

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^{a)} 伊藤貴之^{b)}

概要: 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用活用することは簡単ではない。本報告では、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する。提案手法ではまず、閲覧履歴を構成するWeb ウェブ群を文書内容でクラスタリングし、続いて滞在時間、検索キーワード、アクセス貴重度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、最近のWeb サイトに多く採用されている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。さらに本報告では、レイアウト結果を数値評価することでレイアウトアルゴリズムの妥当性を検証した結果を示す。

CHICA MATSUEDA^{a)} TAKAYUKI ITOH^{b)}

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、ログを書き綴るといった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等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。

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。そして閲覧履歴を用いて自己を振り返ることで、特定の日に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、といった知識復元につながる事が期待される。

本報告では、1日の閲覧履歴を要約するWeb ページ群を一覧表示するシステムを提案する。本手法では、1日の閲覧履歴の中から重要と考えられるWeb ページ群を抽出し、我々が馴染み深い雑誌のようなスタイルでそれらをレイアウトすることで、ユーザの1日を1ページのスクラップブックのように表現する。

本手法に類似するサービスとして、Twitter や Facebook にシェアしたリンクから毎日のまとめをつくる Paper.li^{*1} がある。Paper.li ではマガジンスタイル (2 節で後述) を採用しており、ユーザが興味をもったリンクなどが一覧性の高いデザインでまとめられている。しかし Paper.li ではユーザの友人が公開したリンクを別の友人にもシェアすることを目的としており、ユーザ自身の閲覧履歴の要約表示を目指す本研究とは目的が異なる。

¹ お茶の水女子大学大学院
Ochanomizu University

^{a)} coco@itolab.is.ocha.ac.jp

^{b)} itot@itolab.is.ocha.ac.jp

^{*1} <http://paper.li/>

本報告は2節でマガジンスタイルの定義について、3節で閲覧履歴からの代表ページの選出方法について、そして4節でレイアウト手法について説明する。

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。マガジンスタイルは、その名の通り雑誌のようなレイアウトのことである。

図1のWebページのように、様々な大きさの四角形を多数画面上に敷き詰め、その中に文字や画像を当てはめてコンテンツを表示させるこのスタイルは、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、多くのコンテンツが詰まってもコンテンツの内容を理解しやすい。そのため現在多くのWebページでマガジンスタイルが使われている。



図1 マガジンスタイルを適用しているサイト例 *2

マガジンスタイルは一種の俗称であり、それがどのようなレイアウトであるかの明確な定義は我々の知る限り見当たらない。そこで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、我々は自身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生成方針を以下のように定義した。

定義1: 画面全体を2~4列程度の大きな縦長の領域に分割し、さらに各領域を長方形領域に分割し、各Webページを各長方形領域に埋め込む。

定義2: 各長方形領域にはWebページの重要度に応じた面積を割り当て、さらに画面の左上には重要度の高いWebページを優先的に割り当てる。

定義3: 各長方形領域のアスペクト比に統一感をもたせる。

定義4: 同じ大きさの長方形領域を複数隣り合わせて配置

することで、整って見えるレイアウトを実現する。

これらのうち定義3について検証するために我々は、マガジンスタイルを採用しているニュースサイト *2 にあるの計236個の記事の大きさとそのアスペクト比を測定した。図2はその分布を示した図である。

