様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月　 １ 日現在

機関番号：２４４０３
研究種目：基盤研究(C)（一般）
研究期間：2013～2015
課題番号：２５３３０２９２

研究課題名（和文）多目的遺伝的機械学習手法の並列分散実装

研究課題名（英文）Parallel Distributed Implementation of Multiobjective Genetics-based Machine Learning Algorithms

研究代表者

能島　裕介（Nojima, Yusuke）

大阪府立大学・工学（系）研究科（研究院）・准教授

研究者番号：１０３８２２３５

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：数値データからの知識獲得において考慮すべきことは，知識の精度と分かりやすさである．しかし精度と分かりやすさの間にはトレードオフの関係があり，両者が最適になる知識は一般に存在しない．多目的遺伝的機械学習は，これら２つの目的を同時に最適化することで，精度と複雑性の異なる複数の知識を一度に獲得できる手法である．また，大規模データからの知識獲得が必要とされており，本研究では，多目的遺伝的機械学習の並列分散実装方法を検討する．また，並列分散実装の適用可能性を確かめるために，様々な条件部集合に基づく遺伝的機械学習への適用と，複数データセンタからの知識獲得への適用を検討する．

研究成果の概要（英文）：There are two goals for data mining from numerical data. One is to maximize the accuracy of obtained knowledge. The other is to maximize its interpretability. However, there is a tradeoff between the accuracy and the interpretability. To address this issue, we proposed multiobjective genetics-based machine learning (MoGBML) which can simultaneously handle these two objectives and provide a number of classifiers with different accuracy and interpretability as knowledge. To further extend MoGBML to large data sets, we apply our parallel distributed implementation to MoGBML in this study. In addition, we examine the effects of a various kind of antecedent sets on the performance of our parallel distributed GBML. We also examine the applicability of our parallel distributed implementation to data mining from multiple data centers.

研究分野：計算知能

キーワード：知識獲得　多目的最適化　並列分散実装　進化計算

１．研究開始当初の背景

近年，様々な数値データからの知識獲得が盛んに行われている．多くの知識獲得手法が提案されているが，獲得される知識の分かりやすさを考慮したものは多くない．すなわち，知識の精度（汎化性能）のみが議論されることが多い．数値データからの知識獲得において，得られる知識の分かりやすさは情報の利活用から考えると非常に重要である．例えば，ニューラルネットワークなどのモデルを用いた場合，入出力の関係性を理解することは困難である．一方，If-then ルール形式の知識表現を用いた場合，「もし○○ならば，△△である」という理解可能なルールの集合を獲得することが可能になる．ただし，条件部分が複雑な場合は，If-then ルール単体での可読性が低下する．また，膨大なルール集合として知識が獲得された場合も，直感的にルール集合全体を把握することは困難である．

また，近年のビッグデータ解析ブームによって，大規模データからの知識獲得への期待が非常に高まっている．大規模なデータを扱うことで，より汎化性の高い有用な知識が得られる可能性があるが，処理すべき情報量も多くなることから，計算時間も大幅に増加するという問題も生じる．

以上のことから，大規模データから高精度かつ分かりやすい知識を短時間で獲得することが求められている．

２．研究の目的

遺伝的機械学習は，If-then 形式のルールで構成される識別器を数値データから獲得できる知識獲得手法である．知識獲得に求められる精度と分かりやすさを容易に組み込むことが可能であるという特徴もある．遺伝的機械学習は，確率的多点探索手法である進化計算に基いているため，知識獲得において，学習用データを用いた候補解の繰り返し評価が必要である．そのため，データが大規模化した場合，膨大な計算時間が必要となるという問題がある．

これまで研究代表者は，If-then ルールを言語的に解釈可能にするために，条件部にファジィ集合を用いたファジィ遺伝的機械学習を提案してきた．また，その高速化に，個体群と学習用データを同時に分割し，複数の CPU コアに割り当てて進化を行う並列分散実装を提案してきた．本研究では，これまでの並列分散型進化型知識獲得手法を拡張する．主に，進化型多目的最適化の利用とその並列分散実装の提案と改良を行う．探索性能を改善するために，目的関数の修正や，複数の進化型多目的最適化アルゴリズムの比較を行う．また，均等分割のファジィ集合以外の知識表現を用いた並列分散型遺伝的機械学習の提案と性能比較を行う．さらに，並列分散実装の拡張として，個人情報保護を考慮した複数のデータセンタからの知識獲得の方法を検討する．

