利用者がレレバンス判定に費やす時間

安形輝（亜細亜大学） agata@asia-u.ac.jp

【抄録】レレバンス判定において、判定結果と判定時間の関係に焦点を当てた研究が近年、行われている。しかし、判定結果と判定時間の関係については実験結果が分かれており、判定尺度や提示順を適切に扱ったかには疑問が残る。本研究では、判定尺度や提示順を考慮に入れた大規模な実験を行うことで既往研究よりも明確に判定結果と判定時間の関係を示すことができた。

1. はじめに

1.1 レレバンス判定に費やす時間の既往研究

利用者がレレバンス判定に費やす時間については、主としてレレバンスフィードバックの領域で研究されてきた。情報検索においてレレバンスフィードバックの有効性は SMART システム[1]の時代より明らかであった。インターネットの普及により利用者層が拡大するにつれ、単純な検索式が用いられるようになってきたが、そのような検索でのより特定的な検索式への自動修正にもレレバンスフィードバックは有効であると考えられる。しかし、明示的に利用者に検索結果を評価させる検索システムで実用化されたものは、利用者の認知負荷が高く、成功したものはほとんどない[2]。その反省から、近年、暗黙的な手法（implicit method）により利用者のレレバンス評価や文献に対する関心度を測定する方策が検討されるようになってきた。暗黙的な手法に使われる手がかりとしては、検索結果の閲覧のさいに費やされる時間、スクロール操作、視線の軌跡、マウスの操作などがある。

閲覧時間に関する既往研究としては、情報フィルタリングの領域で Morita らによる研究[3]があるが、より特定的な検索質問に対する情報検索への応用を前提としたものとしては①Kellyらの研究[4]、②Whiteらの研究[5]、③Kellarらの研究[6]がある。以下では①～③の記号で参照していく。

1.2 既往研究の課題と本研究の目的

既往研究には「利用者はレレバントな文献に対する判断にはより長い時間を費やす」という仮説が統計的に確認されたもの（②、③の試行2回目）と確認されなかったもの（①、③の試行 1 回目）がある。異なる結果が得られた原因の一つとして、どの研究も比較的小規模なものであったこと（被験者は①6 人、②16 人、③10 人）が考えられる。

また、レレバンス判定研究では、判定尺度や文献提示順が判定結果に影響を与えることが指摘されてきた。判定尺度については Eisenberg[7]や Bruce[8]が指摘し、レレバンス判定では自由度の高いマグニチュード推定法が望ましいという結果を得た。また、文献の提示順の影響に関しては Eisenberg ら[9]や安形[10]が指摘してきた。この二つの研究では、被験者はレレバンス評価の高いものから低いものへという順序で提示されると、レレバンスの高い文献を過小評価し、逆順の場合、レレバンスの低い文献を過大評価する傾向が見られた。しかし、提示順の影響に関して既往研究①～③で適切に扱われてきたとはいえない。

そこで、本研究では、既往研究で異なる結果となった判定時間と判定結果に関する仮説をより明確に検証するために、大規模なレレバンス判定の実験を行った。また、独自のシステムを用いることで、判定尺度や文献の提示順の影響について分析できるようにしている。

2. 実験環境

レレバンス判定に関する実験は 2004 年 9 月、2005 年 9 月、2006 年 9 月の3回行った。これらの実験は、被験者、検索質問数、提示する文献数以外は同じ条件下で行われた。ここでは 2004 年の実験を中心に結果を紹介する。

