

偏りのある特徴語を考慮した皮肉の検出

肥合 智史, 嶋田 和孝 (九州工業大学大学院)

背景

皮肉検出

皮肉：
あえて意図と反対であったり、遠回しな表現を用いて相手を非難すること

皮肉例：
口先だけで何もしない彼は、政治家の鑑だ。

「～の鑑」という肯定表現を用いて、逆に、非難を示す

感情分析における、より高度な文章理解のための挑戦的課題

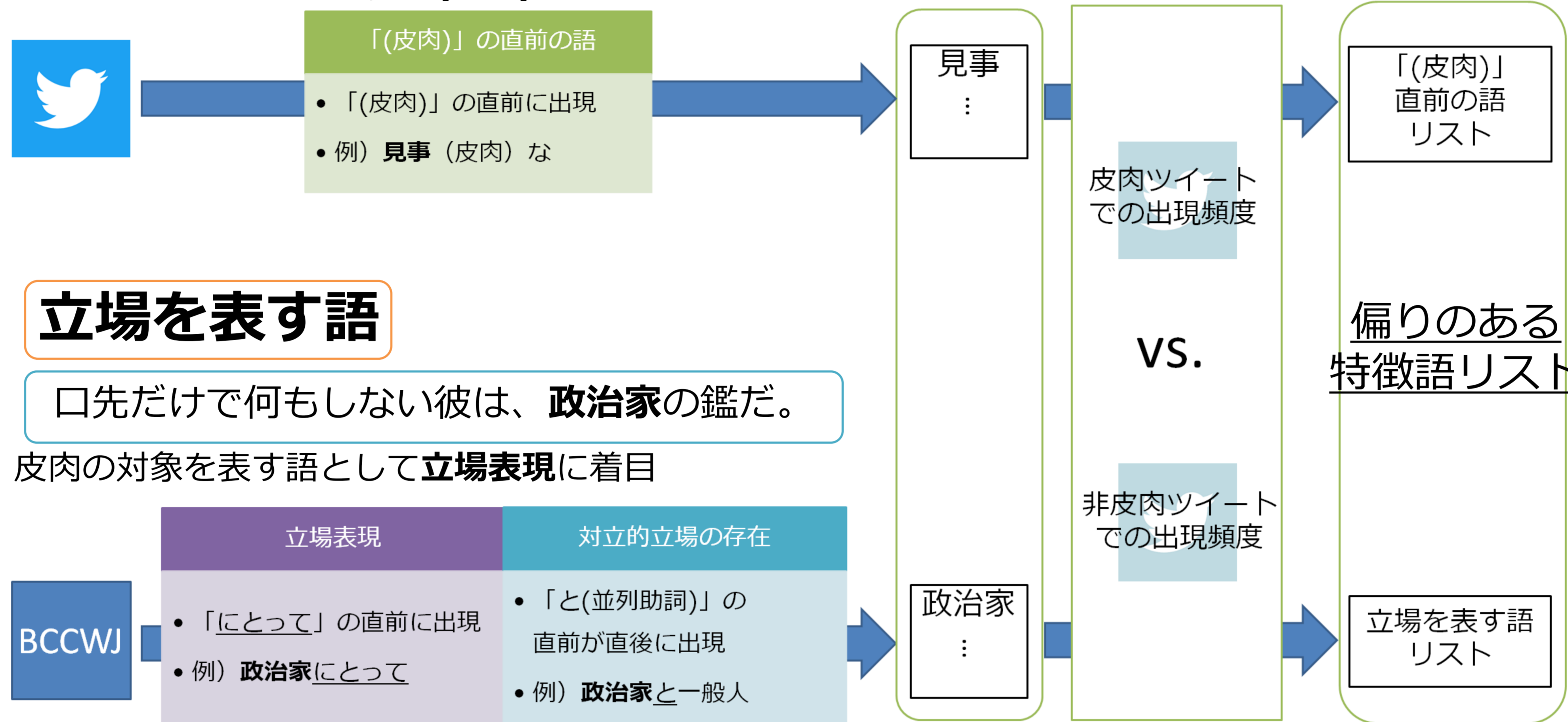
偏りのある特徴語の抽出

二種類の特徴語について、皮肉かどうかの判別に有用な語を絞り込んで抽出

「(皮肉)」直前の語

見事 (皮肉) な部屋の散らかりようですね！

