

Web 情報検索における文脈依存型ブックマーク

4V-1

中島 伸介[†] 黒田 慎介[†] 田中 克己[†]

†京都大学大学院情報学研究所

1. はじめに

ユーザは、Web ページを閲覧した際に、気に入ったページが見つければ、その URL をブックマークとして残す、という行動を取る。通常、このブックマークには、URL、ブックマーク名、日付のみが記録される。

しかしながら、ユーザが採用した検索キーワードや、ブックマークするまでの巡航履歴等の、ブックマークまでの過程（文脈）にこそ、そのブックマークの意味や価値が隠されていると考える。したがって、URL やブックマーク名等の従来のブックマークデータのみならず、ブックマークが持つ“文脈”データを取得することにより、そのブックマークの意味的内容や価値を表現することが可能となり、新しい利用方法の提案や再利用性の向上が期待できると考えた。

そこで本報告では、この“文脈依存型ブックマーク”という概念を提案し、その利用方法と文脈の取得方法について述べる。

2. 文脈依存型ブックマークとその利用

2.1 文脈依存型ブックマーク

ここでいうブックマークの文脈とは、ブックマークしていないものも含めて、ユーザが閲覧した Web ページ集合の内容・閲覧順序・滞留時間や、ユーザの目的、ユーザが入力した検索キーワード、ブックマークに対するユーザの満足度や確信度、さらにはブックマークしたユーザのプロファイルなどを含めたデータ等によって示される、ブックマークの背景および過程を表すものである。

以上から、文脈依存型ブックマークには、URL・文脈データ・ブックマークのランキング値を保持させたものとする。（表 1 参照）

表 1 文脈依存型ブックマーク

①URL データ	②文脈データ
③ブックマークのランキング値	

2.2 文脈依存型ブックマークの利用

・ブックマークランキング値の呈示

文脈データの解析により、そのブックマークの

ランキング値を算出する。これは、ユーザが閲覧した Web ページ集合の中で、ブックマークを付けたページに対するその他のページの類似度と、ブックマークまでの閲覧した Web ページの数に基づいて算出する。（→ 3.1 節参照）

・ブックマーク推薦システムへの利用

文脈データの解析により、ユーザの潜在的な要求を取得し、ユーザに対してブックマークの推薦に利用することが可能である。（→ 3.2 節参照）

・個人の User Profile 自動作成への利用

情報フィルタリングに利用可能な、個人の User Profile を、文脈データおよび閲覧履歴の解析により自動作成する。

・協調作業時の情報共有システムへの利用

ブックマークの文脈とそれに依存した価値を呈示することにより、他のユーザがブックマークを容易に理解することができるため、情報交換や情報共有システムに利用できる。

3. 文脈依存型ブックマークのアルゴリズム

2 節にて、文脈依存型ブックマークおよびその利用方法について述べた。本節では、実際のアルゴリズムについて説明する。

3.1 ブックマークランキング値の算出

ユーザが Web ページ集合 S を閲覧した後に、ある Web ページ x にブックマークを付けたという状況を想定する。ランキングの基礎データとしては、ユーザが閲覧した Web ページ集合内で、ブックマークを付けたページに対するその他のページの類似度の平均と、ブックマークまでに閲覧した Web ページの数に基づいて算出する。ブックマークのランキング値を R_B とすると、 R_B は以下の式で定義される。

$$R_B(x, S) = (1 - \alpha) \cdot |S| + \alpha \cdot \frac{\sum_{y \in S - \{x\}} sim(x, y)}{|S| - 1}$$

ただし、 $sim(x, y)$ は Web ページ x と y の類似度

第 1 項は、ブックマークまでに閲覧した、Web ページの量を示す。第 2 項は、ブックマークを付けたページに対するその他のページの類似度の平均値である。すなわち、ブックマークするページを選択する際に、その Web ページと類似したページを多く閲覧している

場合には、よりシビアな選択を行っていると仮定し、この場合にブックマークのランキング値が高くなるようにしている。ただし、算出式中の α の値に関しては、検討が必要である。

