

# ニュースの内容とコメントの 類似度が投稿するコメントに 与える影響の分析

西原陽子, Sheng Wang,  
Junjie Shan

立命館大学情報理工学部

# 研究背景（1）

ニュースサイトやSNSで個人がコメントを書き込むことが一般的になった

- ・ 米国でニュースサイトを見る55.3%がコメントを読み書きする, 24.6%がコメントは書かないが読む [Stroud, 2016]

コメントから他者の意見を知ることができ, ニュースの理解に役立てることができる [中原, 2022]

こんなニュースがあるんだ.

他の人は, こんなこと考えてるんだ.



## 研究背景（2）

コメントの中には、ニュースと関連が  
高くないものも含まれる[Mozafari, 2019]

- 関連が高いコメント：より深い考察へと発展させる可能性がある
- 関連が高くないコメント：ニュースの理解を混乱させる可能性がある

関連が高いコメントとそうでないものが混在  
するとニュースを読むうえで支障が生じる

こんなニュースがあるんだ。

他の人は、こんなこと考えてる...の？



# ニュースとコメントの関連性に関する既存研究

Mozafariらの研究では、ニュース記事に対するコメントの関連性を分析している[Mozafari, 2019]

- 関連性を類似度で評価。
- **関連性がある**コメントは客観的でニュース内容を主題としたものが多い
- **関連性がない**ものは主観的で、ニュース内容を主題とせず、一般的な単語が多く含まれる

コメントを重要度順にランキングする手法が提案されている[Wei, 2016][藤田, 2018][田淵, 2019]

- 説得力のある順、建設的な順など
- ニュースを読むユーザのニュースの理解を助けることができている

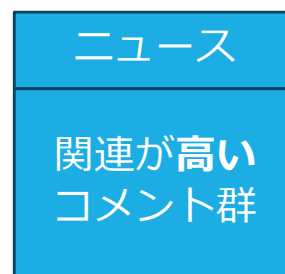
ニュースに対してコメントを書くユーザに対して、提示するコメントを変更することはどのような影響を与えるのか？

# 本研究の仮説

主仮説：人は、ニュースと既に投稿されたコメントを読んで自分の新たなコメントを投稿するならば、既投稿コメントと新コメントは関連する

ニュースとコメント、コメント間の関連はテキストの**類似度**により評価する

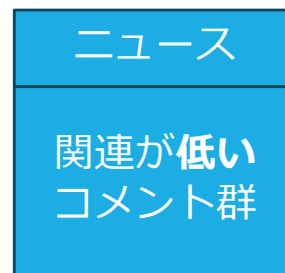
コメントを書くユーザ



+



新コメントは  
関連が**高**くなる？



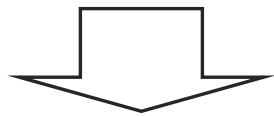
+



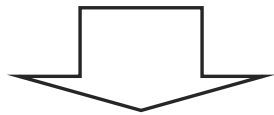
新コメントは  
関連が**低**くなる？

# テキストの類似度を評価する モデルの選定

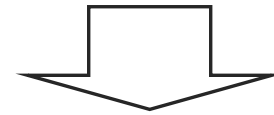
ニュース コメント



埋め込み表現



類似度を評価



類似度で並び替え

言語モデル	パラメータ
Wikipedia- Doc2vec	アルゴリズム:dmpv
	入出力ベクトル長:300
Japanese- RoBERTa	隠れ層のニューロン数: 768
	ニューラルネットワーク: 12層
	入出力ベクトル長: 512

評価手法

コサイン

内積

# 選定方法

被験者アンケートを行い、アンケート結果を用いてモデルを選定

## アンケート手順

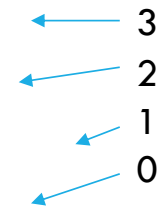
- (1) 被験者を募集し、ニュースとコメントを見せ、関連性を4段階で評価してもらう
- (2) 4パターンのモデルでニュースとコメントの類似度を算出する
- (3) 関連性と類似度の分布から、最も適切な類似度モデルを選定する

