

進捗報告

1 データセット

1.1 楽天トラベルレビュー：アスペクトセンチメントタグ付きコーパス

楽天トラベルレビュー：アスペクトセンチメントタグ付きコーパスとは、楽天グループ株式会社が提供しているデータセットである。日本語レビュー文章とそれぞれの文章について、立地、部屋、食事等の 7 項目のアスペクトに対するポジティブまたはネガティブのタグが付与されている。総データ数は 76624 で、朝食、夕食、風呂、サービス、施設、立地、のポジティブ、ネガティブの 14 個のカテゴリに分類される。今回は 14 のいずれのカテゴリにも属さないデータを除くことで、少なくとも 1 つのラベルに属し、語彙数が 10 以下と 100 以上のデータを取り除いたデータ群にした。総データ数は 50211 である。

2 今週やったこと

東北大学の乾研究室が公開している日本語評価極性辞書^{*1}を用いて総データ 50211 に含まれる各カテゴリ毎での評価語の出現回数を調査した。このうち、各カテゴリにおいてポジティブとネガティブの双方に属するデータについては今回除いて調査した。カテゴリに属するデータと属さないデータの合計がいずれのカテゴリにおいても 50211 にならないのはこのためである。その様なデータを除いた理由は、今後ポジティブであるか、ネガティブであるかの 2 値分類をする予定であるためである。表 1 に各カテゴリに属しているデータにおける評価語と評価語がないデータ数と、カテゴリに属していないデータにおける評価語と評価語がないデータ数について示す。表 1 から、いずれのカテゴリにおいても、ネガティブ評価語よりポジティブ評価語の方が多いことがわかる。特に立地カテゴリに属しているデータではその差が大きいことがわかる。

表 1: カテゴリに属しているデータと属していないデータの評価語の出現回数と評価語がないデータ数

| | 朝食 | 夕食 | 風呂 | サービス | 立地 | 施設 | 部屋 |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| カテゴリに属するデータの数 | 13882 | 10325 | 9179 | 18369 | 6417 | 13355 | 10033 |
| ネガティブ評価語の数 | 613 | 564 | 547 | 955 | 390 | 844 | 630 |
| ポジティブ評価語の数 | 827 | 753 | 693 | 1084 | 601 | 917 | 721 |
| 評価語がないデータの数 | 1958 | 1511 | 1773 | 3210 | 1204 | 2660 | 1680 |
| カテゴリに属さないデータの数 | 35994 | 39643 | 40697 | 31558 | 43614 | 36463 | 39846 |
| ネガティブ評価語の数 | 1332 | 1366 | 1359 | 1111 | 1402 | 1226 | 1337 |
| ポジティブ評価語の数 | 1377 | 1408 | 1426 | 1170 | 1443 | 1346 | 1418 |
| 評価語がないデータの数 | 6611 | 7063 | 6772 | 5343 | 7343 | 5892 | 6875 |

次に、図 1 から図 7 に各カテゴリでの評価語の出現回数が多い順に 20 個抽出して示す。いずれのカテゴリにおいても、ネガティブ評価語よりもポジティブ評価語の出現回数が多いことがわかる。また、カテゴリに属しているデータと属していないデータの評価語は基本的には異なっている一方で、出現回数が上位の評価語は似通っていることも見受けられる。朝食と夕食のカテゴリでは食事に関連すると考えられる評価語の出現回数が多い。その他のカテゴリにおいても、そのカテゴリから想起することが可能な評価語が多く出現している。

^{*1} <https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|---------------|------|----------|------|
| 種類 | 692 | 美味しい | 3185 |
| 大変 | 645 | 満足 | 1867 |
| 残念 | 365 | 良い | 1644 |
| 気 | 170 | おいしい | 901 |
| 値段 | 162 | 豊富 | 415 |
| 非常 | 147 | 満足 する | 409 |
| ボリューム | 139 | 味 | 372 |
| 悪い | 87 | サービス | 356 |
| 価格 | 76 | 最高 | 341 |
| 料金 | 76 | 素晴らしい | 302 |
| 形式 | 66 | 量 | 299 |
| 静か | 57 | 出来る | 259 |
| 美味しい-NEGATION | 53 | いい | 258 |
| 混雑 | 53 | 期待 | 212 |
| 不満 | 51 | 充実 | 193 |
| 文句 | 49 | 楽しめる | 179 |
| 遅い | 44 | 嬉しい | 167 |
| 混む | 40 | 旅行 | 159 |
| がっかり | 40 | 綺麗 | 146 |
| 問題 | 36 | 楽しみ | 144 |

