Reference を使った複雑性の指標

福井大 工、福井県立大 生物資源[A]　村田拓生、工藤清、玉川洋一
中村量空[A]、山川修[A]、森崇

1 イントロダクション

今までの代表的な複雑性の指標は、対象を客観的に見る指標であった。しかし我々は、複雑さは人により感じ方が違うということから、複雑性の指標は対象を主観的に見なければならないと考えている。また我々は、一番複雑なのは困難を有するが理解はできる所だと考えている。つまり、簡単に理解できたり難しすぎて理解できない所は、複雑ではないと考えている。我々が考えているような複雑性の指標を、関心度と呼ぶことにする。

池上と津田が、複雑性の指標 C_2 を提案した [1]。これは、複雑さを測る対象を主観的に見る指標である。この指標は、我々が考えている様な指標になっている。

2 Reference を使った複雑性の指標

2.1 複雑性の指標 C_2 の求め方

池上と津田の提案した Reference を使った複雑性の指標 C_2 は、まず Tiling という作業を行なった後に測られる。Tiling のプロセスは、次の図 1 のようになる。Target Pattern は、複雑度を測る対象である。Reference System は、人間に例えると知識全体を表しており、1 つのセルが 1 つの知識を表す。また、このセルのことを Template と呼ぶ。Target Pattern を Template と同じ大きさに分割し、Reference System から Template を選び、 Target Pattern と同じ大きさになるまで貼っていく。この Template を貼っていく時に、Target Pattern と同じになるように、かつ隣同士の Template が同じものになるように貼っていく。

上記に述べられた Tiling のプロセスは、次に示すような Cost Function を最小化するプロセスになる ：

$$F = \sum_{i,j} < t_i|m_{ij}|t_j > + \lambda \sum_i \mathrm{Ham}(|g_i >, |t_i >) \tag{1}$$

ここで、$|t_i >$ と $|g_i >$ は、それぞれ i 番目の Template と Target Pattern のピクセルの配置を表す。第 1 項目の $< t_i|m_{ij}|t_j >$ は、i 番目と j 番目の Template の間の相互作用を表し、同

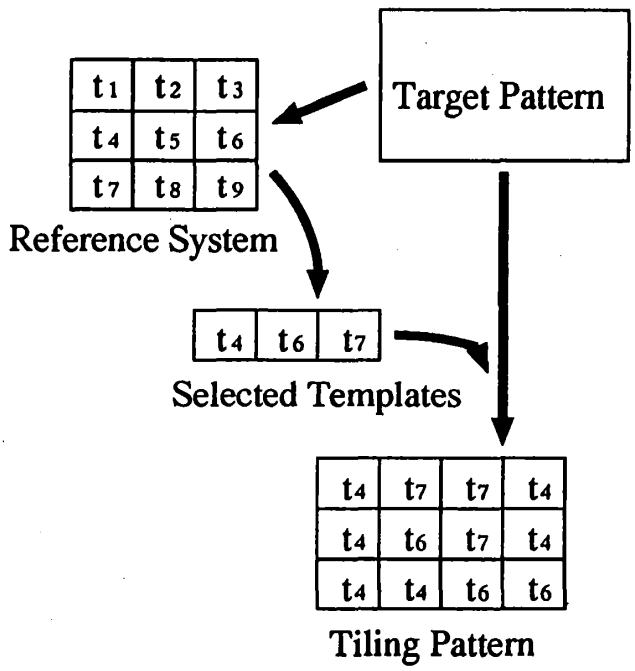

図 1: Tiling のアルゴリズム

じ Template なら -1 、そうでなければ 1 であるとする。第2項目の Target Pattern $|g_i>$ と Template $|t_i>$ の間のパターンの合い具合は、ピクセルの配置の間の Hamming distance で測られる。つまり、Ham$(|g_i>, |t_i>)$ は、$\sum_{j=1}^{m} ||t_{ij}> - |g_{ij}>|$ で与えられる。ここで、 m は 1つの Template の中のピクセルの数である。この時、第1項目の隣同士を合わせようとする効果と、第2項目の Target Pattern と合わせようとする効果の間で Frustration が起こる。この Tiling が終ってから複雑度を測る。またこの時、同じタイプの Template で集まっている領域をクラスターと呼ぶ。

この Frustration を反映した物が、次のような複雑性の指標 C_2 である ：

$$C_2 = \sum_{k=1}^{N_1} \left(\alpha - < \sum_{i=1}^{n_k} \mathrm{Ham}(|g_i>, |t_i>)/n_k >_k \right) \tag{2}$$

ここで、N_1はクラスターの数、$<>_k$ は k 番目のクラスターでの平均、n_k は k 番目のクラスターでのセルの数、α はゼロでない任意のパラメタである。