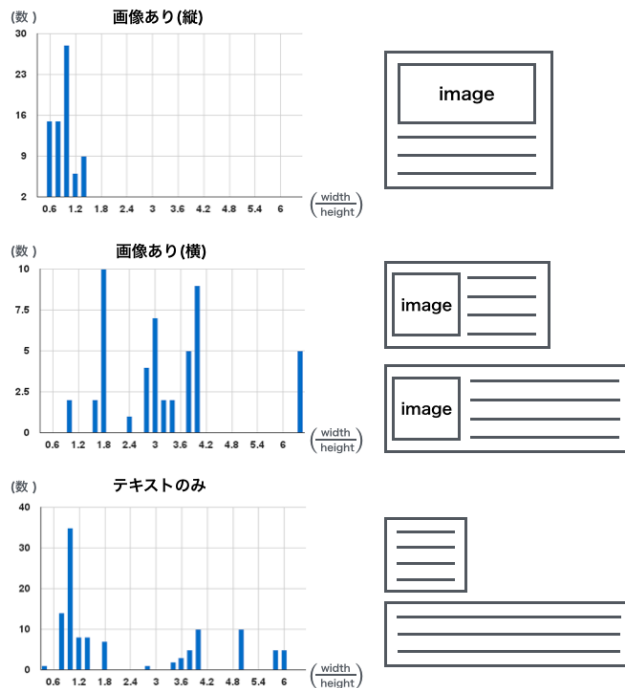


図2 各記事のスペースのアスペクト比の分布

これらをふまえて、本研究では各長方形領域の理想的なアスペクト比を表2のように定義する。なお表2に示した数値のうち、上のものほど優先度が高いとする。

表示タイプ	横 ÷ 縦 (画像位置)
画像有	0.9 (上)
	1.6 (左)
	3.0 (左)
	3.8 (左)
画像無	1.0
	3.8
	5.0

表1 各長方形領域の理想的なアスペクト比

また定義4についても調査した。その結果、マガジンスタイルによってデザインされた現実のニュースサイトを構成する長方形領域のうち95%が、同じサイズの長方形を2個以上並べて配置したものであることがわかった。

*2 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/>)

3. 提案手法

本手法では、ユーザの特定の1日の閲覧履歴の中から代表的な n 個の Web ページを抽出し、そのユーザの1日をニュースサイトのようなページデザインで表示する。なお我々の実装では、その日ユーザが訪れた Web ページ数に応じて、 n の値を6~12程度の範囲であらかじめ調節する。

3.1 代表 Web ページ群の選出

本手法ではまず、1日の履歴を構成する各 Web ページの重要度を計算し、その結果から履歴を要約するための代表 Web ページ群を選出する。ここで文書群の要約の目的で代表文書を選出する手法として、tf-idf 法によって測った重要度の高い文書を重複を避けて抽出することで、満足度の高い結果が得られる傾向があることが知られている [5]。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。以下に代表 Web ページ群の選出手法を示す。

3.1.1 閲覧履歴のクラスタリング

本手法では、1日の履歴にある全ての Web ページの内容を対象として、以下の文書クラスタリング手法を用いて n 個のクラスタに分類する。

- (1) Web ページのコンテンツ内容を Bag-of-Words 表現に変換する。
- (2) 潜在的意味解析 (LSA) を用いて Bag-of-Words を次元削減する。
- (3) 潜在的意味空間に位置する各 Web ページを K-means 法でクラスタリングする。

3.1.2 各ページの重要度計算

続いて本手法では、3.1.1 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。

- m : その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 : アクセスの貴重度 (1日にその Web ページアクセスした回数と今までのアクセス数で割った商と定義する)

現時点での我々の実装では、各 Web ページの重要度 p_p を以下の式で定義している。ここで p_a , p_b , p_c は経験的に決定される定数である。

$$P_p = p_a q + p_b m + p_c m q \quad (1)$$

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。

本手法では各クラスタを構成する Web ページ群に式 (1) を適用し、その値が最大である Web ページをクラスタの代表として選出する。

3.1.3 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。

$$c_p = \sum_{k=1}^{n_c} p_p(k) \quad (2)$$

ここで n_c は当該クラスタに属する Web ページ数、 $p_p(k)$ 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。

3.2 レイアウトアルゴリズム

[1] のように、Web のレイアウトを焼きなまし法などの最適化手法を用いたレイアウト手法はいくつか発表されているが、マガジンスタイルに特化した手法は我々が調べた限りでは見当たらない。本手法では、3.1 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^{*1} のようなデザインで Web ブラウザの一面面に配置する。

3.2.1 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 Squarified Treemap[3] 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3である。Squarified Treemap のレイアウト生成には d3.js を用いている。



図3 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、

^{*1} 例: The New York Times <http://www.nytimes.com/>

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。

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。

3.2.2 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 (3.1.3 節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, \dots 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
 - $ratios_S$: 集合 S 全体の理想的なアスペクト比群 (表 2 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} : 各クラスタが配置される長方形領域の最小横幅 (定数)
 - H_{min} : 各クラスタが配置される長方形領域の最小縦幅 (定数)

3.2.3 クラスタのグループ化

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。

$$C_{pj} \leq [(C_{pi} \times 1.2)] \quad (3)$$

(1)

