

レビューを対象とした信頼性判断支援システムの提案

伊木 惇(中国電力)
亀井清華(広島大学)
藤田 聡(広島大学)

背景(1)

- Amazonや楽天市場などのレビュー
 - Web上の商品購入者の60%
 - ◆評価値やレビューを含んだサイトから購入
 - 購入者の70%
 - ◆レビューを読んだから購入
- レビューは読者の購買意欲を左右する



背景(2)

- ステルスマーケティング(サクラ・スパムの存在)

■サクラ

- ◆商品の評判を意図的に上げる(下げる)ために不当な偽のレビューの投稿を行う悪意を持った投稿者

■スパム

- ◆サクラによって投稿されたレビュー
- ◆レビューを読んだユーザを騙し、誤解させる

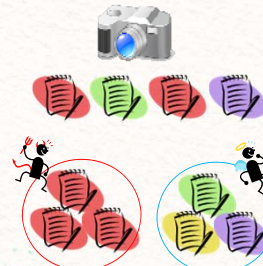


- 各レビューがスパムであるか否かを判断する必要
- ◆スパムらしい=信頼性が低い

関連研究～スパム検知(1)

- Jindaiらの研究

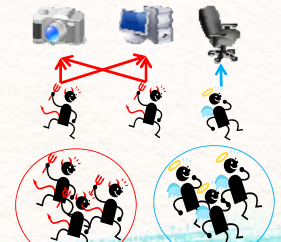
- 複製されたレビューはスパムらしい



[8] Jindai, et al. "Opinion Spam and Analysis", 2008.

- Mukherjeeらの研究

- 特定の商品に、協力してレビューを投稿しているユーザはサクラらしい



[7] Mukherjee, et al. "Spotting Fake Reviewer Groups in Consumer Reviews", 2012.

関連研究～スパム検知(2)

- スパムであるか否かを二値分類する
 - スパムの一面的な特徴を捉えたもの
 - ◆全てのスパムの検知には至っていない
 - スパムとスパム検知はいたちごっこ
 - ◆検知だけでは不十分
 - 精度は人手で判断したものとの比較
 - ◆人手での判断は難しい
 - 信頼性を判断するために役立つ情報が乏しい
 - 必要な情報を個人が十分に収集するのは困難

