

Balanced Random Forest を用いた特徴量抽出に基づく重症度判定システムの構築

Construction of decision system based on the feature extraction using Balanced Random Forest

山田 雄基 † 濱上 知樹 †
Yuki YAMADA Tomoki HAMAGAMI

1 はじめに

近年、高齢化の進行を背景として救急出動の件数が増加傾向にある。また、救急搬送者に占める軽症の傷病者の割合も依然として高く、救急サービスの質の低下が懸念されている。これを受けて日本全国で救急隊の数を増やしているが、平成 25 年において救急出動の件数は 10 年前と比較すると約 22 % と増加しているのに対し、救急隊は約 7 % の増加にとどまっている [1]。つまり、救急隊を増やすだけでは全ての救急要請に対して迅速に対応することは難しく、不要不急な救急要請を減らすことが必要だと考えられる。

そこで、不要不急な救急要請を減らすために、救急要請すべきかの判断を助ける従来システムとして救急受診ガイド [2] が東京消防庁によって導入されている。救急受診ガイドを利用することで、医学的知識の少ない通報者でも救急要請の必要性や自力受診の緊急性を自身で確認することができる。そして、これにより救急搬送者に占める軽症の傷病者の割合を低下させ、重症度の高い傷病者を優先できるという効果がある。しかし、このシステムは医師の経験則をもとに作成されているため、定量的な評価に至っておらず、判定結果のフィードバックが得られにくいという課題がある。また、利用者は該当する症状を提示される症状項目の一覧から探し出し、選択する必要があるため、利用者への負担が大きいためという課題もある。

そこで本研究では、これまで蓄積されてきたコールドトリアージの救命救急データを活用して、重症度に関わる要因を定量的に分析する。そして、それによって項目を提示していくことで判定を出す重症度判定システムを提案する。重症度の判定には二分決定木である CART を用いる。また、Balanced Random Forest によって抽出された特徴量を CART の構築に用いることで、より少ない項目で判定を出せるような利用者への負担が小さいシステムの構築を目指す。

2 救急受診ガイド

東京消防庁は、救急要請すべきかの自己判断を助ける従来システムとして平成 24 年 4 月から救急受診ガイドを導入している。救急受診ガイドはパソコンや携帯電話から利用でき、緊急性を判定するシステムである。

図 1 に救急受診ガイドの判定アルゴリズムを示す。まず、救急受診ガイドは判定の対象となる人の年代 (大人又は子ども) を利用者を選択させる。そして、緊急性の高いとされる症状項目の一覧から順に利用者に提示していく。利用者は提示される一覧から該当する症状を探し出し、一つ選択することで判定結果が得られる。このとき、提示される一覧には 10 個前後の症状項目が含まれている。緊急性に応じた判定結果として、救急要請の必要性や自力受診の緊急性、受診する診療科目を提供する。また、歩行状態などによっては緊急性を一つ上の段階に上げた結果を出力する。

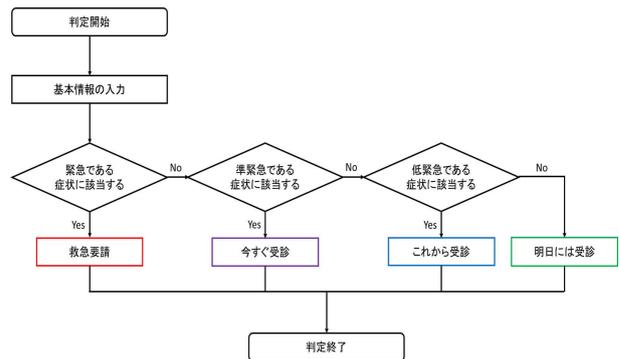


図 1 救急受診ガイドの判定アルゴリズム [2]

