

人文社会ビジネス科学学術院 ビジネス科学研究群 2023年度 春C

# テキストマイニング day 5

# スケジュール

## day 1

- 講義 — テキストマイニング概説 (津田先生)
- 講義 — 自然言語処理の最新動向

## day 2

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

## day 3

- 演習 — テキスト解析 (2)
- 講義&演習 — データ分析 (使い方編)

## day 4

- TextMining Studio の紹介
- 講義&講義 — データ分析 (実践編)

## day 5

- 講義&演習 — データ分析 (実践編)

# テキスト分析（実践編）

(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x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	4.22	4.28	4.11	4.01	4.29
01_登別	4.03	4.27	3.95	3.88	4.31
02_草津	4.19	4.28	4.03	3.92	4.31
03_箱根	4.22	4.15	4.12	3.97	4.22
04_道後	4.16	4.41	4.10	4.00	4.09
05_湯布院	4.52	4.28	4.36		4.21
■B_ビジネス	4.00	4.34	4.10		4.55
06_札幌	3.99	4.37	4.09		4.19
07_名古屋	3.98	4.26	4.06	3.92	4.20
08_東京	3.97	4.34	4.11	3.91	4.16
09_大阪	4.06	4.34	4.14	3.96	4.14
10_福岡	4.01	4.40		3.95	4.24

• 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x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	4.22	4.28	4.11	4.01	4.29	4.26	4.28
B_ビジネス	4.00	4.34	4.10	3.92	3.82	4.06	4.19

# (再掲) 実践的な分析

- 実践1: カテゴリーやエリアごとの**ユーザーの注目ポイント**を押さえる
- 実践2: カテゴリーやエリアごとの**ユーザー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

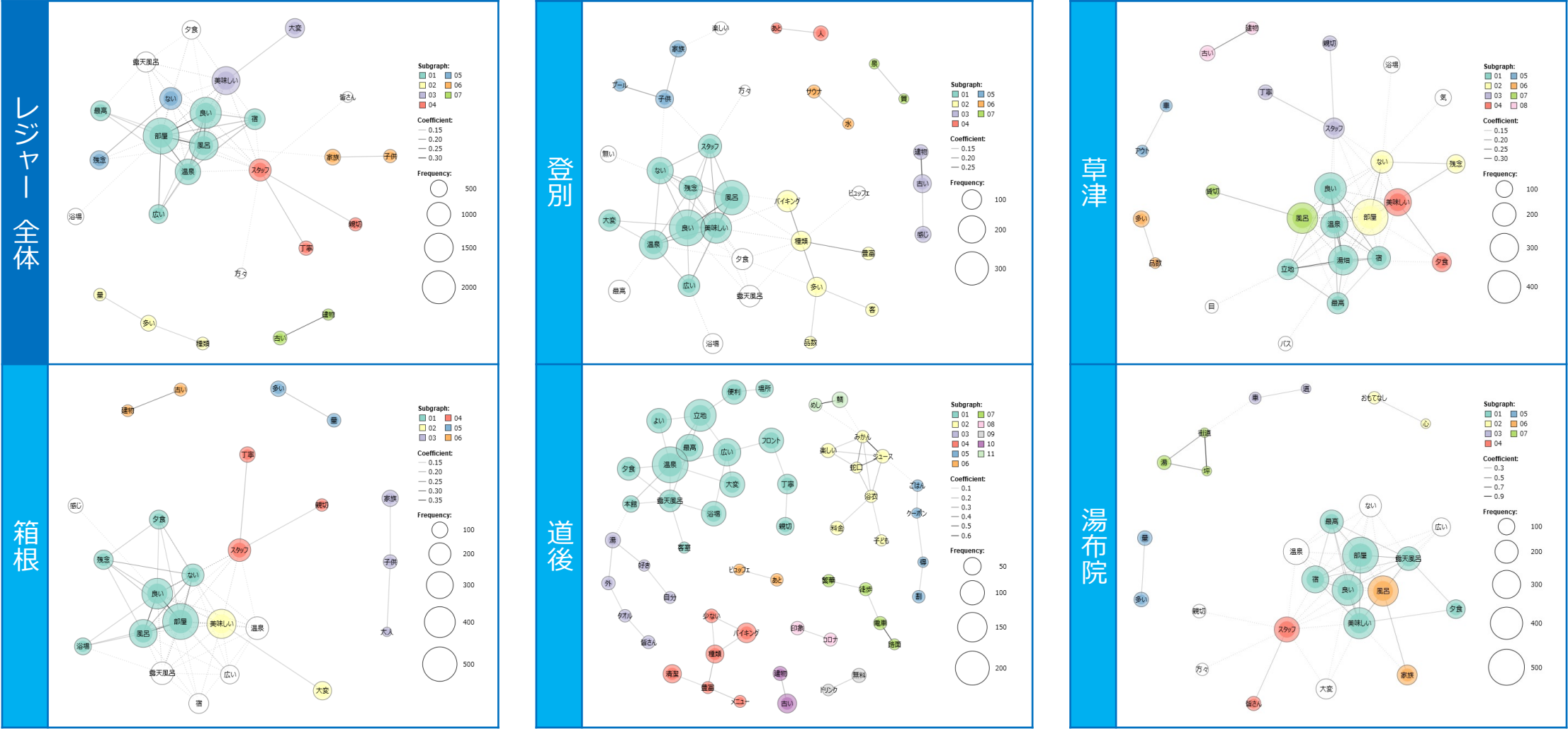
- カテゴリーやエリアごとの注目する観点の評価の違いを確認する
  - ・ カテゴリー「レジャー」と「ビジネス」を比較する
  - ・ カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する
- 手順の一例:
  - ・ カテゴリーやエリアごとの共起NWから、どの観点をどう評価しているか調べる

