

先行研究サーベイの自動化・効率化

シリーズ第2回は、先行研究サーベイの自動化・効率化である。最初に対象を明確化する。生成 AI サービスが生成する対象として2つが考えられる。一つは、研究テーマ領域のサーベイに必要な情報を収集・整理したものであり、もう一つは論文中のサーベイ部分そのものである。

前者については、生産効率性を大幅に引き上げるツール群が発展しており、本稿と次稿では実際の使用例を紹介しながら主要ツールの解説を行う。まず、本稿ではサーベイ作業の分解と、デジタル化にともなう手法の進化について解説し、次稿で具体的に、Elicit、Consensus、Sciteなどの生成 AI ツールを取り上げる。サーベイの方法論に大きな技術革新が生じたこと、研究の質量両面で競争力や生産性の観点から、その活用は必須であることが実感できよう。

執筆を自動化しない理由

後者は、収集した情報に基づき論文を執筆する作業であり、論文作成の完全自動化を目指した研究では、既にシステムとして実装が試みられている。しかし、本稿では2つの理由から論文のサーベイ部分の自動執筆を目指してはいない（いかなるケースにおいてもすべきでない」と主張している訳ではない）。

1番目の理由は、「サーベイ作業は研究テーマを巡る先行研究のマッピングを行う行為であり、研究者の頭に見取り図を作成し、今後の研究のためにこれを記憶させる行為である」ことに因るものである。この作業では、研究者が自ら執筆することが有効であり、そのためには収集した情報の消化が前提となる。ここにも生成 AI 活用の機会が大きく広がっており、今後取り上げる。論文執筆の完全自動化によって情報消化なしでも研究が可能となるが、記憶には定着し難く、対象分野の全体像をつかむ機会も放棄してしまうことになる（加えて、果たしてそれを研究と呼ぶのかという問題もある）。

また、書きながら考える、あるいは作業をループしながら質を高めていくのは、前稿で指摘したように研究の全工程で生じるものであり、特にサーベイ過程では研究の方向性と完成した際の研究価値に直結してくる。それゆえ研究者の関与は重要であると思われる。材料収集

や情報整理・消化には生成 AI サービスを積極的に活用すべきだが、研究者の能力向上・知見蓄積の視点からは執筆は自ら行ったほうがよいといえよう。

2 番目の理由は、ハルシネーションのリスク回避である。生成 AI のコア技術である LLM（大規模言語モデル）の知識量増加と AI エージェント技術の発展により、LLM 登場初期に比べてハルシネーションは大幅に抑制されるようになった。AI エージェントが論文アーカイブなどネット上の専門情報をサーチし、収集した情報を LLM が消化し、必要な情報をさらにサーチして統合し、「思考を深めていく」という Deep thinking/Chain of Thought 技術も、その抑制に貢献している。にもかかわらず、ハルシネーションは LLM のモデル構造に起因する避けがたい事象として発生し続ける（LLM の詳細については、[副島\(2024\)](#)の 2～3 節を参照）。

もちろん、研究者も誤認や情報不足、認知バイアスなど様々な理由でハルシネーションを起こす。既に、生成 AI のハルシネーションは研究者のそれより小さくなっているかもしれない。また、生成された原稿のファクトチェックや批判的検討を生成 AI に繰り返し行わせることでハルシネーションを回避する方法もある。このため、2 番目の理由の重要性は、徐々に縮小していくと推測される。

先行研究サーベ이의意義

研究者には言わずもがなであるが研究に馴染みがない読者のために、先行研究のサーベイにおいて研究者がどのような作業を行っているかを先に解説する。業界の方は次節「サーベイ手法のデジタル進化」へスキップされたい。

まず、サーベ이의意義であるが、あらゆる研究は先行研究の積み重ねの上に成り立っており、少なくとも金融経済の分野でこれを行わない論文は学術研究とはみなされない。ほとんどの人間が思いつくことは、概ね誰かが気が付いて先に研究を行っている。論文が完成して報告を行った際に、「それは先行研究ですでに検証済みであり、同じことを再現検証した以上の付加価値はない」と指摘されるのは、最も避けたいことのひとつである。

