

利用状況の動的変化に対応するブックマーク情報提示システムの提案

北越大輔†, 塩谷浩之‡, 中野良平†

† 名古屋工業大学 〒466-8555 愛知県名古屋市昭和区御器所町

‡ 室蘭工業大学 〒050-8585 北海道室蘭市水元町 27-1

† {kitakosi, nakano}@ics.nitech.ac.jp, ‡ shioya@epsilon2.csse.muroran-it.ac.jp

<http://www-nkn.ics.nitech.ac.jp/~kitakosi/>

Abstract: 不確実性に対して頑健な性質を有する確率モデルを利用したシステムに関する研究が盛んに行われている。我々は、不確実性を含む動的環境等に対して有効に機能する確率モデルの一つであるベイジアンネットワークを混合モデルの形式で利用することで、インターネット利用者の利用状況の変化に柔軟に適応した効果的なブックマーク提示システムを提案する。提案システムを適用することによって、Web 巡回の際に、膨大なブックマークコンテンツから所望のページを探索する時間と精神的負荷を軽減可能となることが期待される。

1. はじめに

不確実性に対して頑健な性質を有する確率モデルを利用した、様々なシステムに関する研究が盛んに行われている。また、より実用的な問題設定として、システムを取り巻く環境の動的変化への柔軟な対応が求められる状況も数多く存在する。本稿で取り上げるインターネット利用環境でも、ユーザの利用状況（時間帯、趣味・嗜好、感情、仕事中 or 休憩中 等）に応じて閲覧したい Web ページは変化し得る。このような不確実性や動的性を含む環境に対して有効に機能すると考えられている確率モデルの一つに**ベイジアンネットワーク (Bayesian Network : BN)**がある。BN は確率変数間の依存関係を非循環性有向グラフの形式で表現した知識表現系ネットワークであり、その構造を利用した様々な確率推論法や、統計データを利用したネットワーク構造学習法が提案されている。

我々は、複数の BN を組み合わせた BN 混合モデルを用いて強化学習エージェントを様々な環境変化へ適応させることを目的としたシステム (**IPMBN**) を提案し、その有効性の検証を行っている[8]。計算機実験および移動ロボットを用いた実験から、BN 混合モデルが多様な環境情報を表現可能であり、IPMBN を実装したロボット (エージェント) が未知環境を含む様々な環境変化へ適応可能となることを確認した。強力な情報表現能力を有する BN 混合モデルは、不確実性や頻繁な環境変化を伴う多くの状況において有効なツールとなることが期待できる。

本研究では、BN 混合モデルのより一般的かつ実用的な適用例として、インターネット利用者の利用状況の変化に適応した効果的なブックマーク (お気に入り) 提示システムの構築を目指す。ユーザの利用状況変化は、文献[8]におけるロボットの環境変化と対応し、利用状況変化に対する効率的なブックマークの提示は、ロボットの環境変化への適応と対応づけられる。多くのユーザは一般的に複数の Web ページをブックマーク登録していると考えられるが、その分量の増加に伴い管理の困難が生じてくると予想される。また、同一の情報端末上で趣味等に関するページや、仕事、勉強等に関連するサイトの双方を閲覧する場合、ユーザのこれまでの利用状況に応じて今後選択される可能性の高いページを“予め選択されやすい状態”にしておくことで、膨大なブックマークコンテンツから所望のページを探索する時間と精神的負荷を軽減し、軽快なネットサーフィンの実現が可能となる。

2. 提案システム

提案システムでは、ユーザの現在の利用状況に基づき、ユーザが今後選択する可能性の高いブックマークコンテンツを確率推論によって求め、それらをユーザが選択しやすい形式で提示することでスムーズな Web 巡回を可能とすることを目指す。ユーザのブックマーク (コンテンツ) 選択方針は、後述する BN 混合モデルによって表現される。提案するブックマーク情報提示システムの枠組を図 1 に示す。一つの BN は特定の同時確率分布を表し、ネットワーク内のノード (確率変数) 間に確率的依存関係が存在する場合、それらの間に有向リンクが付加される。

