

分散表現学習を利用した効率的な特許調査

一文書のベクトル化方法と文書分類への応用

2019.07.04

アジア特許情報研究会¹⁾
花王 知的財産部 安藤俊幸
はやぶさ国際特許事務所 桐山 勉

要旨

近年、word2vecのような単語の分散表現手法やそれを文書のベクトル化に拡張したdoc2vec等の有用性が注目されている。本報では文書、文、単語のベクトル化方法とそのベクトルを用いた機械学習による文書分類、次元圧縮による俯瞰可視化との組み合わせ等の特許調査への応用を事例を基に報告する。

1. 特許文書のベクトル化手法
2. 機械学習による文書分類
3. 特許調査への応用検討



1) アジア特許情報研究会 <https://sapi.kaisei1992.com/>

2) deep-learningを用いた効率的な特許調査
https://sapi.kaisei1992.com/wp-content/uploads/2019/03/N12_ando.pdf

人工知能(AI)の一般的な方法論と特許調査への応用

問題の定式化と解法

解きたい問題 (情報要求)
社会(産業、経済、学術)
等のあらゆる問題

定式化



問題の解

数学的な記述
(最適化問題等)

情報・数理的手法
(最適化・探索法等)

統計・データサイエンス

←コンピュータ
内部の世界

文書データを
コンピュータ内部で
扱えるようにする
→文書ベクトル化

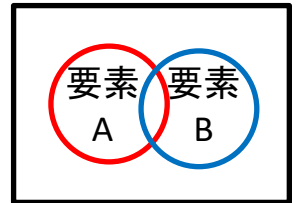
高次元配列:テンソル
テンソルフロー

※現在定式化はほとんど
人が行っている

○特許調査の一般的な方法論

万能のアルゴリズムは無い! NFL定理
目的・データに応じてアルゴリズム選択 ノーフリーランチ定理
外枠

調査目的に応じて
パターン化 (定石・フレーム)



フレーム問題

教師データ	機械学習の種類	
有	予め決められたクラスに クラス分類	数値で予測する 回帰
無	似た者同士をまとめる クラスタリング	特徴量を減らす 次元圧縮

教師データ有の場合

$$y = f(X)$$

y: 目的変数

X: 説明変数

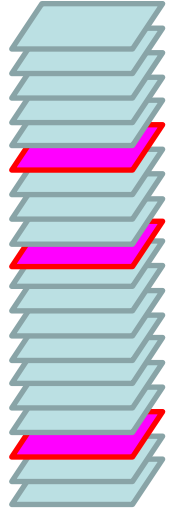
目的変数が
カテゴリーデータ: クラス分類
実数: 回帰

スクリーニング課程の現状と理想

類似度によるスコア計算等

現状

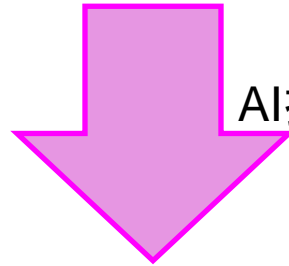
どの文献が当たりか判らないので、片端から読み込む



D2(Y1)

D3(Y2)

D1(X1)



AI技術(機械学習、深層学習、自然言語処理)の活用によるパラダイムシフト

理想

可能性の高い文献から順番に読みたい

先行文献にスコアが付いていれば、スコアの高い文献から読むことができる・・・

発明の構成要素毎に根拠個所を見たい



D1(X1)

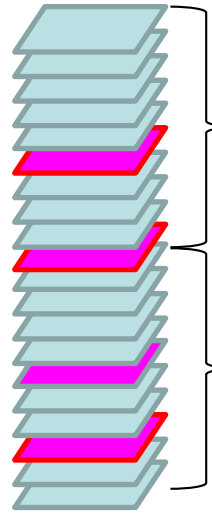
D2(Y1)

D3(Y2)

機械(AI)がなぜ根拠と判定したか理由を知りたい 3

教師有機械学習を用いた効率的な特許調査の課題

検索結果



教師データ
トレーニングデータ
テストデータ

文書を形態素解析してベクトル化
(テンソル化: 高次元配列)