一般に学術研究の価値は、先行研究が手掛けていない分野を「ある観点」で押し拡げたという行為によって創られる。同じ内容を異なる国や時期のデータセットで再検証したという価値もあれば、先行研究の仮定を一つ緩和しても結果は一般性を失わなかった、あるいは逆に結果が異なったという発見も価値である。

更にサーベイが重要な理由がある。論文検索サイトの Google Scholar のトップページには、「巨人の肩の上に立つ」というアイザック・ニュートンの言葉が掲載されている。先人の成果に立脚するほうが、研究テーマの選択と分析視点の模索を効率的に行えることを示している。もちろん、独創性の観点からは自由な発想や妄想とサーベイ作業の往復運動も重要であるが、ベースとなる基盤がない限り独創もない。

このため、サーベイにおいては広範な先行研究の収集、消化、整理が求められる。新分野に取り組む際には、学術論文を数百本単位で検索し、100本単位で読みこむ必要がある。一本の論文のためにこれを行うのは著しく効率が悪いので、多くの場合、専門分野を限定して蓄積を図ることで、膨大な先行研究をカバーしている。

また、最近の研究までバランスよくカバーした上級テキストが出版されると、代表的な論文を個別にあたる手間を省くことが可能となり、大学院での学術訓練や蓄積された英知のキャッチアップ作業にも貢献する。更に専門的な Handbook シリーズ（特定学術分野の発展を網羅的にカバーするために展望論文や解説論文を集めた百科事典的なもので、第一人者らによって執筆され、その学術分野の現在地やコンセンサス、文献レビュー、未解決問題の整理などが得られる）も、こうした理由から作成されている。North-Holland(Elsevier)社の Handbook of Econometrics や Handbook of Industrial Organization などが代表例である。

サーベイ手法のデジタル進化

サーベイの効率化や自動化は、生成 AI 時代になって突然発展したわけではない。そのデジタル進化の歴史を簡単に振り返っておく。

デジタル技術の発展以前は、主要研究論文の巻末参考文献を手掛かりに先行研究のリファレンス・ネットワークを辿ることで、ある分野の研究体系を習得していた。このほか、上述の Handbook シリーズや、Journal of Economic Perspective、Journal of Economic Literature、各専門誌の Annual Review といった展望論文を扱うジャーナルを読み、そこで参照されている文献にもあたるといった方法も、効率的なアプローチとして活用されていた。先端的な研究については、NBER (National Bureau of Economic Research) や CEPR (Centre for Economic Policy Research) などのワーキングペーパー・シリーズの目録にあたるという方法があった。また、その分野の先達に話を伺うといった対人の情報収集もあった。

これらは物理的には紙のジャーナルのページを繰る行為であったが、そうした時代に変革をもたらしたのは電子アーカイブの登場であった。数学・物理分野の電子アーカイブ arXiv に続いて、SSRN や RePEc など経済学分野の電子アーカイブが 1990 年代後半に誕生し、医学の PubMed、工学の IEEE Xplore など著名なアーカイブが発展した (arXiv はその後に対象分野を拡大)。これらにより論文タイトルや筆者名、キーワードによる検索や、詳細分野・コード体系別の一覧チェックが可能となった。

2000 年代になると、JSTOR や ScienceDirect (Elsevier 社) など電子ジャーナル利用の普及により、検索や論文収集が効率化された。しかし、収集はできても、論文の消化と体系だった整理は研究者自らの作業として残されていた。

プレ生成 AI 期の状況

次に生じた変革が検索技術の高度化である。[Scopus](#) (Elsevier 社) や [Web of Science](#) (Clarivate 社、旧: トムソンロイター) などの引用データベース検索サービスが登場した(リンク先は現時点のサービス、いずれも有料)。これらのサービスにより、誰が誰を・どの論文がどの論文を引用しているかといった研究のネットワーク追跡が容易となった。

