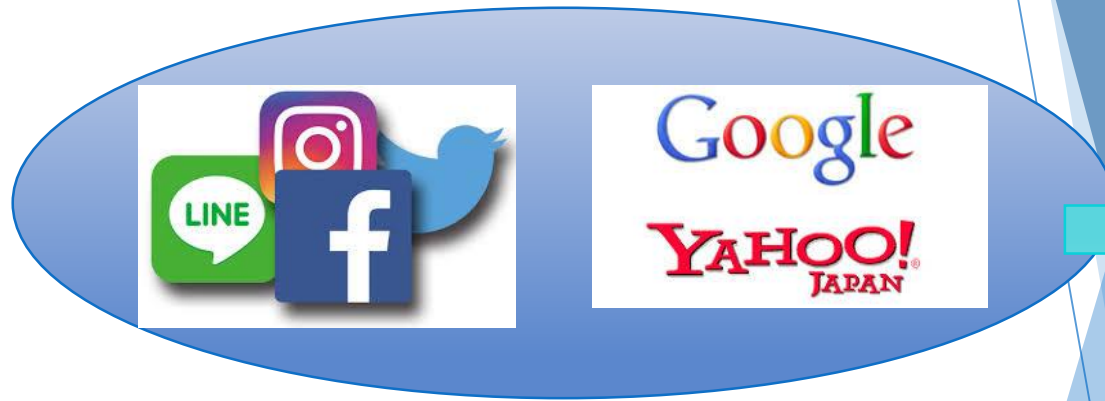


# ニュースにおける異メディア間の 差異情報抽出手法

灘本研究室 11571027 大中ちづる

# 背景

- ▶ 世の中はたくさんのニュースメディアで溢れている



- ▶ それぞれ見るメディアによってニュースに対する意見が異なる

- ▶ **人は好きなサイトやニュースばかり見てしまう傾向がある**



# 問題点

利用者が好ましい  
と思う情報ばかり  
が選択的に  
得られてしま

価値観が似た者同士  
の共感による  
意見

**自分とは違う見方や  
意見を見ることで視野  
を広げる**

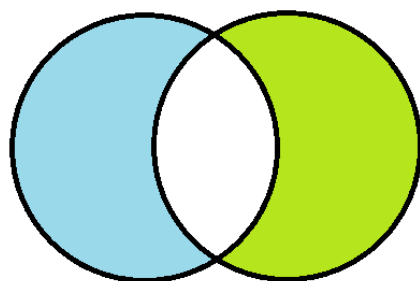
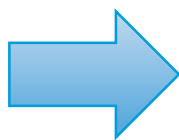
▶ フィルタ

視野が

に回収してしま

# 研究目的

- ▶ ある同じテーマのニュースに対して異なるメディアから一方のメディアにしかない情報(差異情報)を抽出



新聞やテレビ



**差異情報**

ソーシャルメディア

# 差異情報抽出の例

## ▶ テーマ: 「加計学園」

### ツイート

- ・ 安倍首相の説明責任は果たせていない
- ・ 詐欺学園、補助金返せ！

比較

### 社説

- ・ 野党は疑惑の徹底追及をすべき
- ・ 安倍首相は説明責任を果たせ

### ツイート

詐欺学園、補助金返せ！

### 社説

野党は疑惑の徹底追及をすべき

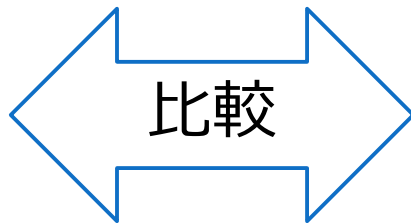
## 本研究の効果

- ▶ ユーザが見ていないメディアからニュースに対する差異情報を提示することで
  - ユーザの視野を広げる材料になる
  - ユーザの情報を分析する能力の向上に繋がる
  - ユーザの情報リテラシーの向上に繋がる