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]“<>カテゴリー-->A\_レジャー”」  
「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,タグ,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位60,共起関係ほど濃い線に」

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]“<>エリア-->01\_登別”」  
「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,タグ,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位60,共起関係ほど濃い線に」

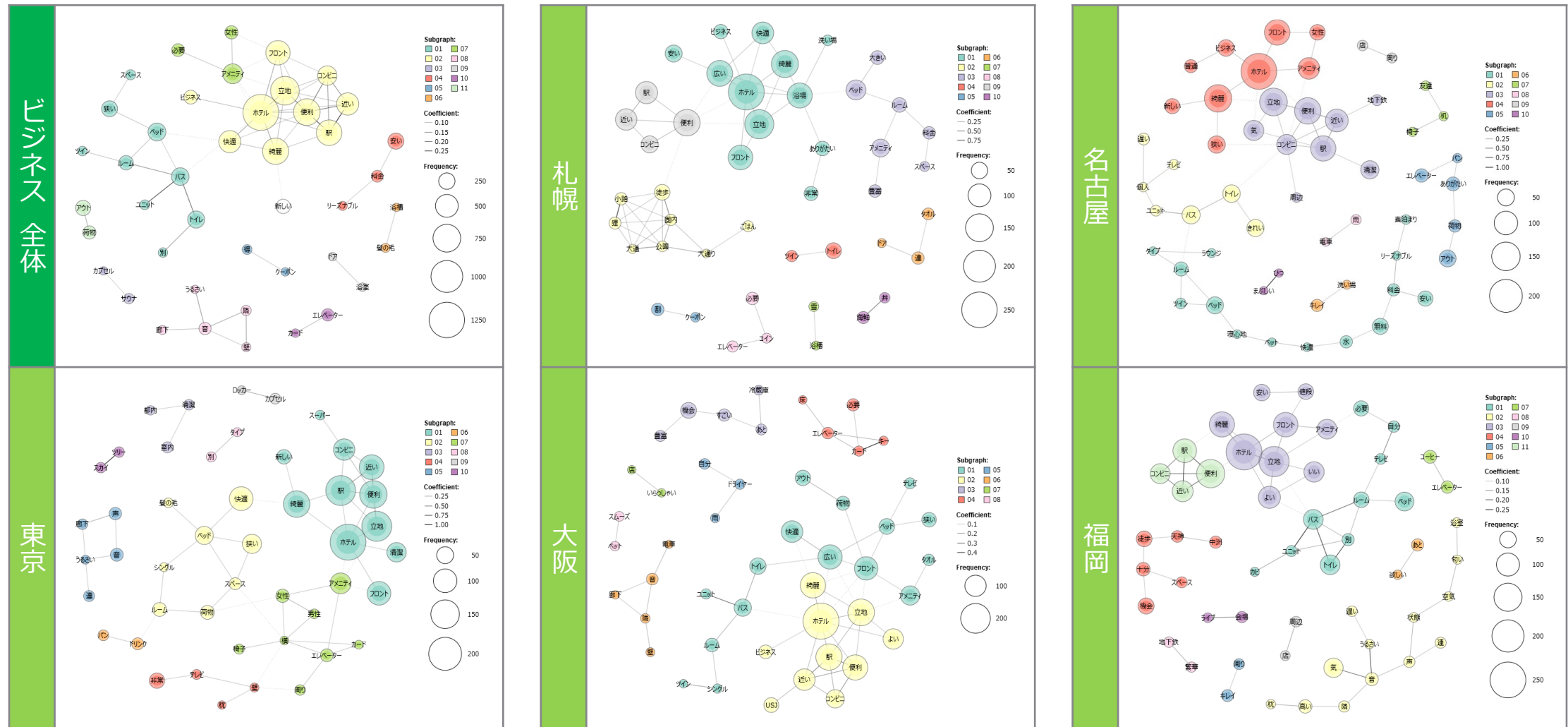
# 実践2 — ユーザー注目ポイントの評価を見る

## ● レジャーと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

## ● ビジネスとエリアごとの共起ネットワーク





## 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
- 手順の一例:
  - ・ 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

### 対応分析を用いる例:

「ツール」→「コーディング」→「**対応分析**」→「コーディングルール・ファイル: **coding-rule\_new2.txt**」「コーディング単位: **H5**」「コード選択: **\*ポジ, \*ネガ, \*総合1-2, \*総合4-5**」「コードx外部変数: **エリア**」 ※ **coding-rule\_new2.txt** については次頁で後述