救急受診ガイドの導入によって、通報者は医療従事者に電話相談することなく、救急要請の必要性や自力受診の緊急性を自身で確認できるようになった。そして、電話相談件数の増加が緩和し、また救急搬送者に占める軽症の傷病者の割合も減少するなどの効果が得られている。しかし、救急受診ガイドは医師の経験則で作成されているため、定量的な評価に至っておらず、利用者は得られた結果をどれほど信頼できるのかわからないという課題がある。また、図 1 より低緊急の症状に該当する場合でも緊急、準緊急である症状項目の一覧から該当する症状を探す必要があり、利用者への負担が大きいためという課題もある。

3 提案システムの概要

救急受診ガイドの課題を踏まえて、定量的な評価ができ、利用者への負担が小さい重症度判定システムを提案する。重症度の判定は軽症か中等症以上かの 2 通りであり、判定の偏りは小さい。これは、軽症の傷病者には自力受診を促し、中等症以上の傷病者には救急要請を勧めるためである。

3.1 提案システムのアプローチ

通報者自身が扱う救急受診ガイドの他に、重症度・緊急性を判定するコールドトリアージ支援システム [3] が平成 20 年 10 月から横浜市で導入されている。これは、指令員が通報者から聴取した年齢や各症状に関する情報を入力することで、アルゴリズムに従ったカテゴリ判定の結果が得られるシステムである。判定に用いられている項目には年齢や性別などの傷病者に関する基本項目と頭痛や意識などの症状項目がある。このとき、通報者は質問に回答したり該当している症状だけを伝えれば済むため、負担は小さい。しかし、このシステムをそのまま転用するには、画面上に表示された全ての項目に対して回答する必要がある、負担が大きくなるという課題がある。

そこで、利用者の負担を小さくするため、コールドトリアージ支援システムにおける判定項目を一つずつ提示していき、利用者は Yes か No で回答するだけで済む提案システムを構築

† 横浜国立大学大学院工学部

する。提案システムでは、重症度に関わる要因を明らかにでき、二分決定木である CART を重症度の判定に用いる。また、CART の学習データに蓄積されてきたコールドトリアージデータを用いて定量的な分析を行う。

3.2 技術的課題

CART はデータマイニングの手法の一つであり、意思決定に必要な項目と判断基準を木の形で表現する手法である。それぞれの要因間の相互関係を理解しやすいという特徴がある。決定木を用いた既存研究として、大腿骨頸部骨折患者の歩行自立に関わる要因を明らかにした研究 [7] や疾患診断のための診療の流れの修正を効果的に行う研究 [8] などがある。

CART を用いることで、関連性の高い項目を順に提示するシステムを構築することができる。しかし、CART の構築では各ノードごとに最適な特徴量が選択されるため、分類性能にほとんど寄与しない特徴量でもノードによっては最適な特徴量として選択されることがある。これによって、木が深くなるという課題がある。このとき、提示すべき項目が多くなり、利用者への負担も大きくなると考えられる。

4 提案システムの構築

先の課題を解決するため、予め分類性能に大きく寄与する特徴量を抽出し、それらを CART の構築に用いる。特徴量を抽出する手法として Random Forest が研究されている [4]。これによって、どの特徴量の組み合わせに対しても分類性能が比較的高い特徴量を効率的に抽出できる。しかし、クラス間でデータ数が偏っているような不均衡データに対して Random Forest [5] を用いた場合、決定木の各ノードにおいて多数クラスの分類性能を優先した特徴量が最適な特徴量として選択されやすいという課題がある。本研究ではコールドトリアージデータが不均衡データであり、また軽症クラスの方が多いため軽症クラスの分類性能を優先した特徴量が選択される可能性がある。そこで、この課題を解決する手法として Balanced Random Forest を用いる。

これを踏まえて、提案システムの構築の流れを以下に示す。

1. コールドトリアージデータに対して前処理を行う。
2. Balanced Random Forest を用いて分類性能に大きく寄与する特徴量を抽出する。
3. 前処理されたデータセットと抽出された特徴量集合を用いて CART を構築する。

