

人文社会ビジネス科学学術院 ビジネス科学研究群 2025年度 春C

テキストマイニングの実践

day 4

スケジュール

day 1

- 講義(後半) – 自然言語処理とLLM

day 2

- 講義 – テキストマイニングの手順
- 講義&演習 – データ理解
- 講義&演習 – テキスト解析 (1)

day 3

- 講義&演習 – テキスト解析 (2)
- 講義&演習 – テキスト分析 (1)

day 4

- 講義&演習 – テキスト分析 (2)

day 5

- テキストマイニングツール紹介 – TMS
- ラップアップ – Q&A

(前回) day 3 – レポート課題

- 以下を PDF ファイルで提出 してください
 - ノートブック **day-3-2.ipynb** の末尾にある「【演習】2022~2023 データセット」に従って、別のデータセット (**rakuten-1000-2022-2023.xlsx.zip**) で作図した「共起ネットワーク図」と「対応分析プロット」のキャプチャ

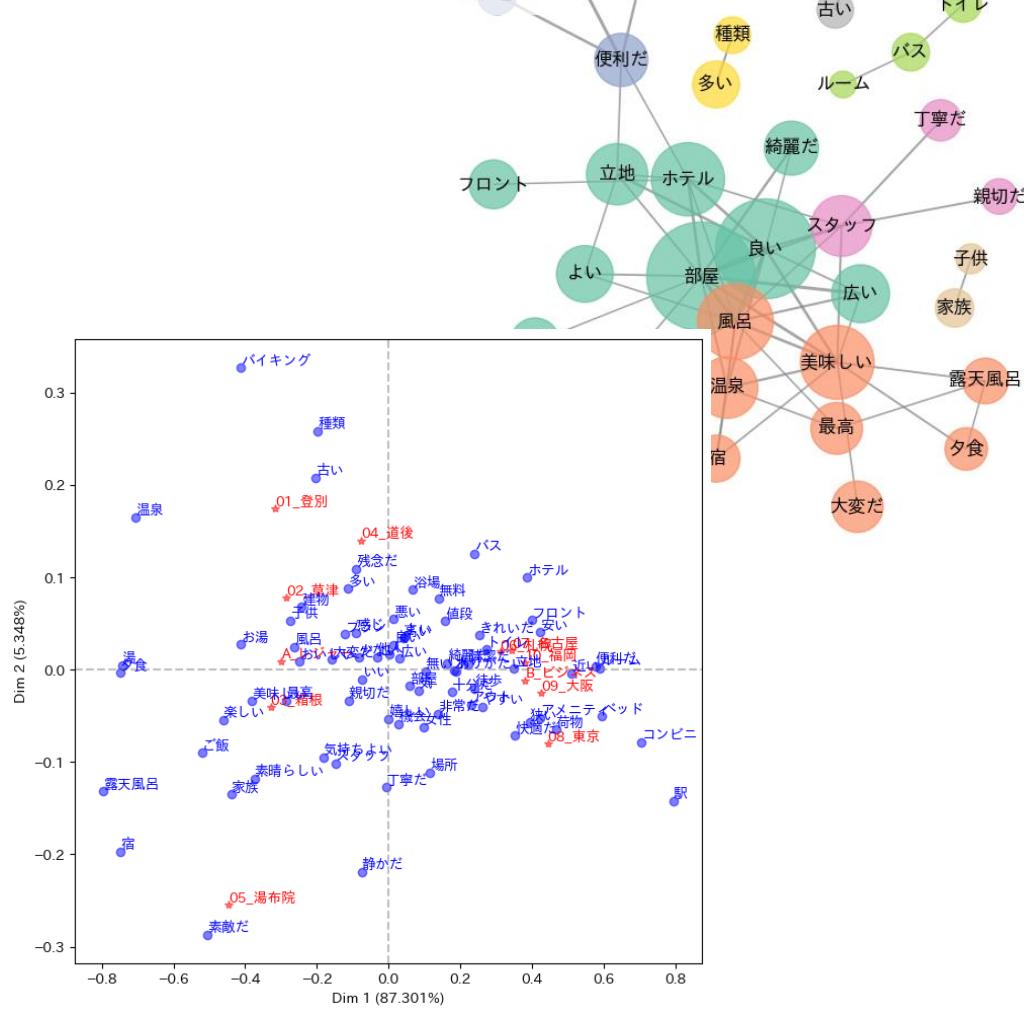
※ 何らかの事情で上記のキャプチャを提出できない場合、本日の講義の感想を文章で記述してください

レポート形式	提出先	期限
PDF	manaba	次回～18:20

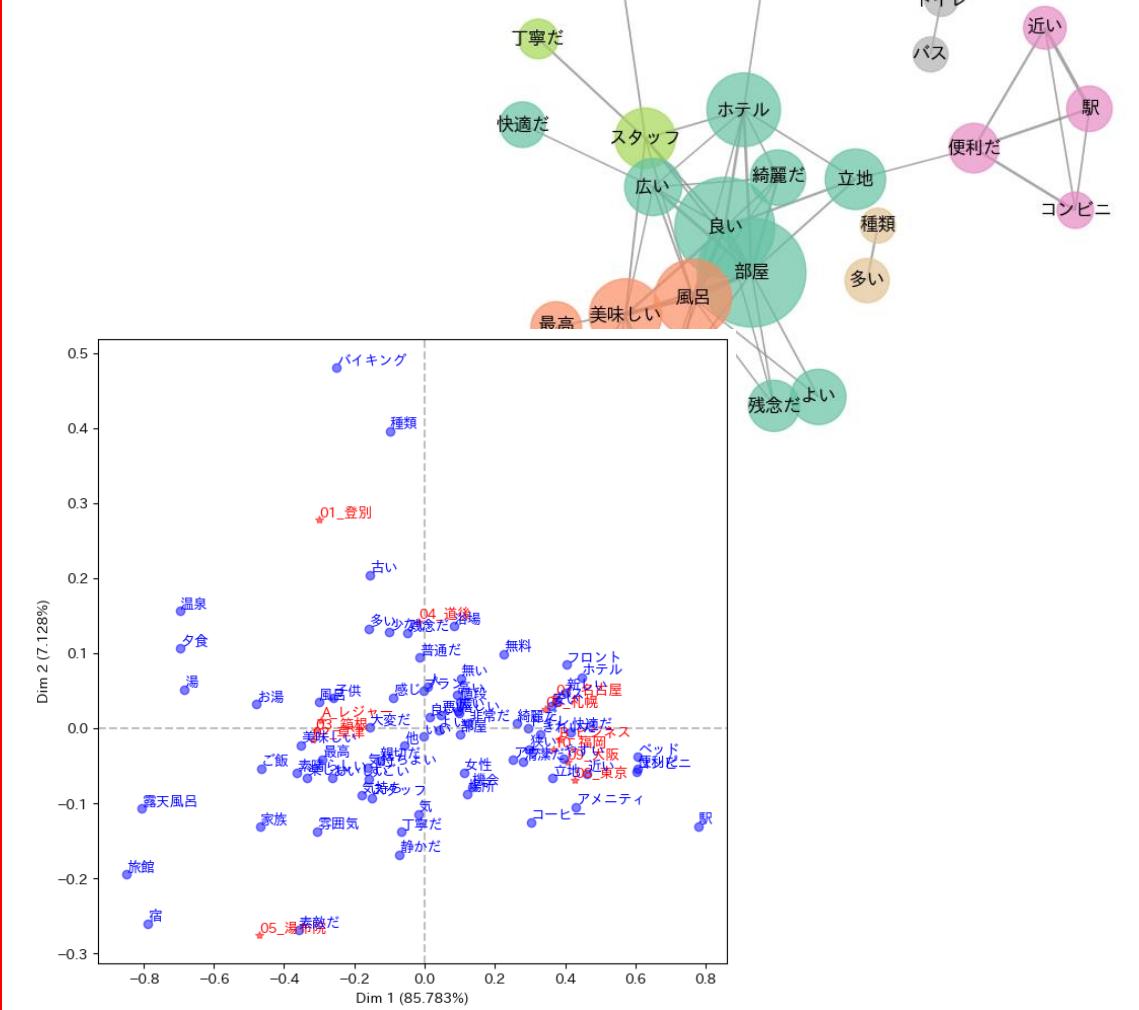
(参考) day3 レポート課題のプロット

コード: day 3-2.ipynb

rakuten-1000- **2024-2025.xlsx.zip**



rakuten-1000- **2022-2023.xlsx.zip**



テキスト分析 (2)

(再掲) テキストマイニングの手順

- データをよく知る
 - データ件数や構成比を集計 → データを理解する
 - 旅行目的別の人気エリアは?
 - 同伴者別の人気エリアは?
 - 数値評価による人気エリアの差異は?
- テーマを設定する
 - 解決すべき課題を決める → 分析目的を明確にする
 - 数値評価が低い原因は?
 - 高評価の施設に学ぶ改善点は?
- テキスト分析に取り組む
 - これら課題を解決するために、テキスト分析を実施

(再掲) 数値評価で違いを見るのは難しい

【再掲】⑧-a 数値評価の平均 (エリア別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂			
■ A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.29	4.50
01_登別	4.07	4.21	3.95	3.90	4.34	4.08	4.16	
02_草津	4.23	4.22	4.07	3.97	4.32	4.20	4.25	
03_箱根	4.24	4.12	4.18	4.05	4.29	4.33	4.26	
04_道後	4.19	4.41	4.07	4.00	4.03	4.19	4.29	
05_湯布院	4.51	4.28	4.37					4.52
■ B_ビジネス	3.98	4.30	4.01					4.13
06_札幌	4.05	4.30	4.09					4.19
07_名古屋	4.00	4.25	4.04	3.89	3.75			4.15
08_東京	3.93	4.38	3.94	3.82	3.70	3.99		4.06
09_大阪	4.01	4.35	4.05	3.93	3.82	4.06		4.18
10_福岡	3.93	4.24	3.86	3.81	3.64	4.01		4.07

- ユーザーの8割が4~5の評価、1~2をつけない→本音が見えない

- 同じ点数でもテキストを見れば差異があるかも

- すべての項目に回答する→どこに注目しているかよくわからない

【再掲】⑧-b 数値評価の平均 (カテゴリ別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13

分析テーマ

- 実践1: カテゴリーやエリアごとの宿泊者の注目ポイントを押さえる
- 実践2: カテゴリーやエリアごとの宿泊者の注目ポイントの評価の違いを見つける
- 実践3: 高評価のエリアに倣って、低評価のエリアを改善するプランを提案する
→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

例) 実践3のまとめ方

対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	・風呂が広い 根拠原文:	・エアコンが臭い 根拠原文:	・... ・...