本稿では、システムを利用するユーザの Web 巡回履歴を、ユーザの利用状況 (例: 趣味での利用時 or 仕事, 研究での利用時) ごとに分類したサンプルデータとして収集し、これらを基に各 BN の結合構造 (図 1 の Nb1 および Nb2) を決定する。構造決定された BN のそれぞれは、典型的な各利用状況でのユーザのコンテンツ系列選択方針に関するある種の知識と見なすことができ、ユーザの T 回前 (本

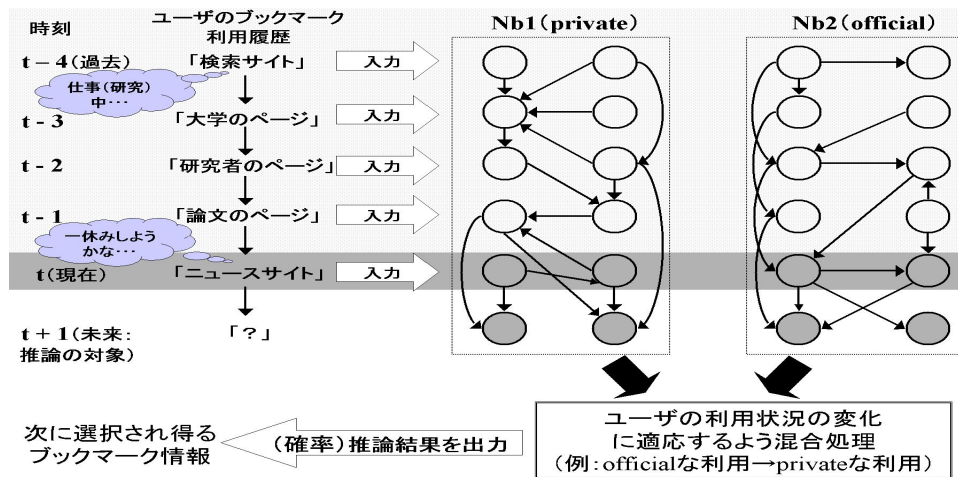


図1 ブックマーク情報推薦機構の基本的枠組

稿では $-T$ 回目と表現)から現在(0回目)までに選択したブックマークを基に、次(1回目)にユーザが選択するコンテンツを推論するために利用される(図2)。図2の右側において、BNの個々のノード(確率変数) $X_{0,-T}, \dots, X_{n,-T}$ は、ユーザが現在より T 回前に選択したブックマークフォルダを表している。 T 回前においてユーザがフォルダ X_1 のコンテンツ B_{10} を選択した場合、 $X_{1,-T} = \tilde{B}_{10}$ (\tilde{B}_{10} : コンテンツ B_{10} に対応する整数値)となり、他の変数 $X_{0,-T}, X_{2,-T}, \dots, X_{n,-T}$ には、“フォルダ未選択”を表す特定の整数値(例:-1)が代入されることとする。BNの構造決定に用いられるデータもこれと同様に表現され、過去 $T+1$ 回前から現在までのブックマーク選択履歴を表した多変量データとなる。

BNによって表現されるユーザのブックマーク選択方針を、現在の利用状況に応じて柔軟に活用可能とするため、本稿では、構造決定された各BNを用いた指数型BN混合モデルを導入する。指数型モデルは確率変数ベクトル $\mathbf{X} = (X_{0,-T}, \dots, X_{n,-T}, X_{0,-(T-1)}, \dots, X_{n,-(T-1)}, \dots, X_{0,1}, \dots, X_{n,1})$ として、以下で表される。

$$P_{BN}^{mix}(\mathbf{X}) = \frac{\prod_{i=1}^M P_{BN_i}(\mathbf{X})^{\alpha_i}}{\sum_{\mathbf{X}'} \prod_{i=1}^M P_{BN_i}(\mathbf{X}')^{\alpha_i}}$$

