

経験に基づいた有用で希少な レビュー情報の抽出手法の提案

甲南大学 知能情報学部 知能情報学科
灘本研究室 11371025 織田 翔真

背景

現在インターネット上には多数のショッピングサイトがある。



amazon



楽^R天
I C H I B A

背景

ユーザが商品を購入する
判断材料の一つとしてレビューがある



背景

レビューではいろいろなユーザが
商品について書いている

書かれている量が少ない経験に基づいた希少な
レビューは見つけにくい

お掃除ロボットのレビュー

★★★★☆ 毎日頑張ってくれています

投稿者 らんまま 2013年8月10日

Amazonで購入

とても可愛くて、つつい目で追ってしまいます。我が家はペットの毛があるので、この大きさのダストボックスではすぐにいっぱいになってしまいます。段差ではセンサーで回避とありますが、前ではなく、後ろに下がった時落ちそうになってます。でも、この価格なので許しちゃえます。可愛いし、色もいいです



経験に基づく

背景

レビューの件数が多いとユーザの負担
になってしまう
→情報を見落としてしまう



CCP【LAQULITO】自動ロボット
掃除機（エントリーモデル）レ
ッドブラック CZ-860-RB
CCP

¥ 12,500 **プライム**

ポイント: 125 pt (1%)
残り2点。注文はお早めに。

こちらからご購入いただけます

¥ 12,500 新品 (6 出品)

¥ 4,300 中古品 (16 出品)