分析テーマ

- 実践1: カテゴリーやエリアごとの宿泊者の注目ポイントを押さえる
- 実践2: カテゴリーやエリアごとの宿泊者の注目ポイントの評価の違いを見つける
- 実践3: 高評価のエリアに倣って、低評価のエリアを改善するプランを提案する
→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

例) 実践3のまとめ方

対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	• 風呂が広い 根拠原文: ... • ...	• エアコンが臭い 根拠原文: ... • ...	• ... • ...

実践1 — 宿泊者の注目ポイントを押さえる

コード: day-4-2.ipynb

● 特徴語の抽出結果の例 (Jaccard係数 Top10)

A_レジャー		数値評価指標		01_登別		02_草津		03_箱根		04_道後		05_湯布院	
部屋	.329	風呂		温泉	.125	温泉	.136	露天風呂	.137	温泉	.120	宿	.165
良い	.309	部屋		風呂	.106	風呂	.130	美味しい	.131	ホテル	.078	美味しい	.149
美味しい	.273	食事		美味しい	.098	宿	.111	良い	.106	立地	.074	露天風呂	.137
風呂	.260	サービス		良い	.092	美味しい	.106	部屋	.105	よい	.066	温泉	.122
温泉	.253	設備		夕食	.085	良い	.103	温泉	.101	浴場	.061	風呂	.119
スタッフ	.155	立地		残念	.080	部屋	.098	風呂	.101	本館	.057	スタッフ	.116
宿	.148			趣柄	.080	スタッフ	.089	夕食	.097	便利	.055	最高	.107
露天風呂	.145							宿	.091	フロント	.054	部屋	.107
最高	.131							スタッフ	.090	大変	.052	夕食	.105
夕食	.123							大変	.086	いい	.047	家族	.095
B_ビジネス		数値評価指標		06_札幌		07_名古屋		08_東京		09_大阪		10_福岡	
ホテル	.216	風呂		ホテル	.096	ホテル	.088	駅	.123	駅	.099	ホテル	.095
駅	.153	部屋		便利	.083	駅	.070	近い	.083	ホテル	.092	立地	.085
立地	.152	食事		立地	.076	便利	.069	便利	.082	便利	.080	便利	.082
便利	.142	サービス		浴場	.070	立地	.064	コンビニ	.081	立地	.077	駅	.080
近い	.108	設備		駅	.070	フロント	.064	ホテル	.081	近い	.075	近い	.066
フロント	.105	立地		快適	.063	アメニティ	.061	立地	.076	コンビニ	.069	フロント	.066
綺麗	.103			綺麗	.062	綺麗	.059	フロント	.061	フロント	.065	快適	.056
快適	.093			広い	.059	快適	.055	アメニティ	.056	快適	.063	コンビニ	.051
コンビニ	.087			フロント	.058	近い	.055	快適	.054	広い	.063	アメニティ	.051
アメニティ	.074			コンビニ	.054	コンビニ	.047	やすい	.052	綺麗	.059	ベッド	.045

数値評価ではすべての項目に回答
→ エリアによっても注目する項目にかなり偏りがありそう

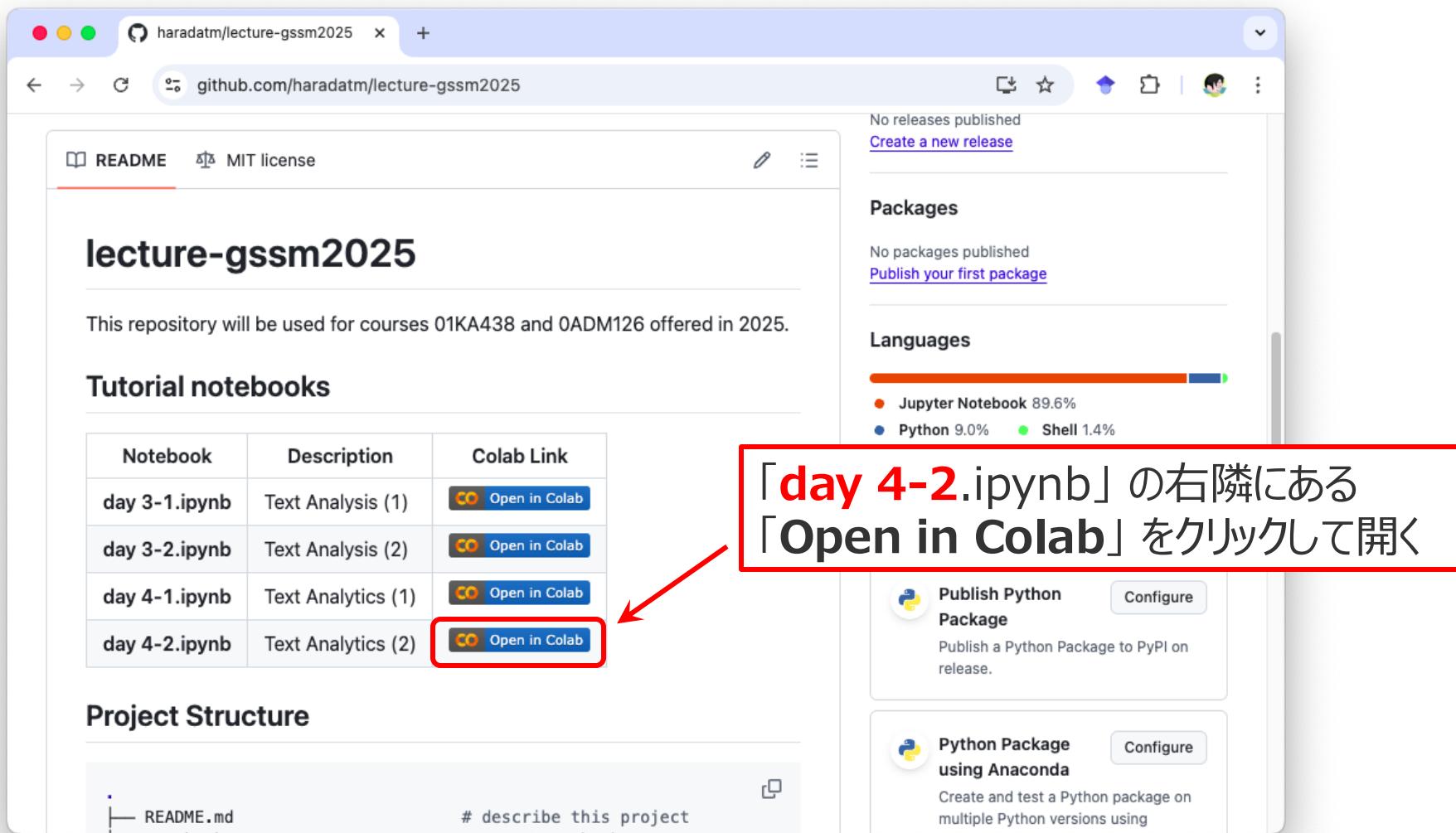
数値評価ではすべての項目に回答
→ レジャーとビジネスでは注目する項目にかなり偏りがありそう

● 特徴語の抽出結果の例 (Jaccard係数 Top10)

A_レジャー		数値評価指標									
		01_登別		02_草津		03_箱根		04_道後		05_湯布院	
部屋	.329	風呂	.125	温泉	.136	露天風呂	.137	温泉	.120	宿	.165
良い	.309	部屋	.106	風呂	.130	美味しい	.131	ホテル	.078	美味しい	.149
美味しい	.273	食事	.098	宿	.111	良い	.106	立地	.074	露天風呂	.137
風呂	.260	サービス	.092	美味しい	.106	部屋	.105	よい	.066	温泉	.122
温泉	.253	設備	.085	良い	.103	温泉	.101	浴場	.061	風呂	.119
スタッフ	.155	立地	.080	部屋	.098	風呂	.101	本館	.057	スタッフ	.116
宿	.148		.080	スタッフ	.089	夕食	.097	便利	.055	最高	.107
露天風呂	.145		.075	湯	.086	宿	.091	フロント	.054	部屋	.107
最高	.131		.074	最高	.084	スタッフ	.090	大変	.052	夕食	.105
夕食	.123		.071	バス	.081	大変	.086	いい	.047	家族	.095
B_ビジネス		数値評価指標									
		06_札幌		07_名古屋		08_東京		09_大阪		10_福岡	
ホテル	.216	風呂	.096	ホテル	.088	駅	.123	駅	.099	ホテル	.095
駅	.153	部屋	.083	駅	.070	近い	.083	ホテル	.092	立地	.085
立地	.152	食事	.076	便利	.069	便利	.082	便利	.080	便利	.082
便利	.142	サービス	.070	浴場	.064	コンビニ	.081	立地	.077	駅	.080
近い	.108	設備	.070	駅	.064	ホテル	.081	近い	.075	近い	.066
フロント	.105	立地	.063	フロント	.061	立地	.076	コンビニ	.069	フロント	.066
綺麗	.103		.062	綺麗	.059	フロント	.061	フロント	.065	快適	.056
快適	.093		.059	快適	.055	アメニティ	.056	快適	.063	コンビニ	.051
コンビニ	.087		.058	フロント	.055	快適	.054	広い	.063	アメニティ	.051
アメニティ	.074		.054	コンビニ	.047	やすい	.052	綺麗	.059	ベッド	.045