2.2 我々の考え方

我々は複雑さは人により感じ方が違うということから、複雑性の指標は対象を主観的に見なければならなく、また一番複雑だと人が感じる物は、理解するのに困難を有するが理解がで

きるという物だと考えている。簡単に理解できたり難しくて理解できない物は、複雑ではないと考えている。このような考え方を反映した複雑性の指標を、関心度と呼ぶことにする。

我々は、池上と津田が提案した複雑性の指標 C_2 が関心度のような物になっているのではないかと考え、異なる Target Pattern を Reference の数を変化させて C_2 を見た時、C_2 がどのように変化するかを調べた。

我々は、次のように Reference System と Target Pattern をそれぞれ３種類ずつ設定した。Reference System としては、Reference の種類の少ない物から、幼児、子供、大人と名前を付けた。幼児、子供、大人の Reference の種類は、それぞれ5、50、500 個である。1 個の Template の大きさは、$3 \times 3 = 9$ ビットである。今回は、それぞれの Reference には特別な意味を持たせてなくランダムに作られた物であるが、大人の Reference には必ず子供の Reference が含まれ、子供の Reference には必ず幼児の Reference が含まれるという条件を持たせている。Target Pattern は、幼児の Reference をランダムに並べたものを絵本、子供の Reference をランダムに並べたものをマンガ、大人の Reference をランダムに並べたものを小説と呼ぶ。Target Pattern の大きさは、$10 \times 10 = 100$ 個の Template である。

上記のように Reference System と Target Pattern を設定した時、複雑性の指標 C_2 が関心度になっているなら、子供の場合、絵本は簡単すぎ小説は難しすぎるので C_2 は低くなり、マンガはちょうどの難しさなので C_2 が高くなる。

3　シミュレーションと結果

上記のように、Reference System と Target Pattern をそれぞれ３種類ずつ設定した時のシミュレーションの結果が、図3である。それぞれの結果では、１０回ずつの C_2の平均とその標準偏差が示してある。複雑性の指標 C_2は、Cost Function の Frustration を反映した指標であるが、λ を変化させていった時、Cost Function が Frustration している所としていない所が出て来る。λ の小さい所と大きい所では、Tiling の仕方が一定になり Frustration は起こらない。しかし、その間では隣同士をそろえようとする効果と Target Pattern と合わせようとする効果により、Frustration が起こる。したがって C_2は、Fruatration が起こっている λ で見なければならない。図3では、$\lambda = 1.6$ で C_2 を測るべきである。

(a) では、幼児の Reference System で３種類の Target Pattern を見た時の C_2 の λ に対する変化である。Frustration が起こるのは $\lambda = 1.6$ であるから、そこで C_2 を見ると絵本が一番高く、小説とマンガが同じくらい低くなっている。(b) では、子供の Reference System で３種類の Target Pattern を見た時の C_2 の λ に対する変化が示してあり、$\lambda = 1.6$ ではマンガが一番高く、小説と絵本が同じくらい低くなっている。(c) では、大人の Reference System で３種

予想

Reference (多↑少)	絵本	マンガ	小説
幼児	高	中	低
子供	中	高	中
大人	低	中	高

Target (少 → 多)

(a)

結果 $\lambda = 1.6$

Reference (多↑少)	絵本	マンガ	小説
幼児	高	中	中
子供	中	高	中
大人	低	高	高

Target (少 → 多)

(b)

図 2: (a)C_2が関心度のような物であると考えた場合の予想。(b)Cost Function での frustration の起こる$\lambda = 1.6$ で測った時の C_2。

類の Target Pattern を見た時の C_2 の λ に対する変化が示してあり、$\lambda = 1.6$ では小説とマンガが同じくらい高く、絵本が低くなっている。これらをまとめたのが図 2の (b) である。

4 まとめ

図 2の (a) に、C_2が我々の考えている関心度のような物だとした時のシミュレーションの予想が示してある。例えば、子供の Reference System では C_2が関心度のような物であると、絵本は簡単すぎ小説は難しすぎるので C_2が低くなり、マンガはちょうどの難しさなので C_2が高くなることが予想されるということを示している。図 2の (b) に、実際のシミュレーションでの結果を示している。C_2は$\lambda = 1.6$ で Cost Function の frustration を良く反映しているので、$\lambda = 1.6$ での結果を示してある。

