

ホームページの関心度指標による構造設計の提案

山縣 修[†] 片岡 信弘^{††} 森末 道忠[†]

[†]広島市立大学情報科学研究科 ^{††}東海大学電子情報学部情報メディア工学科

関心度指標はページ単位に閲覧者の関心を規定するキーワード(語)の出現頻度と情報選択の可能性を示すリンク数(anchor タグ数)で構成される。被験者13人にアンケート調査により、この指標とAHP(階層分析法)によるホームページの満足度の重要度に相関がある事を検証し、提案した関心度指標は閲覧者の情報選択に有用である事を確認した。キーワードのページ単位の度数分布とページ間の遷移行列を対応分析することで、ホームページの構造の設計を支援する手法を提案する。

Proposal for skeleton design by interest-metrics about homepage

OSAMU YAMAGATA[†] NOBUHIRO KATAOKA^{††} MITITADA MORISUE[†]

[†]Faculty of Information Sciences,Hiroshima City University

E-mail: yamagata@imc.im.hiroshima-cu.ac.jp , :morisue@im.hiroshima-cu.ac.jp

^{††}School of Information Technology and Electronics Tokai University

E-mail: kataoka9@keyaki.cc.u-tokai.ac.jp

ABSTRACT.The interest-metrics consists of two elements: the total number of occurrences of the keywords that define the user's interest in each page; and the number of links (i.e., the number of anchor tags) that represents the range of information selection possibility. The analysis of 13 subjects has revealed that there is a correlation between the interest-metrics and the degree to which the homepage satisfies, which has been obtained by the AHP (Analytic Hierarchy Process). Thus, we have verified that the interest-metrics proposed in this paper is useful. Finally we study in skeleton design by using correspondence analysis.

1. はじめに

最近、WWW は数億ものホームページで構成される巨大な情報空間であるが、ホームページの文書構造の非線形性がホームページの閲覧者と作成者の双方に、『迷子問題(disorientation problem)』¹⁾を誘発している。そのため、Web 空間の情報選択を支援する多くの研究あり、Web 空間の視覚化による支援する「HyperSpace」²⁾、閲覧履歴を機械学習し適当なページを推薦する「Web Watcher」⁴⁾あるいはハイパーメディアのオーサリングシステム⁵⁾の研究がある。

本研究は、ホームページから情報選択する指標として“関心度指標”を提案し、この指標によるホームペー

ジの構造設計の可能性を考察する。

関心度指標は、閲覧者がホームページから情報選択する過程で、ページ単位にその関心が満される度合である。被験者13人からアンケート調査し、関心度指標列とAHP⁶⁾によるホームページの満足度の重要度には相関があり、提案する関心度指標がホームページの閲覧者に有効な指標であることが確認できた。

これにより、関心度指標は閲覧者の関心が満される度合を示しているため、高い関心度指標のページからなる関心度指標列の組を重畳しているホームページは利用者の満足を得られると考えられる。そこで、ページ毎の語の分布をクロス表、ページ間のリンク構造を制約

条件として対応分析することで、ホームページの構造を設計する手法を提案する。

以下、本論文では、2章で“関心度指標”を提案し、3章で関心度指標の有用性を検証し、4章でホームページの構造設計を提案する。

2. “関心度指標”の提案

ホームページを構成する要素、“関心度指標”と関心度指標列の定義について述べる。

2.1 ホームページを構成する要素

ホームページを構成するページの情報の種類は大野⁷⁾が述べているように、メニュー、文章、ラベル(文字列)、フォーム、ロゴ・飾り文字、ビットマップ図および他のページを関係付けるリンク(アンカー)の7種類に分類することが出来る。

また、ページの情報は HTML(HyperText Markup Language)で記述され、IMGタグの alt 属性によりその図柄を表現する代替テキストを記述できる。同様に、リンクは A タグの alt 属性によりリンクされた情報の代わりに表示する代替テキストを記述できる。

