ファクトチェックの自動化は可能か？

機械学習に基づく自動ファクトチェックとその多言語拡張

山岸 順一 (コンテンツ科学研究系)

どんな研究？

- 虚偽の主張・情報は、フェイクニュース、プロパガンダ、SNSなど、様々な形で発生
- 手作業のファクトチェックは時間がかかる
- ファクトチェックを機械学習で自動化！

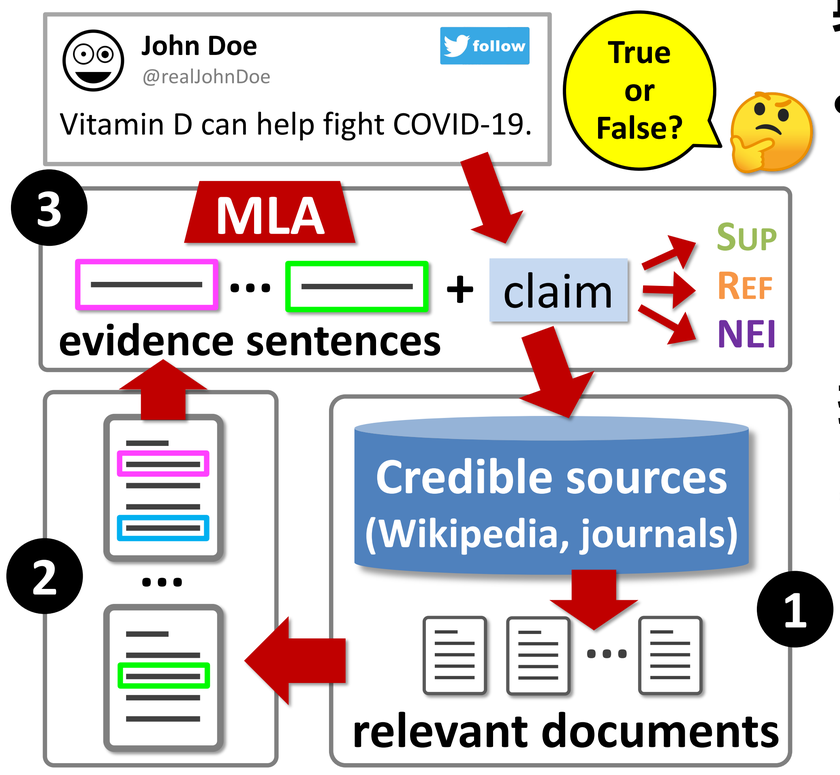
何を検討している？

- 主張の真偽を検証するには？
- ファクトチェックにどう機械学習を使う？
- 他言語の主張を検証するには？
- 表データも一緒に使えるか？

機械学習による自動ファクトチェック

根拠に基づく自動ファクトチェックの流れ：３ステップで構成

1) **ドキュメント検索:** 信頼できる情報源から関連ドキュメントを検索
2) **根拠文選択:**検索により見つかったドキュメント内の根拠となり得る文章を複数選択
3) **主張の検証:** 主張と根拠文を照らし合わせ、以下のどれかを予測:
支持可能, **反論可能**, **情報不足**



英語でのベンチマーク実験

- **FEVER** (Fact Extraction and VERification)：人手により検証された主張18.5万件を含む大規模データセット

提案機械学習方式

- 根拠文選択と主張検証モデルを階層アテンションニューラルネットワーク (**M**ulti-**L**evel **A**ttention) により学習

自動ファクトチェックシステムの多言語化や表データの利用

多言語ファクトチェック

- FEVERデータセットの検証済み主張と根拠文をスペイン語・フランス語・インドネシア語・日本語・中国語に機械翻訳することで **XFEVERデータセットを構築！**

実験結果(精度)

学習手順	英語	スペイン語	フランス語	インドネシア語	日本語	中国語	平均
mBERT	87.9	83.7	84.3	82.6	72.4	82.1	82.2
XLM-R-base	87.7	83.7	81.3	81.9	74.4	78.0	81.2
XLM-R-large	**89.5**	**87.3**	**85.3**	**85.5**	**82.0**	**83.1**	**85.5**

テキストデータに加え、表データも利用

- テキストエンコーダに加え、表エンコーダ（TAPAS, Tapexなど）も利用できる様にネットワーク構造を拡張
- テキストの情報と表データの情報をアテンションネットワークにより自動融合
- **テキストのみの場合よりも予測精度向上**

- **Claim:** Aramais Yepiskoposyan played for FC Ararat Yerevan, an Armenian football club based in Yerevan during 1986 to 1991.
- **Label:** SUPPORTS
- **Evidence:** Aramais Yepiskoposyan_cell_0_6_1, FC Ararat Yerevan_sentence_1

NII
連絡先：山岸研究室／ 国立情報学研究所　コンテンツ科学研究系
TEL：03-4212-2576　Email：ylab-pr@nii.ac.jp