CF-006

健康診断予測における経年予測の解釈性を考慮した予測モデルの構築
Learning Medical Checkup Prediction Model with Interpretable Interannual Prediction

谷口 敦† 玉野 浩嗣†‡
Atsushi Taniguchi Hiroshi Tamano

1. はじめに

多くの企業では社員の生活習慣病を初めとする様々な病気の早期発見や治療、または病気の予防や健康維持を目的とした健康診断が実施されており，健康診断の結果から生活習慣病あるいはその予備群と判断された従業員に対して，保健指導が行われることがある．

保健指導の支援を目的として，それまでに蓄積された健診データを機械学習して将来の検査項目の値を予測し，健康状態を可視化する試みがある．保健指導の対象者が予測される将来の検査項目の具体的な値や傾向をグラフで閲覧することで，自身が将来どのような健康状態になるかを直感的にイメージできるようになり，健康が悪化しないように前もって対処できるようになる．

将来の検査項目の値の予測は回帰問題として捉えることができ，過学習を防ぐための正則化パラメータを用いる場合には，いくつかパラメータを設定したモデルを複数作成した中から，最も精度が良いモデルを選択することが基本的な考え方である．

正則化パラメータの調整によるモデル作成を複数回行った際に，最も精度が良いモデルと比べて十分に精度が良いモデルがいくつか生成されるが，その中には検査項目の値の経年予測の解釈が困難な予測を多く出力するものとそうでないモデルが混在する．そのため，精度の良さのみでモデルを選択すると，経年予測の解釈が困難な予測を多く出力するモデルを選択してしまう可能性がある．

予測の目的を考えると，モデルの選択は単に精度が良いだけでなく，経年予測の傾向の解釈が容易であるかどうかという観点でも行うことが望ましい．しかし，解釈が容易であるかどうかの判断は専門的な知識を要するドメインエキスパートが実施する必要があり，さらに解釈が容易であるかどうかの判断をするためには多くの予測サンプルを確認する必要があるため，ドメインエキスパートには大きな負担がかかる．

そこで，本研究では，ドメインエキスパートに依存しているモデルから出力される検査項目の値の経年予測の解釈が容易であるかどうかの判断を計算可能にするスコアリング方法を提案すると共に，そのスコアリング方法を用いた精度と経年予測の傾向の解釈性の両観点から自動的にモデル選択を行うモデルの構築方法について提案する．

2. 背景

本章では，健康診断と保健指導について，そして保健指導時に使用することを想定した NEC ソリューションイノベータで開発を進めている健診結果予測シミュレーションの詳細について記載する．

† NEC ソリューションイノベータ イノベーションラボラトリ, NEC Solution Innovators Innovation Laboratories
‡ NEC データサイエンス研究所, NEC Data Science Research Laboratories

2.1 企業における健康診断と保健指導

多くの企業で様々な病気の予防や健康維持を目的とした定期健康診断が全従業員を対象に年に 1 度実施されており，その結果，生活習慣病などの健康リスクが高いと判断された従業員に対しては健康診断の事後措置として，産業医や保健師による保健指導が実施されることがある．保健指導では定期健康診断結果の再確認や生活習慣の見直しなど，保健指導の対象者の生活改善に向けた指導を行う．

2.1.1 定期健康診断

企業での定期健康診断は，労働安全衛生法に基づき，全従業員を対象に年に 1 度実施され，一般的に問診と健診が実施される．問診では自覚症状や他覚症状の有無の他に，食事や飲酒，運動などの生活習慣に関しても記録する．健診では身長，体重，腹囲，血圧などの計測，また血液検査を行い，中性脂肪やコレステロール値，血糖値などを測定する．なお，今回予測に使用する問診及び健診データの詳細は後述する．

2.1.2 保健指導

定期健康診断の結果に応じて，産業医などが必要であると判断した場合，健康診断の事後措置として産業医や保健師による保健指導が実施されることがある．例えばメタボリックシンドロームと診断された者やその予備軍を対象に，定期健康診断結果の再確認，生活習慣の見直し，改善目標と改善方法を考えるなど，保健指導対象者の生活改善に向けた指導を行う．