2.2 関心度指標と関心度指標列の定義

関心度指標の定義、関心度指標列の定義と関心度指標列の特徴量の定義について述べる。

(1) 関心度指標の定義

テキストは、alt 属性を正しく使用していれば、ページを構成する各種類から情報を共通に抽出することができるので、閲覧者が関心を示すページはその関心を規定するキーワードの出現頻度が重要な要素である。また、ホームページの作成者はリンク(A タグ)により、情報の単位であるページの集合間を関係付けているので、リンクの数の大小は情報選択の可能性の大小と同時に、情報の有益性も表していると考えられる。一方、閲覧者はこのリンクを選択することにより、ホームページから情報を選択する。このように、ページを記述した HTML のコードが含むテキスト情報とリンク情報はページの特徴を示すと考えられる。

提案する関心度指標は、情報選択の単位であるページごとに、テキスト情報に対応する関心を規定するキーワードの出現数とリンク情報であるリンク(選択肢)数、ページごとの文字総数で構成する。この時、ページを構成する文字数が多ければキーワードの出現度数も大

きくなるので、文字総数で正規化し、リンク数は Top-Page, Next, Up, Previous の典型的なリンクを除いたリンクと各ページに共通したリンクを除く。さらに、演算結果が極端に小さい数あるいは極端に大きい数にならないように4つの定数を定める。

関心度指標 I は各キーワードの出現頻度数 k 、ページを構成する文字総数 m とリンク(A タグ)数 l の3変数と、演算結果が0あるいは0による除算を防ぐための定数 C_1, C_2, C_3 および演算結果を整数化(1以上に)するための比例定数 C_4 の4定数で構成する。以上で述べた関心度指標の定義式を式1に示す。

$$I = C_4 \times \frac{\sum_{i=1}^n k_i + C_1}{m + C_2} \times (l + C_3) \cdots \text{(式1)}$$

この時、 k の添字 i は閲覧者が規定したキーワードの数を表す。

(2) 関心度指標列の定義

関心度指標は閲覧者がその関心に沿って情報選択できるような1ページ単位の指標であり、ホームページの構造に則した関心度指標の分布がわかれば、先々の予測ができる。そのため、「広り」(横)方向を優先して木構造を展開する方法により、ページ集合を1次元にマッピングする。

(3) 関心度指標列の特徴量の定義

関心度指標列は閲覧者の関心により様々な分布を示す。そのため、分布の特徴を共通的に評価する関心度指標列の特徴量を以下のように定義する。

関心度指標列の特徴量 J は関心度指標列を近似する三角形の頂点の y 座標値 h と両底角 θ_1, θ_2 の \tan 値の総和とする。関心度指標列の特徴量の定義式を式2に示す。

$$J = h + \tan \theta_1 + \tan \theta_2 \cdots \text{(式2)}$$

・頂点は y 座標値を関心度指標列の最大値、 x 座標値をそのページ位置とする。

・端点は y 座標値を関心度指標列の最小値、 x 座標値を0とページ位置の最大値の2点とする。

3. 関心度指標とホームページの満足度の解析

アンケート調査、関心度指標列の算出、ホームページの満足度および関心度指標列とホームページの満足度の解析について述べる。

3.1 アンケート調査

アンケート調査は常時インターネットを使用している学生 13 名を被験者として、2001年4月上旬に、アンケート調査した。ホームページは消費材の購買における情報収集を想定し、デジタルテレビと太陽光発電の2種類について、製品シェア上位の2社、合計4社を選択した。

3.2 関心度指標列の特徴量とホームページの満足度の算出

関心度指標列の特徴量とホームページの満足度の重要度の算出について述べる。

(1) 関心度指標列の特徴量の算出

関心度指標は以下の前提により算出する。

- ・キーワードは漢字またはカタカナの単語で中頻度語である動詞および名詞とし、8語から 10 語の範囲で規定する。
- ・式1の m は全角の漢字、カタカナを対象とする。
- ・式1の k と m は HTML のコードファイル(拡張子.html または.htm)から数え上げる。
- ・式1の定数は $C_1=C_2=C_3=0.1$ 、 $C_4=200$ として算出する。