### クロス集計を用いる例:

「ツール」→「コーディング」→「**クロス集計**」→「コーディングルール・ファイル: **coding-rule\_new2.txt**」「コーディング単位: **H5**」「コード選択: **\*ポジ, \*ネガ, \*総合1-2, \*総合4-5**」→「集計」ボタンをクリック→「ヒート」ボタンをクリック ※ **coding-rule\_new2.txt** については次頁で後述

## 実践3 — 改善プランを提案する

### ● コーディングルールをテキストエディタで編集する (例: メモ帳、Notepad++)

#### coding-rule.txt の中身

##### \*ポジ

良い or 美味しい or 広い or 多い or 素晴らしい or 嬉しい  
or 気持ちよい or 楽しい or 近い or 大きい or 気持ち良い  
or 温かい or 早い or 優しい or 新しい or 暖かい or 快い  
or 明るい or 美しい or 可愛い

##### \*ネガ

古い or 無い or 高い or 悪い or 小さい or 狭い or 少ない  
or 寒い or 遅い or 熱い or 欲しい or 暑い or 冷たい or  
遠い or 臭い or 暗い

##### \*風呂1-2

<>風呂-->1 | <>風呂-->2

##### \*風呂4-5

<>風呂-->4 | <>風呂-->5



#### coding-rule\_new2.txt の中身

##### \*ポジ

良い or 美味しい or 広い or 多い or 素晴らしい or 嬉しい  
or 気持ちよい or 楽しい or 近い or 大きい or 気持ち良い  
or 温かい or 早い or 優しい or 新しい or 暖かい or 快い  
or 明るい or 美しい or 可愛い **or 満足**

##### \*ネガ

古い or 無い or 高い or 悪い or 小さい or 狭い or 少ない  
or 寒い or 遅い or 熱い or 欲しい or 暑い or 冷たい or  
遠い or 臭い or 暗い **or 不満 or 残念**