2.2 健診結果予測シミュレーション

NEC ソリューションイノベータでは保健指導時に将来の検査項目の値をシミュレートする 健診結果予測シミュレーション(図 1) を開発している．

図 1 健診結果予測シミュレーション

健診結果予測シミュレーションは生活習慣病に関連する 9 種類の検査項目について予測を行い，例えば保健指導時に

保健師が，保健指導対象者に生活習慣の改善指導を行う際に使用される．システムでは 1 年ごとに将来の検査項目の具体的な値を予測し，将来の健康状態の傾向をグラフで可視化する．その結果，保健指導対象者は自身が将来どのような健康状態になるかを直感的にイメージできるようになるため，自身の行動変容に向けたきっかけを与えることができる．

2.3 検査項目の予測方法

健診結果予測シミュレーションでは検査項目の値の予測を行う．説明変数に定期健康診断で取得可能な過去の検査項目の値 $\boldsymbol{x_k}$ と生活習慣に関する問診項目 $\boldsymbol{x_m}$ を入力に将来の検査項目の値 y を予測するモデルを作成する．本研究では学習器に異種混合学習 [4] を用いる．異種混合学習により構築された予測モデルを式 1 とする．

$$y = f_{FAB}(\boldsymbol{x_k}, \boldsymbol{x_m}) \quad (1)$$

予測シミュレーションでは式 1 の予測モデルに予測を行いたい生活習慣 $\boldsymbol{x_m}$ を入力し，入力した生活習慣で今後過ごしていくと，将来どのような健康状態になるかのシミュレーションを行なうことができる．また，f_{FAB} の学習では L0 正則化が行われ自動で属性が選択される．加えて L2 正則化パラメータのチューニングを行なう．

2.4 検査項目の値の予測結果の解釈性

検査項目の値の予測は通常，回帰問題として扱われ，過学習を防ぐための正則化パラメータを用いる場合には，正則化パラメータの調整によるモデル作成を複数回行い，最も精度が良い時のモデルを選択する．しかしながら，異なる正則化パラメータが与えたれたモデルの中には，ほとんど精度が同じモデルがいくつか生成され，それらの中には，検査項目の値の経年予測の解釈が困難な予測を多く出力するものとそうでないモデルが混在する．例として，運動をあまり行わず，夕食後に週に 3 日以上間食をしているような，食生活をあまり気にしない生活習慣を 3 年続けた場合の体重の経年予測の例を以下の図 2，図 3 に示す．

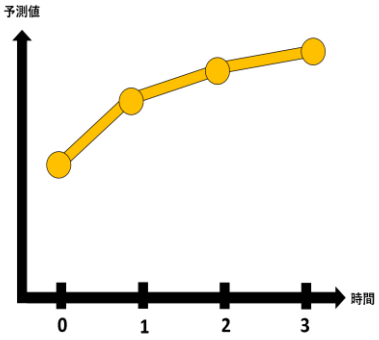

図 2　モデルから出力される経年予測の例 1

産業医や保健師といったドメインエキスパートによると，同じ生活習慣を 3 年間続けた場合，図 2 のグラフのように，検査項目の値の変化の傾向が変わらないのが自然であり、解釈し易い予測結果である．

一方、図 3 のグラフのように変化の傾向が急に変わっているのは不自然であり，解釈することが困難な予測結果であると言える．

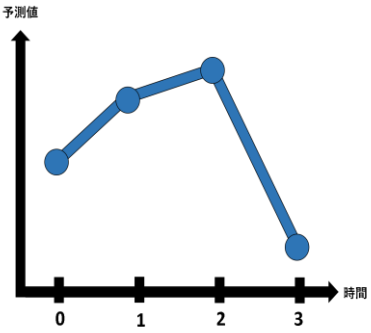

図 3　モデルから出力される経年予測の例 2

予測される検査項目の経年変化の傾向が解釈し易いものであると，保健師が納得の行く保健指導ができるようになり，保健指導対象者の行動変容に向けたきっかけを与えやすくなる．そのため，選択される予測モデルは，単に精度が良いのみではなく，モデルから出力される経年予測サンプルの傾向がユーザにとって解釈し易いものであることが望ましい．