演習用の教材 – Google Colab ノートブック

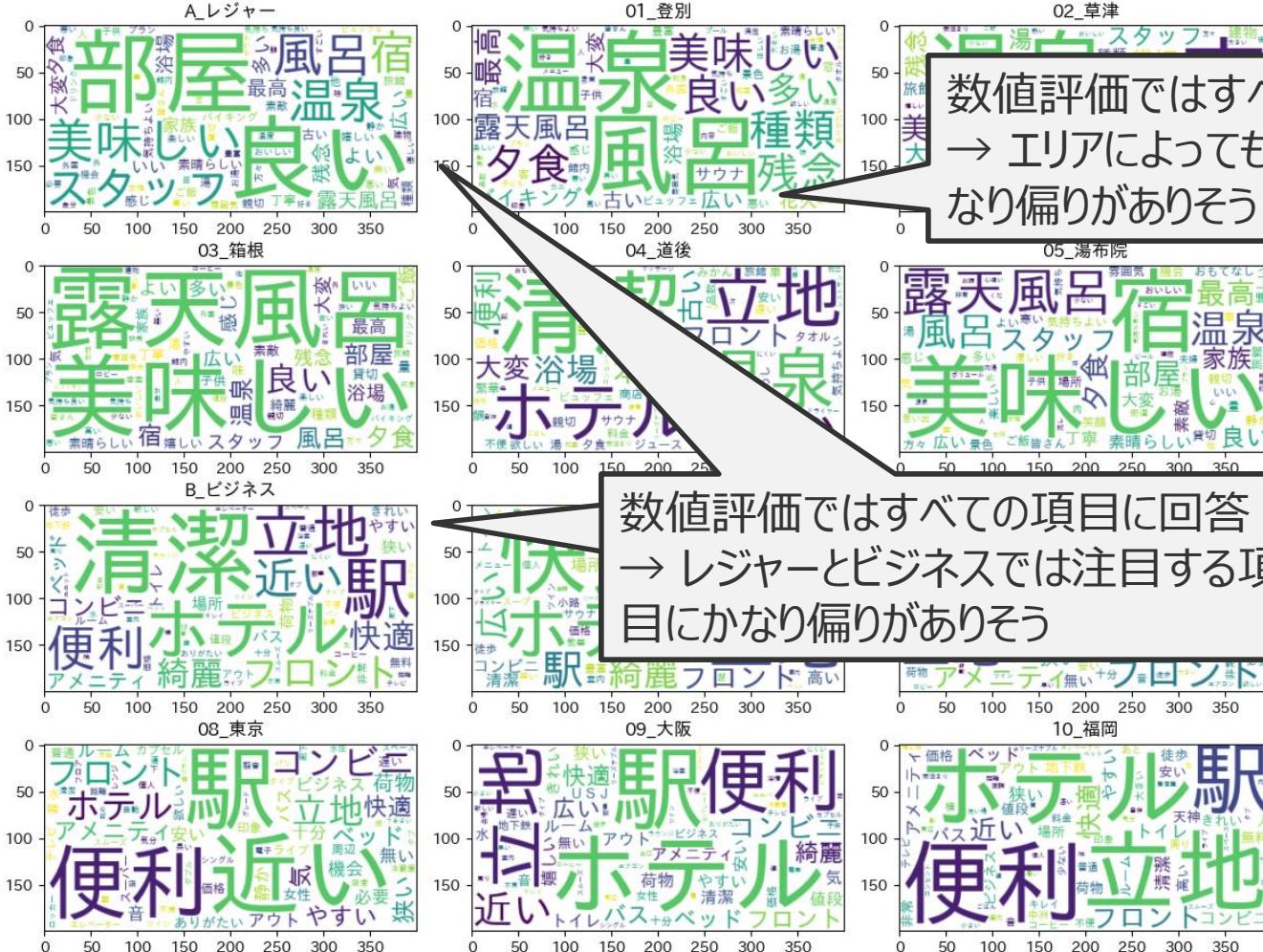
- URL: <https://github.com/haradatm/lecture-gssm2025>



実践1 — 宿泊者の注目ポイントを押さえる

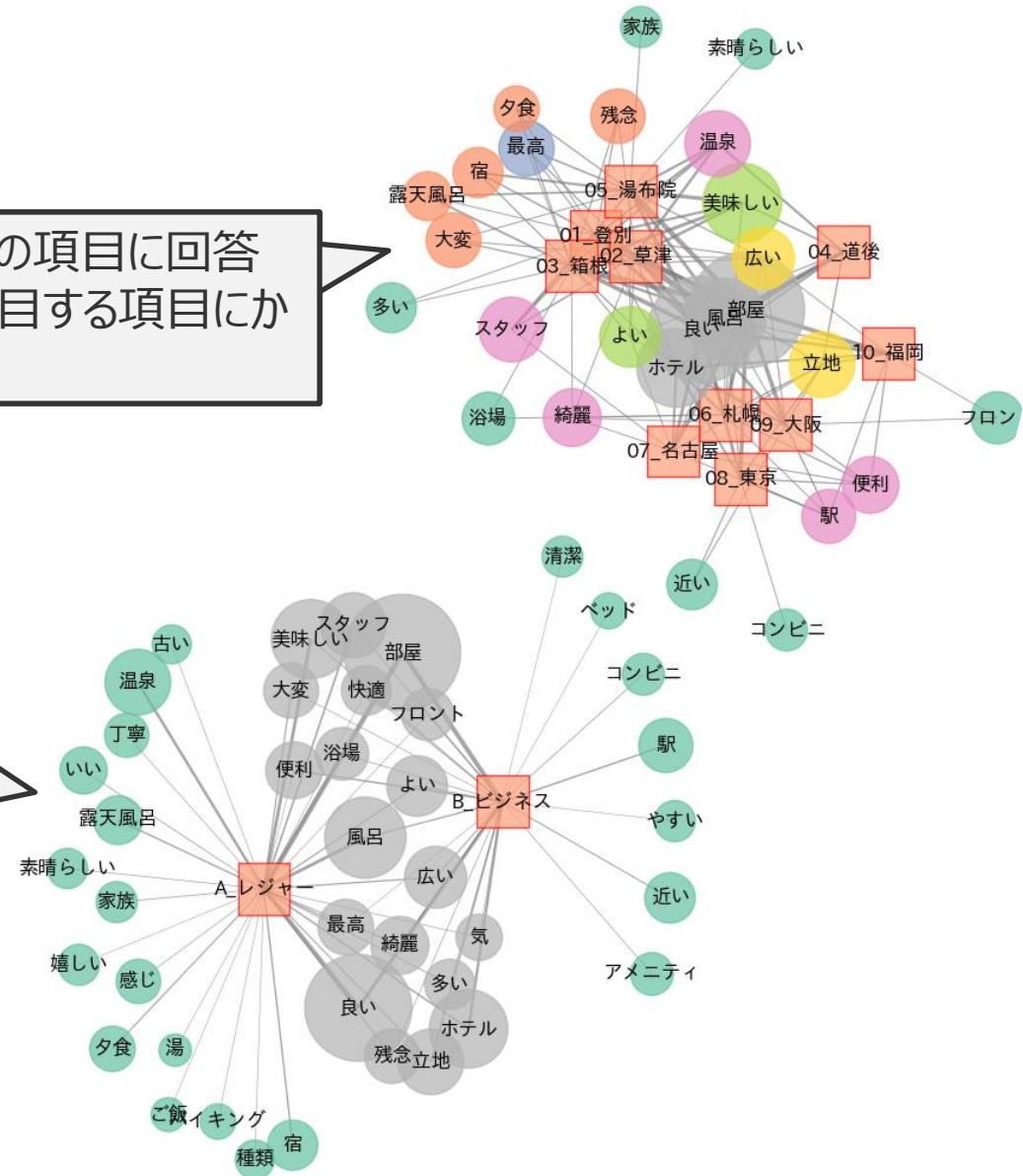
コード: day-4-2.ipynb

● 特徴語の抽出結果の例 (その他の可視化)

