

共起制限と係り受け構造グラフを利用した 宿泊レビューの特徴表現抽出手法

2017/07/8

Koji Tanaka¹, Koichi Tsujii², Taakashi Ikoma³
Akiyuki Sekiguchi⁴, Kazuhiko Tsuda³

1:Hitachi Government & Public Sector Systems, Ltd

2:Nippon Travel Agency Co. Ltd.

3:Graduate School of Business Sciences University of Tsukuba

4:Meiji University

Contents

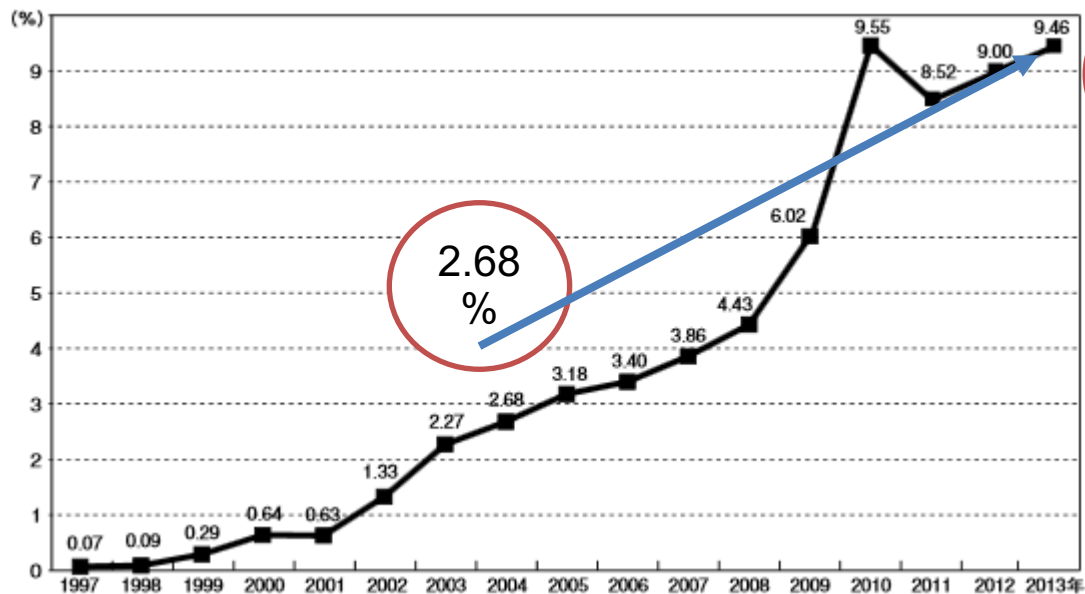
1. はじめに
2. 関連研究
3. 共起制限と係り受け構造グラフ
4. 特徴表現抽出結果
5. おわりに



1. はじめに

1. Introduction

The Internet-sales ratio of a tour company



インターネット宿泊予約サイトの利用者数はこの10年で4倍近い成長を示している

- 宿泊検討者は最も良い条件の宿泊施設を選ぶために、宿泊予約サイトの宿泊レビューを閲覧し、その評価を宿泊施設選択の参考情報として重要視している
- 宿泊施設にとっても、宿泊レビューから改善点やアピールポイントを把握するなど、マーケティングに宿泊レビューを活用することが期待されている

1. はじめに

- 感想コメントは宿泊者が自由に記述する情報のため、表現の揺れが大きい。



類義語辞書を作成し、表現を揃える処理を行う 必要がある [4, 5]



類義語辞書の作成に対する課題

- 手間がかかる
- 宿泊レビュー投稿数が増加した場合に見直しが必要になる
- 想定していない評価対象に関しては類義語辞書が作成されず、想定外の評価対象に関する評価が困難になる