ホーム&キッチン: 全1,839点を見る

新しいバージョンを表示

★★★★☆

281

281件

目的

他のレビューに埋もれてしまった見つけにくい
経験に基づく希少で有用なレビューを抽出する



定義

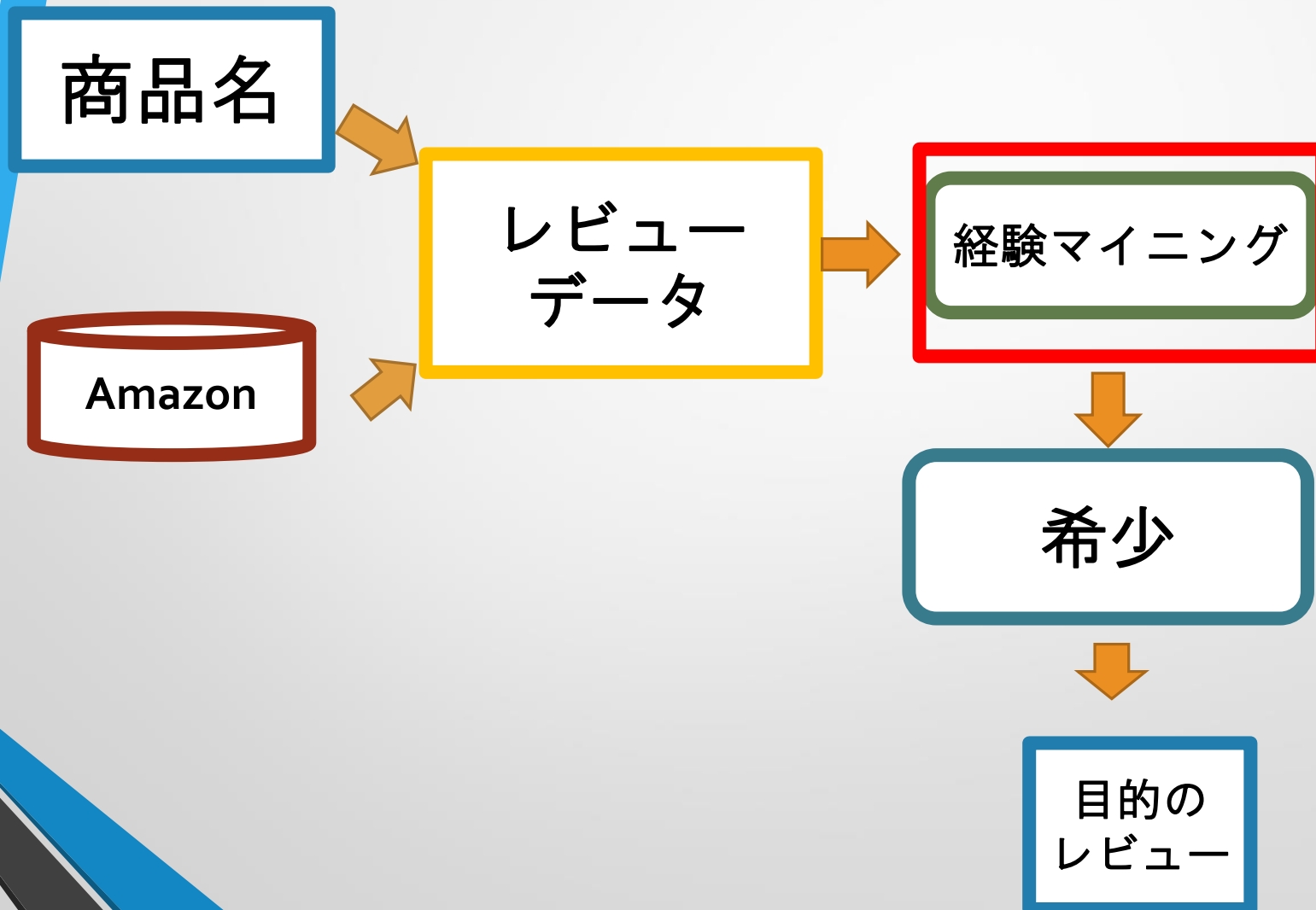
経験

ユーザが実際に使用したことから得られた知識、技能

希少

見落としてしまう可能性のある少数意見

提案手法：全体の流れ



経験マイニング

- (1) a. トピック：どの利用物，サービスに関する経験か
b. 経験主：経験の主体
c. **事態タイプ**：経験情報の核となる事態の種類（e.g. ポジティブ/ネガティブな出来事・状態・性質，入手・利用等の行為）
d. **事実性**：(c) の事態の事実性に関する情報（e.g. テンス，アスペクト，極性，モダリティ情報）
e. 事態表現：経験の核となる事態の表現（典型的には述語項構造）

抽出

本研究ではこの2つを抽出して経験マイニングとします。

乾健太郎，原一夫，“経験マイニング：Web テキストからの個人の経験の抽出と分類”，言語処理学会第14回年次大会論文集，pp1077-1080，2008.3

経験マイニング

事態タイプ

出来事

行為

感情

日本語評価
極性辞書

レビュー文に
出現しているか

ポジネガ値が
閾値以上か

小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203-222, 2005.

ポジネガ値

$$\text{ポジネガ値} = \frac{|P| + |N|}{|A|}$$

- |P| : レビュー文章中の**ポジティブ**な単語の出現回数
- |N| : レビュー文章中の**ネガティブ**な単語の出現回数
- |A| : レビュー文章に出現する単語の総数

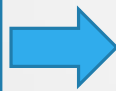
経験マイニング

- (1) a. **トピック**：どの利用物，サービスに関する経験か
- b. **経験主**：経験の主体
- c. **事態タイプ**：経験情報の核となる事態の種類（e.g. ポジティブ/ネガティブな出来事・状態・性質，入手・利用等の行為）
- d. **事実性**：(c) の事態の事実性に関する情報（e.g. テンス，アスペクト，極性，モダリティ情報）
- e. **事態表現**：経験の核となる事態の表現（典型的には述語項構造）

経験マイニング

事実性

話者態度
(モダリティ)



機能表現辞書
「つつじ」



モダリティ情報を用いた条件式を満たす。
↓
除外

経験マイニング

モダリティ情報とは . . .

