

P2-2 機械読解システムの推論過程のベンチマークの構築

近日公開予定

<https://naoya-i.github.io/r4c/>

井之上直也(東北大/理研) Pontus Stenetorp(UCL/理研) 乾健太郎(東北大/理研)

背景: 機械読解データセットに潜むバイアス

- 言語を理解せずとも表層的な手がかりにより回答可能な機械読解問題が多数ある (Sugawara+18, Mudrakarta+18, etc.)
- 他の分野 (含意関係認識, 因果関係認識) でも同様な発見あり (Gururangan+18, Kavumba+19, etc.)
- 問題: システムの性能の過大評価

関連研究

- 方策1: 難易度を上げる (e.g. マルチホップQA)
- 方策2: モデルの内部挙動を評価 (Yang+18, Jansen+19)
- HotpotQA (Yang+18): 回答 + 説明の出力を要求
- 説明: 回答の根拠となる文集合 (Supporting Facts)

■質問: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard...

[4] Mother Love Bone was an American rock band that... [5] The band was active from... [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to... [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's...

■回答: Malfunkshun, ■Supporting Facts: 1, 2, 4, 6, 7

- ⊙回答根拠文は説明としては粗い (回答の理由付けに無関係な情報を含んでいる)

提案: 新しい機械読解タスク R4C

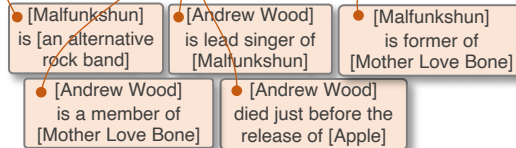
- 出力: 回答 + 説明 Right for Right Reasons Reading Comprehension
- 説明: 回答を導く entity 間の関係の集合 (導出)

■質問: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard...

[4] Mother Love Bone was an American rock band that... [5] The band was active from... [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to... [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's...

導出



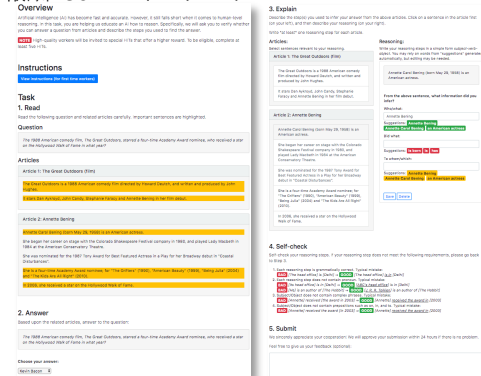
■回答: Malfunkshun

評価尺度

- ROUGE-L (Lin+04) による複数参照導出に基づく評価
- Entity 単位, Relation 単位, 全体での評価
- ⊙回答の導出に無関係な情報を出力するとペナルティ

クラウドソーシングによるデータ構築

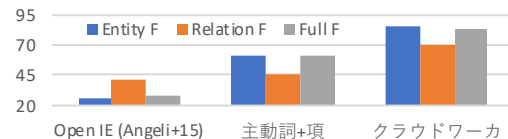
- HotpotQA 4,747 事例に付与 (3参照導出/1事例)
- 事前スクリーニングテストの実施 (効果大)
- 報酬: 30セント/1タスク



- 品質: 付与された導出のうち93.9%から回答を復元可

ベースラインモデルによる評価実験

- Golden Supporting Facts を用いた実験



- 回答根拠文のすべての情報は導出には含まれない
- 人間の性能との隔たりは大きい