関心度指標の特徴量は式2により算出する。つぎに、人間の短期記憶の量とされている数のひとつである8を区間とした移動平均後、式2により算出する。

(2) ホームページの満足度の重要度の算出

ホームページの満足度の決定問題を AHP の階層構造により図1に示す。

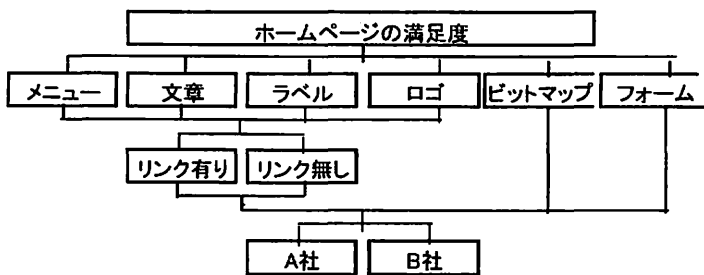


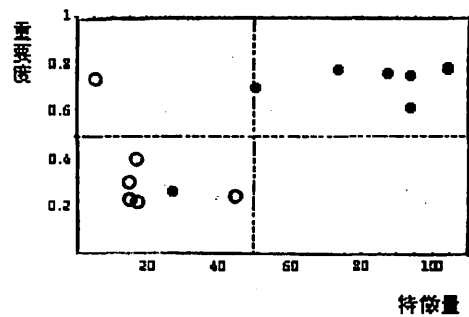
図1 ホームページの満足度決定問題の構造図

図1の最上段は“問題”であり、中の2段は先に述べたホームページを構成する要素に基づく“評価基準”であり、最後の段は“代替案”である。

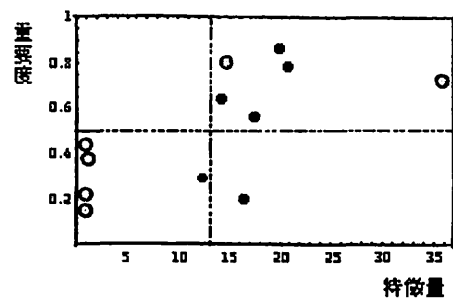
3.3 関心度指標列とホームページの満足度の解析

ホームページの満足度の重要度と関心度指標列の特徴量の相関を図2に示す。図2の(1)が三角形近似、

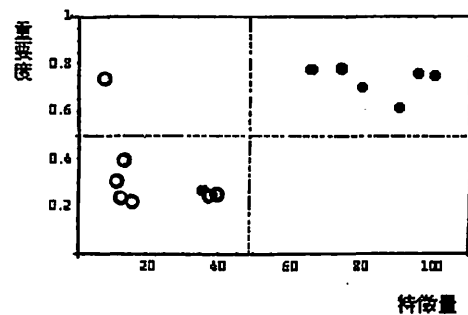
図2の(2)が移動平均後に三角形近似による相関図であり、それぞれの(a)はデジタルテレビの2社に対する被験者7人による相関であり、(b)は太陽光発電の2社に対する被験者6人による相関である。



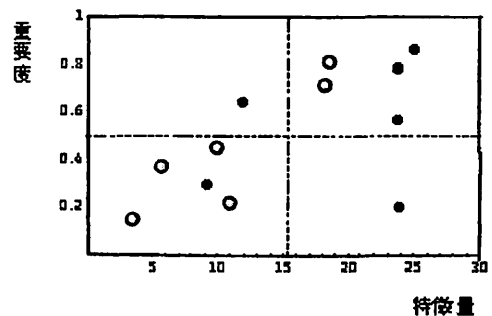
(1)-(a) デジタルテレビ



(1)-(b) 太陽光発電



(2)-(a) デジタルテレビ(移動平均後)



(2)-(b) 太陽光発電(移動平均後)

図2 ホームページの満足度と関心度指標列の特徴量の相関

図2の(1)が三角形近似、図2の(2)が移動平均後に三角形近似による相関図であり、それぞれの(a)はデジタルテレビの2社に対する被験者7人による相関であり、(b)は太陽光発電の2社に対する被験者6人による相関である。各図とも縦軸はホームページの満足度の重要度、横軸は関心度指標列の特徴量である。2つの破線は平均値を原点とするx軸y軸である。黒円はA社とC社、円はB社とD社を示す。

3.4 関心度指標の有用性の考察

図2による近似法ごとにホームページの満足度との相関と関心度指標列の有用性を述べる。

(1) ホームページの満足度との相関

デジタルテレビのホームページの満足度の重要度と関心度指標列の特徴量の相関において以下の事が認められる。

- ・関心度指標列の特徴量の平均とホームページの満足度の重要度の平均を原点とする座標系の第1象限、第3象限に分布する点の数は14点中13点である。近似法による差異はなく、いずれも正の相関が視認される。