ここで M はモデルを構成するBN数、 α_i は混合パラメータ($\alpha_i \in \mathbf{R}, 1 \leq i \leq M, \sum_{\mathbf{X}} \prod_{i=1}^M P_{BN_i}(\mathbf{X})^{\alpha_i} < \infty$)である。混合モデルを構成する各BNは、ユーザが次に選択するブックマークを独立に推論可能である。しかしながら1.節でも述べた通り、ユーザは現在の自身の興味や嗜好、感情等によって動的にブックマークの選択方針を変更し得る。例えば、午前中まで仕事をしていたユーザが休憩に入り、これまでは別のフォルダに属するサイトを閲覧する可能性がある。また、その日の気分によって巡回の順序や、巡回するサイト自体も変化し得る。指数型BN混合モデルでは、ユーザの現在の利用状況に最も適合したブックマーク選択方針を表現するよう混合パラメータ α_i を調節することにより、ユーザが選択する可能性のあるブックマークを推論可能となる。加えて、ユーザの直近の利用履歴に対する現在の混合モデルの尤度変化を監視することで利用状況の変化を認識し、変化後の状況に対する迅速な対応も可能となる。指数型混合モデルと比較して、より一般的に用いられるモデルの一つとして、線形混合モデル($P_{BN}^{mix}(\mathbf{X}) = \sum_{i=1}^M \beta_i P_{BN_i}(\mathbf{X}); \sum_i \beta_i = 1; \beta_i \geq 0$)がある。線形混合モデルは指数型と比較して分布の計算が容易である反面、混合パラメータ β_i に関する制約が多い。指数型モデルは状況変化の認識に伴うパラメータの再学習とモデルの再構成に関して線形モデルよりも多くの計算量を要する反面、 α_i が負の値を取り得るため、BNに対応するブックマーク選択方針の特徴を“逆に”活用することが可能となる。このため、指数型混合モデルは同数のBNを用いて線形混合モデルよりも多様な情報を表現可能であることから、提案システムでは当面、指数型混合モデルのみを用いることを予定している。

本節の最後として、BN混合モデルを用いたブックマーク情報提示システムの動作の流れについて以下に簡潔に示す(ユーザの典型的な利用状況に対応するBNの構造は決定済みであるものとする)。

- (1) ユーザの直近 C 回のブックマーク選択結果を基に、現在のユーザの利用状況に対応した混合モデルを構成すべく、モデル中の BN_i における混合パラメータ α_i を調整。
- (2) 混合モデル、およびユーザの現在までのブックマーク選択履歴を利用して、ユーザが次以降に選択すると考えられるブックマークコンテンツを推論し、結果をユーザに提示。これらの作業と並行して、

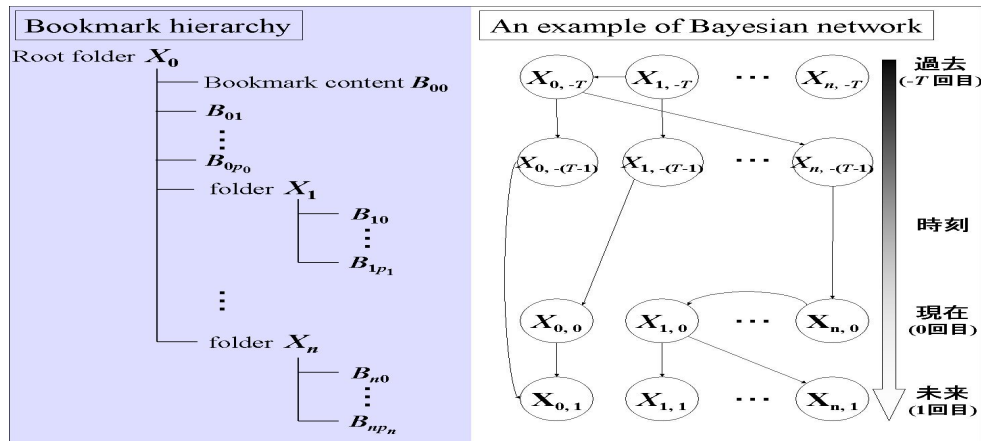


図2 ブックマーク構造とBNとの対応 (p_i : ブックマークフォルダ i に含まれるコンテンツ数)

- ブックマーク選択履歴から混合モデルの尤度を計算し、ユーザの利用状況が変化していないかを判定。
- (3) ユーザの利用状況の変化が認識された場合、変化後の状態に混合モデルを適応させるべく、ユーザの C' 回のブックマーク選択結果を用いて α_i を調整。
- (4) (2)へ戻る。