# 差異情報抽出の研究対象



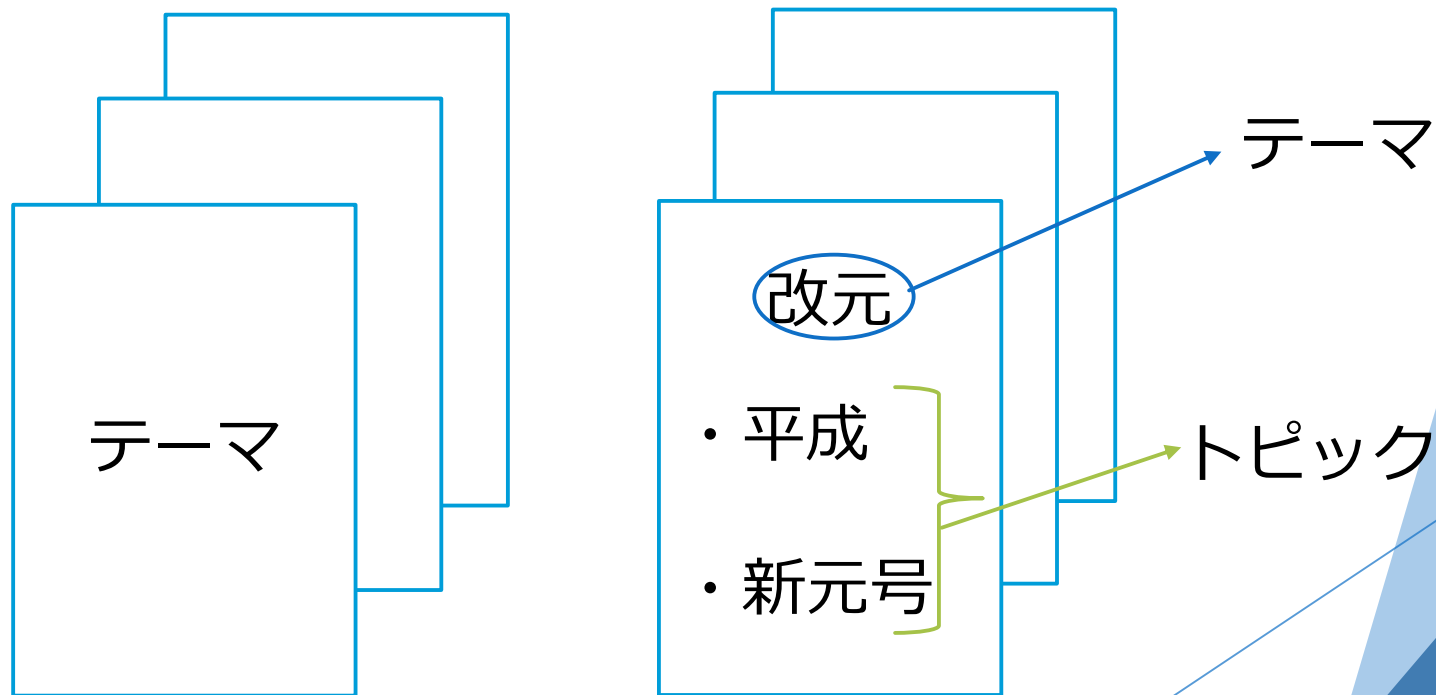
Twitter



新聞の社説

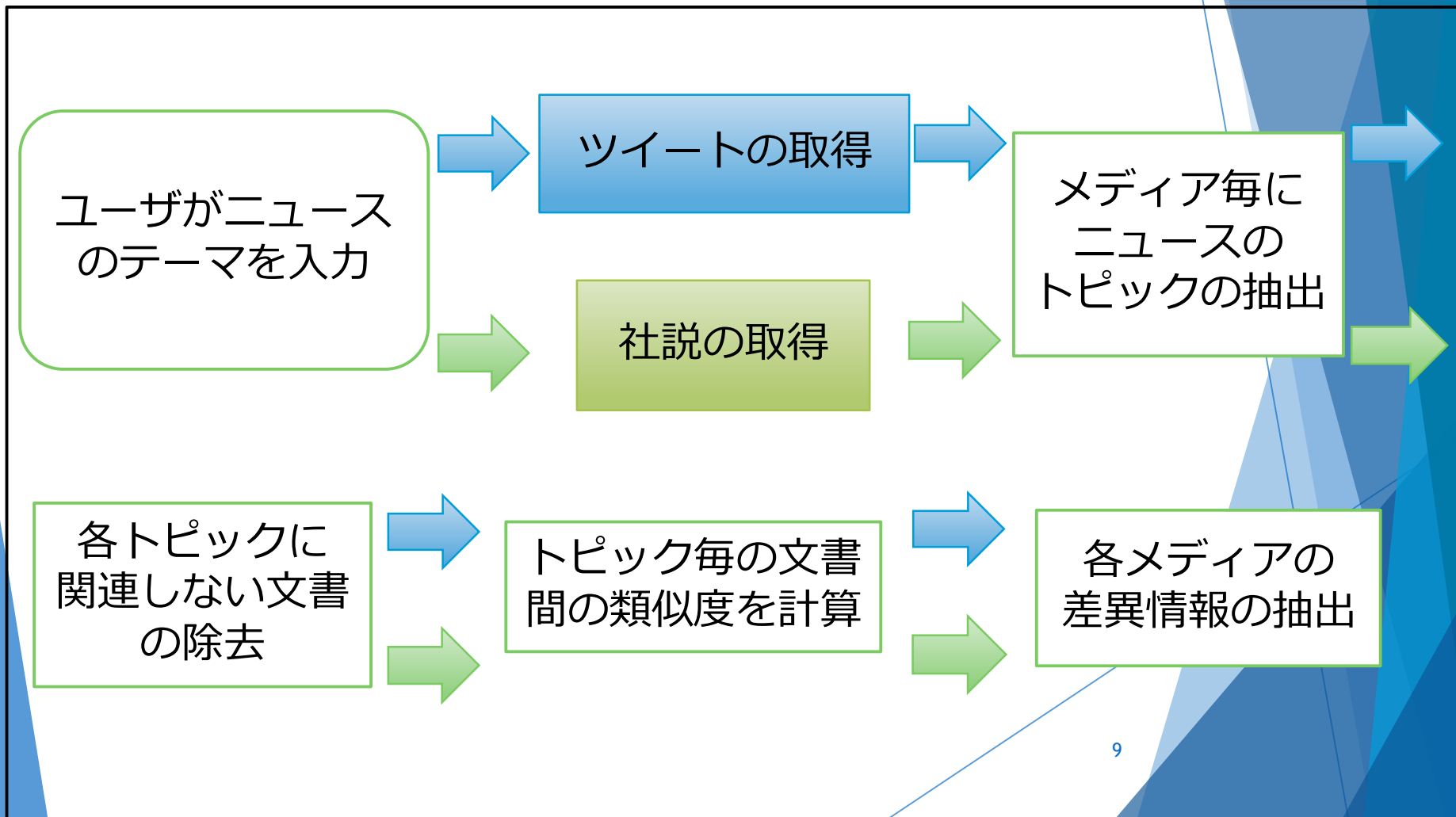
# 用語の定義

- ▶ テーマ: ユーザが入力した単語
- ▶ トピック: テーマのニュースの中に含まれている単語





# システムフロー

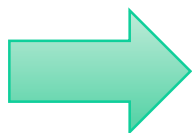
