

ストーリーコンテンツに対する ネタバレの基礎調査とその判定 手法の検討

田島 一樹(明治大学 総合数理学部B3)

中村 聡史(明治大学 総合数理学部, JST CREST)

本研究での貢献

1. 放送時間差によって視聴者の**70%**はネタバレをされてしまう可能性があることが明らかに
2. アニメのネタバレを345件収集、それらを12種類にカテゴリ分けし、その中から「**正体・生死・勝敗人物特徴**」に関する情報を致命的なネタバレとする
3. **勝敗のネタバレの推定は容易**であり、**正体のネタバレの推定は困難**であることが明らかに

背景

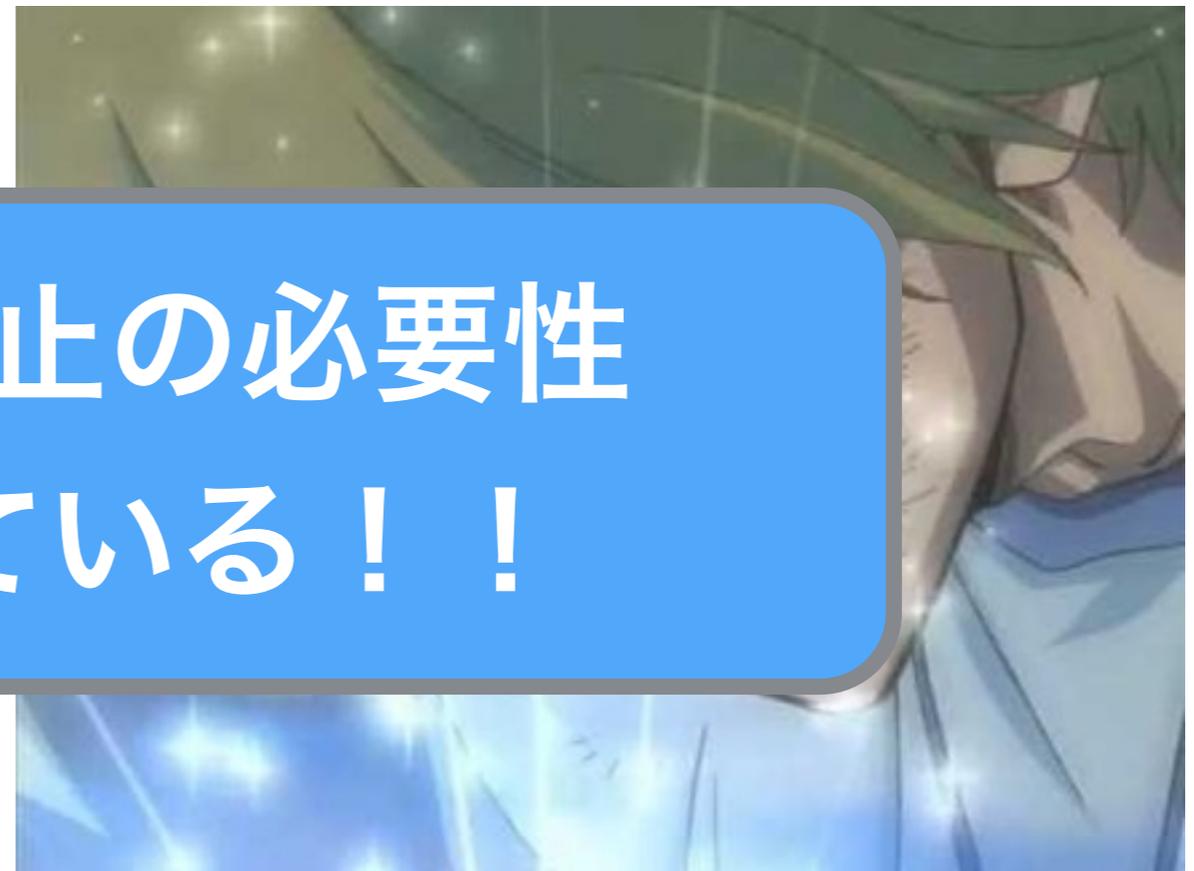
- ・気になっていた黒幕の正体に関する情報をネット上で発見してしまった
- ・登場人物が死ぬかどうかのハラハラする展開だったのに次回予告で結末が分かってしまう

