

菅原 朔

東京大学大学院情報理工学系研究科
大学院生(博士課程)

言語理解の能力に基づく機械読解ベンチマークの構築

§ 1. 研究成果の概要

研究の背景:機械の言語理解能力を測るタスクのひとつとして国語の文章題のような形式の質問が課される機械読解タスクがあります。これまでに提案されている機械読解データセットにおいては、問題に正しく回答するためにどのような能力が必要とされているかの分析が少なく、成績の良いシステムがどのようなシステムを実際に備えていると言えるかについて根拠付けることが難しいという課題がありました。先行研究では人手によるアノテーション等を行って小規模のデータを分析していましたが、昨今使われる大規模なデータセットへの適用が難しく、自動的な分析手法が望まれていました。

研究の概要:「ある能力を行使するために必要となる要素」を課題文中から落としてもシステムが正答できるのであれば、その問いはその能力を行使しなくても解ける可能性が高いと言えます。たとえば課題文の全体の語順をランダムに並び替えても依然としてシステムが正しく回答できるのであれば、システムはその回答の根拠として課題文の語順の理解を必要としなかったのではないかと推測できます。そのような問いにおいては、語順を理解することを含む能力、たとえば統語構造の理解などが必要とされない可能性があり、言語理解の評価には足りない場合があります。以上の考察から既存のデータセットを分析し、多くのデータセットですでに解けている問いでは高度な言語理解が要求されているわけではなさそうである、という観察を得ました。より高度な言語理解の評価を行うためには、問いのデザインをより洗練させることが必要であると考えられます。本研究の成果は人工知能分野のトップ国際会議である AAAI 2020 に採択されました。

課題文の語順シャッフル

C: Chris Ulmer, the 26-year-old teacher in Jacksonville starts his class by calling up **each student individually to give them much admiration and a high-five**. I couldn't help but be reminded of Syona's teacher and how she supports each kid in a very similar way.
 Q: What can we learn about Chris Ulmer?
 A: He **praises his students one by one** (multiple choice)

C: his help a in calling class but Syona's starts each 26-year-old similar **individually** Ulmer, and Chris **admiration** way. Jacksonville kid much I by couldn't them the a to supports of in **student** and teacher **each** be teacher reminded give how she **high-five**. up very
 Q: What can we learn about Chris Ulmer?
 A: He **praises his students one by one** (multiple choice)

POS タグ + ID による語彙の匿名化

C: Immediately behind the basilica is the Grotto, a Marian place of prayer. It is a replica of the grotto at Lourdes, France where the Virgin Mary reputedly appeared to **Saint Bernadette Soubirous** in 1858.
 Q: To whom did the Virgin Mary allegedly appear in 1858 in Lourdes France?
 A: **Saint Bernadette Soubirous** (text span selection)

C: @adverb1 @prep5 @other0 @noun17 @verb2 @other0 @noun20 [...] @other0 @noun20 @prep6 @noun25 @punct0 @noun26 @wh0 @other0 @noun7 @noun8 @adverb3 @verb4 @prep4 @noun27 @noun28 @noun29 @prep2 @number0 @period0
 Q: @prep4 @wh2 @verb6 @other0 @noun7 @noun8 @adverb4 @verb4 @prep2 @number0 @prep2 @noun25 @noun26
 A: @noun27 @noun28 @noun29 (text span selection)