##### \*総合1-2

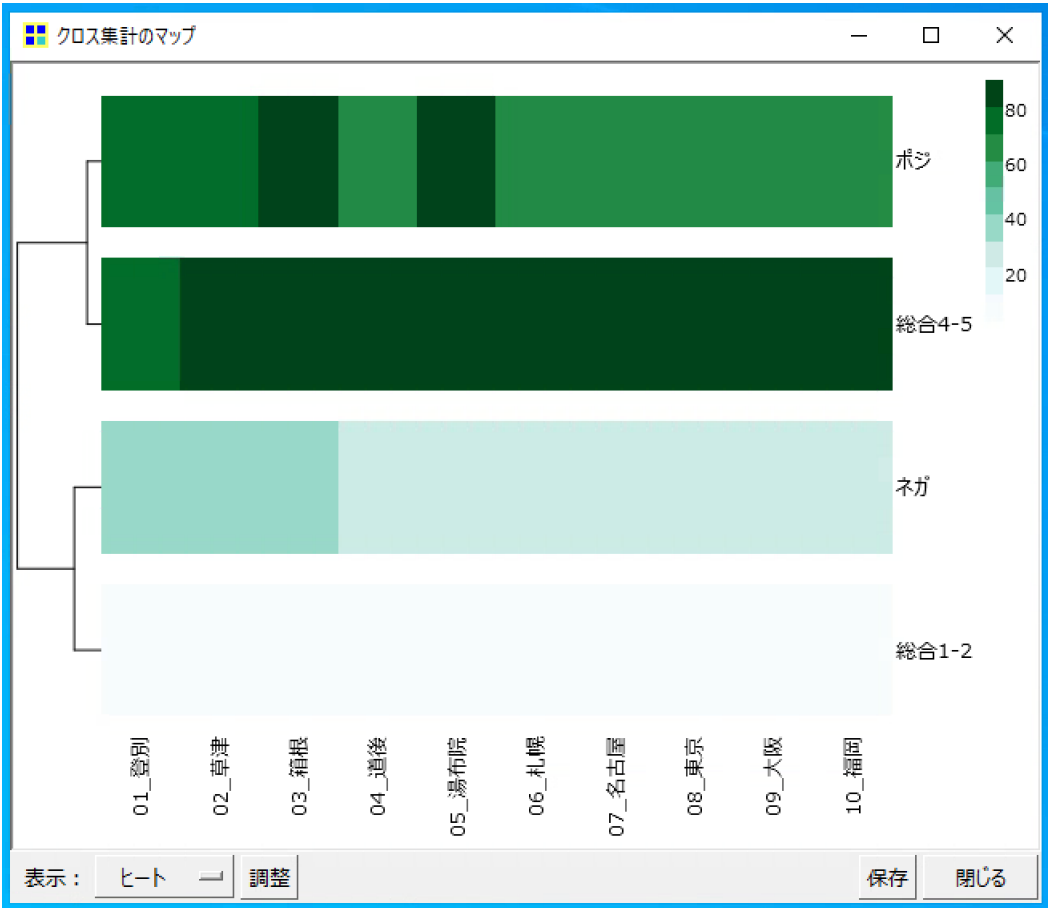
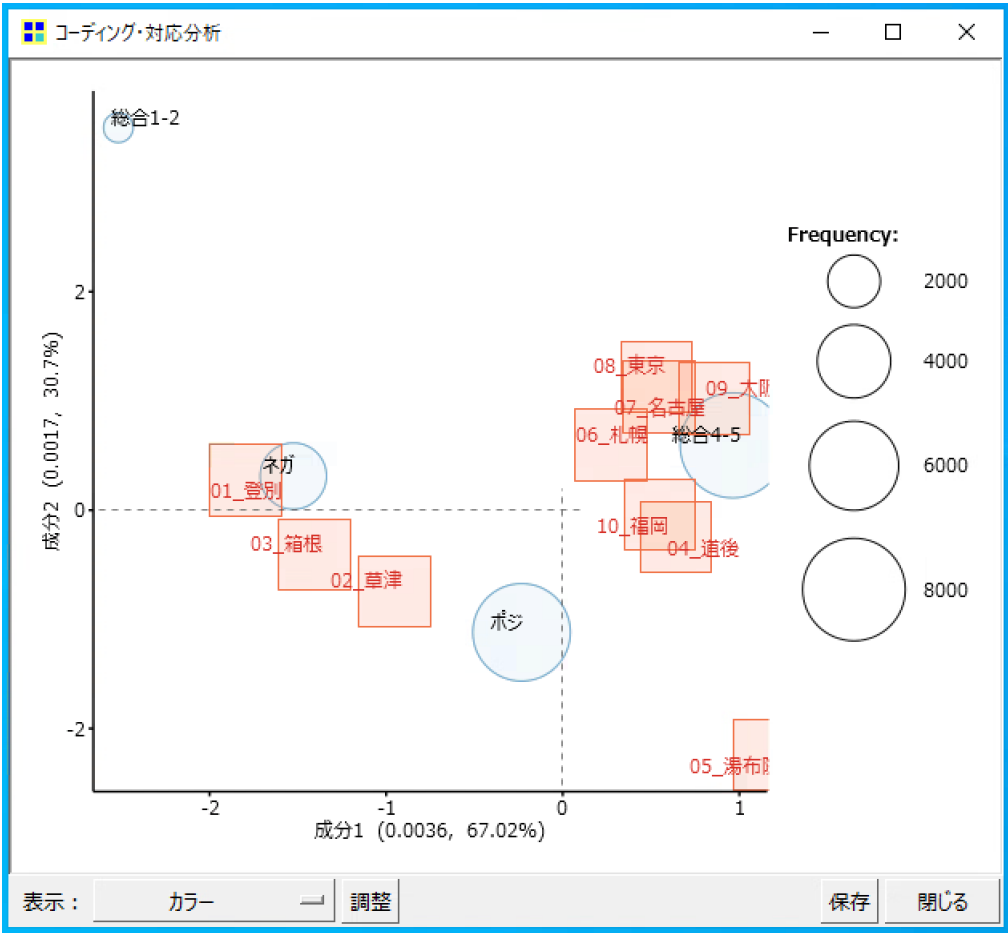
<>総合-->1 | <>総合-->2