予め評価対象を設定することや類義語辞書を作成することなく、宿泊レビュー中の感想コメントから評価に関する特徴表現を抽出する手法の検討を行った

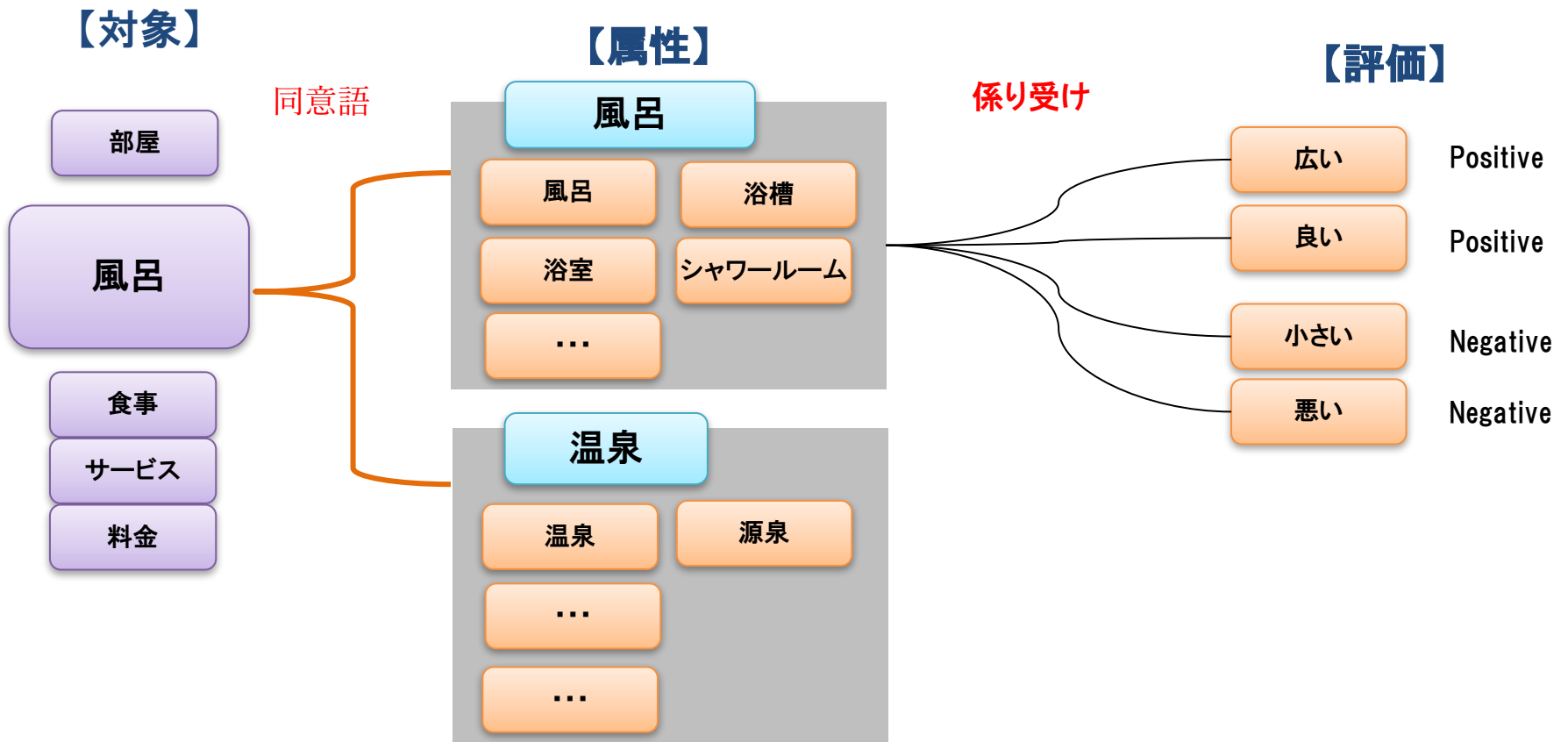


2. 関連研究

2. 関連研究

2.1 A. 宿泊レビュー分析における評価属性辞書

辻井らの研究[6]では「評価属性辞書」作成して、特徴表現を抽出
(小林らの評判分析に関する研究[7]を、宿泊レビューに適用すると共に拡張)



2. Related work

2.2 シソーラス

語の意味的な類似性を表現する辞書として利用される



概念を追加・更新するためには人間の手作業による膨大な手間がかかる



精度の高いシソーラス辞書を低コストで(半)自動的に作成する手法が必要とされている

過去の研究

- 語のクラスタリング手法を用いるもの[10]
- Webページ間のリンク構造を利用して作成するもの [11, 12].

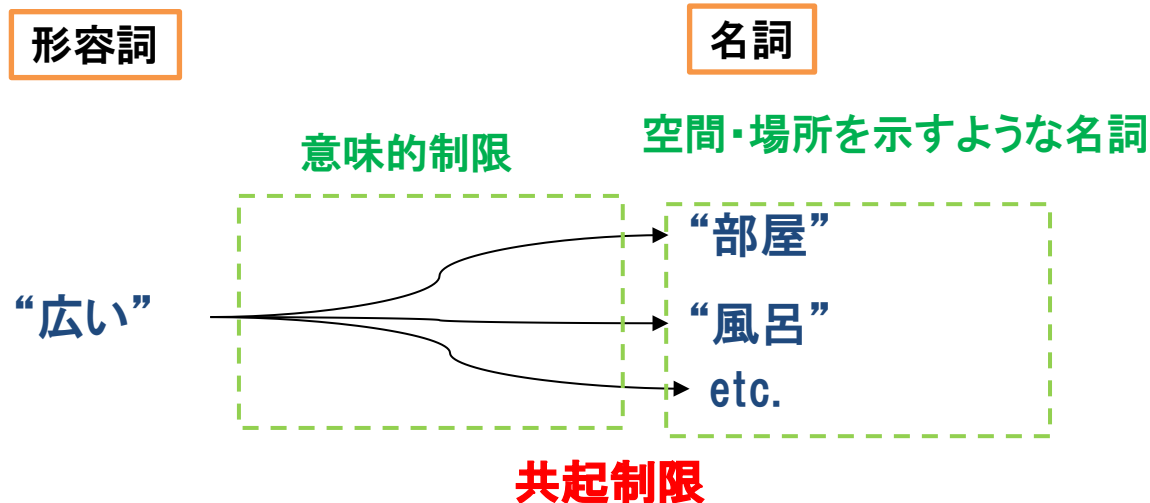
これらの手法は、似た内容の文書間で利用される単語は、意味的な類似性をもった単語が利用されやすいことを利用している