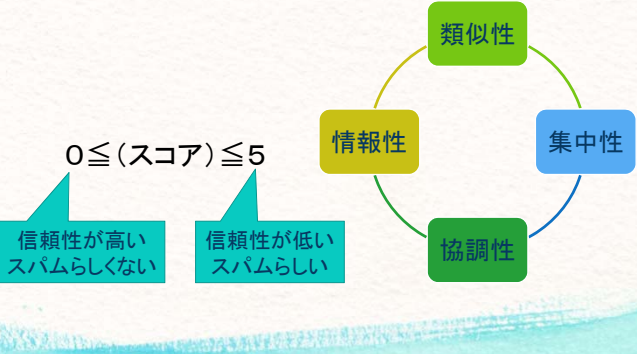
研究目的

レビューを対象とした信頼性判断支援

- 「スパムらしさ」を表す指標をユーザに提示
 - ユーザ自身が信頼性を意識することを促す
 - ユーザ自身が信頼性を判断できるように支援する

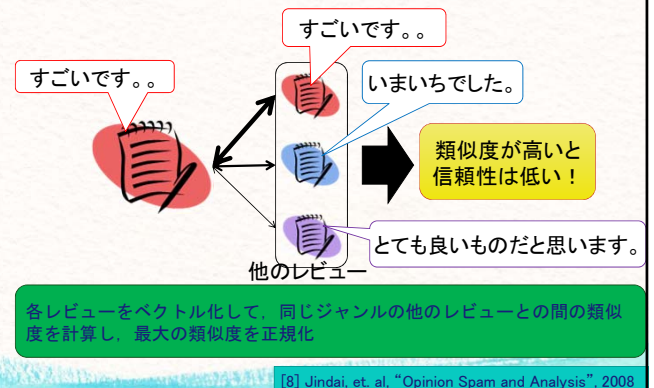
信頼性指標

●各レビューに各指標に関するスコアを提示



類似性

●どの程度、他のレビューの文章と類似しているか



協調性

●どの程度、サクラグループに投稿された可能性があるか

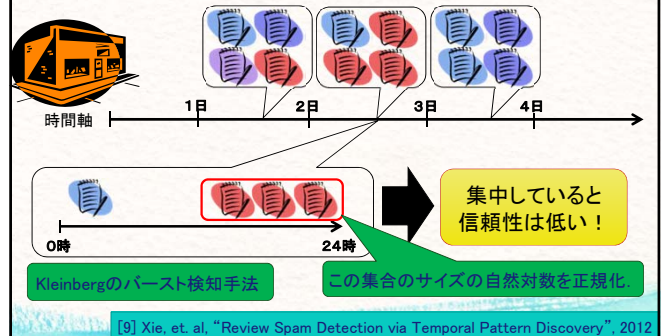
■サクラグループ: 同じグループのメンバーが同じ商品に対して投稿を行うことで、協力して評判を変える



集中度

●どの程度、時間的に集中して投稿されたか

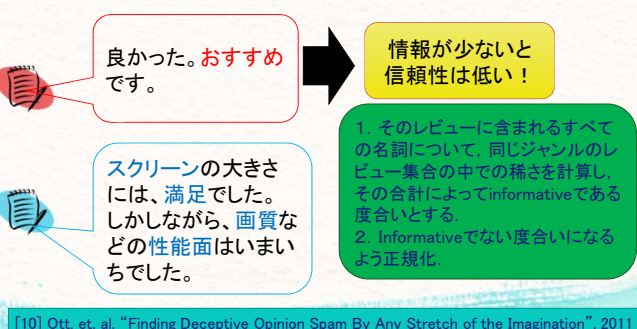
■普段より短時間に多く投稿がされているときのものを検出



情報性

●どの程度、informativeでない文章か

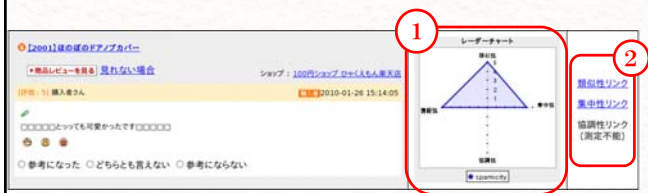
■Informativeなレビュー: 他のレビューであまり使われていない名詞を多く含んだレビュー



提案システム(1)



提案システム(2)



