

# 情報検索におけるユーザモデル

加藤 誠

京都大学大学院 情報学研究科

*kato@dl.kuis.kyoto-u.ac.jp*

**概要** 情報検索を行うユーザのモデルは、ランキング学習や評価指標、対話的情報検索にて用いられ、現在もなお情報検索において中心的なトピックとなっている。本講演では、検索ユーザの理解とそのシステムへの応用という2つの側面から、情報検索におけるユーザモデルを紹介する。

**キーワード** 情報検索, ユーザモデル, 評価指標, 対話的情報検索

## 1 はじめに

「情報検索」という学問領域はしばしば他の領域の1つのアプリケーションとして捉えられることが少なくない。例えば、索引付けのための単語分割やランキングのための言語モデルは自然言語処理、転置索引等の高速化の技術はデータベース、ランキング学習などの仕組みは人工知能などの分野に由来すると思われる。では、「情報検索」に特有の技術は一体何なのであろうか。

Croft, Metzler, Strohman による情報検索の教科書 [3]によれば、情報検索の三大トピックは適合性 (relevance), 評価 (evaluation), 情報要求 (information needs) であるという。この主張がすべての情報検索の研究者に受け入れられているわけではないが、これが三大トピックとして掲げられていることは多くの人にとって意外であるかもしれない。一方で、これら三大トピックに共通する部分に着目することで、「情報検索」と他の学問領域を隔てる部分は何であるのかを窺い知ることができる: どのトピックも情報システムを利用するユーザの行動や思考に関する研究と捉えることができるのである。

例えば、適合性に関しては BM25 のような検索アルゴリズムの研究が含まれる。もし我々がユーザのクエリや適合性判定の曖昧性について何の仮定も持たずに研究を行っていたのであれば、それは単にデータベース中のタプルを与えられたクエリに従って取得するだけのシステムで完結していたであろう。(文書中に出現する全単語、または、文書の ID を指定するような検索である。) 実際には、ユーザのクエリは極めて不完全であり、その適合性判定の基準をクエリだけから推し測ることは極めて困難であるため、多種多様な検索アルゴリズムが提案されてきているのである。評価に関しても同様のことが言える。近年の多くの検索評価指標はユーザの存在を前提としている。現在広く用いられている Normalized Discounted Cumulative Gain (nDCG) [4] においても、

top-heaviness (上位の結果を重視する性質) の根拠をユーザの粘り強さに置いている。(著者らは nDCG がユーザの存在を想定しない場面で用いられることは本意ではないと言っているらしい。) 情報要求については特に説明を要しないものと思われる。

ここで簡単に論じたように、情報検索の三大トピックは情報システムを利用するユーザの行動や思考に関する研究であると言い換えることができる。これがもし正しいのであれば、情報検索の大部分の研究を以下のように捉えることができる:

情報を獲得しようとするユーザの行動・思考を  
理解・推定・活用し、ユーザの情報アクセスを  
効果・効率の面で向上させる研究

つまり、情報検索ではユーザを明示的に問題の枠組みの中に取り入れ、それらの理解自体もまた研究の関心として含んでいるのである。上記の意味では、冒頭で挙げたような他の学問領域よりも、むしろ、ヒューマンコンピュータインタラクションに近いと言えるであろう。

ユーザがどのような行動を行うのか、また、考えるのかについて、抽象化を経て数理的な枠組みの中で利用可能な形式にしたものを本講演ではユーザモデルと呼ぶ。ユーザモデルに関する研究には下記の3つの段階があると著者は考えている:

1. Exploration: 観察によるユーザ理解
2. Explanation: モデルの観測データへの当てはめ
3. Exploitation: モデルの検索システムへの応用

Exploration の段階ではユーザスタディなどを通して、より現実のユーザに近いユーザモデルの構築を目的とする。Explanation では特定のユーザモデルを仮定し、実際の観測データに基づいてその妥当性を検証したり、モデル中のパラメータを調整したりする。Exploitation ではユーザモデルを活用し、検索アルゴリズムや対話的情

報検索システムの最適化、および、ユーザモデルに基づいた評価を行う。本講演では著者が行ってきた研究を交え、上記3つの段階における情報検索研究の紹介を行う。

## 2 ユーザモデルの Exploration

Exploration は認知学や人間行動学的な側面が強く、ユーザスタディなどを伴うことが多い。近年盛んに行われているのは、特定のタスク下において、クエリの数やセッション長などの各種検索に関わる指標がどのように変化するかを観察する研究である。例えば、Liu らは4つの異なる検索タスク (background information collection, interview preparation, advance obituary, and copy editing) においてこれらの指標がどのように変化するかについて報告している [6]。他にも、ドメイン専門家の検索行動に関する研究 [13]、記憶と検索に関する研究 [7]、検索エンジン切り替え行動に関する研究 [12]、難しい検索時の検索行動に関する研究 [1] などがある。興味深いのはこれらの多くが SIGIR, WSDM, CIKM などの情報検索系の会議で発表されていることである。この事からもユーザの検索行動分析が情報検索において重要なトピックであると認知されていることがうかがえる。

