

# 発育時系列データの類似検索による育児ブログ推薦システム

山本 めぐみ<sup>†</sup> 川越 恭二<sup>†</sup>

<sup>†</sup>立命館大学 情報理工学部

## 1 はじめに

近年、子育てに関する疑問や情報交換を行うために、育児ブログを利用する保護者が増加している。育児ブログには、授乳やおむつ替えなどの育児に必要な基本知識や子どもの成長過程などが掲載されている。そのため、育児ブログは育児の詳細を聞けない状況にある親に有益な情報を提供している。

しかし育児ブログは 4000 程度も存在し、利用者が自分の子どもの発育に適した少数の育児ブログをこれらの中から検索することは困難である。また、子育てブログ [1] は既存の育児ブログ検索サイトとして存在するが、利用者の子どもの成長について考慮して作られていないことという問題がある。

そこで本研究では、発育時系列データの類似検索による育児ブログ推薦システムを提案する。掲載されている発育時系列データが子どもの発育状況をもっともよく反映していることに着目した。育児ブログに掲載されている子どもの発育時系列データと利用者の子どもの発育時系列データを比較し、類似する発育状況を検索する。このことにより、利用者は子どもの成長に適した育児ブログの閲覧が可能になる。

## 2 発育時系列データの類似検索

### 2.1 類似検索の基本的な考え方

本研究では、育児ブログに掲載されている子どもの発育時系列データと利用者の子どもの発育時系列データを比較し、類似する発育時系列データを持つ子どもの育児ブログを求める。

発育時系列データとは、子どもの身長や体重などの身体的な発育状況の時間的変化を示すデータである。ただし、単純化するために本稿では、育児ブログに掲載されている子どもの身長データと

体重データの 2 種類の数値を使用する。発育時系列データの時間軸を統一するために時間軸として月齢を使用する。発育時系列データの類似検索を行うために、月齢ごとに発育時系列の数値の差が小さい方を発育状況が類似していると考ええる。

全ての育児ブログが月齢ごとに身長と体重を記載しているわけではない。つまり、身体測定を毎月行わなかったり育児ブログに記載し忘れて、データが抜けている場合がある。このため、データの欠損を補うための補間が必要である。本稿では簡単化のために、欠けているデータの前後の数値を使用して補間を行う。したがって、1 点のみ欠けている場合には前後の平均値がその点の値として用いることとなる。

### 2.2 成長比較因子

成長比較因子とは、子どもの成長を比較するために使用する因子である。本稿では成長比較因子として、1) 身長データ、2) 体重データ、3) 身長・体重データの 3 種類を用いる。いずれも月齢による時間的に変化するデータである。

ここで、身長・体重の成長比較因子として身長  $h(\text{cm})$  と体重  $w(\text{g})$  から、次式 (1) のカウプ指数 [2]  $C$  を用いる。カウプ指数とは、乳幼児の発育状態を知るために実際に用いられている数値である。

$$C = \frac{w}{h^2} \cdot 10 \quad (1)$$

利用者はこのいずれかの成長比較因子から利用者の望む因子を 1 つ選択する。この因子を利用して類似度を算出する。

### 2.3 発育状況の類似度

$R_u = \{r_{u,t}\}$  と  $R_j = \{r_{j,t}\}$  を、利用者の子ども  $P_u$  及び成長記録データベースに登録されている子どもデータ  $P_j$  の時系列データとする。これらの時系列データの数値は利用者が選択した成長比較因子  $S$  に関して生後  $t(0 \leq t \leq tmax)$  カ月目の値である。ここで本稿では  $tmax = 16$  とする。 $R_u$  と  $R_j$  の時系列データの距離  $D(R_u, R_j)$  を (2) で定義する。

$$D(R_u, R_j) = \sum_{t=0}^{tmax} |r_{u,t} - r_{j,t}| \quad (2)$$

Child Care Blog Recommendation System Using Similarity Search of Time Series

Yamamoto Megumi<sup>†</sup>, Kyoji Kawagoe<sup>†</sup>

<sup>†</sup>College of Information Science and Engineering, Ritsumeikan University.

