

ポスター

ポスター8

医療データ分析4（データ・テキストマイニング）

2018年11月23日(金) 17:00 ～ 18:00 K会場(ポスター、HyperDemo) (2F 多目的ホール)

[2-K-4-5] 検索語を取得できないアクセスログを活用した病院 Web サイトの 閲覧モデル構築手法の開発

○鈴木 隆介, 鈴木 哲平, 辻 慎太郎, 小笠原 克彦（北海道大学保健科学院）

【背景と目的】近年、医療機関 Web サイトが、患者の医療機関選択時の情報源となっている。そのため、医療機関は市民のニーズを反映した情報提供を行う必要がある。しかし、医療機関 Web サイト、特に放射線関連 Web サイト情報を閲覧する市民の閲覧行動は明らかになっていない。この原因として、閲覧意図を表す検索キーワードが取得できないログが多く割合を占めることが挙げられる。そこで、本研究では、共起ネットワークを用いて放射線関連キーワードをクラスタリングし、それぞれのクラスターの特徴を求めることで、キーワードを取得できない閲覧行動の閲覧意図を推測することで、より精度の高い閲覧行動モデルを構築することを目指した。【方法】2016年9月1日から2017年8月31日までの北海道大学病院 Web サイトのキーワード取得状況を Google Analytics を用いて調査し、ログを抽出した。次に、検索キーワードが取得できた899件のセッションデータの検索キーワードを用いて共起ネットワークを作成し、ネットワークの形状等から閲覧行動のクラスタリングを行った。閲覧行動のログ情報を用いて、クラスター番号を目的変数、閲覧ページ、ページデプス等を説明変数とした数量化Ⅱ類による要因分析を行った。分析により得られた判別モデルより、検索キーワードが取得できていない17,878件のセッションデータを閲覧ニーズごとにクラスタリングし、ベイジアンネットワークを用いて閲覧行動の可視化を行った。【結果と考察】Web サイトのキーワード取得率は11.4%であり、従来の分析は限定的な分析であることが明らかになった。閲覧意図に合わせて5つの閲覧行動に分類でき、判別率は64.3%であった。この判別モデルより各クラスターの閲覧行動を表したベイジアンネットワークモデルが構築され、放射線治療・PETのモデルでは検索キーワードが取得できたデータより作成したモデルと同様の傾向が観察され、より詳細な閲覧モデルの推測を行える可能性が示唆された。

検索語を取得できないアクセスログを活用した 病院 Web サイトの閲覧モデル構築手法の開発

鈴木隆介^{*1}、鈴木哲平^{*1}、辻真太郎^{*1}、小笠原克彦^{*1}^{*1} 北海道大学大学院保健科学院

Development of browsing-model construction method of hospital Website using access-log which can not acquire search term

Ryusuke Suzuki, Teppei Suzuki, Shintaro Tsuji, Katsuhiko Ogasawara

Graduate school of Health Science, Hokkaido University

The purpose of this research is to construct browsing behavior models with higher accuracy by guessing the intention of browsing behavior that can't acquire keywords. We created the co-occurrence network using Hokkaido University Hospital website access-log, and clustering browsing behaviors by the shape of the network etc. Next, using the log information of browsing behavior, we performed quantification II classification. Bayesian network was constructed using discriminant models obtained by analysis. It can be classified into five browsing behaviors according to the intention of browsing, and the discrimination ratio was 64.3%. From this discriminant model, a Bayesian network model representing the browsing behaviors of each cluster was constructed, and in the radiotherapy model, the same tendency as the model created from the data from which the retrieval keywords were acquired was observed. It was suggested that we could make a detailed browsing model estimation.

Keywords: Web marketing, hospital Website, hospital information, Bayesian network

1. 緒論

健康保険組合連合会の平成 29 年度の「医療に関する国民意識調査」によると、医療機関を選ぶにあたり最も参考にするものは「インターネットの情報」であった¹⁾。また、インターネットの情報源として「医療機関のホームページ」が高い割合を占めていた。インターネット環境の整備が進む現代では、受診する病院、検査・治療選択にインターネットが用いられる比率はさらに上昇していくものと予想される。そこで、医療機関 Web サイトは、市民のニーズを反映した情報提供を行うことが望まれる。鈴木らは、市民の閲覧行動を元にサイトの改善を検討する手法としてベイジアンネットワークを用いた閲覧モデルの構築を提案した²⁾。この手法を用いることで、近年市民の関心が高まっている放射線関連情報提供の改善提案を検討することができたが、放射線関連情報のように限定的な情報提供の改善を検討する際は、市民の閲覧意図を把握できるケースが少なく、閲覧モデルの精度にも影響が出るという問題点が報告されている。この原因として、閲覧モデルを構築するにあたり使用するアクセスログデータのうち、閲覧意図を表す検索キーワードが取得できないログが多くの割合を占めていることが挙げられる³⁾。検索キーワードが取得できないデータの閲覧意図を推測することで、閲覧モデル作成にあたり使用できるサンプルデータ数が向上し、より詳細で精度の高いモデル構築の手法を確立することで、医療機関経営者が容易にサイト改善を検討できることが予想される。

