

# 自然言語処理を用いた Web情報の評価

9班

2018年10月19日

村山喬則 芦田佳樹 池田遼 渡邊竣  
アドバイザー教員 高安亮紀 遠藤靖典

# 発表の流れ

研究背景

自然言語処理技術の概要

先行研究

研究の手法

結果

まとめ、今後の課題

# 背景

- 通信販売サイト、口コミサイトやSNSには商品等へのユーザーからの評価やコメントが蓄積されている
- 企業が商品のレビューに含まれる情報を分析し、マーケティングに利用する例もある

Web上に掲載される  
ユーザーのコメント・評価の  
影響力が高まっている

## 43件のカスタマーレビュー

★★★★☆ 5つ星のうち4.5



## この商品をレビュー

他のお客様にも意見を伝えましょう

カスタマーレビューを書く



広告のフィードバック

## カスタマーの画像



カスタマーの画像をすべて表示

## 43件中1 - 8件目のレビューを表示

トップレビュー

Amazon カスタマー

★★★★☆ 軽い、可愛い、音質良くて保証付き&磁石は助かるー

2018年9月14日

色: 【進化版1.0】 7.5時間音楽再生 | Amazonで購入

今までネックに引っ掛けるタイプの物を使用してましたが、それさえ鬱陶しくてこの商品を購入。シンプルだし、ピンクが可愛くて防水等や保証もあったので買ってみましたが、思った以上の音質でした。

耳が小さい方なので、ハマりきってるのかよくわかりませんがちゃんと聞こえてるし何より軽くて音質良い！それで磁石なのでくちゃくちゃになる事もないので長く使えそうです。

消耗品かもしれませんが、何だかんだ使うものだし程々に高いので早々買えないので大事に使います。良い買い物だったかな？

1人のお客様がこれが役に立ったと考えています

役に立った | 役に立たなかった | コメント | 違反を報告

LIU JING

★★★★★ 便利

2018年9月25日

色: 【進化版1.0】 7.5時間音楽再生 | Amazonで購入

綺麗なピンク色でした！iPhone7を使っているので、充電口とイヤホン口は統一されていたので、充電するときにはイヤホンを使えないのが中々面倒臭いです。このイヤホンだと充電しながら使えますので、便利です！操作も簡単にできるので、いい感じですよ。

1人のお客様がこれが役に立ったと考えています

# 背景

- 口コミサイトは製品だけでなくサービスや企業の情報の主要な入手元となっている
- SNSを活用して新商品の宣伝をするなど広告としての役割も担うようになってきた



twitterより抜粋

消費者・企業からのSNSや口コミサイトへの関心が高まっている

# Web上に存在する情報の問題点

2018年9月20日 00:08 [1162131-1]

満足度	★★★★★	3
デザイン	★★★★★	5
高音の音質	★★★★★	4
低音の音質	★★★★★	4
フィット感	★★★★★	2
外音遮断性	★★★★★	4
音漏れ防止	★★★★★	3
携帯性	★★★★★	5

おいしい

【デザイン】  
以前はur Beatsを使ってました。デザインはシンプルで好きです。

【高音の音質】  
ur Beatsと比べるとシャリシャリと目立ちます。これは、iPhoneのイコライザーをdancelに変えたら収まりました。

【低音の音質】  
低音は、あまり強調されていない気がします。ur Beatsの方が出ていた

【フィット感】  
これが一番今回の難点です。ur Beatsでは、一番小さなイヤークリップを使っていたので今回も一番小さいのを付けたところ小さ過ぎてダメでした。次のサイズでは大き過ぎるし、ur Beatsのだと形状的に合わないらしく…ガッカリです。

【外音遮断性】  
イヤークリップが合わないので若干外音を拾います。

【音漏れ防止】  
イヤークリップが合えば音漏れはしないでしょう

【携帯性】  
カバンの内ポケットに入るくらいケースも小さいので携帯性はいいです。

【総評】  
イヤークリップがバッチリだったら90点くらいではないでしょうか…合うイヤークリップを探したいと思います。 <http://kakaku.com/item/J0000026886/>

## 自動車 評価レビュー 一覧

オートックワンで見積もりした人、成約した人の口コミをチェック!

その車に興味をもったきっかけは?  
見積もりをして、成約に至った決め手は?  
ディーラーとのやり取りや、試乗した際の感想など、  
新車購入するときの参考意見としても役立ててください。

ユーザー試乗レビューは  
こちらから Click

話題の車、人気の車をユーザーが評価  
監修: 桃田健史さん(自動車ジャーナリスト)



3000件中 1~10 件を表示 1 2 3 4 5 ... 300 >

見続したよ カサイさん 三菱 パジェロ

二世世代の家族でレジャーへ出かける際にも使えて、且つ普段の通勤や買い物にも使える車種を探していたところ、パジェロが候補にあがりました。見積もり前の下調べで気に入ったポイントが、スライドドアではなくヒンジドアタイプである事、ガソリンとディーゼルエンジンが選択できる…

選んだポイント: 安全性能 / 走行性能 / 荷室の広さ 投稿日: 2018年08月26日

[カサイさんの評価レビューを読む](#)

見続したよ カサイさん 三菱 アイ・ミーブ

燃費(電費)の良さとボディサイズの手頃さで興味が湧き、見積もりサイトを利用させていただきました。しかも、ミッドシップなコンパクトカーなので、走行性能も良さそうでしたので。カタログが手元になくて、グレードが選べなかった旨のコメントを付けて見積もりをお願いしたところ…

選んだポイント: 走行性能 / 維持費や燃費 投稿日: 2018年08月24日

[カサイさんの評価レビューを読む](#)