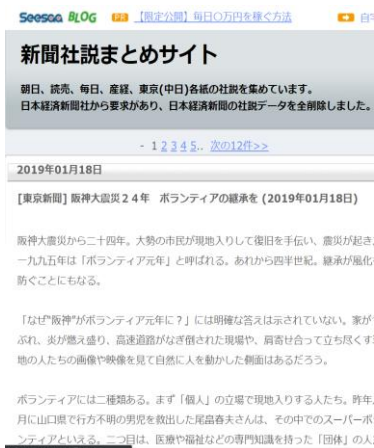


# システムについての説明

- ▶ 社説とツイートの取得方法
- ▶ 社説とツイートからのトピックの抽出
- ▶ 各トピックに関連しない文書の除去
- ▶ 差異情報の抽出方法

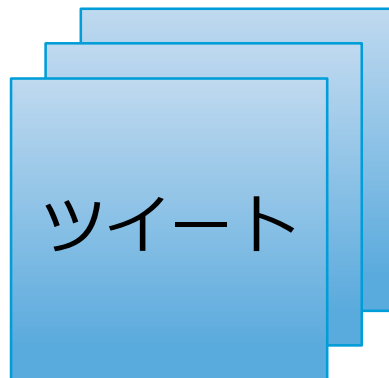
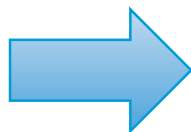
# 社説とツイートの取得方法

