HistoryPaper: ユーザー個人のブラウザ履歴からの代表ページ選択とマガジンスタイルレイアウト

松枝　知香†　　伊藤　貴之†

† お茶の水女子大学　〒 112–8610 東京都文京区大塚 2-1-1
E-mail: †coco@itolab.is.ocha.ac.jp, ††itot@is.ocha.ac.jp

あらまし　インターネットを毎日利用する人の閲覧履歴を要約することは，その人の行動や知識の要約につながると考えられる．しかし，現在のブラウザに実装されている閲覧履歴の表示方法だけでは，そのような情報を有効活用することは簡単ではない．本論文では，1 日の閲覧履歴の中で特に重要であると判断した Web ページ群を抽出し，それらを雑誌のようにレイアウトすることで，ユーザーの毎日の行動や獲得知識を要約表示するシステムを提案する．提案手法ではまず，閲覧履歴を構成する Web ページ群をコンテンツ内容でクラスタリングし，検索キーワード，アクセス回数から定義される重要度を算出して各クラスタから代表 Web ページを選出する．続いて，マガジンスタイルを模倣するレイアウトアルゴリズムによって，代表 Web ページ群を一画面に配置して一覧表示する．提案するシステムは，ユーザの閲覧履歴をライフログとして活用することを可能にし，特定の日におけるユーザの行動を思い出すといった知識復元に役立つことが期待される．

キーワード　ブラウザ閲覧履歴，自動レイアウト，マガジンスタイル，履歴要約，ライフログ

1. はじめに

近年スマートフォンに加え多くのウェアラブルデバイスが誕生し，人々にとってライフログという言葉が身近になった．ユーザの日常生活に関わるデータをライフログとして収集することで，ユーザの行動支援や知的活動支援に活用することができる．ライフログには無意識な記録と，意識的な記録に大きく分類することができる．スマートフォンやウェアラブルデバイスを用いてユーザの位置情報や映像などを取得する方法は前者にあたり，ブログを書き綴るというような自発的な方法は後者にあたる．無意識的な記録はユーザにとっての負担が小さく，継続してデータを記録し続けられるという利点があるが，取得できる情報はデバイスの機能によって制限される．一方，意識的な記録はユーザの意図を多様に表現できるが，ユーザの負担が大きく，継続的な記録が困難な場合も多い [1].

そこで我々が注目したのがブラウザの閲覧履歴である．ユーザはライフログのために意識的にブラウザの閲覧履歴を記録しているわけではない．一方で閲覧履歴には，ユーザ自身の行動や，獲得した知識についての情報を含んでいる．つまり，閲覧履歴をライフログとして扱うことで，無意識な行動でありながら意識的な内容としてライフログを収集することができる．

既存の閲覧履歴に関する研究は，以前に訪れたページをもう一度訪れたい，思い出したいということに重点を置いたものが多い [3] [4]．本研究では，閲覧履歴を特定の何かを思い出すためではなく，ライフログとして自己を振り返り，特定の日におけるユーザ自身の行動を思い出す，ユーザが過去に獲得した知識をもう一度思い出す，といった知識復元のために用いる．

閲覧履歴をライフログとして利用するにあたって，ユーザが普段から気軽に活用できる状態で提供する必要がある．無意識的なライフログは，一般的にひと目で状況が把握できるようにグラフ等で可視化表示することが多い．しかし閲覧履歴は情報量が多く，歩数情報や睡眠状態等の様にユーザがひと目で認識できるよう可視化するのが簡単ではない．閲覧履歴を可視化する研究 [2] も存在するが，一般のユーザにとって理解しやすいものは多くない．そこで注目したのがマガジンスタイルと呼ばれる雑誌風のレイアウトである．マガジンスタイルは普段から雑誌を読むユーザにとっても見慣れたものであり，情報量の多い閲覧履歴をまとめるのに適した可視化方法であると考えられる．

本論文の提案手法では，閲覧履歴から重要な Web ページを代表として抜き出し，それらのページをマガジンスタイルでレイアウトすることでユーザの 1 日を一覧表示するシステムを提案する．

2. マガジンスタイル

多くの Web ページの運営者は，たくさんのコンテンツをいかに見やすく表示させるかという課題に直面している．効率的な Web ページのレイアウトをするにあたって，まずはユーザの閲覧行動を観察することが不可欠である．

Weinreich らの Web ページでのユーザの行動に関する実験 [5] によると，ユーザは完全・連続的には Web ページを読んでいない．ブラウジングは左上から右下へ行われ，ほとんどの場合は左上の 1/4 しか読まれていないことが示されている．このようなユーザの行動を受け，オンラインニュースやショッピングサイトの様な情報量の多いサイトでは，長年マルチカラムレイアウトが用いられてきた．マルチカラムレイアウトは 1 ページに埋め込める情報量が多くなってもユーザが理解しやすいという特徴があるためである [6]．しかし，最近ではマガジンスタ