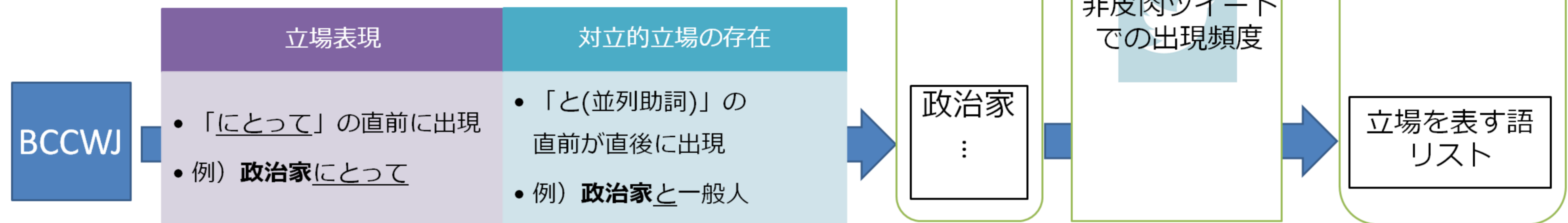
Twitterの投稿における、「(皮肉)」直前の語に着目



立場を表す語

口先だけで何もしない彼は、政治家の鑑だ。

皮肉の対象を表す語として立場表現に着目



BCCWJ：現代日本語書き言葉均衡コーパス

実験・評価

データ

学習データ

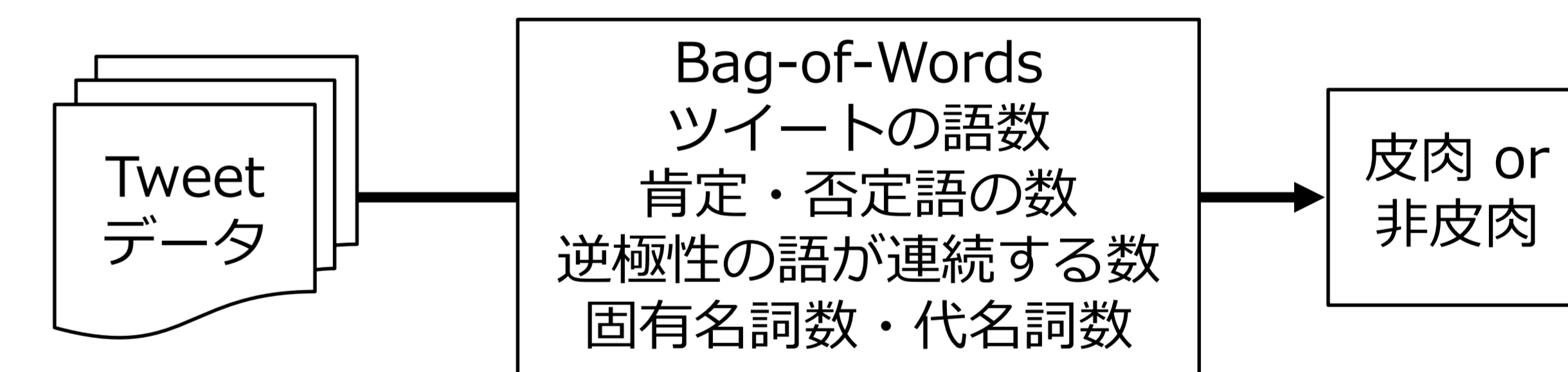
- 正例
 - 「(皮肉)」を含むツイート3120件
 - 「(皮肉)」という表現自体は削除
- 負例
 - 「(皮肉)」を含まない3120件のツイート