2. Related work

2.2 シソーラス

語の共起関係に基づいてシソーラス辞書を作成する研究[8]もある

語の係り受け関係には意味的な制限があることが知られている



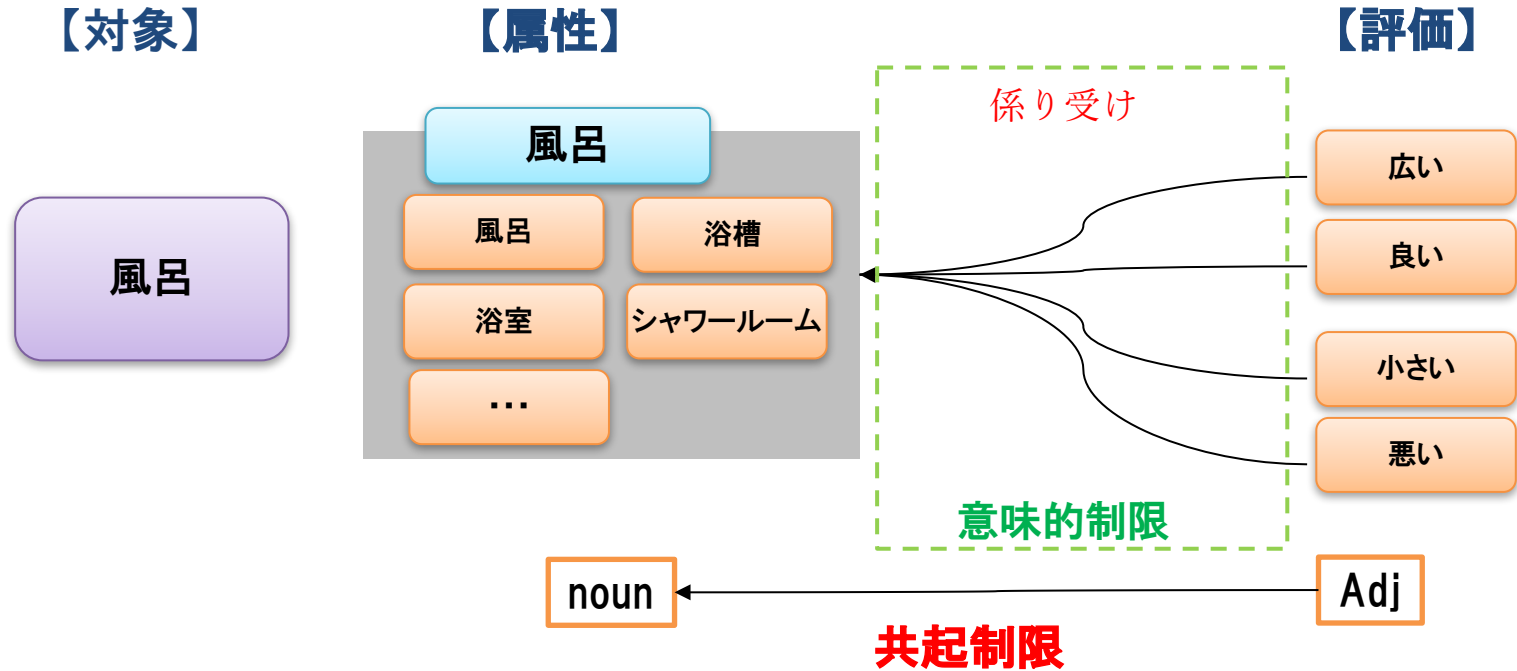
Related work

- 名詞に対して他の単語との係り受け関係の頻度を利用するもの [13]
- 複合語内の名詞に注目して名詞間の接続関係を利用するもの [14]
- 全ての品詞に利用して同義表現を抽出しようとするもの [15].