検索行動の一部とも考えられるが、適合性の概念そのものについても多くの検討が情報科学の分野において行われてきている。最もよく知られている適合性研究は、Saracevic による適合性の分類であると思われる [10]。Saracevic によれば適合性はアルゴリズム的 (algorithmic)、話題的 (topical)、認知的 (cognitive)、状況的 (situational)、情動的 (motivational) 適合性に分類される。このリストの後の方に行くほど、より主観的でありユーザ個人に依存する適合性であると言える。アルゴリズム的適合性は、検索システムによって判断された適合性であり、アルゴリズムによって付与されたスコアを指す。この値が高いからといって必ずしもその文書（もしくは情報）がユーザにとって適合であるとは限らない。話題的適合性は、文書の話題とユーザが欲する話題の合致度を指し、多くの検索アルゴリズムはこの段階における適合性の向上を目指して提案されている。多くのテストコレクションにおいて用いられている適合性は、主題的適合性であり、主観性が極力排除されていることになる。認知的適合性は、ある文書からユーザが認知できる特徴とユーザの期待した認知的特徴との合致度を指し、話題的適合性よりも主観的で、個人に強く依存する適合性である。例えば、わかりやすいブラックホールに関する文書が欲しいと思っているユーザにとって、多くの専門用語をもってブラックホールを解説している文書は話題的適合であると言えるが、認知的適合であるとは言えない。認知的特性の例としては他に、主観的、客観的、具

体的、抽象的などが考えられる。

我々はこれまでユーザが期待する文書の認知的特徴が、クエリにどのような形態で現れるかについて研究を行ってきており、本講演ではこの研究について紹介したいと考えている [5]。この研究の主な成果は、ユーザが期待する認知的特徴はクエリ中にいかなる形でも出現しないことが多い、という実験結果である。例えば、わかりやすいブラックホールに関する文書が欲しいと思っているユーザはただ単に「ブラックホール」というクエリを入力するのである。このことは、認知的適合性を期待するユーザの存在を検索エンジンの目から覆い隠し、認知的適合性を考慮した検索アルゴリズムの設計を困難にしていると思われる。

## 3 ユーザモデルの Explanation

Explanation は Exploration と同時に行われ、特別なモデルを仮定するのではなく、既存の機械学習のモデルを用いて行われることが多い。例えば、検索エンジンの切り替え行動に関する研究 [13] では、ロジスティック回帰モデルを利用して切り替え行動予測を実現している。

本講演では特に興味深い Explanation の例として、近年の検索評価指標の構築方法を紹介したいと考えている。もともと、ユーザの満足度という不確かなものを近似するために多数考案されてきた検索評価指標であるが、近年では特にユーザモデルを明示的に仮定することで満足度を算出するケースが増えてきている。

Mean Average Precision (MAP) は最近になって背後に存在するユーザモデルの存在が指摘されている [8]。Expected Reciprocal Rank (ERR) は新規性と多様性を考慮するために考案された検索評価指標であり、検索結果を上から順に閲覧し、ある程度の適合文書を見つけると満足して検索を止めるというユーザモデルに基づいている [2]。Sanderson らによる研究では実際にユーザの選好と検索評価指標の相関を調べており、これによってユーザモデルの Explanation が行われていると考えることもできる [9]。

より直接的な Explanation の例として、Time-biased Gain (TBG) の提案と共に行われた、ユーザモデルのキャリブレーションが挙げられる [11]。Smucker と Clarke は nDCG などのようにランクに応じて利得が減衰するようなモデルではなく、時間に応じて利得が減衰するモデルを用いて、TBG と呼ばれる評価指標を提案している。TBG では、適合文書を一定の確率でクリックし、その文書を閲覧するというユーザモデルを仮定し、ある順位までにかかる時間を推定している。また、このモデルでは一定時間経過後にユーザが検索を止めるという仮定を置いており、検索を止めるまでに見た適合文書によっ

て評価指標を算出している。これらのユーザ行動の仮定によって、より上位に出現する文書の適合性がより重視されるという top-heaviness や、得られる利得が利得取得回数に応じて減衰していく diminishing return という性質を実現している。この評価指標を構築する際には、ユーザスタディを行って、検索結果のクリック確率や文書閲覧にかかる時間などを計測し、ユーザモデルにて用いられるパラメータを決定している。

