



HOKKAIDO
UNIVERSITY

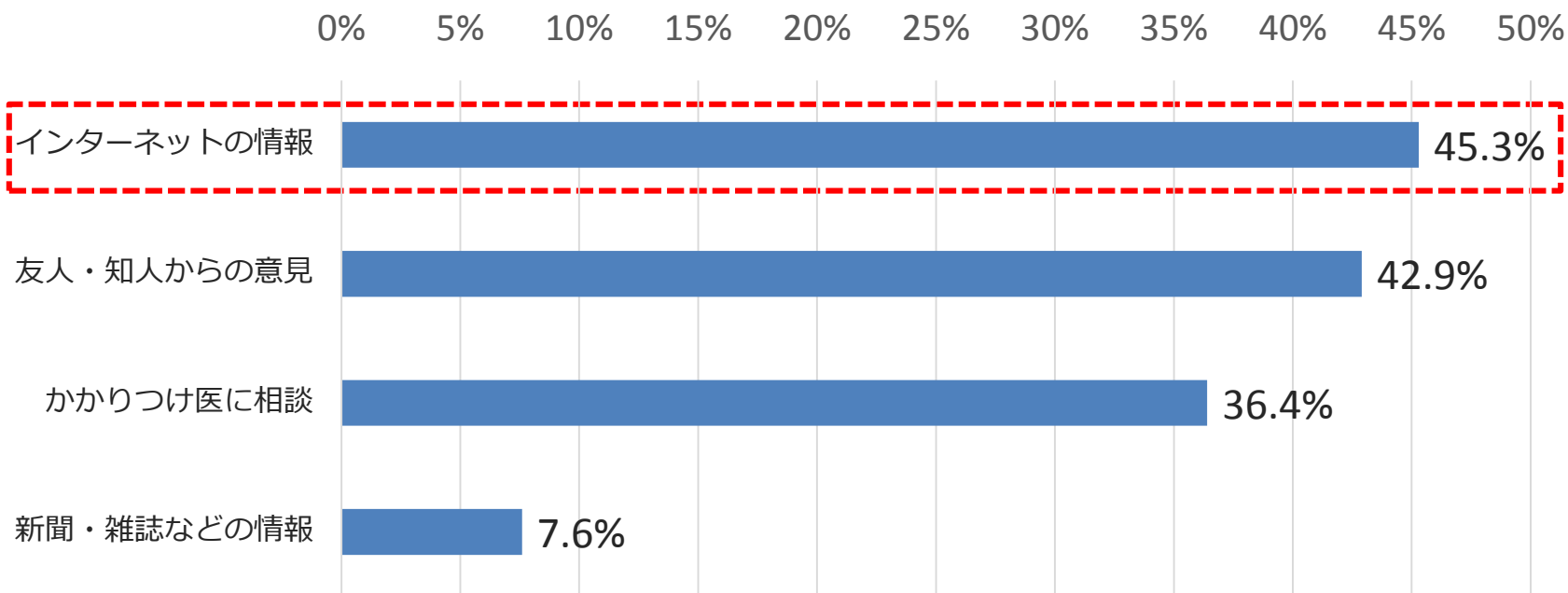
ベイジアンネットワークを用いた 医療機関Webサイトの改善手法の提案

鈴木隆介 鈴木哲平 小笠原 克彦
北海道大学大学院保健科学院

株式会社NTTデータ数理システム
2018年度 学生研究奨励賞

背景 | 市民の医療機関選択

医療機関を選択する際の情報源（複数回答）



市民にとってインターネットは重要な情報源

背景 | 医療機関からの情報発信

■ 広告規制

- 医業における広告は原則禁止
 - 1)医療は人の生命・健康に密接にかかわる
 - 2)きわめて専門性が高い領域を取り扱う
- チラシ・パンフレット・ポスター・バナー広告などの広告は国から規制
- 医療機関Webサイトのみ規制対象外*
- 2018年医療法改正で医療機関Webサイトも広告規制の対象になったが他広告とは別の枠組みにする案が検討

医療機関Webサイトは情報発信の柱

背景 | 医療機関Webサイトアクセス状況

平成27年度
北海道大学病院患者数

外来患者数	746,707人
入院患者数	303,194人

北海道大学病院Webサイト
アクセスログ
(2017.1/1-2017.12/31)

PV数	2,595,312
セッション数	685,907
ユーザー数	398,505

医療機関Webサイトは多くの市民に情報を提供

背景 | Webサイトを用いた経営戦略

- Webサイトマーケティングではプロジェクト定義から改善案実行までの推奨プロセスが提案されている

Dave Chaffey et al. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics, Journal of Direct, Data and Digital Marketing Practice, 2012.

- サイトパフォーマンスを向上させる上でWeb解析は重要視されている

Joel Jarvien, Heikki Karjaluoto. The use of Web analytics for digital marketing performance measurement, Industrial Marketing Management, 2015.

