

集合知定理に基づくクラス判別法の提案と 医療分野への応用に関する検討

吉開範章¹ 栗野俊一¹ 橋口徳一¹ 加澤佳奈² 森山美知子²

概要: ノイズを多く含んだデータに対するクラス判別システムに関する検討を行い、時系列変化も考慮して教師データ自体にノイズが含まれる場合の改善法として集合知定理を活用する方法を示す。その応用分野として、保健指導における患者の重症化リスクを予測する AI を試作し、専門家の予測との一致率を評価した。その結果、人の目視で見逃していた高リスク者を特定可能であり、さらに、提案手法により一致率の予測精度の向上が期待できる事を確認した。

キーワード: 深層学習、人工知能、集合知定理、アンサンブル学習

Class Discrimination Method based on Theory of Collective Intelligence and its Application to Medical Field.

NORIAKI YOSHIKAI^{†1} SHUN-ICHI KURINO^{†1}
NORIKAZU HASHIGUCHI^{†1} KANA KAZAWA^{†2} MICHIKO MORIYAMA^{†2}

Abstract: After considering a class discrimination system using AI with noise data environment, the novel method using the theory of Collective Intelligence and the time series processing is proposed for improving the predicted accuracy rate when the training data itself contains some noises. In order to evaluate the performance, we developed AI that predicts the risk of aggravation of patients in health guidance. The results show that the AI was able to identify high-risk individuals who were visually overlooked, and that the proposed method can improve the accuracy of predicting the correct answer.

Keywords: Artificial Intelligence, Theory of Collective Intelligence, Ensemble Model, Time Series Processing

1. 前書き

深層学習 (Deep Learning: DL) をコア技術とする機械学習システムの飛躍的な発展に伴い、様々な分野への人工知能 (AI: Artificial Intelligence) の活用が始まっている。これまでに、DL との相性の良さから、画像データを対象にした画像認識を行う機械学習システムを活用する多数の応用が実用化されている[1]。それらは、正しい回答が既にわかっており、そのビッグデータを AI に学習させる事により、雑音の多い被データに対して判定精度が高い正解率を得られる事の特徴とする点が共通である。

一方で、現実の問題として、設備・税などの社会で共有している資源の配分、あるいは社会保障・医療・教育などの制度設計のような社会的意思決定を、AI を用いて効果的・効率的に実施しようという動きが出つつある[2]。意思決定を実施する為には、被対象データを、何らかの尺度で分類し、さらに、閾値を設けて意思決定 (例; Yes/No, 病気

/健康) を行う必要があるが、その分野の専門家ですら、絶対正しいとは言えない現象を含む為、従来の教師データの作成法を、そのまま適用する事はできない。さらに、画像データ以外のデータに対しても、そのデータのクラス分けを実施可能な AI を実現する事が必須となる。

テキストデータに対しては、2018年に発表された BERT により言語モデルの自動獲得が、比較的簡単に可能となり、目覚ましい発展を遂げようとしている[3]。

特に医療分野においては、電子カルテシステムに代表されるように、画像情報と共に、膨大な医療テキストデータが既に蓄積されており、これらを利用し、自然言語処理により新たな医学的な知見の発見、診断・診療支援、及び業務支援へ活用する需要は高く、国内・外で検討が進められている[4]。しかし、技術的な問題と共に、テキストデータの書き方、および用語の不統一のようなデータの精度の問題も明らかとなり、実現までにはまだ時間が必要である。

本稿では、クラス分けを実施する AI について一般的に

¹ 日本大学
Nihon University
² 広島大学
Hiroshima University

考察し、医療費の中で大きな比率を占める生活習慣病に対する医療テキストデータを対象に、アンサンブル学習を用いて教師データの精度を向上させる事により、重症化レベル判定精度の大幅な向上が実現でき、将来の医療費の削減可能性を実験的に示したので報告する。

2. クラス分けの特徴と課題

個人や、物、あるいは情報等のクラス分けの対象に関する様々なデータから、その個人や物、あるいは情報に関するランク、等級、危険性、重要度といったある種の事象のクラス分けを行うことが少なくない。

そのクラス分けは、一般に多数のデータに基づいて判定されるものであり、個人や、物、あるいは情報等の対象の事象に関連するデータの種類の専門家の判断によることで、そのクラス分けの信用度も増大する。

例えば、医療関係では対象となる各個人の医療データに基づき、各個人が事象として有する病気の重さや重篤度が、その専門家である医師等の医療従事者によって総合的に判断されることで、その正確度が向上する。