## ▶ 社説



ニュースのテーマを含む社説(段落毎)の取得

## ▶ ツイート



ニュースのテーマを含むツイートの取得

# システムについての説明

- ▶ 社説とツイートの取得方法
- ▶ 社説とツイートからのトピックの抽出
- ▶ 各トピックに関連しない文書の除去
- ▶ 差異情報の抽出方法

# 社説とツイートからのトピック抽出

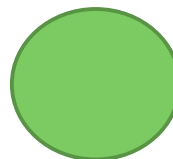
▶ 社説



トピック1      トピック2

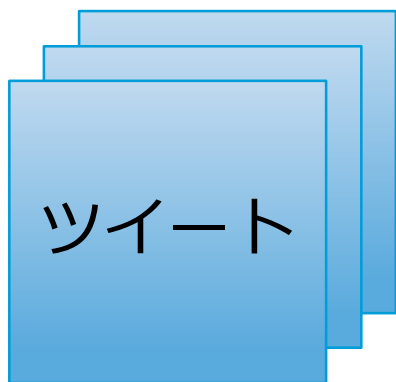


トピック3

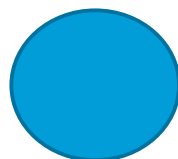


トピックモデルを用いて  
 (Repeated Bisection法)で  
 クラスタリング

▶ ツイート



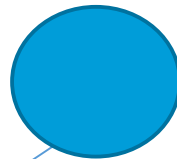
トピック4



トピック5



トピック6

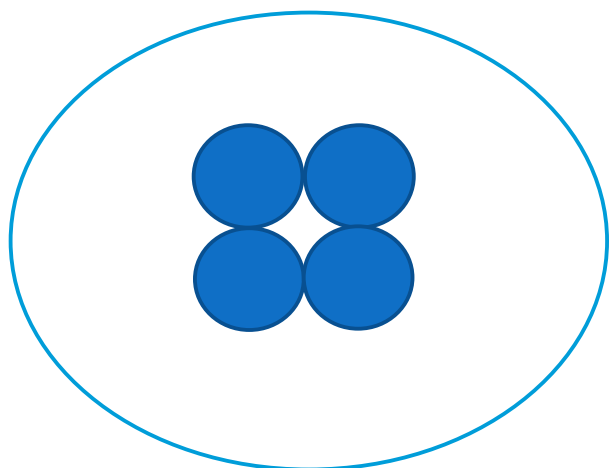


# システムについての説明

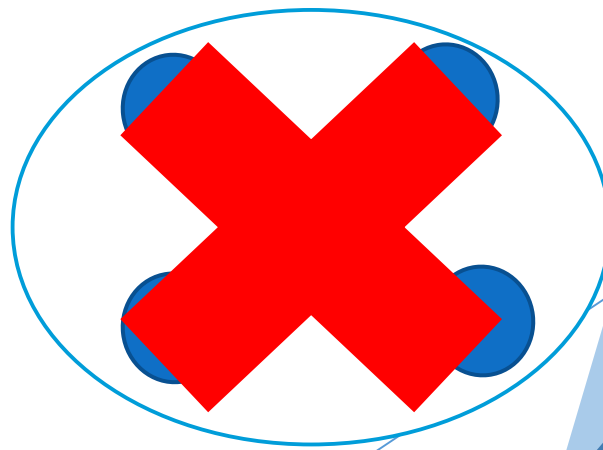
- ▶ 社説とツイートの取得方法
- ▶ 社説とツイートからのトピックの抽出
- ▶ 各トピックに関連しない文書の除去
- ▶ 差異情報の抽出方法