イルと呼ばれる，長方形モジュールが並ぶ雑誌のようなレイアウトが多く採用されている．Nebeling らの研究・実験 [6] によると，ユーザはマガジンスタイルのサイトを見る際に，まず全体を見渡してからどのコンテンツを読むか決めるということがわかっている．マガジンスタイルは，マルチカラムよりも配置が柔軟で，印刷された雑誌や新聞で用いられているようなレイアウトである．ユーザは印刷された雑誌や新聞に馴染みがあるため，マガジンスタイルレイアウトは一覧性が高く，ユーザがどの情報が自分にとって重要か判断しやすいレイアウトである（例: 図 1）．本研究はマガジンスタイルレイアウトの自動生成手法を提案する．新聞の自動レイアウト [9]，汎用的な長方形分割 [10] に関する既存研究は既に多くあるが，マガジンスタイルに特化した長方形分割手法は我々の知る限りで見当たらない．

2.1 マガジンスタイルの定義

マガジンスタイルらしい Web ページの自動生成を実現するために，これらの観察結果にもとづいて，本研究におけるマガジンスタイルの 4 つの生成方針を定義する．

定義（1） 2〜4 程度の縦のカラムに分かれている

定義（2） 重要度が高い記事ほど割り当てられる長方形領域の面積が大きい

定義（3） 同じ大きさの長方形領域が隣り合う

定義（4） アスペクト比は表 1 のいずれかに近くなる

定義（1）（2）は，マガジンスタイルと呼ばれるレイアウトの基本である．しかし，この条件だけではマガジンスタイルを模倣するレイアウトを自動で作成するのは難しい．マガジンスタイルがどのようなレイアウトであるかの明確な定義を我々の知る限り見当たらない．よりマガジンスタイルらしい Web ページレイアウトの自動生成を実現するために，マガジンスタイルを採用しているニュースサイト (注1) と，そのサイト内の計 236 個の記事について調べた．

その結果，マガジンスタイルには同じ大きさの長方形領域が隣り合うという特徴があるということがわかった．調べた記事のうち，94.5%の領域が，同じ大きさの長方形が 2 つ以上隣り合ったレイアウトであった（図 1，2）．特に，長方形領域のアスペクト比が正方形から遠のくほど複数の同じ大きさの長方形が隣合う傾向が見られた．そのため，この特徴を定義（3）として採用した．

図 1　隣り合う同じ大きさの領域例

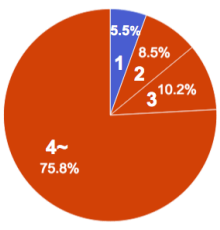

図 2　同じ大きさの領域の個数

(注1)：The New York Times (http://www.nytimes.com/), The New Yorker (http://www.newyorker.com/), Japan Today (http://www.japantoday.com/), sky NEWS (http://news.sky.com/), BBC NEWS (http://www.bbc.com/news/)

さらにマガジンスタイルレイアウトにおける長方形領域のアスペクト比には，ある程度の規則性があることがわかった．例えば極端に縦に細長い長方形領域は，文字の改行があまりに多くなり読みづらいため存在しない．実際にどのような傾向があるか計測するため，記事のアスペクト比の測定を行った．その際に，記事のアスペクト比は画像の有無等によって傾向が違うため，画像が上もしくは下にある記事，画像が左右にある記事，画像のない記事で分けて測定を行った．図 3 は各々のタイプの記事のアスペクト比の分布を示した図である．

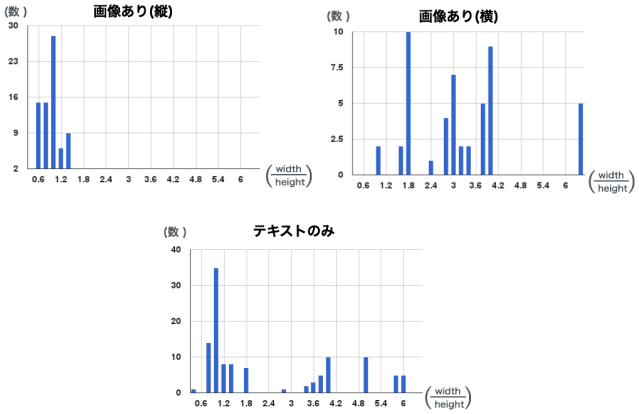

図 3　タイプ別アスペクト比分布

この結果を受け，長方形のアスペクト比がマガジンスタイルらしくなるよう定義（4）を採用した．表 1 は，図 3 の結果を受けて作成した，アスペクト比率の分布が特に多い部分の表である．

表 1　各長方形領域の理想的なアスペクト比

表示タイプ	横 / 縦 (画像位置)	最小連続数
画像有	0.9 (上)	1
	1.6 (左)	1
	3.0 (左)	2
	3.8 (左)	2
画像無	1.0	1
	3.8	2
	5.0	2

なお表 1 に示した数値のうち，上のものほど優先度が高いとする．また，表 1 の最小連続数は，定義（3）を踏まえて最低でも同じサイズがいくつ隣合う必要があるかという値である．

3. 閲覧履歴からの代表ページ選出

本章ではユーザの特定の 1 日の閲覧履歴の中から代表的な Web ページを抽出するアルゴリズムについて説明する．文書の要約手法の 1 つで，tf-idf 法によって測った重要度の高い文を重複を避けて抽出することで，満足度の高い結果がでる傾向があることが知られている [12]．本手法ではこの知見を利用して，閲覧履歴を構成する Web ページ群に対して文書内容の類似度に基づくクラスタリングを適用し，各クラスタの中から最も重要度の高いものを選ぶことで，履歴の要約となる Web ページ群を構成する．