3. 共起制限と係り受け構造グラフ

3.1. 共起制限と係り受け構造グラフ

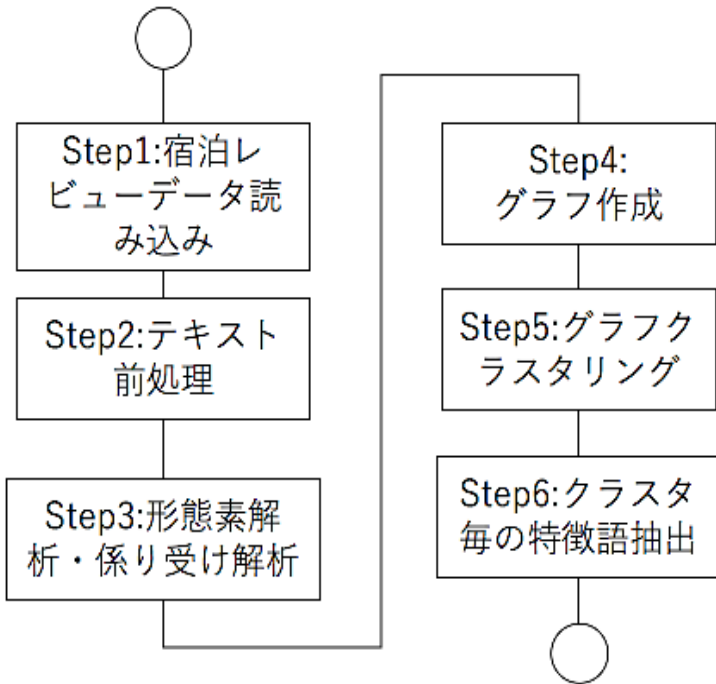


本研究では

係り受け構造をグラフ構造で表現し、そのグラフをクラスタリングことで、宿泊レビュー中の感想コメントから評価に関する特徴表現を抽出する

3.1. Co-Occurrence Restrictions and Dependency Structure Graphs

Detailed procedure



Step 1. 宿泊レビューデータを読み込む

- レビュー対象の宿泊施設名称
- 日本語で記載されている感想コメント

Step 2. テキスト前処理

全角 半角

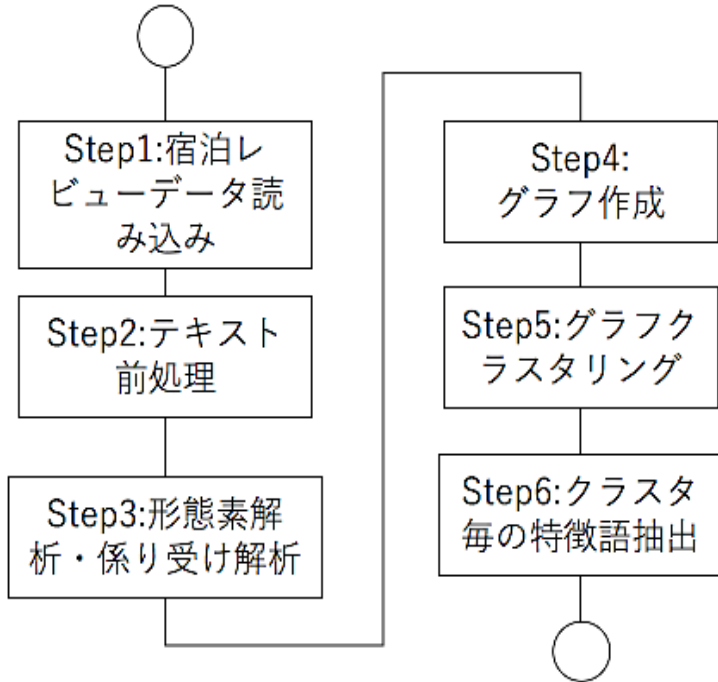
Ex) “AB” → ”AB”

マイナス記号 長音記号

Ex) “キ-” → ”キー”

3.1. Co-Occurrence Restrictions and Dependency Structure Graphs

Detailed procedure



Step 3.形態素解析・係り受け解析

形態素解析器:MeCab [16]

辞書:mecab-ipadic-NEologd [17]

係り受け解析器 :Cabocha [18].