- 医療機関WebサイトのWeb解析を行う研究は少ない

Teppei Suzuki, Yuji Tani, Katsuhiko Ogasawara. Behavioral Analysis of Visitors to a Medical Institution's Website Using Markov Chain Methods, Journal of Medical Internet Research, 2016.

医療機関に適したWebサイト分析手法は提案されていない

目的

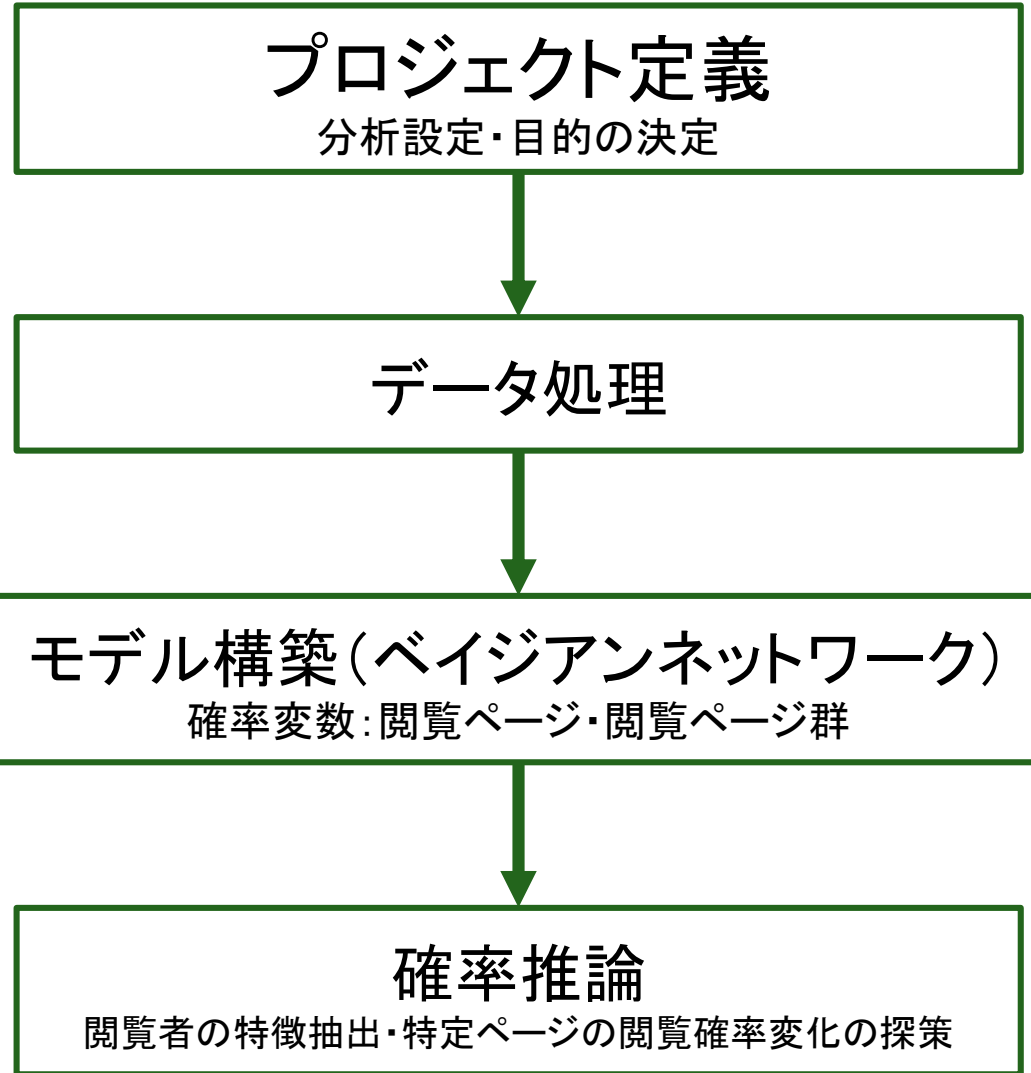
【目的】

医療機関がアクセスログからWebサイト閲覧者の閲覧行動を把握し改善案を検討できる手法の提案



医療機関がWebサイト改善について抱えると予想されるケースを元にベイジアンネットワークを用いて分析

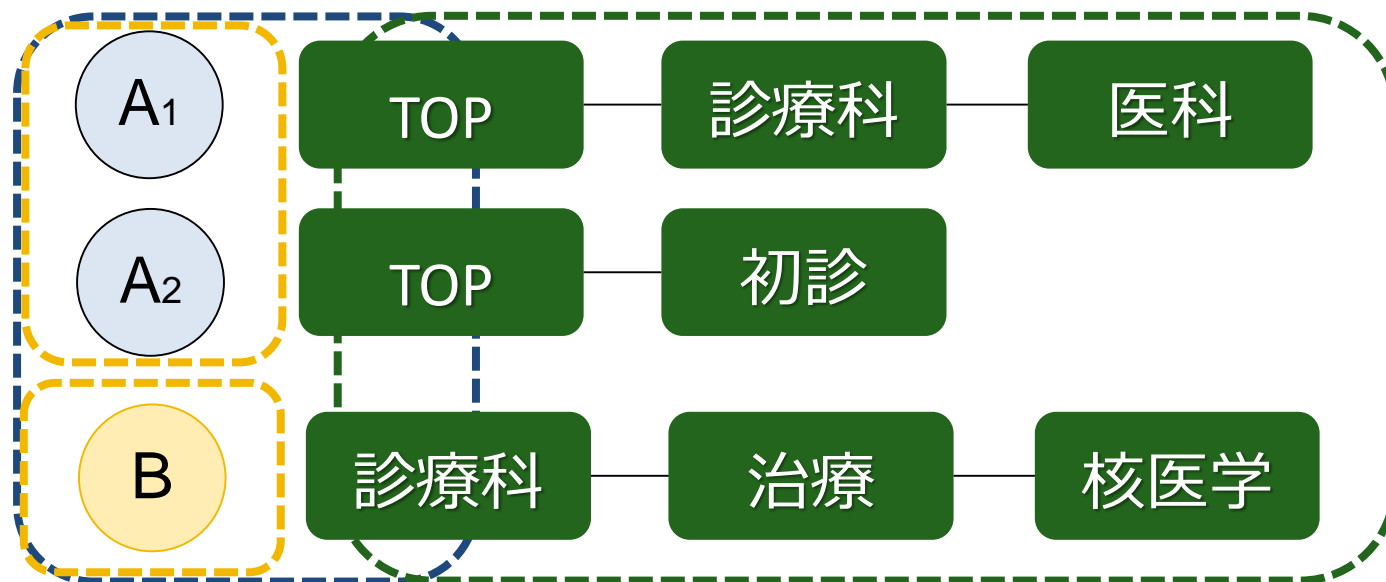