トレーニングデータで
サーチ/ノイズを学習

教師データの準備はどうするの？

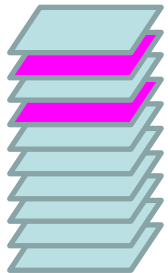
文書分類

- BoW
- TF-IDF
- Ave-word2vec
- doc2vec
- CDV
- Ave-fastText
- SCDV-fastText

- XGB
- SGD
- GNB
- SVC
- SVCrbf
- RFg
- AdaBoost
- MLP

テストデータのサーチ/ノイズ確率を計算 → 2値分類
→ 回帰

スコア付け/仕分け



先行技術調査/無効資料調査

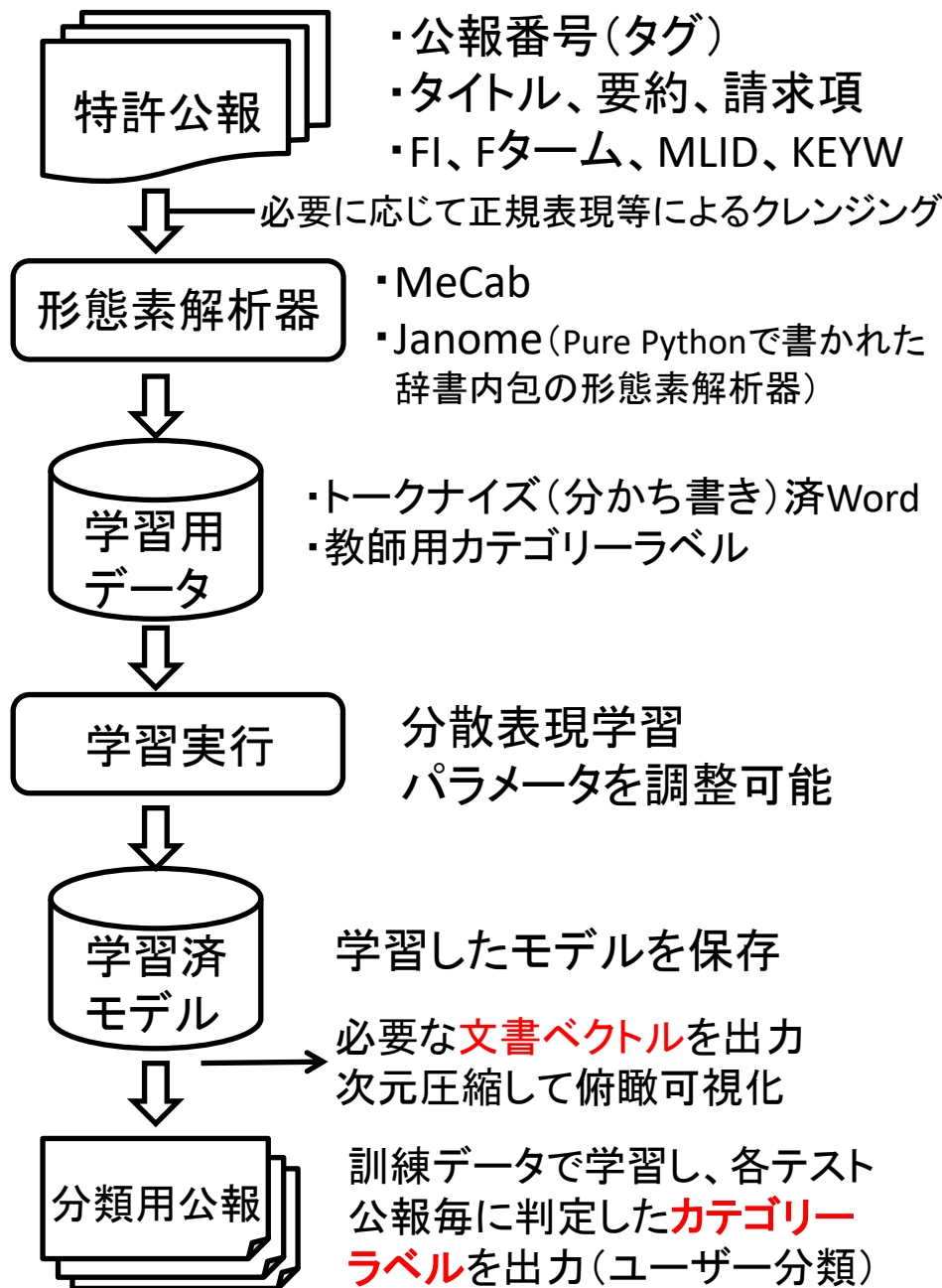
適合確率の高い文献から順番に査読

SDI (Selective Dissemination of Information) 調査

SDIをサーチ特許とノイズ特許に仕分け

仕分け/分類性能を知りたい！

文書のベクトル化処理と文書分類の概要



文書ベクトル化ソース

T: タイトル

A: 要約

C: 請求項

(E: 実施例)

特許分類(FI、Fターム等)をソース(素性)として用いることも可能

文書ベクトル化方法

	略号	モデル名称
①	BoW	Bag of Word
②	TF-IDF	TF-IDF
③	Ave-word2vec	average word2vec
④	doc2vec	doc2vec
⑤	SCDV	Sparse Composite Document Vectors
⑥	Ave-fastText	average fastText
⑦	SCDV-fastText	SCDV fastText

分散表現ベクトル

文書分類方法

	略号	分類アルゴリズム
①	XGB	eXtreme Gradient Boosting
②	SGD	Stochastic Gradient Descent 確率的勾配降下法
③	GNB	naive_bayes.GaussianNB ナイーブベイズ
④	SVC	SVC(kernel='linear') サポートベクトルマシン
⑤	SVCrbf	SVC(kernel='rbf') サポートベクトルマシン
⑥	RFg	Random Forest (gini)
⑦	AdaBoost	AdaBoost
⑧	MLP	多層パーセプトロン