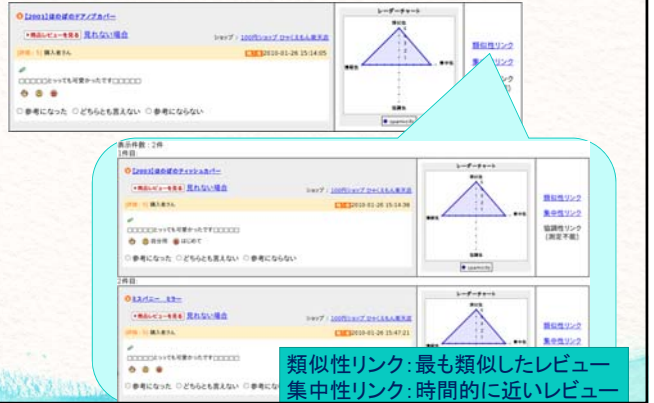
①指標の視覚化

- ・信頼性意識の向上
- ・直観的な判断

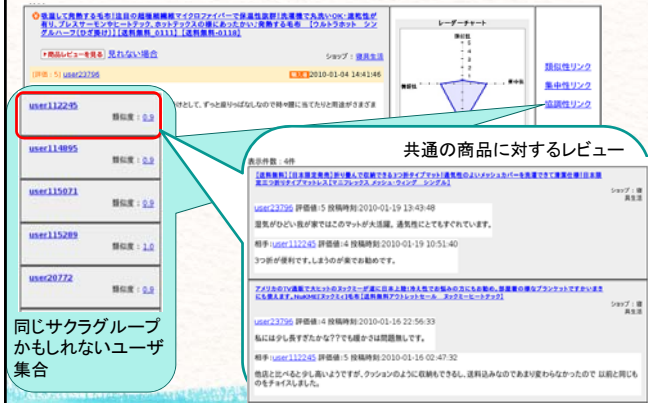
②情報にアクセスするためのリンク

- ・判断材料となる情報を提示

類似性リンク及び集中性リンク



協調性リンク



データセット

- 楽天市場の「みんなのレビュー・ロコミ」データ
- 期間: 2010年1月1日～1月31日の1カ月間
- ジャンルの内訳
 - 本(44,668件)
 - 家電(57,469件)
 - 家具(82,064件)
 - 合計(184,201件)

予備実験

- 各指標に基づいてスパムレビューを抽出する

指標	スコア5の件数	スコア4以上の件数
類似性	9265	926
協調性	67	381
集中性	59	154
情報性	813	1117
類似性+協調性	3	270
類似性+集中性	7	116
類似性+情報性	131	145
協調性+集中性	0	0
協調性+情報性	0	3
集中性+情報性	0	5
類似性+集中性+情報性	0	3

本: 5438
家電: 1497
家具: 2330

集中性は普段から高い(低い)評価ばかり投稿されていると高くなりにくい

事前アンケート

	普段からレビューがスパムか意識してますか?
はい	9(43%)
いいえ	12(57%)

- 被験者21人にレビューを読む際に見る情報を選択してもらった
 - 普段からスパムを意識する人であっても、以下の情報はほとんど見ないことが確認された
 - ◆ レビューの投稿時間
 - ◆ 投稿者履歴
 - ◆ 他の商品のレビュー

実験(アンケート)

- 提案システムを利用しない場合と利用する場合との比較
 - 被験者21人にレビューを1件ずつ、計30件を提示
 - ◆投稿者履歴、店情報、商品情報、他のレビュー等へのアクセス可
 - 「信頼できる」「どちらとも言えない」「信頼できない」の判断を集計
 - 判断の参考にした情報を集計
- システムを利用しない場合の実験から1週間経過後、システムを利用する場合の実験
 - ◆同じレビュー30件を順番を入れ替えて提示

判断結果

- 被験者の信頼できるかどうかの内訳
 - 21人×30件=630回答

	信頼できる	どちらとも言えない	信頼できない
利用しない	49%	22%	29%
利用する	39%	14%	47%

- 被験者間の判断の一致度(21人)

	一致度(Fleissのκ係数)
利用しない	0.22

■値が1で信頼性の判断が容易になっている

参考にした情報

- 参考にした情報(1つのレビューあたりの平均参考人数)

	レビュー本文	投稿者履歴	他の投稿者のレビュー	ショップ履歴
利用しない	18.1人	8.8人	7.6人	1.2人
利用する	16.1人	4.3人	2.9人	0.2人

- 提案システムを利用する場合のみ

類似性		集中性		協調性		情報性
スコア	リンク	スコア	リンク	スコア	リンク	スコア
11.2人	8.0人	5.8人	3.5人	7.1人	3.8人	2.5人

信頼性に対する意識

- 実験前

	普段からレビューがスパムか意識していますか？
はい	9(43%)
いいえ	12(57%)

- 実験後

	信頼性を判断する必要を感じましたか？
はい	20(95%)
いいえ	0(0%)
どちら	

信頼性に関する意識の向上↑

まとめと今後の課題

- まとめ
 - レビューを対象とした信頼性指標とシステム提案
 - 信頼性の意識の向上を促進
 - 有効な判断支援
- 今後の課題
 - 大規模な実験・データセットの利用
 - ◆被験者数やデータセットの期間を増やす
 - 別の指標の検討
 - ◆レビュー本文の意味や評価値との整合性など