3. 研究目的

2.4 節より，健診結果予測シミュレーションにおける予測モデルの選択では，予測精度が高いことに加え，さらに出力される経年予測が解釈できることが必要となる．

しかし，複数の予測モデルの中から，ドメインエキスパートの観点で，経年予測の解釈が困難な予測を最も少なく出力するモデルを選択するには，現状それらの予測モデルから得られる経年予測をドメインエキスパートが目視し，どのモデルが良いかを判断する必要がある．これらの作業は，たくさんの経年予測の結果を目視することになるため，非常に負担が大きい．

そこで本研究では，経年予測の解釈性の判断を計算可能にするスコアリング方法を提案すると共に，その方法を用いて，精度と経年予測の解釈性の両方を備えたモデルの構築方法を提案する．

4. 関連研究

健診結果予測シミュレーションの他にも，山本ら [1] は，血圧やコレステロール値などの健康診断データ 4 年分を機械学習することによって，5 年目の健康診断項目の値が 4 年目よりも大，小，同程度の 3 つのクラスのうちのどのクラスに分類されるかを予測する方法を提案している．また，長谷川ら [2] は状態空間モデルを用いた検査項目の予測を行っている．健康診断データが，時系列であることに着目し，前年度の生活習慣と健診結果が，次年度の健診結果に影響するような状態空間モデルを仮定し，検査項目の結果を予測している．中島ら [3] は生活習慣データの取得から実際に健康状態に現れるまでの遅延時間に着目した手法により，個人ごとに生活習慣と健康状態の間の規則性についてデータマイニングを行っている．

このように，健康診断データを機械学習する研究は多く提案されているが，これまでに検査項目の値を経年予測した際の解釈性について注目した報告は無い．

5. 提案手法

本章では，解釈性の判断を計算可能にするスコアリング方法と，そのスコアリング方法を使ったモデル構築手法について記載する．

5.1 解釈性の評価指標 NAScore

2.4 節で述べたとおり，ある生活習慣を一定期間続けた場合，検査項目の値の経年予測の傾向が変化しないものは解釈し易く，変化するものは解釈することが難しい．加えて，生活習慣を続けた場合，検査項目の値が発散していくことも考えにくい．つまり，ある生活習慣を続けた場合の経年予測は，漸近モデルに近いときに解釈が容易であると考えられる．そこで，経年予測の解釈の難しさを，漸近モデルと予測系列とのフィッティング誤差でスコア化する．(これを以降 NAScore: Non Asymptotic Score と呼ぶ) NAScore は，経年予測が漸近モデルから外れている（解釈が困難である）場合に大きくなり，漸近モデルと近い(解釈が容易である)場合は小さくなる．

NAScore を算出する際に，X＝年数と Y＝検査項目の値として漸近モデルでの回帰誤差を求めると，ユーザが実際に見る表示(図 1)とは異なる状況でのスコア計算となる．そのため，実際のグラフ表示と同じ状況でのスコア計算するために X の値をリスケールする必要がある．

今回提案する評価指標，NAScore のアルゴリズムを Algorithm1 に示す．

Algorithm1： 解釈性評価スコア算出

Input 4 年分の系列データ $D = \{y_1, y_2, y_3, y_4\}$,
x 軸のリスケールパラメータ x_{scale}

Output $Score$

1. $x_n = n\ x_{scale}\quad, n \in \{1,2,3,4\}$
2. $Score\ = \min_{\theta} \sum_n |f(x_n,\theta) - y_n|^2$
   $f(x,\theta) = c + ba^x,\ \theta = \{a,b,c\}$
   ただし、$0 < a < 1$
3. **return** $Score$

5.2 NAScore を用いたモデル構築方法

本章では，5.1 節で記述した解釈性の評価指標である NAScore を用いたモデル構築方法について述べる．

検査項目の予測は回帰問題として扱い，過学習を防ぐために正則化パラメータを用いる．通常，正則化パラメータの調整によるモデル作成を複数回行い，作成されたモデルの中から，精度が最も良いモデルを選択する．しかし，今回の手法では経年予測の解釈性の指標も用いるため，最も精度の良いモデルと比べ，十分に精度が良いモデルの中で，解釈性の低いサンプルの数が最も少ないモデルを選択する．また，本研究では学習器に異種混合学習[4] を用いる．異種混合学習では初期値依存性のある学習アルゴリズムであるため，初期値を変えて複数回リスタートを行う．具体的なモデル構築アルゴリズムを Algorithm2 に示す．

