

1N-05 文献の線形順序づけに従う効率的なブラウジングのための ユーザインタフェース

鈴木 雄清 佐藤 慶三 曲 艶華 伊藤 哲郎

大分大学工学部

1.はじめに

従来から検索システムの多くは、検索結果の文献を質問との関連の高い順にランキングし提示する方法を採用している[1,2]. また、質問と文献間に加え、文献相互間の関係を考慮に入れてクラスタ化し提示する方法も提案されている[3].

我々は、これら双方の利点を取り入れ優れた検索効率が期待できる Acl ブラウジング (Adjacent cluster based browsing) 法を提案してきた[4]. この方法では、文献をそれら相互間の類似度をもとに線形順序づけた後クラスタ化しておく. そしてブラウジング時は、質問との類似度が大きい文献を含むクラスタおよび、これらクラスタの隣接クラスタを逐次調べてゆく形になる. 本報告では、図書館を1次元的に配架している図書館に倣った仕方で、Acl ブラウジングが効果的に行えるようなユーザインタフェースを提案する.

2.検索結果の線形順序づけ

Acl ブラウジングでは文献の線形順序づけが基本となる. 線形順序づけには、文献数 N の検索結果に対し、反射率、対称率を満たす $N \times N$ の大きさの文献間類似度行列 $[s_{uv}]$, $1 \leq u, v \leq N$, から、加えて推移率を満たすような推移行列 $[t_{uv}]$ を求める. そして推移行列を用いて、行(列)が離れるほどできるだけ値が小さくなるよう、もとの類似度行列を並べ替える. 質問も文献の1つと考えると、質問を含め関連の高い文献同士が近くなるような線形順序づけがなされる[5].

推移行列を求めるにあたっては、コスト問題が生じる. これには、文献集合に対してあらかじめ推移行列を求めておき、質問を後で組み込むことで対処する.

3.Acl ブラウジング法

Acl ブラウジング法の流れを以下に示す.

(S1) 線形順序づけされた文献集合に対し、ブラウジング単位としての連続した順にある関連の高い文献の集り、すなわちクラスタ、を求める.

(S2.a) 質問との類似度が高い文献を含むクラスタをランキングし、ランキング順に適合判断する. このとき質問との類似度が最も高い文献、すなわちクラスタの代表、の適合判断を、そのクラスタすべての文献に対する適合判断と見なす.

(S2.b) 適合文献を含むクラスタの隣接クラスタの適合判断も同時に行う.

4.ユーザインタフェース設計

Acl ブラウジングに従うと、効率の良い検索が可能になることが計算機シミュレーションを通じて示されている[4]. しかしながら、この方法をユーザ主導の対話的な環境で利用するには、インタフェースの設計が重要になる. 線形順序づけを反映する文献の全体的な配置に際しては、できるだけ多くの文献のタイトルを同時に表示できるようにしなければならない. ここでは現実の図書館のような、本が左から右へ、上から下へと1列に並べられた本棚状の形態の利用を考える.

4.1.視覚化法

各文献の視覚化にあたっては、タイトルで表わす. 文献同士の相互関係は、線形順序で u 番目と $u+1$ 番目 ($1 \leq u \leq N-1$) の文献間の類似度 $s_{u, u+1}$ に反映されている. これを距離に変換し、求めた値に従って各文献のタイトルを1列に配置する.

計算機シミュレーションから予測されたしきい値 α によってあらかじめデフォルトのクラスタを決め、薄い色をつけておく. ユーザは後で、自身の望む大きさのクラスタに変更してよい.

質問との類似度は、各文献タイトルと共に提示されるアイコンに、質問との類似度に従う色をつけて表現する. さらに、各クラスタにおける代表はアイコンの形状を変えて示す. これらにより、ユーザは質問と関連の高いクラスタの代表文献を容易に見つけられる.

4.2.クラスタの設定

ユーザが望む大きさのクラスタに変更するための機構を設ける. ユーザはデフォルトのクラスタを任意の位置で分割したり、前後クラスタを結合することで全体的なクラスタ化のしきい値 α を変えることができる. このとき、大幅にクラスタが変更されてユーザを混乱させることのないよう、クラスタのしきい値の変更は最小限に留めるようにする.

Designing User Interface for Efficient Browsing of Linearly Ordered Documents

Yusei Suzuki, Keizo Sato, Yanhua Qu and Tetsuro Ito
Faculty of Engineering, Oita University
700 Dannoharu, Oita, 870-1192 Japan

4.3. ブラウジング支援

ユーザは質問との関連が高い文献を含むクラスタ C_u を適合と判断した場合、その隣接クラスタ C_v , $u+1 \leq v$ (および C_v , $u-1 \geq v$) についても適合判断する。上記の視覚化法により、クラスタ間の関係は空白の幅に反映された形となっている。