# 各トピックに関連しない文書の除去

- ▶ クラスタリングした文書集合の中にはニュースのテーマに関連しないクラスタが含まれ、事前に取り除く必要がある
- ▶ クラスタ内の文書の密度に着目



高



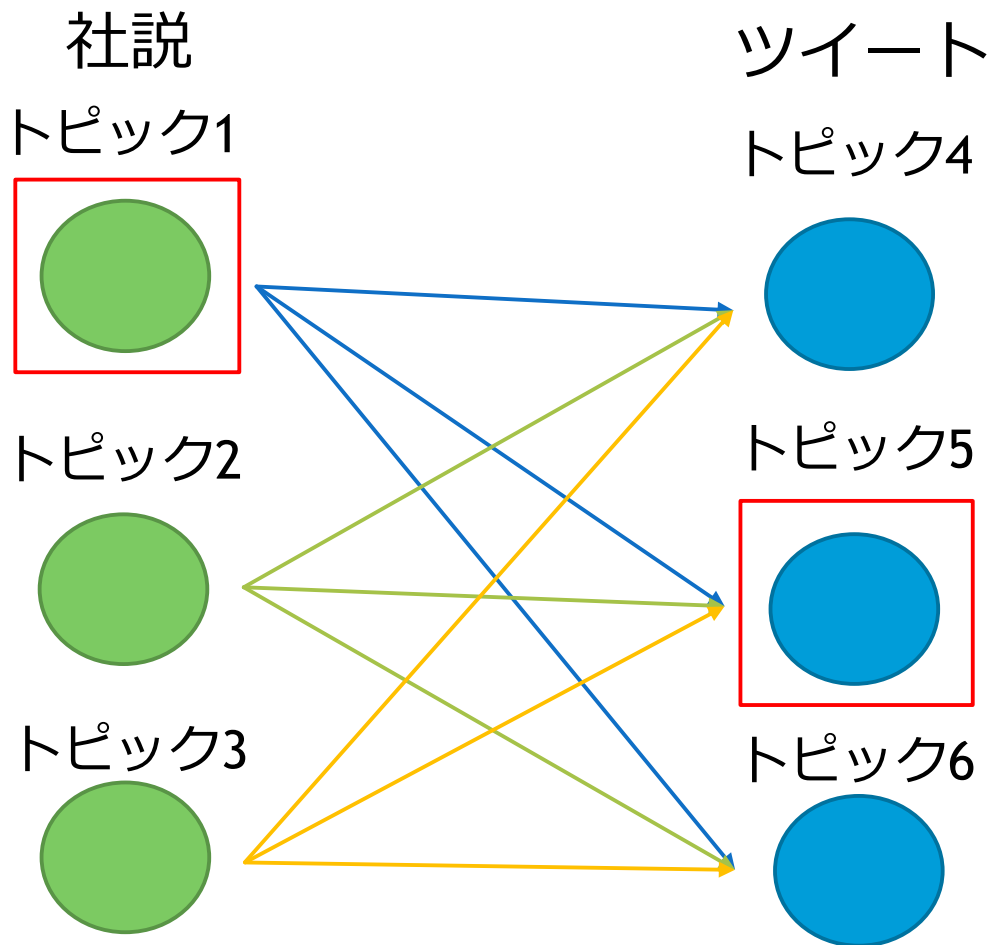
低

# システムについての説明

- ▶ 社説とツイートの取得方法
- ▶ 社説とツイートからのトピックの抽出
- ▶ 各トピックに関連しない文書の除去
- ▶ 差異情報の抽出方法



# 差異情報の抽出方法



どのクラスタに対しても閾値以下の類似度を持つクラスタに含まれる文書群(**差異情報**)

トピック同士でクラスタ間の  
 コサイン類似度を求める

# 文書間の類似度の閾値

- ▶ コサイン類似度の閾値:0.3
- ▶ 社説とツイートの各クラスタ内の文書同士を比較して判断

# 差異情報の提示

## ▶ 差異情報を社説とツイート別に表示

**大阪万博**

← ニュースのテーマ

社説のトピック

社説

← ツイートのトピック  
 ツイート

**科学**

万博のテーマの実現に貢献できるような、質の高い生命科学や医療の研究機関、企業が関西には集積している。万博を通じてそれらが生かされ、さらに発展できることが望ましい。

