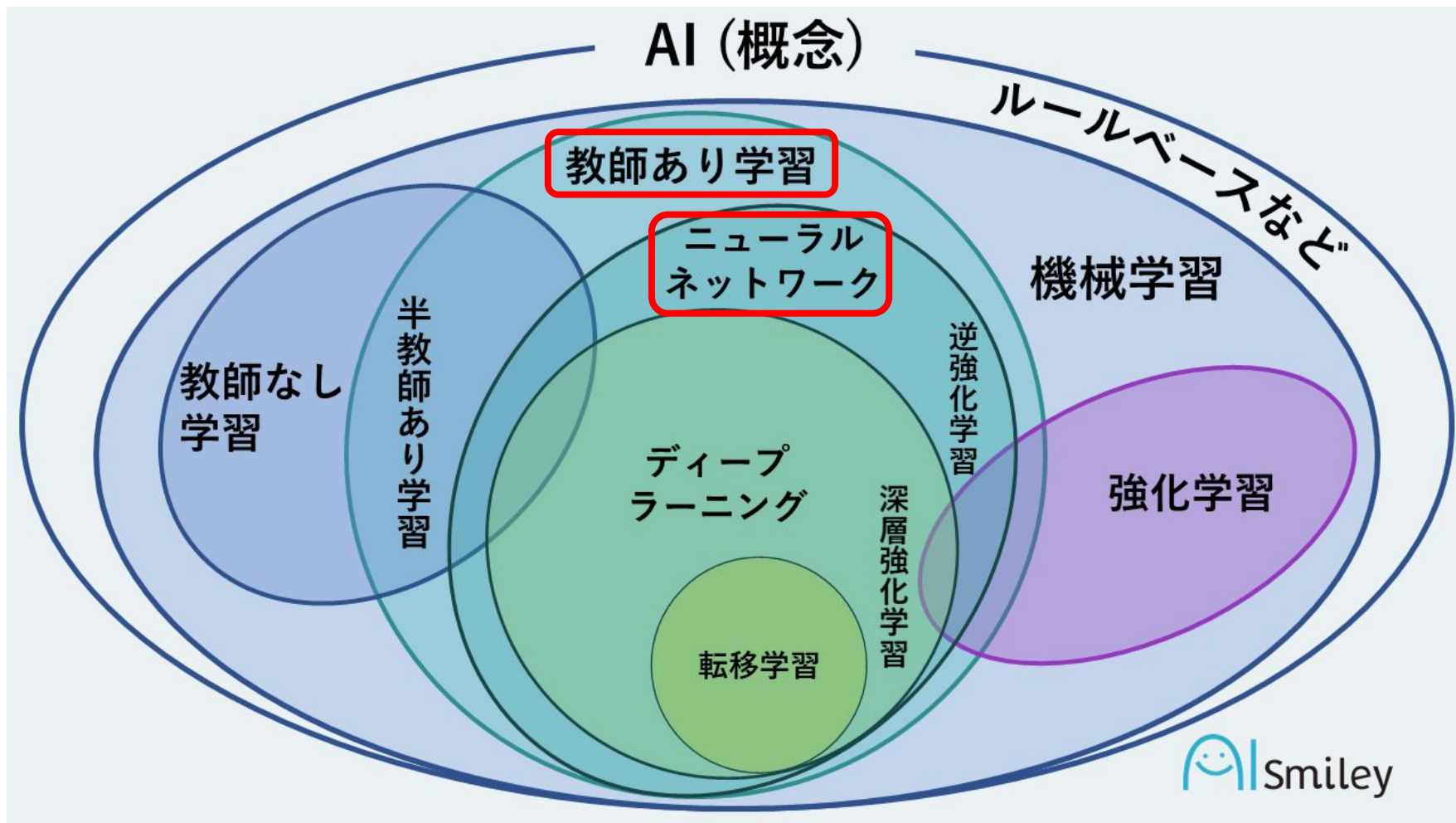


AIを活用した特許検索の現状

2023/7/22
Patentfield株式会社
近藤 和樹

AIと自然言語処理

AIの種類



出典: https://aismiley.co.jp/ai_news/what-is-the-difference-between-deep-learning-and-machine-learning/

自然言語処理とは

自然言語処理 (Natural Language Processing、NLP) は、人間が日常的に使用する言語をコンピューターが理解し、解析できるようにする技術の一つです。これにより、コンピューターは文章を読み取り、意味を把握し、応答する能力を得ます。例えば、検索エンジンの結果の改善、自動翻訳、音声認識などに広く利用されています。

コンピューターで扱える言語としては「プログラミング言語」がありますが、これらは、文法が定義されており、決まったルールに従ってコードを解析します。

一方、自然言語は、独特の「柔らかさ」があり、同じ意味の文章でもさまざまな表現が可能であったり、曖昧さもあります。また、時代の変化に応じて新しい言葉が生まれたりします。

この自然言語特有の「柔らかさ」をコンピューターに理解させる過程で、機械学習が広く利用されています。

また、日本語や中国語は単語の「わかち書き」をしないため、コンピューターに理解させる前段階として、単語の区切りを特定するのにテキストの解析(形態素解析)が必要となります。

英語: My name is Kondo.

日本語: 私の名前は近藤です。

形態素解析とは

形態素解析は、文書や文を単語や音節などの最小単位(形態素)に分割し、その品詞などを識別する処理です。自然言語処理の一環として、特に日本語など単語の境界が明確でない言語において重要な役割を果たします。

形態素解析の結果は、文書の理解や情報抽出、類似文書検索などに利用されます。

例えば、「庭には二羽ニワトリがいる」という文章を以下のように形態素に分解します。

庭(名詞)/
に(助詞)/
は(助詞)/
二(数詞)/
羽(助数詞)/
ニワトリ(名詞)/
が(助詞)/
いる(動詞)

形態素解析には、商用／オープンソースの解析エンジン等が広く利用されています。

- ・MeCab
- ・Janome

形態素(単語)のベクトル変換

形態素解析によって得られた形態素(単語)や文章そのものをコンピューターに処理させるためには、単語や文章を数値に変換する必要があります。

自然言語処理では、単語や文章をベクトル変換することで数値化して、単語や文章の意味や性質、類似度などを計算できるようにしています。

ベクトル変換には複数の手法がありますが、大きく分けると以下の2つです。

・Bag of Words

文章の中に出現する単語の出現数を用いて、文章をベクトル化する方法。文章の中にどれだけ単語が出現したのかに着目するため、文脈は考慮していません。

代表例:TF-IDF

・分散表現

単語が持つ意味に着目してベクトル化する方法。単語の意味をベクトル化することで、同じような意味に近いベクトルを与えることができ、単語同士の関係性もベクトルで表現できます。

代表例:Word2vec、BERT

TF-IDFの仕組み

TF-IDF (Term Frequency – Inverse Document Frequency) とは、文書中に含まれる単語の重要度を評価する手法の1つ

TF (Term Frequency)

ある文書における単語の出現頻度。

文書中に出現する頻度が多いほど、その単語は重要である可能性が高いと評価する。

$$\begin{aligned} \text{tf値}(\text{単語}_{i\text{番目}}, \text{文書}_{j\text{番目}}) &= \frac{\text{文書}_{j\text{番目}} \text{における単語}_{i\text{番目}} \text{の出現回数}}{\text{全文書に含まれる全ての単語の数}} \\ &\quad (i = 1 \cdots n, j = 1 \cdots m) \\ &= (\text{ある文書における、ある単語の出現頻度}) \end{aligned}$$

IDF (Inverse Document Frequency)

ある単語がいくつの文書で使われているかを表しています。

特定の文書にしか登場しないレアな単語であるほど、IDF値は高くなります。

$$\begin{aligned} \text{idf値}(\text{単語}_{i\text{番目}}) &= \log \left(\frac{\text{全文書の数}}{\text{単語}_{i\text{番目}} \text{が含まれる文書の数}} \right) \\ &\quad (i = 1 \cdots n) \\ &= (\text{ある単語の文書間でのレア度}) \end{aligned}$$

TF-IDFの仕組み

TF-IDF = TF(単語の出現頻度) * IDF(各単語のレア度)

「その単語がよく出現し」かつ「その単語がレアなほど」、TFIDFは大きい値を示します。
特定の文章に含まれる各単語のTF-IDF値を求めることで、その文章の特徴ベクトルを計算することができます。特徴ベクトルは、scikit-learnやTensorFlow等の機械学習ライブラリを利用することで計算できます。

