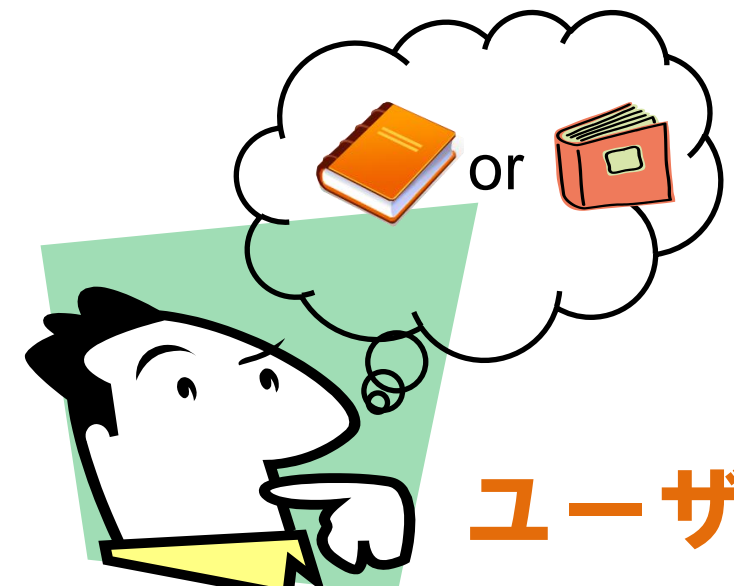


レビュー文書からのクレーム文発見システムの開発

梅沢佑介, 渡辺尚吾(筑波大学), 板谷悠人(筑波大学大学院)