3.1 閲覧履歴のクラスタリング

はじめに，1 日の履歴にある全ての Web ページの内容を対象として，Web ページをクラスタリングする．

まず Web ページのコンテンツ内容・タイトルを形態素解析し，Bag-of-Words 表現に変換する．Bag-of-Words 表現とは，単語の出現順序は考慮せず，出現頻度によって文書をベクトルで表現するモデルである．次に，潜在的意味解析を用いて Bag-of-Words を次元削減する．潜在意味解析は，「車」と「自動車」のように違う言葉でも同じような意味を持つ概念の集合を作成する手法である．

この次元削減された Bag-of-Words を用いて閲覧履歴をクラスタリングする．クラスタリング手法は，各データ間の距離が近いものから同じクラスタに融合していく階層クラスタリングと，分割と評価関数の再計算を繰り返して最適なクラスタを作成する非階層クラスタリングがある．閲覧履歴のクラスタリングを行った先行研究では，階層クラスタリングの 1 つである最短距離法や，それを応用した手法を用いていることが多い [7] [8]．しかし，最短距離法は大きなクラスタを少数作るという傾向があるため，小さい話題の分類結果が適当にならないことが多い．そのため本研究ではクラスタ数が決定している K-means 法を用いてクラスタリングを行う．クラスタ数は n 個とし，ブラウザの大きさによって 8〜16 程度の値になる．

3.2 各 Web ページの重要度計算

続いて本手法では，3.1 節の方法で生成した各クラスタについて，以下の変数を用いて Web ページの重要度を計算する．まずはじめに，ユーザにとってどのような Web ページが 1 日を表すページとしての重要であると考えるか，1 日 10〜100 個の Web ページを閲覧する 10 名にアンケートを行った．その結果を図 4 に示す．アンケートの結果，特に検索キーワードを多く含むページや，普段からは訪れていないページが特に重要なページであると考える人が多いことがわかった．

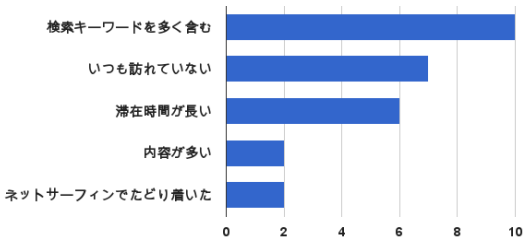

図 4　どんなページが代表ページにふさわしいか

また，履歴を構成する各ページを再度閲覧してもらい，代表ページに相応しいと思うページを選んでもらった．図 5 は，横軸に各ページにアクセスした回数，縦軸に各ページに含まれているその日の検索キーワード数を割り当てたヒートマップである．代表ページに相応しいページの比率が高いほど赤く，代表ページに相応しくないページの比率が高いほど青く表示される．図 5 の左上に行くほど，今まであまり訪問しておらず，その日知りたかったことを含むページ，右下にいくほど普段から頻繁に訪れているページを意味する．図 5 から，被験者は普段は訪問しておらず，検索キーワードを多く含むページが代表ページとして相応しいと考えていることがわかる．これらの 2 つのアンケート結果から，代表ページには，検索キーワードを多く含み，普段からは訪れていないページの重要度が高いという仮説を立てた．

そこで以下の 2 変数を用いて，式 1 を用いて各 Web ページの重要度 p を計算する．この値が最大である Web ページを，クラスタの代表として選出する．

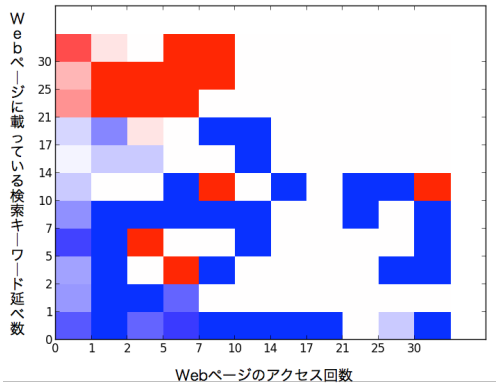

図 5　代表ページに相応しいと考えるページページと相応しくないと考えるページの比率

- m: 1 日に用いた検索キーワードのうち当該 Web ページに載っているキーワードの延べ数
- q: アクセス回数

$$\mathrm{p} = \frac{m+1}{q} \tag{1}$$

この定義により本手法では，普段はアクセスしないがその日にだけアクセスした Web ページ，あるいはユーザがその日検索に用いたキーワードが多く出現する Web ページを，重要な Web ページとして積極的に表示する．

また，アンケートの結果でも滞在時間は重要度に関連する情報ではあるが，滞在時間を取得するのにはユーザの個人情報をサーバーに送る，もしくはローカルサーバを設置する必要があるなど，一般のユーザにとって使いづらいシステムになってしまう可能性があるため，滞在時間情報の取得は行わない．

3.3 各クラスタの重要度計算

続いて本手法では，我々自身の定義による以下の式により，クラスタの重要度 c_p を算出する．

$$\mathrm{c}_p = \sum_{k=1}^{n_c} p_k \tag{2}$$

ここで n_c は当該クラスタに属する Web ページ数，p_k は当該クラスタに属する k 番目の Web ページの重要度 (式 (1) により算出) である．