最後に、各文書の特徴ベクトル同士の距離(コサイン距離)を計算することで、文書間の類似度(コサイン類似度)を求めることができます。

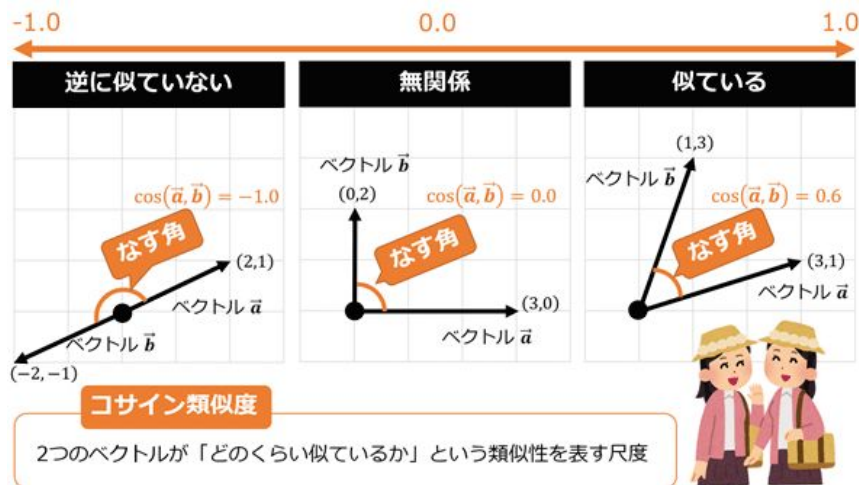
$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos\theta$$

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos(\mathbf{a}, \mathbf{b})$$

ベクトル \mathbf{a} とベクトル \mathbf{b} の内積 = ベクトル \mathbf{a} の L_2 ノルム ×

ベクトル \mathbf{b} の L_2 ノルム ×

コサイン値(ベクトル \mathbf{a} , ベクトル \mathbf{b})



出典: <https://atmarkit.itmedia.co.jp/ait/articles/2112/08/news020.html>

Word2vecにより単語の分散表現を取得

Word2vecは、「単語の意味は、その周辺に現れる単語によって形成される」という分布仮説をベースとし、各単語をベクトル変換して、その距離や角度により単語間の類似性を計算する手法です。

TFIDF値による特徴ベクトルのコサイン角度は、単語が完全に一致しなければ0。

＝類義語は反映してくれない！

例)「私は犬が好き。」と「僕、チワワが気に入っている。」の各文書は類似しない。

Word2vecを組み合わせることで、単語の類似性を考慮して文書の特徴ベクトルを生成することができ、単語の類義語を反映した類似文書の判定ができます。

Word2vecは、2つのニューラルネットワークモデルにより、単語の分散表現を獲得します。

・CBoW (continuous bag-of-words) モデル

周囲のコンテキストからターゲットの単語を予測する。

you say [?] and I say hello. *コンテキスト:say,and

・skip gramモデル

ターゲットを使用して、周囲のコンテキストを予測する。

you [?] goodbye [?] I say hello. *ターゲット:goodbye

Word2vecの弱点: 対義語が苦手。「最近野菜が高い／安い」のように対義語は同じ文脈で出現するので、分布仮説に基づいた手法では似通った分散表現になってしまいます。