また、論文の引用回数、研究者の評価指標である h-index (ジャーナル採択論文の量と、引用回数で測った質の両者を勘案した指標)¹、各ジャーナルのインパクトファクター²など、定量的な影響度分析が行えるようになった。Scopus は、被引用数や著者・機関別の指標を集計する Analyze results 機能を提供しており、Web of Science は同機能に加えて [Citation report](#) による引用分析機能を提供している。これによりサーベイは、「影響度と引用関係に基づく中核的な文献集合の特定」という性質を帯びることになった。そのプロ・コンについては種々の議論がなされている³。

こうした技術発展があって 2010 年代にはネットワーク分析と可視化技術がサーベイに活用され始めた。文献同士の引用ネットワーク、キーワード共起ネットワーク、著者や所属機関のネットワークを可視化し、クラスタリング(研究トピックの塊り群)や中心性指標を算出する。その分析結果から、「どのクラスタがその分野の主要な潮流か」、「どこに新興トピックがあるか」が俯瞰可能となる。こうした手法は、Systematic literature network analysis や Bibliometric analysis (計量書誌学的分析) と呼ばれている。なお、中心指標性などのネットワーク分析については[副島\(2008\)](#)の 2 節を参照されたい。

さて、ここに至っても、論文の読み込みや論文間の関連性の理解は、研究者によって行われてきた。しかし、生成 AI の基盤技術の一つとなった自然言語処理技術(NLP: Natural Language Processing)や機械学習(ML: Machine Learning)の発展は、セマンティック分析と呼ばれる新しい「意味解析」手法をもたらした。キーワードの共起分析では、各論文のキーワードの拾い上げは論文筆者の手によってなされるか(例えば論文要旨の下方に JEL コードとあわせて記載されている)、文献アーカイブ運営者などが行っていた。しかし、NLP の技法のトピックモデリング(LDA: Latent Dirichlet Allocation というモデルが発展契機)を用いると、論文から自動的にトピックが検索でき、かつ、ある論文はトピック A が 60% でトピック B が 40% というような定量分析に基づくラベリングができるようになった。

アーカイブにある膨大な論文をトピック分析すると、トピック毎の規模感や時系列的な隆盛や、キーワードの代わりにトピックを使ったクラスタリングやネットワーク分析が可能となる。さらには、Word2Vec の登場で単語をベクトル表現化する技術が発展した(文脈に単語がどう埋め込まれているかを活用するため埋め込みモデルとも呼ばれ、後の生成 AI に繋がる基本技術となっている)。これにより単語間や文章間の類似度計測が可能となり、新たなクラス

タ化（グルーピング化）の技法がもたらされた。ベクトル表現や埋め込みモデルについては、前出の[副島\(2024\)](#)を参照されたい。

[Semantic Scholar](#) は、Allen Institute for AI（マイクロソフトの共同創業者ポール・アレンが設立）が 2015 年に公表した無料の AI 駆動型の論文検索エンジン・サービスである。単なる検索サービスでなく、セマンティック（意味解釈）技術を用いて、科学の知識構造を計算可能なグラフ（Semantic Graph）に落とし込んでいくプロジェクトである。その特徴は、論文の意味を解析し、類似論文の推薦、重要文献のハイライト、引用関係や影響力に基づくランキング、自動抽出した要約やキーフレーズの提示などを活用する点にある。初期の開発は生成 AI の爆発的成長が生じる前の時代であったため、当初は、例えば要約が抽出的・テンプレート的であり、また解釈の踏み込み度合いも浅く、現在の生成 AI サービスほど能力は高くなかった。しかし、その後、生成 AI 時代を迎えたことで新技法が導入され、高度化が図られていく。