3.4 代表ページの選出結果

本節では筆者のある日の閲覧履歴から代表ページを選出した例を紹介する．筆者がその日に検索したキーワード群を図 6 に

示す．ここでフォントサイズの大きさは検索に用いた回数を示す．表 2 は各クラスタから選出された代表 Web ページである．ここで c_p は 3.3 節で述べたクラスタの重要度である．

図 6　検索キーワード例

表 2　代表ページの選出結果

c_p	Web ページタイトル
399	2D bin packing with javascript and ⟨*canvas*⟩
243	ログの出力 - Log クラス - Android 入門
239	京都大学 永持・趙研究室 「研究内容紹介 〉 詰め込み問題に対する実用的なアルゴリズムの開発」
199	ポートチェック【外部からポート開放確認】
107	heuristic の意味・用例｜英辞郎 on the WEB：アルク
90	1 章 箱詰めの数理 1 1.1 アルゴリズムと計算量 . . . . . .3 - pdf.io
72	SIMPLE IS BETTER THAN COMPLEX. 長方形詰め込み問題の近似解法まとめ
57	よくあるご質問 — レンタルサーバーの WADAX FAQ 詳細
16	FINAL FANTASY　仲間を求めて 公式歌 Ver

図 6 と表 2 を照らし合わせ，実際にどんなページが代表として算出されたのかを検証する．

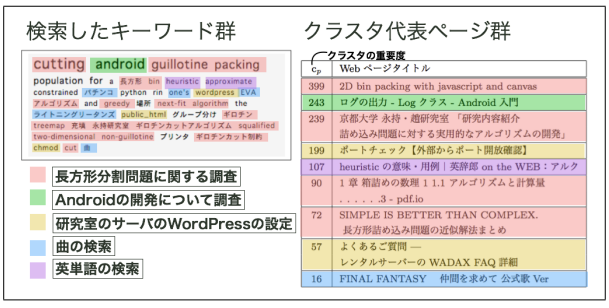

図 7　ページ選出結果

図 7 のうち，赤で塗られたものが長方形分割・詰め込みに関するキーワードと代表ページである．キーワード群のうち長方形分割に関するキーワードは半分弱を占めているが，代表ページも約半分が長方形分割に関連するものであることが見て取れる．また，緑で塗られた部分を見ると，ユーザは Android について何度も調べたことが検索キーワードの文字の大きさから見て取れる．実際に Android に関するページが 1 つ選出されている．黄色で塗られたものは，研究室のサーバの WordPress を設定した時に検索したキーワードである．それらに関するページが 2 つ選出されている．そして，青で塗られたものは曲を検索した時に関するものであり，紫で塗られる部分が英単語について検索したページである．このように，検索キーワードだけ見ても一見任意のキーワード間の関係性や共起性を想像するのは容易ではないが，ユーザの行動に沿って見ていくと，それぞれの行動に関して少しずつ選出できていることがわかる．

4.　レイアウトアルゴリズム

本章では，3. 節に示した手法で選出した n 個の Web ページ群を，マガジンスタイルの Web サイト (注1) のようなデザインで Web ブラウザの一画面に配置する貪欲なアルゴリズムを提案する．

4.1　予 備 実 験

マガジンスタイルのレイアウトアルゴリズムの開発に先立ち，我々は Treemap による配置を試みた．Treemap は二次元の領域を入れ子状に分割することによって，木構造のデータを可視化する手法であり，重要度に比例した面積を各領域に割り当てることができる．また Squalified Treemap [10] を用いることで，全ての長方形領域の形状が正方形に近くなるように配置することが可能である．これを利用して，クラスタ重要度に比例した面積を各ページに割り当てて配置した結果が図 8 である．Squarified Treemap のレイアウト生成には d3.js を用いている．

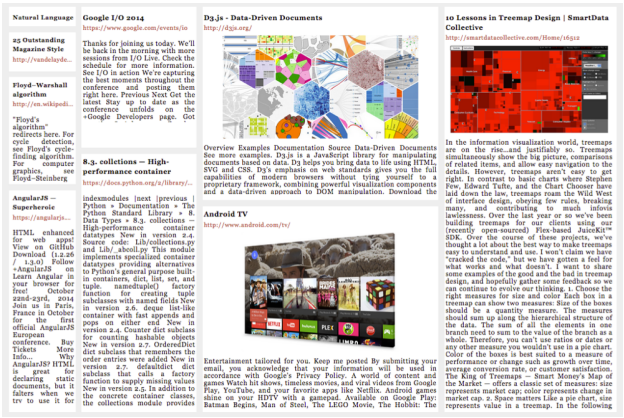

図 8　Treemap による配置実験結果

いくつかの Web ページ群にて同じ実験を試みた結果，我々は以下の問題点に気がついた．

- 小さすぎるスペースには文字が入らず，大きなスペースには余裕がありすぎる．言い換えれば，我々の定義に従って算出したクラスタ重要度に忠実な面積を各ページに割り当てるより，視覚的に好ましい範囲内で面積を調整した上で各ページを表示したほうがよい．
- Squarified Treemap は全ての長方形領域の形状を正方形に近づけることを目標としているが，現実には一部の長方形領域の形状が細長く歪む結果を生みやすい．言い換えればアスペクト比の最悪値が悪く，一部の記事が犠牲になって見難くなるような配置結果を生みやすい．