AIと特許検索

Patentfieldの基本機能の紹介

4つの機能をワンパッケージ化

特許調査～分析に関するさまざまな課題にワンパッケージで対応！

4つの機能を組み合わせて、検索式作成～可視化・分析までの業務を最大80%効率化！



機能紹介A

誰でも使いやすい
特許検索



機能紹介B

手間いらずの可視化



機能紹介C

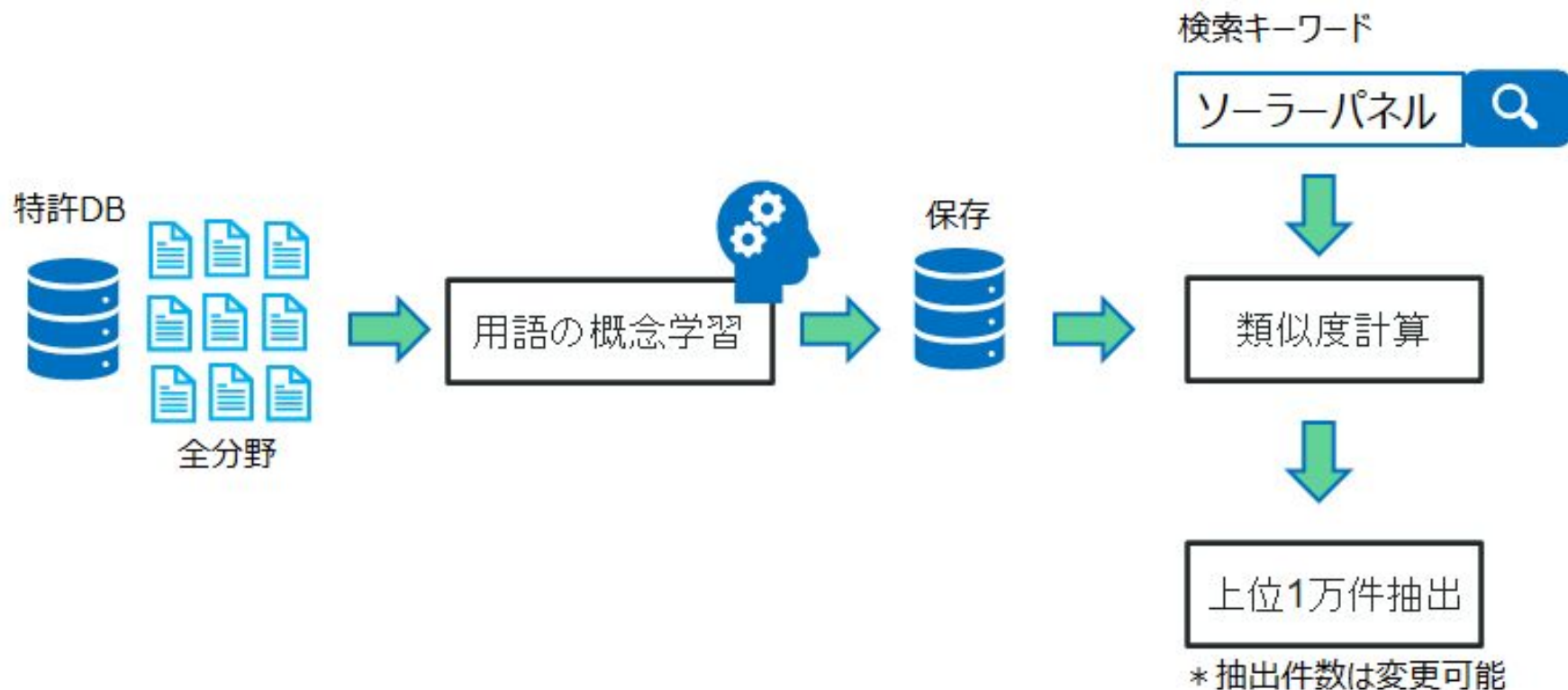
AIが類似を探す
セマンティック検索



機能紹介D

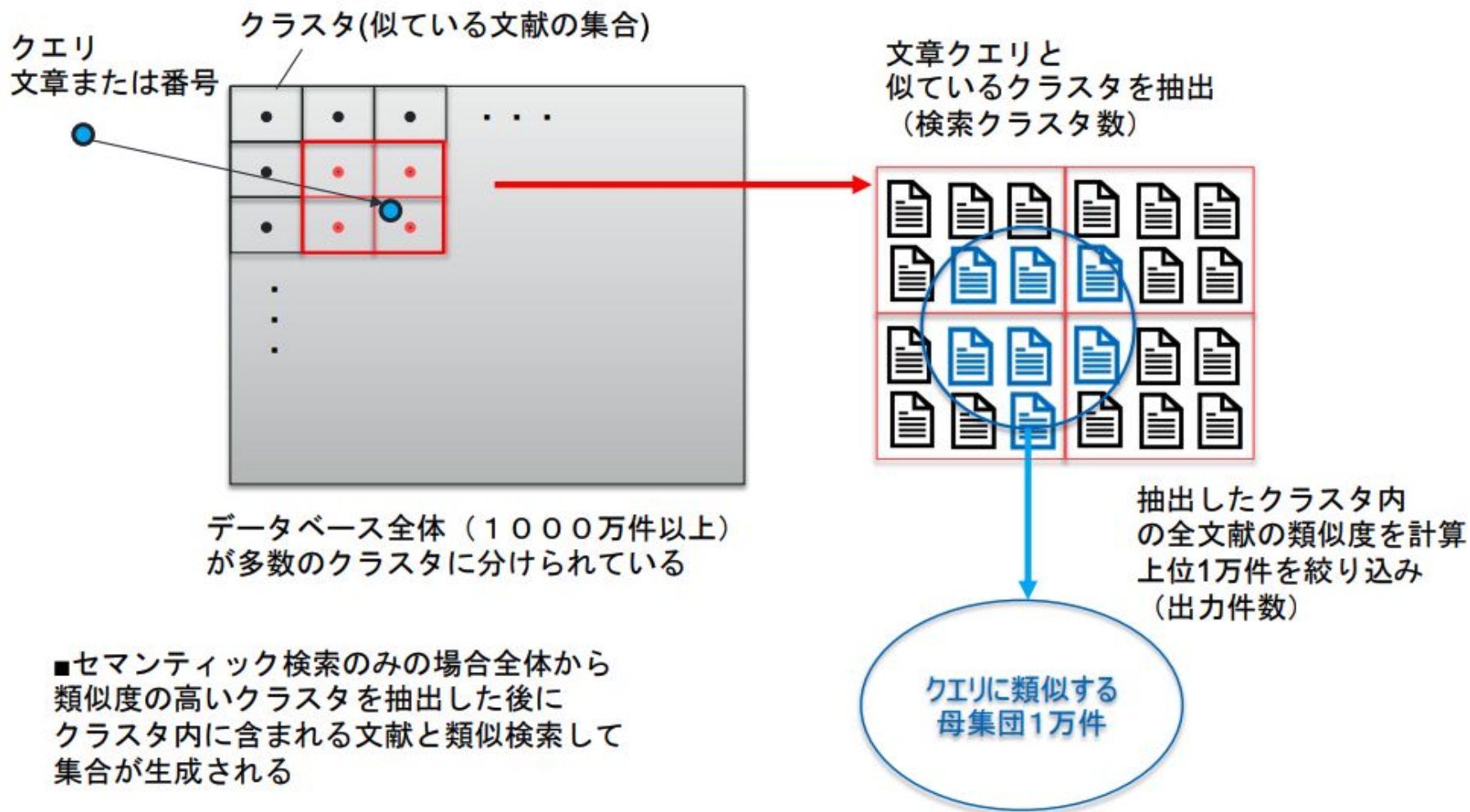
AIがすぐ振り分け
分類予測

AIセマンティック検索の概要



事前に、収録済みの数千万件の特許文書全体から「用語の概念学習」を行っています。
AIセマンティック検索では、検索キーワード(ソーラーパネル) + 類似概念(太陽光パネル、太陽発電パネルetc)で網羅的に文章の類似度計算を行い、収録済みの数千万件の特許全体から、類似する上位n件を抽出しています。

AIセマンティック検索の処理の流れ(検索フェーズ)



■セマンティック検索のみの場合全体から類似度の高いクラスタを抽出した後にクラスタ内に含まれる文献と類似検索して集合が生成される

AIセマンティック検索 テキストによる類似検索

日本語

タイトル/要約/請求の範囲/明細書/審査官キーワード ▾

★ セマンティック検索オプションを表示 ⚙

下腿の正常な歩行動作が困難なユーザーに対しても、歩行状態に応じた適切な下腿への歩行補助力を付与し得るアクチュエータ付き長下肢装具を提供する。本発明に係る長下肢装具においては、制御装置は、一のサンプリングポイントにて状態姿勢検出手段から入力される角度関連信号に基づいて大腿位相角を算出し、当該制御装置に予め記憶されている、大腿位相角と下腿側装具に付与すべき補助力との関係を示す補助力制御データに前記一のサンプリングポイントでの大腿位相角を適用して、前記一のサンプリングポイントにおいて前記下腿側装具に付与すべき補助力を算出し、前記補助力が出力されるように前記アクチュエータの作動制御を実行する。

英語

タイトル/要約/請求の範囲/明細書/審査官キーワード ▾

★ セマンティック検索オプションを表示 ⚙

The present invention provides an actuator-equipped knee ankle foot orthosis in which a control device calculates a thigh phase angle based on an angle-related signal detected by a thigh orientation detecting means at one sampling point, applies the thigh phase angle at that sampling point to an assisting force control data, which is stored in the control device in advance and indicates the relationship between the thigh phase angle and a size of the assisting force to be imparted to a lower leg-side brace, to obtain the size of the assisting force to be imparted to the lower leg-side brace at that sampling point, and executes operational control for an actuator unit such that the assisting force having the size is output.

AIセマンティック検索は、日本語／英語の入力に対応しています。

日本語で検索：日本語で記載された類似公報を検索 (JP/WOJP)

英語で検索：英語で記載された類似公報を検索 (US/EP/WO)

* 日本語横断検索オプションでは、日本語でJP/US/EP/WO/CN/KRの類似公報の検索が可能

AIセマンティック検索 公報番号による類似検索

公報を1件入力する場合の入力例

タイトル/要約/請求の範囲/明細書/審査官キーワード ▾ ★ セマンティック検索オプションを表示 ⚙

特開2017-145243

pub_id:JP2017145243

公報を複数件入力する場合の入力例(どちらでも可)

JP2017145243 JP2017213246 WO2020195447

pub_id:JP2017145243 pub_id:JP2017213246 pub_id:WO2020195447

AIセマンティック検索では、公報番号を入力して類似検索ができます。

外国公報の入力も可能です。

公報番号を入力した場合は、入力公報の「全文」テキストを利用して類似検索をします。

公報番号を入力すると、該当候補が赤枠サジェストに表示されるので、選択すると入力完了です。

公報番号は複数件入力して検索することもできます。

複数公報を入力した場合は、複数公報のすべての全文テキストを合成して、類似検索します。

AIセマンティック検索 対象特徴量

The screenshot displays the AI Semantic Search interface. On the left, a dropdown menu lists various target features (特徴量) for search. The selected feature is "タイトル/要約/請求の範囲/明細書/審査官キーワード". The main search area shows a search bar with the text "とができる太陽光発電パネル水上設置ユニットを提供する". Below the search bar, there is a table with columns for "フィルター(簡易)", "集計", and "スコア". The table is currently empty. A note at the bottom of the search area states: "出現するキーワードを入力して下さい。" and "論理式を設定できます。 例：(L:1 OR L:2) AND L:3".