集合	クラスタ重要度	画像
R_1	45	○
R_2	32	○
R_3	30	○
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	○
G_2	32 30	○
G_3	20	
G_4	11 10	
G_5	5	
G_6	2 1	

図 4 クラスタのグループ化

例として図 4 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, \dots, 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 章に示した定義 4 を満たすことができる。

3.2.4 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トした値を、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、3.2.1 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。

$$area_{R_i} = priority_{R_i} \frac{WH}{W_{min}H_{min}priority_R} \quad (4)$$

図 4(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 5 に示す。

ID	クラスタ重要度	面積占有度	画像
G_1	45	27	○
G_2	32 30	20 18	○
G_3	20	12	
G_4	11 10	7 6	
G_5	5	3	
G_6	2 1	2 1	

図 5 面積占有度の算出

3.2.5 Web ページの配置

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 3.1 節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6に示す。

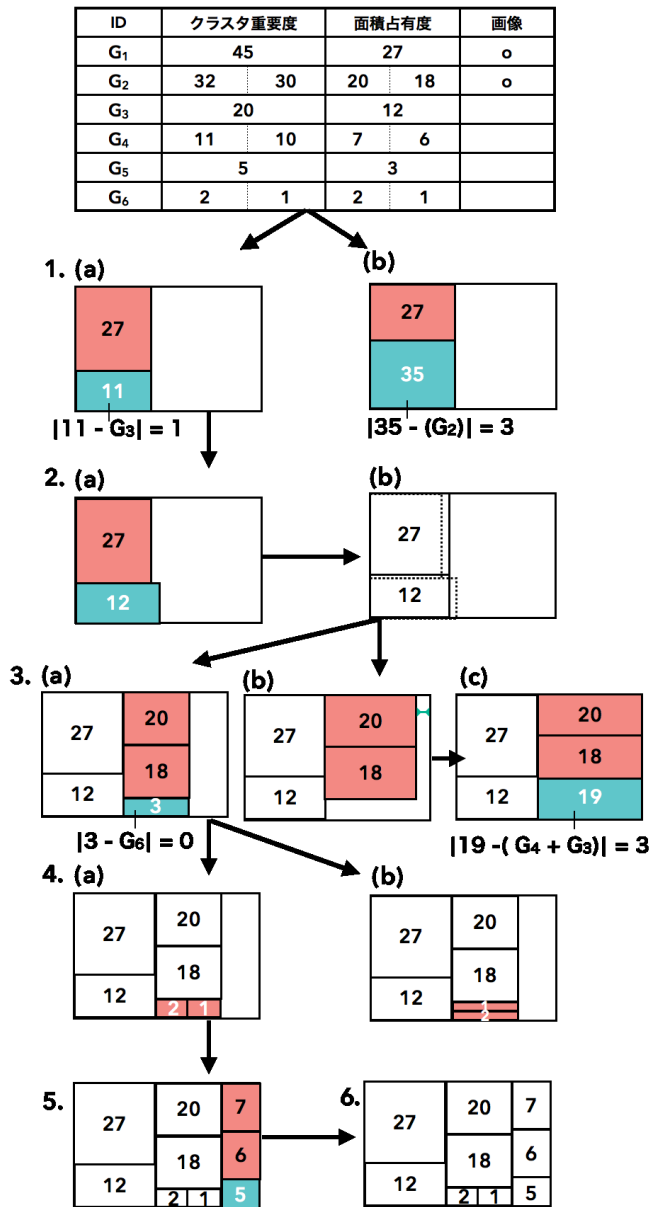


図6 配置アルゴリズム

- (1) $S = G$ とする. $rect_W = W$, $rect_H = H$ とする.
- (2) S 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6(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} 以下である場合には, 図6(3.(b)(c))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6(1.(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6(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6(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.
- (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を, これを合わせるように長方形領域群を変形する. 図6(2.(a))が変形前の例, 図6(2.(b))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6(3.(a))の水色の領域を図6(4.(a)(b))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6(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6(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) (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.
- 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を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.

