

未解決問題の正体

本特集号は、2014年から2015年に行われたProject Next NLP¹と、その最終報告会を兼ねた本会年次大会ワークショップ「自然言語処理におけるエラー分析」の成果を受けて企画された。本号は前号に続く第2弾である。

情報抽出、質問応答、文書要約、対話処理など、言語処理研究はこれまで様々なタスクに取り組んできたが、どのタスクをとっても、十分な精度が得られないまま、どうも60%、70%あたりで伸び悩んでいるように見える。精度「60%」の先の未解決問題の正体は一体何なのか？それはどうすれば解明できるか？それが示唆する次の重要課題は何か？これらが本特集号の問いであった。掲載された全9編の論文はいずれも上の問いに対してそれぞれ個性的な方法で極めて示唆に富む解答を提供している。各研究グループの力作を幅広い研究者・技術者に味わっていただきたい。

さて、「60%」の先の壁の正体は何だったのか？本特集号で確認できたことの一つは、いずれのタスクにおいても、何か一つの根源的な問題がそこにあって、それが解ければ一気に精度が上がるという構造になっている訳ではどうもなさそうだとということであろう。たとえば、照応・省略解析や含意関係認識などで実際の解析誤りを見てみると、「60%」の先には、オントロジカルな語彙知識やパラフレーズ、世界知識が関わる問題の他に、実に多様な問題が少しずつ顔を見せる。未知語の問題、数量の計算の問題、時空間の推論の問題、領域ごとの慣習的スタイル、知覚情報へのグラウンディング、オノマトペ、メタファーなど、雑多な問題が少しずつ混じっていて、それらの誤りのかけ算で精度が伸びない。「60%」の先はかなり性質の違う雑多な問題の組み合わせになっているように見える。本特集号に掲載された多様な分析は、そうした傾向が幅広いタスクに共通に見られることを示している。

「60%」の先の壁が雑多な問題の組み合わせだとすると、我々はそれにどう取り組めばよいのか？雑多な問題の一つ一つはそれぞれ性質も違っており、おそらくかなり異なる解決法を考える必要があるだろう。ところが、どれか一つにしぼって解決法を作ったとしても、その効果が見える問題は評価用データのごく一部にしか現れないので、精度への寄与を定量的に示すのは容易でない。評価用データの正解ラベルにノイズが混じっている場合はなおさらである。そうになると、研究する方も論文を書きにくいので、なかなか元気が出ない。このように、「60%」の先の雑多な問題を個別に解決しようとしても、その効果を測定できる自然なデータセットを作るのが難しい、したがってなかなか研究が進まない、という状況があるのではないか。

上のようなことが起こっているとすれば、今後考えられる方向性は少なくとも3つある。まず

¹ <https://sites.google.com/site/projectnextnlp/>

第1に、未解決問題の分析をその方法論も進化させながら継続し、課題の理解を深めていく必要がある。雑多に見えた問題には隠れた共通項があるかもしれない。課題分析の価値を適正に評価する土壌を学会が醸成していくことも重要であろう。第2に、雑多な問題の各々に対する処理をきめ細かく評価できるデータセットの設計・開発と、それを共有する仕組み、あるいは動機づける仕組みが今後ますます重要になると考えられる。クラウドソーシングによるデータ作成法の発展、データジャーナル等による研究データ出版の動きなど、我々を取り巻く環境もそうした方向を後押しするものと思われる。第3に、技術的には、雑多な問題を同時に解く結合学習に対する期待が今後さらに大きくなる可能性がある。近年、ディープラーニングによって複雑なタスクを end-to-end で実現する試みの報告がにわかに増えているが、こうした動きが「60%」の先の壁を崩すことができるか、個人的には極めて興味深い問いである。

冒頭に述べたように、本特集号が生まれた背景には、Project Next NLP と「エラー分析」ワークショップに対する多くの研究者の賛同と多大な貢献、そして一連の活動を牽引された関根聡氏（ニューヨーク大学）の並々ならぬ労があった。記して感謝の意を表したい。

乾 健太郎（東北大学）