３．研究の方法

(1) 並列分散型多目的遺伝的機械学習

知識獲得では，高精度かつ分かりやすい知識が求められるが，それらはトレードオフの関係にあり，高精度な知識は複雑になり，簡単な知識は精度が低い．また，利用者が持つ背景知識によって，必要とされる知識の分かりやすさは異なる．そこで，進化型多目的最適化アルゴリズムを用いて，精度と複雑性の異なる複数の知識を１回の手法の実行により得られる多目的ファジィ遺伝的機械学習手法を提案してきた．この手法を並列分散化することで，精度と複雑性の異なる知識を短時間で獲得する方法と提案する（図 1）．

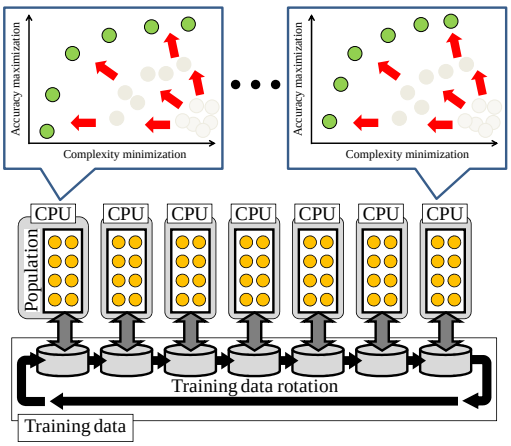

図１　並列分散多目的遺伝的機械学習

また，精度の最大化（誤識別の最小化）にバイアスを掛ける探索方法も検討する．具体的には，誤識別率の最小化 f_1 と複雑性の最小化 f_2 の２つの目的関数を回転行列により修正することで，誤識別率最小化の強調 f_1' と，複雑性最小化 f_2' の緩和を行う（図 2）．

$$\begin{pmatrix} f_1' \\ f_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

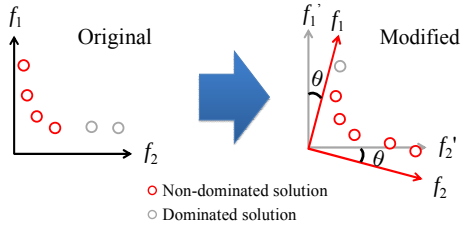

図２　目的関数の修正

これまで多目的ファジィ遺伝的機械学習には，解の優越関係に基づく NSGA-II を用いてきたが，並列分散実装において，別の進化型多目的最適化アルゴリズムが適している可能性がある．そこで，アルゴリズム自体の分散化が容易であると思われる MOEA/D を用いて有効性を検証する．MOEA/D は多目的最適化問題を複数の単一目的最適化に分割して探索を行う手法である．例えば，図 3 の

ように，探索方向を複数のベクトル群で分割することが容易に行うことが可能である．

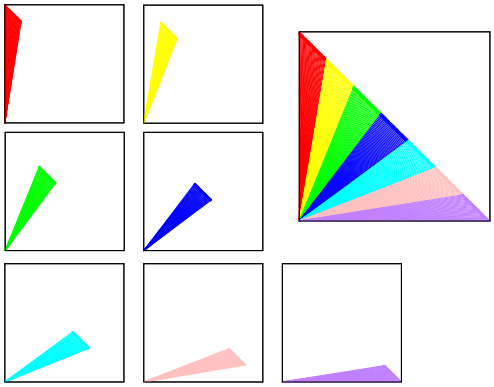

図 3　MOEA/D における探索方向の分割

(2) 区間集合および非均等ファジィ集合を用いた並列分散進化型知識獲得

代表的な区間集合を用いた遺伝的機械学習 GAssist に，並列分散実装を適用し，ベースとなる遺伝的機械学習の違いによる影響を調査する．

また，知識の汎化性を改善するために，区間導出型ファジィ集合を利用した方法と，均等ファジィ分割の位置を後処理として修正する方法を提案し有効性を検証する．

(3) 個人情報保護を考慮した実装

並列分散ファジィ遺伝的機械学習の応用として，複数のデータセンタからの知識獲得手法の提案を行う．各データセンタはデータの内容は公開せず，知識の評価のみ行い，その評価値のみから進化を行うというモデルである．