加えて、個人や、物、あるいは情報等の対象の事象は、時間とともに変化するものであり、その事象の変化に応じて、分類されるクラスも変化する。

したがって、現在や過去のデータに基づいた将来的な事象の変化をクラスの変化として推測することができれば、事前のリスク回避や将来的な準備など、より良い社会的な活動に貢献することができる。

一方、近年のAI技術の向上により、様々なデータに基づいて予測判定の精度の向上が図られている。

しかし、そのAIは、結果出力のための推論と結果との因果関係が明確でないために、因果関係の明確化を求められる分野にとって、その利用の妨げとなっている。

したがって、少なくとも、専門家が見れば説明が出来る結果が示されるAI処理が求められている。

さらに、後述するように、医療従事者のような専門家の判断にもエラーやその時々環境に応じて微妙に変化する揺らぎが存在する。しかも、あるクラス分けを行う場合には、その判断が専門家間で微妙に異なり、それが各専門家の個性ともなって、個性に基づいた回答結果であるクラス分けにも違いを生じる。その個性は尊重すべき点もあるが、その個性によって特異なクラス分けがされると却って予測の精度が低下する事となる。

一方、AIの判断においても、通常のAIに使用される教師データ量と比較すると、専門家が判断するデータは少数であり、AI自体の回答の精度もそれ故、幅を有することとなる。

したがって、専門家の判断の個性を維持しつつ、専門家の回答の正確性と、AIによる判断の正確性との相互を調

整する新たな手法が求められていた。

3. 集合知定理を用いるクラス判別法

3.1 基本処理フロー

図1に、提案するクラス判別システムの基本構成案を示す。基本的には、単年度毎のデータに対して、対象物のクラスを決定し、その結果を次年度のパラメータの一部として、さらに学習を続ける構成とする。さらに、単年度のクラス決定においては、2章に述べたように、多くの不確定要素があるため、単純な処理では、必ずしも精度の高い結果を期待できない。そこで、ここでは、複数の弱学習器（ここでは、ニューラルネットワーク）を組み合わせることで精度を向上させるアンサンブル学習を用いる。

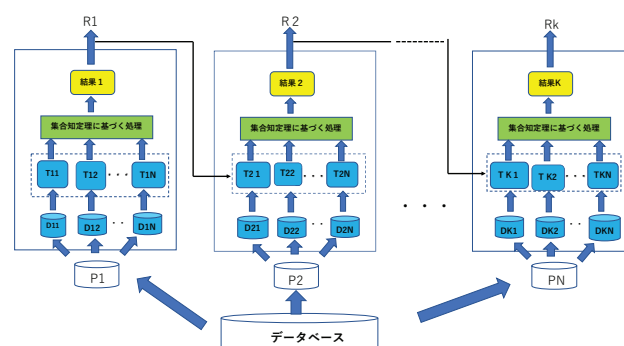


図1 クラス判別手法の基本構成案

Figure 1 Basic Structure of Class Discrimination System.

アンサンブル学習とは、一般に、与えられたデータから複数の学習器を生成し、それを適切に組み合わせることで予測することにより、一台の学習器で予測した場合よりも予測精度を向上させようとするものである。

その弱学習器の構成には、単純に平均値に対する多数決論理で判定を決めるバギング、誤検出率に応じて重みをつけるバディン、複数の決定木から平均多数決をとるランダムフォレストが知られている。

ここでは、構造の単純さから、バギングを対象にする。但し、判定を決定するために、単純に平均値の多数決論理だけではなく、集合知の定理[5]（付録参照）が示すように、集団の精度を向上させるために、分散の大きさも同時に評価して判定することにする。

図1に示すように、全体のデータベースから、単年度毎のデータが取り出される。K年後の予測まで対象にすると仮定し、それらに対して、ランダムサンプリングしてサブデータ群 $(D_{ij}; i=1\sim K, j=1\sim N)$ を構成する。それらのデータに対して、同じ学習アルゴリズムを有する学習器で予測値を出す。それらの出力データに対して、前述の集合知の定理による処理を行い、単年度の最終的な結果を出す。対象年