scikit-learn アルゴリズム早見表

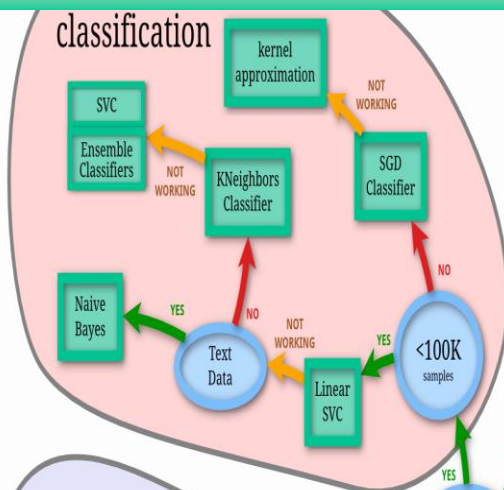
scikit-learn
algorithm cheat-sheet

クラス分類

教師あり

目的変数:
カテゴリー型

応用例:
文書分類

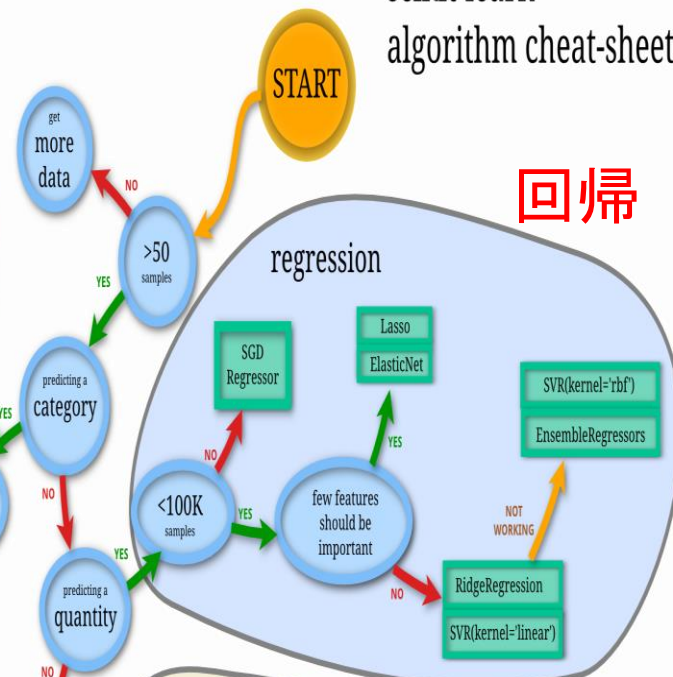


回帰

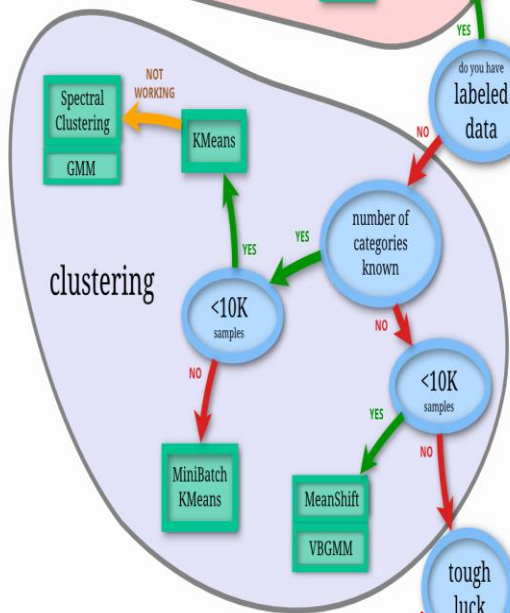
教師あり

目的変数:
連続した数値型

応用例:
売上予測

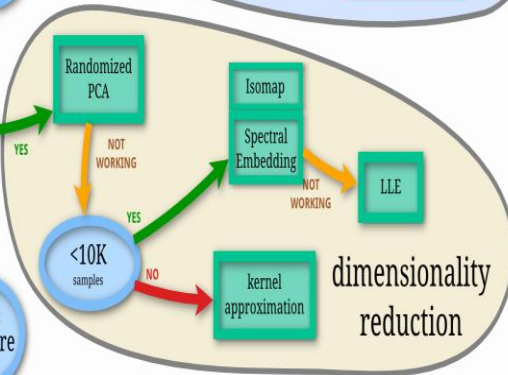


教師なし



教師なし

- PCA
- t-SNE



次元圧縮

Back

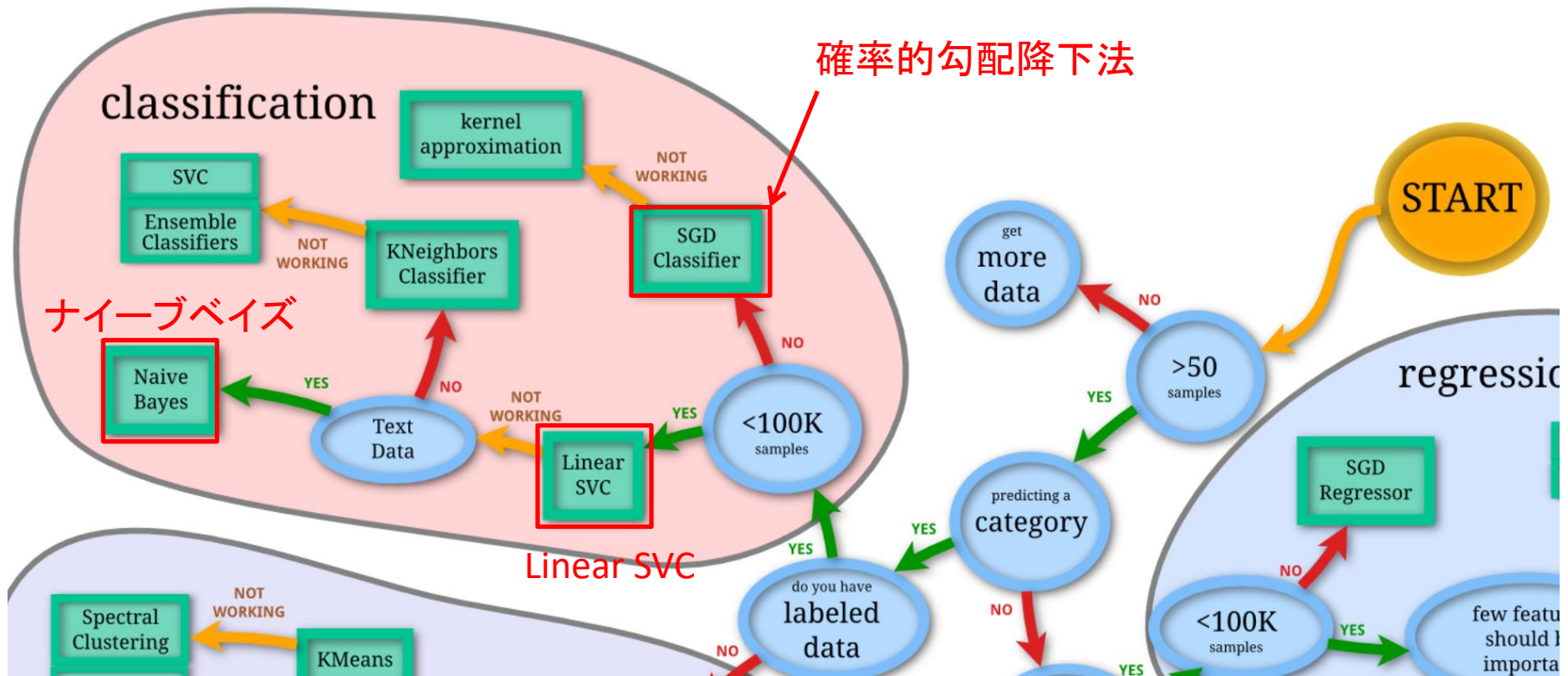


http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

scikit-learn のクラス分類拡大図

クラス分類拡大図

scikit-learn algorithm cheat-sheet



scikit-learn algorithm cheat-sheet

https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

商用特許DBにおけるクラス分類の利用例