2. 目的

共起ネットワークを用いて放射線関連キーワードをクラスタリングし、それぞれのクラスタの特徴を求めることで、キーワードを取得できない閲覧行動の閲覧意図を推測し、閲覧モデル構築の際使用できるサンプルデータ数を向上させることでより精度の高い閲覧行動モデルを構築することを目指した。

3. 方法

2016 年 9 月 1 日から 2017 年 8 月 31 日までの北海道大学病院 (以下、北大病院) Web サイトのログを使用した⁴⁾。

3.1 検索キーワード取得率の調査

取得したアクセスログデータのうち、検索キーワードが取得できたデータの割合を算出した。指標として、任意のユーザーがサイトに訪問した回数を示す「セッション数」を用いた。

3.2 キーワード取得可能データを用いた分析

3.2.1 閲覧意図のクラスタリング

共起ネットワークを作成して、閲覧意図のクラスタリングを行った。共起ネットワークとはテキストの中で用いられた語をノードとして、同時に用いられた語と語を結びネットワークとして表したものである。取得した放射線関連キーワードを形態素分解し、形態素の共起パターンを元にしたネットワークを構築する。構築したネットワークを元に仮クラスタを作成し、重複の多いクラスタ同士を結合させるなどの処理を行った後、同様の検索意図を持つクラスタとして採用した。共起ネットワークの構築に KH Coder⁵⁾を用い、構築条件はサブグラフ検出を、検索キーワードという非常に短文の中から共起関係を探ることから、共起頻度の尺度として利用される Jacard 係数は 0.1 以上、最小出現数は 5 以上とした。

3.2.2 判別モデル式の構築

閲覧意図を推測する手法として数量化 II 類を用いた。数量化 II 類とは目的変数・説明変数ともに質の変数で、目的変数の判別を行う際に使用される。今回、閲覧意図を表すクラスタ番号を目的変数とした。説明変数には PV 数の多い放射線関連ページ 10 ページ分に対する閲覧の有無と、任意の 1 セッションで閲覧したページ数を指すページデプス、ネットワークドメインを用いた。作成した判別モデルを評価する指標として、相関比、判別率を用いた。

3.3 キーワード取得不可データを用いた分析

検索キーワード取得不可データのうち、放射線関連ページを閲覧しているセッションデータを抽出し、判別モデルに投入することでそれぞれのセッションの閲覧意図を推測する。その後クラスタごとに、①検索キーワード取得可能データのみを用いてベイジアンネットワークを構築 (以下、通常モデル) ②検索キーワード取得不可データを用いてベイジアンネットワーク

を構築(以下、拡張モデル)した。ベイジアンネットワークは、不確実性を含む事象の予測や合理的な意思決定、障害診断などに利用することのできるグラフィカルモデルの一種である⁶⁾。クラスタごとのベイジアンネットワークの形状を比較し、判別モデルを用いてサンプルデータ数を向上させる手法の有用性を確認した。モデル構築の際、BayoLink version7.0.1⁷⁾を使用した。モデル構築に Greedy-Search アルゴリズムを採用し、ノードは対象閲覧者の閲覧ページの閲覧確率とした。

4. 結果

期間中の全閲覧セッション数は 663,213 件であり、検索キーワードが取得可能であったセッション数は 75,586 件であった。検索キーワード取得率は 11.4%であり、膨大なアクセスログデータの内閲覧意図が取得できるのはごくわずかであった。

4.1 閲覧意図のクラスタリング

共起ネットワークの構築の結果、図1のようなネットワークを構築することができた。ネットワークの形状を元に、①SPECTシンチ②放射線治療③核医学全般④MRI⑤歯科の5つのクラスタを同様の閲覧意図を持つまとまりとして採用した。

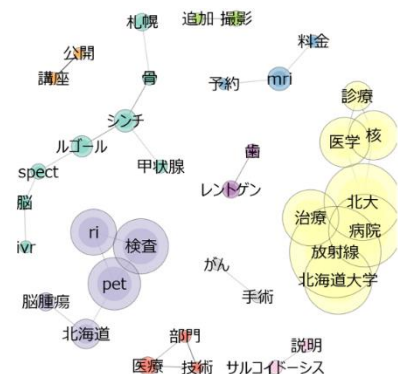


図1 共起ネットワーク(放射線関連キーワード)