タイトル/要約/請求の範囲/明細書/審査官キーワード

セマンティック検索オプションを表示

とができる太陽光発電パネル水上設置ユニットを提供する

出現するキーワードを入力して下さい。

フィルター(簡易)	集計	スコア
-----------	----	-----

論理式を設定できます。 例：(L:1 OR L:2) AND L:3

公報の所定の項目ごとに特徴量を保持し、クラスタリングされています。
ユーザーは、任意の対象特徴量を選択して、類似検索が可能です。

AIセマンティック検索 特徴キーワード

特開2010-273680の特徴キーワード

AH	
1	特徴キーワード(タイトル/要約/請求の範囲/明細書/審査官キーワード)
	{ "MLOXP"=>40.80072, "エピゾーマルベクタ"=>33.56473, "PMXS"=>33.26043, "CREリコンビナーゼ"=>28.83056, "IPS細胞"=>27.55997, "シストロニック"=>25.29804, "OKS"=>25.26369, "LOXP"=>24.24025, "複製開始点"=>23.90792, "NANOG"=>23.89254, "効率改善"=>23.77097, "核初期化因子"=>23.1879, "サザンプロット"=>22.75378, "外来遺伝子"=>22.65265, "樹立"=>22.54192, "レンチウイルス"=>22.16345, "遺伝子導入"=>21.501, "体細胞"=>21.46591, "エピゾーマル"=>21.02227, "CRE"=>20.40215, "SOX"=>20.37026, "MUTANT LOXP"=>20.3128, "マウス由来"=>20.28903, "ESRRB"=>20.25282, "レポータ"=>20.20734, "レトロウイルス"=>20.13776, "ES細胞"=>20.08963, "蒔い"=>19.89043, "自律複製"=>19.81524, "挿入変異"=>19.63149, "LOX"=>19.61423, "LTR"=>19.58419, "SPHI"=>19.39185, "染色体"=>19.15149, "コンストラクト"=>19.07839, "発現ベクタ"=>19.05963, "ピユーロマイシン"=>18.89725, "MSTO"=>18.85407, "MYC"=>18.70915, "プラスミドベクタ"=>18.60364, "キメラマウス"=>18.50517, "エピソーム"=>18.48899, "FBX"=>18.40644, "IPS"=>18.30564, "変異型"=>18.26758, "核初期化物質"=>18.26476, "コロニ"=>18.23507, "クローン"=>18.10747, "口蹄疫ウイルス"=>18.01991, "GENMIC"=>18.00636, "ベクタ"=>17.94963, "MEF"=>17.7813, "プロモータ"=>17.60747, "FLOXED"=>17.26081, "ESRRG"=>17.12731, "反復配列"=>16.97445, "SV40"=>16.9343, "遺伝子"=>16.91741, "変異"=>16.36605, "レトロウイルスベクタ"=>16.07476, "発現阻害剤"=>16.04145, "ウイルスベクタ"=>16.03958, "複製"=>15.8584, "分化"=>15.83992, "上皮細胞"=>15.76008, "選択マーカー遺伝子"=>15.73253, "陽性"=>15.65267, "肝細胞"=>15.63846, "KOS"=>15.49578, "LIN"=>15.4236, "KL"=>15.25601, "アデノウイルス"=>15.07078, "プラスミド"=>14.93593, "ORIP"=>14.88988, "エンハンサ"=>14.88173, "ゲノム"=>14.86628, "-"=>14.85723, "GENOMIC"=>14.75213, "組み換え"=>14.71567, "GFP"=>14.69699, "切り出し"=>14.61564, "CMYC"=>14.50206, "ECORI"=>14.45046, "マイクロインジェクション法"=>14.44097, "細胞"=>14.42784, "フィーダー細胞"=>14.32158, "耐性遺伝子"=>14.25758, "プローブ"=>14.248, "因子"=>14.24774, "STEM CELL"=>14.02154, "継"=>14.00105, "BAMHI"=>13.98646, "HDAC"=>13.95648, "幹細胞"=>13.95048, "プライマ"=>13.9422, "培地"=>13.86226, "OKITA K"=>13.86172, "側プローブ"=>13.83894, "HPV16"=>13.77627, "遺伝子座"=>13.689, "F4"=>13.60325, "タンパク質"=>13.49371, "組換え"=>13.45728, "OCT"=>13.4013, "AAMSTO"=>13.31569, "ハイグロマイシン"=>13.28502, "培養"=>13.16882, "PULSER XCELL"=>13.16625, "野生型"=>13.16616, "該初"=>13.10231, "ゲラチン"=>13.05198, "線維芽細胞"=>13.02091, "培養プレート"=>13.00095, "KLF"=>12.92826, "PCAG"=>12.90811, "核酸"=>12.90657, "PTD"=>12.87969, "PCR"=>12.8419, "マウス"=>12.83325, "ウイルス感染"=>12.82652, "写真"=>12.82342, "GENOMIC DNA"=>12.80267, "SHRNA"=>12.79366, "バンド"=>12.76429, "置換可能"=>12.60093, "発現"=>12.56188, "前駆"=>12.56181, "遺伝子導入効率"=>12.55958, "サザンプロット"=>12.51882, "EBNA"=>12.51543, "スペーサー部分"=>12.50624, "ポリA"=>12.41582, "融合タンパク質"=>12.32557, "解析"=>12.30011, "OKITA"=>12.26053, "ヒト"=>12.24444, "欠失"=>12.21665, "胎仔"=>12.12399, "SIRNA"=>12.06942, "非特許文献"=>12.03439, "付加シグナル"=>12.00517, "性因子"=>11.98172, "一過

各対象特徴量の文ベクトルは、各公報から抽出されたトークン(単語)に対して、各単語の重要度(例:TF-IDF値)を算出しています。そのトークンと重要度を特徴量として文ベクトルに変換しています。

AIセマンティック検索 特徴キーワード

特開2010-273680の特徴キーワード

AG

特徴キーワード(請求の範囲)

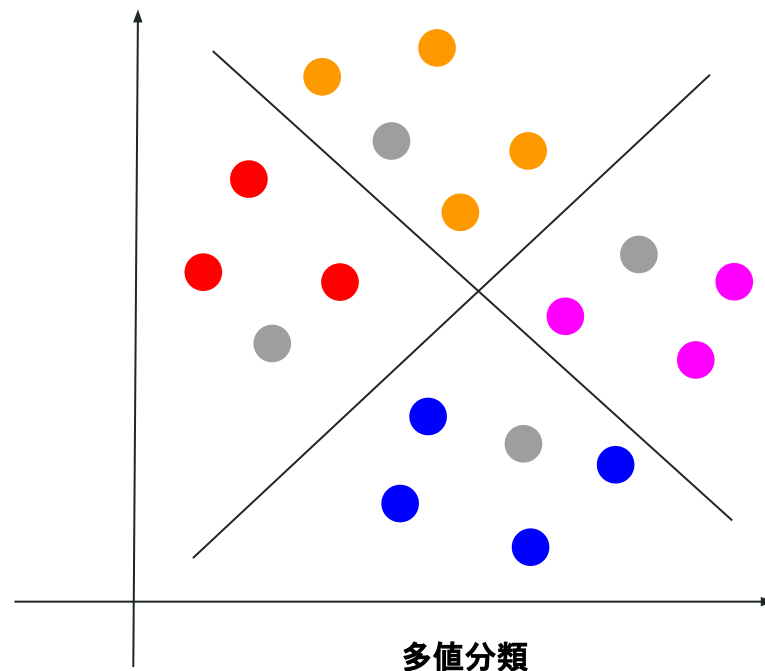
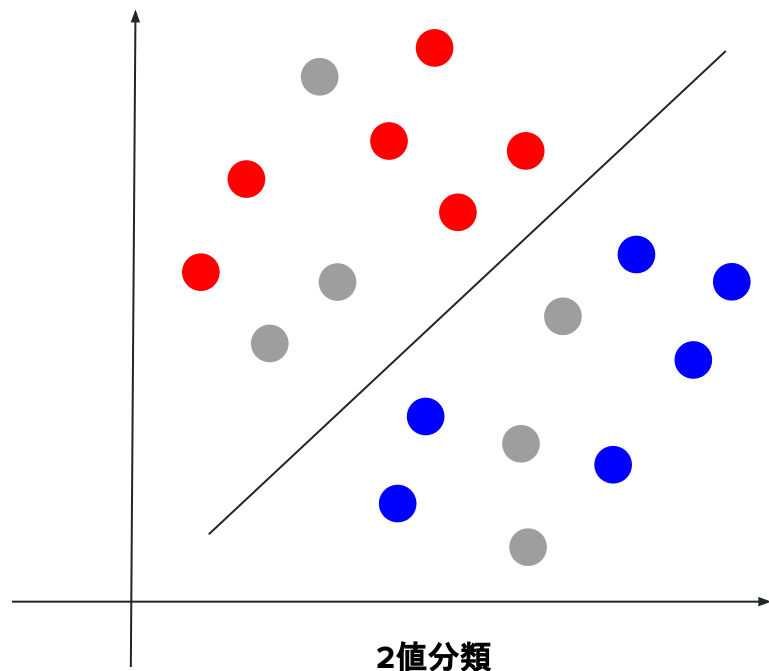
{ "GSCLC"=>45.3674, "精細管"=>37.06875, "PGCLC"=>35.64718, "PLZF"=>33.12561, "GSC"=>32.70486, "PIW"=>29.78941, "精巢"=>27.70647, "生殖巢"=>25.75687, "精子"=>25.56733, "GFR"=>25.28537, "始原生殖細胞"=>25.10976, "授精"=>25.07438, "DDX"=>24.80108, "卵細胞"=>23.96121, "稔性"=>23.90393, "PSC"=>23.77387, "GFRA"=>23.64655, "DAZ"=>23.63564, "幹細胞"=>23.46727, "性細胞"=>23.28016, "表面マーカー"=>22.94326, "精"=>22.48874, "陽"=>22.40737, "移植"=>22.29303, "子孫"=>21.89853, "浮遊培養"=>21.73421, "成体"=>21.09545, "気液界面"=>20.80224, "受精"=>20.74996, "再構成"=>20.52734, "ITGA"=>20.39541, "共培養"=>20.15866, "INTEGRIN"=>19.63717, "多能性幹細胞"=>19.05396, "単離"=>18.96942, "IL4"=>18.91867, "SSEA"=>18.65411, "転写因子"=>18.54219, "生殖"=>18.37259, "顕微"=>18.36896, "細胞"=>18.21165, "様"=>17.96328, "体細胞"=>17.88805, "SCP"=>17.56134, "同等"=>17.41423, "RET"=>17.20411, "インビトロ"=>17.06585, "発現レベル"=>17.02904, "IL2"=>16.56252, "陽性"=>16.45065, "占める"=>16.20276, "定着"=>15.94257, "解離"=>15.80089, "全身"=>15.7987, "培養"=>15.75454, "寄与"=>15.17994, "系列"=>15.11998, "性質"=>14.64376, "誘導"=>14.59643, "哺乳動物"=>14.38255, "増殖"=>14.19214, "由来"=>13.87279, "遺伝子"=>13.49537, "KIT"=>12.89859, "割合"=>12.66556, "正常"=>12.39816, "ID"=>12.31399, "増幅"=>12.07103, "条件下"=>11.99767, "CD"=>11.2579, "得"=>11.18088, "群"=>10.99368, "得る"=>10.76955, "生じる"=>10.68686, "それ"=>10.51628, "維持"=>10.17942, "速度