3.2 ブックマーク推薦システムへの利用

文脈データの解析により、ユーザの潜在的な要求を取得し、ユーザに対してブックマークの推薦に利用する方法を提案する。

まずは、検索キーワードを、ユーザが検索エンジンに与えて、検索結果の中から幾つかの Web ページを閲覧した後に、気に入ったページのブックマークした状況を想定する。

ユーザが入力した初期検索キーワードを

$\{t_1, t_2, \dots, t_n\}$ とする。また、検索結果の中から、ユーザが実際に閲覧した s 個の Web ページを $\{P(1), P(2), \dots, P(s)\}$ とする。この中でユーザが、ブックマークをつけた Web ページを x とする。

ここで、共通のキーワードを含む s 個の Web ページの中から、ユーザが x を選んだのは、Web ページ x に対して他の Web ページと違う特徴を見出したからである。そこで、ユーザが閲覧した s 個の Web ページをその特徴で差別化することにより、この中から x を選択した際にユーザが目にした特徴量を抽出する。この特徴量こそが Web ページ x を選択したユーザの意図を反映したものであると考える。

s 個の Web ページの差別化の方法としては幾つか考えられるが、IDF の範囲を s 個の Web ページに限定した TF-IDF によるキーワード抽出も有効な方法である。この場合、 x の固有のキーワードを、 w_x とすると、ユーザがブックマークした Web ページの特徴は、“検索キーワード $\{t_1, t_2, \dots, t_n\}$ によって、検索された Web ページ集合の中で、キーワード w_x に特徴付けられる Web ページ” であるといえる。

この文脈データの解析により、ユーザの潜在的な要求を満たすような Web ページの新たなキーワードを提示することができる。例えば、 $\{t_1, t_2, \dots, t_n\}$ であった検索キーワードに対して、新たなキーワード w_x を追加して、 $\{t_1, t_2, \dots, t_n, w_x\}$ が新たな問合せとなり、前回よりも絞り込まれた検索を実現する。この作業を

繰り返すことにより、ユーザの意図を反映したより複雑な問合せを導出することができる。そして最終的にはユーザの要求に近い Web ページに到達することにより、新たなブックマークを推薦することができるのである。

検索エンジンを用いた情報検索時を想定して説明したが、自由に Web ページの巡航を行った後にブックマークした場合においても、ブックマークした Web ページに関連して閲覧したページを限定できれば、同様に文脈データを取得し、ブックマークを推薦することが可能である。比較的関連性の弱い文脈ではあるが、日ごとの履歴データを 1 つのブックマークセッションとすることも可能であると考えている。

3.3 文脈依存型ブックマークの問題点

- ① TF-IDF によって抽出された Web ページの固有キーワードが必ずしも Web ページの意味内容やユーザの意図を反映したものととは限らない。
- ② ブックマークした Web ページに対するユーザの満足度を文脈に反映させることができない。
- ③ Web ページを自由に巡航した後にはブックマークした時の文脈情報が十分とはいえない。

①については、閲覧した Web ページ集合を差別化する方法は、TF-IDF である必要はないので、Web ページの意味内容をより反映させることができるような手法を検討する。

②については、ブックマークする際にユーザにブックマークに対する“確信度”のようなものを入力させることも検討する。

4. おわりに

ユーザが Web ページに対しブックマークを付ける際の“文脈”を取得かつ解析することにより、そのブックマークの意味的内容や価値を表現することを可能にする、“文脈依存型ブックマーク”を提案した。

関連研究としては、ユーザの Web ページへのアクセス履歴からブックマークの推薦を行うもの^[1]や、ユーザの Web 巡航履歴を抽出するもの^[2]がある。

今後は、提案した手法の検証や、情報共有システムへの拡張に関する検討を行う予定である。

参考文献

- [1] Wen-Syan Li, et al : PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management. WWW8 / Computer Networks 31(11-16): 1375-1389 (1999)
- [2] Vinod Anupam, et al : Automating Web navigation with the WebVCR. WWW9 / Computer Networks 33(1-6): 503-517 (2000)

表 2 TF-IDF によるキーワード抽出

閲覧した Web ページ	ブックマーク	TF-IDF により抽出されたキーワード				
		w1	w2	w3	w4	w5
P(1)	—					
P(2)	—					
P(3)	○					
P(4)	—					
P(5)	—					