4.1 コールドトリアージデータに対する前処理

コールドトリアージデータとは、コールドトリアージ支援システムで判定に用いられる症状などに関する情報と医師が実際に傷病者を診断した結果からなる。医師が実際に診断した結果をゴールドスタンダード (GS) といい、GS の定義を表 1 に示す。

表 1 ゴールドスタンダード (GS) の定義

GS	定義
死亡	初診時に死亡が確認されたもの
重篤	生命の危険が切迫しているもの
重症	生命の危険の可能性のあるもの
中等症	生命の危険はないが入院を要するもの
軽症	入院を要しないもの

コールドトリアージデータは GS によって 5 つのクラスに分類されている。

提案システムを構築する前に学習データであるコールドトリアージデータに対して処理を行う。まず、軽症クラスと中等症以上クラスの 2 クラスに分ける。そして、データにおける基本項目や症状項目の中には 2 値をとる項目と多値をとる項目があり、判定に CART を用いるため多値をとる項目に対して 2

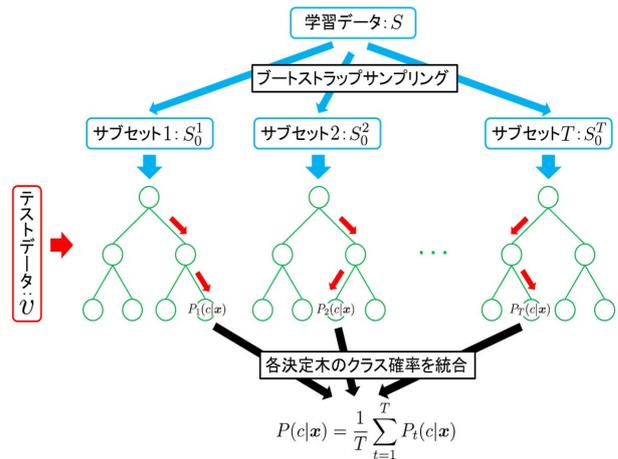


図 2 Balanced Random Forest の学習・判定

値化を行う。

4.2 Balanced Random Forest を用いた特徴量抽出

4.2.1 Balanced Random Forest

Balanced Random Forest [6] は弱学習器に決定木を用いたアンサンブル学習であり、不均衡データの場合に用いられる。決定木の学習に利用されるサブセットの作成方法が Random Forest と異なる。性質として、入力ベクトルの次元が大きくても高速に学習ができ、また高い汎化性能をもつ。図 2 に Balanced Random Forest の学習・判定の概略図を示す。

学習では、まず全学習データ集合 S から複数のサブセットを作成する。ここで、各学習データは d 個の症状などに関する情報 $x = (x_1, x_2, \dots, x_d) \in 0, 1^d$ と診断結果のクラス c からなる。また、判定に用いられる項目が特徴量 $\theta_i (i = 1, 2, \dots, d)$ になる。i 番目の決定木の学習に利用するサブセット $S_0^i \in S$ は、2 クラス間のデータ数が等しくなるようにブートストラップサンプリングによって作成される。多数クラスのデータ数を少数クラスのデータ数に合わせることで、2 クラス間のデータ数を等しくする。ブートストラップサンプリングとはランダムにデータを抽出するサンプリング手法であり、データの重複や抽出されないデータがあることを許している。

そして、サブセットから各決定木を構築する。一般的に決定木の学習では各ノードにおいて全特徴量集合が候補となるが、Balanced Random Forest の各決定木の学習では、ランダムサンプリングされた特徴量集合が候補となる。つまり、ノード j で候補となる特徴量集合 Θ_j は全特徴量集合 Θ からランダムにサンプリングされる。そして、最適な特徴量 θ^* は、候補となる特徴量集合の中から目的関数 I_j を最大とする特徴量が選択される。サンプリングされる特徴量の数の推奨値は、総数の平方根とされている。

$$\theta^* = \arg \max_{\theta \in \Theta_j} I_j \tag{1}$$

決定木が CART の場合、ジニ係数を基準に分割を行う。ジニ係数とは不純度を表す指標であり、0 から 1 の間の値をとる。0 に近いほど純度が高く、データのばらつきが少ない。ジニ係数は以下の式で表される。

$$GI(S_j) = 1 - \sum_{c \in C} P_j(c)^2 \tag{2}$$

ただし、P_j(c) はデータ集合 S_j における各クラス c のデー

タの割合である。そして、目的関数はジニ係数 $GI(S_j)$ の減少量として以下の式で表される。

$$I_j = GI(S_j) - \left(\frac{N_j^L}{N_j} GI(S_j^L) + \frac{N_j^R}{N_j} GI(S_j^R) \right) \quad (3)$$