AG

特徴キーワード(要約)

{ "始原生殖細胞"=>33.61906, "精巢"=>33.40445, "PLZF"=>30.21766, "精子"=>29.26784, "生殖巢"=>28.67244, "生殖"=>28.3126, "DDX"=>26.39011, "幹細胞"=>26.32233, "浮遊培養"=>24.42151, "共培養"=>23.97709, "気液界面"=>23.26675, "再構成"=>22.39445, "多能性幹細胞"=>21.16231, "体細胞"=>21.13167, "性細胞"=>20.85154, "系列"=>20.57623, "細胞"=>20.43431, "様"=>20.19355, "陽性"=>19.9932, "インビトロ"=>19.39519, "培養"=>19.19629, "陽"=>18.66537, "解離"=>17.59411, "誘導"=>16.99866, "単離"=>16.21394, "条件下"=>15.2226, "由来"=>14.48764, "工程"=>12.85835, "得"=>12.05217, "得る"=>11.18561, "方法"=>10.8763, "当該"=>10.32078, "製造"=>10.17186, "含む"=>10.03782, "製造方法"=>9.882659, "られる"=>8.913322, "該"=>8.889297, "中"=>8.64298, "本発明"=>8.388892, "られ"=>7.762846, "し"=>6.732034, "する"=>6.225091, "提供"=>5.611886, "れ"=>5.581204, "こと"=>5.387112, "さ"=>5.003911, "要約"=>4.66154}

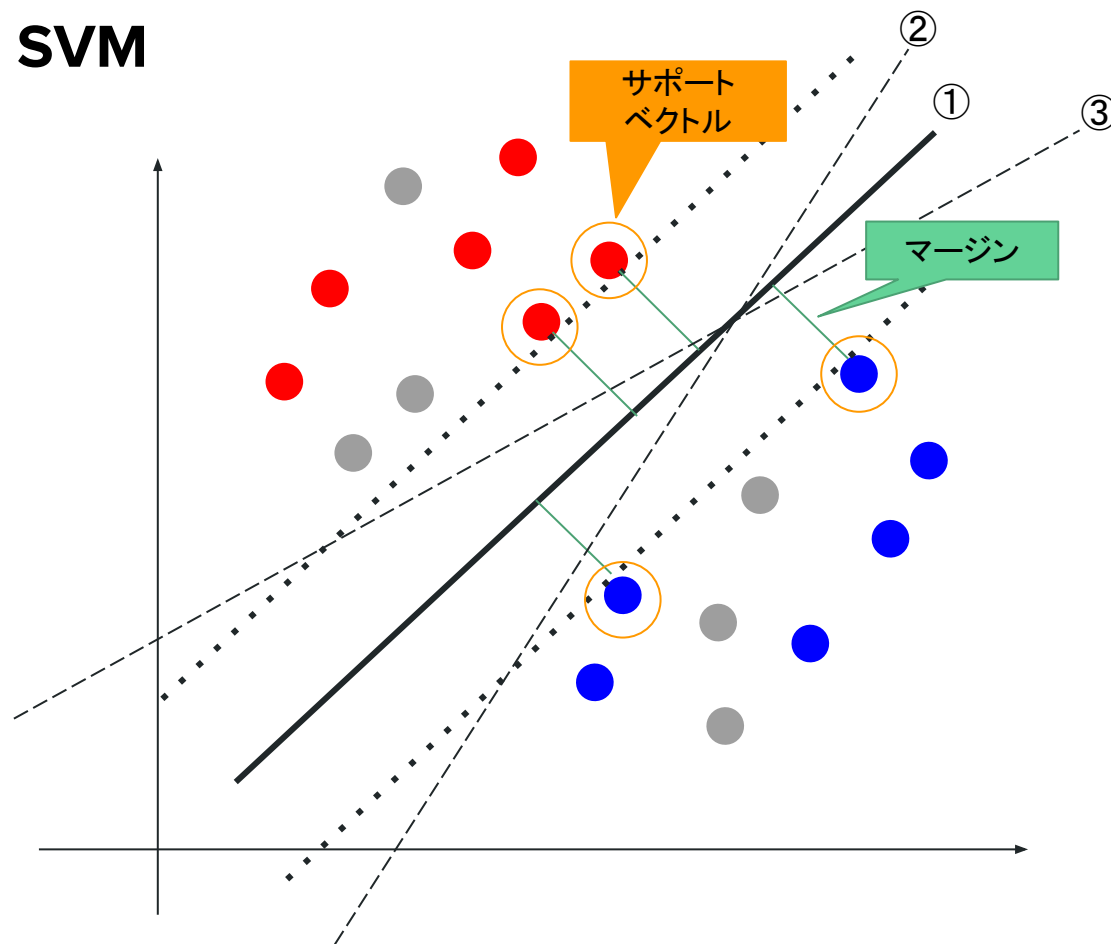
AI分類予測の概要



ベクトル空間上にある各公報の文ベクトルデータに対して、事前に設定した教師データ(色付き)に基づいて、未知の公報(灰色)に対して、SVM(線形分類器)によりデータ間の境界を決定して、データを分類します。また、各ベクトルデータと境界線との分離平面からの距離を正規化することにより、分類予測スコアを算出しています。