2.1 被験者

被験者は亜細亜大学夏期司書講習の情報検索演習において「インターネット検索」の回を受講済みの受講生から募集した。

2.2 検索質問と文献集合

検索質問には利用者の欲しい文献、その背景、検索に使われたキーワードが記述されている。

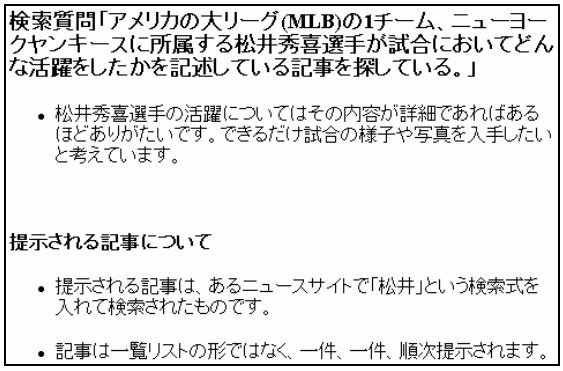

図1 検索質問例

レレバンス判定を行った文献集合は、インターネット上で入手できる新聞記事から構成されている。これらは、実験前に Yahoo!ニュース検索[11]、Google News Beta[12]からキーワードで実際に検索されたものを使用している。一部の記事には写真も含まれている。なお、記事内容は改変していないが、含まれるキーワードが判別しやすいよう、黄色く色づけし表示するよう加工した。

検索質問数と提示した文献数は年毎に異なり 2004 年は 1 問 10 件、2005 年と 2006 年は検索質問3問と提示する文献 18 件(6 件×3 問)から構成されている。

2.3 判定尺度

判定尺度としては、以下の三尺度を用意し、検索質問ごとに無作為に割り当てた。

①5段階尺度

多くの検索実験において伝統的に用いられてきたカテゴリの尺度であり、ここでは、レレバンスの低い方から「全く適合せず」「あまり適合せず」「どちらともいえない」「だいたい適合」「非常に適合」までの5段階としたラジオボタンを用いた(図2)。

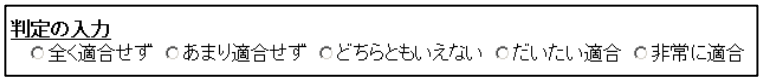

図2 5段階尺度

②スライダー

マウスを使ってハンドルを移動することで、レレバンス判定の値として 0 から 100 までの値を表現することができる(図3)。

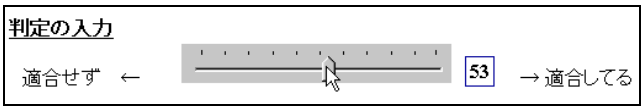

図3 スライダー

③マグニチュード推定法

範囲を決めない数値を記入させるため、制限の少ないテキストボックスを用いた(図4)。

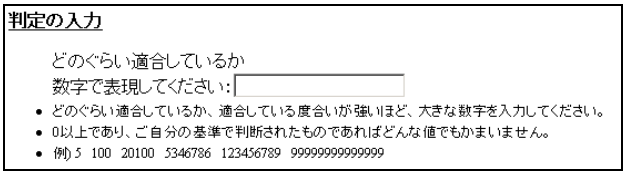

図4 マグニチュード推定法

2.4 提示順序

各検索質問について、調査者があらかじめ判定した「レレバント」から「非レレバント」の順に文献の順序を並べ替えておいた。被験者に対してはこの順序に基づき、検索質問ごとに「レレバント」から「非レレバント」、「非レレバント」から「レレバント」、「混合」という三つの提示順序を機械的に割り当てた。

2.5 実験システム

本実験のために、ウェブ上からアクセス可能な実験システムを独自に構築した。このシステムにより、以下のことが可能となった。

・同時に多数の被験者に対する実験
・3 つの尺度、3 つの提示順の無作為な割当て
・レレバンス判定にかかる時間の記録

2.6 実験手順

レレバンス判定実験は以下のような 7 段階から構成される手順で行った。

① 被験者がウェブ上の実験システムへアクセス
② レレバンス判定実験の説明
③ 年代、性別等の基本的な属性の入力
④ 検索質問の説明
⑤ 文献ごとのレレバンス判定
⑥ 規定の検索質問が終わっていなければ④へ、終わっていれば⑦へ
⑦ 実験への協力の謝辞の表示

3. 実験結果

3.1 基本的な属性

被験者の基本的な属性を表1に示した。20 代が多く、年代があがるにつれ被験者数は少なくなっている。

表1 被験者の属性

		性別		
		女性	男性	不明
年代	20代以下	72	26	0
	30代	27	5	1
	40代	14	7	0
	50代以上	7	4	0
小計		120	42	1
総計		163		