ただし、 S_j , N_j はノード j における学習データ集合と学習データ数である。また、 S_j^L , N_j^L は分割によって左の子ノードに割り当てられる学習データ集合と学習データ数、 S_j^R , N_j^R は右の子ノードに割り当てられる学習データ集合と学習データ数である。

ここで、全てのノードが以下の終了条件を満たしたら決定木の学習を終了する。

- 予め定めた深さに達した場合
- 予め定めたノードにおける最小データ数を下回った場合
- ノードにおいて最適な特徴量 θ^* が存在しない場合

この終了条件により過学習を抑えることができる。そして、分岐が終了した葉ノードではそのノード内の全学習データに対する各クラス c のデータの割合 $P(c|\mathbf{x})$ を計算し、対応付けられる。

判定では、テストデータ \mathbf{v} を全ての決定木に入力する。そして、 t 番目の決定木において、たどり着いた葉ノードに対応付けられたクラス確率 $P_t(c|\mathbf{x})$ を出力する。つまり、 T 個の決定木の出力の平均は以下の式で表される。

$$P(c|\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T P_t(c|\mathbf{x}) \quad (4)$$

判定 c^* は全ての決定木の出力を平均することで統合し、最もクラス確率の高いクラスとする。

$$c^* = \arg \max_{c \in C} P(c|\mathbf{x}) \quad (5)$$

4.2.2 特徴量抽出

特徴量の重要度は分類性能の高さを表す。決定木における特徴量 θ_i の重要度はジニ係数の減少量の平均をとることで算出できる。そして、 T 個の決定木で構成されている Balanced Random Forest では、特徴量の重要度 $I_{BRF}(\theta_i)$ を以下の式で算出できる。

$$I_{BRF}(\theta_i) = \frac{\sum_{t=1}^T I^t(\theta_i)}{\sum_{t=1}^T N^t(\theta_i)} \quad (6)$$

ただし、 $N^t(\theta_i)$ は t 番目の決定木において特徴量 θ_i が選択された回数、 $I^t(\theta_i)$ はジニ係数の減少量の総和である。つまり、Balanced Random Forest における特徴量の重要度は、全ての決定木におけるジニ係数の減少量の平均をとることで算出できる。

図 3 に特徴量抽出のアルゴリズムを示す。特徴量を k 回削除したとき、残っている特徴量集合を $\Theta_{(k)}$ とする。まず、特徴量集合 $\Theta_{(k)}$ を用いて Balanced Random Forest を構築し、各特徴量の重要度を算出する。そして、精度判定を行ってから重要度の最も低い特徴量を削除し、残った特徴量集合 $\Theta_{(k+1)}$ を用いて同様のことを繰り返す。これを特徴量の数が閾値を下回るまで行い、下回った時点で特徴量の削除を終了する。

4.3 不均衡データにおける CART の構築

不均衡データを CART の学習に用いた場合、多数クラスの方に判定が偏るという欠点がある。つまり、コールトリアージデータをそのまま CART の構築に用いた場合には軽症クラスに判定が偏り、中等症以上の傷病者を軽症と判定しやすくなるという課題がある。そこで、多数クラスのデータに対して k -means 法を用いてクラスタリングを行う。そして、少数クラスのデータ数に合うようにそれぞれから一定数サンプリングすることでクラス間のデータ数を等しくし、構築に用いる。これにより、判定の偏りを小さくすることができる。また、ランダムにサンプリングするよりもデータの偏りをなるべく失わずにサンプリングすることができると考えられる。