4.2 判別モデル

目的変数が5カテゴリーの数量化Ⅱ類を行ったため、4つのモデル式が作成された。各モデル式における相関比は表1の通りである。相関比が比較的高いモデル式1・2を閲覧意図の判別に用いることとした。判別率は63.4%であった。

この判別モデルを用いて、検索キーワードを取得できないが放射線関連情報のニーズがあると思われるデータ(n=17,878)の閲覧意図の判別を行った(表2)。

表1 モデル式の相関比

モデル式	相関比
1	0.466
2	0.429
3	0.226
4	0.109

表2 検索キーワード取得不可データ判別結果(n=17,878)

クラスタ名	推定セッション数
①SPECT・シンチ	5,396
②放射線治療	4,484
③核医学全般	2,365
④MRI	5,113
⑤歯科	520

4.3 ベイジアンネットワーク

ベイジアンネットワークを構築した結果、通常モデルと拡張モデルで同様の閲覧ページの関係性を確認できた。図2に示すようにクラスタ②放射線治療では、通常モデルで示された、放射線治療ページ、放射線診断科ページ、PET/CT・RIページの関係が拡張モデルでも描出されている。アクセスページなど通常モデルでは確認できないページも描出された。

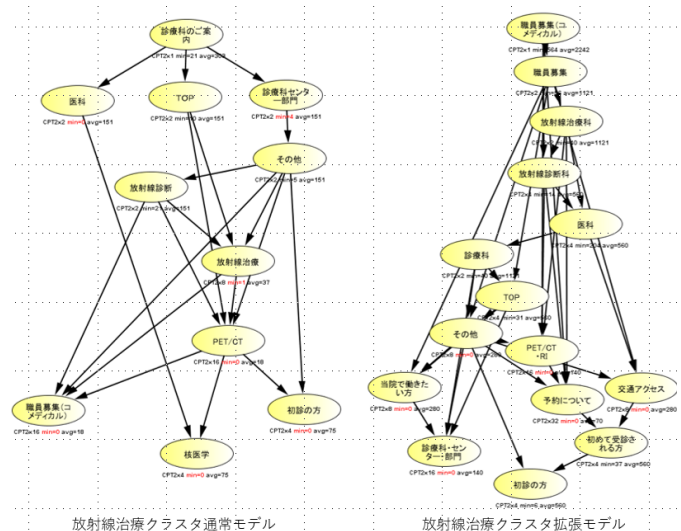


図2 放射線治療クラスタ拡張モデル

5. 結論

本研究では、検索キーワードが取得できないアクセスログの閲覧意図を推測することで、閲覧モデルに使用するサンプルデータを拡張することを目指した。まず、共起ネットワークを用いて同様の検索意図を持つデータのクラスタリングを行う手法を提案した。次に、説明変数に閲覧ページなどを用いて、目的変数である閲覧意図を推測する判別モデルを構築し、検索意図を推測する手法を提案した。結果、判別率63.4%の判別モデルを構築することができた。また、検索キーワード取得可能データから構築したベイジアンネットワークと判別モデルでサンプルデータを拡張して構築したベイジアンネットワークを比較した結果、同様の傾向が見られ、拡張モデルを用いることでより詳細なページの関係性を把握することができた。今後、判別精度の向上を目的として検索意図に寄与する説明変数の探索を行うことで、医療機関がアクセスログを元にWebサイトの改善を検討する手法の有用性を高めていきたい。

参考文献

- 1) 健康保険組合連合会, 医療に関する国民意識調査, 2017.
- 2) 鈴木隆介, 鈴木哲平, 辻真太郎, 小笠原克彦: ベイジアンネットワークを用いた放射線関連Webサイトの閲覧モデル構築, 2017. [https://confit.atlas.jp/guide/event/jcml2018s/session/KB-11/tables?YuLCmXsEXk (cited 2018-Aug-29)]
- 3) Yahoo! JAPAN. Yahoo! JAPAN サービスは常時 SSL (AOSSL) に対応します. [https://about.yahoo.co.jp/info/aossil/ (cited 2018-Aug-29)].
- 4) 北海道大学病院. [https://about.yahoo.co.jp/info/aossil/ (cited 2018-Aug-29)]
- 5) KH Coder. [http://kncoder.net/ (cited 2018-Aug-29)]
- 6) 本村陽一, 岩崎弘利: ベイジアンネットワーク技術, 東京電機大学出版局, 2006.
- 7) BayoLink. [http://www.msi.co.jp/bayolink/ (cited 2018-Aug-29)]