(a) 朝食カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 1224 | 良い | 4562 |
| 大変 | 1070 | サービス | 1180 |
| 気 | 818 | 満足 | 1128 |
| アウト | 422 | 綺麗 | 1040 |
| 非常 | 412 | 便利 | 943 |
| 悪い | 399 | いい | 934 |
| 静か | 312 | 快適 | 810 |
| 問題 | 281 | 出来る | 802 |
| 臭い | 247 | 最高 | 751 |
| 不便 | 237 | 清潔 | 751 |
| 種類 | 208 | 美味しい | 739 |
| 値段 | 193 | 丁寧 | 696 |
| 料金 | 189 | 親切 | 679 |
| 遅い | 176 | 素晴らしい | 614 |
| 疲れ | 173 | 嬉しい | 613 |
| 喫煙 | 148 | きれい | 606 |
| 困る | 145 | 助かる | 561 |
| 不満 | 139 | ため | 501 |
| 仕方 ない | 124 | 無料 | 470 |
| 価格 | 114 | 設備 | 438 |

(b) 朝食カテゴリに属さないデータの評価語

図 1: 朝食カテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|---------------|------|----------|------|
| 大変 | 542 | 美味しい | 2276 |
| 残念 | 295 | 満足 | 1492 |
| 種類 | 246 | 良い | 1261 |
| ボリューム | 154 | おいしい | 584 |
| 値段 | 126 | 量 | 351 |
| 気 | 124 | 満足 する | 348 |
| 非常 | 105 | サービス | 314 |
| 料金 | 63 | 味 | 311 |
| 悪い | 59 | 最高 | 267 |
| 美味しい-NEGATION | 52 | 素晴らしい | 223 |
| 価格 | 50 | 出来る | 213 |
| 静か | 48 | いい | 180 |
| 遅い | 41 | 期待 | 165 |
| 形式 | 38 | 旅行 | 151 |
| 文句 | 36 | 楽しめる | 134 |
| 不満 | 36 | 豊富 | 131 |
| 夜泣き | 33 | 丁寧 | 118 |
| 付け | 33 | 楽しい | 113 |
| がっかり | 31 | 質 | 112 |
| 問題 | 30 | 新鮮 | 109 |

(a) 夕食カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 1301 | 良い | 4964 |
| 大変 | 1171 | 美味しい | 1658 |
| 気 | 864 | 満足 | 1511 |
| 種類 | 671 | サービス | 1226 |
| 非常 | 454 | 綺麗 | 1100 |
| アウト | 440 | いい | 1016 |
| 悪い | 431 | 便利 | 959 |
| 静か | 321 | 出来る | 848 |
| 問題 | 287 | 快適 | 847 |
| 臭い | 246 | 最高 | 827 |
| 不便 | 245 | 清潔 | 781 |
| 値段 | 230 | 丁寧 | 703 |
| 料金 | 204 | 親切 | 695 |
| 遅い | 178 | 素晴らしい | 693 |
| 疲れ | 173 | 嬉しい | 688 |
| 不満 | 154 | きれい | 624 |
| 困る | 151 | 助かる | 575 |
| 喫煙 | 150 | 無料 | 562 |
| 価格 | 142 | ため | 523 |
| 混雑 | 136 | おいしい | 501 |

(b) 夕食カテゴリに属さないデータの評価語

図 2: 夕食カテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 347 | 良い | 1476 |
| 大変 | 312 | 満足 | 933 |
| 気 | 158 | 最高 | 410 |
| 疲れ | 138 | いい | 257 |
| 非常 | 97 | 出来る | 250 |
| 悪い | 95 | 快適 | 231 |
| 種類 | 81 | 綺麗 | 223 |
| 値段 | 68 | 美味しい | 212 |
| 疲れが | 56 | 素晴らしい | 207 |
| 臭い | 55 | 満足 する | 201 |
| 静か | 55 | 清潔 | 196 |
| 問題 | 53 | ゆったり | 171 |
| 排水 | 53 | 質 | 166 |
| カビ | 48 | 気持ち良い | 154 |
| 不満 | 44 | 楽しめる | 153 |
| 価格 | 42 | 景色 | 147 |
| 不便 | 41 | 旅行 | 132 |
| 混む | 39 | 癒す | 131 |
| 混雑 | 38 | きれい | 128 |
| アウト | 37 | サービス | 125 |

(a) 風呂カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 1404 | 良い | 4695 |
| 残念 | 1234 | 美味しい | 3811 |
| 種類 | 844 | 満足 | 2071 |
| 気 | 831 | サービス | 1416 |
| 非常 | 459 | おいしい | 1060 |
| アウト | 419 | 綺麗 | 950 |
| 悪い | 393 | 便利 | 946 |
| 静か | 314 | いい | 932 |
| 値段 | 292 | 丁寧 | 803 |
| 問題 | 260 | 出来る | 796 |
| 料金 | 233 | 親切 | 730 |
| 不便 | 204 | 素晴らしい | 707 |
| 臭い | 200 | 快適 | 707 |
| ボリューム | 193 | 嬉しい | 705 |
| 遅い | 190 | 最高 | 676 |
| 価格 | 151 | 清潔 | 620 |
| 喫煙 | 148 | 助かる | 563 |
| 困る | 145 | 無料 | 553 |
| 不満 | 143 | 味 | 527 |
| 仕方 ない | 117 | きれい | 524 |