5 実験および考察

判定実験によって、Balanced Random Forest を用いた特徴量抽出の有効性を確認した。また、決定木分析によって重症度に関わる要因を確認し、利用者への負担が小さい重症度判定システムの構築が可能であることを確認した。

実験 1 では、特徴量抽出を行い、特徴量削減による精度の比較を行った。また、分類性能に大きく寄与する特徴量を確認した。実験 2 では、抽出前と抽出後の特徴量をそれぞれ用いて決定木を構築し、深さにおける判定精度の比較を行った。実験 3 では、決定木分析によって重症度に関わる要因について確認した。

5.1 前処理

本実験を行う前に学習データである横浜市によるコールトリアージデータに対して処理を行った。まず、GS で 5 つのクラスに分類されているデータを軽症クラスと中等症以上クラスの 2 クラスに分けた。次に、年齢や呼吸状態のような多値をとる項目に対して 2 値化を行った。その結果、項目の数は 90 個となり、コールトリアージデータは 58069 件を実験に用いた。コールトリアージデータのクラス内訳を表 2 に示す。

表 2 コールトリアージデータのクラス内訳

クラス	軽症	中等症以上
件数	31834 件	26235 件

5.2 実験 1: 分類性能に大きく寄与する特徴量の検証

5.2.1 実験方法

特徴量抽出の流れに従って実験を行い、特徴量の数が 40 個になるまで重要度の低い特徴量の削除を行った。ここで、Balanced Random Forest ではランダム性が導入されているため、判定精度にばらつきがあると考えられる。そこで、同様の実験を 3 回行い、その平均をとることで判定精度を算出した。

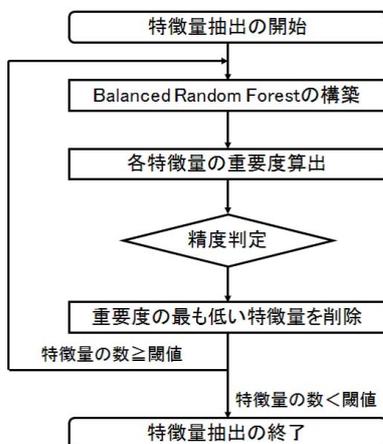


図 3 特徴量抽出のアルゴリズム

判定精度はデータセットを 10 分割して、そのうちの一つをテストデータ、残りを学習データとして交差検定によって評価した。重要度の低い特徴量は特徴量削減による精度変化を顕著にするために 2 つずつ削除した。Balanced Random Forest のパラメータは以下の通りに設定した。

- 決定木の数：60
- 決定木の最大の深さ：40
- ノードにおける最小データ数：10

また、ランダムサンプリングされる特徴量の数は特徴量抽出により特徴量の総数が減少するため、推奨値を基に表 3 に示すように減少させた。

表 3 ランダムサンプリングされる特徴量の数

特徴量の総数:M	ランダムサンプリングされる特徴量の数
81 < M ≤ 90	10
64 < M ≤ 81	9
49 < M ≤ 64	8
40 < M ≤ 49	7

5.2.2 実験結果と考察

まず、図 4 に抽出前の特徴量集合を用いて Balanced Random Forest を構築し、重要度を算出したときの下位 10 項目を示す。この結果より、最も重要度の低い項目が「その他の熱傷」であることがわかる。ここで、「その他の熱傷」に該当しているデータは全データの内、29 件であった。これより、該当しているデータが少ない項目は重要度が低くなる傾向があると考えられる。

次に、図 5 に特徴量を削除したときの精度変化の結果を示す。この結果より、各クラスの精度は偏ることなく下がっていくことがわかる。これは、Balanced Random Forest を用いたことにより、不均衡データの影響を受けずに分類性能の低い特徴量を削除することができたためだと考えられる。また、特徴量の数が 68 個になるまでは少しずつ精度が下がり、それ以降では大きく精度が下がっていることがわかる。これより、特徴量の数が 68 個になるまでは重要度の低い特徴量を削除することができたと考えられる。よって、90 個のうち分類性能に大きく寄与する特徴量として最大で 68 個までに削減できることがわかる。