Step 4. Step.3の解析結果からグラフデータを作成

3.1. Co-Occurrence Restrictions and Dependency Structure Graphs

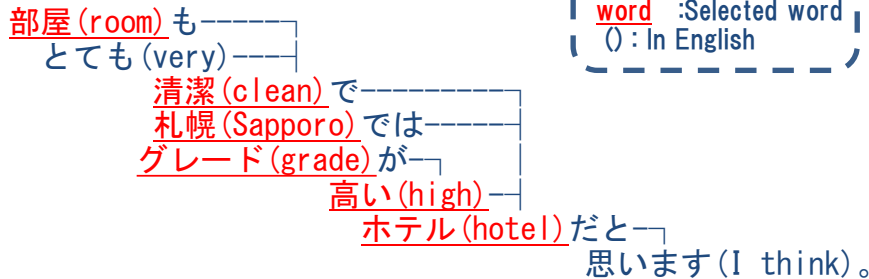
Step 4. Step.3の解析結果からグラフデータを作成

<Data>

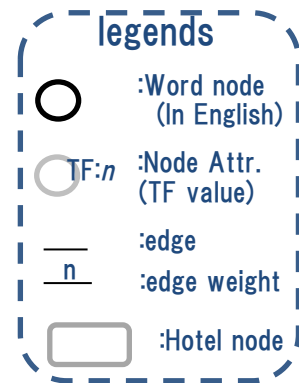
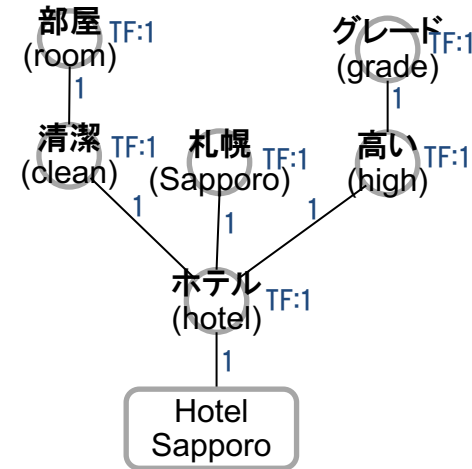
Hotel Name: Hotel Sapporo
 Impression comment : "部屋もとても清潔で、札幌ではグレードが高いホテルだと思います."
 ("The room was very clean, I think the hotel is a high grade hotel in Sapporo.")



<Result of dependency analysis>



<Graph>



【Property】

【Evaluation】

noun

Adj

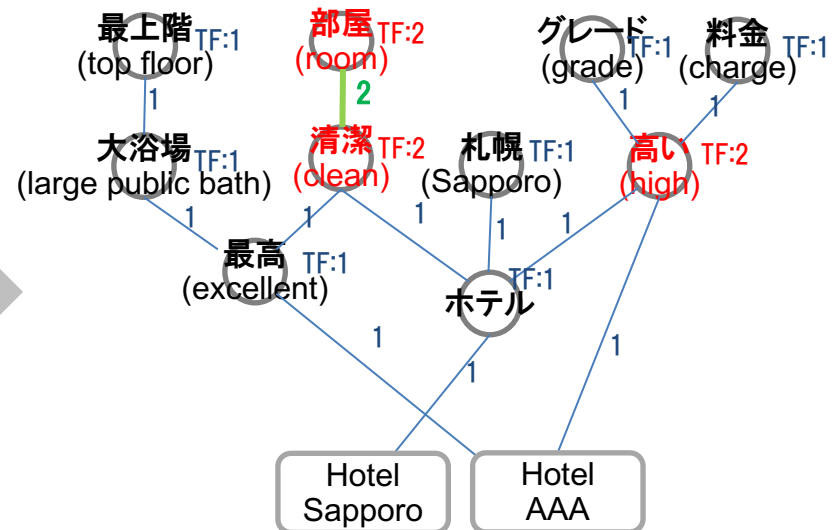
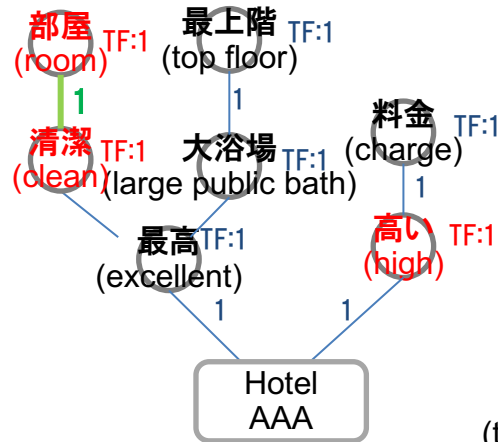
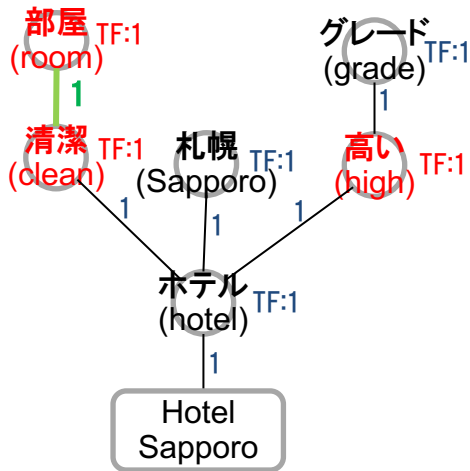
部屋 (room)
 グレード (grade)

清潔 (clean)
 高い (high)

3.1. Co-Occurrence Restrictions and Dependency Structure Graphs

Hotel Name: Hotel Sapporo
 Impression comment: "部屋もとても清潔で、札幌ではグレードが高いホテルだと思います."
 ("The room was very clean, I think the hotel is a high grade hotel in Sapporo.")

Hotel Name: Hotel AAA
 感想コメント: "部屋も清潔で、最上階にある大浴場も最高でした。でも、料金が高い。"
 ("The room was clean, and the large public bath on the top floor was excellent. But the charge was high.")