<https://autoc-one.jp/review/>

各消費者がそれぞれの価値観で評価しているため、信頼性に欠ける


レビューがコメントのみであり、製品の良し悪しが一目では理解できない

# 研究目的

Web上に載せられた商品レビュー等は  
簡単には判断できない



コメントなど文章で寄せられた評価を  
自然言語処理を用いることで定量化



点数化やレーダーチャートなど  
多くの人に理解しやすい評価指標を与える

# 分析手法

(実装は基本Python)

## Web情報の取得

Webスクレイピング



## 自然言語処理

形態素解析

Janome



感情分析

感情極性対応表の利用



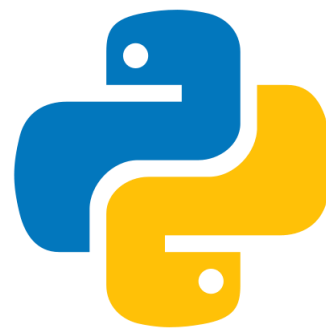
# Webスクレイピング

ウェブサイトから情報を抽出するコンピュータソフトウェア技術

Selenium(<https://pypi.org/project/selenium/>),

ブラウザ毎のWebDriverを利用。

ブラウザを自動で操作し，情報を取得





# 形態素解析

形態素：意味を持つ表現要素の最小単位（言葉の元素）

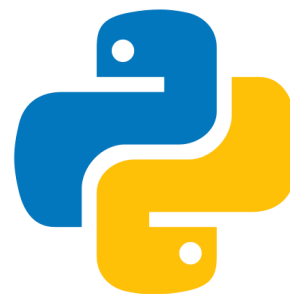
日本語はわかち書きされないため、複雑

ラティス上の経路予測、条件付き確率場（CRF）

例：Janome(Python)

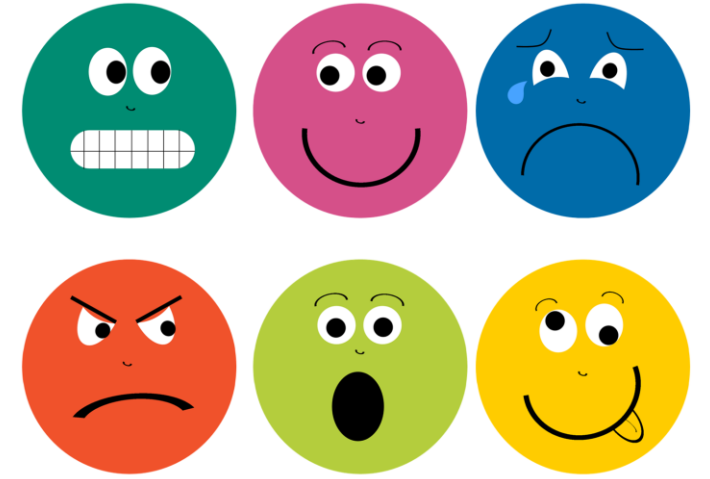
```
>>> for token in t.tokenize(u'リスク工学グループ演習9班が発表します'):
...     print(token)
...
リスク 名詞,一般,*,*,*,*,リスク,リスク,リスク
工学   名詞,一般,*,*,*,*,工学,コウガク,コーガク
グループ 名詞,一般,*,*,*,*,グループ,グループ,グループ
演習   名詞,サ変接続,*,*,*,*,演習,エンシュウ,エンシュー
9      名詞,数,*,*,*,*,9,*,*
班     名詞,接尾,助数詞,*,*,*,班,ハン,ハン
が     助詞,格助詞,一般,*,*,*,が,ガ,ガ
発表   名詞,サ変接続,*,*,*,*,発表,ハッピーウ,ハッピーウ
し     動詞,自立,*,*,サ変・スル,連用形,する,シ,シ
ます  助動詞,*,*,*,特殊・マス,基本形,ます,マス,マス

>>> tokens = t.tokenize(u'リスク工学グループ演習9班が発表します', wakati=True)
>>> tokens
['リスク', '工学', 'グループ', '演習', '9', '班', 'が', '発表', 'し', 'ます']
```



# 感情極性対応表

ポジティブ・ネガティブの極性値  
言語学的知見に基づく  
人手により評価の確かさが保証されている



PN Table（東工大高村ら）では極性辞書として、  
ネガポジの2極性値を-1~+1

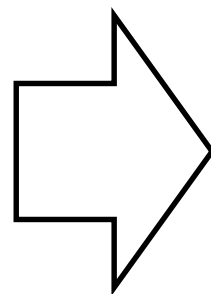
[http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html)

# 極性辞書

## PN Table

高村らが作成した辞書で、単語ごとに-1~+1の間の値の極性情報が対応されているデータベース  
言語学的知見に基づき、人手により評価の確かさが保証されている。

優れる:すぐれる:動詞:1  
良い:よい:形容詞:0.999995  
喜ぶ:よろこぶ:動詞:0.999979  
褒める:ほめる:動詞:0.999979  
めでたい:めでたい:形容詞:0.999645  
賢い:かしこい:形容詞:0.999486  
善い:いい:形容詞:0.999314  
適す:てきす:動詞:0.999295  
天晴:あっぱれ:名詞:0.999267  
祝う:いわう:動詞:0.999122  
功績:こうせき:名詞:0.999104  
賞:しょう:名詞:0.998943  
嬉しい:うれしい:形容詞:0.998871  
喜び:よろこび:名詞:0.998861  
才知:さいち:名詞:0.998771



PN Tableに記載されていない単語  
については極性値を0とする

# 先行研究

Web文書集合からの意見抽出と着眼点に基づく要約生成[2], 立石ら(2004)