ここで、隣接クラスタをブラウジングする際のしきい値 β ($\beta \leq \alpha$) を設け、 $s_{v-1,v}(s_{v+1,v})$ が β より大きい間ブラウジングする。 β の値は、計算機シミュレーションから予測しておく。 $s_{v-1,v}(s_{v+1,v})$ が β の値より小さいクラスタ間には、空白に色をつける。これにより、ユーザは隣接クラスタをどこまでブラウジングすればよいかを視覚的に判断できる。

ユーザがまだブラウジングしていないクラスタのうち、質問との関連が最も高い文献を含むクラスタが収められた棚を画面中央に表示する機構も設ける。ユーザがこの隣接クラスタをブラウジングすることを想定し、すべてが画面上に1度に表示されるようこれら棚の数を調整する。

4.4. プロトタイプ

上記で述べた仕方にに基づき、ユーザインタフェースのプロトタイプシステムを JAVA 言語で作成した。スクリーンショットを図1に示す。

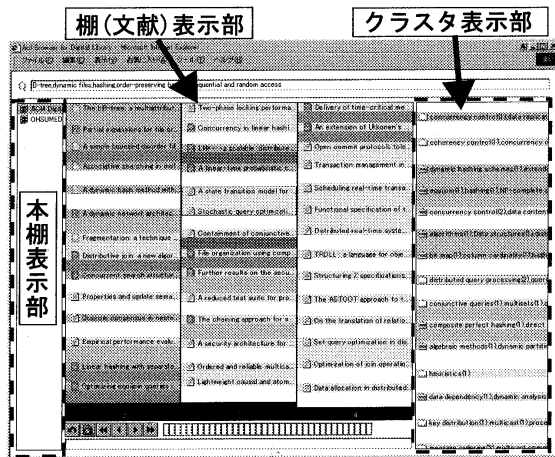


図1. Acl ブラウジングのためのユーザインタフェース

ユーザの趣味や専門分野など関連性の高い一連の文献集合を本棚として提示する。本棚の各棚では、1つのクラスタがなるべく2つの棚にまたがらないように配置する。さらに、各クラスタ中の文献が持つキーワードを出現頻度順に羅列して提示し、ユーザの適合判断を支援する。

5. 評価実験

文献の線形順序付けを固定しておく、質問によっては、質問との関連の高い文献を含むクラスタが分散して配置されてしまうことになる。ブラウジ

ング時、ユーザは棚の間を前後、左右に大きく移動しなければならない。これに対処するため、質問が与えられる度にクラスタの並べ替えを行う。

並べ替えの単位は、しきい値 β より大きな $s_{v-1,v}(s_{v+1,v})$ を持つ文献を持つ隣接クラスタの集りで、それに含まれる文献と質問との類似度順に行う。この方法の利点は、質問が違ってても、隣接クラスタの集りの中での文献配置は変わらない。また、ブラウジングする隣接クラスタの順序も変わらず、ブラウジングされるそれぞれのクラスタも同じになる。

並べ替えの効果を確かめるため、並べ替え前と後の Acl ブラウジングでのクラスタ間の移動数を再現率がそれぞれ 0.2, 0.3 になるまで調べた。但し、隣接クラスタの移動は数えないものとする。対象データには ACM 関連雑誌に掲載された論文約 500 篇を用い、任意の文献 12 を質問として与えた。質問に対する適合文献は、用語間の概念的な関係を考慮した類似度で高い値を与えるものとした[6]。

結果を表1に示す。並べ替えによってブラウジング時のクラスタ間の移動が大幅に減少し、効率良くなっていることがわかる。

	並べ替え前		並べ替え後	
	0.2	0.3	0.2	0.3
Q1	1050	1108	19	44
Q2	887	1840	69	329
Q3	1641	1847	68	130
Q4	1643	2554	61	93
Q5	817	3128	66	232
Q6	489	1839	25	96
Q7	713	1015	22	42
Q8	358	1505	44	122
Q9	484	917	26	44
Q10	138	181	20	42
Q11	127	236	28	51
Q12	938	985	31	50
平均	773	1429	39	106

表1. 並べ替えによる Acl ブラウジングでのクラスタの移動数の変化

6. おわりに

文献の線形順序づけに従う Acl ブラウジングのためのユーザインタフェース設計について述べた。今後の課題として、プロトタイプシステムの有効性の検証や、WWW 上に実装するための改良が考えられる。

参考文献:

- [1] AltaVista, <http://www.altavista.com/>
- [2] Infoseek Japan, <http://www.infoseek.co.jp/>
- [3] Marti Hearst and Jan Pedersen, Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, Proceedings of the 19th Annual International ACM/SIGIR Conference, Zurich, August 1996.
- [4] 曲艶華ほか: 電子図書館における効率的な探索のための検索結果配置, 平成10年度電気関係学会九州支部連合大会, 539, 1998.
- [5] 伊藤哲郎: 情報検索, 昭見堂, 1986.
- [6] 中島誠ほか: 用語間の概念的関係を考慮した測度による文献のランキング, 電子情報通信学会論文誌 D-1, Vol. J82-D-1, No. 3, 1999.