4．研究成果

(1) 並列分散型多目的遺伝的機械学習

多目的ファジィ遺伝的機械学習に，これまで提案した並列分散実装を直接適用した場合，計算時間は大幅に改善するものの，高精度かつ複雑な知識が獲得できないことが明らかになった．これに対して，目的関数を修正し，誤識別率の最小化にバイアスを掛けることで，これまで取れなかった高精度かつ複雑な知識が獲得できることを明らかにした．ただ，複雑な知識を個体群に含む場合，計算時間に大きな影響を与えることも明らかになった．

図 4 から図 6 に数値実験結果の一例を示す．数値実験では，7CPU コア用いて並列分散実装を行った．すなわち，個体群と学習用データはそれぞれ 7 分割し，個別の CPU コアに割り当てた．図 4 は，Segment データの評価用データに対する誤識別率である．Non-parallel は非並列非分散の通常の多目的ファジィ遺伝的機械学習の結果である．0 ° は並列分散型多目的ファジィ遺伝的機械学習の結果である．15 ° は目的関数を 15 ° 回転させた並列分散型多目的ファジィ遺伝的機械学習の結果である．図 4 から，並列分散実装により，複雑な知識が獲得できていないことが分かる．それに対して，目的関数を回転させることで，並列分散実装でも，非並列非分散と同様の精度と複雑性を持つ複数の知識が獲得されていることが分かる．さらに，もっと複雑な知識（20 ルール以上）の獲得も確認できるが，ほとんど誤識別率が改善していないことも分かる．

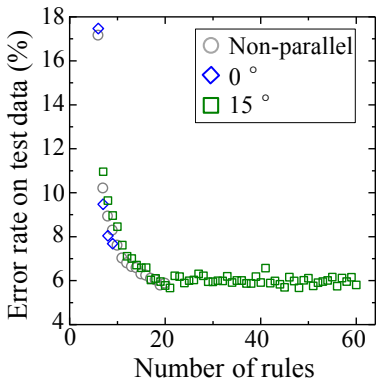

図 4　Segment データの評価用データに対する誤識別率

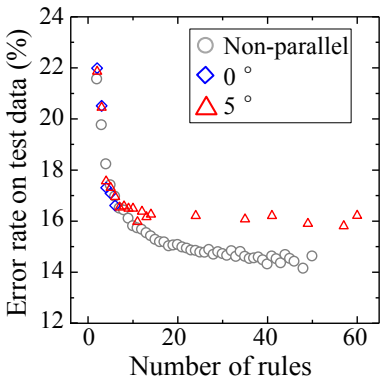

図 5　Phoneme データの評価用データに対する誤識別率

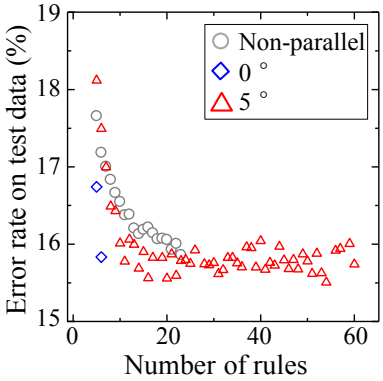

図 6　Satimage データの評価用データに対する誤識別率

図 5 は，Phoneme データの評価用データに対する誤識別率である．非並列非分散と同様

の精度と複雑性を持つ複数の個体はルール数 10 位上でほとんど獲得できなかった．

図 6 は，Satimage データの評価用データに対する誤識別率である．他のデータと異なり，単純に並列分散実装を行うだけで，非並列非分散で獲得された知識をすべて優越する知識を獲得できるたことが分かる．また，目的関数を修正することで，さらに，誤識別率が低い知識を獲得できた．

このように，多目的ファジィ遺伝的機械学習の並列分散実装の性能は，データに大きく依存するが，大きく改善できる可能性も高いことも明らかになった．

図 7 に，MOEA/D と NSGA-II の比較実験結果を示す．実験には Penbased データを用いた．MOEA/D は図 3 のようにベクトルを分割し，それぞれ部分個体群として並列分散実装を行った．また異なるスカラー化関数を用いて影響を調査した．NSGA-II の方が 2 つの目的に対して優れた多数の知識が獲得できていることが確認できる．また，MOEA/D は，スカラー化関数の違いにより，探索の方向性が異なることが明らかになった．MOEA/D には，調整可能な種々のパラメータが存在し，またベクトルの構成方法に関しても検討の余地が残っており，これらは今後の課題である．