度が、2年目以降の場合は、前年度の結果を、その年度の入力データとして、学習器の入力データとして利用する。

4. 保健指導に用いる重症化リスク予測 AI の試作と評価

4.1 重症化予測 AI の目的と活用イメージ

クラス判別法の医療への応用として、生活習慣病に対する健康関連情報を元に、翌年の重症化リスクを予測する AI を試作・評価した[6]。

評価の実施は、広島大学大学院医系科学研究科が中心となり、専門医、看護師の意見、診療ガイドラインをもとに保健指導により重症化予防が可能な疾患群とその重症度(予測される結果)を定義し、医療保険者が有する医療・介護レセプト、健診データから重症化に影響を及ぼす予測因子と特徴量を決定した。次に、広島県内自治体の国民健康保険、後期高齢者医療者制度の被保険者のデータ 10,585 件に対し、実際に保健指導をおこなっているエキスパート(医師・保健師・看護師)が目視で重症化リスク・スコアリングが行われた。

重症化リスク予測 AI による医療費削減イメージを述べる。

健康診断を受診した人の健康状態を、クラスとして 4 つの層に分け、各層の人数構成と、医療費との関係で示す事ができる。

下の層から、クラス 1、クラス 2、クラス 3、そしてクラス 4 とクラス番号付けを行い、クラス 1 は、今後も健康増進を期待するもの、クラス 2 は軽度のもの、クラス 3 は中度のもの、そしてクラス 4 を、重度のものとする。

この 4 つの層は、人への病気の危険度を図る上でも利用可能あり、重症化のレベルとしても把握でき、最上位は重症化レベルが最大であり、救急入院・集中治療、さらに死亡までを含むクラスである。また、最下層は、重症化レベルとして、健康上のリスクは、ほぼ無いと判断できる。したがって、このクラス分けはリスク分けでもあり、リスク管理として利用することができる。

このクラス分けを行った場合に、全体の人数に占める各層の人数の割合は、下から、50%、30%、15%、5% といったピラミッド状の分布を形成する。一方、全体の医療費の占める各層の医療費は、最上位層のクラス 4 で、40% を占め、その下に行くにしたがって、35%、20%、5% の逆ピラミッドの分布を形成する事が知られている[7]。

したがって、人数構成が 5% や 15% となるクラス 4 やクラス 3 の発生を予想し、その予防ができれば、重度となる可能性のある人へのより適切なケアを可能とし、全体の

医療費の削減も可能となる。

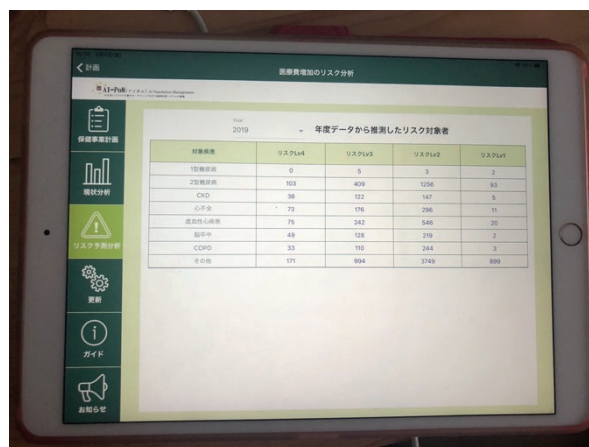


図 2 I-Pad での試作システムの画面例

Figure 2 Screen of Developed System on I-Pad.



図 3 重症化リスク予測 AI の出力例

Figure3 Screenshot of Output of Class Discrimination AI

4.2 設計環境

設計はプラットフォームとして、Google Cloud Platform を使い、ディープラーニングのライブラリーとして TensorFlow を使用した。

ニューラル・ネットワークとしては DNN とし、入力チャンネルは医療データの検討から 5 3、出力チャンネルはリスクレベルから 4 とした。中間層については、バイズ最適化を用いた Hyperparameter Tuning[8]を使って、階層が 1 から 5 までの範囲で自動的に決定する方法を採用した。

また重み付けに関する初期値設定は、Truncated normal distribution[9]を元に、-1 から 1 までの範囲でランダムに重み付けを実施した。

4.3 試作システムの評価と考察

試作システムは、単年度データを対象にする構成とし、またアンサンブル学習の効果評価を目的に、実験を行った。

(1) 実験1

各種医療機関のデータサーバから、「診療情報明細書」、「介護給付明細書」、「特定健診結果」及び「被保険者マスタ」のデータを収集して、250人の医療データを取得し、2014年の医療データを医療に従事する医師やその他の専門家に目視してもらい、翌年度の2015年時点での各人の重症化のリスクレベルを決定した。

次に、そのリスク番号を1～4の番号に振り分け、リスク番号1～4を伴った各人の医療データを学習用データとして、単一構成のAIに入力した。さらに、その学習したAIに対して、250人の中から2014年の50人の医療データを入力して、2015年の重症化リスクとして予想した。