例えば、生成 AI の大規模言語モデル（LLM）で著名な BERT というモデルが 2018 年に登場しているが（[副島 2024](#) 参照）、Allen Institute for AI は BERT のモデル構造を利用し、膨大な数の科学論文を学習させることで SciBERT という科学論文特化型の LLM を翌 2019 年に作成・公表している。科学専門用語の辞書から作り直しているため、専門的な文章や単語のトークン化において汎用モデルの BERT より優れている⁴。こうした工夫が、学術論文に対するセマンティック能力の向上をもたらした。SciBERT の成功を受けて、BioBERT（PubMed の論文を用いて汎用 BERT に「追加」学習を実施）、MathBERT、ClinicalBERT（臨床記録であるカルテから学習）などが登場している。

生成 AI 時代のサーベイ技術

生成 AI の爆発的な技術進化は、BERT の次に登場した著名モデル Transformer によって始まった。これが、その後の GPT や Claude、Gemini、Llama など普及モデル躍進の礎となっている。LLM モデルの飛躍的進化は、サーベイに特化した生成 AI サービスの登場と機能発展をもたらしている。

以下、その代表例を示す。大別すると、大量の論文を読んで情報を抽出し整理する生成 AI サービス（Elicit、Consensus 等）と、前出の引用などの繋がりを可視化する生成 AI サービス（引用の関係評価を含む、Scite、Connected Papers 等）に分けられる。後者における生成 AI 時代ならではの「関連性がある研究間で結果が同意見か、それとも相反するものであるか」といった論文の内容解釈に基づく情報が活用されている点である。

なお、サーベイは論文を読み込む行為と不可分であり、SciSpace のような読むことにも重点を置いた生成 AI サービスや、Research Rabbit のように参考文献の作成作業と連動した生成 AI サーベイも存在する。まずは各サービスの概要を紹介する。

① Elicit

自然言語で質問を投げかけると（例：日本の家計の資産選択行動を実証分析した研究は、どのような切り口を用いて、どのような分析結果を得ているか？）、関連論文を探し出し、それらの特徴を一覧表として提供してくれる。対象データ、使用した手法、結果、限界点など、指定した項目を論文から抜き出して一覧表を作成する。論文の内容を読み込む必要があるため、LLM による情報の抽出と集約を行い、Semantic Scholar 等のデータベースを利用する点が特徴である。

② Consensus

コンセンサスの有無を判定する検索エンジンで、Yes/No で答えられる質問を与えると、関連論文を分析し、「60%の論文が Yes と言っている」といった指標（コンセンサス・メーターと呼称）を算出する。Yes/No 以外に Possibly や Mixed がある。研究者がこれから検証しようかとプランニングしている仮説が、先行研究でどの程度支持されているか、あるいはまだ注目されていない仮説であるのか（＝抽出された論文が少ない）を確認するのに便利である。Semantic Scholar 以外のデータベースとして OpenAlex や、Consensus 独自の学術論文 web クローラーを用い、まず、関連性の高い 1,500 本の論文を抽出し、さらに高精度モデルでトップ 20 本に絞り込んだうえで、与えられたテーマに関するコンセンサスを導出する。このほか、個別の論文から主要な発見（Key Takeaway）を 1 行で抜き出す機能や、論文内容をチャットで問える機能も有している。

③ Connected Papers

キーワード検索ではなく、1 本の種論文（Origin Paper）から関連論文を探知するツールである。指定した種論文と引用関係や共起関係が強い論文を、グラフ構造としてビジュアル化する。中心の種論文の近くにある論文ほど関連性や重要度が高く、色や距離で類似度が表される。定量評価には、共引用分析（Co-citation）と書誌結合法（Bibliometric Coupling）が用いられており、手法的には古い技術であるが、直感的でモダンな UI となっている。

④ Litmaps

ネットワーク図に時間軸（タイムライン）が加えられている。左から右へ時間が流れるマップ上に論文を配置し、誰の研究が、後の誰に継承されたかという研究の系譜（引用の連鎖）を可視化する。研究テーマの歴史的発展を追い、最新の派生研究を見落とさないための有益な生成 AI サービスとなっている。

