

データマイニング

Data mining

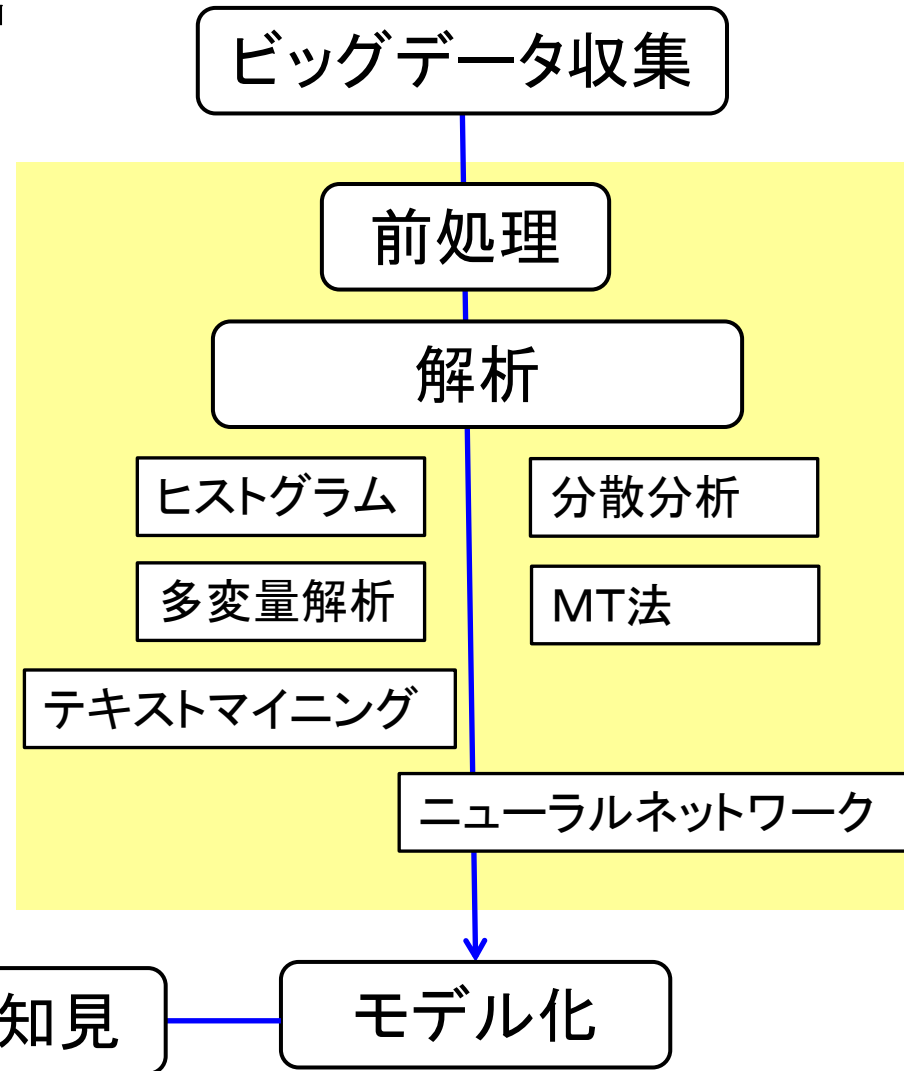
データ発掘

ガラクタ



データマイニング

進化
宝



セキュリティを強化したテキストマイニング ビジネス版を公開 (分析結果サンプル)

解析したいテキストを入力する

1つの文書を解析

2つの文章を比較

Excelクロス集計 (ベータ版)

音声入力

サンプル: 太宰治『走れメロス』

フォーム入力

ファイルをアップロード

履歴から選択

クリックしてみてください

品質というと「統計」を知らないと仕事ができないという方が居られます。確かに、知らないよりは知っていた方が良いでしょうね。私は「感性」「気づき」が一番大事なスキルだと思っています。感性が低い私が言うのも説得力がありませんが。

一番重要なスキルは、**五感**を使った「感性」です。視覚・聴覚・嗅覚・味覚・触覚ですね。例えば、機械や**データ**の異常を見つけるためには**五感**を十分に働かせる必要があります。「いつもと違う」音、臭い、振動、数字の並びに如何に気が付くかのセンスが必要です。