Algorithm2： 解釈性を考慮したモデル構築

Input 正則化パラメータ $\lambda_k\ (k = 1, \ldots, K)$,
学習用データ D_{train}, モデル選択用データ $\mathrm{D}_{\mathrm{select}}$,
リスタート回数 T, 精度閾値 α, 解釈性閾値 β

Output モデル W

1. **for** k **= 1 to** K
2. **for** t **= 1 to** T
3. 学習用データD_{train}と正則化パラメータλ_kからモデルW_t^kを学習
4. モデル選択用データD_{select}を使いモデルW_t^kの RMSE を計算
5. **end for**
6. $W_1^k, \ldots, W_T^k$の中で RMSE が最も良いモデルを選択し，これをW_{best}^k する．
7. **end for**
8. $W_{best}^1, \ldots, W_{best}^K$ の中から，RMSE が最も良いモデルを選択．これをW_{best}^{best} とする．
9. W_{best}^{best} の RMSE との差がα% 以内のモデルを $W_{best}^1, \ldots, W_{best}^K$から選択し，これを$W_{best}^p$ (p = 1, … , P)とする．
10. **for** p **= 1 to** P
11. W_{best}^pについてモデル選択用データD_{select}を使い 1～3 年後までの検査項目の値を予測する．ただし，1～3 年後までの生活習慣に関する説明変数は予測を行う時点のものと同一にして予測を行う．
12. 各予測系列に対して NAScore を算出する．
13. NAScore が閾値βを超える予測系列を解釈性の低いサンプルとしてカウントする．
14. **end for**
15. W_{best}^p (p = 1, … , P)の中から，解釈性の低いサンプル数が最も少ないモデルW を選択する．
16. **return** W

6. 評価

6.1 NAScore の評価

5.1 節で提案した経年予測の解釈性の評価指標 NAScore による解釈性の判定と従来手法 (保健師) による判定 を比べ，高い精度で判断できているか，解釈性の低いサンプルの数の計算時間が短縮されるかどうかを評価する．

6.1.1 評価方法

評価には，Hba1c[%]，空腹時血糖(FBS)[mg/dL]，HDL コレステロール（HDL）[mg/dL], LDL コレステロール（LDL）[mg/dL]，中性脂肪（TG）[mg/dL]，収縮期血圧[mmHg]，拡張期血圧[mmHg]，体重[kg]，腹囲[cm]の 9 つの検査項目の実測値についての 2011 年～2014 年の 4 年分のデータを用いる．サンプル数は空腹時血糖（FBS）で 928 サンプル，空腹時血糖（FBS）以外の検査項目は 464 サンプルを用いた．まず，4 年分の経年実績データを生活習慣が変化しなかった時のものとみなし，保健師が解釈容易であるか否かの正解ラベル付けを行う．次に 5.1 節で提案した NAScore によっ

てスコアを算出し，そのスコアが予め設定した閾値を超えたかどうかで，解釈が容易であるか否かの 2 値判別を行った．その結果を比較し，判別結果の正誤を判断する．また，実行時間については保健師によるラベル付け時間と，NAScore の計算による判別にかかった時間を比較する．