この実験で抽出された特徴量集合を決定木の構築に用いる。ここで、68 個より少ない特徴量を抽出し、それらを用いた場合、分類性能に大きく寄与する特徴量も削除しているため精度は抽出前よりも下がると考えられる。一方、68 個以上では抽出前と同等の精度が得られると考えられるが、どの特徴量集合を用いたときに最も木が浅くなるかはわからない。そこで、抽

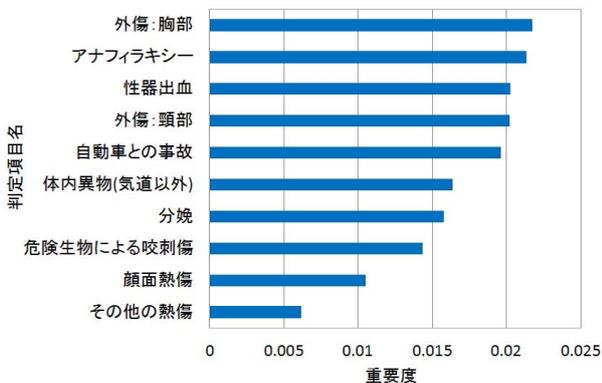


図 4 重要度の低い 10 項目

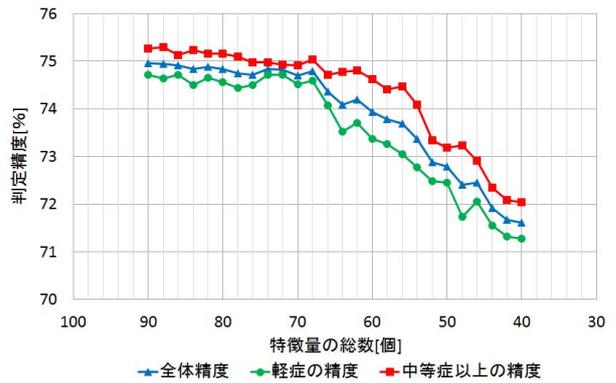


図 5 特徴量抽出による精度変化

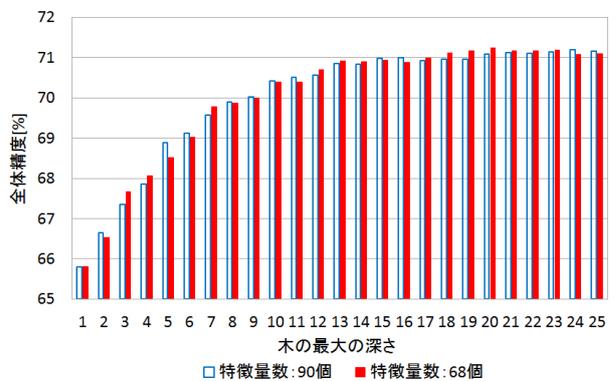


図 6 決定木の最大の深さと精度

出された特徴量集合をそれぞれ決定木の構築に用いて、比較する必要がある。

5.3 実験 2: Balanced Random Forest を用いた特徴量抽出の有効性の検証

5.3.1 実験方法

少数クラスのデータ数をクラスタ数とし、k-means 法を用いて多数クラスのデータに対してクラスタリングを行った。そして、それぞれからデータを一つずつサンプリングすることで、クラス間のデータ数を同じにして実験を行った。決定木のノードにおける最小データ数は、Balanced Random Forest のパラメータと同様に 10 と設定した。また、実験 1 と同様に交差検定によって判定精度を評価した。

抽出前の 90 個の特徴量と抽出された 68 個の特徴量をそれぞれ用いて、決定木を構築した。このとき、最大の深さは 1~25 まで 1 刻みに変化させた。また、実験 1 によって抽出された特徴量の数を 90~68 まで変化させ、決定木を構築した。このとき、最大の深さは 15~30 まで 1 刻みに変化させた。

