

O11 GA を用いたファジィルール集合の帰納的学習

本多功一 元木達也

新潟大学大学院自然科学研究科

1 はじめに

与えられた事例に関連する特徴値に基づいて診断／分類する問題に対して

IF 温度が低いか中くらい \wedge 天気が晴れか曇り
 $\wedge \dots$ THEN 開催

IF 温度が低いか極低か高いか極高 $\wedge \dots$
 THEN 中止

といった形の IF-THEN ルールを集めてファジィルール集合を構成する研究が行われている。ファジィルールの条件部には言語的曖昧性が含まれており、ルールを実際に適用するためには実際に得られた特徴値 (e.g. 温度が 10℃) と言語的値 (e.g. 温度が低い) との対応付けを行う必要があるが、ファジィ研究においては、これは次の図に例示されるように、実際に得られた特徴値が言語的値をもつ度合いを関数 (メンバーシップ関数という) で表すことによって行う。

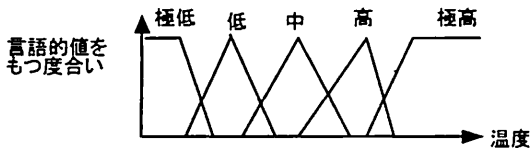


図1 メンバーシップ関数

ファジィルール自体は我々人間の知識に近いものであるため、これまで色々な問題に対して数多くのファジィルール集合が専門家との協力で作り出されており、Wang 達[1]は、同じ問題に対して作られた複数のファジィルール集合 (とメンバーシップ関数の族) を GA を用いて統合してより良いものを作り出す手法を提案している。彼らの手法はファジィルール集合とメンバーシップ関数族の組を一つの個体と考えてピッツバークアプローチ風に GA 操作を行うものであり、それなりの成果も上がっているが、一方では、結論部の違うルール間で遺伝子組み替えが起こったり、それまでに出来上がっていたファジィルール間の協調関係が壊されたり、といったこともあり、改善の余地もかなり残っているものと思われる。また、専門家が構築したファジィルール集合なしで GA を用いてどの程度のものを探索できるのかについても興味がある。

そこで、本研究では、ランダムな個体 (ファジィルール) 生成から始めて、ルールの結論部ごとにファジィルールの集団を GA で進化させ、それらの結果を基にファジィルール集合を構成する手法について検討する。実験には、AI レポジトリに事例データが置かれている肝炎診断問題[2]を用いる。

2 ファジィルール集合の帰納学習

ファジィルール集合の帰納的学習は、与えられたメンバーシップ関数の族を固定した上で、訓練事例に基づいて次のように行う。

学習手順1 (与えられたメンバーシップ関数族に対してファジィルール集合を学習)

1. ルールの結論部ごとに、ファジィルール (条件部が $\bigwedge_{F: \text{特徴}} \bigvee_{\alpha: \text{特徴値}} F = \alpha$ という標準形をしていると仮定) の集団を与えられた訓練事例に基づいて GA で進化させる。その際、適合度は次の3つの量の積とする。

・ルールの適用できる訓練事例に対する正解率

$$\text{必要性} = \sum_{\substack{c: \text{そのルールが} \\ \text{正解をだす事例}}} \frac{1}{\text{正解をだすルールの個数}}$$

$$\frac{1}{1 + \text{ルールが適用できるが間違ってしまう事例の数}}$$

2. 訓練事例をすべてカバーするだけのファジィルールを、ステップ1の最終集団から選ぶ。(適合度の高い順に必要なだけ選んでいく。)
3. ステップ2で選ばれたファジィルールの統合／一般化を C4.5[3]の "pessimistic error rate" を利用して行う。
4. ステップ3で得られたファジィルールに優先順位を付け、また、適用できるルールがなかった時のために、デフォルトの結論を決めておく。これらは、訓練事例に対する誤分類が少なくなるように行われるが、組み合わせ爆発もあるので、ここでは C4.5[3]に倣って、結論部が同じルールはひとまとめにし、それらの間の優先順位をつけることだけを考える。