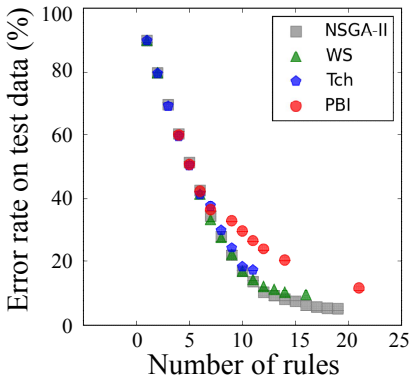

図 7　異なるスカラー化関数を持つ MOEA/D と NSGA-II の比較

(2) 区間集合および非均等ファジィ集合を用いた並列分散進化型知識獲得

まず，区間集合を用いた代表的な遺伝的機械学習手法である GAssist の並列分散実装を行った．学習用データと個体群の分割数を変え，ファジィ遺伝的機械学習の並列分散実装との比較を行った．図 8 に Penbased データの評価用データに対する正答率の比較結果を示す．まず GAssist に関して，7 分割において汎化性能の改善がみられた．また，分割数を増やしても，ファジィ遺伝的機械学習もともに大幅な改悪が見られなかった．並列分散実装による高速化は，分割数の 2 乗倍になることから，汎化性能を悪化させずに，計算時間の大幅な短縮が可能であることが明らかになった．

次に，並列分散ファジィ遺伝的機械学習で得られたファジィ If-then ルールに基づく知識において，ファジィ集合の位置を後処理として進化計算により最適化した結果を表 1 に示す．ファジィ集合の位置を最適化することで，学習用データおよび評価用データに対して正答率が改善できることが確認できた．

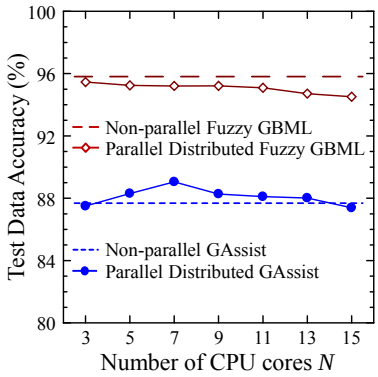

図 8　並列分散 GAssist と並列分散ファジィ遺伝的機械学習の比較

表 1　ファジィ集合の位置の最適化前後の正答率の比較

Data name	Parallel Distributed GBML		Genetic Lateral Tuning	
	Training	Test	Training	Test
Penbased	97.56	96.45	**97.70**	**96.77**
Phoneme	85.33	83.93	**85.80**	**84.21**
Satimage	85.36	84.38	**85.54**	**84.54**
Segment	95.63	94.00	**96.16**	**94.26**

(3) 個人情報保護を考慮した実装

図 9 に示すような状況を想定し，数値実験を行った．Dataset D_1 のみフルアクセス可能であり，他のデータ集合は，知識の評価のみ行える状況である．Dataset D_1 の情報を元に，ファジィ If-then ルール集合を解候補として作成し，その解の評価のみ他のデータで行うことで，Dataset D_1 だけで知識獲得を行う場合よりも，汎化性能の高い知識が得られることが明らかになった．同等の情報を保持する複数のデータセンタ（例えば，病院など）に，知識評価用のコンピュータを設置し，遺伝的機械学習を実行することで，個人情報を保護した状態で汎化性能の高い知識が得られる可能性を示した．

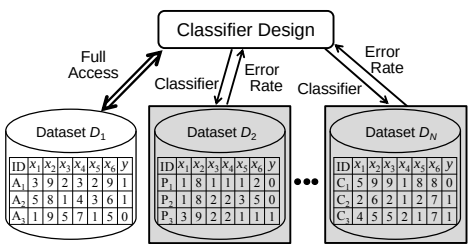

図 9　個人情報保護を考慮した複数データ集合からの進化型知識獲得

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

① **能島裕介，石渕久生，遺伝的機械学習を用いた大規模データからの知識獲得，システム／制御／情報，査読なし，第 57 巻 10 号，**pp. 421-426 (2013 **年 10 月**)