6.1.2 評価結果

9 つの検査項目について，経年実績のグラフの解釈が容易であるか否かの判断を NAScore で行った際の ROC 曲線を図 4 に示す．

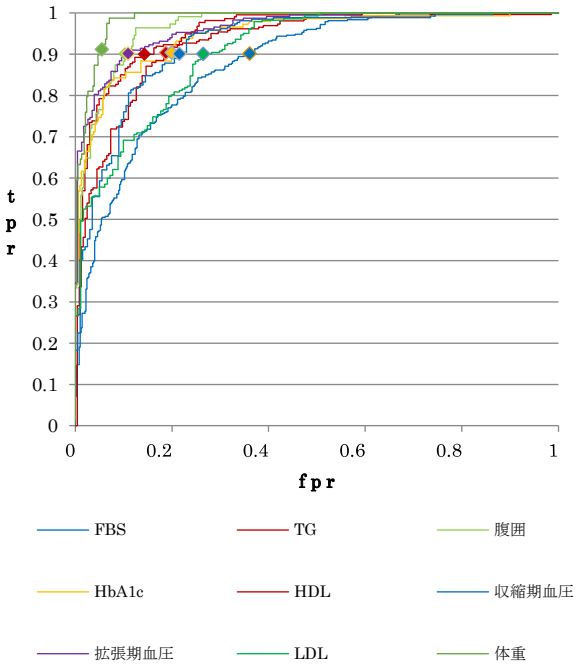

図 4　NAScoreによる判別のROC曲線

また，TPR＝0.9 に固定したときの各検査項目の FPR を表 1 に示す．

表 1　TPR=0.9 のときの各検査項目の FPR

検査項目	TPR=0.9のときのFPR
HbA1c	0.199
FBS	0.361
HDL	0.187
LDL	0.295
TG	0.142
収縮期血圧	0.215
拡張期血圧	0.109
体重	0.055
腹囲	0.102

NAScore で判別した際に，実際には解釈性の低いサンプルではないが，解釈性の低いサンプルであると判別されてしまう割合である FPR の平均は 0.18 ほどで，高い精度で経年予測の傾向の解釈が容易か否かを判別できることがわかる．また，保健師が，検査項目の値の経年予測の傾向の解釈が容易であるか否かを判断する際，927 サンプルに対して約 5 時間 30 分かかっていたのに対して，NAScore での算出時間は約 13.7 秒であった．

6.2 モデル構築方法の評価

本節では 5.2 節で提案した NAScore を用いたモデル構築方法が，精度だけからモデル選択を行う方法（従来方法）に対し，解釈性の低いサンプルが少ないモデルを安定して選択できるか否かの評価を行う．

6.2.1 評価方法

データセットとして 1 年に 1 回実施される定期健康診断の 2009 年～2014 年の 6 年分，2540 名分のデータを用いる．その中から，学習データに 2009 年～2011 年のデータ 904 名分，モデル選択用データに 2009 年～2011 年のデータ 226 名分，テストデータに 2012 年～2014 年のデータ 1410 名分を用いた．目的変数に Hba1c[%]，空腹時血糖(FBS)[mg/dL]，HDL コレステロール（HDL）[mg/dL]，LDL コレステロール（LDL）[mg/dL]，中性脂肪（TG）[mg/dL]，収縮期血圧[mmHg]，拡張期血圧[mmHg]，体重[kg]，腹囲[cm]の 9 つの検査項目を設定し，それぞれ予測モデルを構築する．説明変数には，1 年前や 2 年前の過去の検査項目の値，定期健康診断に取得可能である生活習慣に関する問診項目のうち，特に検査項目に関係のあると考えられる 24 項目を設定した．問診項目には例えば厚生労働省の標準的な質問表にあるような「睡眠で休養が十分とれている」というような睡眠に関する項目や「人と比較して食べる速度が早い」や「朝食を抜くことが週に 3 回以上ある」といった食事に関する項目，「1 日 30 分以上の軽く汗をかく運動を週 2 日以上 1 年以上実施しているか」というような運動習慣に関する項目を使用した．また，前処理として，欠損値のあるあるサンプルは除外し，ドメインエキスパートが外れ値と判定する中性脂肪が 300 [mg/dL] 以上のサンプル，空腹時血糖 200 [mg/dL]以上のサンプルを除外し，血圧や脂質，糖尿への服薬をしているサンプルは除外した．

学習器には異種混合学習[4] を用い，検査項目の予測モデルを学習する．Algorithm2 の正則化パラメータは，L2 正則化パラメータとする．従来手法，提案手法共に，6.2.1 節で述べた前処理を行った後，従来手法は予測精度のみでモデルの選択（Algorithm2 のW_{best}^{best}を出力する方法），提案手法は Algorithm2 の方法により，モデルの構築を行う．従来手法，提案手法が出力するそれぞれのモデルに対し，同じテストデータによる評価として精度と解釈性の低いサンプル数を算出し，比較する．

学習アルゴリズムに初期値依存性があり，同じデータを使用しても，すべての結果に異なる値が生じるため，従来手法，提案手法共に比較をするまでの一連の流れを 10 回ずつ行い，提案手法が安定して解釈性の低いサンプル数が少ないモデルを構築できるかどうかを評価する．なお，提案手法，従来手法ともに 12 種類の L2 正則化パラメータ，リスタート回数 T＝10 を用い，提案手法のモデルの選択アルゴリズムの閾値 α は 5 %，解釈性の閾値 β は TPR=0.9 となる閾値を用いた．