嘘でしょ！？
～が犯人

【悲報】

可愛い顔
んてこと

ネタバレ防止の必要性
が生じている！！



考えられる対策①

- ・そもそも番組を観るまではTwitterなどのSNSを使わない

Twitterは友人・知人との**コミュニケーション**や**情報収集(ニュースなど)**のための**ツール**として必要不可欠なので**非現実的**



考えられる対策②

- ・ハッシュタグ付きの投稿をブロック



あさひな(通信制限によりツイ禁 @... 1h

#sazaesan

これちょっとかっこよくなったけど要
はサザエさんだろww

Details

- ・何の話題か明示
- ・他の投稿も検索可能



うら @urarander

1h

負けた #sazaesan

Details

**ハッシュタグを付けずに
投稿する人も多く、不完全**

考えられる対策③

The Twitter Mute Button [Golbeck 2012]

- ・海外における時差によって生まれるTwitter上でのネタバレを問題に
- ・スポーツやドラマに関する投稿内の単語を登録し、それらが含まれる投稿をミュートすることでネタバレを防止

ネタバレでない投稿までブロックする

→ 日常的なコミュニケーションまで妨げ

られてしまう可能性が高い

に投稿されてる…



た

ネタバレに対する反応



おかべ@ミルキィ幕張(連番者募集) @rafie1206

33分

シャーロットのリアタイ間に合わないから無限にネタバレがTLに流れ込んでくるやつだ



個人的な事情



かるら@期末前低浮上 @rara_poco

2015/06/29

岐阜はうた☆プリ明日なんだよ(-ω-)クッ

まじ東京勢お願いだから黙ってほしい
これ以上ネタバレすんな

見なきゃいいじ
ロワーさんがツ



放送局毎の放送日時の違い

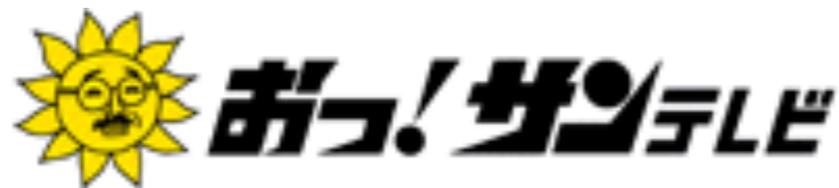
研究目的

1. 放送時間差によりネタバレされる可能性がある視聴者の規模を明らかにする
2. アニメにおける致命的なネタバレを明らかにする
3. 致命的なネタバレの判定がどの程度できるかを明らかにする

1. 放送時間差によりネタバレされる可能性がある視聴者の規模の調査

放送時間差の原因

- アニメ番組は深夜帯に放送されるため、放送局による放送時間のバラつきが多い
- 番組を放送する局が少ない地域が存在する



TOKYO MX
©TM/KK All Rights Reserved.

調査方法

- アニメ20本(2015年放送開始)を無作為に選定
- 各アニメの**番組名**・**放送局**・**放送時間**を取得
- 放送局が放送する地域の**世帯数**を取得