図 2を見て分かるように、予想と結果の一致が良い事が分かる。つまり、この複雑性の指標 C_2は、我々の考えている関心度のような物になっている事が分かった。しかし、必ずしも結果が予想通りにはならなかった。つまり、完全には我々の考えている関心度のような物にはなっていない。その理由として考えられる事は、Cost Function の第 1 項目の隣同士を合わせようとする効果がほとんど効かないという事である。複雑性の指標 C_2の場合、隣同士が全く同じ物でなければクラスターとして考慮されないが、実際には違う物同士で意味をなす事が多いので、その点を改良する事により、より我々の考えている関心度のような物が出来るのではないかと考えている。

参考文献

[1] T.Ikegami and I.Tsuda, 'Measuring Complexity: A Relative Aspect of Pattern Recognition', *International Journal of Modern Physics C*, Vol. 3, No. 3 (1992) 447-458

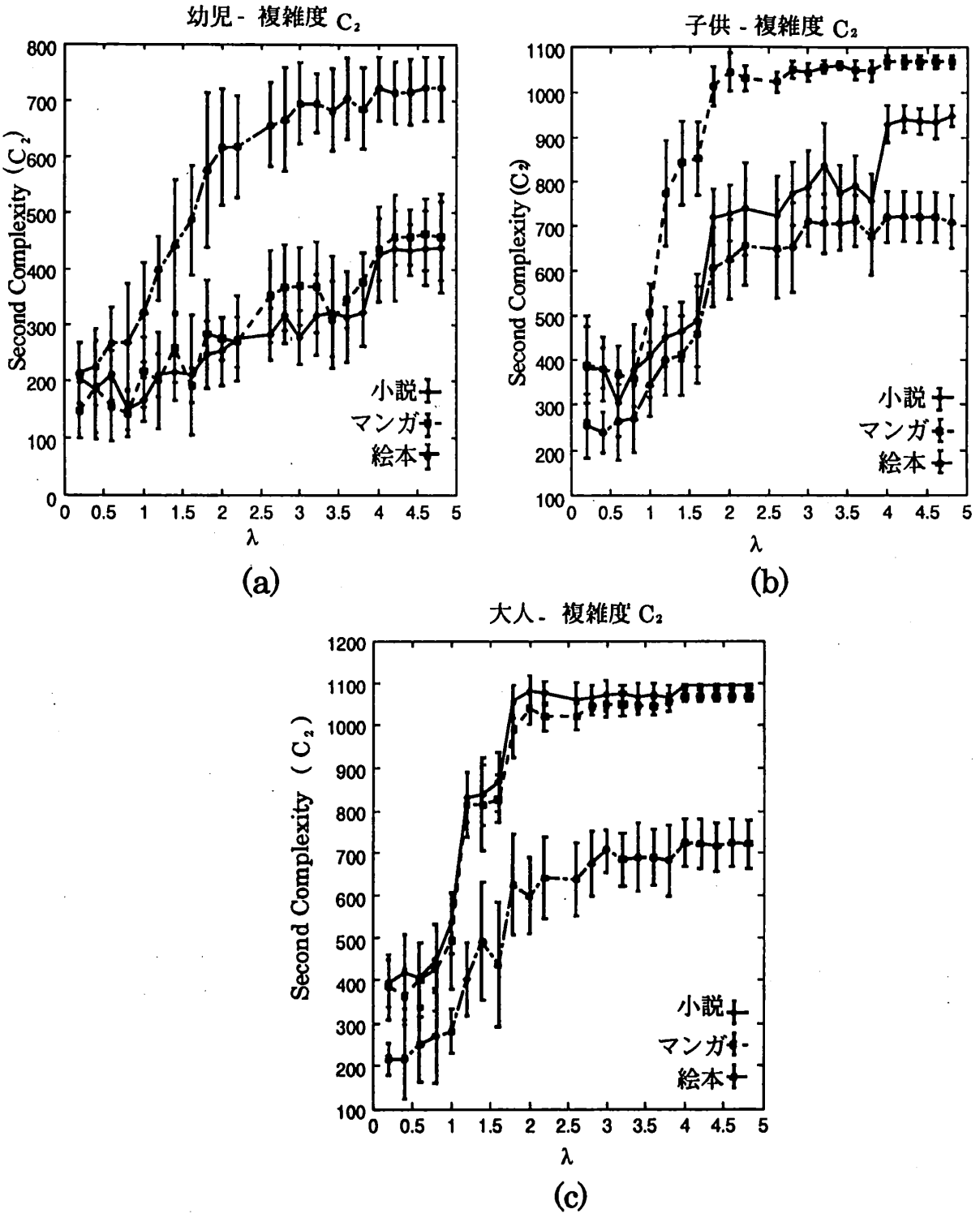

図 3: (a)Reference System に幼児を設定した時、(b)Reference System に子供を設定した時、(c)Reference System に大人を設定した時、それぞれで３種類の Target Pattern を測ったものを示す。