単位:人数

表2 提示順、尺度別査被験者数

		提示順			合計
		レ⇒非	非⇒レ	混合	
尺度	5段階	18	17	18	53
	スライダー	17	19	19	55
	マグニチュード	18	18	19	55
合計		53	54	56	163

単位:人数

各被験者に検索質問と文献集合が提示されるときに、実験システムは無作為に提示順、尺度を割り当てる。その分布は表2のようになった。この表で、「レ⇒非」はレレバントな文献から非レレバントな文献を提示した場合、「非⇒レ」はその逆順に提示した場合、「混合」はレレバントな文献と非レレバントな文献を混ぜて提示した場合である。

3.2 被験者の属性と判定時間

表3は被験者の属性と各文献の判定時間平均の関係を表したものである。被験者は全体として1文献を平均 21.4 秒で判定したことになる。

性別から見た場合、女性よりも男性が短い時間で判定を行うことがわかる。また、年代別では 20 代の被験者は他の年代より短い時間で判定を行うことがわかる。

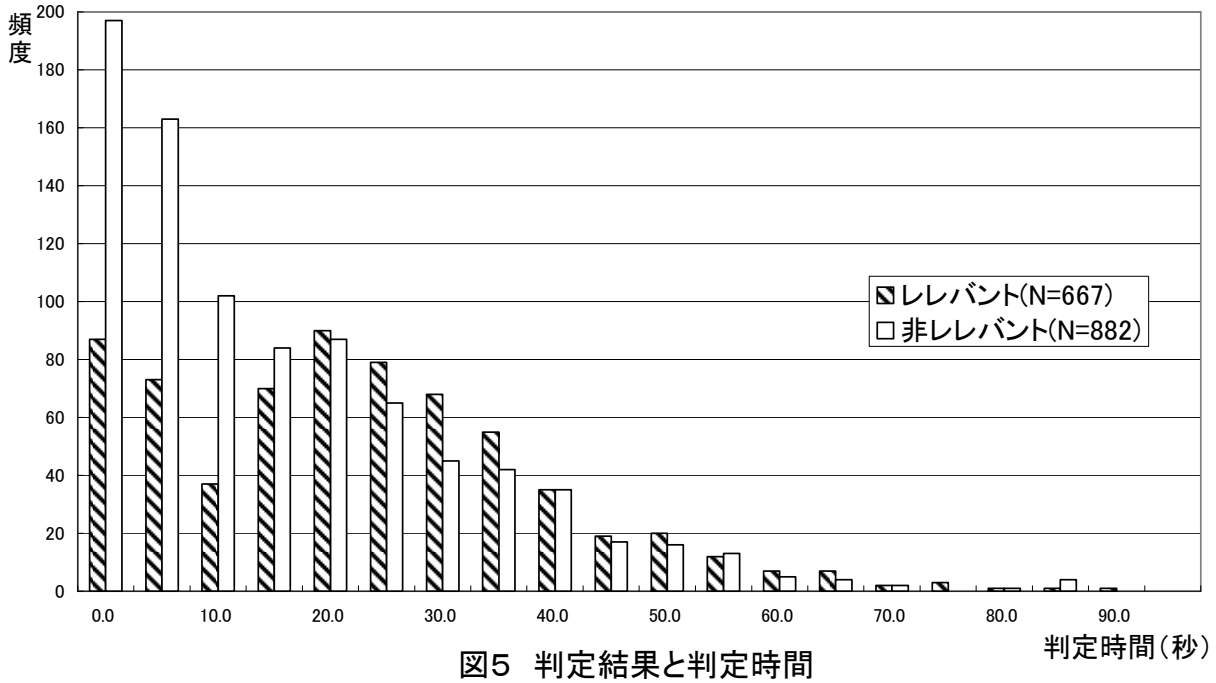

図5 判定結果と判定時間

表3 被験者の属性と判定時間

		性別			全体
		女性	男性	不明	
年代	20	19.53	16.95	N/A	18.84
	30	25.67	22.74	9.62	24.74
	40	27.60	21.51	N/A	25.57
	50	26.62	19.12	N/A	23.89
全体		22.26	18.61	9.62	21.24