五感を鍛える方法を3つお勧めします。

- ・イメージが描ける本を読む
- ・ライブを見に行く、聴きに行く。
- ・自然に触れる

二番目のスキルは、**データ**を読む力ですね。

特別な力ではありません。先ずは、違いに気づくか気が付かないかです。こんな**データ**があります。いかがですか？ 何も感じませんか？ 規格に入っているから問題なし、ですか？

確かに規格内に入っているので、日々見ている限りではおかしくはありません。しかし、規格ギリギリの数値が連続しています。あまりにも綺麗すぎませんか？ 「疑いの眼(まなこ)」で見ることも必要です。

変化を見る手法としては、既に**管理図**という確率した手法が世の中にはあります。QC7つ道具の一つでもあります。7つの中でも、**ヒストグラム**及び**管理図**は特に有効な道具であると考えます。

数字がたくさんある**データ**の処理方法として「統計」が有効ですが、統計処理の前に**データ**を「**ヒストグラム**」に描いてみることをお勧めします。なぜか？

平均値と**標準偏差**を統計ソフトで計算させて安心する方が居られます。では、この2つの**ヒストグラム**をご覧ください。何れも平均値と**標準偏差**は同一です。**ヒストグラム**を見ると、片方は2つのグループが存在する可能性を表しています。

聴く 確か 技能 同一 一番
 入る 臭い 番目 品質 知る 統計処理
 必要 お勧め 平均値 使う
 振動 何れ 気が付く 居る 2つ 道具 描ける
 嗅覚 ヒストグラム 行く 触れる 違う
 低い 規格 五感 管理図 標準偏差
 鍛える 気づく 聴覚 感性 統計 働かせる 描く
 見つける 感じる qc 7つ 表す 触覚 方法 読む
 付く 手法 数字 データ 有効
 考える 視覚 見る おかしい

単語出現頻度

文章中に出現する単語の頻出度を表にしています。単語ごとに表示されている「スコア」の大きさは、与えられた文書の中でその単語がどれだけ特徴的であるかを表しています。通常はその単語の出現回数が多いほどスコアが高くなりますが、「言う」や「思う」など、どの文書にもよく現れる単語についてはスコアが低めになります。

[【スコアの目的と算出方法について】](#)

名詞	スコア	出現頻度
データ	0.75	5
ヒストグラム	29.32	4
五感	6.83	3
規格	4.83	3
統計	3.76	3
感性	2.15	3
スキル	0.32	3
必要	0.08	3
管理図	17.29	2
標準偏差	12.68	2
7つ	2.43	2
手法	1.26	2
有効	0.65	2
お勧め	0.62	2
道具	0.37	2