その結果、50人のうち、専門家による2015年度の判定結果とAIによる2015年度の予測との間で、8件(人)が相違した。

さらに、50人の2015年の医療データからその年の重症度を、ベテラン専門家によって判断してもらい、その実際の結果を専門家とAIの両予測データと比較したところ、相違する8件のうち、AIが予想した2件の重症度と実際の結果とが一致した。

このことから、小規模なデータではあるが、AIの有効性の可能性が確認された。

(2) 実験2

次に、2014年度の因子データを有する3299件のデータ(教師データ:2699件、検証データ:600件)を作成した。

そのデータを、25人の専門家に表示し、2014年度のデータに基づいた専門家による臨床推論を行った。

臨床推論により得られた2015年度のリスク予測は、各患者IDごとの学習用データとして、実験1で用いたAIに入力して学習させた。

2015年年度のリスク予想に使用されるデータとして、収集・抽出した2014年度実データである6707件の被クラス判定データをAIにより判定した。

AIが予測した結果の正確性を確認するために、専門家による2015年度の6707件の実データに基づいた臨床推論からリスク分けを行い、その結果(2015年実データに基づく教師クラス実判定)とAI予測(2015年クラス予測)とを比較した。

結果は、クラスの一致する割合が58.90%であった。

以上から、ある年度のデータから翌年度の重症度の予測の可能性があることが理解できたが、より高い精度での予測も必要である事がわかった。

(3) 実験3

実験2に用いた教師データの内容を調査した結果、専門家によっては、重症度予測にばらつきがあり、他とかけ離れた重症度予測となっており、それが教師データの揺らぎとして、AI予測の精度にも影響を与えていると判断した。

そこで、図4に示すように、教師データに対するアンサンブル学習を行い、異常と判別された重症度予測を排除した教師データを再構成し、それに基づく再度のAI予測から最終的な重症度予測を行うこととした。(図4参照)

実験2で説明した2699件の教師データと、収集・抽出した6707件の被クラス判定データとを利用し、10個のAIの各判定データに対して集合知定理に基づき、互いの判定データと一致しない数が7以上の場合であり、さらに、AI同士間の判定の標準偏差が0.5未満である判定データを削除することとした。

表1に、削除した教師データの例を示す。

結果として、2699件の全体の教師データから異常値を有するとして判断された教師データを削除して、2228件の新たな教師データを作成した。

その新たな教師データにより再学習させたAIを使い、2014年度の6707件の被クラス判定データから、新たな判定結果を得た。

AIから出力された6707件の2015年度クラス予測データと、2015年度の実データに基づく専門家によるクラス実判定とを比較した結果、クラスの一致率は、83%以上の著しい向上が見られた。

以上のように、専門家による判定データ自体にバラツキがあり、その影響から、クラス判定は大きな影響を受けることが明らかになった。また、その改善のために、アンサンブル学習法が有効であることを、実験により示した。