研究背景

Web上のレビュー文書

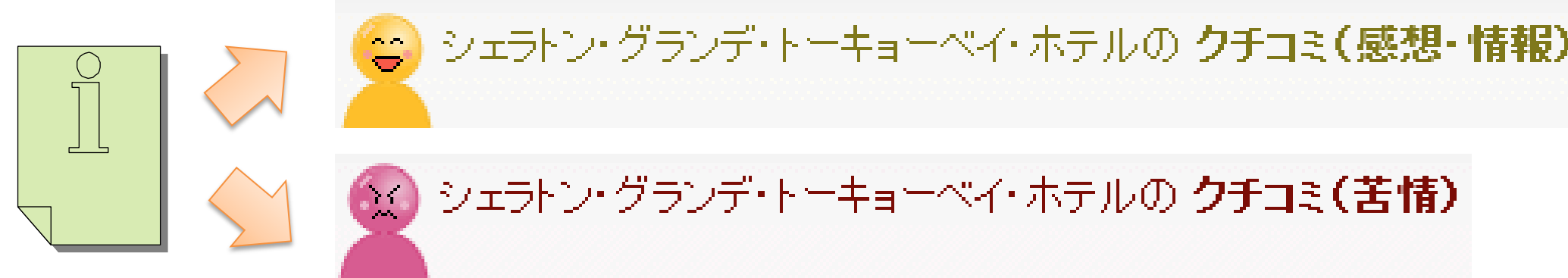


商品購入、サービス利用の際の意思決定支援

マーケティングの材料、商品・サービス改善のための情報として

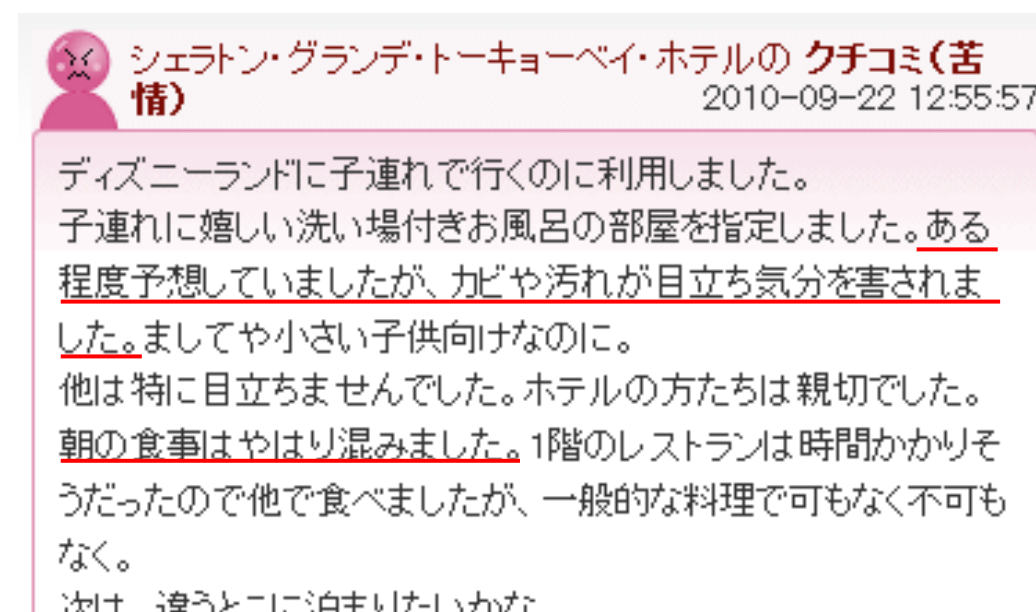
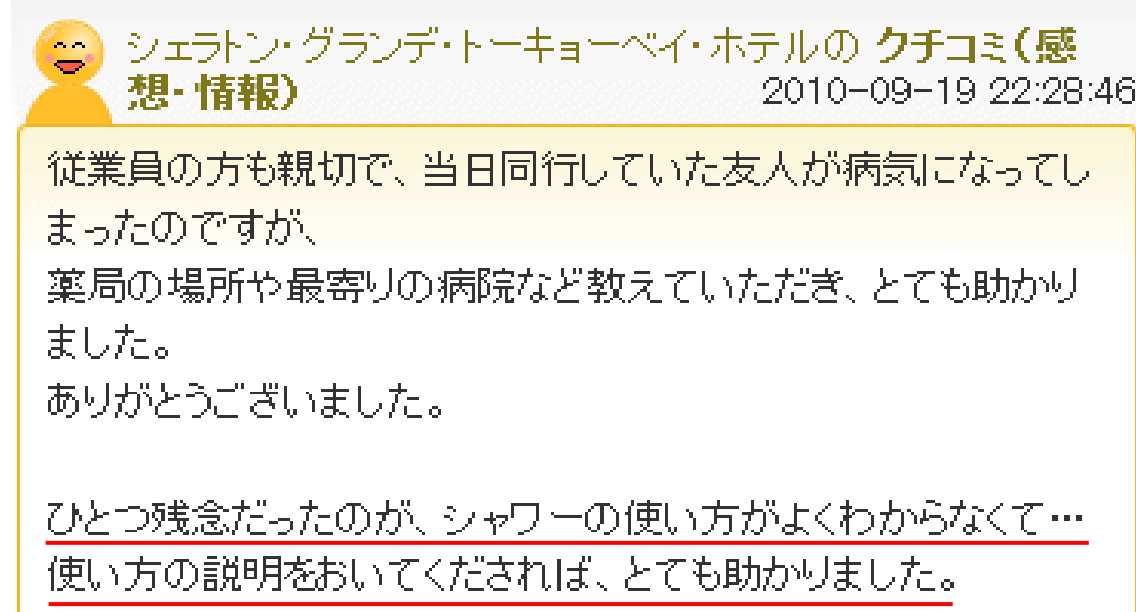
レビューをさらに利用しやすい形に