特許の全出願を5行×7列=35分野に大きく分類

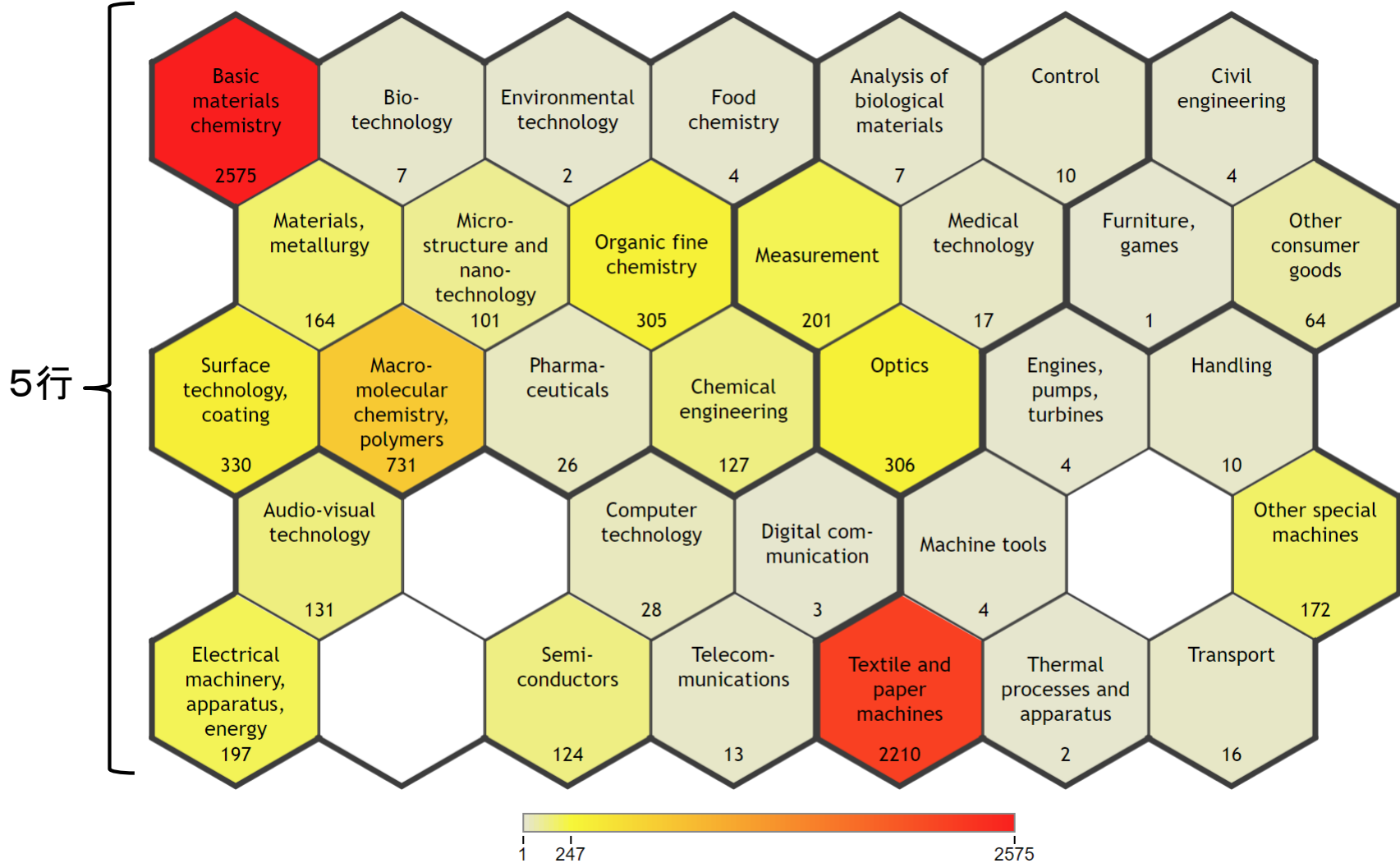
2019.03.25 時点 Orbit.com

IPCの約6万分類→35分類に集約

7列

4J039GA24 and CN

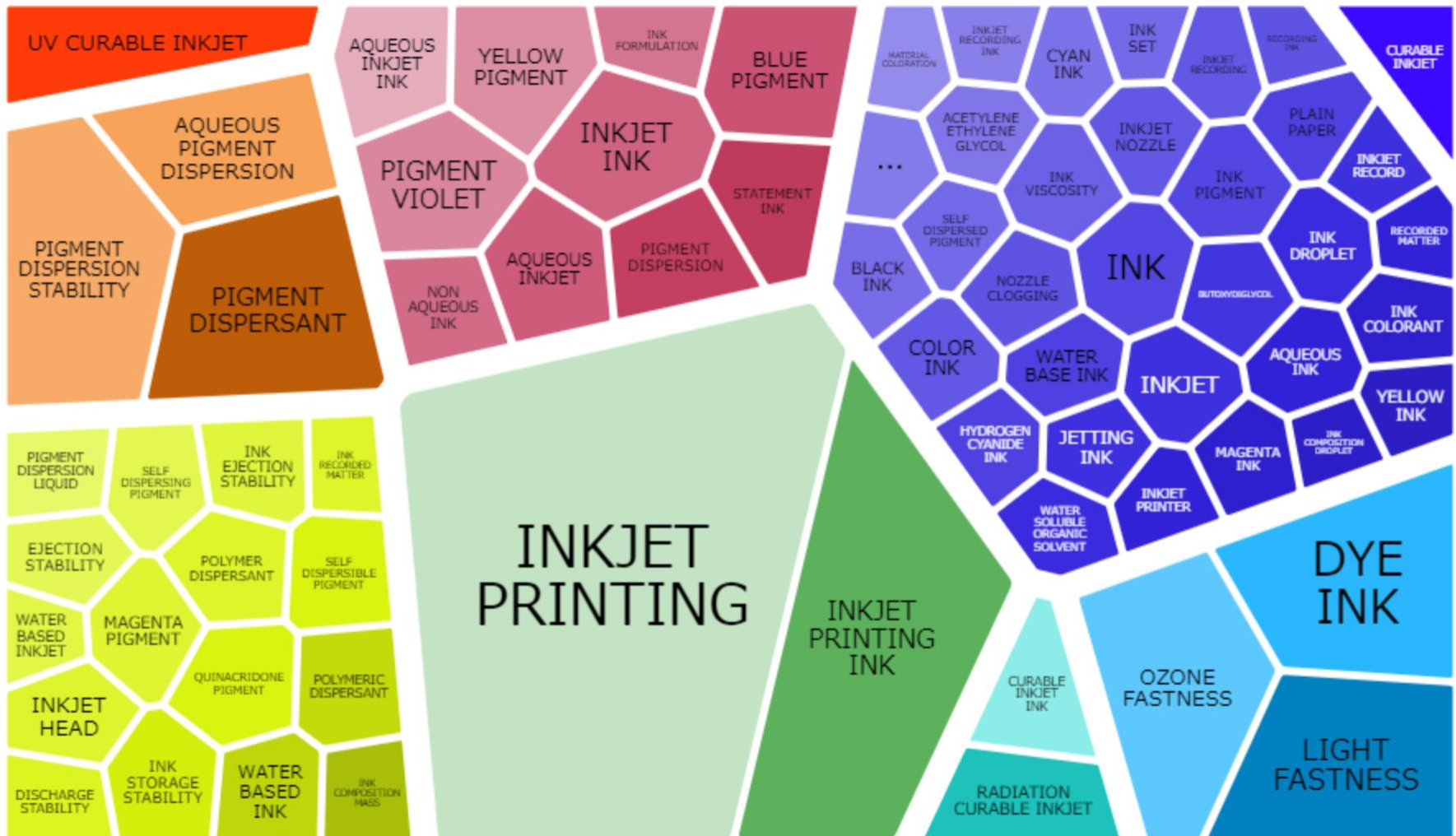
2584件



商用特許DBにおけるクラスタリングの利用例

2019.03.25 時点 Orbit.com
4J039GA24 and CN

2584件

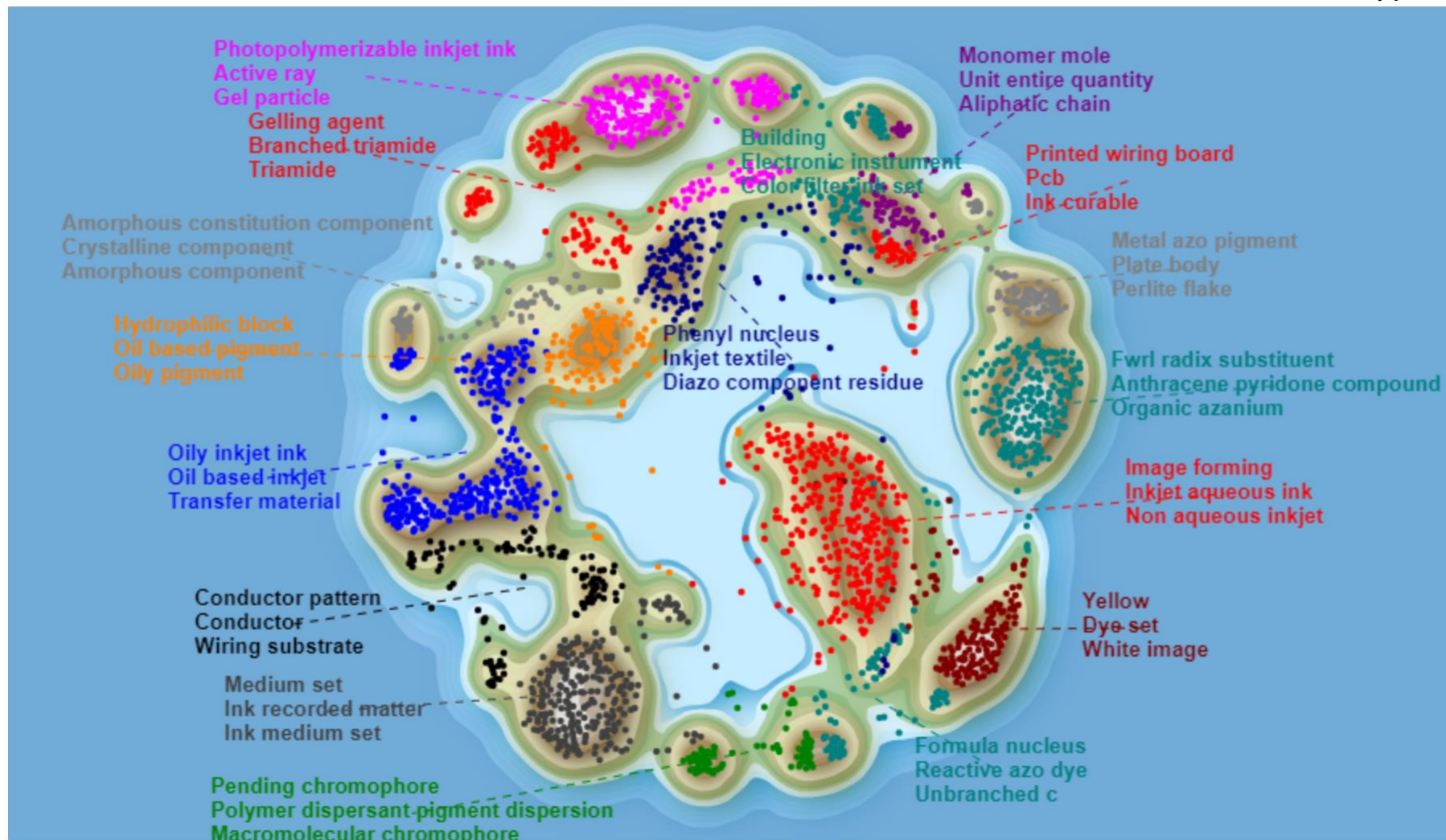


商用特許DBにおける次元圧縮とクラスタリングの利用例

2019.03.25 時点 Orbit.com
4J039GA24 and CN

テクノロジークラスターごとにみるランドスケープ

2584件



文書ベクトル化 × 文書分類検討概要

文書ベクトル化

	略号	モデル名称
①	BoW	Bag of Word
②	TF-IDF	TF-IDF
③	Ave-word2vec	average word2vec
④	doc2vec	doc2vec
⑤	SCDV	Sparse Composite Document Vectors
⑥	Ave-fastText	average fastText
⑦	SCDV-fastText	SCDV fastText