##### \*総合4-5

<>総合-->4 | <>総合-->5

# 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)



## 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
- 手順の一例:
  - ・ 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

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]”<>カテゴリー-->A\_レジャー”」  
「Search Entry: \***ポジ**」AND検索「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位=60,  
**共起関係ほど濃い線に**」

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]”<>エリア-->01\_登別”」  
「Search Entry: \***ポジ**」AND検索「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位=60,  
**共起関係ほど濃い線に**」

## 実践3 — 改善プランを提案する

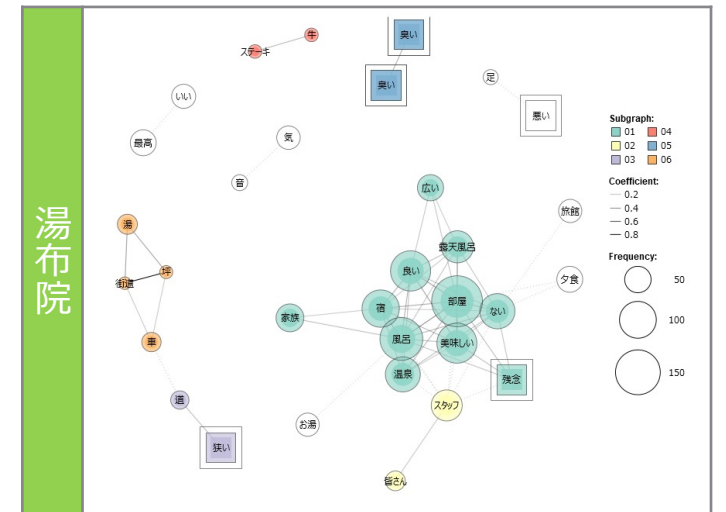
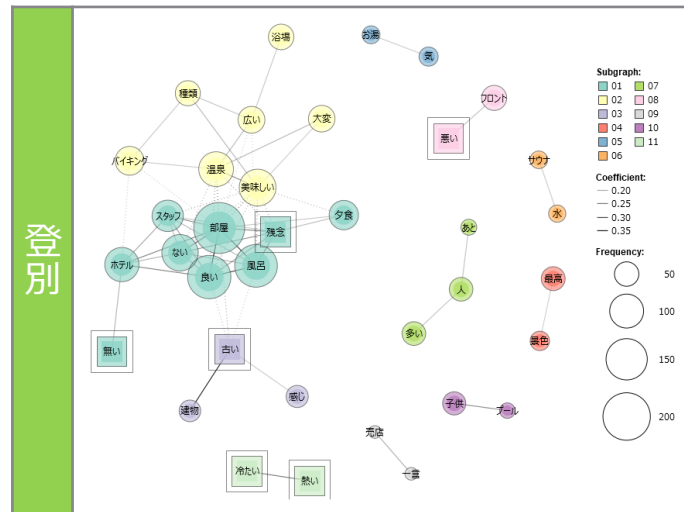
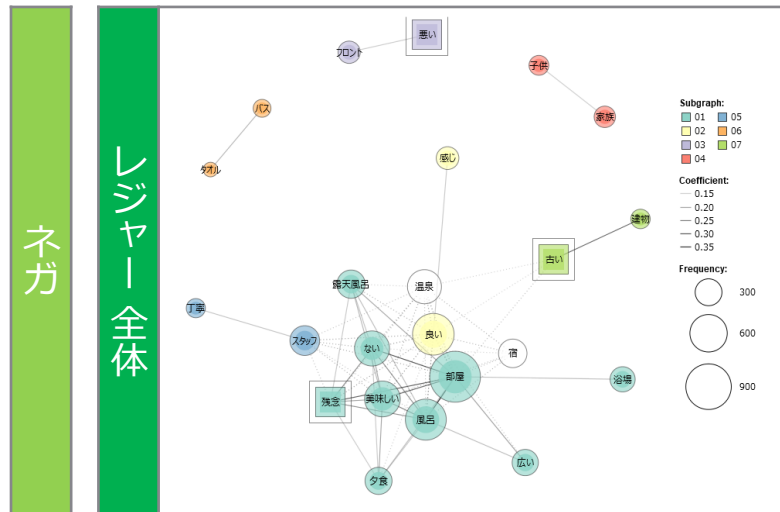
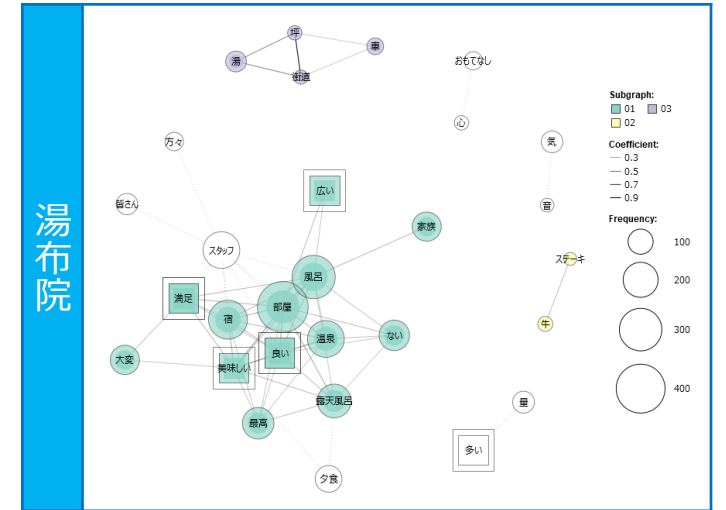
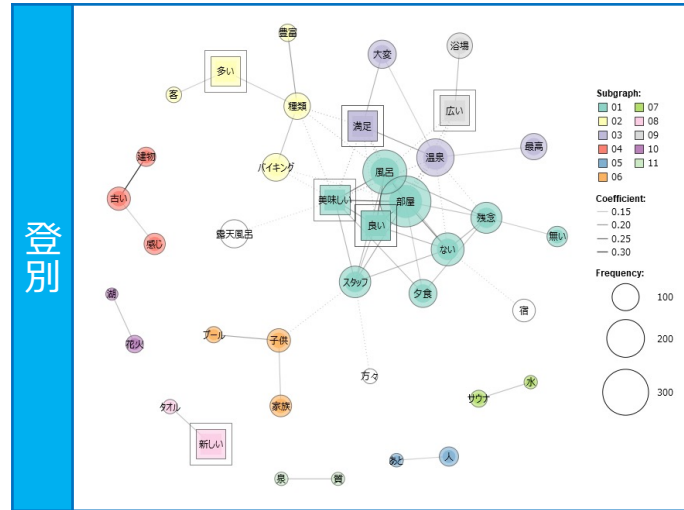
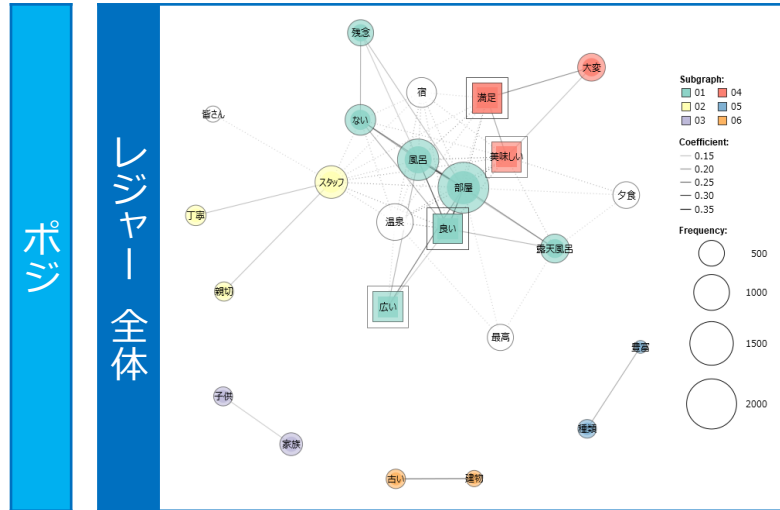
- カテゴリーやエリアごとの**ネガティブ意見**の違いを確認する
  - ・ 対照的な2エリアを選択する
  - ・ 選択したエリアと、そのカテゴリー(「レジャー」or「ビジネス」)を比較する
- 手順の一例:
  - ・ カテゴリーやエリアごとの共起NWから、何を**低(悪)評価**しているかを調べる