(注1)：例: The New York Times http://www.nytimes.com/

つまり Squarified Treemap による配置は，2.1 節での定義のうち，定義（1）（2）を満たすが，定義（3）（4）を満たすのは難しいことがわかった．

これらの問題点を踏まえて，Web ページ群のレイアウトをマガジンスタイルにより近づけるために我々が開発したアルゴリズムを，以下に説明する．

4.2 Web ページ群のデータ構造

本節でのアルゴリズムの説明に先立ち，以下のようにクラスタ等の集合を定義する．

- R: クラスタの集合 (3. 章に示した処理で生成する)
- R_i: i 番目のクラスタ
- G: 1 個以上のクラスタで構成されるクラスタグループの集合
- G_i: i 番目のクラスタグループ
- C: 1 個以上の任意のクラスタグループ G_i,G_j,... で構成される集合 (i,j は任意の自然数)
- S: R，G，G_i，C のいずれかの集合

また，以下の変数を定義する．

- $|S|$: 集合 S に含まれる集合の数
- $area_S$: 集合 S 全体の面積占有度
- $priority_S$: 集合 S 全体の合計クラスタ重要度
- $ratio_S$: 集合 S 全体の理想的なアスペクト比群 (表 1 参照)
- $type_R$: クラスタ画像の有無 (ブール値)
- W: 配置に用いる画面の横幅 (定数)
- H: 配置に用いる画面の縦幅 (定数)
- W_{min}, H_{min}: 各クラスタが配置される長方形領域の最小横幅，最小縦幅 (定数)

4.3 クラスタのグループ化

(1)

集合	クラスタ重要度	画像
R_1	45	o
R_2	32	o
R_3	30	o
R_4	20	
R_5	11	
R_6	10	
R_7	5	
R_8	2	
R_9	1	

(2)

集合	クラスタ重要度	画像
G_1	45	o
G_2	32 30	o
G_3	20	
G_4	11 10	
G_5	5	
G_6	2 1	

(3)

集合	クラスタ重要度	面積占有度	画像
G_1	45	20	o
G_2	32 30	15 15	o
G_3	20	11	
G_4	11 10	7 7	
G_5	5	4	
G_6	2 1	3 2	

図 9　クラスタのグループ化と面積占有度

本手法では，クラスタの重要度が近いものどうしを同じ大きさの長方形領域で隣接表示するために，クラスタを重要度でグループ化する．本手法では以下の式を満たす場合に R_i と R_j の 2 クラスタを同一グループに所属させる．ただし，R_i が既に他のグループに所属している場合，$type_{R_i}$ と $type_{R_j}$ が違う場合には，この処理を適用しない．

$$priority_{Rj} \leqq \lceil (priority_{Ri} \times 1.2) \rceil \tag{3}$$

例として図 9 において，(1) のような重要度を持つクラスタ群は (2) のようにグループ化される．結果としてクラスタ群が n_G 個のクラスタグループを構成する際に，本節ではその各々を G_1....G_{n_G} と呼ぶ．このようにしてクラスタグループを構成することで，2.1 節に示した定義（4）を満たすことができる．

4.4 各クラスタの面積占有度

続いて，各クラスタの重要度をデフォルメした値を，配置する長方形領域の面積占有度として算出する．各クラスタの重要度をそのまま面積占有度としない理由は，2.1 節であげた問題点を解決するためである．ここで極端に面積が小さいと，先述したように Web ページから埋め込める情報が非常に小さくなり，記事の内容を視認しにくくなる．そこで本手法では，以下の式でクラスタの面積占有度を算出する．ここで，x はデフォルメのための定数であり，現在 1.3 としている．

$$area_{R_i} = \left(\frac{priority_{R_i}}{W_{min}H_{min}} \frac{WH}{priority_R} \right)^{\frac{1}{x}} \tag{4}$$

図 9(2) に示したクラスタグループの集合に対して，式 (4) を用いて面積占有度を算出した例を，図 9(3) に示す．

4.5 配置アルゴリズム

続いて本手法では，前節で算出した面積占有度に従って Web ブラウザのウィンドウを長方形分割し，その各領域に 3. 章に示した手法によって選ばれた Web ページ群を配置する．本節ではその配置アルゴリズムを説明する．また，配置アルゴリズムをイラスト化した例を図 10 に示す．

（1） $S = G$ とする．$rect_W = W$，$rect_H = H$ とする．

（2） S の中で面積占有度の平均が最大であるクラスタグループを S から抜き出し，S_{top} とする．S_{top} を構成する各クラスタの面積を算出し，各種のアスペクト比 $ratio_{S_{top}}$ に応じた形状で仮配置する．図 10 (1.(a)(b)) に 2 種類のアスペクト比で仮配置した例を示す．なお，仮配置によって画面の右端に残る空き領域の幅が W_{min} 以下，下に残る空き領域が H_{min} 以下である場合には，図 10(3.(a)，4.) に示すように，その隙間を埋めるように S_{top} を横に引き伸ばして仮配置する．