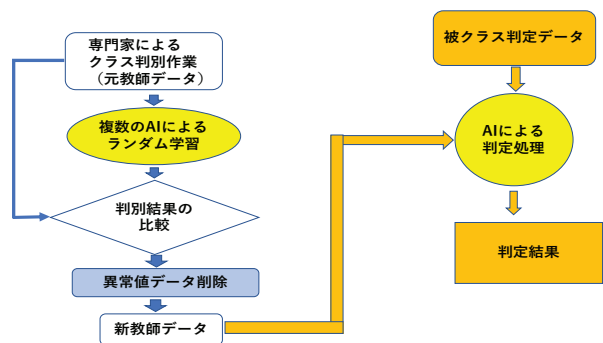


図4 アンサンブル学習による医療従事者の判断の揺らぎを評価する実験フロー

表1 異常値として削除された教師データの例

Table 1 Example of training data eliminated as outliers

患者コード	エキスパート	AI評価2	AI評価3	AI評価4	AI評価5	AI評価6	AI評価7	AI評価8	AI評価9	AI評価10	AI標準偏差	相違
1000000000253	2	1	1	1	1	1	1	1	1	1	0	10
1000000000333	1	2	2	2	2	2	2	2	2	2	0	10
1000000000483	1	2	1	2	2	2	2	2	2	2	0.3	9
1000000000506	2	1	1	1	1	1	1	1	1	1	0	10
1000000000642	2	3	3	2	3	3	3	2	3	3	0.4	8
1000000000754	2	1	1	1	1	1	1	1	1	1	0	10
1000000000859	3	2	2	2	2	2	2	2	2	2	0	10
1000000000974	2	1	1	1	1	1	1	1	1	1	0	10
1000000001017	2	1	2	1	1	2	1	1	1	1	0.4	8
1000000001074	2	1	1	1	1	1	1	1	1	1	0	10
1000000001509	3	2	3	2	3	2	2	2	2	3	0.45825757	7
1000000001556	2	1	1	1	1	1	1	1	1	1	0	10
1000000001736	2	2	1	1	1	1	2	1	1	1	0.4	8
1000000001776	2	1	1	1	1	1	1	1	2	1	0.3	9
1000000001894	1	2	2	2	1	2	2	2	2	1	0.4	8
1000000001939	2	1	1	1	1	1	1	1	1	1	0	10

5. まとめ

医療データのような、データ自体にノイズが含まれる環境下におけるクラス判別システムに関する考察を行い、教師データにノイズが含まれる場合の改善法として集合知定理を応用するアンサンブル学習法が有効であることを実験的に示した。

しかし、実用化のためには、より高い精度が必要であり、またデータ全体に対するアンサンブル学習における性能評価も必要である。さらに、図1に示すような時系列処理も含めた全体的な性能評価を今後実施する予定である。

謝辞 本研究は、日本医療研究機構「AIを活用した保健指導システム研究推進事業」における研究開発の委託、助成を受け実施した。

なお本研究は、広島大学医学部が主管となり、株式会社 OKEIOS、株式会社データホライゾン、株式会社 DPP ヘルスパートナーズと共に実施した。

参考文献

- [1] 松尾豊 人工知能 ディープラーニング 編 ニュートンプレス (2019) .
- [2] 馬場雪乃、櫻井祐子 特集「人工知能と社会的意思決定」にあたって 人工知能学会誌 35巻4号 (2020年7月)
- [3] Devlin, A., Chang, D. J., Lee, M. and Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proc. 2019 Conf. of the North American Chapter of the Association for Computational Languages (NAACL) : Human Language Technologies (HLT), vol.1, pp.4171-4186 (2019)
- [4] 今井健 医用自然言語処理分野における人工知能技術応用、人工知能学会誌 35巻4号 pp.474-479 (2020年7月)
- [5] 西垣通 集合知とは何か - ネット時代の「知」の行方 - 中公新書 (2013)
- [6] 加澤佳奈、渡邊拓人、玉城雄也、森山美知子、栗野俊一、吉開範章、木原康樹、保健指導による介入効果が高い慢性疾患患者の抽出を支援する重症化リスク予測 AI の開発、第 39 回医療情報学連合大会 vol. 3-G-1, pp.462-464 (2019)

- [7] Steenkamer, B.M., Drewes H. W., Heijink, R., Baan, C. A., Strujis, J. N., Defining Population Health Management: A Scoping Review of the literature, Popul Health Manag 2017, vol.20, no.1 pp.74-85 (2017)
- [8] Nabi, J. Hyper-parameter Tuning Techniques in Deep Learning, <https://towardsdatascience.com/hyper-parameter-tuning-techniques-in-deep-learning-4dad592c63c8> (2019)
- [9] Su, Q., Liao, X., Chen, C., Carion, L., Nonlinear Statistical Learning with Truncated Gaussian Graphical Models, Proc. of the 33rd International Conference on Machine Learning, vol.48, (2016)

付録

付録 A.1 集合知定理

次式で表すことができる。

$$\text{集団誤差} = \text{平均個人誤差} - \text{分散値}$$

集団誤差は、ある集団のメンバーで推定した値の平均と正解との差である。平均個人誤差は、各メンバーの誤差の平均値、そして分散値は、各メンバーの推定値のばらつきである。

メンバー*i* の推定値を $X(i)$ ($i=1,2,\dots,N$)、集団推定値を A 、真値を R とした時

$$\text{集団誤差} = (A-R)^2$$

$$A = \{X(1)+X(2)+\dots+X(N)\} / N$$

$$\text{平均個人誤差} = \{ (X(1)-R)^2+(X(2)-R)^2+\dots+(X(N)-R)^2 \} / N$$

$$\text{分散値} = \{ (X(1)-A)^2+(X(2)-A)^2+\dots+(X(N)-A)^2 \} / N$$

で表すことができる。

この定理が示すところは、集団における個人の推測誤差 (右辺第1項) は、多様性 (右辺第2項) によって相殺され、集団として正解に近い推測ができる、というものである。