立石らは掲示板等のWebサイトから意見を抽出し、着眼点と評価値の軸で分類し、レーダーチャートを作成することで意見の要約を行なった。

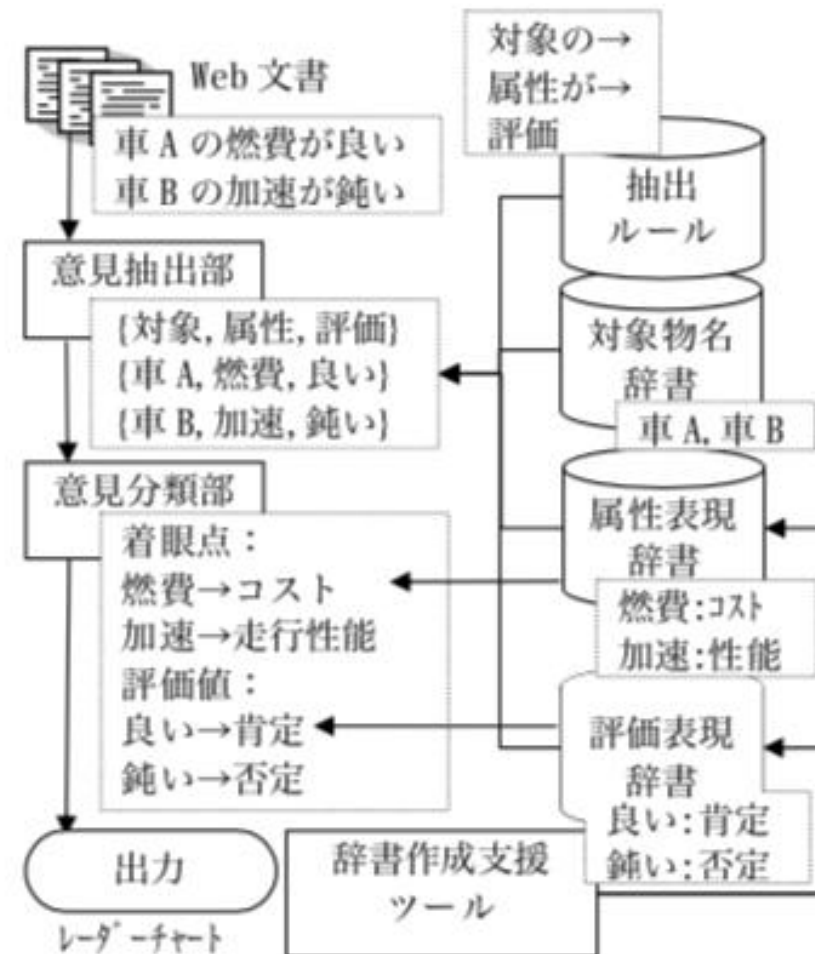
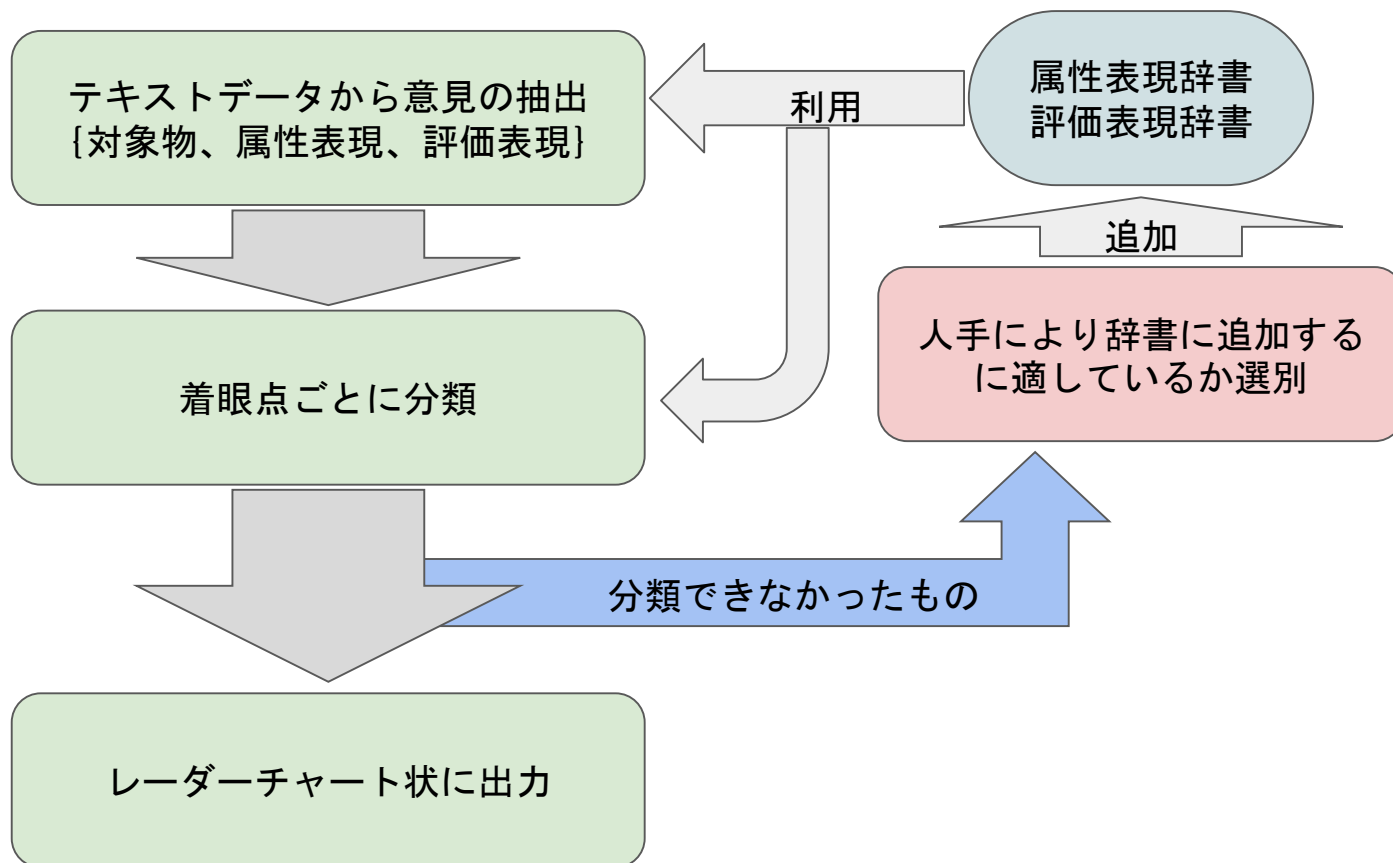