同様に、太陽光のホームページの満足度と関心度指標列の特徴量の相関において以下の事が認められる。

- ・関心度指標列の特徴量の平均とホームページの満足度の重要度の平均を原点とする座標系の第1象限と第3象限に分布する点の数は12点中11点から10点であり、近似法による差異はあるが、いずれも正の相関が視認される。

(2) 関心度指標列の有用性の考察

図2より、関心度指標列の特徴量の近似法によらず、関心度指標列の特徴量とホームページの満足度の重要度に相関があると考えられる。その結果以下の2点可言える。

- ・関心度指標列の特徴量の考え方から、関心度指標列の特徴量は関心度指標に比例するので、関心度指標の大小とページの満足度の大小にも正の相関が有る。
- ・閲覧者が関心度指標の大きいページを優先して情報選択する事は効率的な情報選択法の一つである。

4. ホームページの構造設計

関心度指標列の重畳による設計と対応分析による構造設計の評価について述べる。

4.1 語の分布の重畳による設計

利用者がホームページ中で連続して閲覧するページ数とページ群の重畳について述べる。

(1) 閲覧ページ数

Schneiderman はハイパーテキストの黄金律⁸⁾の中で“ユーザが必要とするのは何時でも断片の一部である。”としているが、2製品のホームページの1ヶ月間のWeb アクセスログを分析して得た連続して閲覧するページ数を示した表1より、その95%が10ページ未満であることが確認できる。

表1 連続して閲覧するページ数

ページ数	累積閲覧数				
	製品 A	製品 B	製品 C	製品 D	製品 E
1	64.0%	81.8%	51.0%	52.9%	71.0%
2	73.5%	89.8%	68.6%	65.7%	80.3%
3	80.8%	93.3%	77.8%	72.5%	84.5%
4	87.4%	95.2%	83.9%	79.4%	86.7%
5	91.7%	96.4%	87.6%	84.3%	88.9%
6	94.5%	97.3%	90.9%	88.7%	91.5%
7	96.3%	97.9%	93.5%	92.6%	92.9%
8	97.9%	98.5%	94.7%	95.1%	94.1%
9	98.7%	98.7%	95.8%	96.1%	95.0%
10	99.3%	99.0%	97.0%	98.0%	96.0%
11	99.4%	99.1%	97.8%	98.5%	96.7%
12	99.5%	99.2%	98.4%	99.5%	97.3%

(2) ページの重畳

関心度指標列の定義と Schneiderman の3つの黄金律より、関心度指標の高い数ページで構成するページ群を重畳させたホームページは閲覧者の満足度が高いと考えられる。しかし、ページ群を重畳する過程で、関心度指標列の分布形状が様々に変化し、ページ間の関連が設計者の意図した関連を維持しているとは限らない。一方、関心度指標列はその定義よりページ間のリンク構造を維持していないので、ページ間の関連は関心度指標の高・低しか表現できず、ホームページの設計者には情報不足である。そこで、ホームページの構造をテキスト構造を表すページ毎の語の分布とリンク構造を表すページ間の連結関係よりページ間の関連を2次元に可視化することで、ホームページの構造の設計を支援する手法を考察する。

4.2 対応分析によるホームページの構造の可視化

ホームページの構造の定義式とホームページの構造

の可視化の手法を述べる。

(1) ホームページの構造の定義

先の項で述べたホームページの構造の定義式を式3に示す。

$$[S] = [T] \times [L] \dots (\text{式3})$$

この時、Sはホームページの構造の定義行列、Tは行を語の種類、列をページとする語の分布行列(クロス表)、Lは行、列ともページとするリンク行列である。

(2) ホームページの構造の可視化

ホームページの構造の定義行列より、語の分布行列は目的変数が無く、カテゴリーによる分割表であり、リンク構造を制約条件と考えられる。そのため、ホームページの構造の可視化(分割表の数量化)はデータ解析手法として制約付き対応分析法(あるいは双対尺度法)を選択する。