まだ未発表ではあるが、我々が最近行った Explanation の例として、クエリ推薦が検索行動に与える影響に関する研究を紹介する。我々はクエリ推薦を利用するユーザモデルを構築し、TBG と同様にユーザスタディを通して、 $k$  個のクエリ推薦によってどれくらい検索行動に遅延が発生するのか調査を行った。実験はデスクトップとモバイル環境下で行われ、両者の間ではいくつかの興味深い違いが観測されている。

#### 4 ユーザモデルの Exploitation

Exploitation の例として、我々が最近行った 2 つの研究について触れる。1 つは質問回答を通じた対話的情報検索における質問生成の最適化についてであり、もう 1 つはクエリ推薦と検索結果多様化の同時最適化についてである。

与えられたクエリに対して質問を生成し、その質問への回答を元にクエリを自動的に修正する、会話的適合フィードバック (conversational relevance feedback, CRF) を我々は提案している。例えば、「中華 飲食店」というクエリに対して、「予算はどれくらいですか?」という質問をユーザに与え、その回答によって検索結果を改善するという仕組みである。CRF において最も効果的な質問生成を行うために、ユーザを確率的モデルによって記述し、CRF を用いたユーザの期待利得の最大化問題として質問選択問題を定式化した。この研究において重要な点は、やはりユーザモデルの構築である。我々のユーザモデルは、各質問への回答可能確率や回答確率などをパラメータとして含み、これらをクエリや回答などから推定しつつ、最適な質問選択に用いている。

クエリ推薦と検索結果多様化の同時最適化では、あるクエリ推薦集合を提示した際にどのようにユーザが振る舞うかをモデル化し、最も多くのユーザが満足できるようなクエリ推薦・検索結果を生成する方法について提案を行っている。この研究では先ほど述べた、クエリ推薦が検索行動に与える影響を測るユーザスタディを通して、ユーザモデルのパラメータを決定している。このユーザモデルに従うユーザが最大の期待利得を得られるよう、クエリ推薦を選択し、同時にランキングを決定している。

これら 2 つの研究の例を通して、ユーザモデルが特に

対話的情報検索を改善する上でどのように役立つかを示したいと考えている。

#### 5 まとめ

本講演では、情報検索とはどのような研究領域であるのかについて簡単に紹介を行い、情報検索におけるユーザの重要性について述べた。また、ユーザモデルがどのような段階を経て構築されるのかについて整理し、各段階においてどのような研究が行われてきているのかについて紹介した。これらの解説を通して、より多くの方が情報検索という分野に興味を持ち、NTCIR (NII Testbeds and Community for information access Research)<sup>1</sup> や SIGIR (Special Interest Group on Information Retrieval)<sup>2</sup> 等への日本からの参加者が増えることを期待している。

#### 参考文献

- [1] A. Aula, R. M. Khan, and Z. Guan. How does search behavior change as search becomes more difficult? In *CHI*, pages 35–44, 2010.
- [2] O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *CIKM*, pages 621–630, 2009.
- [3] W. B. Croft, D. Metzler, and T. Strohman. *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2010.
- [4] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *TOIS*, 20(4):422–446, 2002.
- [5] M. P. Kato, T. Yamamoto, H. Ohshima, and K. Tanaka. Investigating users' query formulations for cognitive search intents. In *SIGIR*, pages 577–586, 2014.
- [6] J. Liu, M. J. Cole, C. Liu, R. Bierig, J. Gwizdka, N. J. Belkin, J. Zhang, and X. Zhang. Search behaviors in different task types. In *JCDL*, pages 69–78, 2010.
- [7] S. Ochiai, M. P. Kato, and K. Tanaka. Re-call and re-cognition in episode re-retrieval: A user study on news re-finding a fortnight later. In *CIKM*, pages 579–588, 2014.
- [8] S. Robertson. A new interpretation of average precision. In *SIGIR*, pages 689–690, 2008.
- [9] M. Sanderson, M. L. Paramita, P. Clough, and E. Kanoulas. Do user preferences and evaluation measures line up? In *SIGIR*, pages 555–562, 2010.
- [10] T. Saracevic. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6):321–343, 1975.
- [11] M. D. Smucker and C. L. Clarke. Time-based calibration of effectiveness measures. In *SIGIR*, pages 95–104, 2012.
- [12] R. W. White and S. T. Dumais. Characterizing and predicting search engine switching behavior. In *CIKM*, pages 87–96, 2009.

<sup>1</sup><http://research.nii.ac.jp/ntcir/index-ja.html>

<sup>2</sup><http://sigir.org/sigir2017/>

- [13] R. W. White, S. T. Dumais, and J. Teevan. Characterizing the influence of domain expertise on web search behavior. In *WSDM*, pages 132–141, 2009.