（3） 仮配置した長方形の下にできる長方形の空き領域（図 10(1.(a)(b)，3.(a)(b)) の水色の部分) の面積占有度を求め，それと最も画面占有度の近いクラスタグループの 2 つまでの組み合わせを選ぶ．これを各々の仮配置結果に対して反復し，空き領域とクラスタグループとの面積占有度の差が最も小さい仮配置結果を採用する．図 10(1.) の例では (a) を採用し，G_1 と G_3 を一列に表示している．また，図 10(3.) の例では (a) を採用し，G_2 と G_6 を一列に表示している．以上の処理によって配置されたクラスタグループ群を S から抜き出し，それらの集合を以下 S_{opt} と称する．

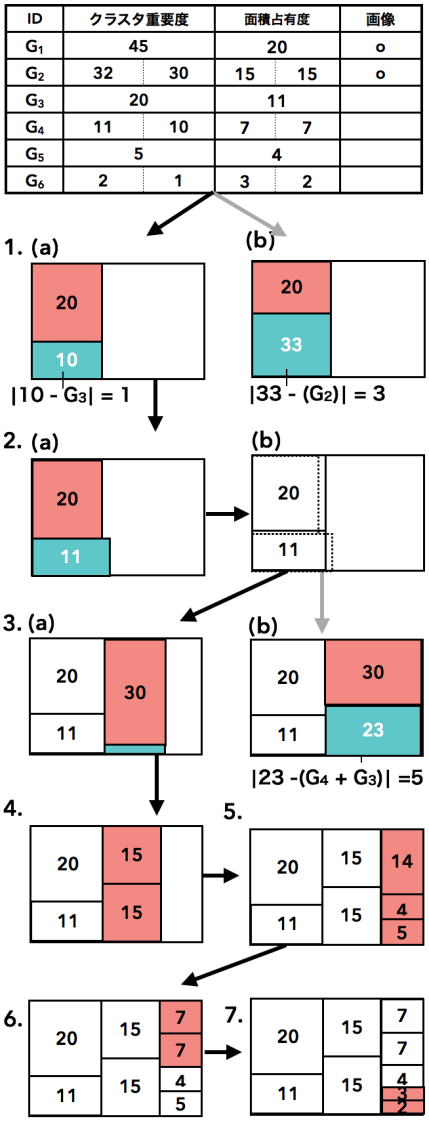

図 10　配置アルゴリズム

（4）　このようにして一列に配置された長方形領域群は，多くの場合において右端の稜線の位置が合わないので，これを合わせるように長方形領域群を変形する．図 10(3.(a)) が変形前の例，図 10(4.) が変形後の例である．このとき，変形前の長方形領域群に対応するクラスタ群の面積占有度の合計を $area_{before}$, 変形後の面積占有度の合計を $area_{after}$ とするとき，$area_{before} = area_{after}$ が成立する位置に右端の稜線を合わせるものとする．

（5）　S_{opt} について，以下の処理を実施する．

（a）　$rect_W = W_{S_{top}}$，$rect_H = rect_H - H_{S_{top}}$ とする

（b）　$|S_{opt}| \geq 3$ の場合 $|S_{opt}|$ を S として (2)〜(3) を適用する．この処理によって，長方形分割された一部分をさらにマガジンスタイルで再分割したような結果が得られる．

（c）　$|S_{opt}| < 3$ の場合 $|S_{opt}|$ を S として (6) を適用する．結果として，図 10(5.(a)) の 14 の領域を図 10(6.) のように再分割する．

また S についても，以下の処理を実施する．

（a）　$rect_W = rect_W - W_{S_{top}}$ とする

（b）　$|S| \geq 3$ の場合 (2)(3) を適用する．結果として図 10(3.(a)(b)) のように，長方形領域群の新しい一列を生成する．

（c）　$|S| < 3$ の場合 (6) を適用する．結果として図 10(5.) のように，右端の一列を長方形領域で埋めることができる．

（6）　S について以下の処理を行う．

（a）　S が G_i の場合，面積を $|S|$ 個に均等に縦分割もしくは横分割し，アスペクト比が理想に近い方を採用する．

（b）　S が C の場合，面積を A_{C_i} の比率で縦分割もしくは横分割する．続いて，C_i の各々を S として以下の処理を適用する．

- $|S| \geq 3$ の場合 (2)(3) を適用する．
- $|S| < 3$ の場合 (6) を適用する．

3．において，組み合わせを 2 つまでに限定しているのは，この要求が一般的に NP 困難で有ることが知られているためである．2 つまでの組み合わせに限定することにより，計算量は $O(|R|^{|R|})$ から $O(|R|^2)$ になる．また，これによってカラムの横幅が大きくなりすぎないという利点がある．

このアルゴリズムを用いることで，定義（1）に沿って画面全体を大きく縦に分割し，定義（2）に沿って重要度に概ね沿った大きさで各ページを表示する．また重要度の高いページは画面の左上に配置されやすくなる．またクラスタのグループ化によって定義（3）を満たし，各長方形領域において好ましいアスペクト比のうち 1 個を適応的に選択してそれに近づけることで定義（4）を満たす．