レビューを文書ごとにラベル付け



研究目的

文レベルでのクレーム発見



(感想・情報) であってもクレーム文が存在

(苦情) であってもクレーム箇所は一部

クレームを文単位で扱う長所

負担軽減



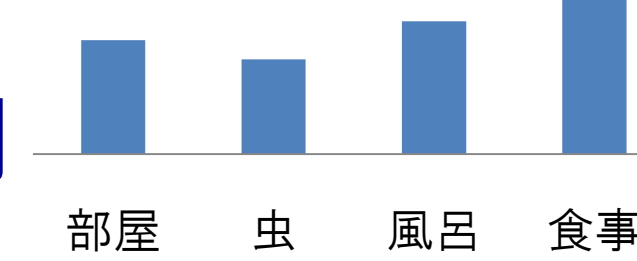
大量の文書をくまなく読み、クレーム箇所を探し出す負担を軽減

クレーム補完



非クレーム文書からクレーム文を発見・補完

二次利用



統計・定量的分析が行いやすく書き込み頻度の高いもの → 優先的に対応など

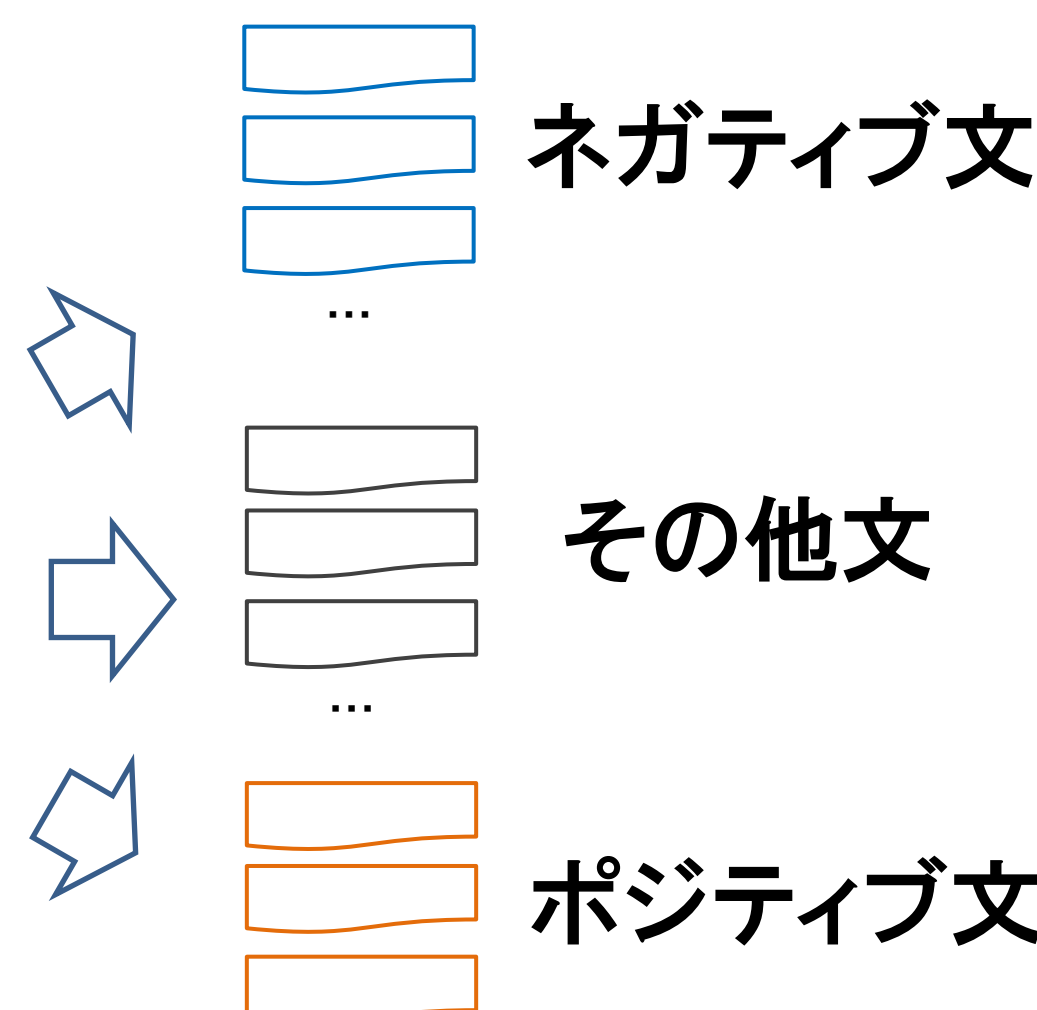
クレーム文発見の流れ

教師データの準備



文単位に分割

評価表現辞書を用いて分類



ナイーブベイズで学習・分類

ネガティブ文集合、非ネガティブ集合(ポジティブ+その他)を教師データとして学習、レビューをクレーム文/非クレーム文に分類

さらに評価表現辞書でも、テストデータを分類しておく(次に利用)