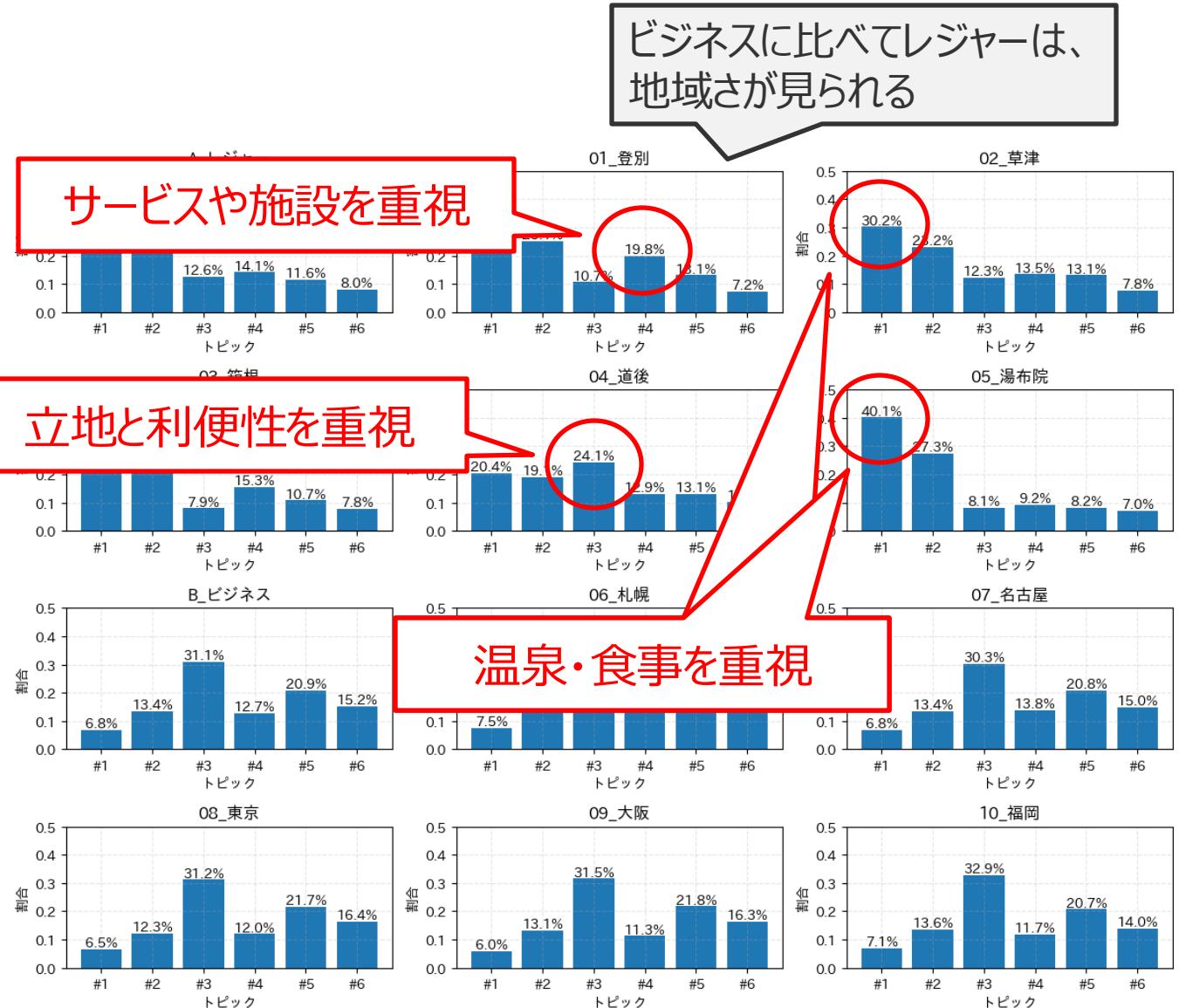
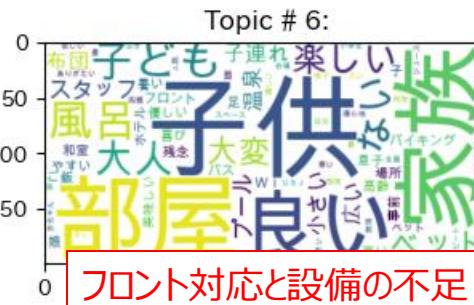
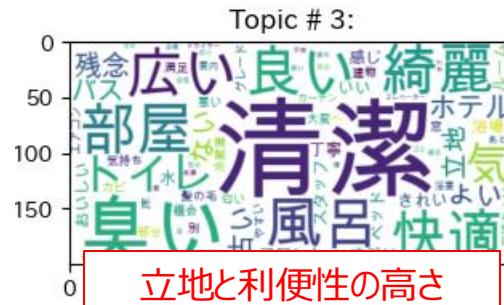
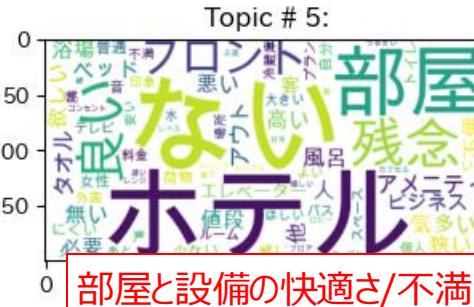
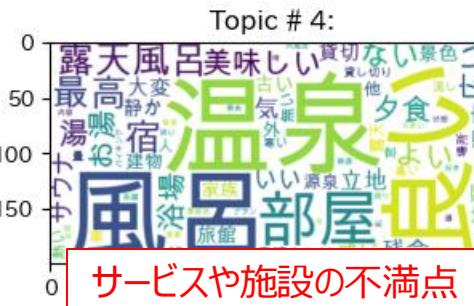
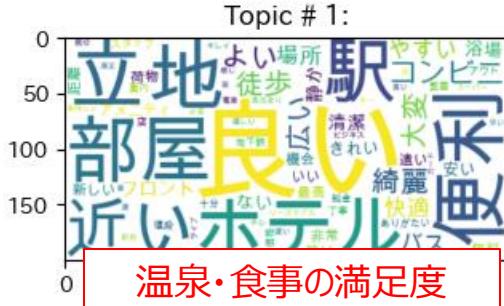


数値評価ではすべての項目に回答
→ エリアによっても注目する項目にか
なり偏りがありそう

数値評価ではすべての項目に回答
→ レジャーとビジネスでは注目する項目にかなり偏りがありそう



● トピック抽出結果とトピック割合



分析テーマ

- 実践1: カテゴリーやエリアごとの宿泊者の注目ポイントを押さえる
- 実践2: カテゴリーやエリアごとの宿泊者の注目ポイントの評価の違いを見つける
- 実践3: 高評価のエリアに倣って、低評価のエリアを改善するプランを提案する
→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

例) 実践3のまとめ方

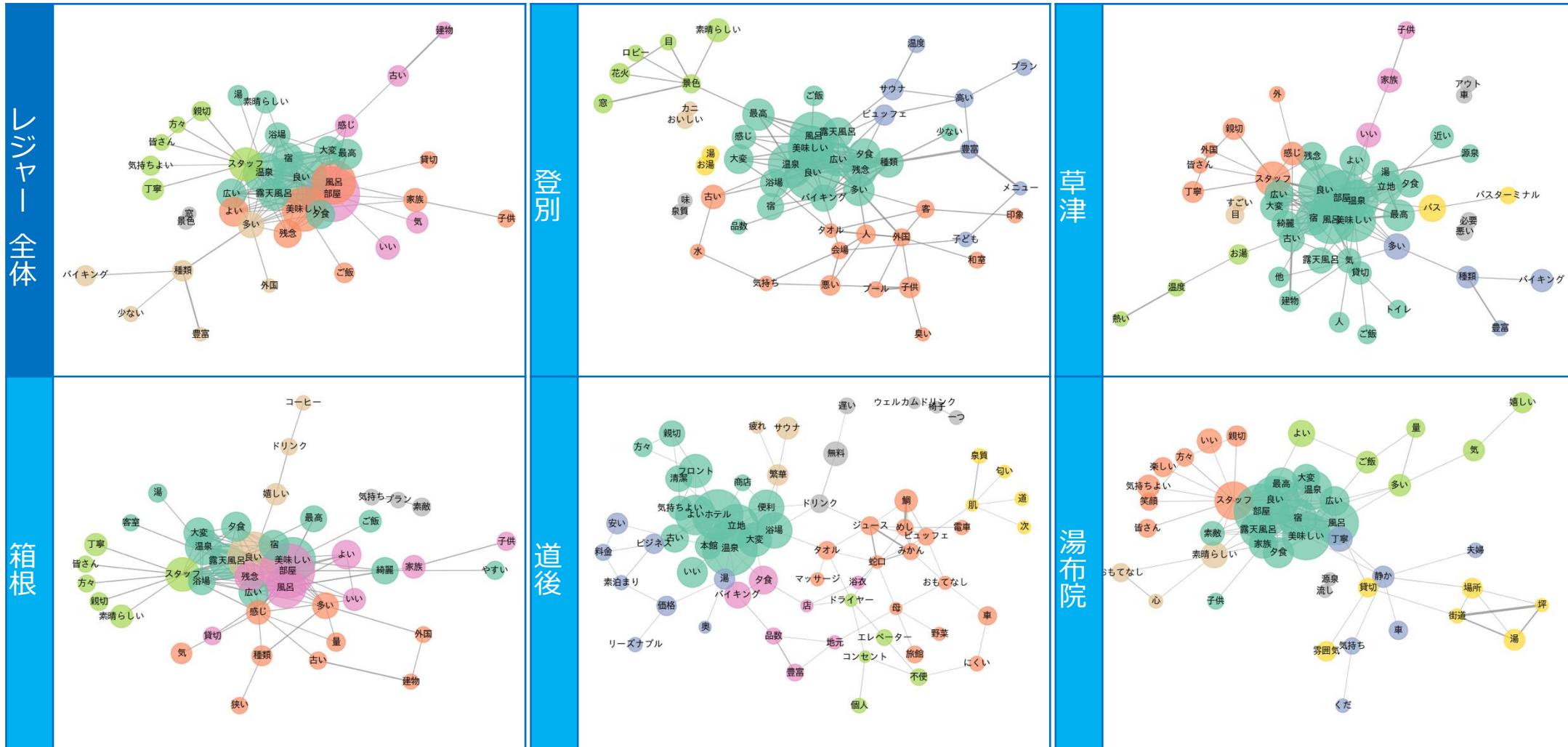
対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	• 風呂が広い 根拠原文: ... • ...	• エアコンが臭い 根拠原文: ... • ...	• ... • ...

