

マッチング処理の検討
永松 秀人

1 前月からの課題

- 全数探索を用いたマッチング処理
- GA を用いたマッチングプログラムの作成

2 全数探索

ダミーの特徴量データベース、および入力特徴量を仮定しマッチングの処理を全数探索的に行った。一致度合いの指標として用いた関数を式 (1) に示す。w は特徴量を使うか使わないかという重み (1 or 0) である。

$$Fitness(i) = \frac{1}{\sum w(i, j) \{a(i, j) - a(X, j)\}^2} \quad (1)$$

また、今回の実験で使用した特徴量は 4 つである。つまり、考えるべき重み係数の組み合わせは 15 通り (0000 は除く) となる。

特徴量データベースには 1~5 までの商品があり、入力される未知の特徴量も 4 つである。データベースと未知入力商品は 1 対 1 に対応していると仮定する。

2.1 実験結果

各入力データごとに、データベース上の商品 ID とのマッチングを重み係数の組み合わせすべてにおいて行った。もっとも Fitness 値が高くなった商品 ID の累積を Table 1 に示す。Input1=ID1, Input2=ID2... と対応するデータを作成したので、Table 1 上の対角線上がもっとも累積値が大きくなるはずである。

Table 1 マッチング結果

	ID=1	ID=2	ID=3	ID=4	ID=5
Input1	13/15	2/15	0/15	0/15	0/15
Input2	3/15	10/15	0/15	2/15	0/15
Input3	0/15	12/15	2/15	0/15	1/15
Input4	4/15	1/15	0/15	10/15	0/15
Input5	2/15	4/15	1/15	0/15	8/15

Table 1 において、ほぼ予測通り対角線上で一番累積値が大きくなっている。

このように、全数探索の実験では想定した商品 ID とのマッチングにおいて Fitness 値が一番高くなるのが重み係数の組み合わせの半数以上に達した。

3 GA を用いたマッチング

対象とする問題は、目的関数 (式 (1)) を最大化する問題と捉えることができる。ここで、作成すべき GA の概要を示す。

- 各重み $w(i, j)$ の値を並べて N ビットの 2 進記号列を遺伝子型とする。(N=特徴量数)
- 遺伝子型をそのまま表現型とする。(重さ 0, 重さ 1)
- 最大化問題, 目的関数が非負なので, Fitness (式 (1)) がそのまま適応度となる。

3.1 マッチング結果

Input1 とすべての商品とのマッチングを行い、もっとも Fitness 値を大きくする重み係数の決定を GA で行った。その結果を Table 2 に示す。求められた重み係数は仮定した特徴量と一致する。また、この結果は全数探索より得られた結果と一致した。

Table 2 Fitness の最大値

	Fitness	重み係数 (w)
ID=1	0.019231	0110
ID=2	0.004878	0011
ID=3	0.000987	1100
ID=4	0.000659	1010
ID=5	0.001000	1010

4 今後の課題

全数探索の結果の一部を Fig. 1 に示す。横軸が重み係数の平均値、縦軸が Fitness 値である。

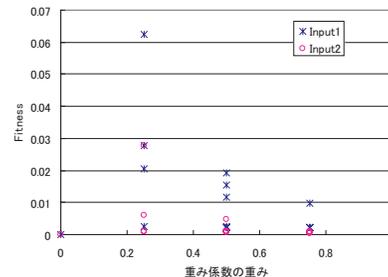


Fig. 1 全数探索の結果

重み係数中の 1 の数を減らせば Fitness の値は大きくなる。しかし、ある程度の特徴量数を使わないとマッチングに意味がなくなってしまう。よって、Fitness 関数の定義を変更する必要がある。