「なぜ大阪で『健康・長寿』の万博なのか」という疑問もある。大阪府は「生命科学の研究機関や企業が関西圏に多い」などと説明するが、それでは説得力が乏しい。

2度目の大阪万博の標語は「いのち輝く未来社会を築こう」。関西の企業が第一に万博を

**"テーマソング"**

大阪万博の頃ですね。万博そのものよりあのテーマソングが流れるたびに耳を塞いで逃げ出していました。

大阪万博のテーマソングは流石に覚えてませんねえ…。行ってもないし…(幼稚園の頃かな…

ゆず 解散か？って話も出てるけど 東京五輪のテーマソングをゆずが担当っていう嬉しい報告だと良いね♪♪ 前回の東京五輪、大阪万博のテーマソングを三波春夫だったから今回の東京五輪、大阪万博のテーマソングをゆずで！！って発表だと良いな♪♪ゆずゆず今後の活動に

# 差異情報の提示

ニュースのテーマ

**カルロスゴーン**

社説のトピック

ツイートのトピック  
 ツイート

社説

**行為**

容疑事実通りであれば、投資家への重大な背信行為である。

◆投資家への背信行為だ

容疑事実通りなら、日産への重大な背信行為である。日産の取締役会は、一連の経緯をどこまで把握していたのか。銀行側の対応の是非が問われる可能性もある。

日産は、ゴーン被告には複数の不正行為があると主張する。

**敗北**

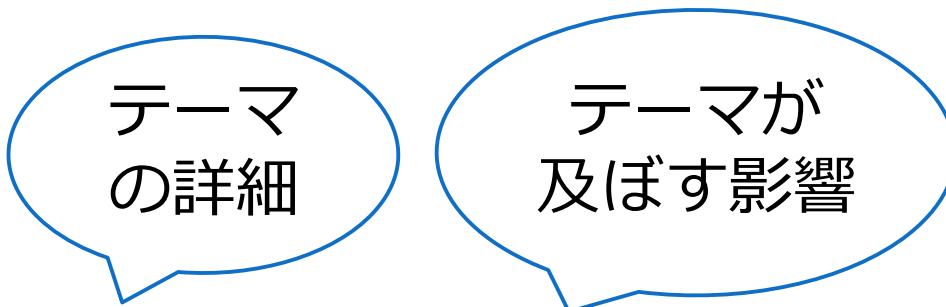
日産のクーデター失敗で西川社長が明智光秀になる日 ゴーン再逮捕も特捜部敗北の危機 WSJの報道内容は眉唾としても、逮捕後に聞こえてくる情報や再逮捕容疑などが、『検察大丈夫か？』ばかりなんだよね、本当に。

報酬をもらって、申告しないなら理解できるが、退任後の約束だけで（たとえあるとしたら）まだ貰っていないのに申告しなければならぬか？日産のクーデター失敗で西川社長が明智光秀になる日 ゴーン再逮捕も特捜部敗北の危機

# 実験

- ▶ テーマ:「大阪万博」,「カルロスゴーン」
- ▶ 実験条件
  - 「大阪万博」  
(社説(段落毎):496件、ツイート:423件)
  - 「カルロスゴーン」  
(社説(段落毎):616件、ツイート:606件)

# 結果



テーマ	社説	ツイート	適合率
大阪万博	松井 ギャンブル テーマ	2兆 活性 夢洲	0.48(社説) 0.38(ツイート)
カルロスゴーン	虚偽 責任 拘留	特捜 支配 組織	0.21(社説) 0.56(ツイート)

社説: テーマにピンポイントなトピック  
 ツイート: テーマから派生したトピック

# 結果と考察

## ▶ 結果

- ・ 社説はニュースのテーマに直接関する内容
- ・ ツイートはニュースのテーマと間接的に関連する内容

## ▶ 考察

どちらのメディアにもニュースのテーマに関連しない  
文書が存在

→ コサイン類似度の閾値の決定の改善が必要

# まとめと今後の課題

## ▶ まとめ

同じニュースのテーマにおける異メディア間の差異情報の抽出

- 社説とツイートの比較
- 社説とツイートからのトピックの抽出
- 差異情報の抽出手法

## ▶ 今後の課題

- 閾値の決定方法の改善
- 時事関連のテーマ以外での実験