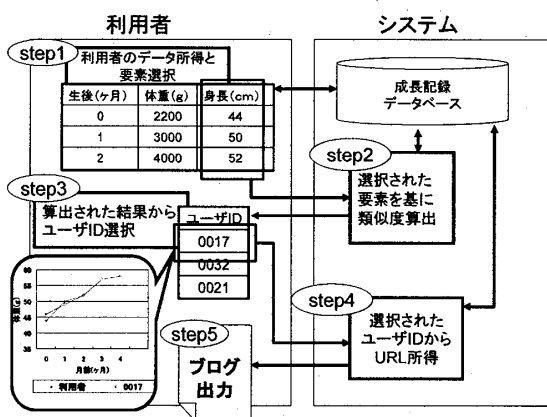


図 1: 育児ブログ推薦システム

類似度  $Sim(R_u, R_j)$  は、次の (3) 式に示すように  $D(R_u, R_j)$  の逆数とする。

$$Sim(R_u, R_j) = \frac{1}{D(R_u, R_j)} \quad (3)$$

### 2.4 検索対象の絞込み

2 種類の方法で算出対象の絞込みを行う。まず、一つ目は月齢に基づいて絞り込む。成長記録データベースに登録されている子どもの集合  $P$  における子ども ( $p_j \in P$ ) の月齢  $k_j$  が利用者の月齢  $k_u$  よりも大きい月齢の子どものみに限定する。なぜなら育児ブログ推薦システムは、利用者が子育てをする中で疑問に感じたことを解決するために育児ブログを提示する。そのため、育児ブログを推薦する対象は利用者の子どもの月齢より大きい必要がある。限定する検索対象のユーザの集合  $L_u$  を、以下の条件式 (4) によって示す。

$$L_u = \{p_j | k_j > k_u, p_j \in P\} \quad (4)$$

次に利用者の子ども  $P_u$  と比較する子ども  $P_j$  の各々の月齢での値  $r_{u,t}$  と  $r_{j,t}$  の差の絶対値  $|r_{u,t} - r_{j,t}|$  が最大値  $maxr_{i,t}$  と最小値  $minr_{i,t}$  の差の  $\alpha$  % に含まれている子どもだけに絞る。これは、子どもの発育時系列データが月齢ごとに偏りがあるからである。絞り込んだ対象の集合  $M_u$  を以下の式 (5) より算出する。

$$M_u = \{p_j | |r_{u,t} - r_{j,t}| < \frac{\alpha (maxr_{i,t} - minr_{i,t})}{100}, p_j \in L_u\} \quad (5)$$

## 3 発育時系列データの類似検索による育児ブログ推薦システム

### 3.1 育児ブログ推薦システムの流れ

利用者の子どもと類似している成長を持つ子どもの育児ブログを推薦するシステムを開発した。その流れを図 1 に示す。

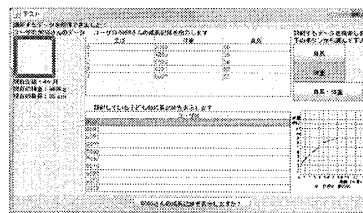


図 2: 育児ブログ推薦システムの実行例

[Step1] 利用者は、利用者 ID を入力する。システムは、成長記録データベースに問い合わせ、利用者の子どもの成長記録を表示させる。そして利用者は、成長比較因子を選択する。

[Step2] システムは、利用者が入力した成長比較因子を基に、子どもの成長記録データベースから子どもの発育状況の類似度を算出する。さらにシステムは、類似度算出で得られた子どものユーザ ID を類似度の高い順に並び替える。最後に、指定人数の子どものユーザ ID を類似度順に出力する。

[Step3] 利用者は、表示されたユーザ ID の中から参考にしたいユーザ ID を選択する。

[Step4] システムは、[Step3] で利用者に選択されたユーザ ID の URL をデータベースから所得して利用者に提示する。

[Step5] 利用者は、選択したユーザ ID の育児ブログの閲覧が可能となる。

### 3.2 育児ブログ推薦システムの実行例

実行例を図 2 に示す。左の項目には利用者の現在の子どもの身体状況と写真が掲載され、中央の項目の上部にはこれまでの成長記録を表で示す。中央の下部には 2 章で検索した結果を表示させる。右の項目の下部には利用者の子どもと、検索結果の中より選択したユーザ ID の子どもの発育時系列データをグラフで表示する。最後に利用者は、下部にあるボタンを押すと選択されたユーザ ID の育児ブログが自動的に表示される。

## 4 おわりに

今後は、本稿で提案した育児ブログ推薦システムの評価実験を行う。さらに、成長比較因子を身長と体重だけに絞るのではなく、子どものはいはいや立つといった運動機能にも着目する。運動機能にも着目することで利用者の子どもにより適した育児ブログが提供できると考える。

### 参考文献

[1] 株式会社ムラウチドットコム: “にほんブログ村”, <http://www.blogmura.com/>  
 [2] 毛利子来, 山田真: “育児児典 暮らし編”, 岩波書店, 2007 年