⑤ Scite

引用数ではなく引用のされ方（意図）を確認する生成 AI サービスである。ある論文が引用さ

れている文脈を AI が読み取り、その引用論文が支持 (Supporting) されているのか、逆の結果が示されているのか (Contrasting)、単に言及 (Mentioning) されているだけなのか、3 つに分類して引用数をカウントする。これを Smart Citations と呼称している。多くの論文で引用されているが、後続研究によっても指示され続けている研究なのか、テーマ着眼は早かったものの分析結果については批判的な評価が定着している研究なのかでは、位置づけが異なってくる。

⑥ Research Rabbit

論文版 Spotify と形容できる文献探索ツールである。Connected Papers の繋がり分析に加え、推薦機能 (レコメンド) が強化されている。種論文から探すのではなく、複数の論文が入ったフォルダ (コレクション) を利用者が最初に作り、その傾向や特徴を基準に論文収集を進めるものである。UX としては、インタラクティブなネットワーク図が表示され、ノードに当たる各論文をクリックすると、その論文の概要が表示されるだけでなく、その論文を中心とした新たなネットワークが再描画される。Rabbit が穴から穴へと飛び移るように、大量の文献群が探索できる。Zotero などの文献管理アプリケーションと同期させて利用すると効率的である。

⑦ SciSpace

Elicit や Research Rabbit が検索・発見に力点を置いているのに対し、論文を読みこむ、論文を書くことまでカバーしたオールインワン型のプラットフォームである。以前は「Typeset」という名称で、論文のフォーマット整形サービスとして知られていたが、AI 機能を取り入れてリブランディングされている。Copilot 機能が使えるため、英論文であっても日本語で対話しながら読み進められる。質問を与えると関連論文を見つけ出し、トップ数本の論文から抽出した情報を比較表にする点は Elicit に類似している。ウェブブラウザ (Chrome) の拡張機能に組み込み、Google Scholar、arXiv などの文献サイトを参照しながら SciSpace の AI 解説を呼び出せる。

ここで紹介した生成 AI サービスを組み合わせると、例えば以下のようなサーベイ作業が考えられる。まず、Consensus に複数の仮説を検証させて結果を眺め、大まかな学説の傾向を掴む。Elicit では、分析手法やデータセット、結果などを比較表にして、精読すべき論文のリストを作る。とくに研究出発の参照点となる論文については、Connected Papers を用いて関連論文を追い、Litmaps でその後の展開を検証する。後続研究から否定されていないかも Scite で確認する。

次稿では、上記のサービス群の一部について具体的な利用事例を紹介する。

なお、こうしたパッケージ化された SaaS 型の生成 AI サービスを用いなくても、例えばネッ

トワーク検索を行わせるプロンプトを作成し、これを AI エージェント機能が強化された最近の汎用生成 AI サービス（例えば、Gemini や ChatGPT）に与えることで、自分流のサーベイエンジンを自作できる。図表 1 がその雛形である。ただし、上述のような作りこまれた便利なサービスが多数存在しているため、まずはこれらの利用から始めてみるのが効率的であろう。フリーで利用できるサービスもある。

図表 1 引用分析ベースのサーベイ：スクリプトの骨子例

STEP 1：リサーチテーマや大まかな範囲の設定
STEP 2：文献データの収集
<ul style="list-style-type: none"> ・データベースを指定する：Google Scholar、利用可能であれば Scopus、Web of Science ・データ取得：論文タイトル、著者、ジャーナル、発行年、引用数、参考文献、要約、キーワード
STEP 3：引用関係データの構築（Reference から Citation Matrix へ）
<ul style="list-style-type: none"> ・被引用行列（Citation Matrix）：引用有無の 0/1、もしくは引用数 ・共引用ネットワーク（Co-citation Network） ・共著者ネットワーク（Co-authorship） ・VOSviewer、CiteSpace、Bibliometrix（R パッケージ）などを利用
STEP 4：ネットワーク分析（中核文献の特定）
<ul style="list-style-type: none"> ・ネットワーク指標による分析：中心性指標（重要論文の抽出）、In-degree（被引用数）、Betweenness centrality（橋渡し論文）、Eigenvector centrality（影響力の高い文献に引用される文献） ・クラスタリング分析：Louvain 法（コミュニティ検出）、モジュラリティ最大化、K-means ・時間軸分析（研究潮流の変化）：クラスターの時系列遷移、テーマの新旧の判別（average publication year）
STEP 5：クラスタごとの意味づけ（質的内容分析）
<ul style="list-style-type: none"> ・主要文献の読解 ・研究貢献（理論 / 実証 / データ / 手法） ・共通テーマの抽出 ・クラスタ間の関係性の整理（補完/競合/進化） ・ギャップ（未研究領域）の明確化