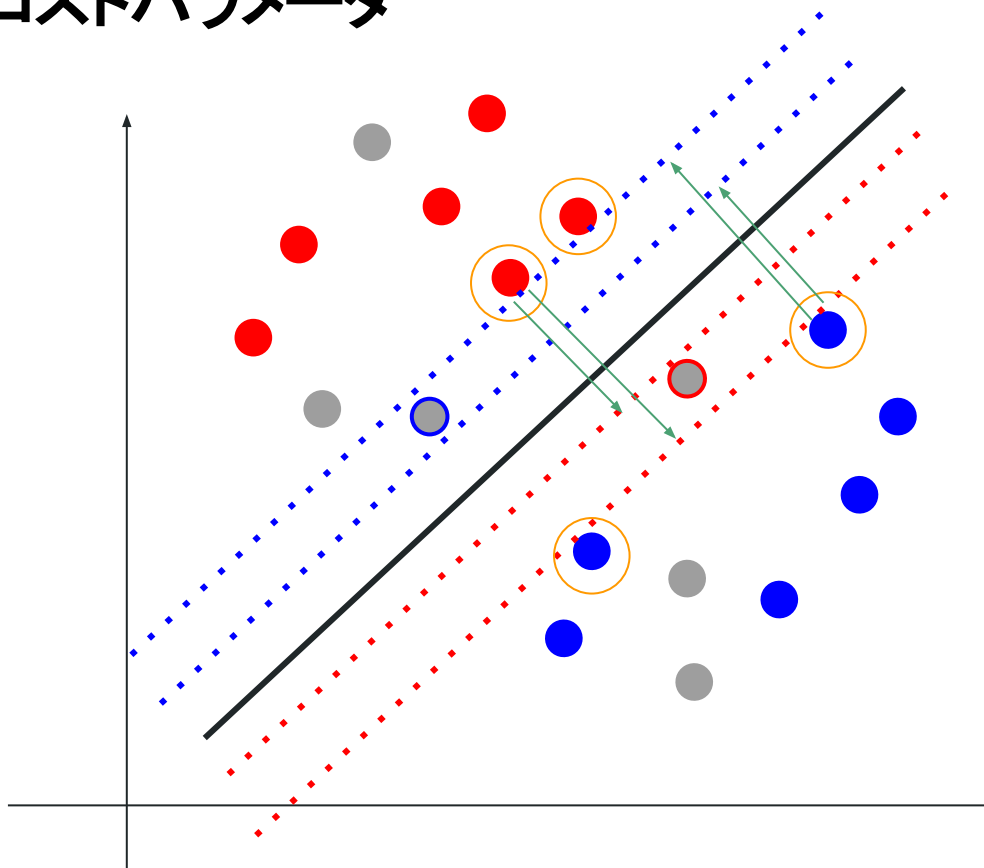
2値分類:-1から1 多値分類:0から1

AI分類予測 SVM



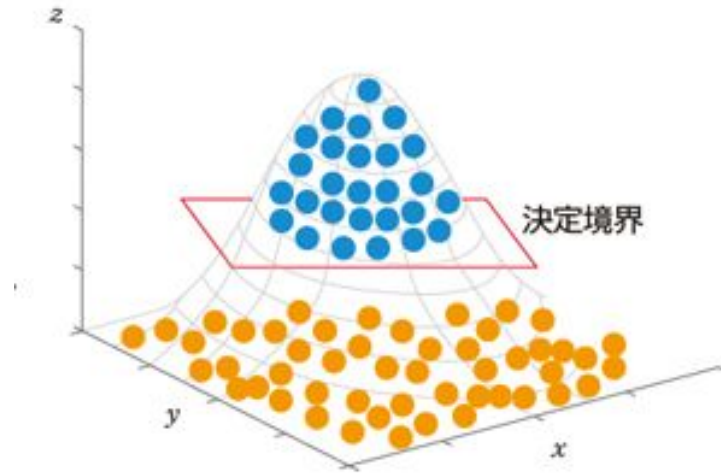
SVMは、境界付近にあるデータ(サポートベクトル)を用いてマージンが最大になるような境界線を決
定し、データの分類を行います。
サポートベクトルに対してマージンを最大化できるポイントで境界線を引くことで、境界付近に存在する
未知のデータの誤分類を防ぐことができます。

AI分類予測 コストパラメータ

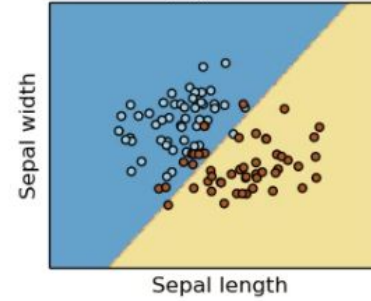


実際のデータには、線形分離不可能なデータも存在します。その場合に、どれくらい誤分類を許容するかをコストパラメータ C を用いて設定します。 C の値が小さいほど誤分類を許容し、大きいほど誤分類を許容しなくなります。一般的には、 C が小さいほど未知のデータに対する汎化性が向上すると言われています。

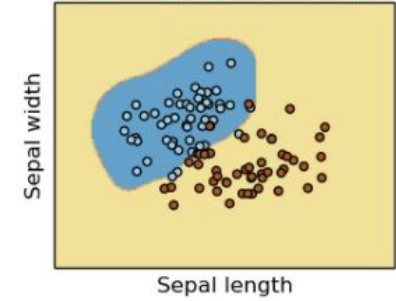
AI分類予測 カーネルパラメータ



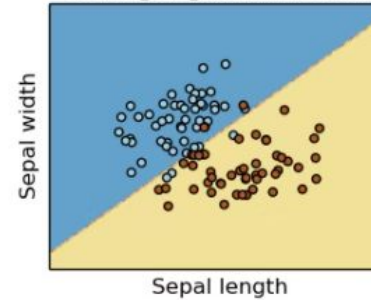
C: small, gamma: small



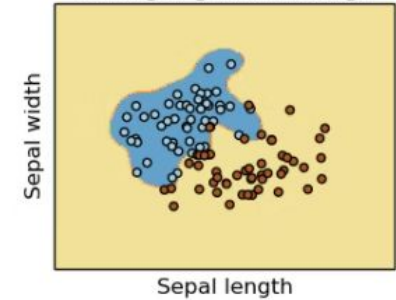
C: small, gamma: large



C: large, gamma: small



C: large, gamma: large



出典: https://qiita.com/sz_dr/items/f3d6630137b184156a67

カーネル法を利用することで、SVMの線形アルゴリズムを非線形化して複雑な境界決定ができます。カーネル法では、カーネル関数を利用することで、高次元空間で学習した決定境界を利用して、低次元空間で非線形な境界を得ることができます。

カーネルパラメータが大きいほどより複雑な境界になり、小さいほどシンプルになります。

AI分類予測 対象特徴量

AIラベル設定

プレ予測モデル

対象特徴量

教師 7 重み

タイトル/要約/請求の範囲/明細書/審査官キーワード

タイトル/要約/請求の範囲

請求の範囲

請求の範囲 トップクレーム

可視化

予測します。 [マニュアル](#)

AIセマンティック検索と同様、AI分類予測でも対象特徴量の選択が可能です。

(セマンティック)の項目では、特徴キーワード+類似キーワードを含めて分類予測を行います。

AI分類予測 パラメータチューニング

教師データ	予測データ	可視化	検索結果の分類設定	学習パラメータチューニング
-------	-------	-----	-----------	---------------

AIのチューニングを行います。

チューニングに使用するデータ

教師データの一部 予測データ

チューニング対象特微量(最大10)

タイトル/要約/請求の範囲/明細書/審査官キーワード(セマンティック) タイトル/要約/請求の範囲(セマンティック)
技術分野/背景技術/課題/効果/解決手段(P)(セマンティック) タイトル/要約/請求の範囲/明細書/審査官キーワード
タイトル/要約/請求の範囲 技術分野/背景技術/課題/効果/解決手段(P)

- C: コスト(誤分類許容度)
- Sigma: カーネルパラメータ
- 2 値優先ラベル

直近1週間の実行回数 3 回

この機能はサーバー負荷が高いため、1週間に実行できる回数が制限されています。
高頻度でのチューニングが必要な場合はお問合せ下さい。

チューニング開始

最適パラメータを反映

チェックしたパラメータを反映

最適パラメータ

精度: 0.74 対象特微量: タイトル/要約/請求の範囲 C: 0.25

- 精度: 0.73 対象特微量: タイトル/要約/請求の範囲/明細書/審査官キーワード(セマンティック) C: 50 Sigma: 5
- 精度: 0.701 対象特微量: タイトル/要約/請求の範囲(セマンティック) C: 100 Sigma: 1
- 精度: 0.701 対象特微量: 技術分野/背景技術/課題/効果/解決手段(P)(セマンティック) C: 0.5 Sigma: 5
- 精度: 0.701 対象特微量: タイトル/要約/請求の範囲/明細書/審査官キーワード C: 0.0001
- 精度: 0.74 対象特微量: タイトル/要約/請求の範囲 C: 0.25
- 精度: 0.721 対象特微量: 技術分野/背景技術/課題/効果/解決手段(P) C: 0.25

AI分類予測で利用する**パラメータ及び対象特微量の最適値を自動で探索**し、その最適値を利用することで、予測精度の向上を図ることができます。

PatentfieldのAI活用事例

1. 類義語の入力サポート

「類似キーワード」「類似キーワード拡張」で、日本語/英語の類義語入力をサポート

2. 出願前調査

AIセマンティック検索で特許分類を選定
検索式母集団に対して「関連順」で類似順にソート

3. 国内外の無効調査

無効化したい特許の公報番号を入力して簡単類似検索
タグ利用による重複調査を排除した調査効率化

4. 開発者自身で特許検索

開発者自身で開発テーマの先行技術調査
教師データを利用して調査効率化

5. 手軽に技術動向分析

プレスト会議等で、
その場ですぐにマップを作成して議論を活性化

6. SDIのノイズ除去

毎月の確認件数を大幅削減
開発部門にも素早く共有

7. 社内分類付与

AIが社内分類を自動付与
作業負担を大幅削減

8. 社内分類でマップ作成

社内分類を軸にマップ作成
AI予測ラベルで競合分析等も可能

9. 関連KWを利用した用途探索

特徴キーワードで、保有技術の多用途展開
の可能性をさぐる

10. 特徴KWを利用した2社分析

特徴キーワードで、自社・競合企業の
強み/弱みを可視化分析

1.類義語の入力サポート

AIレベル設定

詳細検索

保存条件 一覧 保存 履歴 さがす **類似キーワード** 類似分類 分類検索 ヘルプ

+ 保存条件をセット + 番号を一括入力

キーワードを入力してください

キーワードを入力してください	
	ソーラーパネル
1	ソーラパネル
2	太陽電池パネル
3	太陽光パネル
4	太陽光発電パネル
5	発電パネル
6	太陽発電

キーワードを入力してください

キーワードを入力してください	
	"solar panel"
1	SOLAR PANEL
2	SOLAR PANELS
3	PHOTOVOLTAIC PANELS
4	SOLAR PHOTOVOLTAIC
5	SOLAR MODULES
6	PHOTOVOLTAIC MODULES

