

D-05

Web アクセシビリティ問題点の共起性分析

Associations Analysis on Webpage Accessibility Problems

仙波 元樹 田中 悠輝 岡田 英彦
Motoki Semba Yuki Tanaka Hidehiko Okada

1. はじめに

Web の普及に伴い、Web アクセシビリティの重要性がますます高まっており、設計ガイドラインの策定[1]、規格化[2]、検査手法・自動化ツールの開発[3,4]などが活発に進められている。例えば、JIS X 8341-3[2]に記載のガイドラインへの適合性を検査可能なツールを用いれば、検査対象の Web ページにおいてガイドラインに適合しない箇所（問題点）を発見することができる。

本研究では、このような問題点の「共起性」を分析する。ここで問題点の共起とは、異なる種類の問題点（ガイドライン A に適合しない問題点と、A とは異なる別のガイドライン B に適合しない問題点）が、同じ検査対象ページ上で発生することを指しており、共起性とはその共起の程度のことを指している。ある特定のガイドラインの組合せについて問題点が共起しやすいことがわかれば、その原因を調べることにより、問題点の少ないページを作成する上で注意すべきことがわかると期待される。本研究では、JIS X 8341-3 の 5 章に記載のガイドラインに関して、問題点の共起性を分析する。

2. 分析方法

2.1 分析に用いるデータ

問題点の共起性を分析するために、まず、問題点の事例データを集めた。この事例データは、「Web ページ W_i に、ガイドライン G_j に関する問題点が P_{ij} 個含まれている」というデータであり、 W_i を行、 G_j を列、 P_{ij} の値を i 行 j 列のセルの値とすれば、行列形式のデータとなる。 W_i としては既存の Web サイトのトップページを用いることとし、「Yahoo!カテゴリ」に掲載されたカテゴリ 14 種類のそれぞれについてランダムに 100 サイトをサンプリングした。これら 1400 サイトのトップページについて、 P_{ij} の値を求めるために、自動検査ツール WebInspector 5.11[6]を利用した。本ツールは、JIS X 8341-3 の 5 章に記載のガイドライン 39 項目中、22 項目の診断が可能とされており、本研究で得たデータ P_{ij} の j の値は 1~22 である。つまり、今回集めたデータは 1400 行 22 列の行列形式のデータであり、行列の要素の値は 0 以上の整数である。

2.2 問題点共起の調べ方

問題点の共起性を分析する方法として、本研究ではアソシエーション分析を用い、上記の事例データからアソシエーションルールを抽出することとした。得られるルールは、

$$X \rightarrow Y$$

の If-Then 形式をしており、 X は $\{G_j\}$ の部分集合の and 結

合によって表現される条件部、 Y は G_j のいずれか 1 つで表現される結論部である。

今回は、共起性分析の初期的な調査として、問題点の「数の大きさ」は考慮せず、問題点の「有無」に注目することにした。つまり、「どのようなガイドラインの問題点を含んでいるページは、別のどのようなガイドラインの問題点も合わせて含んでいることが多いか」を調べることとした。そこで、アソシエーション分析への入力となるデータは、元の $\{P_{ij}\}$ ではなく、以下のルールで $\{P_{ij}\}$ を 2 値化した $\{Q_{ij}\}$ とした。

$$\text{If } (P_{ij} = 0) \text{ then } Q_{ij} = 0, \text{ else } Q_{ij} = 1.$$

つまり、ページ W_i がガイドライン G_j に関する問題点をまったく含んでいなければ「問題点なし」として $Q_{ij}=0$ 、1 つ以上含んでいれば「問題点あり」として $Q_{ij}=1$ とした。

$\{Q_{ij}\}$ のデータからアソシエーションルールを抽出する手段として、本研究では、統計解析ソフト「R」を利用した[7]。

3. 分析結果

2 章に記載の方法で抽出されたアソシエーションルールのうち、支持度 (support) の値が高い上位 10 件のルールを表 1、確信度 (confidence) の値が高い上位 10 件のルールを表 2、リフト (lift) の値が高い上位 10 件のルールを表 3 にそれぞれ示す。例えば、表 1 の No.1 は、

If (5.4b の問題点がある) Then (5.4a の問題点がある)

というルールを、表 1 の No.2 は、

If (5.4b の問題点がある) and (5.9a の問題点がある)
Then (5.4a の問題点がある)

というルールを表している。表 1~表 3 において、 S 、 C 、 L の列の値はそれぞれ、同じ行が表すルールの支持度、確信度、リフトの値である。

これらの表からわかることとして、まず、ルールの結論部に出現しているガイドラインが特定のガイドラインに集中していることが挙げられる。ルール 30 件のうち、5.4a が 24 件で他より多く、残りは 5.4e が 4 件、5.9a が 2 件であった。ここで、ページ全体の 1400 件のうち、5.4a の問題点があるページは 774 件 (55.3%)、5.4e の問題点があるページは 700 件 (50.0%)、5.9a の問題点があるページは 921 件 (65.8%) であった。いずれも全体の 50% を超えており、問題点があるページの多いガイドラインであることから、もともと結論部に出現しやすい。これらが結論部に出現するルールのうちリフトの値が比較的小さいルールは、たとえ確信度の値が大きくても、「共起性」はあまり高くないと言える。これに該当するルールとしては、例えば、表 1 の No.6 や No.8 のルールが挙げられる。

表 1: 抽出されたアソシエーションルール X→Y
(支持度 S の上位 10 件)