実践2 – ユーザー注目ポイントの評価を見る

- カテゴリーやエリアごとでの注目する観点の評価の違いを確認する
 - カテゴリー「レジャー」と「ビジネス」を比較する
 - カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する
- 手順の一例:
 - カテゴリーやエリアごとの共起NWから、どの観点をどう評価しているか調べる

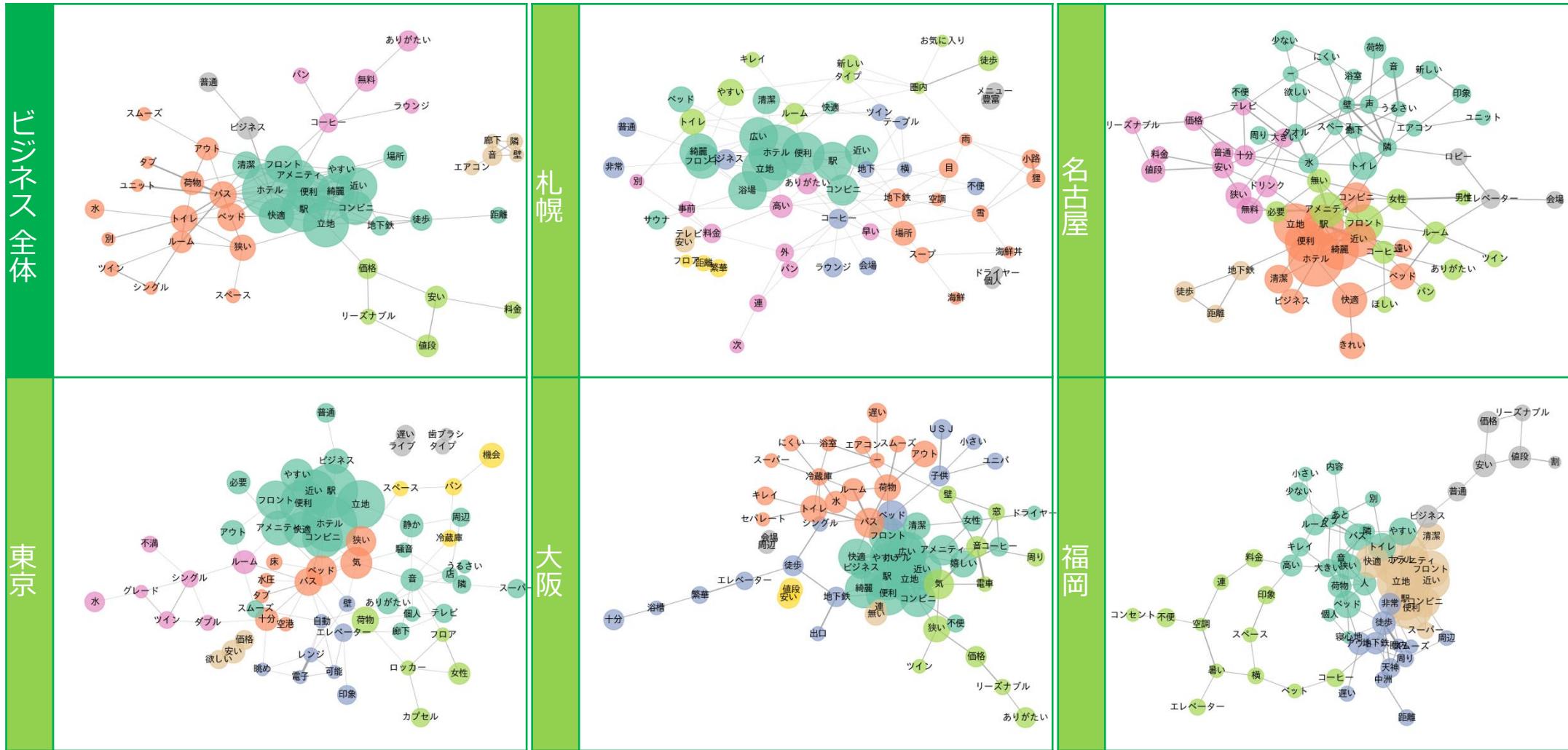
実践2 – ユーザー注目ポイントの評価を見る

- カテゴリー「レジャー」におけるエリアごとの共起ネットワーク



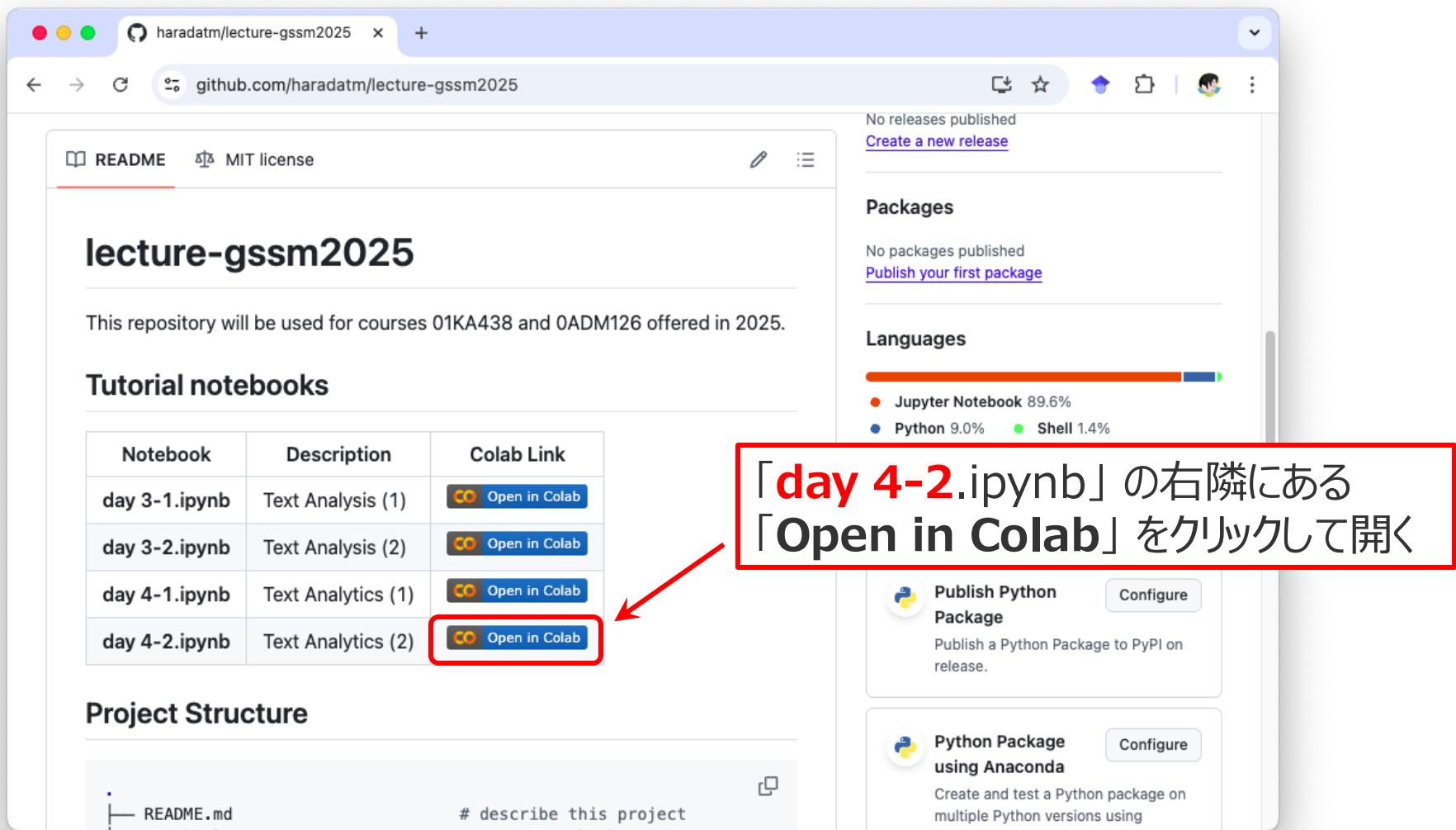
実践2 – ユーザー注目ポイントの評価を見る

- カテゴリー「ビジネス」とエリアごとの共起ネットワーク



演習用の教材 – Google Colab ノートブック

- URL: <https://github.com/haradatm/lecture-gssm2025>

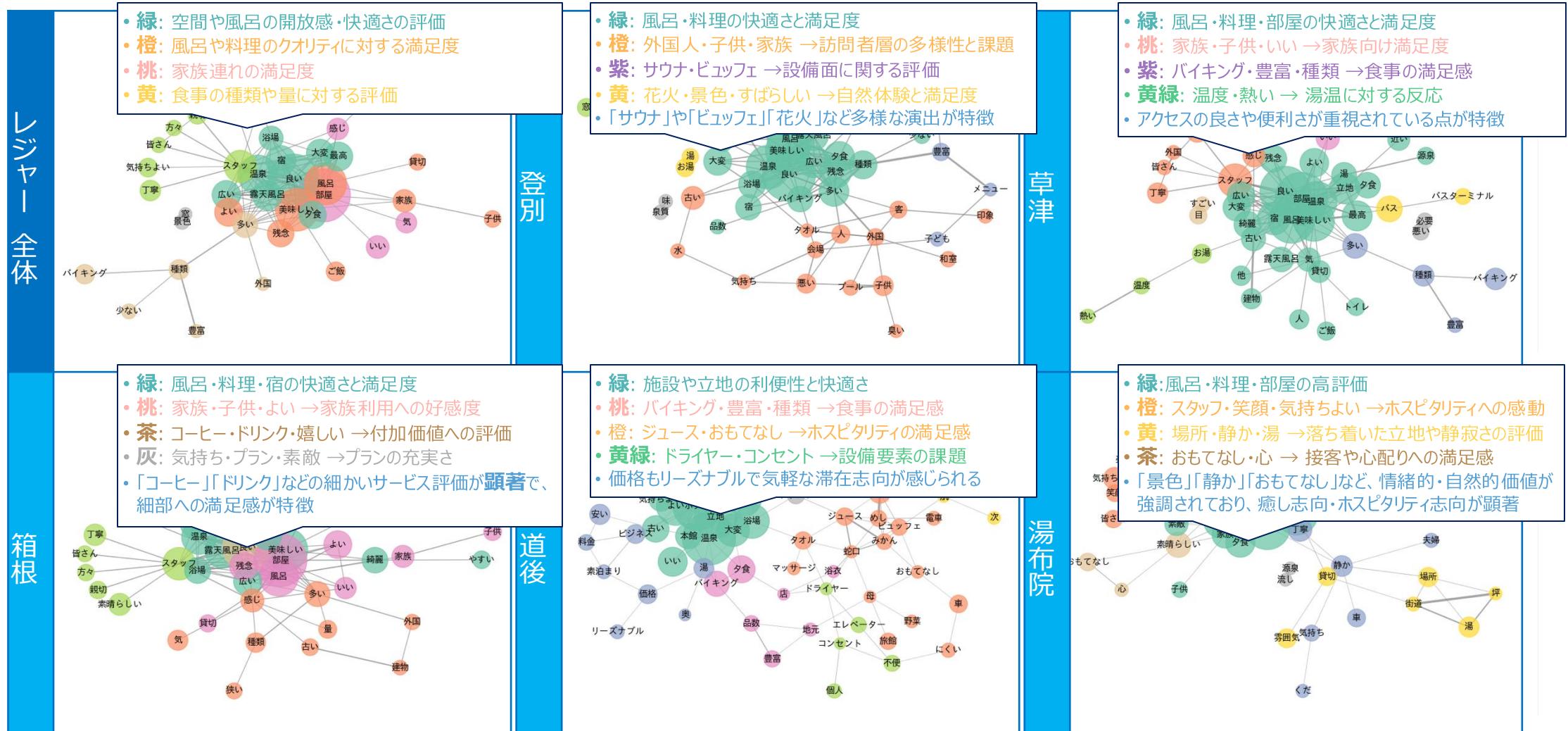


演習 – 評価の違いを確認する

- 個人ワーク (~19:40 ※休憩込み)
 - カテゴリー「レジャー」or「ビジネス」について、作成した共起ネットワーク図をもとにエリアごとの**宿泊者の注目ポイントの評価の違い**を見つけてください