図 立石らによる研究のシステム構成([2]より抜粋)

# 意見抽出による評価属性ごとの定量化手法

立石ら[2]が意見情報抽出と着眼点に基づく要約生成を行った際に用いた手法を基に意見情報から評価がされている箇所を抽出、属性ごとに評価をまとめる

例：商品レビュー（対象：車）

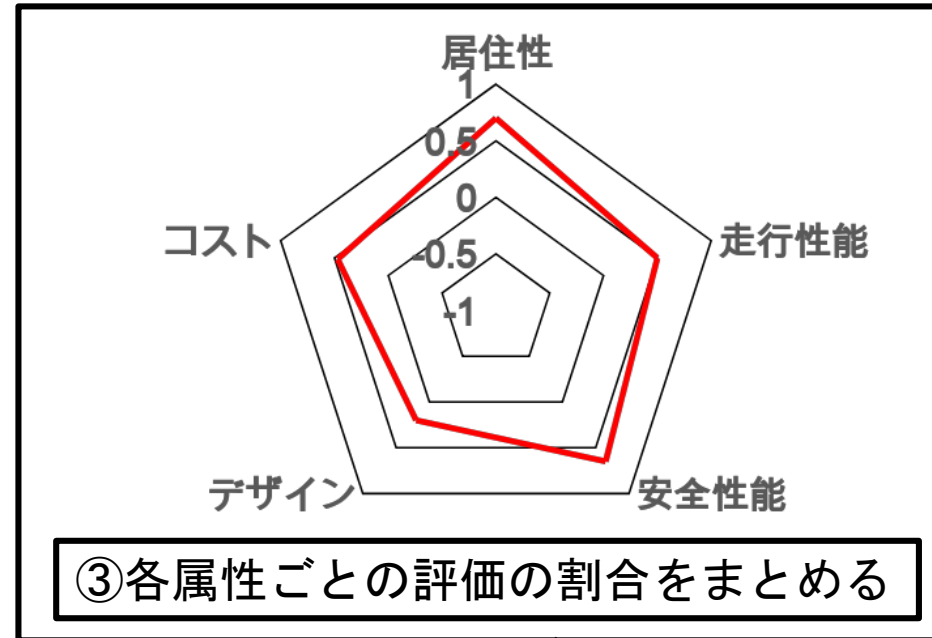
「～～内装が綺麗で良いけど、～～エンジンがうるさい～」

①抽出（構文解析と形態素解析を利用）  
{内装, 良い}、{エンジン, うるさい}

②辞書の作製

表 属性表現、評価表現の辞書

評価軸	属性表現
居住性	内装、荷物、…
走行性能	エンジン、…
安全性能	…
⋮	⋮
評価値	評価表現
positive	良い、いい、…
negative	うるさい、…

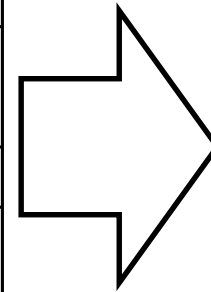


# 意見抽出ルール

構文解析を用いて意見抽出ルールに基づき対象についての意見が記述されている箇所を抽出

表 意見抽出ルール

抽出ルール	例
(属性:が/は/も/の/に/ を/で)→(評価)	デザインが良い, 外観も好き
(評価:<連体修飾>) →(属性)	良いデザイン
(属性)=(評価)	デザイングッド
(属性:の)→(*:が/は /も)→(評価)	デザインの質が良い
(属性:も/や/と/、)→ (*:も/は/が/で)→(評価)	デザインも広告も良い



**{属性表現, 評価表現}  
の形式で抽出**

例:

{内装, 良い}

{エンジン, うるさい}

# 属性表現辞書、評価表現辞書の作製

意見抽出ルールを用いて抽出された属性表現と評価表現の組からそれぞれの辞書を作る

属性表現、評価表現の出現頻度が大きいものから評価軸（レーダーチャートの軸）となる項目を人手により選択

## 属性表現候補：出現回数

音': 1108, 'の': 1042, '\*¥n': 904, 'イヤホン': 828, '音質': 370, '商品': 342, 'もの': 335, '値段': 297, 'こと': 245, '222, '低音': 217, 'これ': 206, '感じ': 205, '割': 202, 'さ': 158, '部分': 128, 'よう': 120, 'レビュー': 118, '製品': 109, 'ため': 102, 'ほう': 100, 'ん': 99, '用': 98, '物': 96, '時': 96, '高音': 94, 'cospa': 92, '自分': 92, 'フィット感': 91, '感': 88, 'ところ': 86, '色': 85, 'こちら': 82, '気': 82, 'それ': 80, 'コストパフォーマンス': 79, '的': 74, '音楽': 74, '点': 73, 'こだわり': 68, 'イヤ': 65, '耐久性': 61, '音漏れ': 61, 'ケーブル': 60, '買い物': 60, '使用': 58, '事': 57, '購入': 57, '中': 55, 'とき': 53, '1': 52, 'デザイン': 51, '心地': 50, '問題': 50, '前': 48, '音域': 47, '品質': 47,