3. 関連研究および今後の方針

インターネット利用時における、ユーザの嗜好に応じた適切な情報提供のための研究は、インターネットの普及に伴い重要性を増してきている。提案手法と同様、ブックマーク情報を利用した主な研究例としては、複数のブックマーク情報間の関連性を利用したものが報告されている。佐保田ら[2]は、ブックマークに含まれる単語の出現頻度とブックマーク構造を利用した協調フィルタリングによるWebページ推薦手法を提案している。同様の情報を用いて濱崎ら[4]は、複数ユーザの持つ興味のうち、共通性のあるものを抽出する共通話題ネットワークを構築した。双方ともブックマークの階層構造を、ユーザの嗜好や関心を反映した情報と捉え、複数ユーザの保持するブックマークの類似性に注目してページ推薦や話題抽出を実行している。Jung[7]は、ブックマーク構造をBNと見なし、内容の類似したブックマークフォルダ同士をクラスタリングした上で、フォルダ間の関連性を考慮したユーザの嗜好検出を行った。一方、ユーザのアクセスパターンを解析し、ブックマークの自動追加やリンク切れコンテンツの削除、他者とのブックマーク共有、およびキーワードによるブックマーク自動分類等の機能を実装した拡張ブックマークシステム (PowerBookmarks) も報告されている[9]。また、丹羽ら[3]は近年普及し始めたソーシャルブックマークとその基盤となる Folksonomy の枠組みを利用したページ推薦システムの構築手法を提案した。Folksonomy ではユーザ自身がブックマークに内容を表すタグを付加するため、同内容のページへ付加するタグがユーザによって異なる“ゆれ”の問題が生じるが、タグ情報をクラスタリングすることによりこれを回避している。

ブックマーク以外を用いた研究例としては、文献[1]や[5]、[6]等がある。[5]では、ユーザのWeb巡回履歴を基にその嗜好や閲覧目的を解析し、ユーザの興味に沿った動的Webページ生成のためのプロトタイプシステムが提案された。Cadez[6]らも文献[5]と同様、Web巡回履歴を解析し、ユーザがどのような巡回経路を辿るのかを推論により示すシステムを構築した。双方とも履歴を基に動作するシステムとなっているが、[5]では履歴中のコンテンツ出現頻度や情報の新鮮度 (履歴中の出現場所) に着目したのに対し、[6]ではコンテンツの出現順に着目して1次マルコフモデルの混合モデルを導入し、その後のユーザの移動先を確率推論により求めている。内山ら[1]は、ユーザの閲覧したWebページで選択されたリンク元、およびリンク先周辺の文字情報からユーザの興味を抽出し、検索システムへの問い合わせとして用いることで、ユーザに対して有用なページを半自動的に提示する手法を提案した。

本稿で提案するブックマーク情報提示システムは、文献[2]や[4]、[9]のように他ユーザのブックマーク情報を用いず、ユーザ本人のブックマークとその巡回履歴のみを用いる。巡回履歴の利用という点で、提案システムはLiによって考案されたPowerBookmarks[9]と類似している。PowerBookmarksは様々な拡張機能の実現のため、Proxyサーバやデータベース等、多くの構成要素を組み込んだ大規模システムとなっているのに対し、提案システムは基盤となる数個のBNから構成され、これらの指数型混合モデルを用いることで、より効率的で広範な情報表現を可能としている。また、ソーシャル

ブックマークを用いた研究[3]のように、ユーザが事前に各ブックマークへタグを付与する必要もない。なお、文献[7]ではページ推薦のため、本研究と同様に BN を用いているが、具体的には非循環性有向グラフとしての構造的特徴のみを利用し、確率モデルとしての特徴は十分に活用されていない。

混合モデルの導入、確率推論の利用という点で、提案システムは文献[6]とも類似しているが、[6]における混合モデルは線形であり、情報表現能力は指数型モデルより劣る上、混合モデルの最大構成要素数はユーザのページ巡回パターン数と等しく、適切な要素数の設定を要する。提案システムにおいても適切な BN 数の設定は問題となり得るが、指数型混合モデルを利用することで線形混合より多様な情報表現が可能となるため、性能面で深刻な問題とはならないことが予想される。