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]”<>カテゴリー-->A\_レジャー”」  
「Search Entry: \***ネガ**」AND検索「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位=60,  
**共起関係ほど濃い線に**」

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]”<>エリア-->01\_登別”」  
「Search Entry: \***ネガ**」AND検索「集計単位:H5」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位=60,  
**共起関係ほど濃い線に**」

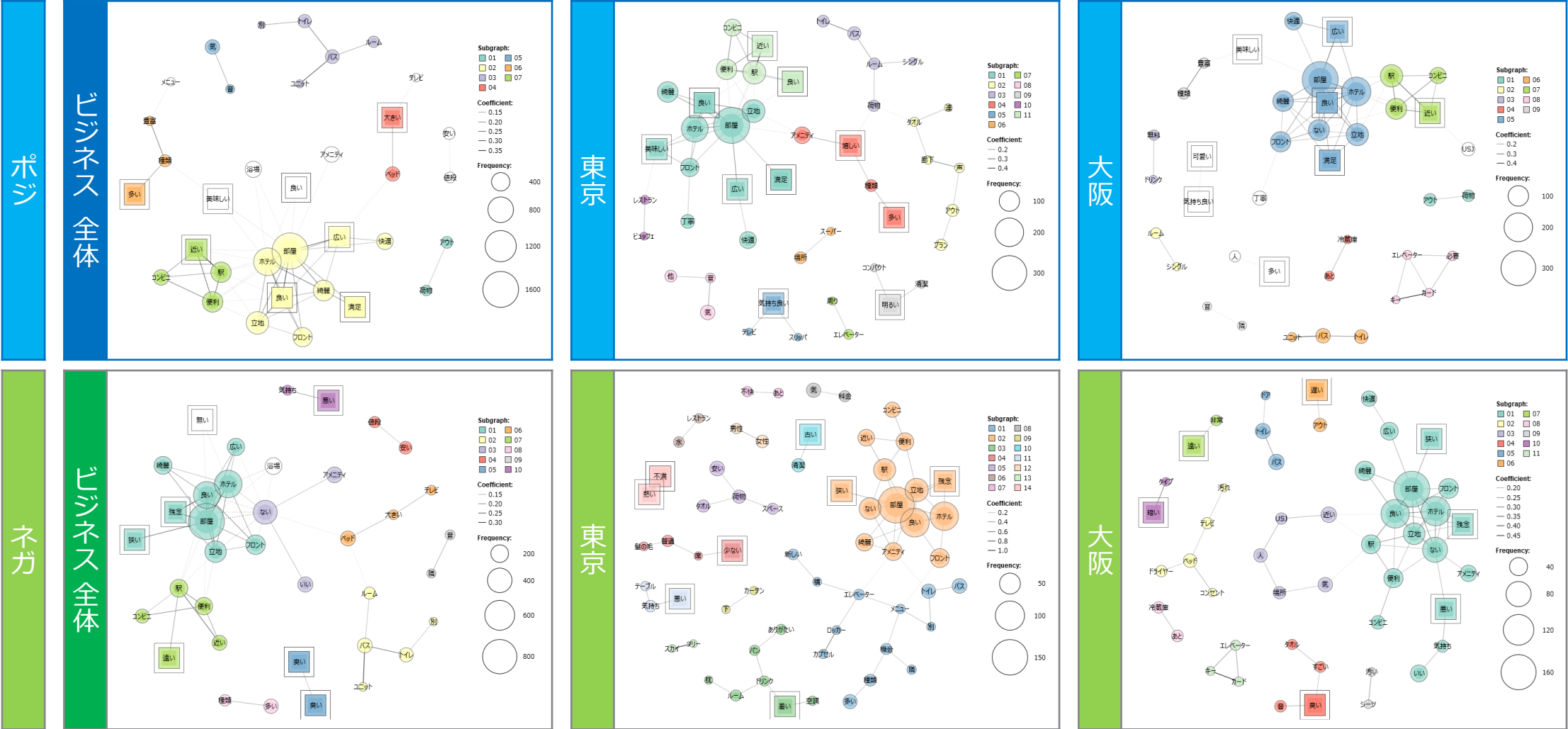
## 実践3 — 改善プランを提案する

## ● 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：東京と大阪のポジネガ比較



- 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.15)
- 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.17)
- 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.19)
  - 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 — 改善プランを提案する

---

- 個人ワーク (～20:20)

- 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

- グループ討論 (～20:50)

- 個人ワークで発見した、2エリアの特徴や改善プランについてグループ内で討論する
- グループ討論の内容を反映し、結果の整理 をブラッシュアップする

## day 5 – レポート課題

- 以下を PDF ファイルで提出 してください