4.3 ホームページの構造の可視化の事例

Rで記述された対応分析法のパッケージソフトウェア(CoCoAn)⁹⁾により、模式的な事例とデジタルテレビの事例、太陽光発電の事例を示す。

(1) 模式的な事例

標準偏差を 0.1 とし、平均を単調に移動させた正規分布による 15 組と一様乱数の 2 組で構成する模式的な分割表を対応分析した結果を図3に示す

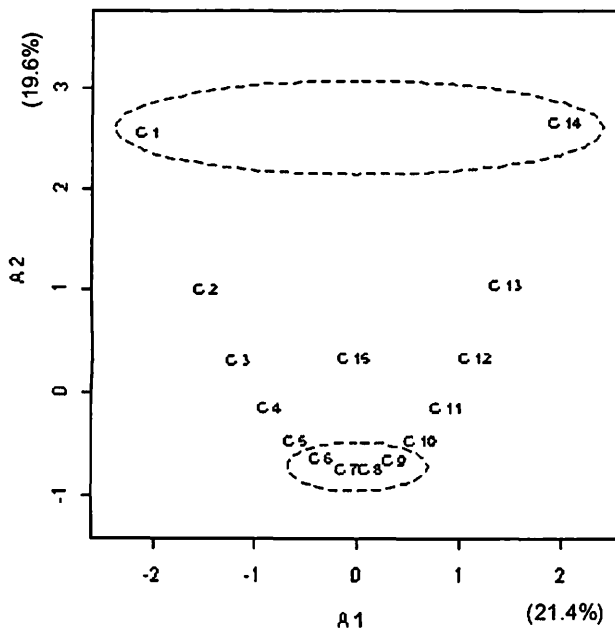


図3 模式的な事例

図3の横軸は第1軸で寄与率は 21.4%で、縦軸は第2軸で寄与率は 19.6%である。文字c1からc15 は列の位置を表す。この図からc1とc14は距離が離れ、c6、c7、c8とc9は距離が接近しているが、分割表を作成した正規分布の平均の大小に対応している。このことは、{c1, c14}と{c6, c7, c8, c9}の組の比較においても同様である。

(2) デジタルテレビの事例

図4はデジタルテレビのホームページの解析結果を示す。図4の横軸は第1軸で寄与率は28.1%で、縦軸は第2軸で寄与率は21.9%である。文字c1からc13は列の位置を表す。この図から以下の事が推察される。

- ・{c1,c2,c3,c7} はトップページ、製品一覧、機能一覧および製品一覧のページであり、話題の広がり(網羅性)を示している。
- ・{c4,c5,c6} はデジタル型の製品説明、{c10,c11} は製品共通の機能詳細の説明、{c12,c13}はアナログ型の製品説明であり、話題の深さ(詳細性)を示している。

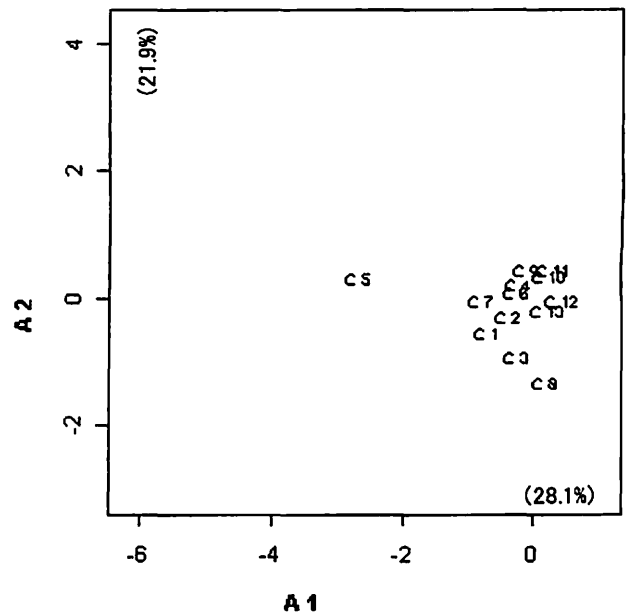


図 4 デジタルテレビの事例

(3) 太陽光発電の事例

図5に太陽光発電のホームページの解析結果を示す。図5の横軸は第1軸で寄与率は30.9%で、縦軸は第2軸で寄与率は26.6%である。文字c1からc11は列の位置を表す。これらの図から以下の事が推察される。