¹ ある研究者の h-index は、次の条件を満たす最大の数が h であることを指す。条件「少なくとも h 本の論文が各々 h 回以上引用されている」。ジャーナル採択論文の本数と引用件数の両者が h 以上であることが必要である。Google Scholar では、研究者が登録していれば h-index や引用回数、i10-index が自動で計算され、誰でも参照できる。突出した引用回数を持つ論文を高評価する g-index という指標もある。なお、h-index は参照する論文データベース（DB）の規模に左右される。Web of Science は Clarivate の DB が対象で、採用基準が上級のジャーナルになるため同指標は低く（厳しく）出る。Scopus は Elsevier の DB を参照し、やはり低めとなる。Google Scholar は NBER など WP や書籍を含むため高く計測される。このように参照 DB が異なる指標は横比較できない点に留意が必要である。

² インパクトファクターは、ある学術ジャーナルに掲載された論文がどれだけ引用されているかを示す指標で、学術誌の影響力や知名度を定量的に測るために用いられる。Clarivate の Journal Citation Reports で毎年公表されている。ただし、指標としての機能面での問題点以外にも、インパクトファクターが研究活動にもたらすマイナス面が指摘されている。この点は、前述の h-index のような研究者評価の指標も同様である。

³ こうしたアプローチのメリットは、1) 中核文献の客観的な特定、2) 研究者の選択バイアスや見落としの回避、3) 新規テーマの盛り上がりの早期発見・パラダイムシフトのキャッチ・研究者コミュニティの可視化、4) これらを通じたサーベイの全

一般的な質向上が挙げられる。負の側面としては、1) 必ずしも引用数が論文の重要性を示すわけではないが、引用偏重の評価システムを作り上げてしまう点、2) 先行研究のマッピングを研究者の脳内から外部サービスにアウトソーシングしてしまうため、論文内容やその背景の解釈、論文の繋がり・発展経緯の理解などが浅くなる点、3) 非主流で評価されてないが未来を変えるような突破的アイデアが評価されにくくなり、スター論文やビッグネームへの集中が強化される点、4) 引用操作や戦略的引用の影響を受けやすい点が挙げられる。メリット・デメリットの両者を理解して利用するというアプローチが適切と思われる。

⁴ 用語や文章の適切な切り方は、分割された言葉を高次元ベクトル空間で表現する際に決定的に重要となる。例えば、局所漸近混合正規性を、局所漸近 | 混合 | 正規性と分割すると別概念になってしまう。局所 | 漸近 | 混合正規性が正しい切り方である。不完全 | 識別下 | における一般化 | モーメント | 法推定量もだめで、不完全識別 | 下 | における | 一般化モーメント法 | 推定量が正しい。なお、標準的なベクトル表現では、英語の場合、1単語が1トークンとしてベクトル化される（近年ではサブワードに分割したうえでベクトル化するケースが一般化している）。それゆえ、専門用語辞書活用においては、あえて複数単語をセットにし、これを1トークンとしてベクトル表現にすることになる。

参考文献

副島豊 (2024) 「生成 AI ウォークスルー：基本技術、LLM、アプリケーション実装」 SBI 金融経済研究所「所報」6号 pp.51-104

副島豊 (2008) 「コール市場の資金取引ネットワーク」日本銀行金融研究所「金融研究」27 卷 2 号 pp.47-99