方法 | 研究の流れ



方法 | 指標

閲覧者がサイトに
訪問した数

セッション数
= 3



ユーザー数 = 2

サイトに訪問した
閲覧者の数

PV数 = 8

閲覧者が閲覧した
ページ数

方法 | 研究対象データ

■ データ

-北海道大学病院Webサイト

(<http://www.huhp.hokudai.ac.jp/>) のアクセスログ

•モデル構築用

2018年4月1日～7月31日 (4ヶ月)

ページビュー数 : 939,029 セッション数 : 257,165

•検証用

2018年8月1日～9月30日 (2ヶ月)

ページビュー数 : 487,178 セッション数 : 145,722

■ 使用ソフト

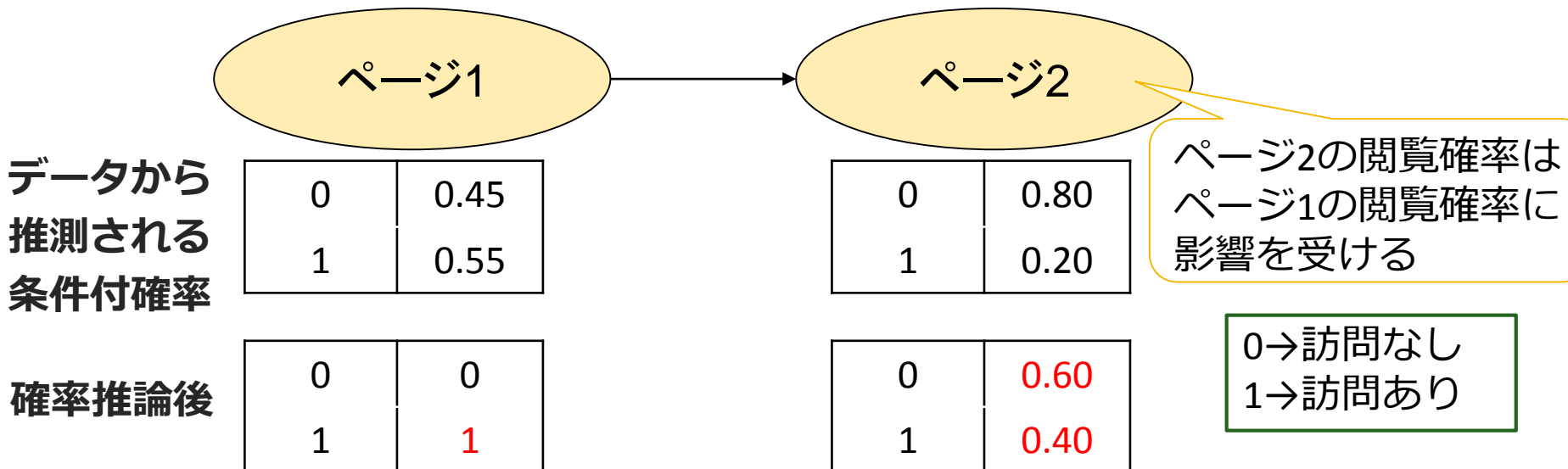
Google Analytics, BayoLink Ver.7.01 (NTTデータ数理システム社製)