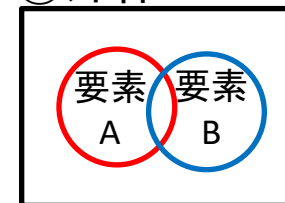
分類体系

筆頭FI	内容	件数
C09D 11/00	インク	923
B41M 5/00 A	記録方法	175
C09D 11/30	インクジェットインク	150
C09D 11/322	顔料インク	99
C09D 11/38	非高分子添加剤	76
C09D 17/00	顔料ペースト	64
C09D 11/326	顔料分散剤	51
C09D 11/328	染料	45
G02B 5/20 101	カラーフィルター	39
C09D 11/34	ホットメルト	38

3種類の分類の粒度での検討予定

- ①発明の主題レベル(筆頭FI)
- ②発明の構成要素レベル(Fターム)
- ③明細書の文言記載レベル

①外枠



文書分類

略号	分類アルゴリズム	概要
XGB	eXtreme Gradient Boosting	Gradient Boosting(勾配ブースティング)とRandom Forestを組み合わせたアンサンブル学習(複数の分類器を集めて構成される分類器)
SGD	Stochastic Gradient Descent 確率的勾配降下法	連続最適化問題に対する勾配法の乱択アルゴリズム。目的関数が、微分可能な和の形である事が必要。
GNB	naive_bayes.GaussianNB ナイーブベイズ	単純ベイズ確率モデルに基づく分類器
SVC	SVC(kernel='linear') サポートベクトルマシン	線形入力素子を利用して 2 クラスのパターン識別器を構成する手法である。訓練サンプルから、各データ点との距離が最大となるマージン最大化超平面を求めるという基準(超平面分離定理)で線形入力素子のパラメータを学習する。
SVCrbf	SVC(kernel='rbf') サポートベクトルマシン	RBFカーネルを用いたサポートベクトルマシン
RFg	Random Forest(gini)	決定木を弱学習器とする集団学習アルゴリズム
AdaBoost	AdaBoost	AdaBoostは前の分類器の間違いに応じて調整された次の分類器を作るアルゴリズム
MLP	多層パーセプトロン	順伝播型ニューラルネットワークの一分類である。MLPは少なくとも3つのノードの層からなる。

計3098

交差検証

8分割交差検証

実務では10分割交差検証がよく使われる

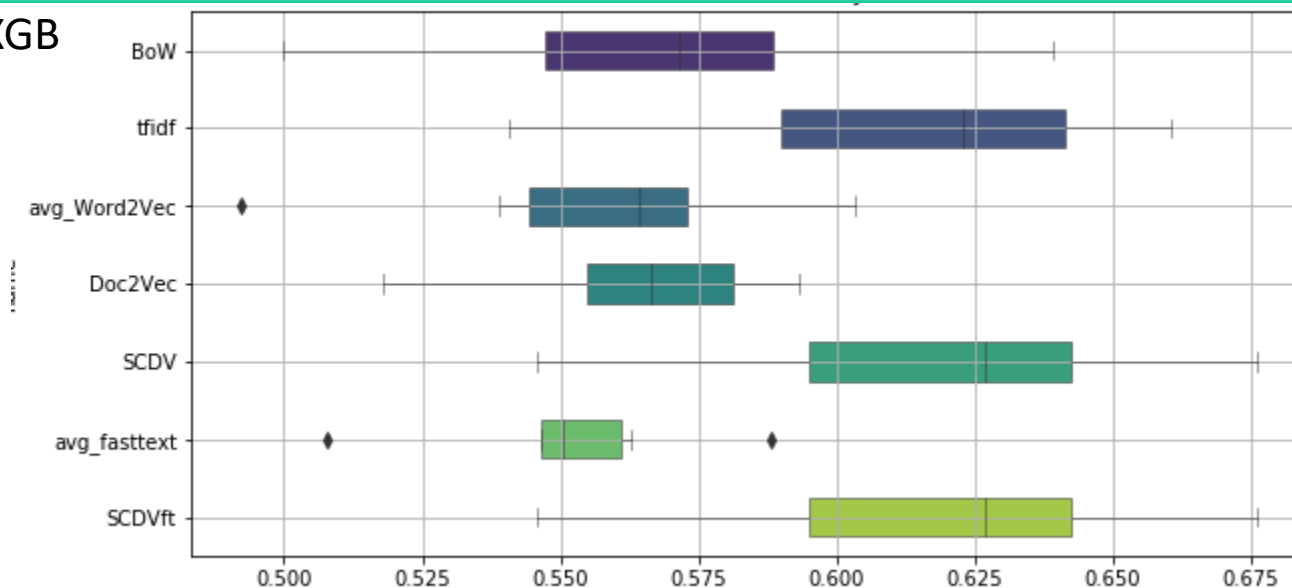
8分割交差検証								
	1	2	3	4	5	6	7	8
1回目	訓練	訓練	訓練	訓練	訓練	訓練	訓練	テスト
2回目	訓練	訓練	訓練	訓練	訓練	訓練	テスト	訓練
3回目	訓練	訓練	訓練	訓練	訓練	テスト	訓練	訓練
4回目	訓練	訓練	訓練	訓練	テスト	訓練	訓練	訓練
5回目	訓練	訓練	訓練	テスト	訓練	訓練	訓練	訓練
6回目	訓練	訓練	テスト	訓練	訓練	訓練	訓練	訓練
7回目	訓練	テスト	訓練	訓練	訓練	訓練	訓練	訓練
8回目	テスト	訓練	訓練	訓練	訓練	訓練	訓練	訓練

交差検証(交差確認)(こうさけんしょう、英: Cross-validation)とは、統計学において標本データを分割し、その一部をまず解析して、残る部分でその解析のテストを行い、解析自身の妥当性の検証・確認に当てる手法を指す。データの解析(および導出された推定・統計的予測)がどれだけ本当に母集団に対処できるかを良い近似で検証・確認するための手法である。

