

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すべての組み合わせについて行った. 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.

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

また, 今回の実験で使用した特徴量は 8 つである. このとき, 考えるべき重み係数の組み合わせは 255 通り (00000000 は除く) となる.

特徴量データベースには 1~10 までの商品があり, 入力される未知の特徴量も 8 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 の定義

式 (1) 中に示す w は使用した特徴量の個数ごとに变化させる値である. 今回行った実験では, w を用いた特徴量数の 2 乗とした.

2.2 実験結果

各入力データごとに, データベース上の各商品 ID とのマッチングを重み係数のすべての組み合わせについて行った. その結果, 各商品 ID においてもっとも $Fitness$ 値が高くなる重み係数の組み合わせがわかる. また, このとき得られる $Fitness$ 値は, 入力データがどの商品とどれだけ似ているのかという指標でもある.

Table 1 入力を商品 9 と想定したマッチング結果

商品 ID	重み係数	$Fitness$ 値
商品 5	00000100	10.000000
商品 3	00000001	10.000000
商品 9	01000111	1.777778

ここで, それぞれの入力データにおいて $Fitness$ 値が高かった商品 ID の上位 3 つまでを調査した. その一例

を Table 1 に示す. 結果, すべての入力データにおいて, 上位 3 つまでに意図した商品 ID を得ることができた. しかし, 意図した商品 ID の $Fitness$ 値が一番高くなるということにはならなかった.

2.3 制約条件

マッチングを行う際に全体として似ているものを良しとするのか, 1 つ, 2 つの特徴量がぴったり一致しているものを良しとするのかによって, $Fitness$ 関数を調整しなければならない. 具体的な特徴量の様子を Fig. 1 に示す.

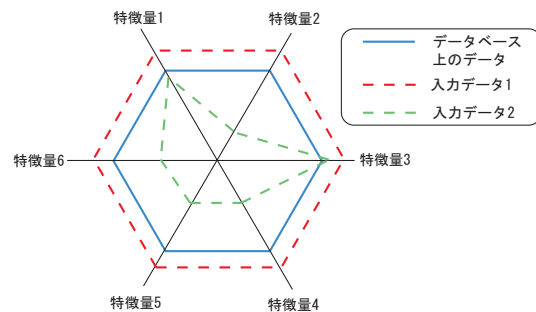


Fig. 1 特徴量の比較

入力データ 2 では, 特徴量 1, 3 のみでマッチングを行うとデータ 1 に比べて $Fitness$ 値は高くなる, 逆にすべての特徴量でマッチングを行うなら, 入力データ 1 はデータ 2 よりも $Fitness$ 値は高くなる.

考えられることとしては, いくら $Fitness$ 値が高くて, 特徴量 1 つでのマッチングでは信頼性は得られない. 最低でも全特徴量の半分ほどはマッチングに用いるのが妥当ではないかと考えられる. そこで, 最低でも全特徴量のうち半分の特徴量はマッチングに使用しなければならないという制約条件をつけマッチングを行うことにした.

すべての組み合わせについてマッチング処理を行った結果, 各入力データにおいて, 想定した商品 ID との $Fitness$ 値がもっとも高い値を示すこととなった.

3 今後の課題

- 実際の商品の特徴量を用いたマッチング処理
- システム側とのドッキング作業
- 自動化のためのプログラムの作成