(b) 風呂カテゴリに属さないデータの評価語

図 3: 風呂カテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 731 | 良い | 2437 |
| 残念 | 530 | サービス | 1282 |
| 気 | 411 | 満足 | 1093 |
| アウト | 398 | 清潔 | 757 |
| 悪い | 229 | 丁寧 | 749 |
| 非常 | 228 | 親切 | 739 |
| 臭い | 138 | 綺麗 | 576 |
| 値段 | 116 | 美味しい | 541 |
| 問題 | 112 | 素晴らしい | 477 |
| 遅い | 111 | 笑顔 | 454 |
| 料金 | 94 | 出来る | 415 |
| カビ | 89 | 快適 | 391 |
| 種類 | 85 | いい | 389 |
| 不満 | 84 | 嬉しい | 343 |
| 静か | 79 | きれい | 328 |
| 喫煙 | 76 | 気持ち | 323 |
| 汚れ | 70 | 助かる | 315 |
| 忘れる | 68 | 旅行 | 274 |
| 落ちる | 67 | 満足 する | 254 |
| 不安 | 66 | 楽しい | 245 |

(a) サービスカテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 1051 | 良い | 3745 |
| 大変 | 991 | 美味しい | 3476 |
| 種類 | 845 | 満足 | 1911 |
| 気 | 572 | おいしい | 990 |
| 非常 | 330 | 便利 | 894 |
| 静か | 290 | 最高 | 857 |
| 悪い | 260 | いい | 805 |
| 値段 | 244 | 出来る | 646 |
| 不便 | 209 | 綺麗 | 603 |
| 問題 | 206 | 快適 | 555 |
| ボリューム | 187 | 味 | 481 |
| 料金 | 172 | 豊富 | 465 |
| 疲れ | 157 | 景色 | 441 |
| 価格 | 130 | 嬉しい | 439 |
| 困る | 114 | 素晴らしい | 439 |
| 臭い | 111 | 楽しめる | 432 |
| 遅い | 105 | 量 | 423 |
| 不満 | 105 | 満足 する | 416 |
| 仕方 ない | 96 | 無料 | 391 |
| 壁 | 85 | ゆったり | 376 |

(b) サービスカテゴリに属さないデータの評価語

図 4: 朝サービスカテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 308 | 良い | 980 |
| 静か | 141 | 便利 | 735 |
| 残念 | 113 | 満足 | 610 |
| 非常 | 111 | 最高 | 207 |
| 不便 | 65 | いい | 191 |
| 気 | 64 | 観光 | 155 |
| 値段 | 60 | 出来る | 142 |
| 問題 | 59 | 満足 する | 142 |
| 価格 | 45 | 素晴らしい | 136 |
| 悪い | 42 | 行ける | 125 |
| 困る | 39 | 快適 | 121 |
| 料金 | 33 | 旅行 | 121 |
| アウト | 30 | 景色 | 120 |
| 不満 | 29 | 機会 | 114 |
| 迷う | 28 | 美味しい | 113 |
| 遅い | 25 | 繁華 | 112 |
| 文句 | 23 | サービス | 107 |
| 不安 | 21 | 綺麗 | 106 |
| 疲れ | 20 | ため | 95 |
| 疲れる | 20 | 楽しめる | 75 |

(a) 立地カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 1509 | 良い | 5260 |
| 大変 | 1417 | 美味しい | 3911 |
| 気 | 930 | 満足 | 2417 |
| 種類 | 917 | サービス | 1437 |
| 悪い | 452 | おいしい | 1083 |
| 非常 | 449 | 綺麗 | 1081 |
| アウト | 427 | いい | 1008 |
| 値段 | 300 | 出来る | 921 |
| 臭い | 251 | 最高 | 885 |
| 問題 | 251 | 快適 | 826 |
| 料金 | 235 | 丁寧 | 801 |
| 静か | 209 | 素晴らしい | 783 |
| ボリューム | 199 | 清潔 | 781 |
| 遅い | 194 | 嬉しい | 758 |
| 疲れ | 178 | 親切 | 745 |
| 不便 | 175 | きれい | 590 |
| 不満 | 162 | 無料 | 578 |
| 喫煙 | 150 | 助かる | 538 |
| 価格 | 148 | 満足 する | 531 |
| 混雑 | 145 | 味 | 531 |