方法 | ベイジアンネットワーク

■ ベイジアンネットワーク

変数間の定性的な依存関係をグラフ構造に、定量的な関係を条件付確率で表したモデル

- ① データからモデル構造を決定できる
- ② 任意の変数にエビデンスを入力することで他変数の事後確率を求めることができる（確率推論）



方法 | データ処理 ①

①データ抽出

検索キーワード・ページデプスなど8つの指標を指定してアクセスログを抽出

検索キーワード	市町村	アクセス日	ネットワークドメイン	ページデプス	ランディングページ	ページタイトル	PV数
北大病院	札幌	4月1日	a	3	TOP	TOP	1
北大病院	札幌	4月1日	a	3	TOP	診療科	1
北大病院	札幌	4月1日	a	3	TOP	内科	1
北大病院	札幌	4月5日	b	1	TOP	TOP	1
放射線	千歳	5月3日	c	2	放射線科	放射線科	1
放射線	千歳	5月3日	c	2	放射線科	検査	1

方法 | データ処理 ②

②データ整形

セッションごとの閲覧ページ、確率変数として用いるページを抽出するためにIDを割り当てる

セッションID	検索キーワード	市町村	アクセス日	ネットワークドメイン	ページデプス	ランディングページ	ページタイトル	ページID
1	北大病院	札幌	4月1日	a	3	TOP	TOP	1
1	北大病院	札幌	4月1日	a	3	TOP	診療科	2
1	北大病院	札幌	4月1日	a	3	TOP	内科	4
2	北大病院	札幌	4月5日	b	1	TOP	TOP	1
3	放射線	千歳	5月3日	c	2	放射線科	放射線科	5
3	放射線	千歳	5月3日	c	2	放射線科	検査	9

方法 | データ処理③

③データ行列作成

IDを用いて、ベイジアンネットワークで利用できるデータ形式に変換する

セッションID	ページID	ページID		
		1	2	3
1	1	1	1	0
1	2	0	0	0
1	4	0	0	0
2	1	1	0	0
3	5			
3	9			

0→訪問なし
1→訪問あり

データ処理をプログラムで自動化

方法 | 閲覧モデル① -全体モデル-

設定

医療機関Webサイトは会員制サイトではないため閲覧者属性ごとの閲覧行動把握が難しい

→**閲覧者属性ごとの閲覧行動の違いを把握する**

使用データ

1セッションにつき2ページ以上ページを閲覧したデータ

確率変数

サイト構造を参考にページのクラスタリングを行いクラスタをノードとして用いる（例：外来患者ページ群、アクセスページ群）

確率推論

閲覧者属性ページ群のそれぞれの状態を1とした確率推論

方法 | 閲覧モデル② -患者モデル-

設定

ECサイトなどで売上貢献を確認する指標として一般的に利用されるCVR（Conversion Rate）の設定が医療機関では難しい
→新規患者獲得のポジションにあたるページの探索

使用データ

1セッションにつき2ページ以上ページを閲覧したデータ
医療関係者・採用を対象としたページを閲覧していないデータ
「募集」「求人」「実習」などの検索キーワードを用いていないデータ

確率変数

対象データにおけるPV数上位30ページ

確率推論

「交通アクセスページ」「診療日ページ」の指定条件化における確率推論
各診療科ページの感度分析

方法 | モデル構築

■ モデル構築

モデルを選択する際の基準：赤池情報量基準 (AIC)