レーダーチャートの  
評価軸の項目となる

表 属性表現、評価表現の辞書

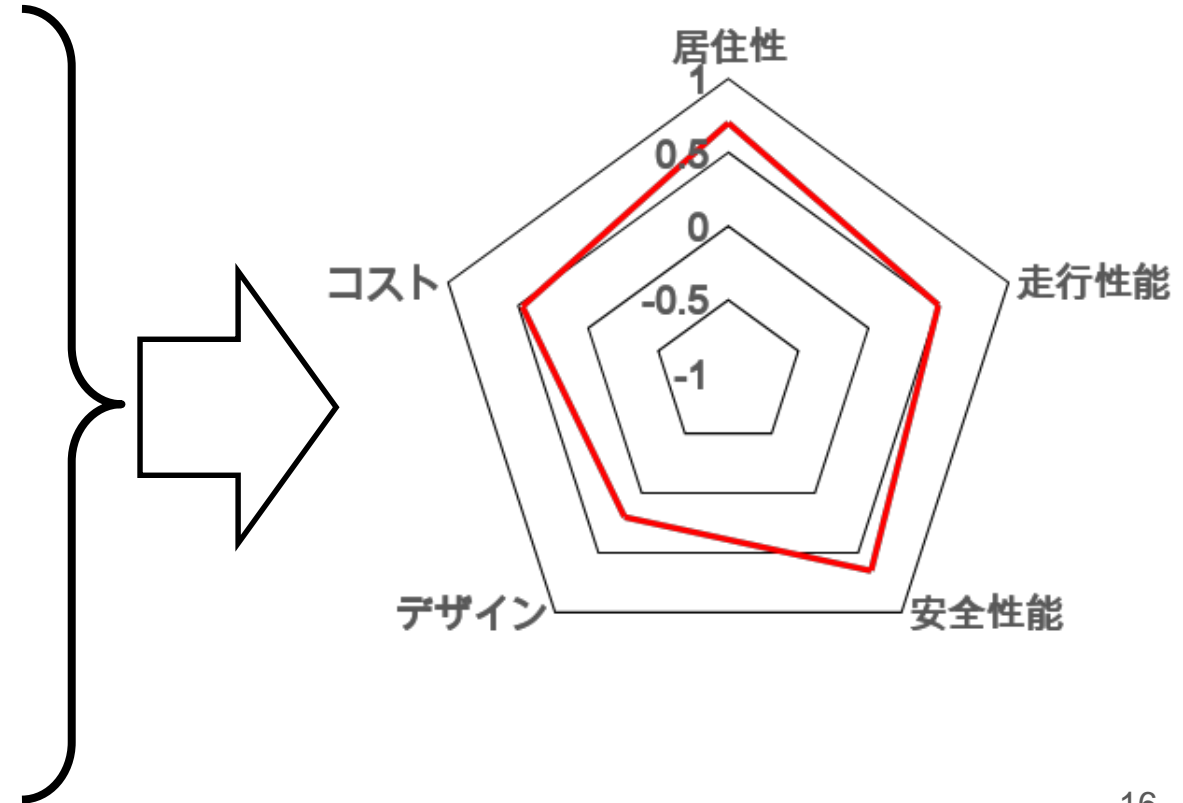
評価軸	属性表現
居住性	内装、荷物、...
走行性能	エンジン、...
安全性能	...
⋮	⋮
評価値	評価表現
positive	良い、いい、...
negative	うるさい、...