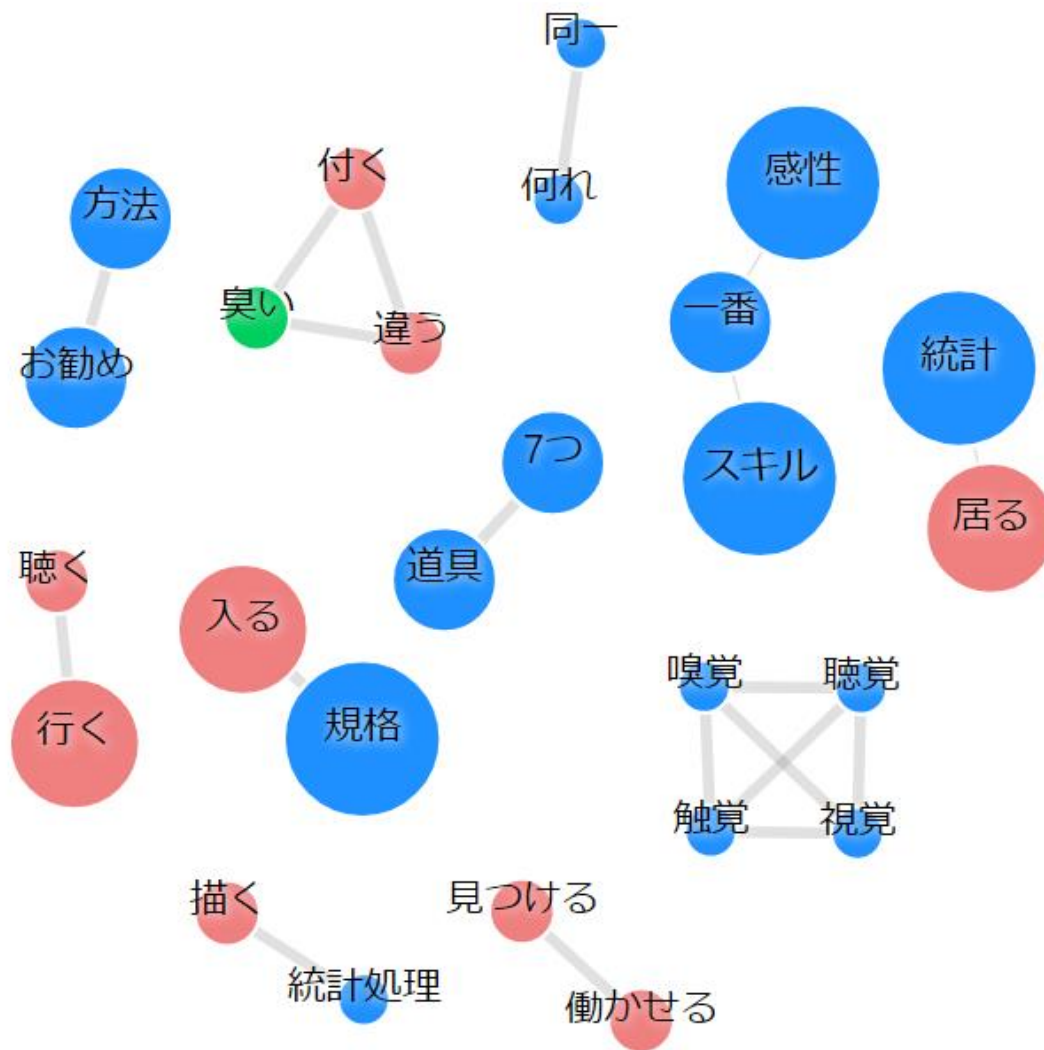
動詞	スコア	出現頻度
知る	0.02	3
見る	0.01	3
居る	0.04	2
気づく	0.03	2
読む	0.02	2
入る	0.01	2
行く	0.00	2
働かせる	0.49	1
気が付く	0.27	1
表す	0.17	1
鍛える	0.08	1
触れる	0.03	1
描ける	0.03	1
付く	0.02	1
見つける	0.01	1

