークであったキーワード 200 種に対して，SVM がトータルで 28 分であったのに対し，深層学習では 2 時間 51 分 (171 分) であった。一方 BNN は，キーワードが少数の時は計算速度は速かったが，増加することで，その計算時間は著しく高くなり，darch を遥

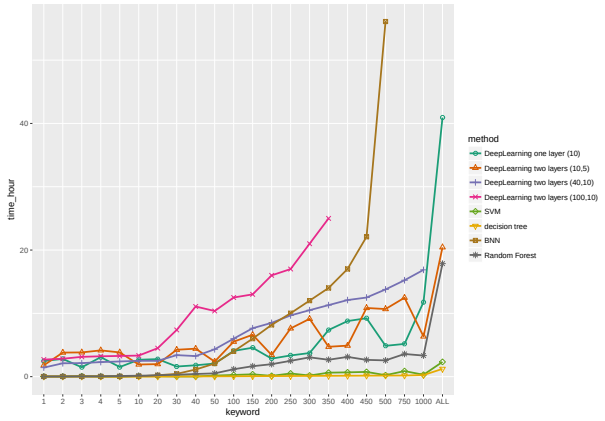

図 4: 各手法の実行速度

かにしのぐものとなった。今回はデフォルトのパラメータの設定を用いたが，今後，中間層を含めたパラメータの設定等で高速化を図れるかどうか検討する必要かある。

しかし，時間的コストを考えた場合，決定木でまず変数を選択するという方法は，今回のような分類問題では悪くない選択であろうと考えられる。その上で，さらに性能の向上を求める場合に，深層学習にて分類器を構築するというプロセスを踏むということの方がよいのかもしれない。

4.3 全プロセスの実行時間

キーワード絞り込みを 100 とした各プロセスの平均実行時間は表 3 の通りである。今後，キーワード絞り込みのスピードアップを進める必要がある。

表 3: 各プロセス平均実行時間

プロセス	平均実行時間
データ抽出	00:13:13
タームの行列作成	00:01:55
キーワード絞り込み	00:20:23
分類モデル構築 (Random Forest)	00:00:07

5. おわりに

本研究では，テキストに形態素解析を適用した後，対応分析による DPC 毎のキーワードの選定を行い，選ばれたキーワードに深層学習を適用し，キーワードによる DPC コーディングを行う分類器を構築した。昨年度登録された退院時要約を用いた検証結果から，SVM が最大 70%程度の正答率であったのに対し，決定木が 80%以上，Random Forest が 90%以上の正答率を上げ，深層学習でも，90%以上の精度を上げることができた。本データでは実行速度，性能の点を考えれば，Random Forest を分類器構築方法として選択するのが最適であると考えられたが，Deep Learner についてはパラメーター調整の余地がある。今後，データから抽出しうる潜在的な知識構造を念頭に置きながら，さらなる検証を行う予定である。

謝辞

本研究は日本医療研究開発機構・臨床研究・治験推進研究事業 15lk1010003h0001 「医用知能情報システム基盤の研究開発」の助成によった。

参考文献

[Drees 13] Drees, M.: Implementierung und Analyse von tiefen Architekturen in R, Master's thesis, Fachhochschule Dortmund (2013)

[Karatzoglou 04] Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A.: kernlab – An S4 Package for Kernel Methods in R, *Journal of Statistical Software*, Vol. 11, No. 9, pp. 1–20 (2004)

[Kim 09] Kim, J.-H.: Estimating Classification Error Rate: Repeated Cross-validation, Repeated Hold-out and Bootstrap, *Comput. Stat. Data Anal.*, Vol. 53, No. 11, pp. 3735–3745 (2009)

[Liaw 02] Liaw, A. and Wiener, M.: Classification and Regression by randomForest, *R News*, Vol. 2, No. 3, pp. 18–22 (2002)

[MeSH] MeSH, : Medical Subject Headings 2017, U.S. National Library of Medicine

[Sebastiani 02] Sebastiani, F.: Machine Learning in Automated Text Categorization, *ACM COMPUTING SURVEYS*, Vol. 34, pp. 1–47 (2002)

[Srinivasan 01] Srinivasan, P.: MeSHmap: a text mining tool for MEDLINE, in *Proc AMIA Symp*, p. 642 (2001)

[Therneau 15] Therneau, T. M. and Atkinson, E. J.: *An Introduction to Recursive Partitioning Using the RPART Routines* (2015)

[Venables 02] Venables, W. N. and Ripley, B. D.: *Modern Applied Statistics with S*, Springer, New York, fourth edition (2002), ISBN 0-387-95457-0

[外池昌嗣 大熊智子 増市博 大江和彦] 外池昌嗣 大熊智子 増市博 大江和彦 荒牧英治：退院サマリ文章可視化システムの構築, 言語処理学会第 15 回年次大会

[三浦 10] 三浦康秀, 荒牧英治, 大熊智子, 杉原大悟 外池昌嗣, 増市博, 大江和彦：電子カルテからの副作用関係の自動抽出, 言語処理学会第 16 回年次大会, pp. 78–81 (2010)

[石田 16] 石田基広：RMeCab, `http://rmecab.jp/wiki/index.php?RMeCabFunctions` (2016)

[鈴木隆弘 横井英人 井宮 淳 里村 洋一 04] 鈴木隆弘 横井英人 井宮 淳里村 洋一 小野大樹：テキストマイニングによる退院時要約自動分類の試み, 医療情報学, Vol. 24, No. 1, pp. 35–44 (2004)