

2050年の学術情報処理

Scholarly Information Processing in 2050

難波英嗣 (中央大学)

Hidetsugu Nanba (Chuo University)

私をはじめ国際会議に論文を投稿した 1990 年代は、今思えば随分のんびりした時代であった。インターネットはすでに普及していたが、論文投稿は郵送だった。査読者の人数分だけ論文を印刷し、封筒に入れ、所定の住所に送っていた。郵送先が日本国内ならばともかく、海外ともなると自分の論文が到着するまでにそれなりに日数がかかる。少しでも日数を短縮するため、EMS(国際スピード郵便)の使い方もすぐに覚えた。査読結果の通知も郵送ならば、最終稿の送付も郵送。今考えると本当に気の長い話である。それから間もなく、インターネット上で論文を投稿できるようになった。はじめてネットで論文を投稿した時は、あまりの手軽さに、感動というよりも、本当にちゃんと投稿できたのか?と逆に心配になった。

今日、学術情報は異常とも言うべきスピードで増加・流通している。筆者が関係する分野では、国際会議に採択された論文の著者は、論文の最終稿が完成すると arXiv.org 上で公開するのが一般的になっている。それらをダウンロードして読めば、国際会議が始まる前にはおおよその研究トレンドを知ることができる。論文だけでなくソースコードや実験データも github.com や paperswithcode.com 上で公開され、実験環境の再現すら簡単にできてしまうことも少なくない。さらに、Google Scholar などの論文

データベースにキーワードを登録しておけば、そのキーワードを含む論文がデータベースに登録されるとメールで通知してくれる。

このように、学術論文の流通という点では、以前では考えられないほど恵まれた状況にあるが、良い点ばかりではない。論文を読むスピード以上に入手できてしまうので、読めない論文が増える一方である。その結果、国際会議参加者の参加報告を聞いたり、国際会議論文の輪読会に参加したりすることになる。論文紹介記録を qiita.com や github.com にアップロードし、Twitter 等で告知してくれる人もいるので、筆者もありがたく活用させてもらっている。しかし、自分が興味ある研究を誰かが都合よく報告してくれるとは限らないので、結局は自分で頑張っ

て論文を読むしかない。

学術情報を対象にした情報処理で現在できることとできないことは、研究者が直面している状況にほぼ対応している。すなわち、学術論文の検索、分類、推薦などの流通に関する技術は、十分実用レベルに達していると言っても過言ではない。一方で、そこから先の、各論文の内容をある程度理解する必要のある処理、例えば、特定分野の論文集合から重要論文を見つける、論文を査読する、特定分野の論文集合を分析し要約する、何か新しい発見をし、論文を執筆する、といった処理は、研究レベルでは様々な

取り組みがあるものの、実用化への道のりは遠い。

「現在の情報処理技術でできること/できないこと」は人間の情報処理能力と無関係ではあるまい。例えば、ある論文が情報検索に関するものかどうかは、情報検索分野の専門家であれば、論文を見た瞬間に判断できるであろう。しかし、その論文の査読となると、明らかに採択基準に満たない論文を除けば、その分野の専門家でも採否の判断にはそれなりに時間がかかる。研究室の学生に渡した論文の内容を、学生がきちんと理解したかどうかを確かめるには、筆者の分野では、その論文の内容をきちんとプログラムとして実装できるかどうかを見れば良い。論文からプログラムを自動生成する試みも研究レベルでは存在するが、実用化レベルには程遠い。

深層学習をはじめとする機械学習は、大量のデータが存在するタスクでは、その能力を遺憾なく発揮する。しかし世の中には機械学習するにはデータが少なくても物量作戦が効かないタスクも無数に存在する。人間はデータ数が少なくても効率的に学習するのが得意である。言語を獲得するのに、何十 GB も何百 GB も文書を読む必要はない。おそらく、入力された情報を、頭の中で汎用性が高い形で保存できるからだと思うのだが、その仕組みがよくわからないからこそ、現在は物量作戦に頼らざるを得ないように思う。

現在の情報処理技術でできることとできないことの溝は深く、今の物量作戦に頼る手法の延長では、溝を埋めることはできないと筆者は思っている。情報基礎とアクセス技術研究会 (IFAT) の前身である情報学基礎研究会の設立趣意書とも言うべき論文 [1] を読むと、知識の獲得や体系化がスコープのひとつになっていることがわかる。それから 30 年経った現在でも、この問題は十分に解決されたとは言えない状況のまま残されている。これから 30 年、知識の獲得や体系化についてさらに議論され、人間と計算機との溝が少しでも埋まることを、IFAT 研究会の主査として願っている。

参考文献

[1] 藤原謙, 情報学基礎のスコープ, 情報処理学会研究報告情報学基礎, 1-1, pp. 1-3, 1986.

難波 英嗣 (正会員)

nanba@kc.chuo-u.ac.jp

2001 年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了. 博士 (情報科学).
2019 年より中央大学理工学部教授, 情報基礎とアクセス技術研究会主査. 自然言語処理, 特許情報処理, 観光情報学に関する研究に従事.

概要

学術論文の検索、分類、推薦など、学術情報の流通に関する技術は現在実用レベルに達しているが、論文の内容をある程度理解する必要のある処理については、実用化への道のりは遠い。深層学習をはじめとする機械学習は、大量のデータが存在するタスクでは、その能力を遺憾なく発揮する。しかし世の中には機械学習するにはデータが少なくても物量作戦が効かないタスクも無数に存在する。現在の情報処理技術でできることとできないことの溝は深く、今の物量作戦に頼る手法の延長では、溝を埋めることはできない。これから 30 年、知識の獲得や体系化についてさらに議論され、人間と計算機との溝が少しでも埋まることを、IFAT 研究会主査として願っている。