4.6　レイアウトの実行結果

図 8 に示した配置実験にも適用した図 9 のデータを用いて，提案手法を実行した例を図 11 に示す．画面全体を大きく縦に分割し，それをさらに分割して，重要度に応じた面積を割り当てて各ページを表示していることがわかる．また，同じ大きさのページを並べたレイアウトが随所に見られることがわかる．以上により，提案手法が 2. 章で論じた定義に近いレイアウトを実現できていることがわかる．

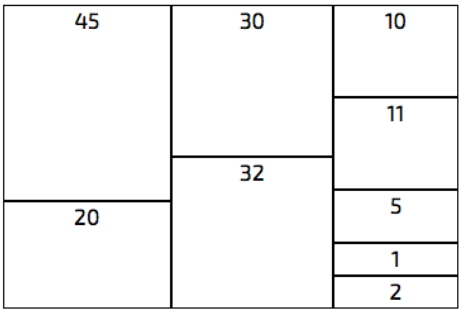

図 11　提案手法の実行結果

5.　レイアウトの評価関数

提案手法の定量評価および他の手法との比較のために，レイ

アウトの評価関数を以下のとおり 2 つ定義する．以下の説明では，i 番目の長方形のアスペクト比を r_i と称する．

（1） 表 1 で定義した理想のアスペクト比からの差異の総計 D_{ratio} を，式 (5) で示す．ただし，長方形領域中に画像がある場合 $x = \{0.9, 1.6, 3.0, 3.8\}$，画像がない場合 $x = \{1.0, 3.8, 5.0\}$ である．

$$D_{ratio} = \sum_{i=0}^{N} (\min\{f(x,i)\})^2 - 1 \tag{5}$$

ただし

$$f(x,i) = \begin{cases} \dfrac{x}{r_i} & (x > r_i) \\ \dfrac{r_i}{x} & (x \leqq r_i) \end{cases}$$

（2） 理想のアスペクト比と長方形領域のアスペクト比との差異の最悪値 W_{ratio} を，式 (6) で定義する．

$$W_{ratio} = \max\{D_{ratio}(i), i = 1, 2, \cdots, N\} \tag{6}$$

いずれも値が 0 に近いほど評価が良いとする．

4.1 節で紹介した Treemap によるレイアウト結果と，提案手法でマガジンスタイルを生成したレイアウト結果について，評価を比較する．

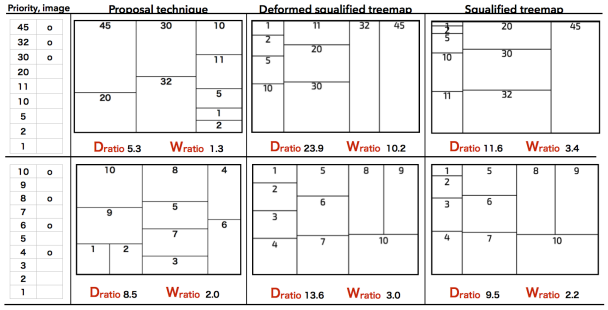

図 12　アルゴリズムの比較

Squarified Treemap，重要度のデフォルメを適用した Squarified Treemap，提案手法，の 3 通りの手法を対象として，2 種類の長方形群を適用した際のレイアウト生成結果，および D_{ratio}, W_{ratio} の算出結果を，図 12 に示す．図 12 の上は，Squarified Treemap ではレイアウトに歪みができやすいデータセットであり，下は Squarified Treemap でも比較的綺麗にレイアウトされるデータセットである．デフォルメの有無にかかわらず，Squarified Treemap は右下に正方形から遠い長方形ができやすいため，アスペクト比の最悪値 W_{ratio} が極端に悪くなる場合がある．本手法の評価は，重要度のデフォルメを適用した Squarified Treemap や Squarified Treemap よりも，D_{ratio}，W_{ratio} 共に良くなっていることが見て取れる．

6. 閲覧履歴を用いたライフログ

本章では提案手法 3. 章で選出した代表ページ群を，4. 章で紹介したアルゴリズムを用いて配置する．1 画面に 1 日の情報を埋め込むことによって，その日の出来事を要約し，一覧表示する．

6.1 クラスタ代表ページのマガジンスタイルへの埋め込み

4. 章で作った長方形のスペースに，それぞれのクラスタの情報を埋め込んでいく．スペースには，以下の様な情報を埋め込む．

- クラスタ代表ページのタイトル
- クラスタ内で重要度が高い 4 つまでの記事のタイトル
- クラスタの Web ページが多く含む検索キーワードのタグ
- OGP タグ (注2) の description の内容もしくはコンテンツの要約

コンテンツの要約については，様々な要約手法が提案されているが，ここでは Edmundson が提案したリード法 [11] とよばれる，文章の先頭から一定の字数抜粋する単純な要約手法を用いている．また，スペースが足りない場合は，上の要素ほど重要度が高いとし，スペースに入る分のみを配置する．