文の並びを利用したフィルタリング

ポジ文、ネガ文同士は隣接しやすいと仮定し、ラベルの付け替えを行う

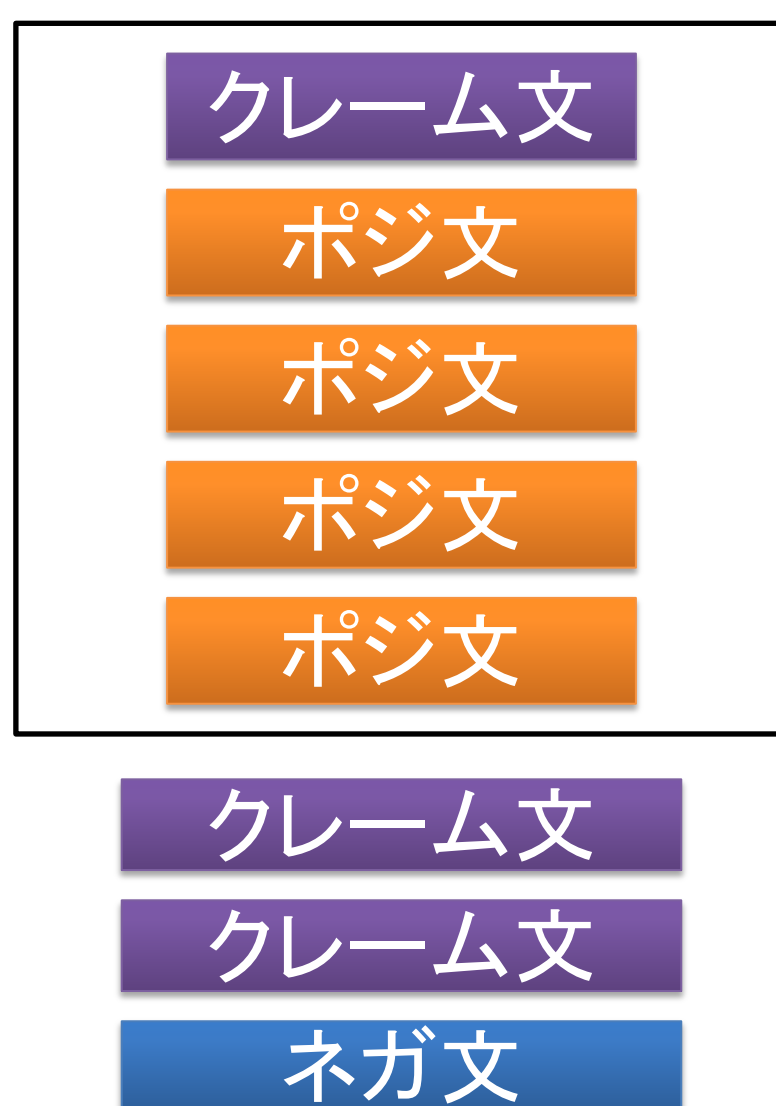
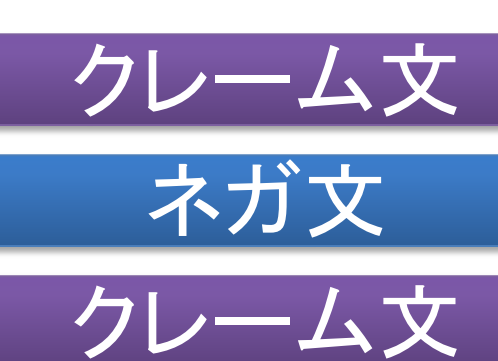
フィルタリングA