- ・ {c1,c5,c7,c8,c9}はトップページ、太陽光発電の説明一

覧、太陽光発電の目的、電力会社への売電および太陽光発電の仕組みの説明であり、話題の深さ(詳細性)を示している。

・ {c2,c3,c4,c10}は共通事項の一覧、共通事項の説明、アンケート(フォーム形式ページ)および会員募集の説明であり、直接の因果は無い。

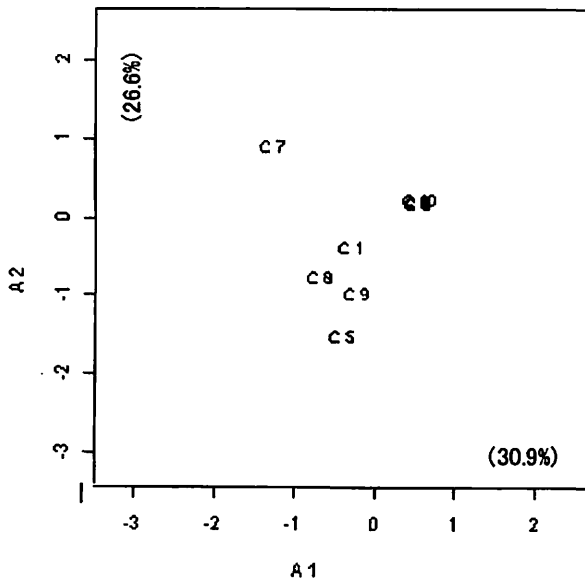


図5 太陽光発電の事例

4.4 考察

今回はホームページの構造を語の分布だけで記述し、対応分析することで、ページ間の関連としてページの遠近および階層構造の縦軸として網羅性、横軸とし

参考文献

- 1) Jeff Conklin :Hypertext: An Introduction and Survey. IEEE Computer 20 September 1987,pp.17-41.
- 2) Andrew Wood,Nick Drew,Russell Beale,Bob Hendl:Hyper Space:Web.browsing.with.Visualisation.http://www.cs.bham.ac.uk/~amw/hyperspace/www95/.1995
- 3) Thorsten Joachims,dayne Freitag,Tom Mitchell : Web Watcher: A Tour Guide for the World Wide Web. http://almond.srv.cs.cmu.edu/afs/cs.cmu.edu/project/theo-6/Web-agent/www/project, 1995.
- 4)http://www.fujixerox.co.jp/tandd/Tr96/Takeshi_Shimizu/

て詳細性を推測する事が出来た。

今後、デジタルテレビと太陽光発電の事例で、リンク構造を制約条件として対応分析する必要があるが、現時点で提案した手法はホームページの構造設計を支援できると考えられる。

5. おわりに

本論文は、ホームページから情報を得ようとする時に、閲覧者とその関心に沿って情報選択することが出来る指標として関心度指標を提案した。つぎに、消費材の購買行動を想定したアンケート調査により、関心度指標列を算出し、ホームページの満足度を AHP により評価した。この結果、関心度指標は閲覧者のホームページの情報選択に有用である事が確認できた。最後に、ホームページの作成のために、対応分析によりページ間の遠近を評価しながらホームページの構造の設計を支援する手法の可能性を考察した。

本論文で提案した関心度指標は閲覧者とその関心に沿って情報選択することが出来る指標であり、この指標の考えに基づいたホームページの構造の設計を支援する手法は、その一つであることを確認できた。

今後の課題は制約条件つき対応分析によりリンク構造の変化とページの遠近の変化を評価することである。

T_Shi101J.html

- 5) Salton,Buckley:Term Weigthing Approaches In Automatic Text retrieval. Information Processing & Management,Vol.24(5),Page515-523.
- 6) 刀根薫 :ゲーム感覚意思決定法.日科技連出版社,1998
- 7) 大野健彦 :Web 画面における情報選択と視線の関係 電子情報通信学会技術報告 HIP2000-11 Page31-35,2000
- 8)http://mail.gnome.org/archives/gnome-gui-list/1998-November/msg00074.html
- 9) http://cran.r-project.org/