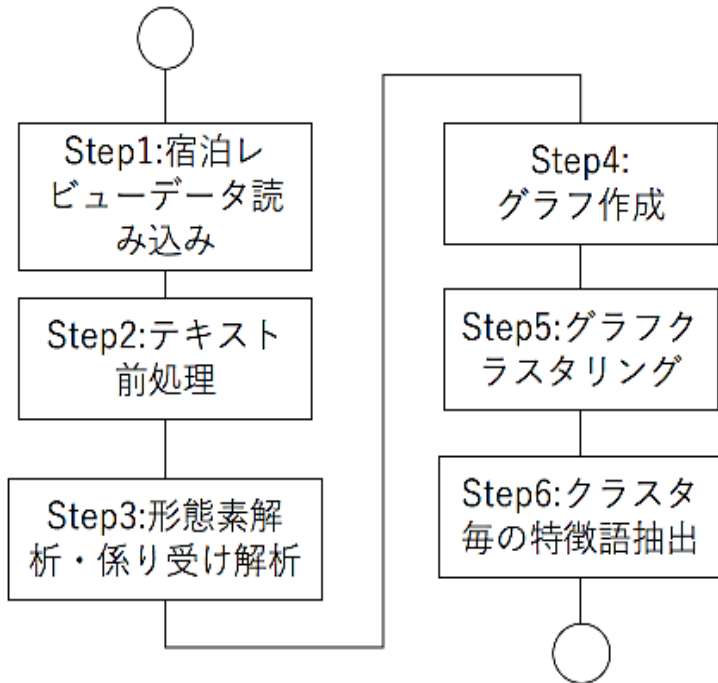
legends

- :Word node
- TF:n :Node Attr (TF value)
- :edge
- n :edge weight
- :Hotel node



3.1. Co-Occurrence Restrictions and Dependency Structure Graphs

Detailed procedure



Step 5. グラフクラスタリング.

Clustering Method

- R. Lambiotte et al

“Laplacian dynamics and multiscale modular structure in networks.” [19]

Step 6. 特徴語抽出

クラスタ毎にノードのアトリビュートに設定したTF値が高いノードを抽出



4. 特徴表現抽出結果

4. Result of extracting feature terms

Data:

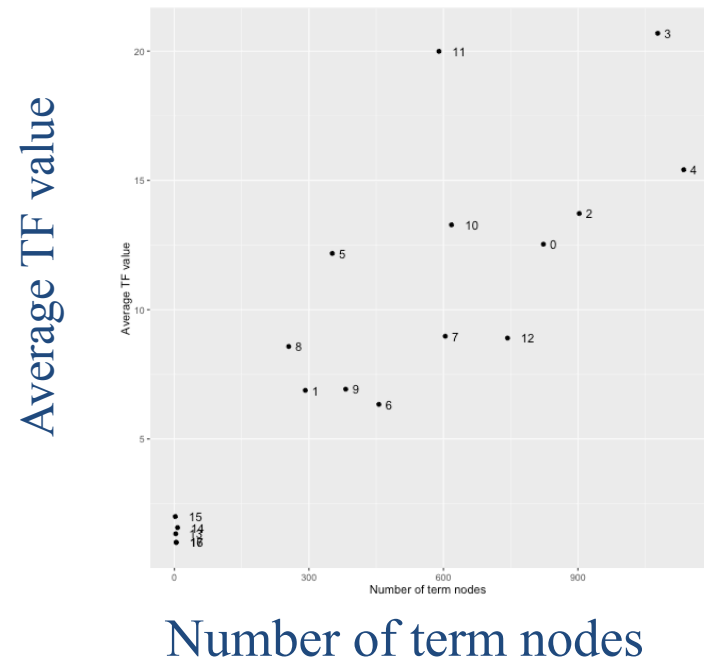
- ・辻井らの研究[6]で利用されたデータと同じもの。
- ・ビジネスエリアの50施設、感想コメント数は4986件

TABLE1.

Step 4で作成されたグラフの特徴量.

Nodes	8,302
Edges	56,071
Average Degree	13.508
Diameter	9
Average Path Length	3.217

クラスタ毎の単語ノード数と平均TF値



4. Result of extracting feature terms

TABLE II. Top 20 feature terms of high TF values in each cluster.