6.2 HistoryPaper の実行結果

代表ページの選出・レイアウトの実行結果を踏まえて作った閲覧履歴の要約「HistoryPaper」の例を以下に示す．

図 13 は，ユーザが 1 日クラスタリングや機械学習について調べていたのがよく分かる結果になっている．

6.3 ユーザテスト

20 代女性 10 人に HistoryPaper を実際に利用してもらい，ユーザアンケートを行った．数日前の行動について，HistoryPaper を用いることによって忘れていたけれど思い出した行動があったか，という質問に対し，9 人が思い出すことがあったと答え，HistoryPaper が 1 日を振り返ることの一助になることがわかった．

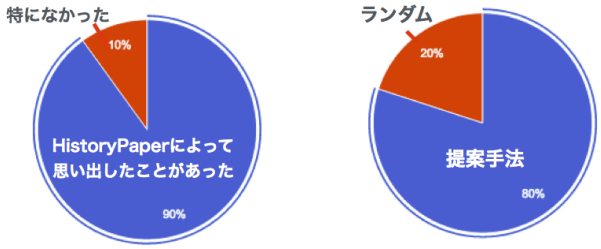

図 14　ユーザアンケートの結果
(左): HistoryPaper を用いることで思い出すことがあったか
(右): 代表ページの選出手法はどれが良かったか

また，以下の 3 つのうちどの選出手法を用いた結果が好みかのアンケートをとった．

- 提案手法での代表ページ選出
- 閲覧履歴全体からランダムな選出
- クラスタリングした後にクラスタの中から 1 つずつランダムに選出

その結果，8 人は提案手法を選んだものの，2 人は閲覧履歴全体からのランダムな選出を選んだ．ランダムを選んだユーザは，思いがけないページが表示されていたのが良かったと答えていた．これによって，3.2 節でのアンケートでは見えづらかった

(注2)：Open Graph Protocol

図 13　HistoryPaper 実行結果例 1

が，ユーザは調べたかった内容のページ以外にも，ネットサーフィンでたどりついたページなども選出する必要が感じられた．これは今後の課題としていきたい．

7.　まとめと今後の課題

本論文では，1 日の閲覧履歴の中で特に重要であると判断した Web ページ群を抽出し，それらを新聞のようにレイアウトすることで，ユーザーの毎日の行動や獲得知識を要約表示するシステムを提案した．また Treemap によるレイアウトと提案手法によるレイアウトを比較することで，提案手法によるレイアウトの妥当性を検証した．なお，本報告で提案したレイアウトアルゴリズムは Github(注3) で公開している．

現段階の我々の研究では，レイアウトアルゴリズムに関して，レイアウトとしての美しさを評価できていないと考える．そこで美しさに関する評価基準を新しく定義した上で，レイアウトに関するユーザテストを実施したい．また，実際に運用されているマガジンスタイルのサイトに表示されている各長方形領域の面積を測定し，これをクラスタ重要度とみなして提案手法でレイアウトして評価する，という実験も行いたい．

文　　献

[1] 相澤清晴. ”1. ライフログの実践的活用: 食事ログからの展望 (< 特集 > ライフログ).” 情報処理 50.7, 2009. pp. 592-597.

[2] Hoeber, Orland, and Joshua Gorner. ”Browseline: 2d timeline visualization of web browsing histories.” IEEE 13th International Conference on Information Visualisation, 2009. pp. 156-161

[3] Hailpern, Joshua, et al. ”YouPivot: improving recall with contextual search.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011. pp. 1521-1530.

[4] Li, Ian, et al. ”Here’s what i did: sharing and reusing web activity with ActionShot.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010. pp. 723-732.

[5] Weinreich, Harald, et al. ”Not quite the average: An empirical study of Web use.” ACM Transactions on the Web (TWEB) 2.1, 2008. p. 5.

[6] Nebeling, Michael, Fabrice Matulic, and Moira C. Norrie. ”Metrics for the evaluation of news site content layout in large-screen contexts.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011. pp. 1511-1520.

[7] 長野翔一, 高橋寛幸, and 中川哲也. ”ユーザの要求変化に着目したウェブ閲覧履歴の分類方式 (分類).” 情報処理学会研究報告. 自然言語処理研究会報告 2008.90, 2008. pp. 65-70.

[8] 山口雄大, 新美礼彦, and 小西修. ”Web 検索効率改善のための Web 履歴の分類とグループ化.” The 23rd Annual Conference of the Japanese Society for Artificial Intelligence. 2009.

[9] Gonzlez, Jess, et al. ”Web newspaper layout optimization using simulated annealing.” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 32, No. 5, 2002. pp. 686-691

[10] Bruls, Mark, Kees Huizing, and Jarke J. Van Wijk. Squarified treemaps. Springer Vienna, 2000. pp. 33-42

[11] Edmundson, Harold P. ”New methods in automatic extracting.” Journal of the ACM (JACM) 16.2, 1969. pp. 264-285.

[12] Genest, Pierre-Etienne, et al. ”A symbolic summarizer for the update task of tac 2008.” Proceedings of the First Text Analysis Conference, Gaithersburg, Maryland, USA. National Institute of Standards and Technology. 2008.

[13] 松枝, 伊藤, HistoryPaper: ユーザー個人のブラウザ履歴を用いた毎日の可視化, ARG 第 4 回 Web インテリジェンスとインタラクション研究会, 2014.

(注3)：https://github.com/cocodrips/articlemap-js(JavaScript), https://github.com/cocodrips/magazine-style-layout(Python)