

研究課題別事後評価結果

1. 研究課題名： 言語理解の能力に基づく機械読解ベンチマークの構築

2. 個人研究者名

菅原 朔（国立情報学研究所コンテンツ科学研究系 助教）

3. 事後評価結果

本研究では、自然言語処理システムがいかなる自然言語処理能力を有しているかを評価するために適切なベンチマークデータを整備することを目標としている。

ベンチマークデータの構築においては、各設問の解答に必要な能力を明確化し、それら設問への解答の質によっていかなる自然言語処理能力をシステムが有しているかを詳細に評価することを可能にする。

その目標に向けた試みとして、例えば入力文章からある要素を削除することで、その要素が解答に必要であるかを検証したり、データセットを用いた評価に対する意味づけを強化するために心理測定学の知見を用いて既存のデータセット設計がどれほど要件を満たしているかについて検討するといった研究を行ってきた。

これらの研究成果は AAIL2020 や EACL2021 といった人工知能および自然言語処理のトップ会議に採択されている。

研究を進めるうえで、設問を収集するためのウェブサイトを整備し、海外との共同研究も進めるなど、自立した研究者として積極的に研究を推進してきた。

現在、自然言語処理用のモデルは年々巨大化し、進化が止まらない状況であるが、それらシステムが実際にどのような能力を持ち、本当に人間が求めている能力を有しているかは慎重に評価する必要がある。

本研究はその問題に対する一つの有用なアプローチを与えており、今後も重要性が増すものと考えられる。一方で、今後もデータサイズのスケールが指数関数的に増大する分野であるがために、いかに本研究によるアプローチをスケールさせるかは今後の課題となろう。