モデル構造決定法：Greedy Search アルゴリズム

■ モデル評価

「0」の比率が低い「TOPページ」の予測を2つの指標を用いて評価

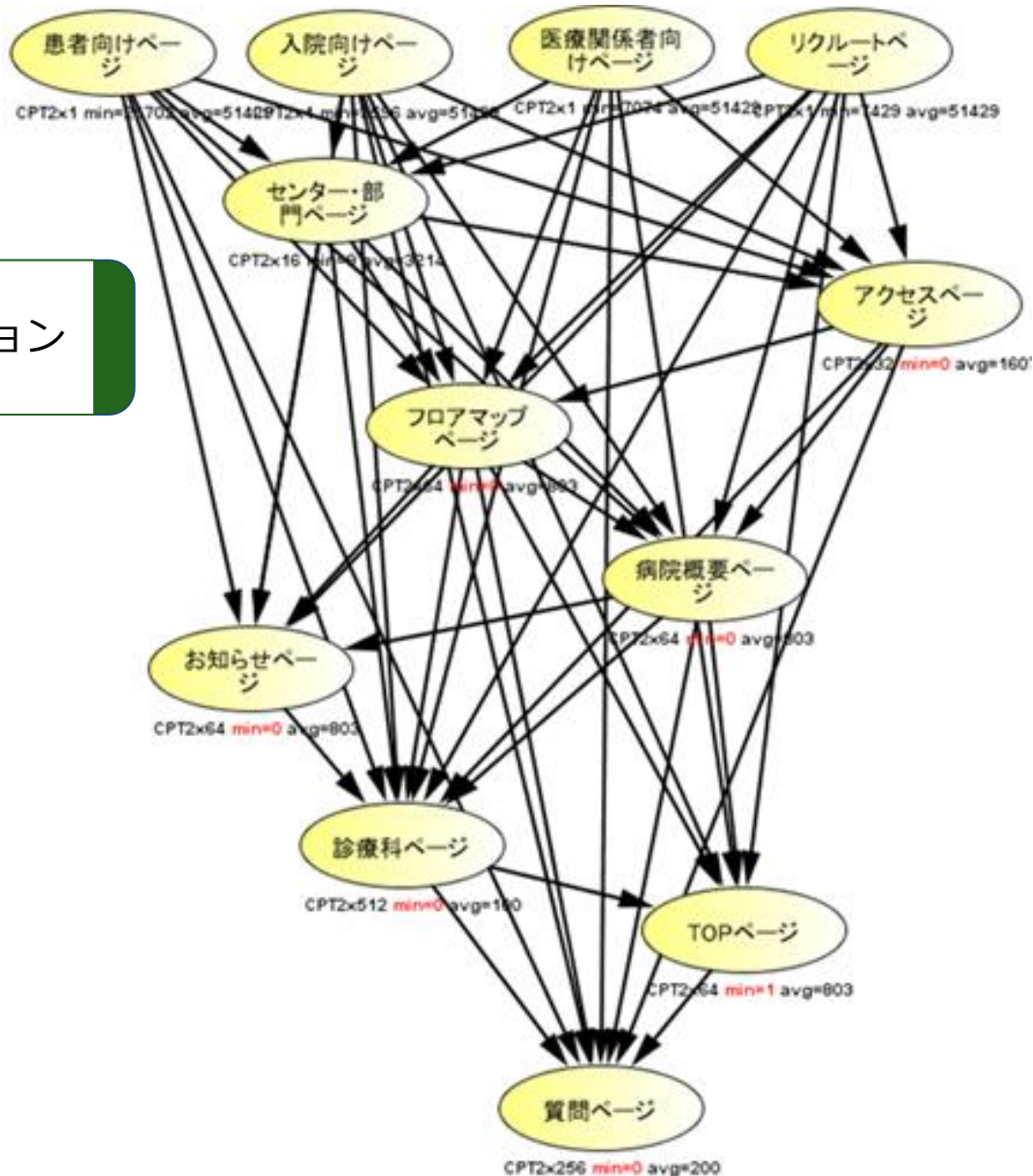
正解率 正しく推定が行われた割合

F値 再現率と適合率を総合的に評価する指標

$$F - \text{measure} = \frac{2 \textit{precision} * \textit{recall}}{\textit{precision} + \textit{recall}}$$

結果 | 全体モデル

対象 102,858セッション



結果・考察 | 全体モデル

表：ページ群閲覧確率変化量の比較

ページ群	推論前	閲覧者属性ページ群			
		外来ページ 閲覧	入院ページ 閲覧	医療関係者 ページ 閲覧	リクルート ページ 閲覧
フロアマップページ	12.74%	-0.02%	+14.96%	-3.45%	-6.68%
アクセスページ	11.59%	+0.55%	+13.57%	-4.33%	-7.66%
病院概要ページ	11.22%	-0.56%	+1.88%	-0.84%	-1.68%
センター・部門 ページ	10.59%	-4.33%	-3.49%	+2.78%	-0.16%
診療科ページ	61.29%	+8.66%	-33.93%	-19.06%	-35.24%
質問ページ	8.24%	+4.80%	+4.13%	+0.78%	-1.89%
お知らせページ	8.24%	-4.87%	-4.41%	-3.24%	-4.10%