実験 実験には肝炎診断問題を用いた。この問題は、18 の特徴から患者が肝炎であるかどうか判定する問題である。155 個の事例のうち 124 個を訓練事例、31 個をテスト事例とし、また、メンバーシップ関数をランダムに設定して学習手順 1 を 10 回実行した。得られた 10 個のファジィルール集合（とメンバーシップ関数族の組）のうちテスト事例に対して最良のもの、最悪のもの、および 10 個の平均的な振る舞いを表 1 に示す。Wang[1]の研究では、10 人の専門家がそれぞれ作ったファジィルール集合はテスト事例に対して正答率が 0.73~0.80 であり、それを統合したファジィルールの正答率は 0.91 であった。本研究の方法は Wang 達によって統合で得られたファジィルール集合の正答率には届かなかったが、専門家達よりも正答率の高いものを作ることができることがわかる。

表 1 学習手順 1 の結果

	訓練事例に 対する正答率	テスト事例に 対する正答率
最良のルール集合	0.983871	0.870968
10 個の平均	0.955564	0.816129
最悪のルール集合	0.959677	0.677419

3 メンバーシップ関数の学習

表 1 を見ると、最良と最悪のファジィルール集合では、テスト事例に対する正答率に大きな差があることがわかる。これは、ランダムに選んだメンバーシップ関数の妥当性が、得られるファジィルール集合の良さに大きく影響を与えているためと考えられる。そこで、以下のようにメンバーシップ関数の学習も考えた。

学習手順 2 (メンバーシップ関数も学習)

1. 複数のランダムに生成されたメンバーシップ関数族について、それぞれ優れたファジィルール集合を学習手順 1 を用いて求める。
2. ステップ 1 で求めたファジィルール集合の中で最も優れたファジィルール集合を使って、妥当なメンバーシップ関数の族を GA を用いて探索する。
3. ステップ 2 で得られたメンバーシップ関数の族に合ったファジィルール集合を再度、学習手順 1 を用いて構成する。

実験 肝炎診断問題に対して学習手順 2 を 10 回実行した。その際、ステップ 1 では 10 組のランダムなメンバーシップ関数族に対して学習手順 1 を適用し、ステップ 2 では（メンバーシップ関数族の）集団の大きさを 10000 とし GA を用いた。得られたファジィルール集合とメンバーシップ関数

族の組のうち、テスト事例に対して最良のもの、最悪のもの、および平均の振る舞いを表 2 に、そして、ステップ 2 終了時点での同様の結果を表 3 に示す。表 2 によると、10 回の実行のうち最良の場合は確かに学習手順 2 の方が学習手順 1（表 1）よりも良い結果になっていて、メンバーシップ関数の学習に効果が期待できることが分かる。しかし、10 回の実行の平均を見ると、メンバーシップ関数をランダムに設定して学習手順 1 を適用した結果（表 1）より、学習手順 2 の結果（表 2）の方が悪くなっている。これは、表 3 で訓練事例に対する正答率が悪いことから判断して、学習手順 2 のステップ 2 において、メンバーシップ関数族の GA 探索が十分に機能していないためと考えられる。

表 2 学習手順 2 の結果

	訓練事例に 対する正答率	テスト事例に 対する正答率
最良の実行	1.000000	0.903226
10 実行の平均	0.972355	0.780635
最悪の実行	0.943548	0.645161

表 3 学習手順 2 の途中結果（ステップ 2 終了時）

	訓練事例に 対する正答率	テスト事例に 対する正答率
最良の実行	0.903226	0.887097
10 実行の平均	0.900806	0.787742
最悪の実行	0.879032	0.709677

4 まとめ

今回、クラス別にファジィルール集合を求め、満足な実験結果が得られた。次にメンバーシップ関数の学習について考え実験したが、現状ではまだ満足な結果が得られていない。これからは、メンバーシップ関数の学習についてさらに研究を進めたい。

参考文献

- [1] C.-H. Wang, T.-P. Hong, S.-S. Tseng, "Integrating Fuzzy Knowledge by Genetic Algorithms," *IEEE Trans. on Evolutionary Computation*, vol.2, pp.138-149, 1998.
- [2] <ftp://ftp.ics.uci.edu/pub/machine-learning-database/hepatitis>
- [3] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.
- [4] C.-H. Wang, T.-P. Hong, S.-S. Tseng, "A Hybrid Genetic Knowledge-Integration Strategy," *1998 IEEE International Conference on Evolutionary Computation*, pp.587-591, 1998.