表 2　従来手法と提案手法の解釈できないサンプル数と RMSE の相対比

試行回数	HDL			LDL			TG		
	従来手法 解釈できない サンプル数	提案手法 解釈できない サンプル数	提案手法 RMSE相対比 （従来手法＝1）	従来手法 解釈できない サンプル数	提案手法 解釈できない サンプル数	提案手法 RMSE相対比 （従来手法＝1）	従来手法 解釈できない サンプル数	提案手法 解釈できない サンプル数	提案手法 RMSE相対比 （従来手法＝1）
1	65	16	1.013	70	3	1.015	49	7	1.033
2	64	12	1.054	72	3	1.021	78	5	1.065
3	57	9	1.055	15	3	1.041	78	7	1.050
4	27	15	1.009	39	3	1.021	78	7	1.052
5	60	15	1.020	70	3	1.019	80	7	1.079
6	91	15	0.995	73	3	1.014	88	7	1.054
7	90	15	0.994	62	3	1.017	10	7	1.003
8	91	15	0.991	52	3	1.017	91	5	1.067
9	60	15	1.014	26	3	1.031	79	7	1.052
10	44	34	0.950	71	4	1.021	84	7	1.051

6.2.2　評価結果

従来手法と提案手法の評価をそれぞれ 10 回ずつ行った結果を表 2 に示す．また，従来手法と提案手法の解釈性の低いサンプル数の平均と標準偏差を図 5 に，RMSE の相対比の平均と標準偏差を図 6 に示す．

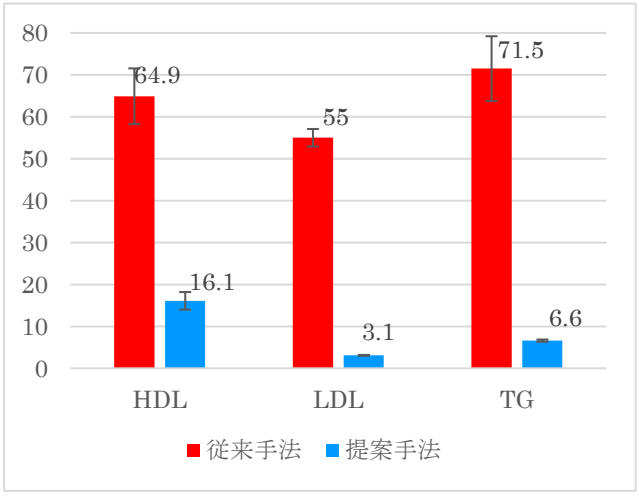

図 5　従来方法と提案方法の解釈性の低いサンプル数

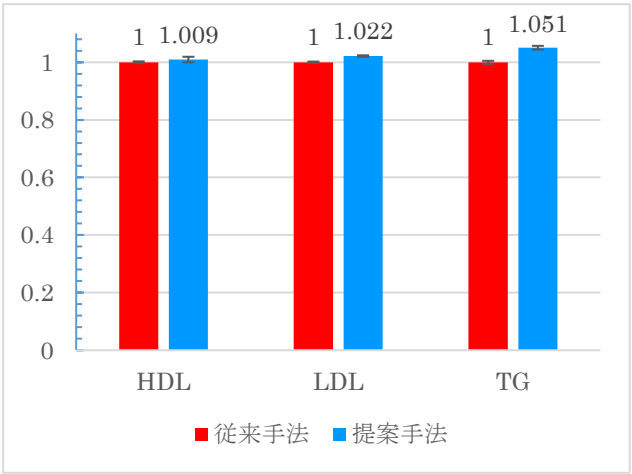

図 6　従来方法と提案方法の RMSE の相対比

HDL，LDL，TG 以外の検査項目の解釈できないサンプルはいずれも従来手法において 3 サンプル以下しか生成されないため，省略する．

従来手法ではほとんどの試行で解釈性の低いサンプル数が多いが，提案手法では LDL，TG で 10 回中 10 回，HDL で 10 回中 9 回と，経年予測の傾向の解釈が容易であるモデルを安定して選択可能であることがわかる．また提案手法の精度は，従来手法に比べ HDL，LDL，TG 共に平均で約 5 % の範囲で精度劣化を抑えることができ，従来手法の精度と同等であることが示された．よって提案手法は従来手法から精度を維持したまま，解釈性の低いサンプル数を削減可能であることがわかる．