5.3.2 実験結果と考察

まず、図 6 に抽出前と抽出された 68 個の特徴量をそれぞれ用いたときの最大の深さにおける判定精度を示す。この結果より、木を深くするほど精度は上がる傾向にあり、収束していることがわかる。これより、木の深さと精度はトレードオフの関係であり、多くの症状を提示した方がより高い精度が得られやすいが、多すぎると利用者への負担が大きくなると考えられる。つまり、同等の精度であれば浅い木の方がより良いと考えられる。

次に、図 7 に精度が収束している部分を拡大した結果を示

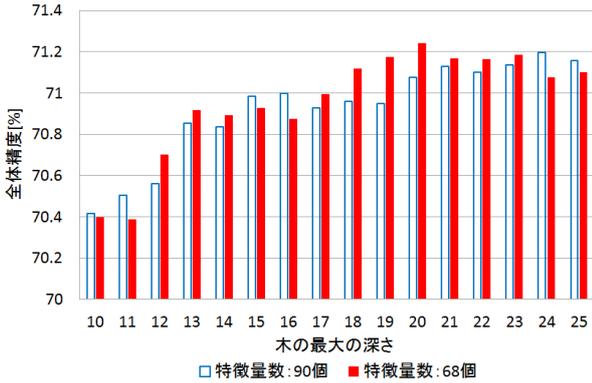


図7 決定木の最大の深さと精度 (拡大)

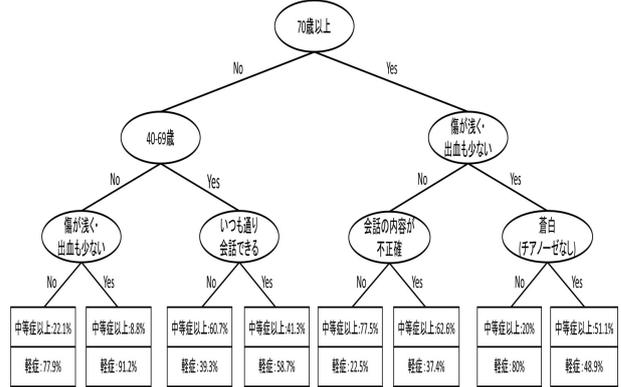


図9 重症度に関する決定木 (深さ:3)

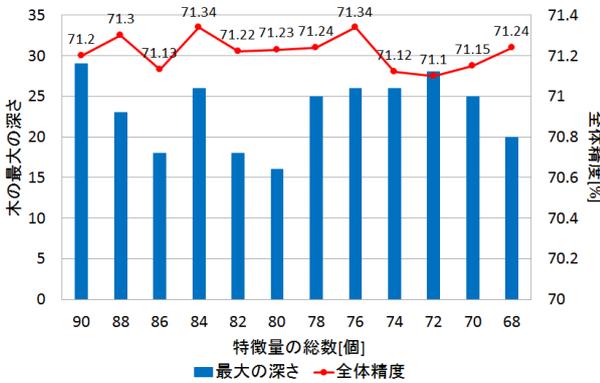


図8 特徴量抽出による最大の深さと精度

す。この結果より、抽出前の特徴量集合を用いた場合は深さが24のときに最も精度が高く、抽出後の特徴量集合を用いた場合は深さが20のときに最も精度が高くなっていることがわかる。これは、抽出前の特徴量集合には分類性能にほとんど寄与しない特徴量もあり、そのような特徴量が組み合わせによって最適特徴量として選択されたために、木が深くなってしまったと考えられる。一方、抽出後の特徴量集合ではそのような特徴量が除いてあるため、木を浅くできることが明らかになった。