提案システムは、他ユーザとのブックマーク比較を通じた知識発見や、新たな Web 情報推薦を目的とする従来研究とは異なり、ユーザの Web 巡回時に得られるブックマーク選択履歴情報のみをデータとし、不確実性や状況変化に対して有効な BN 混合モデルを用いて、ユーザ自身の現在有している知識（ブックマーク情報）の効率的利用の実現を目的とするところに特色がある。また、従来のブックマーク推薦技術やソーシャルブックマークの枠組と、提案システムを組み合わせることで、利用状況に応じて個人のブックマークを最大限活用しつつ、必要に応じて外部から適切な Web ページの推薦も受けられる、より効率的な Web 巡回支援システムの構築が可能となることが期待される。

提案システムを実現するにあたり、今後は主に (i) 推論結果の提示方法と(ii) データの収集法について検討が必要となる。(i) について、提案システムでは信念伝搬法等の確率推論法を用いてユーザが選択する可能性のあるブックマークを予測し、それを“予め選択されやすい状態”とすることで Web 巡回の効率化を図るが、予備実験の結果や他の文献等を参考に、ユーザにとって選択しやすい形式について調査する必要がある。現状では、予測結果をブックマーク表示領域の上部へ移動させることや、コンテンツを表示するテキストのサイズや色等を変更することによって注目を引く手法を検討している。また、予測結果のユーザへの提示範囲(上位何番目まで提示するか等)についても検討する。(ii) において、BN の構造決定用データはユーザが事前に Web 巡回した履歴から生成される。その際、ユーザに典型的な利用状況(工作中 or 休憩中)を想定してもらった上で、利用状況ごとにデータを収集する方法や、通常の使用状況と同設定でユーザに Web 巡回してもらい、得られたデータから各利用状況に対応するデータを分類する方法等が考えられる。前者は、所望のデータが容易に得られる反面、ユーザの実際の利用形態とは乖離している。対照的に後者は、日常的な Web 利用に即した状況でのデータ収集を可能とするが、得られたデータを利用状況に応じて分類するのは困難であり、巡回するブックマークコンテンツごとに、ユーザに“仕事”や“プライベート”等のタグを付与してもらう必要が生じることが予想されるため、より効率的、かつ実用に即したデータ収集法の考案が重要となる。

参考文献

- [1]内山紀明, 鈴木優, 川越恭二, 西村俊和, 利用者のリンク選択を考慮した Web ページ検索手法, 信学会第 17 回データ工学ワークショップ, 2006.
- [2]佐保田圭介, 波多野賢治, 宮崎純, 吉川正俊, 上村俊亮, ブックマークの階層構造を考慮した協調フィルタリングによる Web ページの推薦手法, 信学会第 15 回データ工学ワークショップ, 2004.
- [3]丹羽智史, 土肥拓生, 本位田真一, Folksonomy マイニングに基づく Web ページ推薦システム, 情報論, Vol. 47, No. 5, pp. 1382-1392, 2006.
- [4]濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継, Bookmark からの共通話題ネットワークの発見手法の提案とその評価, 人工知能学会誌, Vol. 17, No. 3, pp. 276-284.
- [5]福村真哉, 中野賢, 春本要, 下條真司, 西尾章治郎, 視聴履歴に基づくデジタルコンテンツの個人化された提示手法, 信学会第 13 回データ工学ワークショップ, 2002.
- [6]I. Cadez, D. Heckerman, C. Meek, P. Smyth, S. White, Visualization of Navigation Patterns on a Web Site Using Model Based Clustering, Data Mining and Knowledge Discovery, Vol. 7, pp. 399-424, 2003.
- [7]J. J. Jung, J. Yun, G. Jo, Collaborative Information Filtering by Using Categorized Bookmarks on the Web, Proc. 14th Int. Conf. Applications of Prolog, pp. 343-357, 2001.
- [8] D. Kitakoshi, H. Shioya, and R. Nakano, Stochastic Information Expressed in an Exponential Mixture Model of Bayesian Networks, Proc. SCIS & ISIS 2006, pp. 636-643, 2006.
- [9]W. Li, Q. Vu, D. Agrawal, Y. Hara, H. Takano, PowerBookmarks: A System for Personalizable Web Information Organization, Sharing and Management, Computer Networks, Vol. 31, No. 11-16, pp. 1375-1389, 1999.