#	Top 20 terms
0	<p>便利, 駅, 近い, 立地, 場所, 最高, コンビニ, 食事, 博多駅, 繁華街, 条件, 地下鉄, 車, 大阪駅, 移動, 買い物, 周辺, 徒歩, 直結, 銀座</p> <p>Convenient, Station, Near, Location, Place, Excellent, Convenience store, Food, Hakata Station, Busy street, Condition, Subway, Cars, Osaka Station, Transportation, Shopping, Neighborhood, Walk, Direct access, Ginza</p>
1	<p>アメニティ, 充実, タオル, 空気清浄機, シャンプー, アメニティー, セット, 香り, 設置, 強い, 難点, 備え付け, 空気, AROMA, リセッシュ, 交換, リンス, 歯ブラシ, 清浄, グッズ</p> <p>Amenity goods, Perfect, Towels, Air cleaner, Shampoo, Amenity goods, Set, Fragrance, Equipped, Strong, Shortcoming, Equipped, Air, Aromatic, Air freshener, Replace, Hair conditioner, Toothbrushes, Clean, Goods</p>
2	<p>朝食, 満足, 大変, 美味しい, プラン, バイキング, 多い, おいしい, 特に, 種類, 無料, パン, レストラン, 豊富, メニュー, 少ない, ありがたい, 和食, 料理, 店</p> <p>Breakfast, Satisfied, Very, Tasty, Plan, Buffet, Many, Tasty, Especially, Kinds, Free, Bread, Restaurant, Abundant, Menu, Few, Appreciate, Japanese food, Food, Store</p>
3	<p>部屋, 広い, お部屋, 風呂, ベッド, 綺麗, 狭い, 快適, 清潔感, 清潔, ルーム, きれい, バス, 夜景, トイレ, 案内, 十分, 大きい, 素敵, シャワー</p> <p>Room, Large, Room, Bath, Bed, Neat, Small, Comfortable, Cleanliness, Clean, Room, Tidy, Bath, Night view, Bathroom, Guidance, Enough, Large, Lovely, Shower</p>
4	<p>利用, 宿泊, 出張, 初めて, 旅行, 機会, 名古屋, いつも, 札幌, 是非, 大阪, 子供, よく, 観光, 東京, ぜひ, 一人, 疲れ, 福岡, 2人</p> <p>Use, Stay, Business trip, First time, Trip, Occasion, Nagoya, Always, Sapporo, Definitely, Osaka, Children, Often, Sightseeing, Tokyo, Definitely, Alone, Fatigue, Fukuoka, Two people</p>

Feature terms related with

Access

Amenity

Food

Room

Occasion

4. Result of extracting feature terms

#	Top 20 terms
5	<p>値段,料金,安い,高い,駐車場,かなり,人,価格,得,リーズナブル,コストパフォーマンス,ポイント,割,金額,駐車,レベル,設定,正直,通常,追加</p> <p>Charge, Charge, Low, High, Parking lot, Fairly, People, Charge, Economical, Reasonable, Cost performance, Point, Discount, Charge, Parking, Grade, Setting, Honest, Normal, Additional</p>
6	<p>サービス,嬉しい,気分,感動,感激,結構,贅沢,帝国ホテル,ケーキ,ビックリ,心,サプライズ,本当にありがとうございました,新聞,欲,文句なし,浴衣,リッチ,堪能,個人的</p> <p>Service, Pleasing, Mood, Impressed, Impressive, Quite, Luxurious, Imperial Hotel, Cakes, Surprised, Heart, Surprising, Very thankful, Newspaper, If possible, Perfect, Yukata (robe), Rich, Satisfied, Personal</p>
7	<p>チェックイン,予約,チェックアウト,荷物,お願い,遅い,早い,到着,スムーズ,確認,事前,12時,台風,早め,急遽,おかげ,ちゃんと,満喫,夜中,じゃらん</p> <p>Check-in, Reservation, Checkout, Luggage, Request, Late, Early, Arrival, Smooth, Confirm, In advance, 12 o'clock, Typhoon, Earlier, In a hurry, Thankful, As wished, Satisfied, Midnight, Jalan (hotel reservation)</p>
8	<p>エレベーター,安心,カード,フロア,必要,キー,混雑,セキュリティ,鍵,ラウンジ,配慮,クラブ,システム,扉,セキュリティー,エレベータ,エグゼクティブ,ロッカー,女性専用,防音</p> <p>Elevator, Secure, Card, Floor, Necessary, Key, Crowded, Security, Key, Lounge, Care, Club, System, Door, Security, Elevator, Executive, Locker, Ladies only, Soundproof</p>
9	<p>大浴場,温泉,うれしい,最上階,サウナ,露天風呂,地下,入浴,1階,魅力,可能,窮屈,天然,脱衣,部分,24時間,露天,表示,シティホテル,畳</p> <p>Large public bath, Hot spring, Pleased, Top floor, Sauna, Open-air bath, Underground, Taking a bath, 1st floor, Appealing, Possible, Narrow, Natural, Undressing, Partly, 24 hours, Open-air, Signs, City hotel, Tatami mat</p>

Feature terms related with

Charge

Customer experience

Check-in / Check-out

Security / Safety

Bath / Hot spring

4. Result of extracting feature terms

#	Top 20 terms
10	ホテル, 感じ, いい, 他, 新しい, 雰囲気, ビジネスホテル, 古い, 普通, さすが, 建物, 口コミ, 楽しい, 客, 自体, 施設, 評価, 内装, おしゃれ, 好き Hotel, Impression, Good, Other, New, Atmosphere, Business hotel, Old, Ordinary, As expected, Building, Reviews, Pleasant, Guests, Itself, Facilities, Evaluation, Interior, Fancy, Like
11	良い, 対応, フロント, よい, スタッフ, 接客, 丁寧, 非常, 親切, 女性, 素晴らしい, 気持ちよい, すごい, 本当に, アクセス, 印象, 気持ち, 顔, 方々, 気持ち良い Good, Service, Front desk, Good, Staff, Service, Courteous, Very, Kind, Lady, Wonderful, Pleasant, Fantastic, Really, Access, Impression, Feeling, Face, Employees, Comfortable
12	ない, 残念, 無い, 気, 悪い, 期待, 臭い, 問題, 自分, 不便, もう, 不満, 説明, 水, まあ, 一つ, 申し分, 仕方ない, 足, タバコ Unavailable, Disappointed, Unavailable, Intend, Bad, Expectation, Smell, Problem, I, Inconvenient, Anymore, Unsatisfactory, Explanation, Water, So so, One thing, All right, Unavoidable, Feet, Tobacco