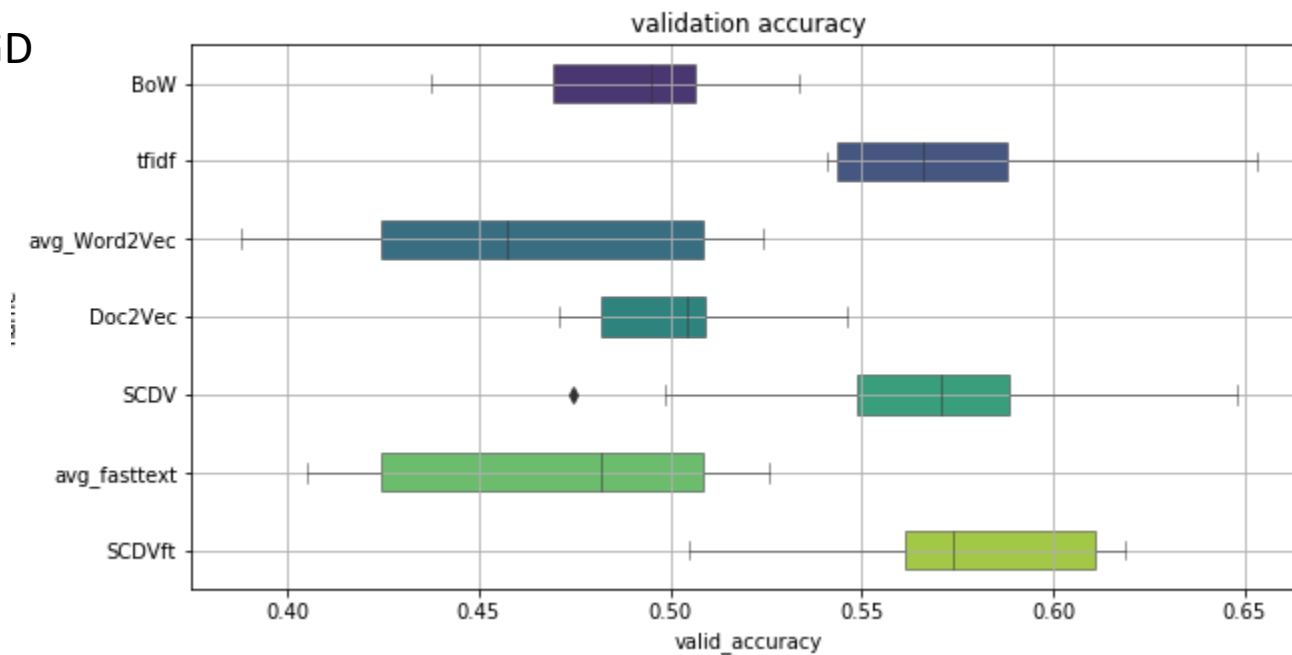
最初に解析するデータを「訓練事例集合(training set)」などと呼び、他のデータを「テスト事例集合(testing set、テストデータ)」などと呼ぶ。

