P5-12 論理的根拠に基づく頑健な機械読解に向けて

原口 大地 (JAIST)　白井 清昭(JAIST)　井之上 直也 (JAIST/理研)

背景

- 機械読解モデルにおける頑健性の課題(Jia+17, Sen+20)

Article: Super Bowl 50
Paragraph: *"Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV."*
Question: *"What is the name of the quarterback who was 38 in Super Bowl XXXIII?"*
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

(Jia+17 より引用)

- 説明性の課題(Paranjape+20, Inoue+21)

関連研究

- Adversarial Training(Jia+17, Jiang+19): 自動生成した敵対的事例によるモデル強化

→ 頑健性向上に焦点

- Chen+22: 説明機構追加による説明性・頑健性の向上

→説明性は向上、しかし頑健性の向上は限定的
本研究: Explain-then-predictによるアプローチを再検討

仮説

1. ExplainerがShortcut reasoningしている
2. Shortcut reasoningしないExplainerを作成することで，説明性と頑健性の向上した機械読解モデルを実現できる

仮説1の例と検証方法

- Shortcut Type Ⅰ: 質問に依存

e.g. "*What is the name of…*" → 名前のある文を選択

- Shortcut Type Ⅱ: 文書に依存

e.g. 固有表現が含まれる文を選択

→入力の質問・文書を編集，説明の変化を分析

Edited Question:
- First half　*"What is the name of the quarterback"*
- First word　*"What"*
- No question　*NaN*
- Shuffle　*"Where was the largest single capital investment IBM made have been built?"*

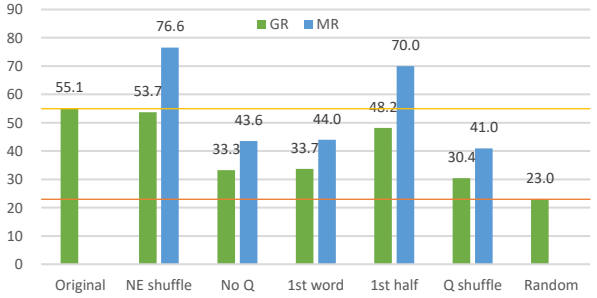
Edited Document (NE shuffle):
"39 became the first quarter- back ever to lead two different teams to multiple Champ Bowl XXXIV. He is also the oldest quarterback ever to play in a 38 at age Super Bowl. The past record was held by Super Bowls, who led the Broncos to victory in Super Bowl XXXIII at age Jeff Dean and is currently John Elway and General Manager. Quarterback Peyton Manning had jersey number 37 in Denver's Executive Vice President of Football Operations."

- モデル: VIB (Paranjape+20)
- データセット: MultiRC (Khashabi+18)

検証

評価尺度

- GR: 正解の説明に対する文単位のF1
- MR: 編集前と後の説明の一致度(F1)



- GRはすべての編集でRandomを上回る

→一定のshortcutが想定される

- NE shuffleのGRはほぼ下落せず・MRも高い

→文書中の固有表現の内容を考慮していない

今後の課題

- 文書の編集による検証の追加
- 異なるデータセット・モデルによる検証
- 新しいExplainerの提案(仮説2)

アイデア: 文書のパターンマッチ解決・質問後方の読解