例) **番組名**：俺物語！！

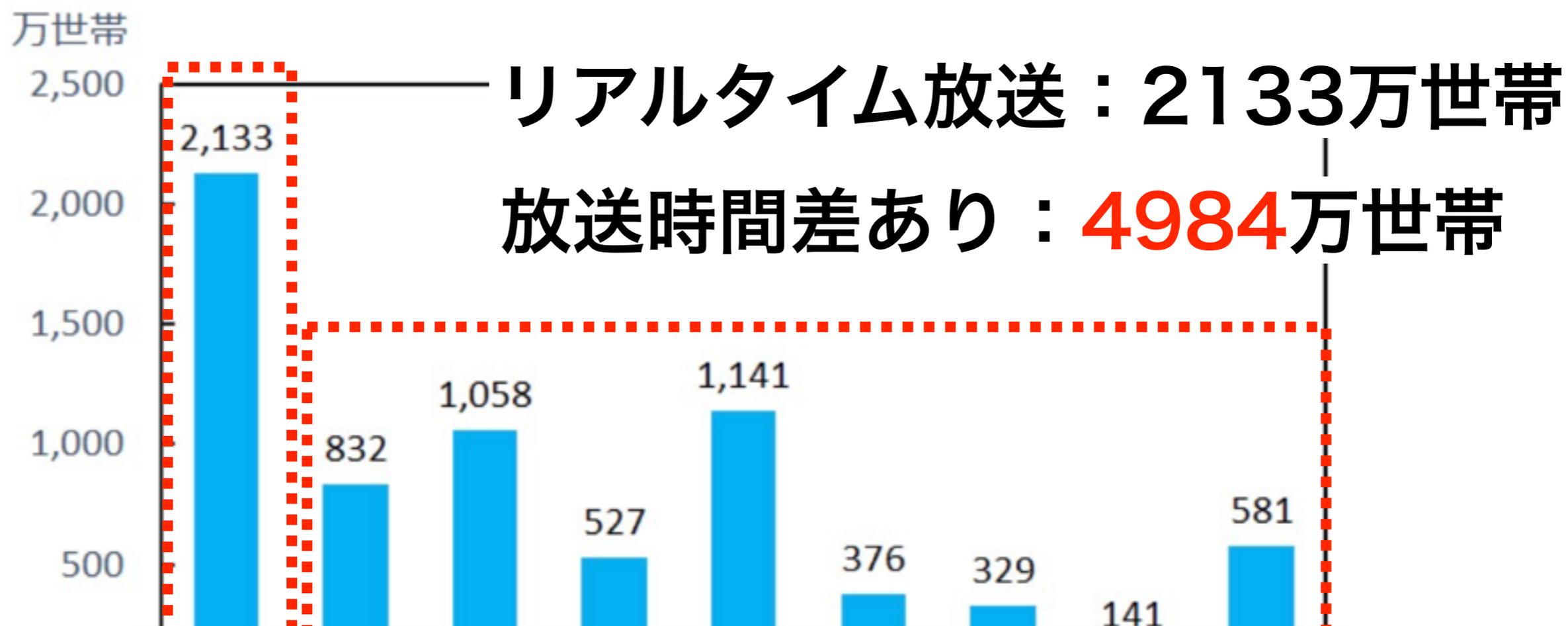
日本テレビ：04/08(水) 25:29～

ミヤギテレビ：04/17(金) 25:57～ など



世帯数：90.2万世帯

調査結果

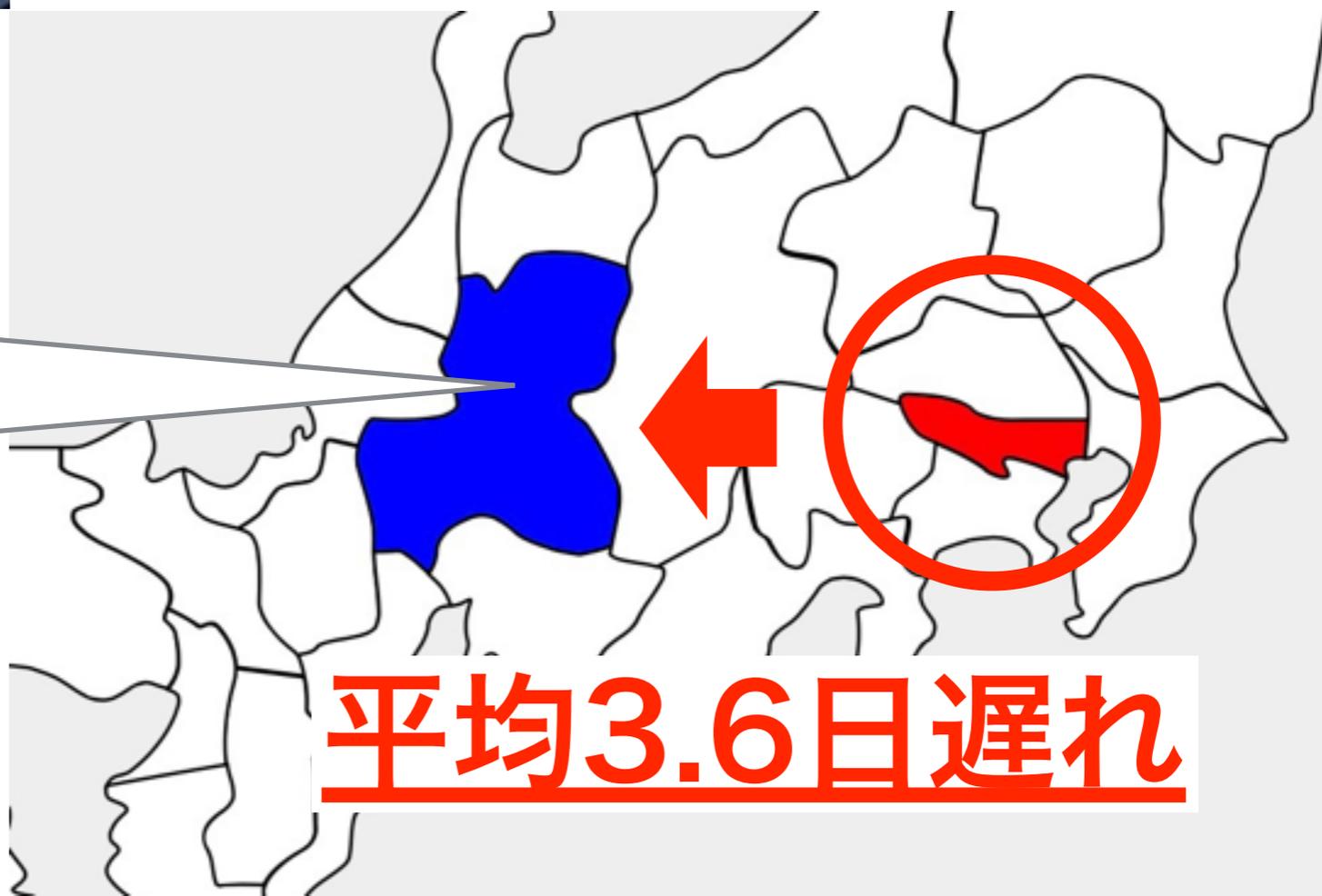