文書ベクトル化手法 × 分類アルゴリズムの正解率

① XGB

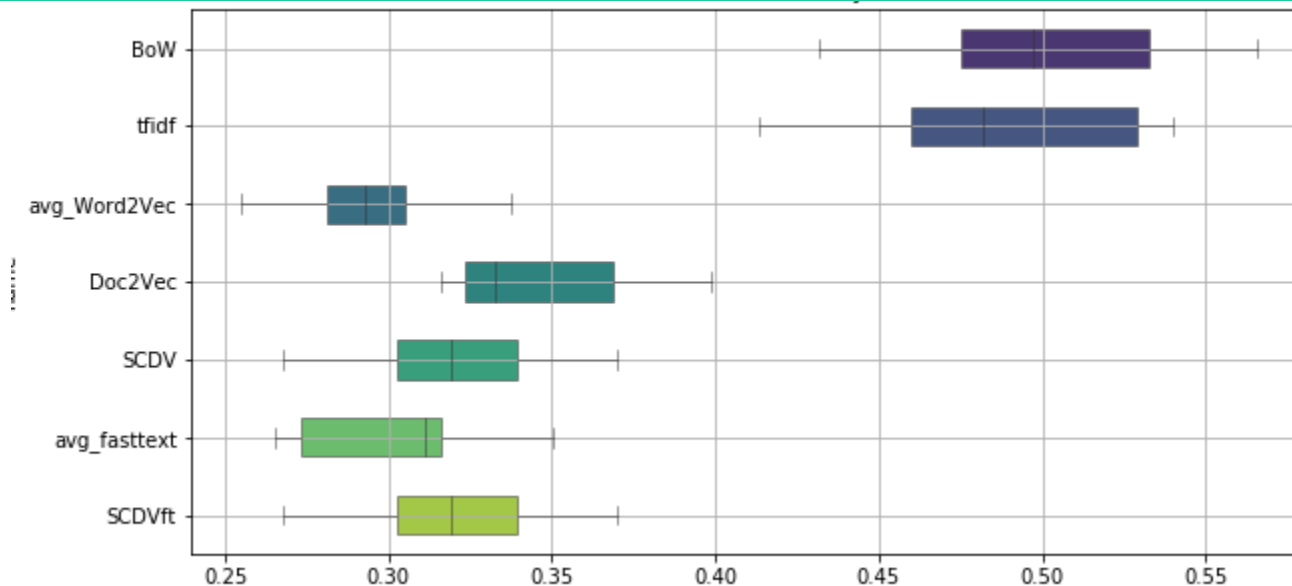


② SGD

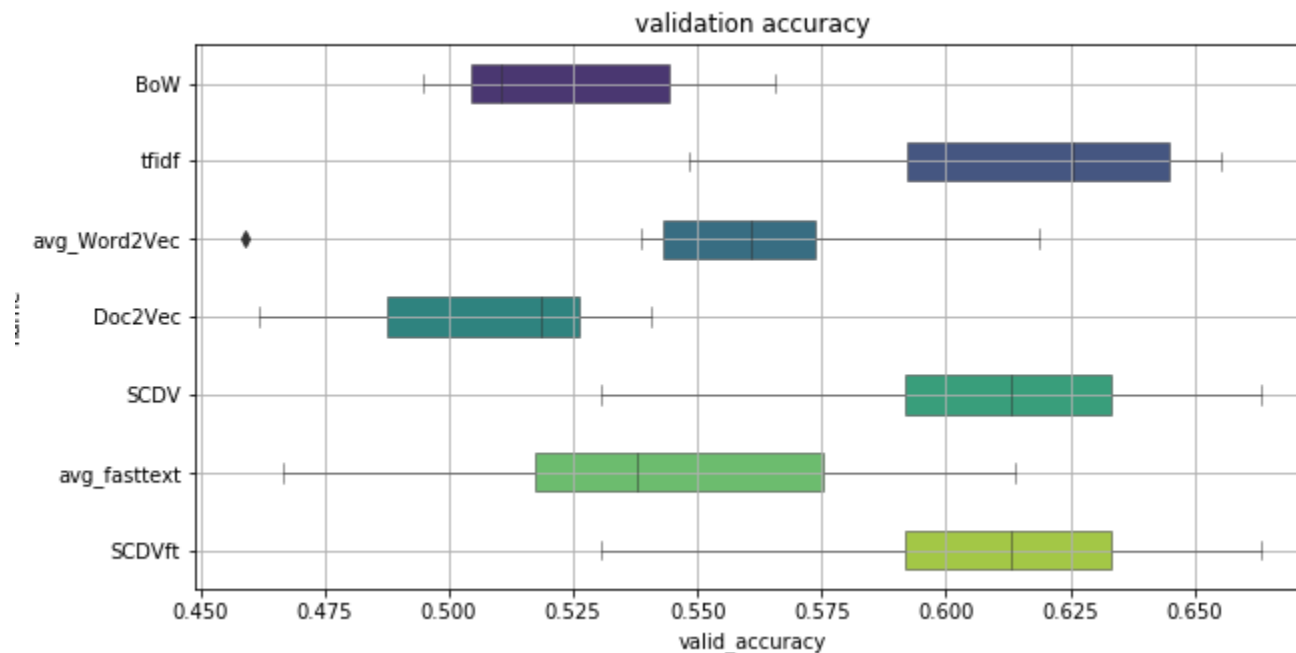


文書ベクトル化手法 × 分類アルゴリズムの正解率

③ GNB

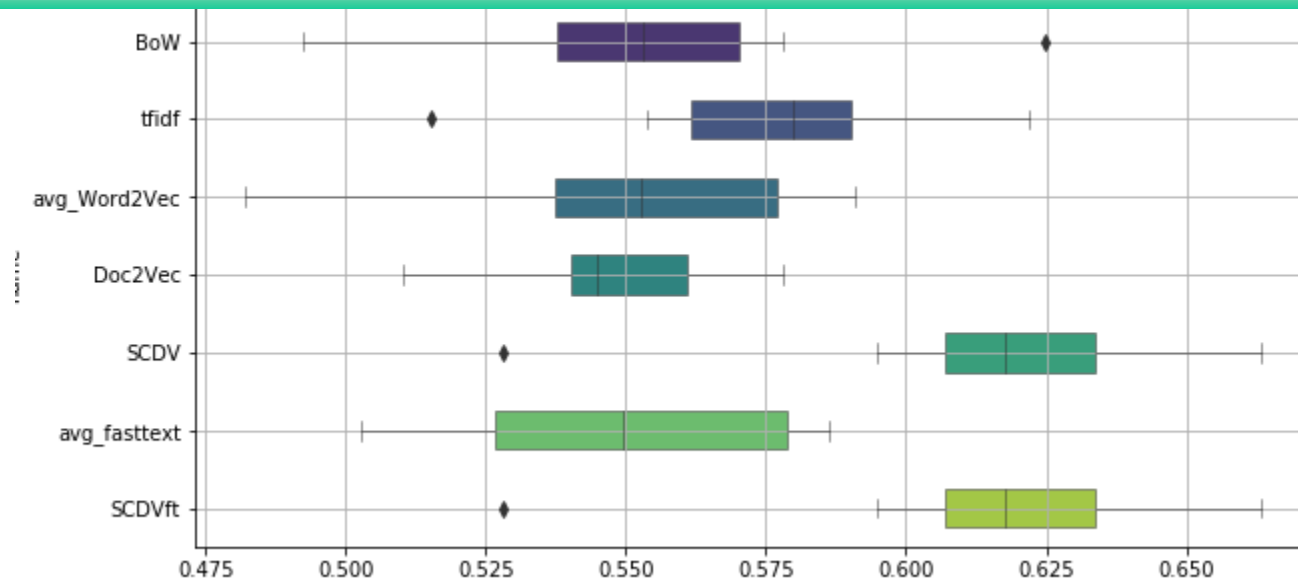


④ SVC

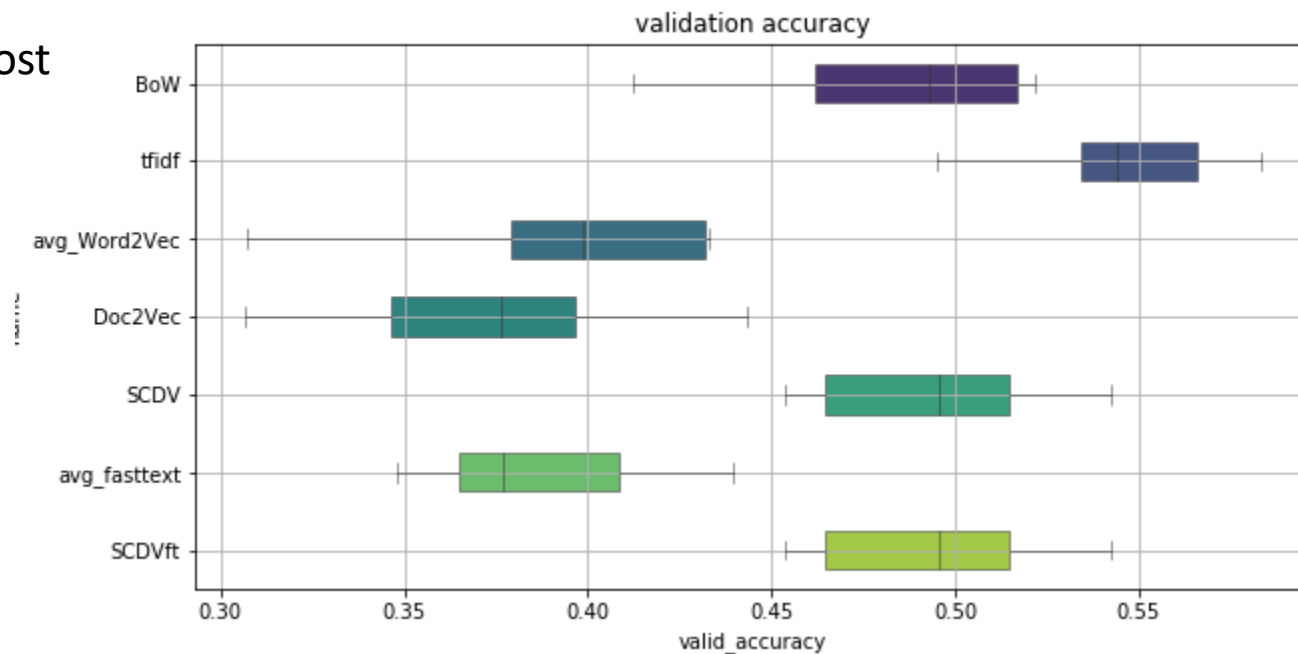


文書ベクトル化手法 × 分類アルゴリズムの正解率

⑥RFg

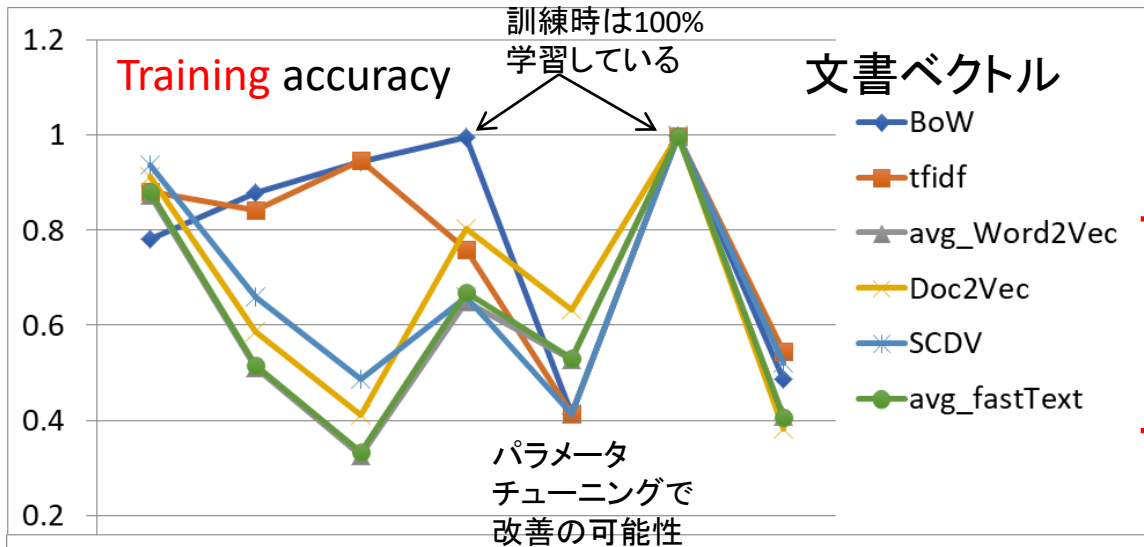


⑦AdaBoost



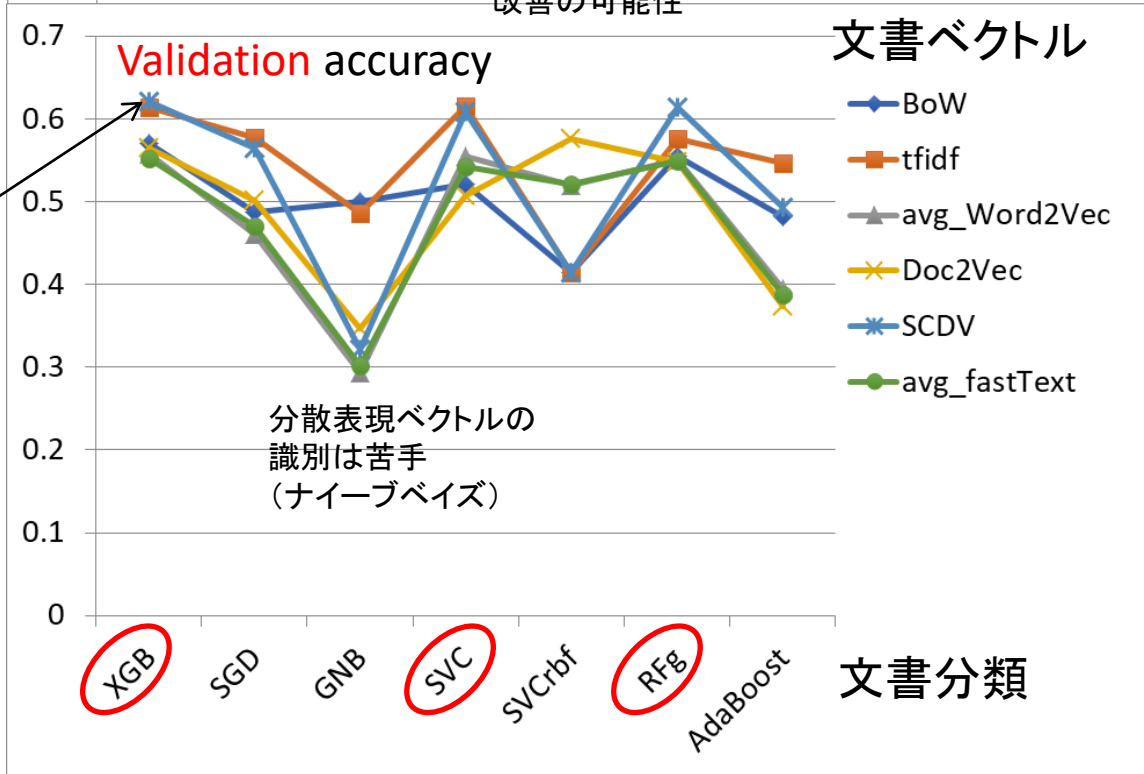
文書ベクトル × 文書分類の正解率比較

訓練



分散表現ベクトル

テスト



※汎化能力が重要
未知のテストデータに対する分類(識別)能力

入学試験に例えると
問題集・模試の点数が訓練。
入学試験本番の点数がテストに相当する。

※文書分類は
ほぼデフォルト条件

カテゴリー別の正解率比較

w2v XGB					
カテゴリー	件数	正解率	精度	再現率	F値
インク	926	0.66	0.51	0.66	0.57
インクジェットインク	150	0.17	0.45	0.17	0.24
顔料インク	198	0.08	0.36	0.08	0.12
顔料分散剤	51	0.06	0.23	0.06	0.09
染料	45	0.15	0.58	0.15	0.23
着色剤	17	0.19	0.25	0.19	0.21
ホットメルトインク	38	0.23	0.50	0.23	0.30
非水性溶媒	38	0.03	0.06	0.03	0.04
多色インクセット	37	0.16	0.46	0.16	0.22
非高分子添加剤	76	0.00	0.00	0.00	0.00
記録方法	239	0.40	0.58	0.40	0.47
その他	1283	0.82	0.68	0.82	0.74

d2v XGB					