No.	X			Y	S	C	L
1	5.4b			5.4a	0.33	0.82	1.48
2	5.4b	5.9a		5.4a	0.26	0.87	1.58
3	5.3e	5.4b		5.4a	0.22	0.82	1.49
4	5.4b	5.4e		5.4a	0.19	0.81	1.47
5	5.6c	5.9a		5.4a	0.17	0.82	1.48
6	5.2c	5.4a		5.9a	0.16	0.82	1.24
7	5.2c	5.9a		5.4a	0.16	0.81	1.46
8	5.2f			5.9a	0.16	0.81	1.23
9	5.3e	5.4b	5.9a	5.4a	0.16	0.87	1.58
10	5.1a	5.4b		5.4a	0.16	0.87	1.57

表 2: 抽出されたアソシエーションルール X→Y
(確信度 C の上位 10 件)

No.	X			Y	S	C	L
1	5.1b			5.4e	0.13	0.93	1.85
2	5.2c	5.4b	5.9a	5.4a	0.11	0.92	1.66
3	5.1a	5.4b	5.9a	5.4a	0.13	0.91	1.64
4	5.4b	5.6c	5.9a	5.4a	0.12	0.90	1.63
5	5.3a	5.4b	5.9a	5.4a	0.11	0.90	1.63
6	5.4b	5.6c		5.4a	0.15	0.89	1.60
7	5.2c	5.4b		5.4a	0.13	0.88	1.59
8	5.1a	5.3e	5.4b	5.4a	0.11	0.87	1.58
9	5.4b	5.9a		5.4a	0.26	0.87	1.58
10	5.3e	5.4b	5.9a	5.4a	0.16	0.87	1.58

表 3: 抽出されたアソシエーションルール X→Y
(リフト L の上位 10 件)

No.	X			Y	S	C	L
1	5.1b			5.4e	0.13	0.93	1.85
2	5.2c	5.4b	5.9a	5.4a	0.11	0.92	1.66
3	5.3a	5.9a		5.4e	0.15	0.83	1.65
4	5.1a	5.4b	5.9a	5.4a	0.13	0.91	1.64
5	5.4b	5.6c	5.9a	5.4a	0.12	0.90	1.63
6	5.3a	5.4b	5.9a	5.4a	0.11	0.90	1.63
7	5.3a	5.4a	5.9a	5.4e	0.12	0.82	1.63
8	5.4b	5.6c		5.4a	0.15	0.89	1.60
9	5.2c	5.4b		5.4a	0.13	0.88	1.59
10	5.1a	5.3e	5.4b	5.4a	0.11	0.87	1.58

また、ルールの条件部に出現しているガイドラインも特定のガイドラインに集中しており、5.4b が最も多く 22 件、5.9a が続いて多く 16 件であった。総じて、これらのガイドラインに関する問題点が、他のガイドラインよりも相対的に共起しやすいことがわかる。例えば、確信度 C の値およびリフト L の値がともに他より相対的に高いルールは、表 2 および表 3 の No.1 のルール：

If (5.1b の問題点がある) Then (5.4e の問題点がある)

であった。ここで、5.1b および 5.4e のガイドライン[2]は、それぞれ次のような内容である。

● 5.1b : ウェブコンテンツには、アクセス可能なオブジェクトなどの技術を使うことが望ましい。

● 5.4e : アクセス可能ではないオブジェクト、プログラムなどには、利用者がその内容を的確に理解し操作できるようにテキストなどの代替情報を提供しなければならない。また、アクセス可能なオブジェクト又はプログラムに対しても、内容を説明するテキストなどを提供することが望ましい。

本研究で用いた自動検査ツール[6]は、5.1b の検査において、<object>、<embed>、<applet>の使用有無などをチェックしており、5.4e では<embed>に対する<noembed>の有無、<applet>への alt 属性の指定の有無などをチェックしている。したがって、5.1b に問題があること (<applet>などを使用していること) と、5.4e に問題があること (<applet>に alt 属性がないこと) は、論理的なつながりがあり、表 2 および表 3 の No.1 のルールは意味的には意外性のあるルールとはいえない。ただし、<object>、<embed>、<applet>の使用については 5.1b および 5.4e の両方に注意すべきということがこのルールからもわかり、このルールはリフトの値が他より相対的に大きいことから、既存のページは<object>、<embed>、<applet>を使っていればそれらについて 5.4e に記載の配慮ができていないケースの割合が多いことがわかる。

4. まとめ

本研究では、アソシエーション分析の方法を用いて、Web アクセシビリティの問題点に関するアソシエーションルールの抽出を試みた。現時点では、まだ抽出されたルールの報告に留まっており、今後、これらのルールをもとに問題点の共起性を分析し、「あるガイドラインに問題があれば別のガイドラインの問題も含んでいることが多い」という事実から、その原因の分析や、それらの共起の発生を防ぐために注意すべきことを検討する必要がある。またこのほかに、2.2 節に記載の{Qij}のデータではなく元の{Pij}のデータを用いたルール抽出なども今後の課題である。

参考文献

- [1] W3C, Web Content Accessibility Guidelines 2.0.
<http://www.w3.org/TR/WCAG20/>
- [2] JIS X 8341-3, 高齢者・障害者等配慮設計指針 - 情報通信における機器、ソフトウェア及びサービス - 第 3 部: ウェブコンテンツ (2004)
- [3] W3C WAI, Complete List of Web Accessibility Evaluation Tools. <http://www.w3.org/WAI/ER/tools/complete>
- [4] Allied-Brains Accessibility Online, アクセシビリティリンク集, 2-2.点検ツール.
<http://www.aao.ne.jp/accessibility/links/2-2.html>
- [5] Yahoo!カテゴリ
<http://dir.yahoo.co.jp/>
- [6] WebInspector : 富士通
<http://jp.fujitsu.com/about/design/ud/assistance/webinspector/index.html>
- [7] 金明哲: R によるデータサイエンス - データ解析の基礎から最新手法まで, 森北出版 (2007)