- ・類似キーワード機能では、入力したキーワードに対して、PatentfieldのAIが学習済みの類義語（類似キーワード）がリスト表示されます。（リストの上位にあるほど類似度が高いことを意味します）
- ・日本語/英語に対応しており、日本語を入力すると日本語の類義語／英語で入力すると英語の類義語が表示されます。
- ・キーワード検索時の類義語調査の効率化/漏れ防止が図れます。

2.出願前調査（AIセマンティック検索で特許分類を選定）

ユーザが視聴したコンテンツとユーザが所持する端末とを効果的にマッチングするマッチングシステム

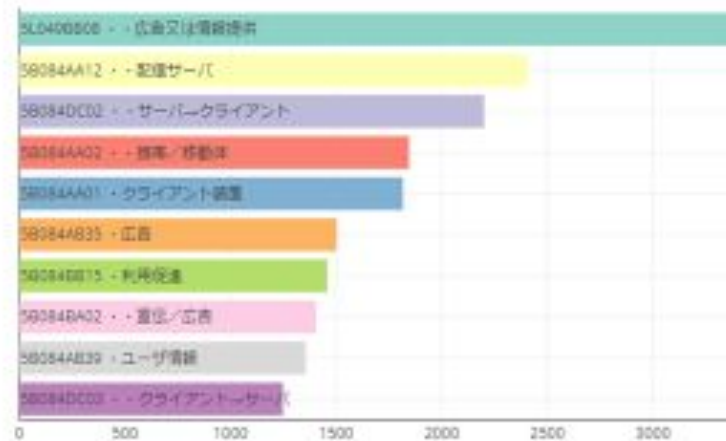


AIセマンティック検索でプレ検索（調査テーマの関連特許を検索）

IPC



Fターム



- AIセマンティック検索でプレ検索を行い、調査テーマの関連特許をざっくり検索する
- 検索結果1万件の関連特許に対して、マップ機能で関連特許群の特許分類を可視化
- 関連特許群1万件に付与されている特許分類ランキングで、本番検索時の特許分類を選定
- **新規分野等での検索実行時の特許分類選定を効率化**

2.出願前調査（関連順で再ソート）

フィルター(詳細)	▼ フィルター(簡易)	集計	スコア	フィルター(スコア/ラベル)
1	AND ▼ 名称/要約/請求の範囲/特許書/審査官フリーワー			広告配信部 OR 広告決定部 OR 広告提出部 OR 広告記憶部 OR 広告送信部
2	AND ▼ Fターム			5L0498B08
3	AND ▼ F			G06Q30/02380

+ フィルター条件を追加



10 1 スコア 0.881



マッチングシステム

【目的】 【課題】 ユーザが検索したコンテンツとユーザが所持する権利とを効果的にマッチングするマッチングシステムを提供する。【解決手段】 実施形態によれば、権利とコンテンツとをマッチングするマッチングシステムは、記憶部と、第1の受信処理部と、第1の受信処理部と、コンテンツ特定部と、マッチング部と、を備える。記憶部は、音を含むコンテンツの開始日時を保持する。第1の受信処理部は、前記開始日時に基づくタイミングで、権利に、前記権利の外部の音の取得を開始させる通知を送信する。第1の受信処理部は、前記権利から、外部の音を受信する。コンテンツ特定部は、前記音情報に基づくコンテンツを特定する。マッチング部は、特定された前記コンテンツと、前記権利とをマッチングする。【選択図】 図1

● 特許出願公開係属中

出願番号	公開日	公開番号	公開日	早期審査 (JP・US) 0
JP2020171890A	2020-10-12	JP2021015187	2021-02-04	
特許庁件数 (JP・US) 0		引続件数 5		代理人/特許事務所 徳業特許法律事務所
出願人 デジタル・アドバタイジング・コンソーシアム		発明者 徳久 昭彦, 杉田 光雄		

AIスコア ▲ UP ▼ DOWN

検索履歴 < 関連特許 特許利用マップ

- ・検索式作成後に検索を実行（277件ヒット）し、調査テーマに一番近い10番目の公報の「関連順」をクリック
- ・検索母集団277件が、「関連順」をクリックした10番目の公報の類似順（関連順）に再ソートされます。
- ・検索母集団が類似順に表示されるため、スクリーニングの効率化が図れます。

3.国内外での網羅的な無効調査

AIセマンティック検索&タグ機能を利用した国内外の無効調査の効率化



- ・無効化したい特許の公報番号を入力して、当該公報に類似する特許を検索
- ・セカンドオピニオン/本格調査前の予調査として利用
- ・日本語横断オプションでは、JPの公報番号を入力して、JP/US/EP/CN/KR/WOを対象に、国内外を横断して類似特許の検索が可能
- ・タグ機能を利用して、「確認済」等を記録しておけば重複チェックを予防して調査を効率化
- ・作業グループ共有オプションでは、タグを共有してチーム内で調査結果をリアルタイムに共有可能
チームで無効調査を行う際に、チーム間での重複調査を防止

4.開発者自身で特許検索&分析

短い文章やキーワードを使って3STEP関連特許検索



【STEP 1】

- 調べたいテーマに関する思いついたキーワードや簡単な文章で検索

【STEP 2】

- 検索結果の上位10~20件から、調査テーマに最も近い特許を1件見つける
- その特許の「関連特許」をクリックすると、その特許公報の全文を使用してAIセマンティック検索が実行される

【STEP 3】

- 検索結果の上位10~20件から、調査テーマに最も近い特許を1件見つける
- その特許の「関連順」をクリックすると、その特許に類似する順に検索結果がソートされる

- 最初に入力した検索文章に依存せず、3STEP操作で調査テーマの関連特許を検索
- 知財部に聞かなくても、「とりあえず自分で簡単に調べてみる」ことで社内の知財活動を活性化

5.手軽に技術動向分析

3STEP関連特許検索の結果で手軽に技術動向分析



- これまでの技術動向分析は、分析対象の母集合作成+マップ作成に膨大な労力がかかる
- 3STEP関連特許検索の結果を利用して、マップの軸を選択するだけで簡単にパテントマップを作成可能
- 開発企画のプレスト会議等で、その場ですぐに様々なマップ作成が可能となり、知財情報をミックスして議論を活性化

6.SDIのノイズ除去

AI分類予測「2値分類」を利用したSDIのノイズ除去



↑ 検索システムに取り込んで、SDIの内容をチェック

Patentfield

SDIメール配信



AI分類予測「2値分類」で、SDIのノイズをAIが判定



AI予測結果のエクセルDL

J	K	L	M	N	O
IPC	テーマコード	リーガルステータス (JPN/PADO/C)	審査記録種別コード (JPN)	AI予測スコア	AI予測レベル
C12M5/071 C12M5/077			A99	0.535692048	1.0
C12M5/0783 C12M5/10			A99	0.48019012	1.0
C12M5/00 C12Q1/6664					
C12Q1/04 C12M15/01	46029				
C12N15/10 C12M1/04	46063	実開中	A99	-0.329073787	-1.0
C12M1/20 C12Q1/6836					

PF上でSDI内容をチェック

- ・過去1年分(目安)のSDI結果を教師データとしてセットして、AIが毎月のSDIのノイズを判定
- ・他社検索システムのSDI公報リストをインポートしてAI予測も可能。予測結果のエクセルDLにも対応
- ・ノイズを除去した状態でSDI内容をチェックすることにより、チェック件数を削減して業務を効率化

Patentfield

7.社内分類付与

AI分類予測「多値分類」「多ラベル分類」を利用した社内分類の自動付与



↑ 検索システムに取り込んで、付与内容をチェック



AI分類予測で、最適な社内分類をAIが予測付与

特許出願公開非絶査定

出願番号	出願日	公開番号	公開日
JP2005142226A	2005-05-16	JP2006318341	2006-11-24
特引用件数 (JP - US) 8		引用件数 2	
出願人		発明者	代理人
ソニー		竹内 英人	は