# 評価軸ごとの評価をまとめる

評価軸の項目ごとの評価表現をPositive 1, Negative -1 として平均する。  
レーダーチャート状に項目ごとの評価値をまとめる。

レビュー(例)	居住性	走行性能	安全性能	デザイン	コスト
この車は内装が綺麗で良いけど、 高速道路ではエンジンがうるさい	1	-1	0	0	0

■ ■  
■ ■  
■ ■

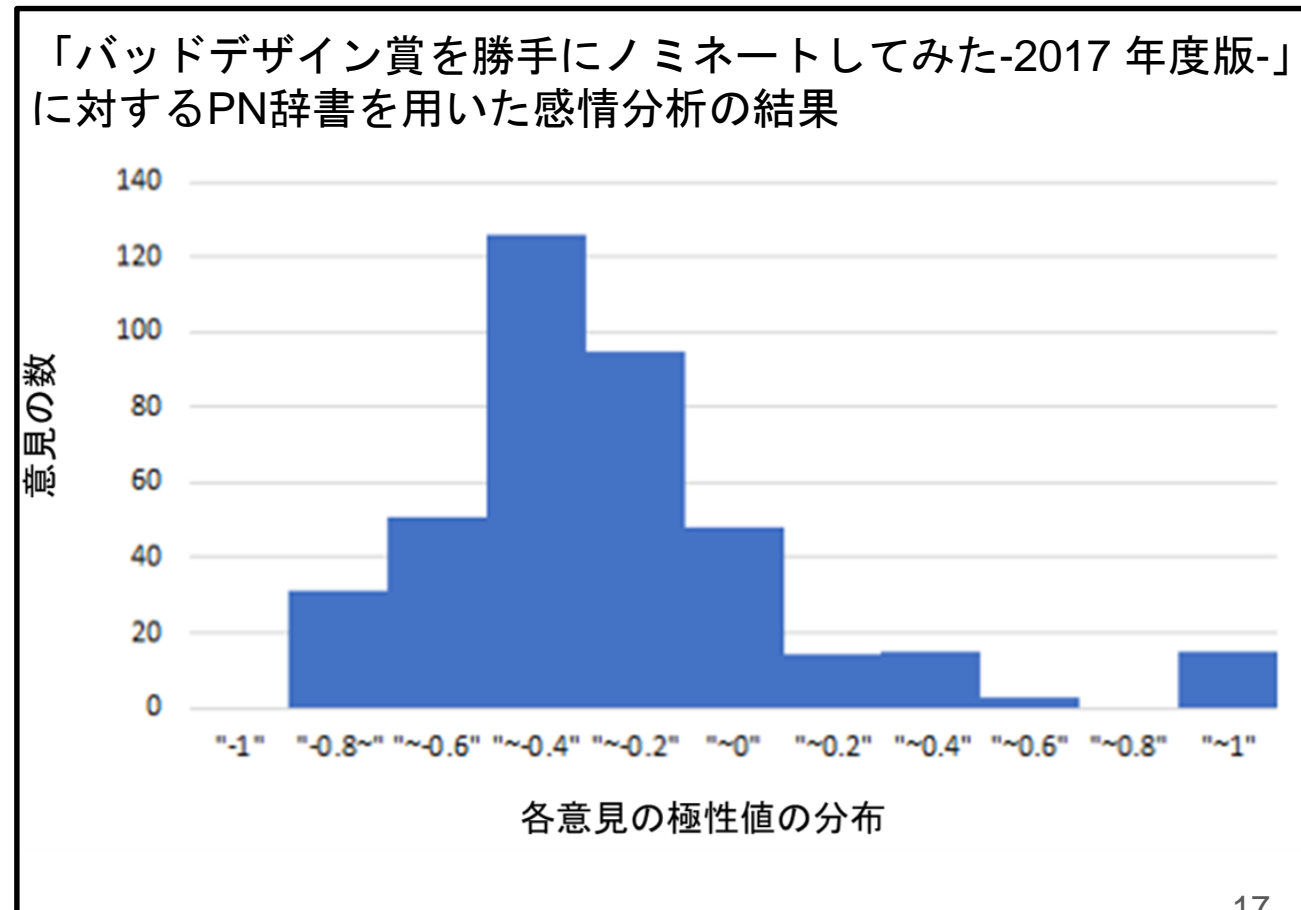




# 結果：PN値を用いたテキストデータの定量化

中間発表時は、ブログ記事に対するtwitter上のコメントにPN辞書を用いた感情分析を行った

- ブログ記事のコメントに対して一定の評価を与えることが出来た
- 一方で評価が既存の辞書に依存しているため、文書ごとに意味が異なるような単語などに対応出来ていない



# 結果: 意見抽出による評価属性ごとの定量化\_対象データ

対象データは「評価軸の多様さ」と「レビュー数の多さ」からイヤホンを選択した

< 「イヤホン」の検索結果に戻る



画像にマウスを合わせると拡大されます

パナソニック カナル型イヤホン ホワイト RP-HJE150-W

パナソニック(Panasonic)

★★★★☆ 3,628件のカスタマーレビュー | 38人が質問に回答しました

Amazon's Choice イヤホン

価格: ¥ 856 ✓prime

この商品の特別キャンペーン スマートスピーカーEcho Dotが35%OFFの3240円 2件

Amazonクラシックカード新規ご入会で9,000ポイント

新品の出品: 34 ¥ 856より 中古品の出品: 3 ¥ 349より

色: ホワイト



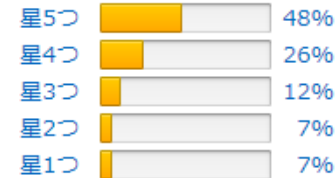
- ノイズキャンセリング:非対応
- サラウンド:非対応
- 最大入力:200 mW
- 重量:3 g
- ワイヤレス非対応
- コードの長さ:約1.2m
- 形式: 密閉型ダイナミック・ステレオインサイドホン
- 付属品: イヤーピース(XS/S/Mサイズ)

> もっと見る

## カスタマーレビュー

★★★★☆ 3,628

5つ星のうち3.9



すべてのカスタマーレビューを見る (3,628)

## 特長別の星評価

使いやすい

★★★★☆

5つ星のうち3.8

耐久性

★★★★☆

5つ星のうち3.8

ノイズ除去

★★★★☆

5つ星のうち3.4

3628件のレビュー全体を  
「5段階評価の平均」と「評価値の分布」  
で要約している

# 結果: 意見抽出による評価属性ごとの定量化\_候補の抽出

抽出ルールに基づき、属性表現と評価表現の候補を抽出する

## 抽出ルール (再掲)

抽出ルール	例
(属性: が/は/も/の/に/ を/で) → (評価)	デザインが良い, 外観も好き
(評価: <連体修飾>) → (属性)	良いデザイン
(属性) = (評価)	デザイングッド
(属性: の) → (*: が/は /も) → (評価)	デザインの質が良い
(属性: も/や/と/、) → (*: も/は/が/で) → (評価)	デザインも広告も良い

候補表現を抽出

## 属性表現候補 : 出現回数

音': 1108, 'の': 1042, '\*¥n': 904, 'イヤホン': 828, '音質': 789, '方': 399, '耳': 370, '商品': 342, 'もの': 335, '値段': 297, 'こと': 245, '私': 235, 'コード': 222, '低音': 217, 'これ': 206, '感じ': 205, '割': 202, 'さ': 197, '価格': 182, '人': 158, '部分': 128, 'よう': 120, 'レビュー': 118, '製品': 116, '非常': 109, '評価': 109, 'ため': 102, 'ほう': 100, 'ん': 99, '用': 98, '物': 96, '時': 96, '高音': 94, 'cospa': 92,

## 評価表現候補 : 出現回数

'良い': 1400, 'いい': 841, 'ない': 688, '安い': 583, '購入': 530, '\*¥n': 389, '高い': 331, '悪い': 315, '音': 259, 'よう': 209, '無い': 192, '使用': 188, 'よい': 187, '耳': 177, '音質': 172, '小さい': 160, '大きい': 153, '的': 148, '細い': 146, 'フィット': 132, '痛い': 131, '多い': 129, '十分': 129, 'イヤホン': 126, '値段': 123, '満足': 116, 'すごい': 111, '断線': 110, '気': 101, '問題': 98, 'これ': 97, 'やすい': 92, '長い': 86, 'クリア': 85, '欲しい': 84, '軽い': 81, '付属': 81, '用': 78, '以上': 76, '低音': 75, '強い': 72, '感': 72, 'オススメ': 67, '素晴らしい': 67, '普通': 66, '弱い': 65, '安っぽい': 64, '最高': 64

