

# Yahoo!知恵袋を利用した貪欲な情報探索を促す問いかけの発見

静岡大学大学院 総合科学技術研究科

齊藤史明 山本祐輔

## 目的

ウェブ検索中に、思わず調べたくなる質問を問いかけるシステムを作成したい。  
そのために、Yahoo!知恵袋に投稿された質問を学習器を利用して評価する手法を提案する。

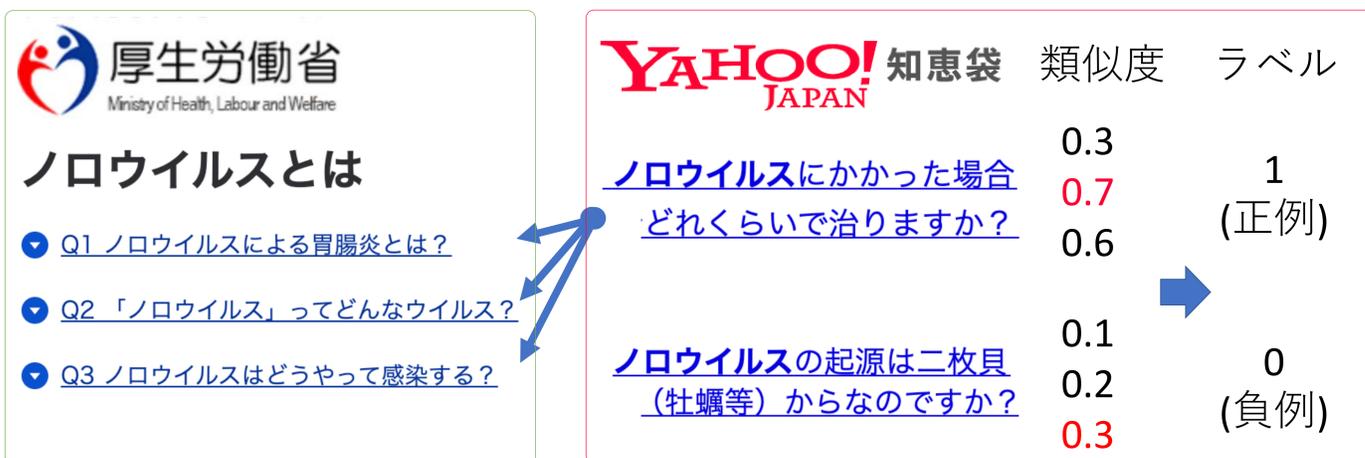


学習器による評価



## 質問を評価する学習器の作成

2値分類問題として教師あり学習による学習器を作成する。  
正解データとして公的機関の用意したQ&A集とのcos類似度を利用する。



## 特徴量

- 学習済みBERTモデル(東北大学:乾研究室<sup>[1]</sup>)によるベクトル
- Bag-of-Wordsベクトル(トピックの影響を受けるため名詞以外)
- 質問に投稿された回答数
- 質問文群でのTextRankの数値

## 学習器

- LightGBMのLGBMClassifier<sup>[2]</sup>

## 性能評価

特徴量	F1値
回答数	0.380
TextRank	0.493
回答数+TextRank	0.512
回答数+TextRank+BERT	0.832
回答数+TextRank+Bag-of-Words	0.663
回答数+TextRank+BERT+Bag-of-Words	0.832

## 提案手法による問いかけの発見

トピック：健康食品

	回答数	TextRank	回答数+TextRank+BERT+BoW
1位	ノンカフェインのコーヒーはネットでしか手にはいりませんか？薬局などの健康食品コーナーなどに行けばありますか？	糖鎖って効くのでしょうか？今話題の糖鎖という健康食品？がありますが、例えば赤ちゃんのミルクや離乳食などに溶かして飲ませると、発達や発育の手助けになったりとかするのでしょうか？	健康食品店は詐欺なのですか？
2位	〇〇の健康食品を飲みました。薬じゃないのは分かってますが効果があるのか無いのか、いまわかりません。確認する方法ありますか？	史上、最強の健康食品はなんだと思いますか？	胃を強くするには、どんな健康食品が良いのでしょうか？
3位	私、〇〇市内の腎臓病患者です。蛋白質や塩分の制限が必須です。スーパーで買物するにも一苦労です。患者の方ならよくご存じだと思いますが〇〇市内でおすすめのスーパーか健康食品店などご存じありませんか？	健康食品など成分を濃縮している物って残留農薬も濃縮されますか？	健康食品は健康に良いのですか？

## まとめ

特徴量としてBERTが特に有効であることが分かった。  
今後は、実際に情報探索中のユーザに問いかけを表示することが情報探索の態度・行動に影響を与えるかユーザ実験を行う予定である。

## 参考

[1] <https://github.com/cl-tohoku/bert-japanese>

[2] <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html>