AIラベル: 予測 | 過半数禁止: 0.535

PF上で付与内容をチェック

AI予測結果のエクセルDL

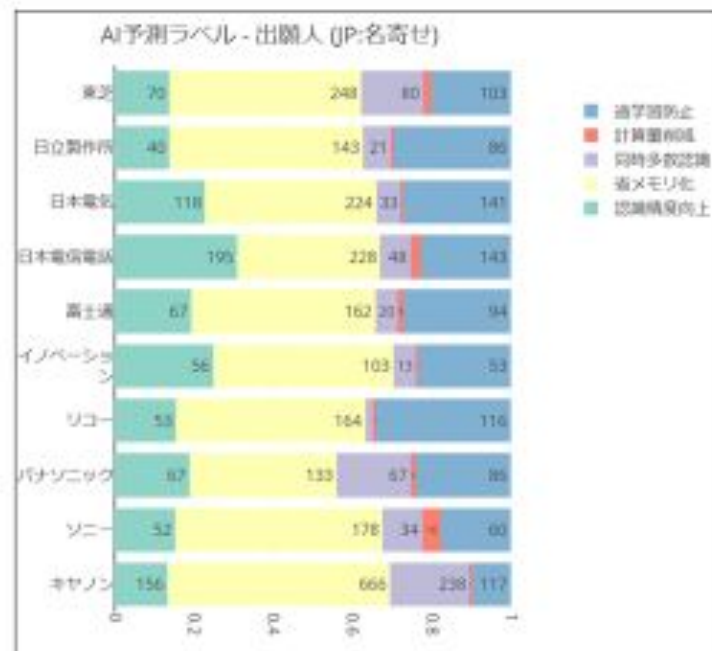
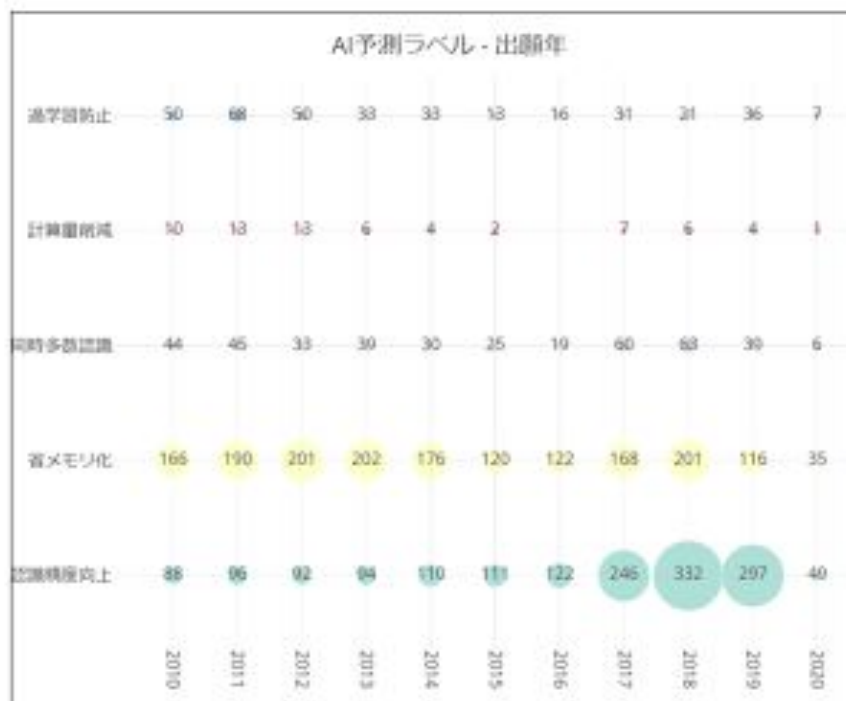
J	K	L	M	AI予測スコア	AI予測ラベル
IPC	ターミナル	USPCクラス	リーガルステータス (JP/N/A/OOGL)		
03B7T100	N086		審査済未付	0.80802038	近接検査向上
03B7T100 G06N09/08	SLO6 M07B		公開中	0.808013587	近接検査向上
03B7T100 G06T1/00				0.8	近接検査向上
03B7T100 G06T1/054	SLO6		権利維持		
03B7T100 G06T1/00	N086 A02G		審査済未付	0.808017154	近接検査向上

- ・過去の社内分類付与リストを教師データとしてセットして、AIが社内分類を予測付与
- ・他社検索システムの公報リストをインポートしてAI予測も可能。予測結果のエクセルDLにも対応。
- ・これまで人手で精査して付与していた社内分類付与をAIの自動付与で業務工数を大幅削減。



8.社内分類でマップ作成

AIの予測ラベルを利用した社内分類ベースのマップ作成 1



- ・社内分類ベースの Patent マップ作成が可能となり、競合企業のポートフォリオを社内分類を軸にマップを作成して分析が可能
- ・作業グループ共有オプションで、構築した教師データを他のIDにリアルタイムで共有可能となり、担当者が構築した教師データを利用して、関連部門のIDで社内分類ベースのマップ作成も可能

9. 関連KWを利用した用途探索

フィルター(詳細) 簡単フィルター(簡易)

1 AND 請求の範囲(出願/付与) グルタチオン

+ フィルター条件を追加

キーワード検索で、探索したい化合物/材料名で検索



関連キーワード

対象特許群 技術分野(JP)

- 組成物 8762.09
- グルタチオン 7155.75
- 出願 5698.48
- 分野 5289.32
- 特許出願 4677.31
- 細胞 4040.2
- タンパク質 4032.35
- 利益 3921.15
- 方法 3904.73

- 栄養補助食品 402.71
- 周産 401.97
- シスチン 401.63
- コラーゲン 399.6
- + 育毛剤 399.27**
- 供与 398.94
- 液体 398.24
- 製剤 397.9



検索 17件

権利状態選択: グルタチオン

スコア設定: 上位のマッチ度

検索条件設定: フィルター(詳細) AND claims: グルタチオン AND technical: 育毛剤

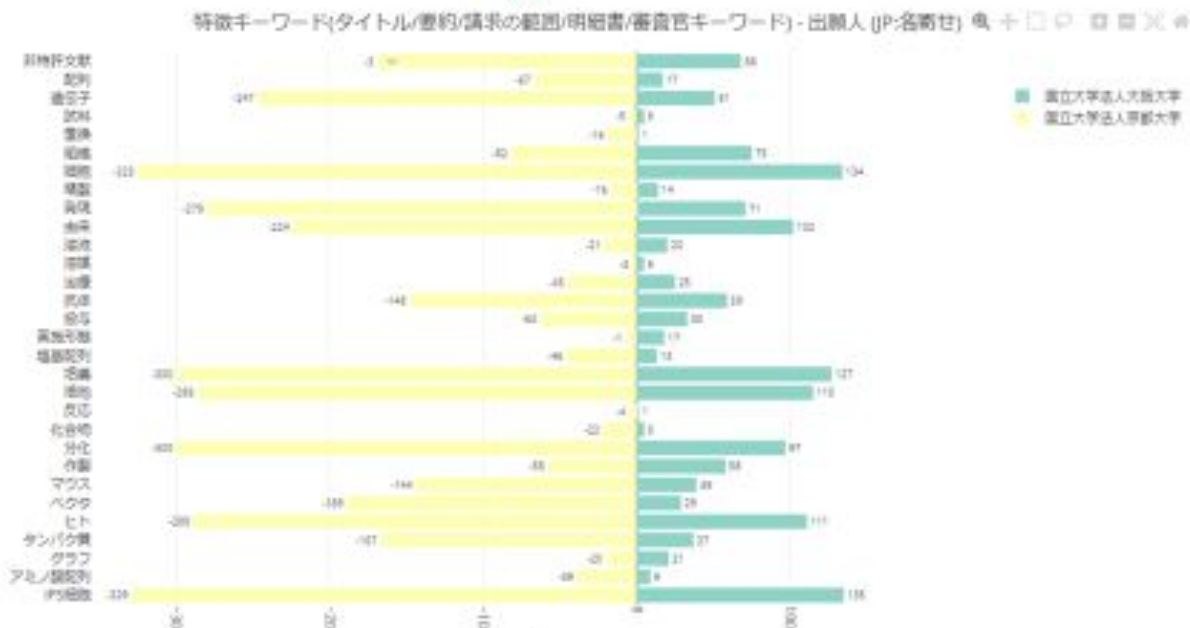
- 探索したい化合物/材料名 (グルタチオン) の検索結果 (4118件) を表示
- 関連キーワードの対象特徴量「技術分野(JP)」から用途に関するキーワードをクリック
- 所定の操作で再検索を実行後、グルタチオンの「育毛剤」に関する特許17件が抽出される
- 関連キーワードで大きな母集団の中から効率的に用途関連キーワードの抽出が可能

10.特徴KWを利用した2社分析

フィルター(詳細)		▽ フィルター(簡易)		集計
1	⌵	AND	出願人/譲受人/権利者 (出願/付与/最新)	国立大学法人京都大学
2	⌵	OR	出願人/譲受人/権利者 (出願/付与/最新)	国立大学法人大阪大学



分析したい2社の出願名で検索実行



- ・「特徴キーワード」×「出願人」のコンパスマップで、両社特許の母集団に出現する単語を可視化
- ・2社間で出現単語を比較することで、**技術的な強み/弱みや注力度を可視化して分析が可能**