右図のような場合、クレーム文を非クレーム文にラベル付け替え



フィルタリングB

右図のような場合、ネガ文をクレーム文にラベル付け換え



評判分析の基本技術

評価表現辞書

評価表現

ポジティブ/ネガティブな意味合いを持つ表現

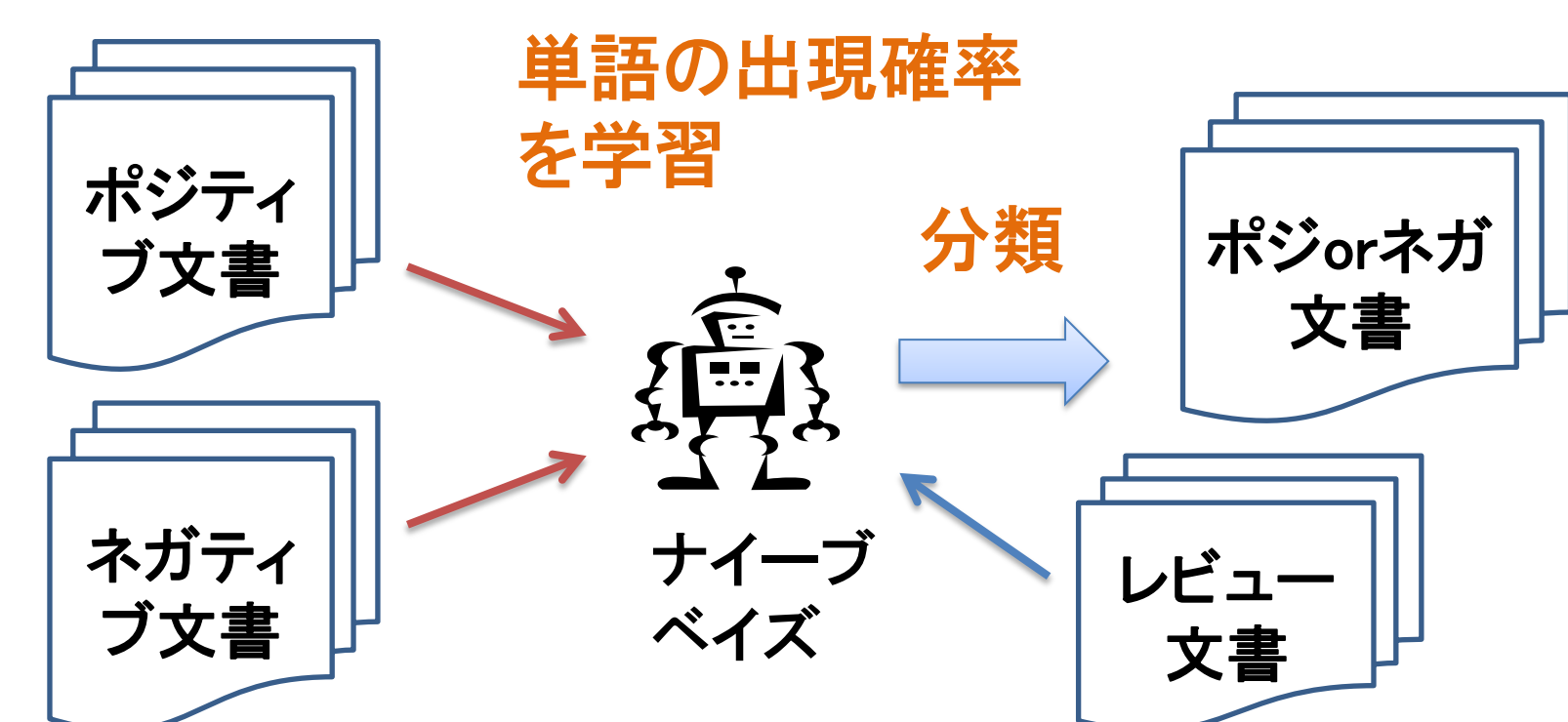
(例)“このリンゴは不味い”

ネガティブ文 評価表現の出現比率により文のポジ/ネガを判定

表現	品詞	P/N
優れる	動詞	Positive
めでたい	形容詞	Positive
不満	名詞	Negative
不味い	形容詞	Negative

Pos : 760words
Neg : 862words

ナイーブベイズ分類器



$$P(c | d) = P(c)P(d | c)$$

$$= P_c P \left(\prod_w n_{w,d} \right) \frac{\left(\sum_w n_{w,d} \right)!}{\prod_w n_{w,d}!} \prod_{w \in V} q_{w,c}^{n_{w,d}}$$

n: 単語の出現回数

q: 単語の出現確率 (多項モデル)

評価実験

評価データ

楽天トラベル

シェラトン・グランデ・トーキョーベイホテルのレビュー1060件(5493文)

苦情文 743文 非クレーム文 4750文(人手でラベル付け)

実験結果

手法	適合率	再現率	F値
評価表現辞書	75%(224/300)	30%(224/743)	0.430
ナイーブベイズ(教師データに感想・情報記事と苦情記事を使用)	54%(296/552)	40%(296/743)	0.459
評価表現辞書+NB	58%(503/862)	68%(503/743)	0.627
評価表現辞書+NB+フィルタリングA	63%(495/786)	67%(495/743)	0.648
評価表現辞書+NB+フィルタリングA,B	60%(528/877)	71%(528/743)	0.652

課題・今後

事実文、要求文の発見

評価表現を含まない文(事実文、要求文等)

今回の手法では発見しきれず。

事実文の例: “室内設備にヘアキャップが無かった。”

要求文の例: “スタッフの数を増やすなどの対応をして欲しかった。”

宿泊施設応答の利用

宿泊施設応答から謝罪文を検出、謝罪文とペアとなるユーザーレビュー中の文を取り出す

教師データとして利用、など