ネタバレをされる可能性のある視聴者は全体の**70%**

岐阜県高山市の場合



リアルタイムで放送しないアニメ
(2015年6、7月放送開始)
→11個のうち**5個**



平均3.6日遅れ

2. ストーリーコンテンツにおける
致命的なネタバレとは？

ネタバレの基礎調査

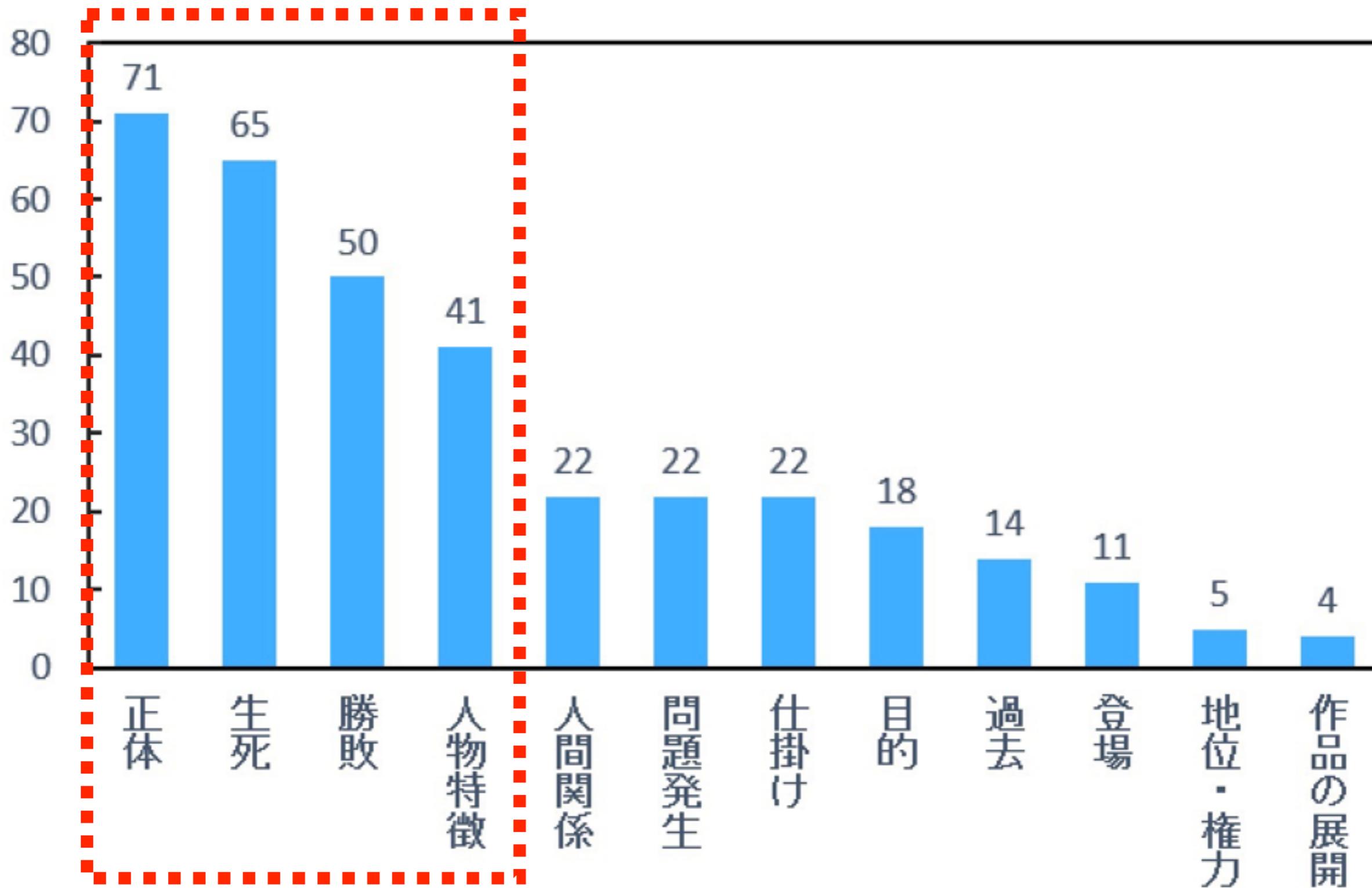
- ネットバシの収集
 - ネットバシ投稿サイト (<http://netabare.meguri.org/>)と学生100名以上に対するアンケート結果を利用
 - 126個のアニメにおけるネタバレを345件収集