# 結果: 意見抽出による評価属性ごとの定量化\_カテゴリ分け1

抽出した属性表現候補を着眼点に基づき人手で属性カテゴリに分類する

## 属性表現候補：出現回数

'音': 1108, 'の': 1042, '\*¥n': 904, 'イヤホン': 828, '音質': 789, '方': 399, '耳': 370, '商品': 342, 'もの': 335, '値段': 297, 'こと': 245, '私': 235, 'コード': 222, '低音': 217, 'これ': 206, '感じ': 205, '割': 202, 'さ': 197, '価格': 182, '人': 158, '部分': 128, 'よう': 120, 'レビュー': 118, '製品': 116, '非常': 109, '評価': 109, 'ため': 102, 'ほう': 100, 'ん': 99, '用': 98, '物': 96, '時': 96, '高音': 94, 'cospa': 92, '自分': 92, 'フィット感': 91, '感': 88, 'ところ': 86, '色': 85, 'こちら': 82, '気': 82, 'それ': 80, 'コストパフォーマンス': 79, '的': 74, '音楽': 74, '点': 73, 'こだわり': 68, 'イヤ': 65, '耐久性': 61, '音漏れ': 61, 'ケーブル': 60, '買い物': 60, '使用': 58, '事': 57, '購入': 57, '中': 55, 'とき': 53, '1': 52, 'デザイン': 51, '心地': 50, '問題': 50, '前': 48, '音域': 47, '品質': 47,



属性カテゴリ	属性
総合	'イヤホン', '商品', '製品', '買い物', '質', 'クオリティ', '性能'
音質	'音', '音質', '高音', '重低音', '中音', '音響', '音色', '音域', 'サウンド', '音色'
デザイン	'デザイン', 'デザイン性', '色', '見た目', '形', '形状', 'カラー', 'レッド', 'ブルー', 'ピンク色'
付け心地	'フィット感', 'フィット', 'フィット性', '装着感', '装着', '心地', 'つけ心地'
コスパ	'コストパフォーマンス', 'コスパ'

# 結果: 意見抽出による評価属性ごとの定量化\_カテゴリ分け2

抽出した評価表現候補を着眼点に基づき人手で評価カテゴリに分類する

## 評価表現候補：出現回数

'良い': 1400, 'いい': 841, 'ない': 688, '安い': 583, '購入': 530, '\*¥n': 389, '高い': 331, '悪い': 315, '音': 259, 'よう': 209, '無い': 192, '使用': 188, 'よい': 187, '耳': 177, '音質': 172, '小さい': 160, '大きい': 153, '的': 148, '細い': 146, 'フィット': 132, '痛い': 131, '多い': 129, '十分': 129, 'イヤホン': 126, '値段': 123, '満足': 116, 'すごい': 111, '断線': 110, '気': 101, '問題': 98, 'これ': 97, 'やすい': 92, '長い': 86, 'クリア': 85, '欲しい': 84, '軽い': 81, '付属': 81, '用': 78, '以上': 76, '低音': 75, '強い': 72, '感': 72, 'オススメ': 67, '素晴らしい': 67, '普通': 66, '弱い': 65, '安っぽい': 64, '最高': 64, '少ない': 64, '好き': 61, '音楽': 58, 'おすすめ': 55, 'コード': 55, '帯': 54, '残念': 54, '高価': 53, '価格': 53, '非常': 52, '必要': 52, '今': 52, '0': 50, '低い': 49, '一番': 49, '安価': 45, '何': 44, '装着': 44, 'そう': 44, '期待': 43,



評価カテゴリ	評価表現
positive	‘良い’, ‘いい’, ‘よい’, ‘高い’, ‘素晴らしい’, ‘すばらしい’, ‘最高’, ‘好き’, ‘十分’, ‘綺麗’, ‘凄い’, ‘豊富’, ‘抜群’, ‘可愛い’, ‘シンプル’, ‘かわいい’, ‘良好’, ‘嬉しい’, ‘感動’
negative	‘悪い’, ‘微妙’, ‘悪意’, ‘安っぽい’, ‘弱い’, ‘少ない’, ‘低い’, ‘残念’, ‘最低’, ‘酷い’, ‘荒い’, ‘不安’, ‘ダメ’, ‘貧弱’, ‘おかしい’, ‘イヤ’, ‘最悪’