結果・考察 | 全体モデル

■ 外来患者の特徴

診療科ページ群や質問ページ群の閲覧確率が高い

■ 入院患者の特徴

フロアマップ・アクセスページ群、質問ページ群の閲覧確率が高い

■ 医療関係者の特徴

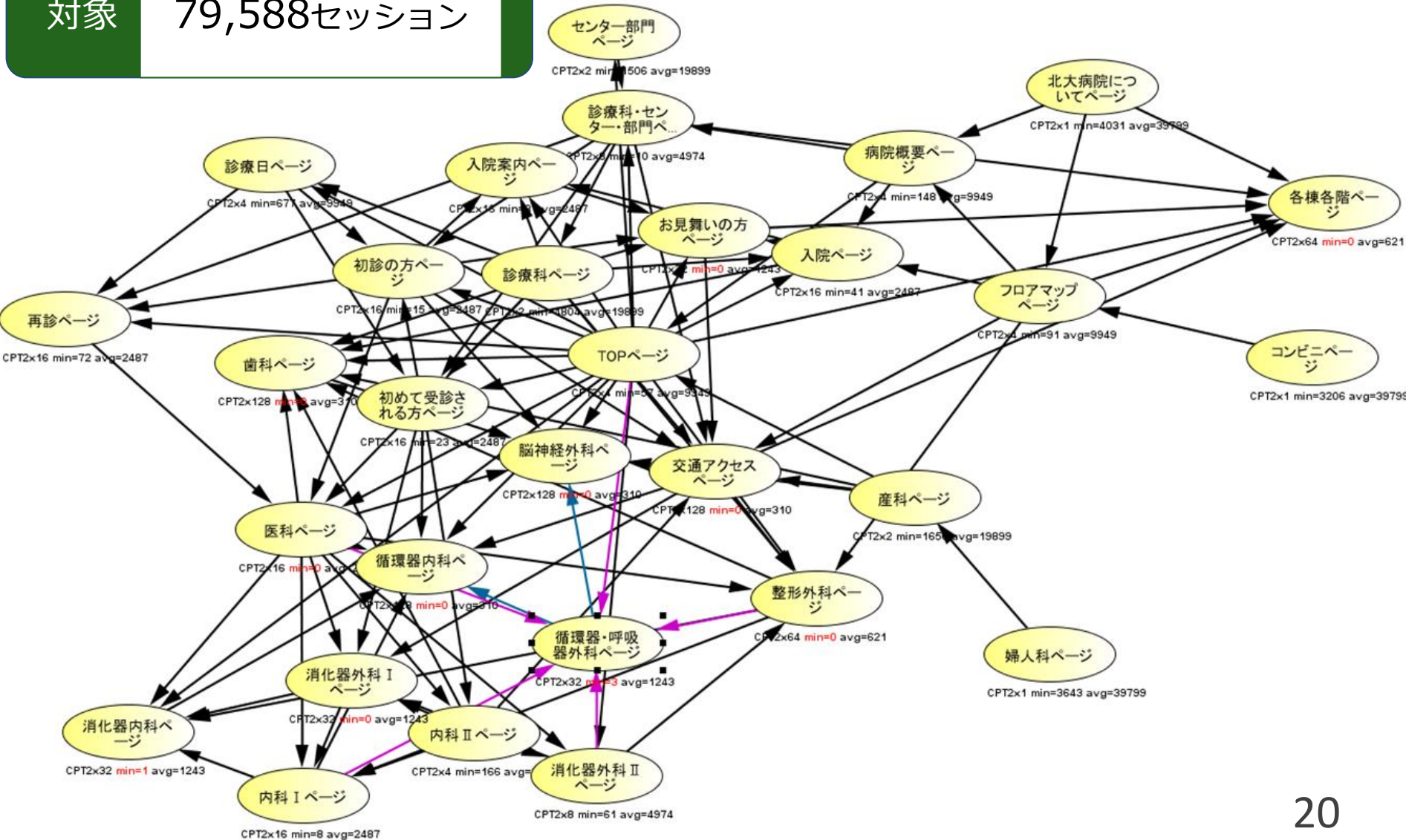
センター部門ページ群の閲覧確率が高い

【改善案】

- 入院患者ページ群とフロアマップページ群・質問ページ群をリンクする
- センター部門ページ群と診療科ページ群の掲載コンテンツのすみ分け

結果 | 患者モデル

対象 79,588セッション



結果・考察 | 患者モデル

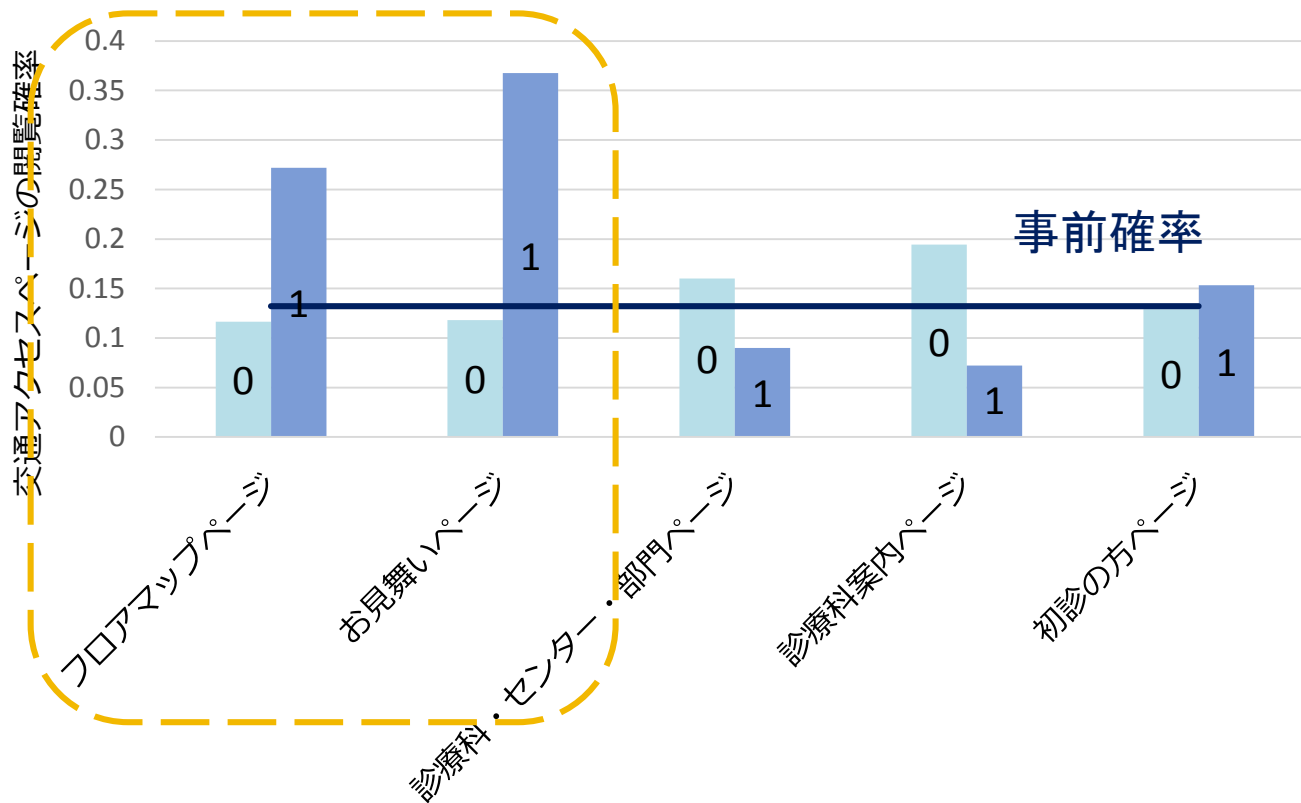
表：各診療科ページと相互情報量の大きいページ

診療科ページ	相互情報量の大きいページ
内科 I ページ	内科 II ページ、医科ページ、消化器内科ページ
内科 II ページ	内科 I ページ、医科ページ、診療科案内ページ
整形外科ページ	診療科ページ、医科ページ
消化器内科ページ	医科ページ、内科 I ページ、消化器外科 I ページ、診療科案内ページ
消化器外科 I ページ	消化器外科 II ページ、医科ページ、消化器内科ページ
消化器外科 II ページ	消化器外科 I ページ、医科ページ
循環器内科ページ	循環器・呼吸器外科ページ、医科ページ、診療科案内ページ
循環器呼吸器外科ページ	医科ページ、循環器内科ページ
婦人科ページ	産科ページ
産科ページ	婦人科ページ
脳神経外科ページ	該当ページなし

