

WWWにおけるユーザの共通話題ネットワークの発見手法

Discovery of Shared Topics Network among Users by using WWW bookmarks

谷口雄一郎[†], 松塚健^{†,1,2}, 武田英明^{††,†}, 木戸出正継[†]

Yuichiro Taniguchi, Takeshi Matsuzuka, Hideaki Takeda, Masatsugu Kidode

奈良先端科学技術大学院大学 情報科学研究科[†]

Graduate School of Information Science, Nara Institute of Science and Technology

国立情報学研究所^{††}

National Institute of Informatics

Abstract: In this paper, we propose a technique to find a shared topics network(STN) among users and describe kMedia system based on this technique. STN enables us to find shared interested topics each other. We define WWW browser's bookmark folders as topics. STN is visualized with the undirected graph of related folders. The relationship among folders is computed by the similarity among web pages included in folders. The similarity among web pages is computed by measuring shared keywords included in these web pages. We have conducted an experiment with kMedia system and confirmed this technique correctly finds relationship among users.

1 はじめに

我々が必要な情報を獲得する際、その情報の内容に応じてアクセスする情報源を適切に選択し、効率良い情報獲得を行っている。同様に、人とのコミュニケーションを通して情報を獲得する場合もその情報に興味をもつ人を捜し出し、質問、議論が出来ることが望ましい。そのためには、人がどういった情報に興味をもっているか、またその興味と他人の興味がどのように繋がっているかを知ることが重要である。そこで、本研究では人の興味とその間の関係を発見する手法を提案し、その有効性を検証する。

2 話題の共通性の発見

我々は人が興味をもつ話題間の共通性に注目する。ある話題について人から情報収集するとき、共通性のある話題に興味をもつ人々に質問すると効率的な情報収集が行える。また、共通する話題に興味をもつ人と人との関係を発見することによって、その話題に関する新たなコミュニティの形成も期待できる。

話題の共通性の発見には、1) 各人の興味ある話題の発見が難しい、2) 話題による情報分類が人によって異

なる、3) 複数の人の話題の共通性を調べるのが困難、といった問題がある。そこで本研究では、各人が興味を持つ話題に関する知識を WWW ブラウザのブックマークから抽出し、それらの共通性を調べる。さらに、各人の話題の共通性をわかりやすく、かつ興味ある話題を正確に把握するために、共通話題で結ばれたユーザ間のネットワークを発見する手法を提案する。

共通話題ネットワークとはそれぞれのユーザが興味を持つ話題とそれらの間の共通性を表現した無向グラフである。ユーザはユーザノード、話題は話題ノードで表現される。ユーザが興味をもつ話題は、ユーザノードから話題への一本のリンクで表現され、話題の共通性は話題と話題の間のリンクで表現される。

本研究ではブックマークフォルダ名をユーザが興味を持つ話題とみなし抽出する。ブックマークフォルダは、ユーザがある視点からの共通性に基づき分類した複数の URL を保持する。フォルダ名はユーザによって付与されるため、保持する URL の話題を適切に表現できていると考えられるからである。

3 kMedia システム

提案手法を kMedia システム (以下、kMedia) に実装した。kMedia はサーバクライアントシステムを構成する。クライアントはユーザのブックマークデータの解析、及び結果の出力を行い、サーバは WWW ページの共通性、ユーザ間の話題の共通性の計算を行う。

まず、クライアントはユーザにブックマークデータを要求する。次にブックマークデータ中の各々の URL が指す WWW ページへアクセスし、含まれるキーワード

⁰連絡先: 〒 630-0101 生駒市高山町 8916-5

Tel (0743)72-5265 Fax (0743)72-5269

mail: yuichi-t@is.aist-nara.ac.jp

http://ai-www.aist-nara.ac.jp/

¹現在、凸版印刷株式会社マルチメディア事業部

²Currently, TOPPAN PRINTING CO., LTD. MULTI-MEDIA DIV.

を抽出し、キーワードとその出現頻度情報をブックマークデータに追加する。キーワード抽出には文字種から判断して切り出す手法 [1] を用いる。キーワード情報を追加したブックマークデータをサーバに送信する。

サーバは受け取ったブックマークデータから WWW ページ間の共通性、及び話題の共通性を計算する。WWW ページ間の共通性は、二つの WWW ページの対で共有するキーワードの個数で決定される。キーワードの個数がある閾値以上である場合に二つの WWW ページの間に共通性があると判断し、その WWW ページ対にマーク付けを行う。ユーザのブックマークデータにある WWW ページの集合から取り得る全ての WWW ページ対にこの処理を行う。