Feature terms related with

Overall evaluation

Staff service

Unsatisfaction

Orthographical variants having the same meaning appeared in a cluster

#1 Amenity goods

アメニティ
アメニティー

#3 Room

部屋
お部屋
ルーム

#4 Definitely

是非
ぜひ

#5 Charge

値段
料金
価格

#8 Security

セキュリティ
セキュリティー

#8 Elevator

エレベーター
エレベータ

#11 good

良い
よい

#11 Service

対応
接客

#12 Unavailable

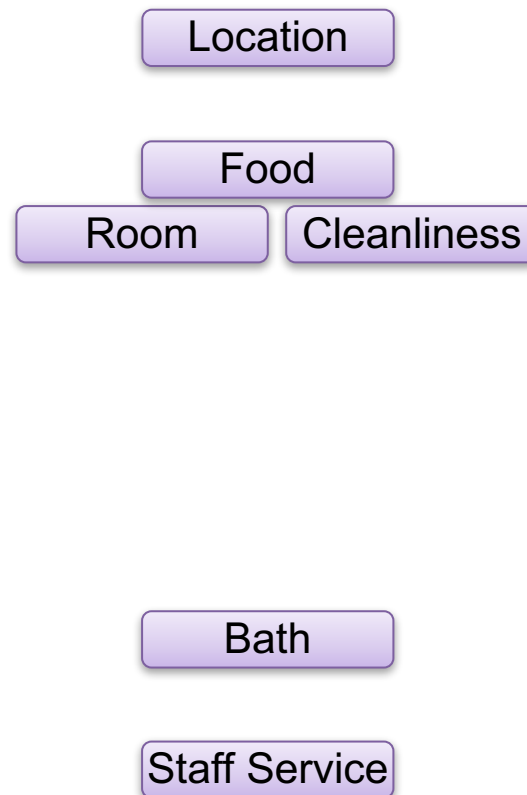
無い
ない

TABLE III
PRESUMED TOPICS OF CLUSTERS

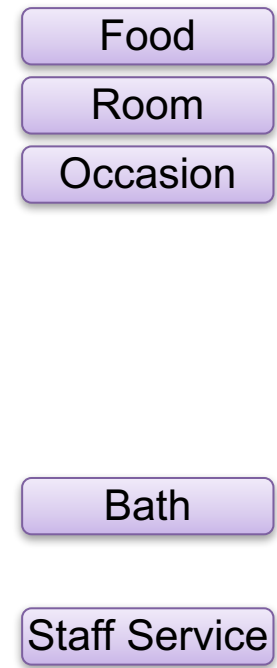
#	Topic
0	Access
1	Amenity
2	Food
3	Room
4	Occasion
5	Charge
6	Customer experience
7	Check-in / checkout
8	Security and safety
9	Bath / hot spring
10	Overall evaluation
11	Staff service
12	Unsatisfaction

<Subject> in reputation analysis

The study by Tsujii et al.



The study by Tanabe et al.





5.Conclusion

5.終わりに

類義語辞書を作成し、表現を揃える処理を行う 必要がある[4, 5]

類義語辞書の作成に対する課題

- ・ 手間がかかる
- ・ 宿泊レビュー投稿数が増加した場合に見直しが必要になる
- ・ 想定していない評価対象に関しては類義語辞書が作成されず、想定外の評価対象についての評価が困難になる


予め評価対象を設定することや類義語辞書を作成することなく、宿泊レビュー中の感想コメントから評価に関する特徴表現を抽出する手法を検討

予め評価対象を設定することや類義語辞書を作成することなく、宿泊レビュー中の感想コメントから評価に関する特徴表現を抽出することができた。

今後の作業:

- ・ レジャーエリアの宿泊施設のレビューデータなど記載内容の特徴が異なると思われるレビューデータで、本稿で述べた手法を適用し評価
- ・ 推定の自動化 (ex:ノードの中心性などの指標を活用)

END



共起制限と係り受け構造グラフを利用した 宿泊レビューの特徴表現抽出手法

2017/07/8

Koji Tanaka¹, Koichi Tsujii², Taakashi Ikoma³
Akiyuki Sekiguchi⁴, Kazuhiko Tsuda³

1:Hitachi Government & Public Sector Systems, Ltd

2:Nippon Travel Agency Co. Ltd.

3:Graduate School of Business Sciences University of Tsukuba

4:Meiji University