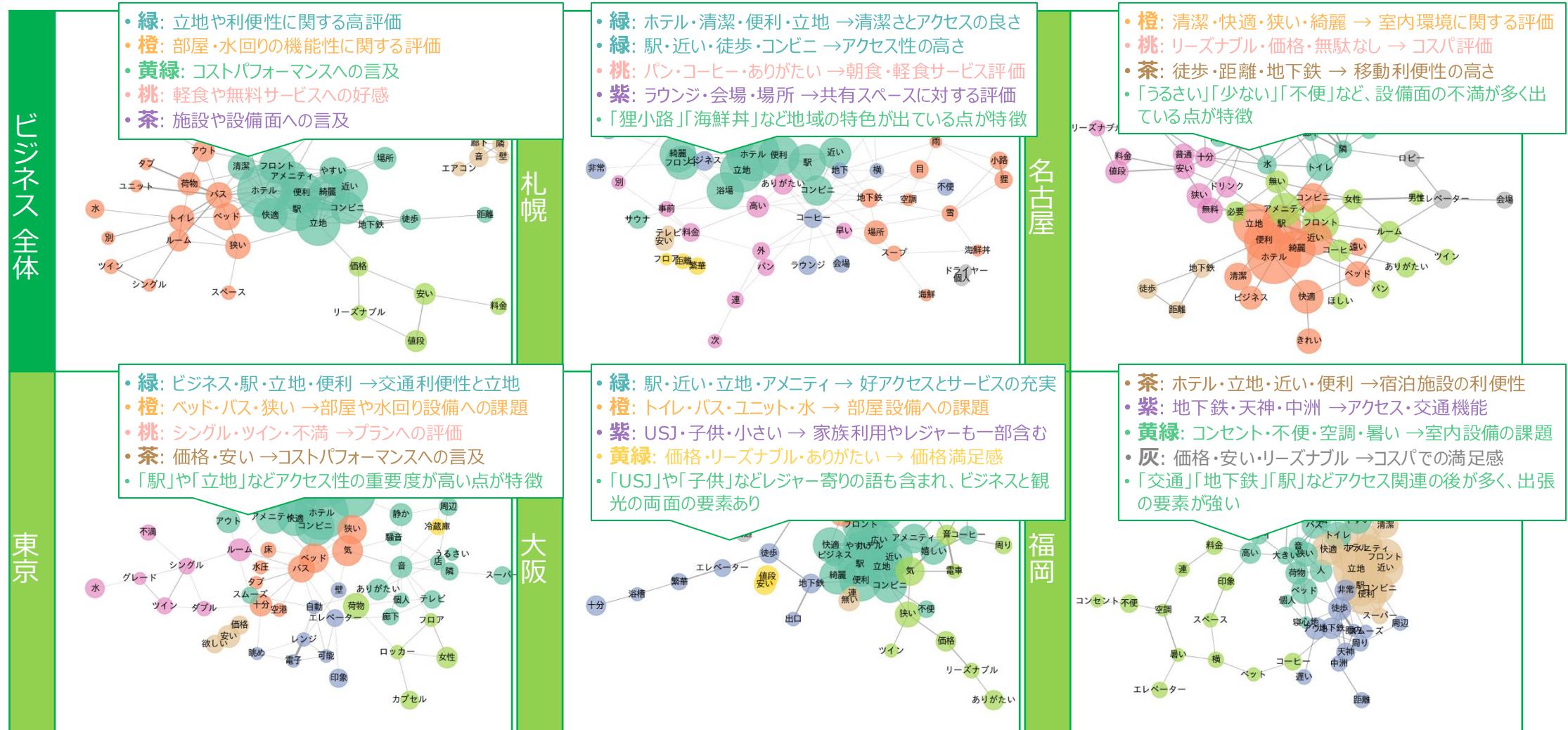
(例) 実践2 – ユーザー注目ポイントの評価を見る

● カテゴリー「レジャー」におけるエリアごとの共起ネットワーク



(例) 実践2 – ユーザー注目ポイントの評価を見る

● カテゴリー「ビジネス」とエリアごとの共起ネットワーク



(再掲) 分析テーマ

- 実践1: カテゴリーやエリアごとの宿泊者の注目ポイントを押さえる
- 実践2: カテゴリーやエリアごとの宿泊者の注目ポイントの評価の違いを見つける
- 実践3: 高評価のエリアに倣って、低評価のエリアを改善するプランを提案する
→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

例) 実践3のまとめ方

対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	・風呂が広い 根拠原文:	・エアコンが臭い 根拠原文:	・... ・...

実践3 – 改善プランを提案する

- カテゴリーやエリアごとのポジ/ネガ意見の違いを確認する
 - 対照的な2エリアを選択する
 - 選択したエリアと、そのカテゴリー（「レジャー」or「ビジネス」）を比較する
- 方法の例：
 - カテゴリーやエリアごとのコーディングや数値評価から、対照的な2エリアを選ぶ

実践3 – 改善プランを提案する

●コーディングルール(KH Coder用語)を作成する

単語のまとめ上げ

ポジティブワード

```
coding_pos = ["良い","美味しい","広い","多い","素晴らしい","嬉しい","気持ちよい","楽しい","近い","大きい","気持ち良い","温かい","早い","優しい","新しい","暖かい","快い","明るい","美しい","可愛い","満足"]
```

ネガティブワード

```
coding_neg = ["古い","無い","高い","悪い","小さい","狭い","少ない","寒い","遅い","熱い","欲しい","暑い","冷たい","遠い","臭い","暗い","うるさい","ない","無い","残念","改善","不満"]
```

外部変数のまとめ上げ

- 数値評価「総合」の値が 1 or 2 → 「総合1~2」
- 数値評価「総合」の値が 4 or 5 → 「総合4~5」

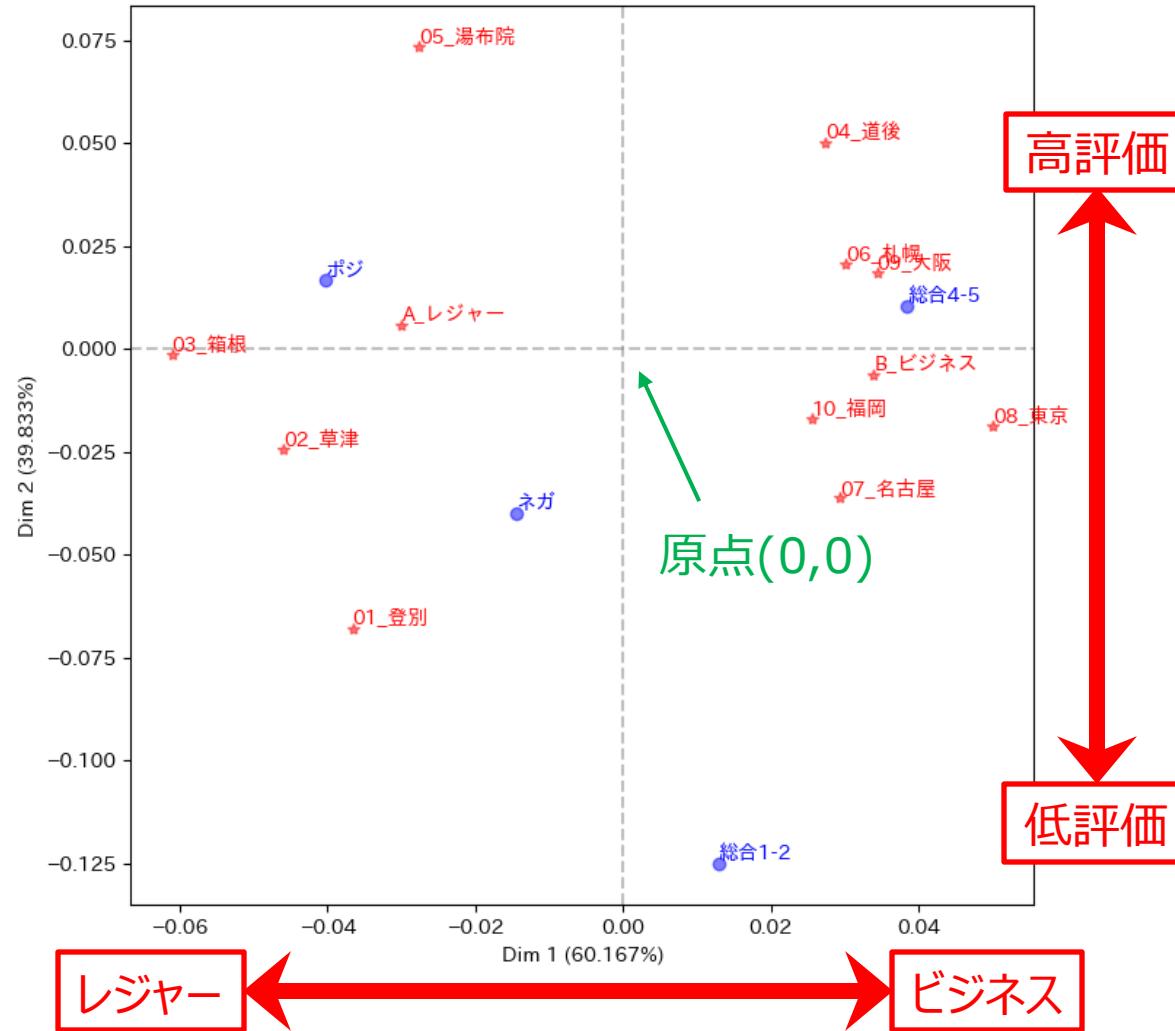


		表層	ポジ	ネガ	総合1-2	総合4-5
カテゴリ	エリア	文書ID				
A_レジャー	01_登別	1	1	0	0	1
		2	0	1	0	1
		3	1	1	0	1
		4	1	0	0	1
		5	1	0	0	1
...	
B_ビジネス	10_福岡	9996	1	1	0	1
		9997	0	0	0	1
		9998	1	1	0	1
		9999	1	0	0	1
		10000	1	0	0	0