例：「きっと雨が降るだろう」

「雨が降る」ということに対する筆者の**推測**

モダリティ情報

経験マイニング モダリティ

経験ではないモダリティ
「推量」・「願望」・「疑問」
の単語の出現回数

$$|G| - |E| \geq |S|$$

経験であるモダリティ
「過去」・「逆説」の
単語の出現回数

「。」「.」改行で
分けられた
文の個数

経験マイニング

レビュー
データ

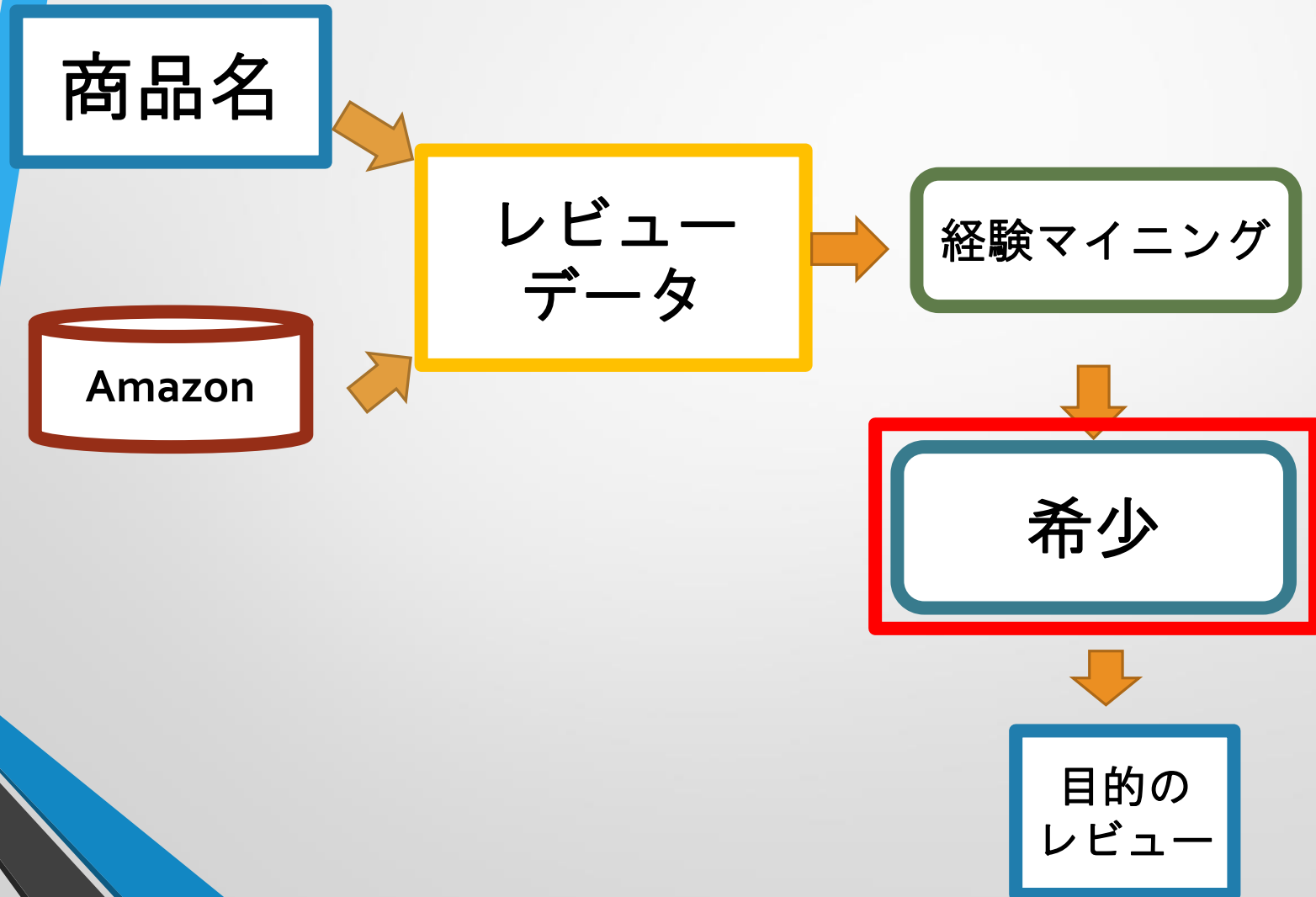
事態タイプ

事実性

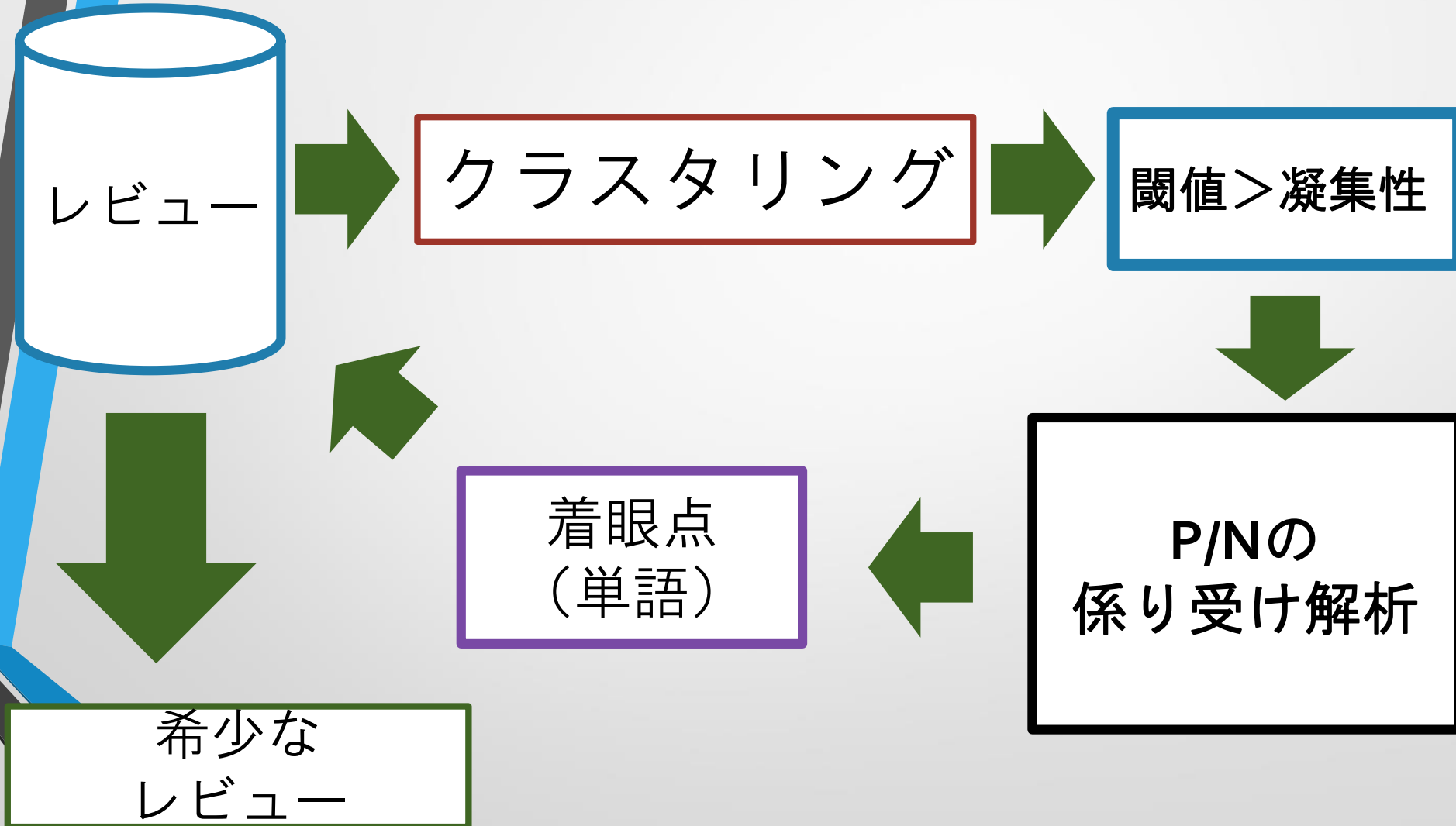
経験マイニング



提案手法：全体の流れ



希少



凝集性

$$A_i = \sum_{x \in C_i} \left(\frac{x \cdot c_i}{|x| |c_i|} \right)^2$$

A_i : 凝集性