(b) 立地カテゴリに属さないデータの評価語

図 5: 立地カテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 残念 | 498 | 良い | 1637 |
| 大変 | 392 | 満足 | 790 |
| 気 | 302 | 綺麗 | 454 |
| 悪い | 159 | サービス | 447 |
| 非常 | 154 | 設備 | 398 |
| 不便 | 130 | いい | 367 |
| 静か | 106 | 快適 | 339 |
| 問題 | 101 | 無料 | 317 |
| 臭い | 90 | 嬉しい | 304 |
| 種類 | 87 | 出来る | 299 |
| 値段 | 85 | 清潔 | 295 |
| 料金 | 83 | 美味しい | 264 |
| 不満 | 79 | 助かる | 244 |
| アウト | 70 | きれい | 222 |
| 困る | 64 | 充実 | 218 |
| 仕方 ない | 56 | 雰囲気 | 213 |
| 喫煙 | 55 | 最高 | 196 |
| 価格 | 54 | ため | 191 |
| カビ | 46 | 便利 | 185 |
| 遅い | 45 | 素晴らしい | 185 |

(a) 施設カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 1319 | 良い | 4561 |
| 残念 | 1105 | 美味しい | 3754 |
| 種類 | 840 | 満足 | 2210 |
| 気 | 675 | サービス | 1093 |
| 非常 | 400 | おいしい | 1054 |
| アウト | 387 | 最高 | 891 |
| 悪い | 329 | いい | 818 |
| 値段 | 272 | 便利 | 785 |
| 静か | 260 | 丁寧 | 766 |
| 問題 | 199 | 出来る | 761 |
| ボリューム | 193 | 素晴らしい | 730 |
| 料金 | 183 | 親切 | 713 |
| 遅い | 176 | 綺麗 | 701 |
| 臭い | 164 | 快適 | 594 |
| 疲れ | 155 | 味 | 507 |
| 価格 | 137 | 清潔 | 501 |
| 不便 | 116 | 満足 する | 493 |
| 混雑 | 113 | 嬉しい | 478 |
| 不満 | 109 | 豊富 | 462 |
| 混む | 103 | 笑顔 | 446 |

(b) 施設カテゴリに属さないデータの評価語

図 6: 施設カテゴリに注目した評価語の出現回数

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 371 | 良い | 1369 |
| 残念 | 310 | 満足 | 945 |
| 気 | 257 | 綺麗 | 763 |
| 静か | 152 | 快適 | 579 |
| 臭い | 135 | 清潔 | 502 |
| 非常 | 115 | きれい | 431 |
| 値段 | 95 | 出来る | 270 |
| 悪い | 89 | 景色 | 263 |
| 問題 | 77 | 最高 | 248 |
| 喫煙 | 74 | ゆったり | 234 |
| 壁 | 65 | 美味しい | 216 |
| 不満 | 57 | 素晴らしい | 214 |
| 価格 | 55 | 満足 する | 199 |
| 料金 | 51 | いい | 197 |
| 落ちる | 40 | サービス | 173 |
| 角 | 40 | 素敵 | 167 |
| カビ | 39 | 旅行 | 146 |
| 疲れ | 37 | 十分 | 143 |
| 仕方 ない | 36 | 設備 | 137 |
| 暗い | 31 | キレイ | 136 |

(a) 部屋カテゴリに該当するデータの評価語

| ネガティブ評価語 | 出現回数 | ポジティブ評価語 | 出現回数 |
|----------|------|----------|------|
| 大変 | 1350 | 良い | 4836 |
| 残念 | 1293 | 美味しい | 3808 |
| 種類 | 904 | 満足 | 2065 |
| 気 | 715 | サービス | 1370 |
| 非常 | 445 | おいしい | 1067 |
| アウト | 434 | いい | 994 |
| 悪い | 401 | 便利 | 916 |
| 値段 | 262 | 最高 | 842 |
| 問題 | 225 | 出来る | 792 |
| 不便 | 222 | 丁寧 | 778 |
| 料金 | 216 | 親切 | 717 |
| 静か | 214 | 素晴らしい | 705 |
| 遅い | 201 | 嬉しい | 702 |
| ボリューム | 192 | 無料 | 584 |
| 疲れ | 160 | 助かる | 570 |
| 混雑 | 145 | 味 | 526 |
| 困る | 145 | 豊富 | 493 |
| 価格 | 137 | 楽しめる | 476 |
| 不満 | 133 | 満足 する | 473 |
| 混む | 119 | ため | 472 |

(b) 部屋カテゴリに属さないデータの評価語

図 7: 部屋カテゴリに注目した評価語の出現回数