評価データ

- 正例
 - 「(皮肉)」という表現を削除した状態で、人手でアノテーション
 - 109件のツイート
- 負例
 - 「(皮肉)」を含まない109件のツイート

ベースライン

先行研究の素性を採用した合議なしの分類



先行研究

機械学習による二値分類

データ

Twitterを利用
ハッシュタグ「#sarcasm」を用いて、皮肉を含む投稿を収集

素性

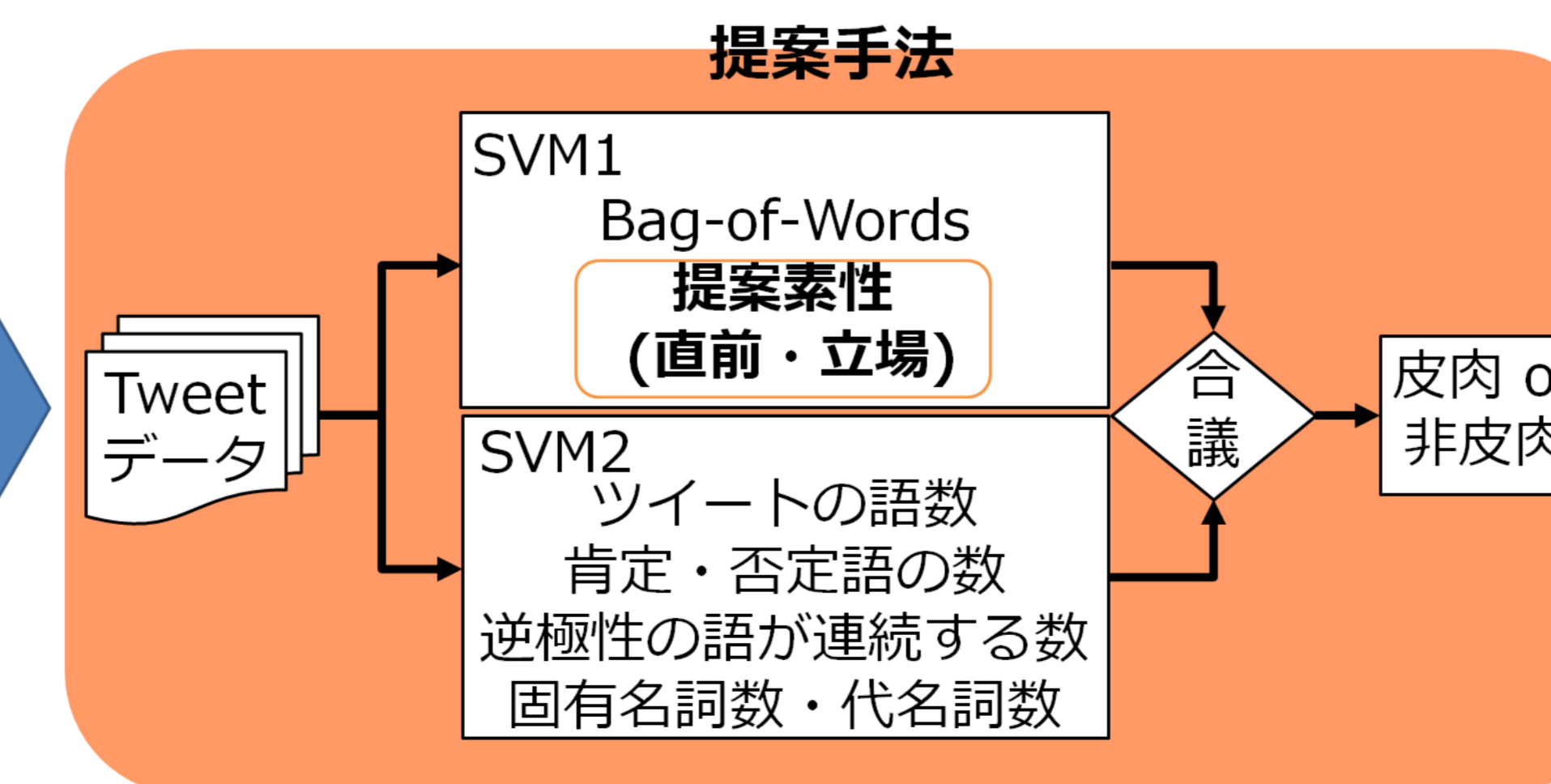
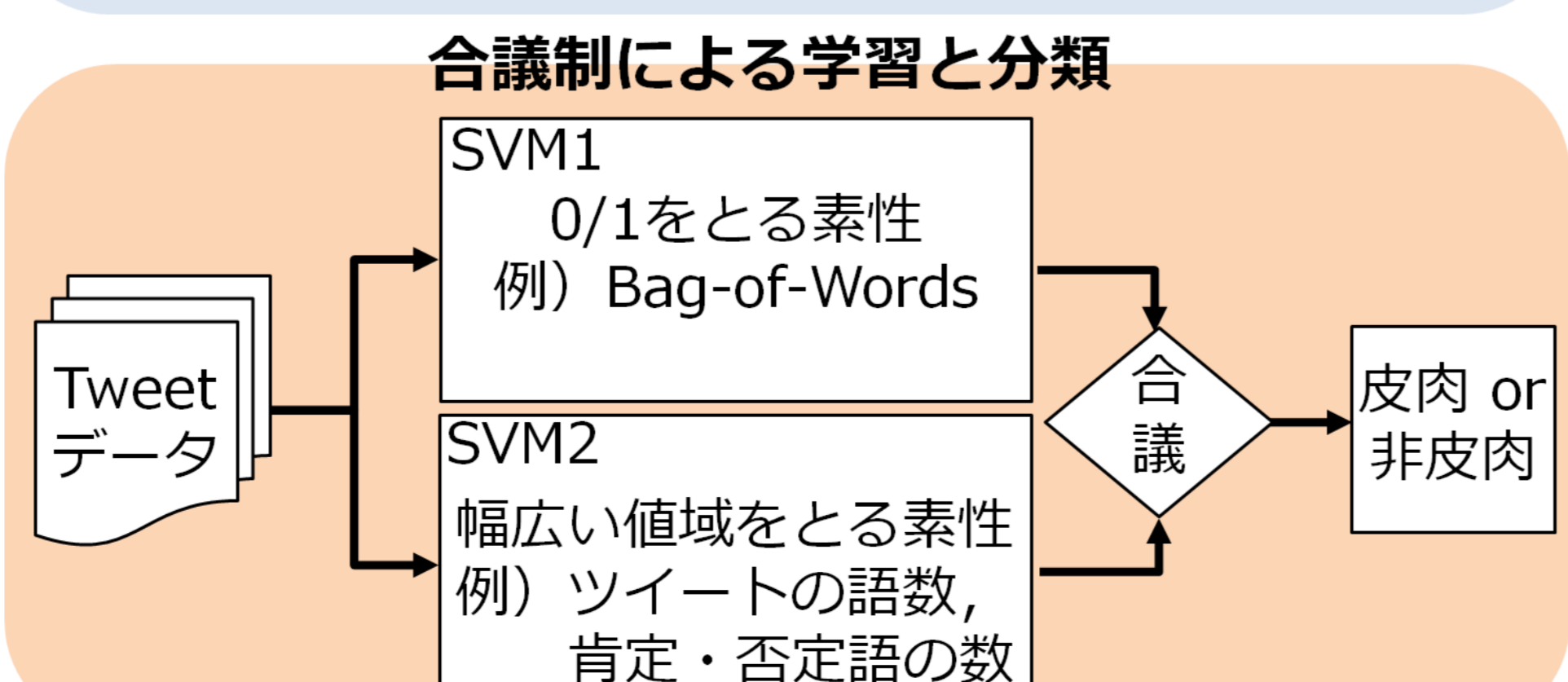
無視するなんて、君はいい性格してる。

文書中の極性語	皮肉の対象の存在
<ul style="list-style-type: none"> 肯定・否定語 異なる極性語の連続 (Joshiら 2015) 	<ul style="list-style-type: none"> 固有名詞の有無 代名詞の有無 (Karouiら 2015)

提案手法

素性に「(皮肉)」直前の語や立場を表す語を利用

- 先行研究：素性
- Bag-of-Words
 - ツイートの語数
 - 肯定・否定語の数
 - 逆極性の語が連続する数
 - 固有名詞数
 - 代名詞数
- 提案：
- 「(皮肉)」直前の語それぞれの有無(0/1)
 - 立場を表す語それぞれの有無(0/1)



結果・まとめ

手法	適合率	再現率	F値
ベースライン	0.74	0.74	0.74
合議	0.82	0.81	0.81
合議+直前	0.86	0.85	0.85
合議+立場	0.85	0.84	0.84
合議+直前+立場	0.85	0.84	0.84

ベースラインと有意差あり

ベースライン・合議と有意差あり

「(皮肉)」直前の語と立場を表す語は皮肉の検出に有効