9880 rows × 4 columns

実践3 – 改善プランを提案する

- 対照的な2エリアを選択する(出力例)



見方：

- 原点付近は特徴がない
 - 原点から遠い方が特徴的 (以下は仮説であって、原文で確認すること)
 - 湯布院は圧倒的に高評価?
 - 登別は圧倒的に低評価か?
 - 道後はレジャーよりもビジネスに近い (出張で行く場所か)
 - 箱根は圧倒的にレジャー
 - 第2固有値までの累積寄与率は $60.17 + 39.83 = 100\%$ で非常に高く、これら2軸でデータをよく表現できている

実践3 – 改善プランを提案する

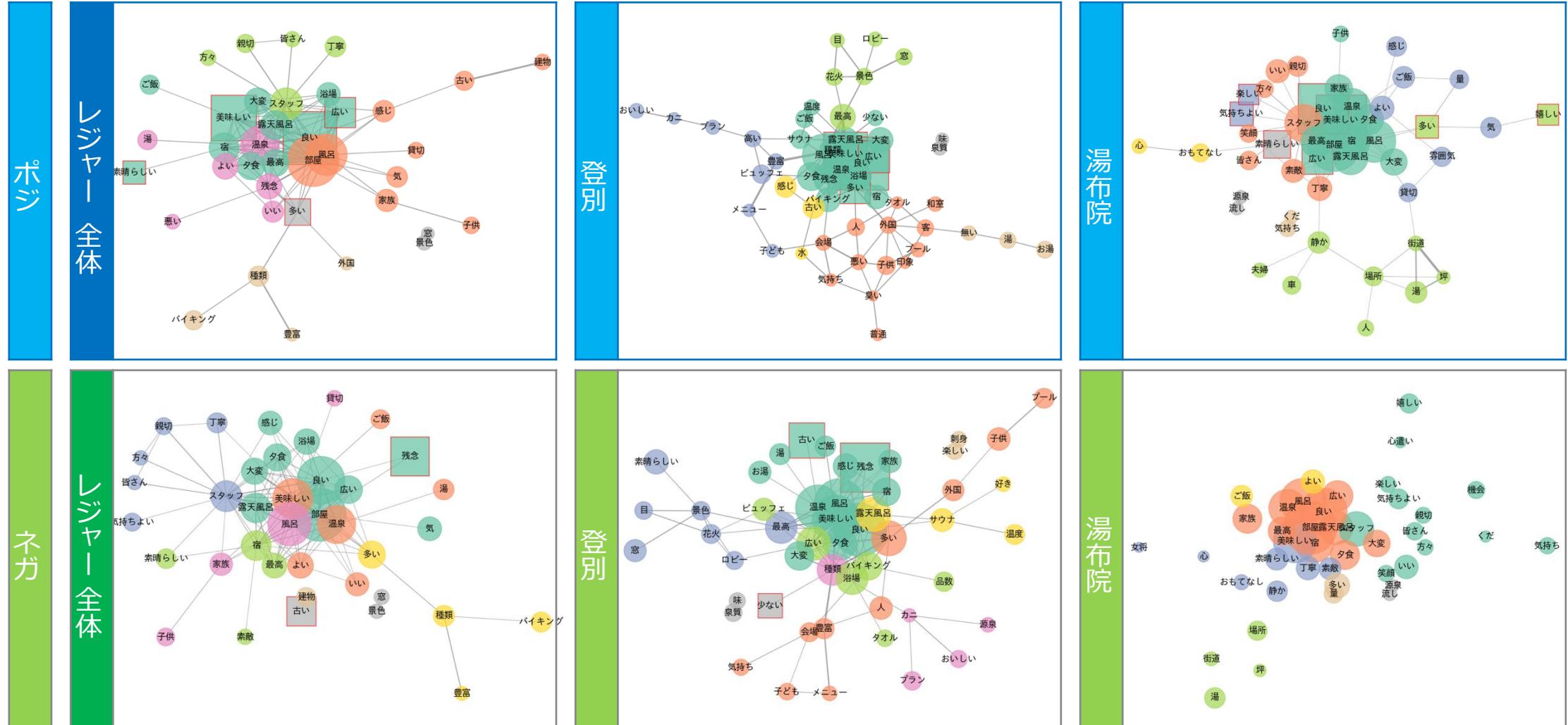
- カテゴリーやエリアごとのポジティブ意見の違いを確認する
 - 対照的な2エリアを選択する
 - 選択したエリアと、そのカテゴリー（「レジャー」or「ビジネス」）を比較する
- 方法の例：
 - カテゴリーやエリアごとの共起NWから、何を高(好)評価しているかを調べる

実践3 – 改善プランを提案する

- カテゴリーやエリアごとのネガティブ意見の違いを確認する
 - 対照的な2エリアを選択する
 - 選択したエリアと、そのカテゴリー（「レジャー」or「ビジネス」）を比較する
- 方法の例：
 - カテゴリーやエリアごとの共起NWから、何を低(悪)評価しているかを調べる

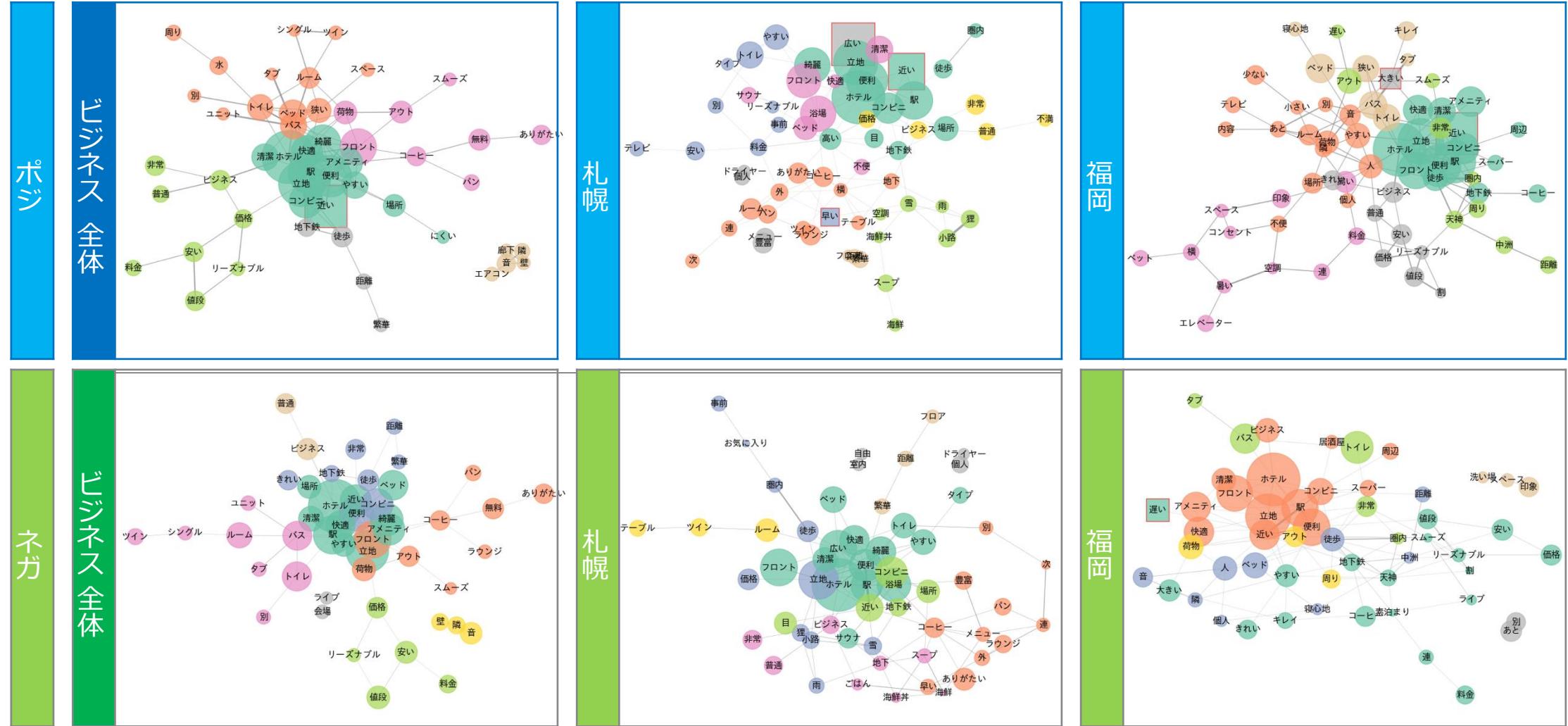
実践3 – 改善プランを提案する

● 例：登別と湯布院のポジネガ比較



実践3 – 改善プランを提案する

● 例：札幌と福岡のポジネガ比較



(参考) ChatGPT を利用した読み解き

● ただし、間違いも多くチェックは必要（演習は人力で読み解いてください）

プロットをキャプチャーして、ChatGPTに貼り付け（できるだけ図中の文字が見える大きさでキャプチャーを取るのがベター）

プロットの色ごとに読み解き

実践3 – 改善プランを提案する

● 結果の整理

- 主張を支持するプロットと、ユーザーの生の声(原文)を使って解釈する
 - エリア X が評価されている点は何か?
 - エリア Y の課題は何か?
 - エリア Y の改善に向けた提案?