3.3 代表ページの配置

最後に、3.1 節に示した手法で選出した代表ページを、3.2 節に示した手法で生成した長方形領域に配置する。ここでそれぞれの長方形領域には、Web ページのタイトル・URL・コンテンツの要約を表示する。コンテンツの要約には、本文冒頭の数行を抽出することで文章の要約に用いるリード手法 [6][7] を適用し、本文冒頭の 300 字程度を抽出して用いている。

4. 結果と考察

4.1 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 7 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 4.1 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 3.1.3 節で述べたクラスタの重要度である。



図 7 検索キーワード例

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

表 2 代表ページの選出結果

4.2 レイアウトの実行結果

図 3 に示した配置実験にも適用した図 4 のデータを用いて、提案手法を実行した例を図 8 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。

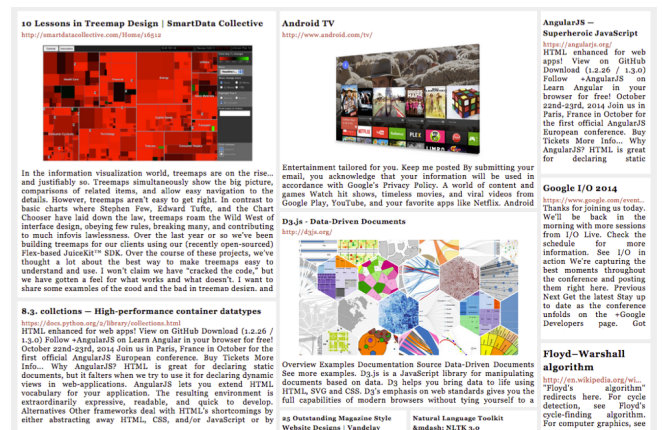


図 8 提案手法の実行結果

4.3 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 2 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。