話題の共通性は以下のように決定される。前述の処理で共通性があると判断された WWW ページ対の一方があるユーザのブックマークフォルダに含まれ、他方が他のユーザのブックマークフォルダに含まれている場合、両者のブックマークフォルダから成るフォルダ対にマーク付けを行う。マーク付けされたフォルダ対の個数がある閾値以上の場合、そのフォルダ対の間に話題の共通性があると判断する。

最後に、WWW ページ間の共通性からは関連する推薦 WWW ページ情報を生成し、話題の共通性からは共通話題ネットワークを生成する。推薦される WWW ページは他のユーザのブックマークデータから選ばれる。これらの情報をクライアントに送信する。

クライアントは、受け取った情報から共通話題ネットワークと推薦 WWW ページを表示する。図 1 は kMedia によって提示された共通話題ネットワークの一例である。この図からは、ユーザ 'nishimori' の話題 '研究関係' とユーザ 'matsuzuka' の話題 '学術系' の間の共通性、また 'Linux' と 'Unix' の間の共通性などを読み取ることができる。

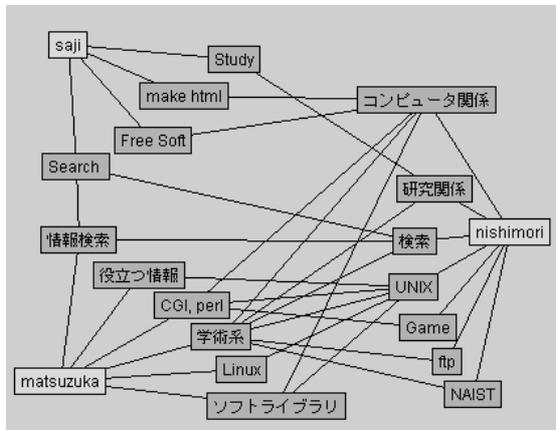


図 1: 共通話題ネットワークの一例

4 評価実験

kMedia が提示した話題、WWW ページの共通性が、ユーザにとって有益なものであるかについて評価実験を行った。被験者のブックマークデータを kMedia に与え、得られた共通話題ネットワーク、および推薦 WWW ページに対し、以下の三点から評価させた。評価は 5 段階で行い、最良の結果には評価 5 を、最悪の結果には評価 1 を付けさせた。

1) 推薦 WWW ページの関連性 自分のブックマークに対してシステムから推薦されたすべてのブックマ

クの関連性を 1 つずつ評価させた

2) 話題の共通性 自分の話題と他人の話題との関連性³について評価させた

3) ユーザ間の関連 システムが提示した共通話題ネットワークによって結ばれた、自分と他のユーザー一人一人の話題の共通性が妥当であるかを評価させた

表 1 に被験者のブックマークデータと kMedia の結果を、表 2 に話題の共通性の評価を示す。推薦 WWW ページの評価は評価 1 が最も多かった。ページどうしを関連付ける話題が抽象的な情報であり、文章中に明示的に現れることが少ない。そのため、文章解析による関連性の発見が難しく、高い評価を得られなかったと考えられる。

一方、話題の共通性については表 2 にみられるように評価 5 が最も多く、話題の共通性発見において有効性を示している。この結果は、個々の情報に対する粗い理解を通して得られる、話題の共通性や人どうしの関連性と言った抽象度の高い情報が人間にとって有意なものとなる、という本手法の性質を示している。

	ユーザ A	ユーザ B	ユーザ C
解析ページ数	263	185	240
話題数	13	17	5
推薦ページ数	345	513	454
共通話題数	10	10	3

表 1: 各被験者のブックマークデータと解析結果

	評価 5	評価 4	評価 3	評価 2	評価 1
ユーザ A	3	3	2	0	2
ユーザ B	3	3	2	2	0
ユーザ C	2	0	1	0	0
合計	8	6	5	2	2

表 2: 話題の共通性の評価

5 結論

本研究では、ユーザ間の情報交換を促進するために共通話題ネットワークの発見手法を提案し、kMedia システムに実装した。kMedia の評価実験をおこない、本手法の有効性を示した。

提案手法では、共通話題ネットワークの発見にユーザのブックマークデータを使用した。評価実験の結果から、このような人のもつ知識と機械的な関係発見の処理の組み合わせが効果的といえる。

今後は、評価の低かった WWW ページの推薦にテキスト処理の工夫や他の手法との併用を施すことによって、個々の情報の類似性の発見を強化する必要がある。

参考文献

- [1] 片岡 充照, 今中 武, 水谷 研治, 若見 昇. テキスト情報を対象としたキーワード抽出と関連情報収集システム. 日本ファジィ学会誌, Vol.9, No.5, pp.710-717, 1997.

³ 共通話題ネットワーク上で、自分の話題から一本のリンクで結ばれた他人の話題すべて