収集したネタバレの分析例

- 分析の結果、下の12種類のカテゴリのように分類。それぞれの投稿件数をカウント

フリーザがサイボーグ化して復活
正体/生死/人間関係/人物特徴/仕掛け/目的/
(ドラゴンボールZ)
勝敗/問題発生/過去/登場/地位・権力/作品の展開

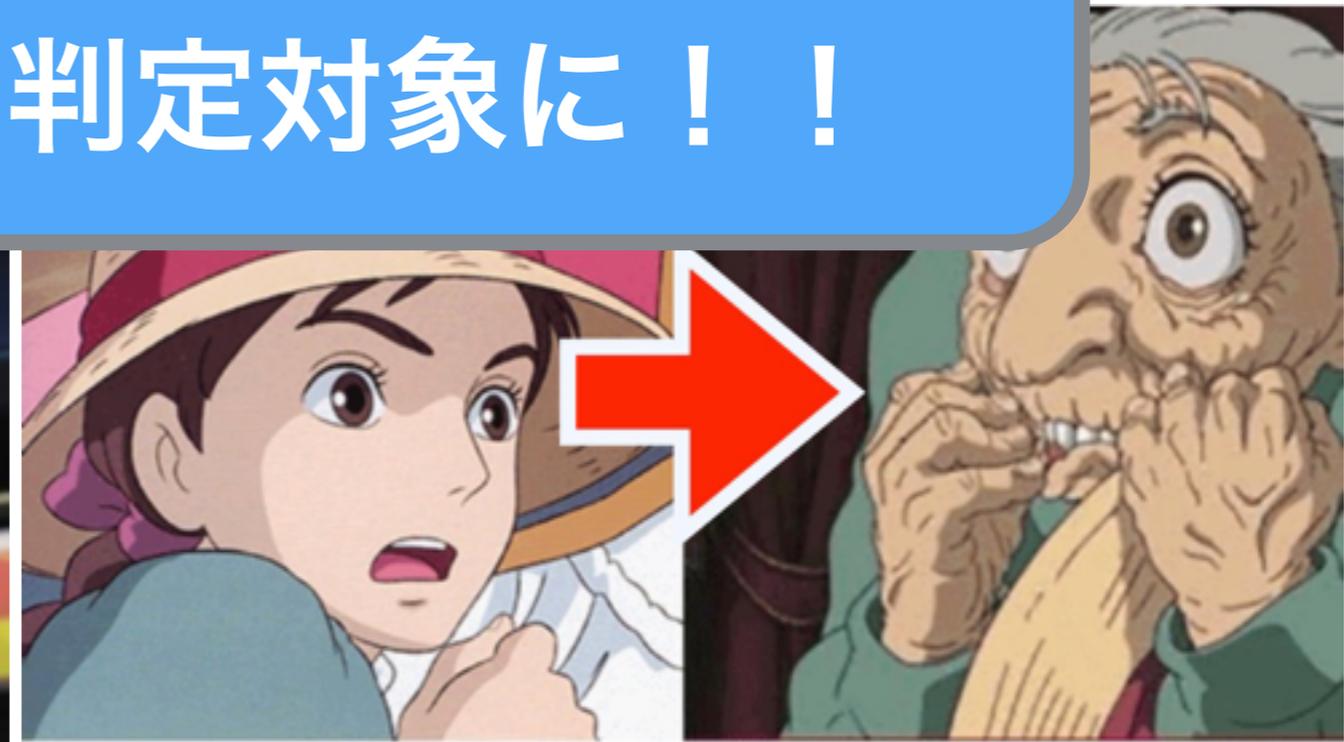
各カテゴリの投稿数



4 カテゴリ



これら4つのカテゴリを致命的な
ネタバレとして判定対象に！！



3. 致命的なネタバシの判定
がどの程度できるのか？

評価用データセット構築

- ・ カテゴリごとに4作品ずつ選定
例) コナン→主人公が犯人を暴く
- ・ アニメに対するハッシュタグの指定し、各作品のツイートを収集
#conan, #jojo_animeなど
- ・ 著者が各カテゴリに該当するネタバレかそうでないかを分類し、投稿にフラグ付け

正体	名探偵コナン 第...
	六花の勇者 第1...
	Charlotte 第9話
	監獄学園 第9話
生死	Fate/stay night
	Charlotte 第7話
	ジョジョの奇妙な...
	戦姫絶唱シンフォ...
勝敗	黒子のバスケ 第...
	Fate/stay night
	Charlotte 第4話
	遊☆戯☆王ARC...
人物特徴	Fate/stay night
	Charlotte 第7話
	ニセコイ: 第12話
	がっこうぐらし!

判定の手順

ネタバレとなる投稿→正例

ネタバレにならない投稿→負例

1. 正例、負例に分類したツイートを
単語ベクトル化
2. 各カテゴリにおけるSVMの分類器を作成し、ネタバレかどうかの判定をする

単語ベクトルの生成

- 収集したツイートを形態素解析(Mecabを使用)
- 形態素に対しIDを割り当て、1行内での出現回数を記録

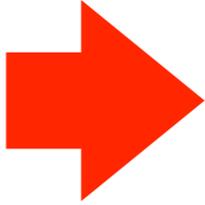
1:犯人	2:は	3:あの	4:医者	...
1	1	1	1	...
1	0	0	1	...
0	0	0	1	...

単語ベクトル生成手法

- 単語ベクトルの生成時に形態素の処理を行うことによって精度の向上を図る3つの手法を提案
 - ベースライン手法
 - 正規化手法
 - 人物名の一般化手法

ベースライン手法

- ・ 品詞は**名詞、動詞、形容詞、副詞、連体詞**の5つを利用
- ・ 活用形のある単語は全て**原形に置換**

例) 死な
死に
死ぬ)  死ぬ

人物名の一般化手法

- ・登場人物名が作品や話数ごとに大きく異なる

- ・**主要人物** 敗れる！！！！(ドラゴンボール)
- ・**主要人物** 敗れる！！！！(ジョジョの奇妙な冒険)

登場人物名を一般的な語に置換

悟空、ベジータ、フリーザ → 「主要人物」

ウーロン、ランチ、ゴズ → 「モブ」

評価実験

- 各カテゴリで選定した4つのアニメの内3つを学習用、残り1つを評価用に
 - 4回検定を行い推定平均を出す
(4-Fold Cross-Validation)