〔学会発表〕（計 12 件）

① Y. Nojima and H. Ishibuchi, “Effects of parallel distributed implementation on the search performance of Pittsburgh-style genetics-based machine learning algorithms,” *2016 IEEE Congress on Evolutionary Computation*, **査読あり**, 8 pages, Vancouver (Canada), July 24-29, 2016. **(発表確定)**

② **武村周治**, **能島裕介**, **石渕久生**, **多目的ファジィ遺伝的機械学習におけるアルゴリズムの違いによる探索性能への影響**, **第 9 回進化計算シンポジウム**, **グリーンホテル三ヶ根（愛知県・西尾市），**2015 **年** 12 **月** 19 **日～**20 **日**

③ **高橋佑治，能島裕介，石渕久生，並列分散型多目的ファジィ遺伝的機械学習における目的関数の回転，第** 31 **回ファジィシステムシンポジウム，電気通信大学（東京都・調布市），**2015 **年** 9 **月** 2 **日～**4 **日**

④ H. Ishibuchi and Y. Nojima, “Handling a training dataset as a black-box model for privacy preserving in fuzzy GBML algorithms,” *2015 IEEE International Conference on Fuzzy Systems*, **査読あり**, 8 pages, Istanbul (Turkey), August 2-5, 2015.

⑤ Y. Takahashi, Y. Nojima, and H. Ishibuchi, “Rotation effects of objective functions in parallel distributed multiobjective fuzzy genetics-based machine learning,” *10th Asian Control Conference*, **査読あり**, 6 pages, Kota Kinabalu (Malaysia), May 31-June 3, 2015.

⑥ Y. Nojima, Y. Takahashi, and H. Ishibuchi, “Application of parallel distributed implementation to multiobjective fuzzy genetics-based machine learning,” *7th Asian Conference on Intelligent Information and Database Systems*, **査読あり**, Part I, pp. 462-471, Bali (Indonesia), March 23-25, 2015. Best Regular Paper Award

⑦ Y. Nojima, Y. Takahashi, and H. Ishibuchi, “Genetic lateral tuning of membership functions as post-processing for hybrid fuzzy genetics-based machine learning,” *7th International Conference on Soft Computing and Intelligent Systems and 15th International Symposium on Advanced Intelligent Systems*, **査読あり**, pp. 667-672, **北九州国際会議場（福岡県・北九州市），**December 3-6, 2014.

⑧ Y. Takahashi, Y. Nojima, and H. Ishibuchi, “Hybrid fuzzy genetics-based machine learning with entropy-based inhomogeneous interval discretization,” *2014 IEEE International Conference on Fuzzy Systems*, **査読あり**, pp. 1512-1517, Beijing (China), July 6-11, 2014.

⑨ Y. Nojima, P. Ivarsson, and H. Ishibuchi, “Application of parallel distributed implementation to GAssist and its sensitivity analysis on the number of sub-populations and training data subsets,” *14th International Symposium on Advanced Intelligent Systems*, **査読あり**, 10 pages, Daejeon (Korea), November 13-16, 2013. Best Session Paper Award

⑩ H. Ishibuchi, M. Yamane, and Y. Nojima, “Rule weight update in parallel distributed fuzzy genetics-based machine learning with data rotation,” *2013 IEEE International Conference on Fuzzy Systems*, **査読あり**, 8 pages, Hyderabad (India), July 10-15, 2013.

⑪ Y. Nojima, “Parallel distributed multiobjective evolutionary algorithms,” *7th International Conference on Intelligent System Application to Power Systems*, **査読なし**, **明治大学（東京都・千代田区），** July 1-4, 2013.

〔図書〕（計 0 件）

〔産業財産権〕
○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕
ホームページ
http://www.cs.osakafu-u.ac.jp/~nojima

国際会議でのチュートリアル講演
Tutorial on “Genetic fuzzy systems and its application to data mining” at 2015 Conference on Technologies and Applications Artificial Intelligence, Tainan, Taiwan, Nov. 20-22, 2015.
http://taai2015.nutn.edu.tw/yusukenojimachunhaochen-tutorialspeaker

６．研究組織
(1)研究代表者
能島　裕介（Nojima Yusuke）
大阪府立大学・工学研究科・准教授
研究者番号：10382235

(2)研究分担者
なし

(3)連携研究者
なし