## 被験者

- 20代の男女30名、1つのコメントを3~5名で評価

## 使用したデータ

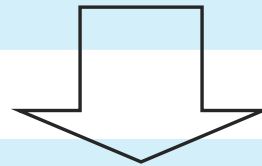
- Yahoo!ニュースの「国際」カテゴリから3件、コメントは $90 \times 3 = 270$ 件



# 選定結果

言語モデル	計測手法	決定係数
Wikipedia-Doc2vec	コサイン	0.0997
	内積	0.1345
Japanese-RoBERTa	コサイン	<b>0.1608</b>
	内積	0.0852

Japanese-RoBERTaを用いるコサイン類似度モデルでの決定係数が最も高くなった



したがって、Japanese-RoBERTaコサイン類似度モデルを用いて仮説を評価する実験を行う



# 仮説の評価実験

## 主仮説：

- 人は、ニュースと既に投稿されたコメントを読んで自分の新たなコメントを投稿するならば、既投稿コメントと新コメントは関連がある

## 実験手順

- ニュースとコメント集合を用意し、ニュースとコメントの類似度を評価。類似度が高いコメント集合と低いコメント集合を作成する。
- 被験者に、ニュースと作成されたコメント集合を示し、新しいコメントを書いてもらう
- 被験者に、ニュース内容、示されたコメント内容についてのクイズに回答してもらう
- 示したコメントと新しいコメントの類似度を評価する

# 評価実験の詳細

## 使用したニュース

- Yahoo!ニュースのカテゴリ「国際」「国内」「エンタメ」「スポーツ」の4つから、5件ずつ、合計20件

## 使用した提示コメント

- ニュースとの類似度が高いコメント、低いコメントをそれぞれ20件ずつ
- コメントの平均文字数はほぼ同じ（高いコメント集合：92.77、低いコメント集合：87.96）