c_i : クラスタのセントロイド

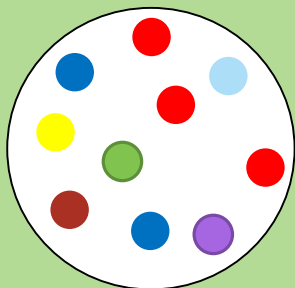
x : クラスタに含まれるレビュー文

弱い

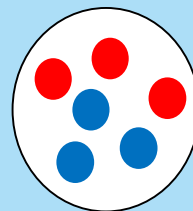
強い

凝集性が弱いクラスタ

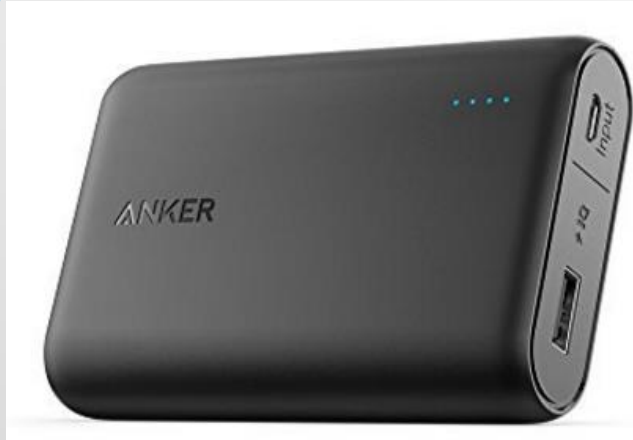
レビュー文同士の
関係性が薄い為
少数意見が集まっている



レビュー文同士の
関係性が濃い為
頻出したレビュー文が
集まっている



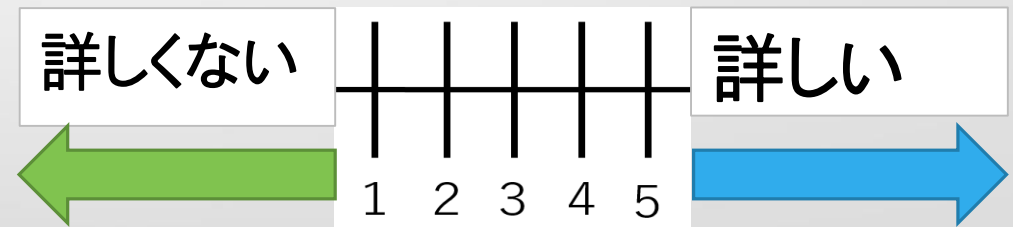
評価実験



210件のモバイルバッテリーの商品説明と商品レビューを10人に読んでもらい、商品レビューが経験であるかを判断してもらい

10人の結果の平均値と本手法の経験マイニングの結果から適合率と再現率を算出する。

経験談を含まない：○
経験談を含む：1～5



実験結果

アンケートから得た結果の平均値が閾値
 ” 2”、” 2.5”、” 3” 以上
 の場合を正解データとした場合の各閾値の再現率と適合率

平均値	2	2.5	3
再現率	0.963855	0.746988	0.626506
適合率	0.358744	0.278027	0.233184

商品に対してではない
 ポジティブ、ネガティブの単語が原因

まとめ

事態タイプと事実性から経験マイニング
クラスタリングした結果から希少なレビューを
抽出する手法を提案した

今後の課題

デマ情報に対する対策
経験マイニングのポジティブ・ネガティブ
の抽出手法の改善

まとめ

事態タイプと事実性から経験マイニング
クラスタリングした結果から希少なレビューを
抽出する手法を提案した

今後の課題

デマ情報に対する対策
経験マイニングのポジティブ・ネガティブ
の抽出手法の改善