# 結果: 意見抽出による評価属性ごとの定量化\_評価例

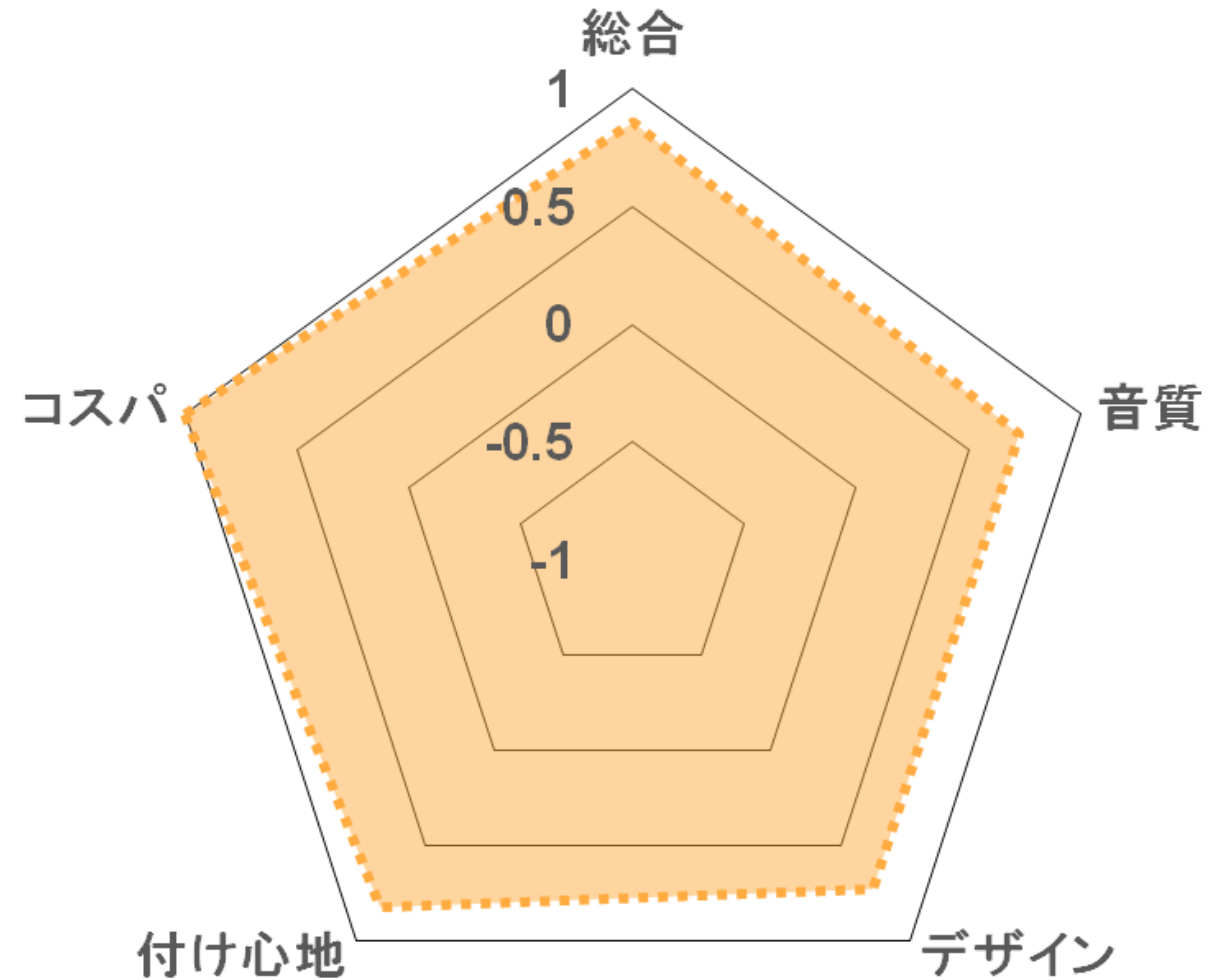
作成した属性表現辞書と評価表現辞書を用いて、カテゴリごとのレビューに評価付けを行う

レビュー	総合	音質	デザイン	付け心地	コスパ
<u>安い割に乗りのいい音で良い.</u> <u>装着感も良好で, デザインも安</u> <u>っぽくない.</u>	0	1	1	1	0
<u>デザインはいいが,安っぽい音質</u> <u>です.</u>	0	-1	1	0	0

# 結果: 意見抽出による評価属性ごとの定量化\_結果の可視化

得られたカテゴリ別の評価をレビュー全体で平均し、結果を可視化する

- テキストデータに対して複数の評価軸で評価付けを行うことが出来た
- 一方で抽出ルールが限定的であるため、評価が付けられなかったデータも多数存在した (次スライド)



# 結果: 意見抽出による評価属性ごとの定量化\_評価率

3600件のレビュー全体の内、評価付けが行えた割合を評価率とし、属性ごとに算出した

- 複雑な文法や口語的な表現などのルールベースでの意見抽出が難しいレビューの存在が評価率の低下に繋がっていると考えられる

	総合	音質	デザイン	付け心地	コスパ
評価回数	360	817	90	103	54
評価率	0.10	0.23	0.02	0.03	0.01



## まとめ・課題

- 極性辞書を用いた手法は比較的簡単に解析が可能であり、商品レビューからSNS上の情報まで幅広い対象への対応が可能であるが、辞書への依存性が高く対象によって辞書の選定が必要である。
- 意見抽出による評価属性ごとの定量化を行う手法では、意見情報であるテキストデータから評価が記述されている箇所を抽出し、評価項目と評価の割合をレーダーチャート状で示すことができた。一定の工程は人手による判別が必要であるが、多種多様で多量の意見情報から全体の意見を定量化するという目標は達成できている。改善点としては人手による工程の削減が挙げられ、属性表現の単語の類似性を分析し、その結果をもとに評価項目を選び出す方法が考えられる。

# 参考文献

- [1] 東京工業大学 高村大也. "単語感情極性対応表", 高村大也, [¥url{http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html}](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html), (参照 2018-06-04).
- [2] 立石健二, 福島俊一, 小林のぞみ, 上出将行, 高橋哲朗, 乾孝司, 藤田篤, 乾健太郎, 松本裕治. Web文章集合からの意見情報抽出と着眼点に基づく要約生成. 情報処理学会研究報告情報学基礎(FI). 2004, no. 93, p. 1-8.

# 係り受け解析

より正確に評価可能

本学では、-----D  
例年、-----D  
省エネ及び-D |  
CO-D |  
2排出削減に-D  
資するとともに、-----D  
夏季における-D  
教職員の-D  
心身の-D  
リフレッシュ及び---D  
今夏の-D  
電力抑制に-D  
対応する-D  
ことを---D  
目的と-D  
し、-----D  
附属病院を-D  
除く---D  
すべての-D  
事業場において-----D  
下記の-D |  
期間、-----D |  
夏季における-D |  
休業を-D |  
実施いたしますので-D  
お知らせします。