形容詞	スコア	出現頻度
臭い	0.04	1
低い	0.02	1
おかしい	0.01	1
良い	0.00	1

感動詞	スコア	出現頻度
---	---	---
---	---	---
---	---	---
---	---	---

共起キーワード

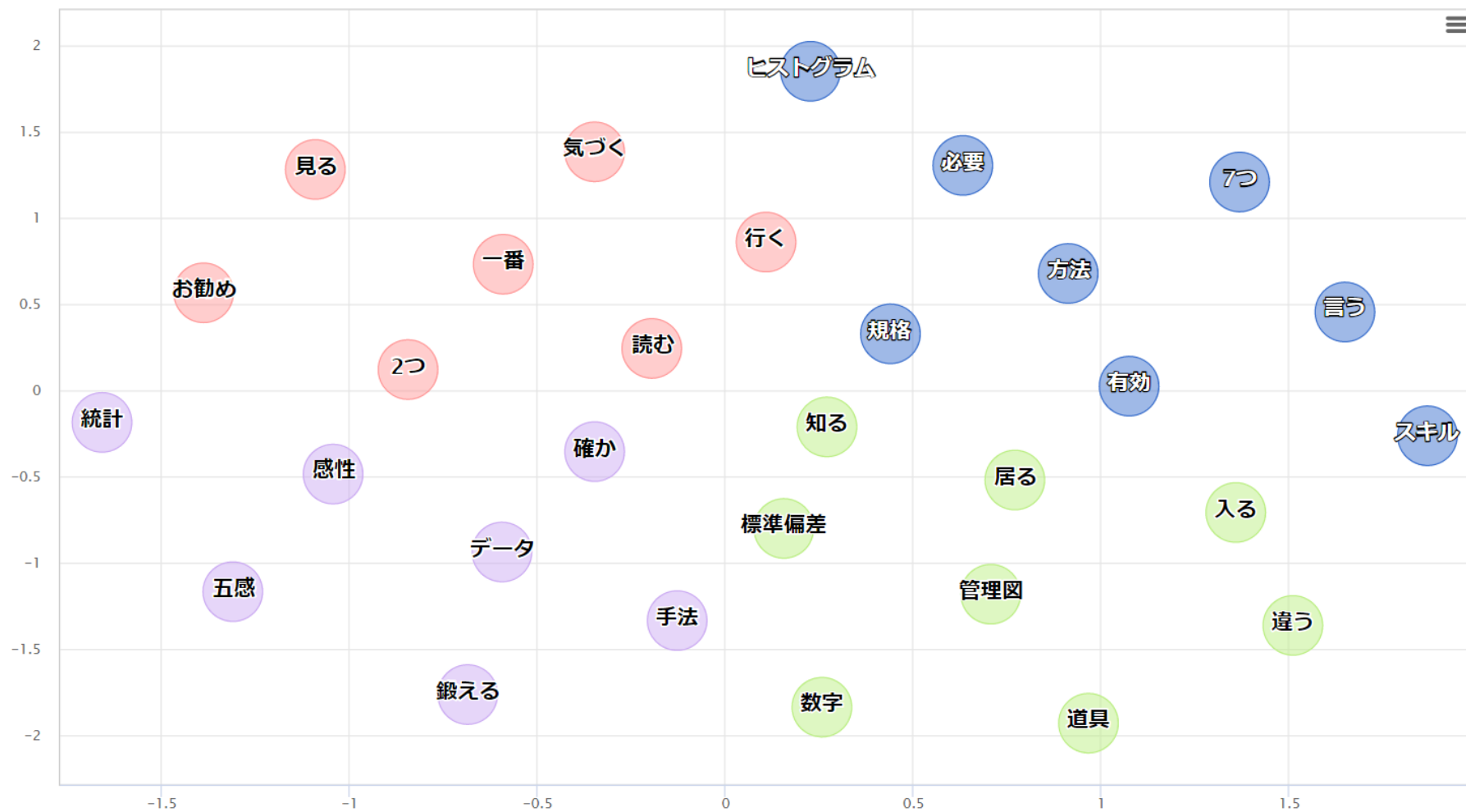
文章中出现する単語の出現パターンが似たものを線で結んだ図です。出現数が多い語ほど大きく、また共起の程度が強いほど太い線で描画されます。 [【共起とは？】](#)