さらに、図8に抽出された各特徴量集合における最も高い判定精度が得られたときの最大の深さとそのときの判定精度を示す。この結果より、抽出された特徴量を用いた方が全体的に抽出前より木が浅い段階で同等の精度またはそれ以上が得られていることがわかる。また、抽出された80個の特徴量を用いたときの深さが16で最も浅くなり、84個と76個の特徴量を用いたときの精度が71.34%で最も精度が高くなっていることがわかる。これより、提案システムの構築において精度を重視するか利用者への負担を小さくすることを重視するかによって、用いる特徴量集合を変えられることが明らかになった。

これらの結果より、Balanced Random Forestを用いた特徴量抽出の有効性が示された。

5.4 実験3: 重症度の関わる要因の検証

5.4.1 実験方法

実験2と同様にk-means法を用いてクラス間のデータ数を等しくしてから決定木を構築した。特徴量の数抽出された68個とし、最大の深さは3とした。

5.4.2 実験結果と考察

図9に決定木分析をした結果を示す。四角には軽症クラスと中等症以上クラスのデータの割合、丸には選択された項目が示されている。この結果より、重症度に関わる要因として最も大きい項目が「70歳以上」であることがわかる。また、決定木を用いたことで重症度に関わる要因の組み合わせとそれに伴う重症度が明らかになった。重症度の要因を明らかにできることを確認するために深さを3としたが、深くすることでより複雑な要因の組み合わせを表現することができる。

しかし、「70歳以上」でNoと回答した場合、次に聞かれる項目としては「40~69歳」であることがわかる。つまり、本来は年齢に関する項目として一回聞けば済むはずが、CARTを用いたために似たような項目をもう一度聞く必要があることが明らかになった。

6 おわりに

本稿では、定量的な評価ができ、利用者への負担が小さい重症度判定システムを提案した。

Balanced Random Forestを用いた特徴量抽出の有効性を確認するために2つの実験を行った。実験1では、特徴量抽出によって分類性能に大きく寄与する特徴量の数を確認された。実験2では、抽出前の特徴量集合をCARTの構築に用いた場合と抽出された特徴量集合を構築に用いた場合の比較を行った。その結果、抽出したことで木を浅くすることができ、同等または抽出前より高い精度が得られることが明らかになった。つまり、特徴量抽出により利用者への負担を小さくできることが確認された。

また、重症度に関わる要因を確認するために実験を行った。実験3では、決定木分析によって重症度に関わる要因を明らかにし、順に項目を提示する提案システムの構築が可能であることが確認された。

今後は、判定にCARTを用いたことによって生じた課題の解決方法を模索し、判定精度の向上方法を検討し、重症度の判定に加えて緊急性も判定できるシステムを検討する予定である。

参考文献

- [1] 総務省消防庁: 平成26年版 消防白書, pp.170-187, 2015.
- [2] 東京消防庁: 東京版救急受診ガイド, <http://www.tfd.metro.tokyo.jp/hp-kyuimuka/guide/08b1/index.html>, 2012.
- [3] 横瀬賢人, 濱上知樹: 決定木を用いた近傍空間内でのサポートベクターマシンによるコールトリアージ判定支援の検討,

計測自動制御学会システム・情報部門学術講演会講演論文集, 2012.

- [4] 三品陽平, 嶋崎克也, 藤吉弘亘: Random Forest の寄与率を用いた効率的な特徴選択法の提案, 第19回画像センシングシンポジウム, 2013.
- [5] L. Breiman: Random Forests, Machine Learning, vol.45, pp.5-32, 2001.
- [6] C. Chen, A. Liaw, L. Breiman: Using Random Forest to Learn Imbalanced Data, Technical Report, No.666, 2004.
- [7] 新井智之, 金子志保, 藤田博暁: 大腿骨頸部骨折患者の歩行自立に必要な要因—決定木分析による検討—, 日本老年医学会雑誌, vol.48, No.5, pp.539-544, 2011.
- [8] 五福明夫, 山崎晃, 光岡和彦, 橋詰博行: 疾患診断における知識の表現と効果的な手順の構成, システム制御用学会論文誌, vol.20, No.3, pp.130-132, 2007.