次に LDL について，正則化パラメータを変えて学習したモデルに対しモデル選択データD_{select}を使って評価した時の精度（最も良いものを 1 としたときの RMSE の相対比）と解釈性の低いサンプル数のグラフを図 7 に示す．HDL，TG についても図 7 と同様のグラフ傾向であるため，省略する．

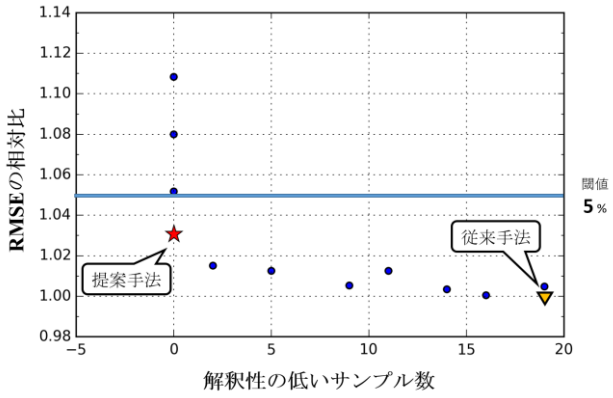

図 7　モデル選択データにおける，各W_{best}^{k}の精度と解釈性の低いサンプル数

従来手法では図 7 の三角印のように，予測精度のみでモデルを選択していたため，解釈性の低いサンプルを多く生成してしまうモデルを選択してしまう．しかし，提案手法では図 7 の星印のような，最も RMSE が良いモデルから閾値以内の精度で，解釈性の低いサンプル数が最も少ないモデルを選択することができる．

7. まとめ

本研究では，モデルから出力される検査項目の値の経年予測の傾向の解釈が容易であるか否かを計算可能な評価方法である NAScore を提案し，NAScore を用いたモデルの構築方法を提案した．提案した評価方法である NAScore により，従来ドメインエキスパートである保健師にしか行えなかった，経年予測の傾向の解釈が容易であるか否かの判断を，高い精度でかつ短時間に行えるようになった．

従来手法と提案手法である NAScore を用いたモデルの構築方法をそれぞれ 10 回ずつ施行した結果，従来手法では経年予測のサンプル傾向の解釈が困難なモデルを選択してしまう場合があるのに対し，提案手法では解釈が容易であるモデルを安定して選択することができた．また，提案手法の精度は，HDL，LDL，TG 共に平均で約 5%以内で収まり，従来手法の精度と同等であることが示された．

今回はモデル選択基準に経年予測の解釈が容易であるかという，解釈性を考慮した指標を取り入れた．今後の課題として，解釈性を考慮したその他の指標，例えば，ある検査項目が上昇すると必ず他のある検査項目も上昇するといった検査項目間の関係性に関わるものや，翌年の予測値の変化幅が急激に増加，減少しないなどの評価も検討していきたい．

また，検査項目の予測精度向上のために，取得できていない他の生活習慣のデータを収集し，説明変数に盛り込むことやデータの取得間隔を狭め，検査項目の値の細かな推移を取得することも検討していきたい．

参考文献

[1] 山本有紗，石川由羽，梅田智広，城和貴，“時系列健康診断データからの未病予測の一手法”，研究報告数理モデル化と問題解決 (MPS), (2017)

[2] 長谷川 嵩矩，新井田 厚司，山口 類，井元 清哉，“L1 正則化付き状態空間モデルを用いた健診結果の将来予測”, 2017 年度統計関連学会連合大会, (2017).

[3] 中島 緋沙恵，竹内 裕之，“個別健康管理システムにおけるデータマイニング手法の研究”，第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2018)， K6-2, (2018).

[4] 藤巻 遼平，森永 聡，“ビッグデータ時代の最先端データマイニング”， NEC 技報 Vol65, No.2, pp.81-85, (2012).