2次元マップ

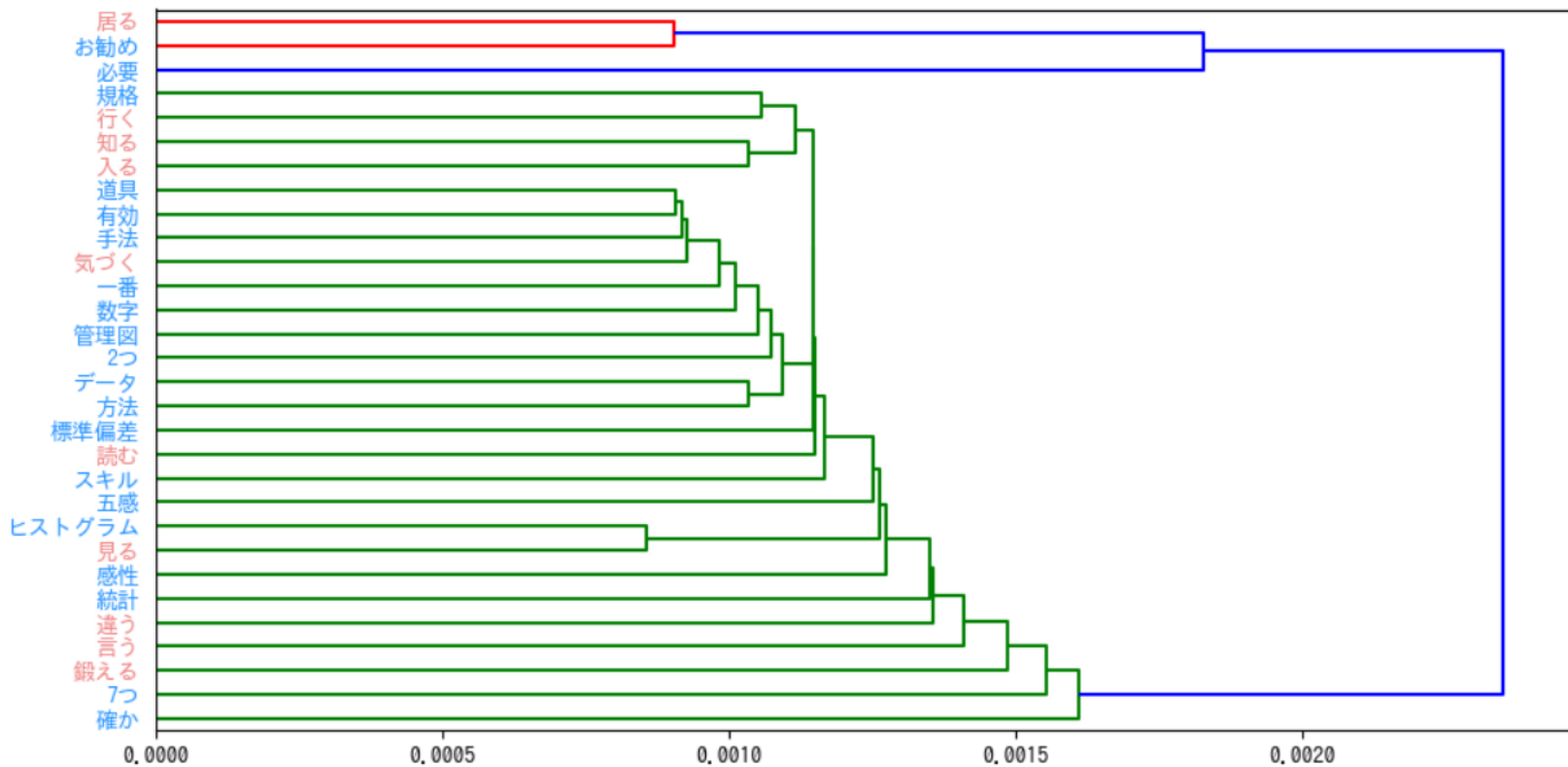
文章中での出現傾向が似た単語ほど近く、似ていない単語ほど遠く配置されています。距離が近い単語はグループにまとめ、色分けしています。

[【2次元マップの詳細説明】](#)



階層的クラスタリング

文章中での出現傾向が似た単語をまとまりとしてとらえられるよう樹形図で表したものです。グループは色分けして表示しています。 [【階層的クラスタリングの詳しい説明】](#)



アドレス <https://api.ce-cotoha.com/contents/index.html>

COTOHA API

お知らせ | イベント | ログイン

COTOHA APIについて API一覧 解析デモ 活用事例 料金プラン 開発者向け ▾

無料登録

日本最大級の 日本語辞書を活用

自然言語処理
音声処理
APIプラットフォーム

今すぐ無料登録 →

NTT Communications

14の自然言語処理・音声認識APIをご提供

言語のなかでも特にむずかしいとされる日本語の高精度な解析が可能なCOTOHA API。
テキスト解析や対話エンジン(テキスト/音声など)にご利用いただけます。



構文解析

入力された文から、「誰（何）」が「どうした」とい
う情報を形態素と呼ばれる単語単位で切り出すことが
できるためデータマイニングなどに応用可能です。



照応解析

対話エンジンとの対話ログの解析などで、「彼/彼女」
などのログ解析にあまり意味のない単語を先行詞に置
き換え、より精密なログ解析を実現することが可能で
す。



音声認識

ファイル形式やストリーミング形式で音声データを受け取り、テキスト化したデータを返すことが可能です。また、音声認識専用のユーザ辞書登録が可能です。



03

今すぐ解析デモを体験

「構文解析+照応解析」のデモを体験いただけます。
ぜひお試しください。

[デモを体験する](#)

まずは、どの部署、どの業務においても「理想、あるべき姿、何をしたいのか」つまり

解析

まずは、どの部署、どの業務においても「理想、あるべき姿、何をしたいのか」つまり「目的」を明確にすることが大事です。

この本を「目的を明確にするか否か」で理解度は格段に向上するはず。