リンク設計や診療科案内ページのページデザイン検討

結果・考察 | 患者モデル

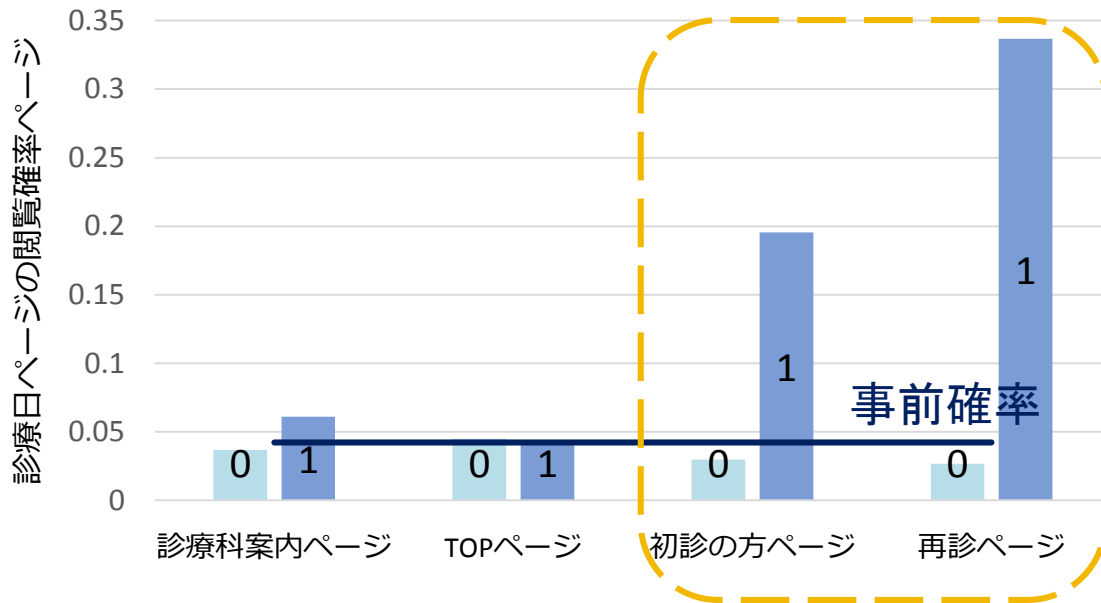
図：交通アクセスページの閲覧確率推移



「お見舞いの方ページ」, 「フロアマップページ」 閲覧が大きな影響
→新規患者獲得の基準となるページには適していない

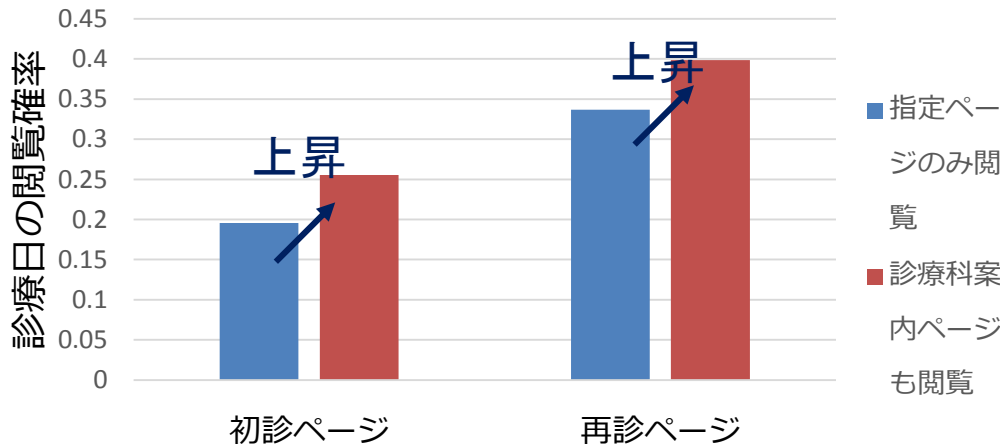
結果・考察 | 患者モデル

図：診療日ページの閲覧確率推移



「初診ページ」, 「再診ページ」
閲覧が大きな影響

→新規患者獲得の基準
ページに適している



新規患者獲得戦略を
立案する際に応用
できる

結果・考察 | 患者モデル

表：再診・初診ページ閲覧者の閲覧確率比較

ページ	再診ページ閲覧者の 閲覧確率	初診ページ閲覧者の 閲覧確率
TOPページ	48.8%	44.3%
診療科ページ	50.3%	45.7%
医科ページ	23.6%	21.5%
診療科・センター・部門 ページ	37.7%	31.2%
交通アクセスページ	12.8%	15.3%
フロアマップページ	10.1%	10.0%
病院概要ページ	8.8%	8.2%
センター部門ページ	7.7%	6.6%
北大病院についてページ	5.3%	5.1%

閲覧確率が
高いページ

閲覧傾向の分析結果からリンク設計・掲載コンテンツの再考

結果 | モデル評価

表：全体モデルの精度評価

正解率		0.647
F値	0	0.619
	1	0.671

表：患者モデルの精度評価

正解率		0.690
F値	0	0.679
	1	0.700

正解率は先行研究と比べても遜色のない値

立岡恵介, 吉田哲, 宗本順三. 店舗内の購買行動のベイジアンネットワーク分析, 日本建築学会計画系論文集第73巻代634号, 2633-2638, 2008.

考察 | ベイジアンネットワークの利用について

- 明確な仮説情報なしでモデル構築できる
- 直感的に閲覧行動を理解できる
- 特定の閲覧者の行動を分析できる



Webマーケティングが確立していない医療機関でも利用しやすい解析手法

ベイジアンネットワーク構築のPoint

- ① 確率変数の設定
- ② 確率変数間の因果関係の決定
- ③ 確率推論



BayoLinkを用いることで、適切な確率変数設定と確率推論シナリオを用意することでさらに簡単に分析可能

今後の課題

■ 正確な閲覧モデル構築

- トラッキングコード未設定（リンク先）ページに対しても同様の分析を行う
- モデルの変数選択を的確に行う

■ 本手法の有用性の検証

- 構築したモデルから検討した改善案を実行し成果が出ることを確認する
- 他医療機関を対象として本手法を用いた分析を行う

結論

目的

医療機関がアクセスログからWebサイト閲覧者の閲覧行動を把握し改善案を検討できる手法の提案

結果と展望

- ベイジアンネットワークを用いた閲覧モデル利用の提案
- 現場が抱えるであろうケースをもとに2つのモデルを構築
- 医療機関の閲覧者属性ごとの閲覧行動の把握から診療科ページの関連性の分析が可能
- **目的に応じてデータ整形・モデル構築・確率推論を行うことで様々な課題の分析を行える可能性**
- 今後はPDCAを回し、他医療機関にも適用することで本手法の有用性を検証したい