## 被験者

- Yahoo!クラウドソーシングで募集した合計600名
- 1人の被験者は、5つのニュースとコメント集合が割り当てられ、それぞれに新しいコメントを書いた

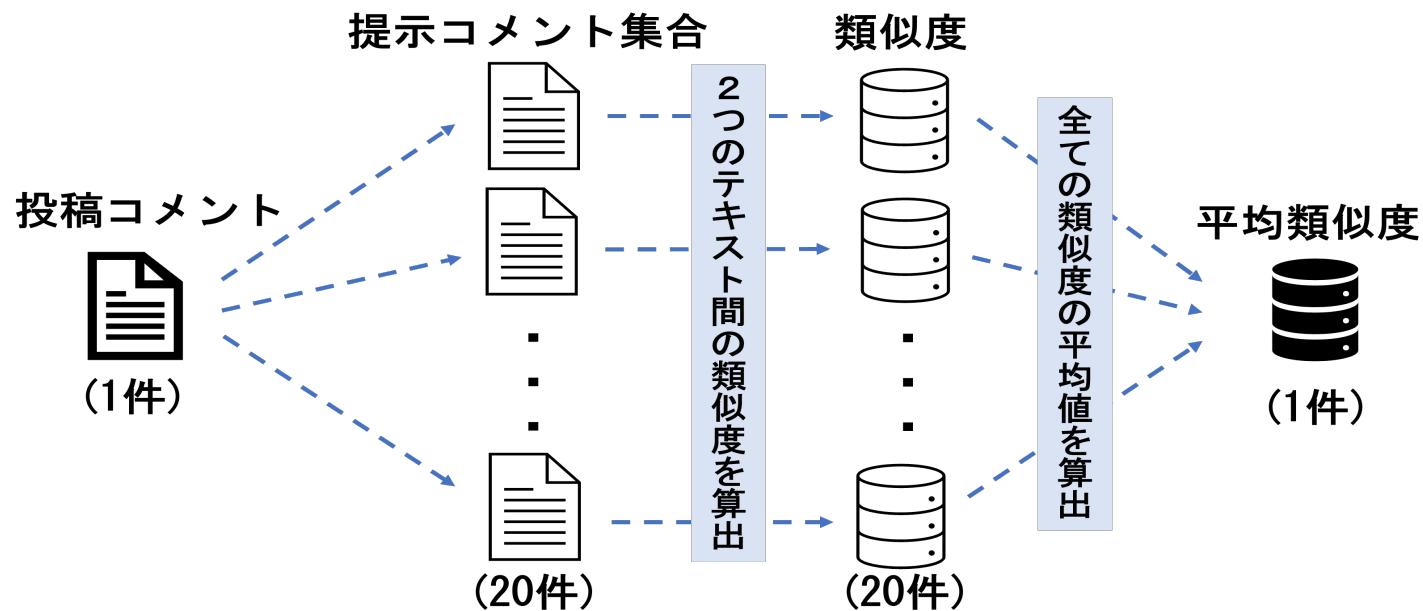
# 評価項目

## 1. 提示したコメントと新コメントの類似度

- 提示されたコメントを参考にした上で，新コメントを書いたかを評価する

## 2. 新コメントとニュースの類似度

- 提示するコメントが異なると，新コメントとニュースの類似度が異なるかを評価する



# 実験結果 1

## 提示コメントと新コメントの類似度

カテゴリ	類似度が高いコメントを提示	類似度が低いコメントを提示
国際	0.6281	0.6226
国内	0.5963	0.6008
エンタメ	0.6130	0.6061
スポーツ	0.5808	0.5936
平均	0.6058	0.6064

提示コメントが異なっても新コメントの類似度は異ならなかった  
新コメントは提示コメントに影響を受けないと考えられる

## 実験結果 2 : 提示コメントを参考にして書かれた新コメントと ニュースの類似度の結果

提示コメントとの類似度が、全体の類似度の平均値よりも高い場合に  
新コメントは提示コメントを参考にして書かれたと判断し、コメントを抽出

カテゴリ	類似度が高いコメントを提示	類似度が低いコメントを提示
国際	0.7252	0.6005
国内	0.6800	0.6172
エンタメ	0.7016	0.6285
スポーツ	0.6667	0.5899
平均	<b>0.6949</b>	0.6100

提示されたコメントを参考にして書かれた新コメントはニュースとの関連が変化する  
可能性が見られた

# 提示コメントを参考にして書かれた新コメントの例

ニュースタイトル：

ローラスケート韓国代表がゴール前にガッツポーズして金メダルと兵役免除を逃す

## 投稿コメントの例

類似度が低いコメントを提示

日本人だってやりかねないなと思いました。モラルのない若者も多いですし。

韓国の話題になるとコメントしてる人たちの本音が出てると思う。

なんでコメントする人たちはこんなに韓国の選手をバカにしてるのか、なぜそんなに上から目線なのか、人間の醜い部分を見た感じである。

類似度が高いコメントを提示

金メダルが銀メダルに変わってしまったのはとても残念でした。団体競技なので最後まで緊張感をもってやればチームにも迷惑が掛からなかったと思います。

ゴールまでが競技であることを理解してスポーツに臨んでほしい。

可哀想だが、最後まで気を緩めてはダメだと思った。

# おわりに

本発表では、「人は、ニュースと既に投稿されたコメントを読んで自分の新たなコメントを投稿するならば、既投稿コメントと新コメントは関連がある」かを調べた実験を報告した

ニュースとの類似度が高いコメントと低いコメントを提示したときに、新しく書かれるコメントの類似度が影響を受けるかを調べた

実験の結果2つの可能性が分かった

- ニュースと既投稿コメントを読んだ上で新しいコメントを投稿するように指示しても、多くの方は既投稿のコメントはあまり参考にしない。
- コメントを参考にした場合は、ニュースとの関連が高くなるコメントが得られる可能性がある