    - 演習で作成した共起ネットワーク図(P.15,17)と結果の整理(P.19)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
- ※「プロット」のキャプチャは Jupyter の出力でも EXCEL でも構いません
- ※ 何らかの事情で上記の提出ができない場合は、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	8/4 ~18:20

## Q&A

# 参考資料

## ● KH Coder

- ・ 樋口耕一. 社会調査のための計量テキスト分析—内容分析の継承と発展を目指して【第2版】 KH Coder オフィシャルブック. ナカニシヤ出版, 2020.
- ・ 樋口耕一. テキスト型データの計量的分析 —2つのアプローチの峻別と統合—. 理論と方法, 数理社会学会, 2004, 19(1): 101-115.
- ・ 牛澤賢二. やってみよう テキストマイニング —自由回答アンケートの分析に挑戦!. 朝倉書店, 2019
- ・ 樋口耕一. 動かして学ぶ! はじめてのテキストマイニング: フリー・ソフトウェアを用いた自由記述の計量テキスト分析 KH Coder オフィシャルブック II.ナカニシヤ出版, 2022.

## ● Windows環境によるデータ収集方法の参考

- ・ テキストマイニングソフトを利用した新未来洞察手法の研究. 第10分科会, (財)市場創造研究会. [[発表スライド](#)]

## ● R を使った参考書

- ・ 金明哲. "テキストデータの統計科学入門." 岩波書店, 2009.
- ・ 石田基広. "RMeCab によるテキスト解析. R によるテキストマイニング入門." 森北出版, 2008, 51-82.

## ● 他のツールを使った参考書

- ・ 那須川哲哉. "テキストマイニングを使う技術/作る技術: 基礎技術と適用事例から導く本質と活用法." 東京電機大学出版局, 2006.
- ・ 上田隆穂, 黒岩祥太, 戸谷圭子. "テキストマイニングによるマーケティング調査." 講談社, 2005.

## ● 統計解析を中心とした参考書

- ・ 前田忠彦; 山崎誠. 言語研究のための統計入門. くろしお出版株式会社, 東京, 2013.