例)

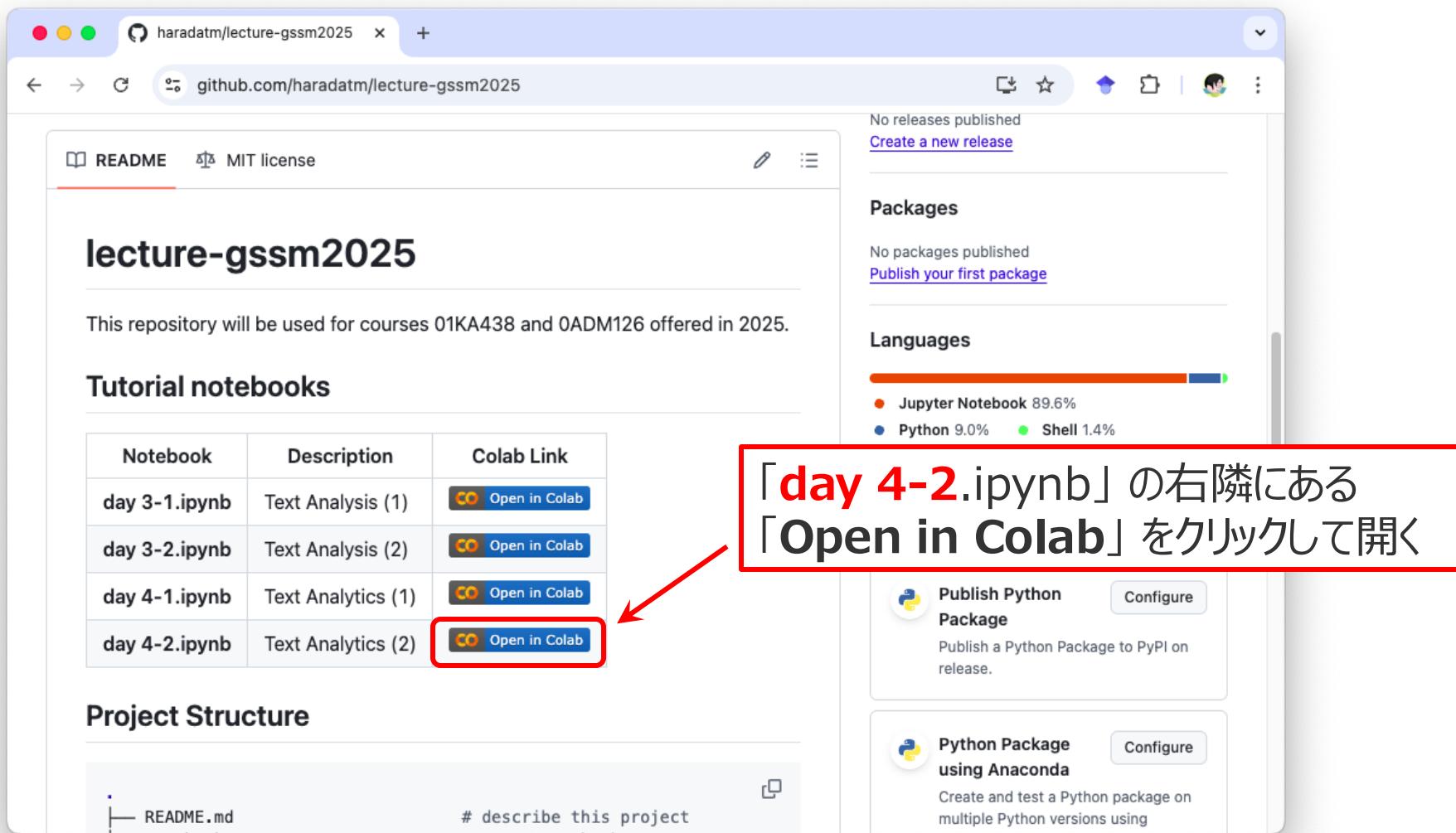
対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	• 風呂が広い 根拠原文: ... • ...	• エアコンが臭い 根拠原文: ... • ...	• ... • ...

演習 — 改善プランを提案する

- 特徴語とポジティブ意見の共起ネットワーク図を作成し、エリアによってポジティブ意見(とその背景)がどう異なるかを比較することで、何がどう評価されているかを確認する
(→P.XXX,XXX)
- 特徴語とネガティブ意見の共起ネットワーク図を作成し、エリアによってネガティブ意見(とその背景)がどう異なるかを比較することで、何がどう評価されているかを確認する
(→P.XXX,XXX)
- 高評価エリアに倣って、低評価エリアを改善プランを提案する (→P.XXX)
→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

演習用の教材 – Google Colab ノートブック

- URL: <https://github.com/haradatm/lecture-gssm2025>



演習 — 改善プランを提案する

● 個人ワーク (~20:50)

- ・ 比較対象の2エリアを選択する
- ・ 共起ネットワーク図を作成し、何がどう評価されているかを確認する
- ・ 高評価のエリアに倣って、低評価のエリアを改善するプランを提案する

→ 注意: プロットによる可視化と宿泊客の生の声(原文)を使って解釈する

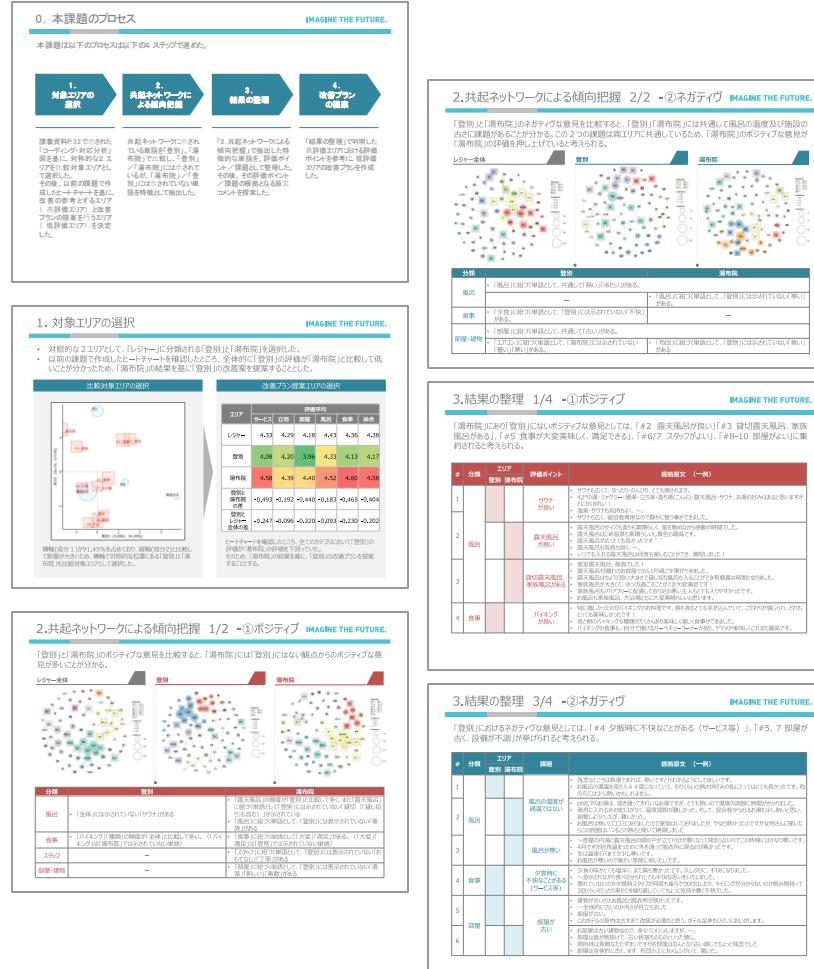
day 4 – レポート課題

- 以下を PDF ファイルで提出 してください
 - 作成した2エリアの「ポジネガ意見の共起ネットワーク図」のキャプチャ (P.XXX)
 - 作成した「結果の整理」の表(P.XXX)を用いて、選択したエリアの改善プランを提案してください
- ※ 何らかの事情で上記2つを提出できない場合、本日の講義の感想を文章で記述してください

レポート形式	提出先	期限
PDF	manaba	8/8～23:59

(参考) 分析レポートの例

● R04年度 受講者のレポート



4. 改善プランの提案

- 「部屋が古い」といったネガティブ意見が多く見受けられるが、施設そのものの改修には多額のコストかかるため、まずはコストがそれほどからないと思われる内容から改善していくことを提案する。
- 具体的には、「1. スタッフの再教育」、「2. 貸切風呂の導入」、「3. 一部設備の新調」、「4. コンセプトの見直し」が挙げられる。

「登別」の現状

1. スタッフの再教育

「湯布院」は「スタッフがよい」傾向が見られたが、「登別」ではその傾向が見られなかった。また、ネガティブ意見においては「夕食時のスタッフの対応が不快」の傾向が見受けられた。

2. 貸切風呂の導入

「湯布院」には「貸切露天風呂、家族風呂がある」といったポジティブ意見が見受けられたが、「登別」にはその傾向は見られなかった。

3. 一部設備の新調

「空調が適切ではない」といった傾向が見られた。

4. コンセプトの見直し

「部屋が古い」といった傾向が見られた。また、「登別」には、「湯布院」にはあった「部屋がよい」というポジティブ意見は見受けられなかった。

IMAGINE THE FUTURE.

改善案

スタッフの教育を充実させる。また、サービスの良いスタッフに対するインセンティブを発生させる等、サービス向上のモチベーションになるような人事制度にする。

例：接客マニュアルの見直し、人事制度（インセンティブ、昇給、昇格等）の改革 等

温泉設備の新規導入には多額の費用がかかるため、既存の浴場を時間帯によって貸切にするなど、貸切風呂の導入を行う。

空調設備等の経年劣化がある設備の新調を行う。

施設のコンセプトを見直し、「古さ」を生かした内装とする。
例：「古民家」「大正モダン」等をコンセプトにし、備え付け家具等のインテリアの見直す