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

ただし

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

(2) 理想のアスペクト比から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。

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

いずれも値が 0 に近いほど評価が良いとする。

4.4 他のレイアウト手法との比較

3.2 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。

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 9 に示す。

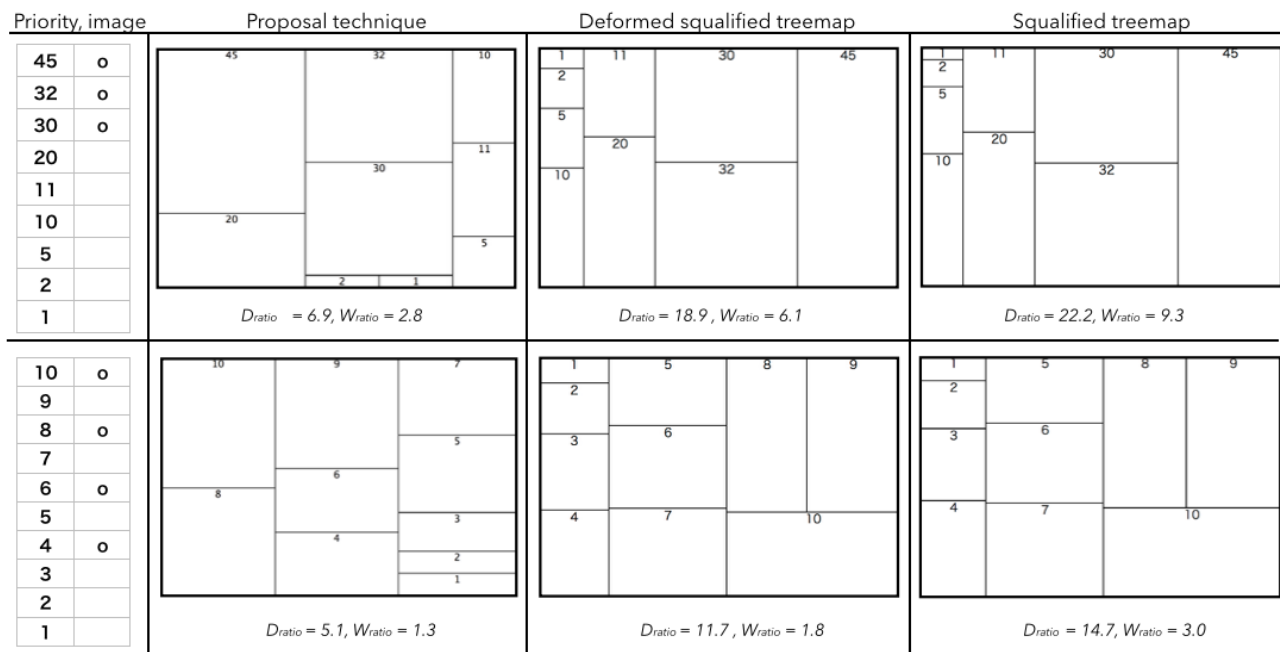


図9 アルゴリズムの比較

Squarified Treemap は重要度のデフォルメを適用しない限り、長方形領域の面積は重要度に忠実に比例する。しかし、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。重要度のデフォルメを適用した Squarified Treemap は、極端に小さい長方形領域がなり、 D_{ratio} , W_{ratio} の評価がともに改善されている。本手法は、重要度のデフォルメを適用した Squarified Treemap よりも、さらに D_{ratio} , W_{ratio} の評価がよくなっていることが見て取れる。

4.5 考察

4.5.1 代表 Web ページの選出に関する考察

4.1 節で示した代表 Web ページの選出結果には、筆者がその日特に何度も検索していた「Android」「長方形 packing」「guillotine cutting」に関する Web ページが重要度の高いクラスタの代表として適切に選ばれている。また、それらの検索の合間に検索した音楽に関するクラスタも選出されていることも見て取れる。しかし、現時点では筆者自身の閲覧履歴からの代表 Web ページ選出結果について考察したのみで、第三者による評価実験はまだ実施していないため、今後の課題としたい。

4.5.2 レイアウトに関する考察

与えられた重要度を忠実に保つことで非常に小さな面積の長方形領域を生成した場合、Squarified Treemap においても本手法においても細長い長方形が生成され、 D_{ratio} や W_{ratio} が悪化する可能性がある。一方で、本手法ではレイアウト処理において代表 Web ページに対して重要度に比

例した面積を与えるのではなく、多少のデフォルメを適用することで長方形領域における掲載内容の読みやすさを高めている。しかし、あまりにも大きく重要度をデフォルメすると、どれが重要なページであるかを視認しにくくなる可能性がある。よって、どこまでデフォルメすることが許容できるか、実験と評価を行う必要がある。

一方で我々は、レイアウトの美的完成度も重要な評価基準であると考え。しかし我々が本報告で設定した D_{ratio} , W_{ratio} の2つの評価値のみでは、レイアウトの美しさやユーザにとっての記事の読みやすさは評価できていないと考える。そこで今後の課題として、長方形分割に基づくレイアウトに関連する先行研究を参考にして、別の評価値を定義した上で評価と考察を重ねたい。

5. まとめ・今後の課題

本報告では、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^{*4} で公開している。

现阶段の我々の研究では、代表 Web ページの選出結果に関する評価実験が充分でない。そこで今後の課題として、まず代表 Web ページの選出方法に関するユーザテストを実施したい。またレイアウトに関しては、面積やアスペク

*4 <https://github.com/cocodrrips/MagazineStyleLayoutAlgorithm>

ト比だけでなく、他の観点からもレイアウト結果の評価を進めたい。特に現時点では、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。

参考文献

- [1] Gonzalez, J., Rojas, I., Pomares, H., Salmern, M., Merelo, J. J. Web newspaper layout optimization using simulated annealing. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 32(5), 686-691 (2002).
- [2] Genest, P. E., Lapalme, G., Nerima, L., Wehrli, E., du Langage, T., A symbolic summarizer for the update task of tac 2008, *The First Text Analysis Conference* (2008).
- [3] Bruls, M., Huizing, K., Van Wijk, J. J., Squarified treemaps, *Data Visualization 2000*, 33-42 (2000).
- [4] Wartenberg, C., Holmqvist, K. Daily Newspaper Layout-Designers' Predictions of Readers' Visual Behaviour-A Case Study. *Lund University Cognitive Studies*, 126, 1101-8453(2005).
- [5] Lin, C. Y., Hovy, E. Identifying topics by position. In: *Proceedings of the fifth conference on Applied natural language processing*. Association for Computational Linguistics, 283-290 (1997).
- [6] Edmundson, H. P. New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2), 264-285 (1969).
- [7] Litvak, M., Last, M., Friedman, M. A new approach to improving multilingual summarization using a genetic algorithm. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, p. 927-936 (2010).
- [8] 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).