例) 正体に関するネタバレなら



学習用

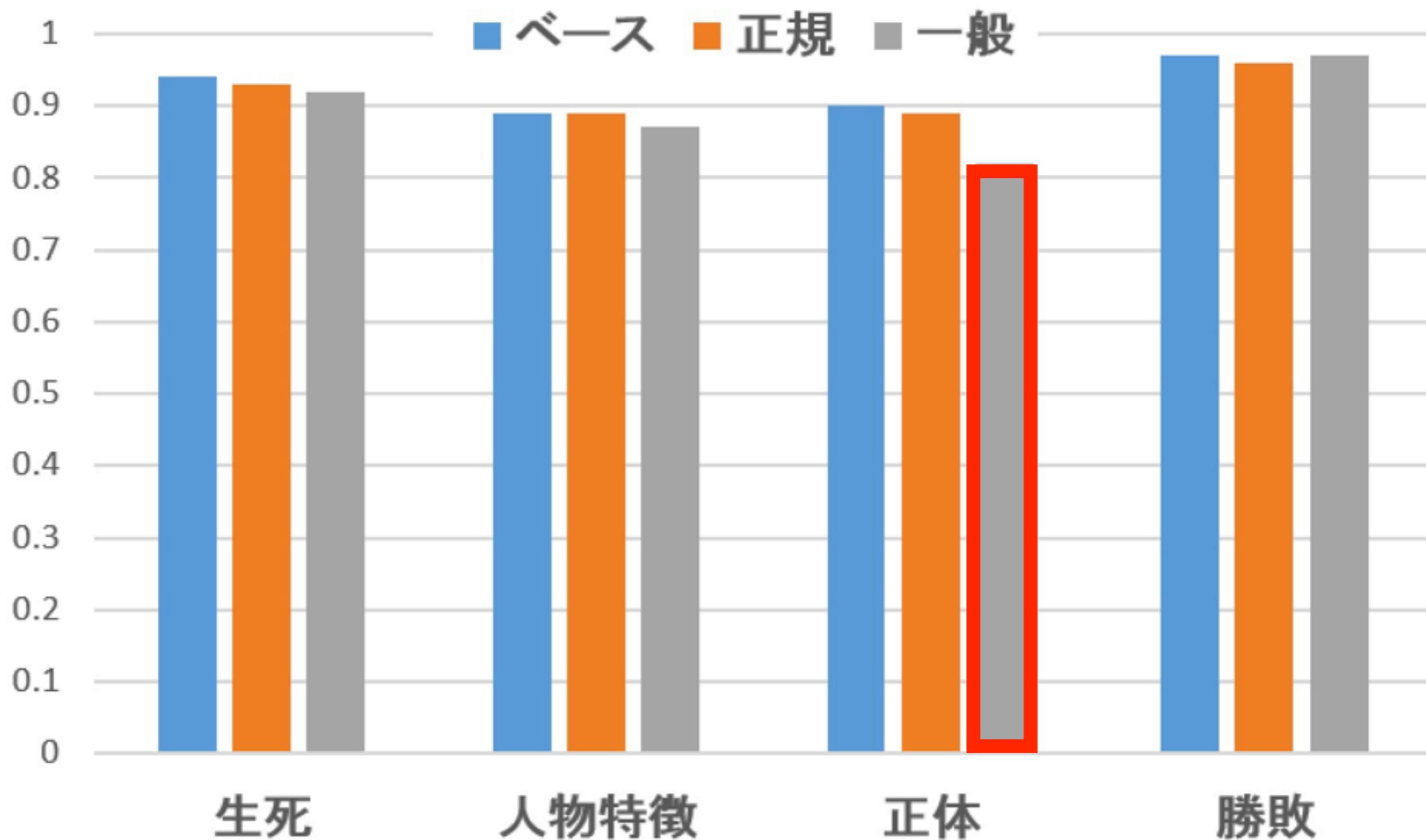
評価用

判定精度の算出

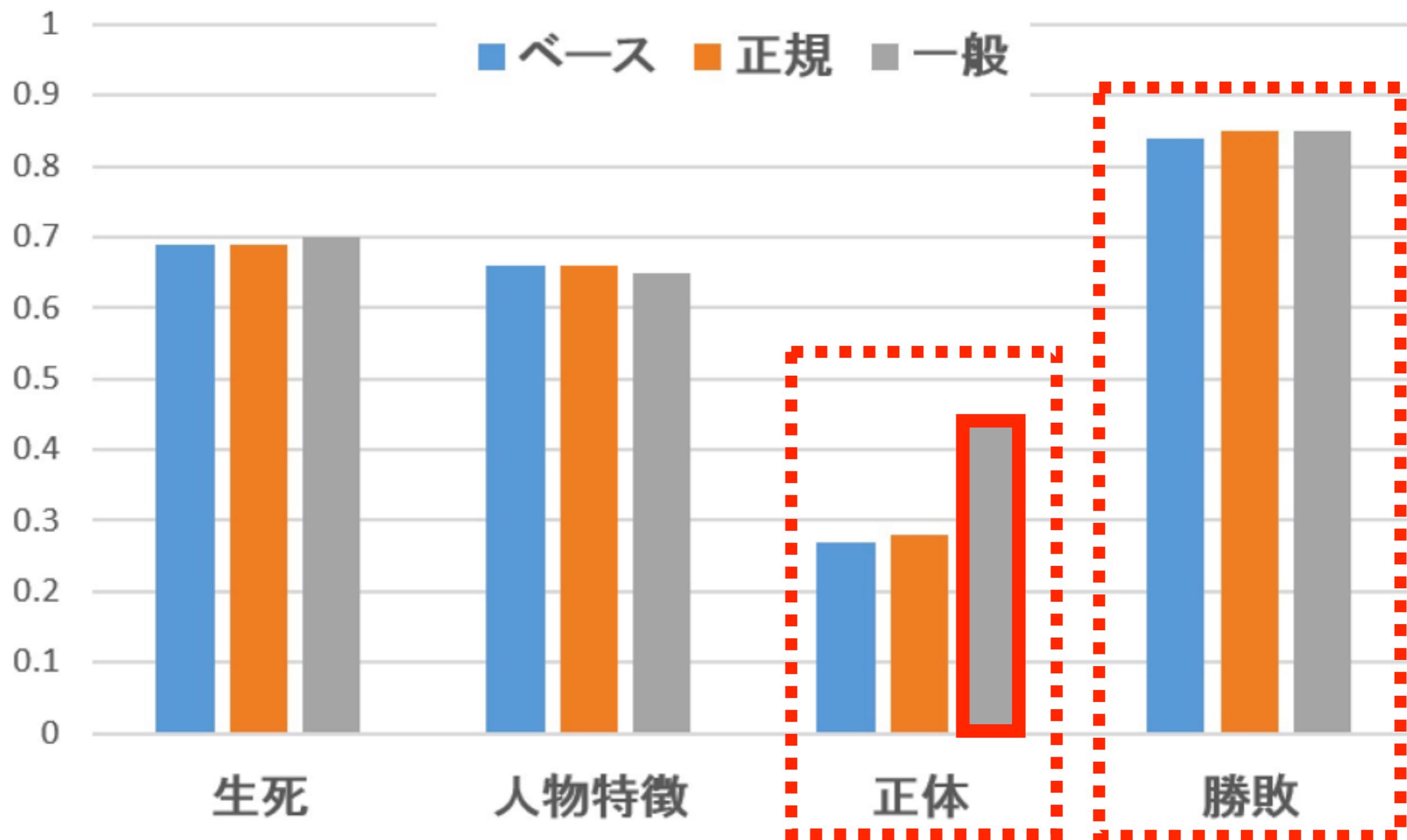
評価尺度は適合率、再現率を利用

		判定	
		正例	負例
答	正例	○	×
	負例	×	○

結果 (適合率)



結果（再現率）



判定しにくい投稿例

- **正体**

スパイ、ストーカー、被害者だった

- **生死**

良い奴だったよ…、タフすぎるw

- **勝敗**

敗因、サヨナラホームラン

- **人物特徴**

パワーアップ、～モード

結果

- 正体に関するネタバレは判定しづらい
 - 特徴語となる人物の正体にあたる単語(特に名詞)が作品ごとに異なることが多い
- 勝敗に関するネタバレは判定しやすい
 - 事実の表現のバリエーションが少ない

考察

- 人物名の一般化手法ではキャラクターの名前以外のだ名も網羅することによって精度の向上が期待できる