単位:秒

3.3 判定尺度、提示順と判定時間

判定尺度ごと、提示順ごとの判定時間を示したものが表4である。

表4 判定尺度、提示順と判定時間

		提示順			全体
		非⇒レ	レ⇒非	混合	
尺度	5段階	19.89	20.72	17.98	19.52
	スライダー	22.09	24.88	22.18	22.98
	マグニチュード	21.99	20.73	20.79	21.16
全体		21.37	22.06	20.36	21.24

単位:秒

3.5 判定結果と判定時間

判定尺度ごとに判定結果の値は異な、同一に扱うことができないため、最大値と最小値を使い、判定値が、0(レレバンス評価＝低)から 1(レレバンス評価＝高)の間に入るように正規化を行った。また、正規化を行った結果、判定値が 0.5 となったものはどちらでもないと判断し、除去した。

表5 判定結果と判定時間

		レレバント	非レレバント
出現頻度		667件	882件
判定時間	平均	24.92秒	18.38秒
	分散	276.00	253.60

結果を示したのが表5である。判定時間に関して既往研究と同様に、t 検定を行ったところ、有意水準 $p<0.005$ でレレバントな文献に対する判定時間と非レレバントな文献に対する判定時間の平均には差があることが確認された。つまり、レレバントな文献に対しては、判定により長い時間がかかったことがわかる。

判定時間ごとの出現頻度を示したものが図5のグラフである。非レレバントな文献に対しては10秒以内に終わった判定が多い一方、レレバントな文献に対しては判定時間が20秒を越えるものも多いことがわかる。

既往研究よりも大規模な実験を行うことで、レレバントな文献に対する判定は非レレバントな文献に対する判定よりも長い時間がかかることを明確に示すことができた。今後は、判定尺度や提示順の影響等、実験データのより詳細な分析を行う予定である。

■ 謝辞

本実験にご協力くださった亜細亜大学司書講習受講生と関係者の皆様に深く感謝致します。

【注・引用文献】

1) Salton, G.; Buckley, C. Improving retrieval performance by relevance feedback. Journal of the American Society for the Information Science. vol.41, no.4, 1990, p.288-297.
2) 初期の goo(http://www.goo.ne.jp/)はレレバンスフィードバックを実装していた。
3) Morita, M; Shinoda, Y. "Information filtering based on user behavior analysis and best match text retrieval". Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, 1994, p.272-281.
4) Diane K.; Belkin, N.J. "Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback". Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001, p.408-409.
5) White, R. W.; Ruthven, I.; Jose, J. M. "The use of implicit evidence for relevance feedback in web retrieval". Proceedings of 24th BCS-IRSG European Colloquium on IR Research, Lecture notes in Computer Science vol.2291/2002, 2002, p.93-109.
6) Kellar, M.; Watters, C.; Duffy, J.; Shepherd, M. "Effect of task on time spent reading as an implicit measure of interest". In Proceedings of ASIS&T 2004 Annual Meeting, 2004, p.168-175.
7) Eisenberg, Michael B. Measuring relevance judgment. Information Processing and Management. vol.24, no.4, 1988, p.373-389.
8) Bruce, H. W. Cognitive view of the situational dynamism of user-centered relevance estimation. Journal of the American Society for Information Science. vol.45, no.3, 1994, p.142-148.
9) Eisenberg, Michael B.; Barry, Carol. Order Effects : A study of the possible influence of presentation order on user judgment of document relevance. Journal of the American Society for Information Science. vol.39, no.5, 1988, p.293-300.
10) 安形輝. "レレバンス判定に対する文献提示順と判定尺度の影響". 第 53 回に本図書館情報学会研究大会要綱. 2005, p.169-172.
11) http://nsearch.yahoo.co.jp/bin/search
12) http://news.google.co.jp/