- ネタバレとなる投稿に特有な単語が話数ごとに異ならないようにカテゴリ分けを工夫すると高い再現率が期待できる

今後の課題

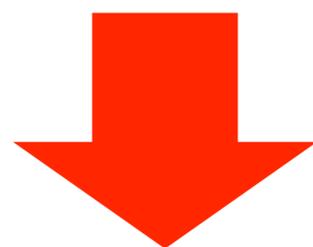
- データセット構築の見直し
 - 著者以外の者との協力
 - データ数を増やす
- 精度の低かったカテゴリ分けの見直し
- 扱うストーリーコンテンツの量を増やす
 - 再現率の向上
- ネタバレ防止のシステムの実装を想定

システム実装の想定

本研究での評価用データセット

ネタバレ(アニメ)となる投稿：ネタバレでない投稿
= 1 : 1

ネタバレとなる投稿：通常の投稿 = ? : ?



実際のTwitter上における割合で判定精度を算出

まとめ

- 放送時間を考慮したネタバレの防止が必要であることが明らかに
- 人々は正体、生死、勝敗、人物特徴に関する情報を致命的なネタバレと見なしている可能性が高いことが明らかに
- 話数が異なっても特徴となる単語が似通うカテゴリは判定しやすく、逆の場合は判定しづらい