カテゴリー	件数	正解率	精度	再現率	F値
インク	926	0.67	0.52	0.67	0.58
インクジェットインク	150	0.04	0.24	0.04	0.07
顔料インク	198	0.07	0.51	0.07	0.11
顔料分散剤	51	0.04	0.25	0.04	0.07
染料	45	0.08	0.44	0.08	0.14
着色剤	17	0.00	0.00	0.00	0.00
ホットメルトインク	38	0.23	0.54	0.23	0.31
非水性溶媒	38	0.03	0.13	0.03	0.04
多色インクセット	37	0.03	0.13	0.03	0.05
非高分子添加剤	76	0.00	0.00	0.00	0.00
記録方法	239	0.36	0.60	0.36	0.44
その他	1283	0.84	0.65	0.84	0.73

FI-SVC-TF					
カテゴリー	件数	正解率	精度	再現率	F値
インク	926	0.74	0.74	0.74	0.74
インクジェットインク	150	0.63	0.69	0.63	0.66
顔料インク	198	0.36	0.66	0.36	0.46
顔料分散剤	51	0.10	0.56	0.10	0.17
染料	45	0.06	0.38	0.06	0.11
着色剤	17	0.56	0.58	0.56	0.54
ホットメルトインク	38	0.80	0.71	0.80	0.74
非水性溶媒	38	0.57	0.56	0.57	0.55
多色インクセット	37	0.16	0.31	0.16	0.20
非高分子添加剤	76	0.10	0.52	0.10	0.16
記録方法	239	0.63	0.56	0.63	0.59
その他	1283	0.87	0.75	0.87	0.80

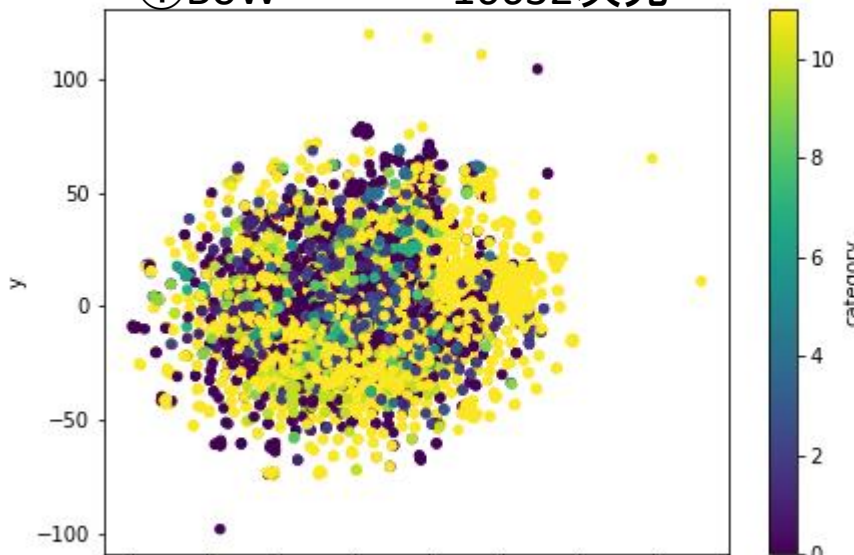
TF-IDF XGB					
カテゴリー	件数	正解率	精度	再現率	F値
インク	926	0.75	0.56	0.75	0.65
インクジェットインク	150	0.14	0.52	0.14	0.21
顔料インク	198	0.18	0.56	0.18	0.27
顔料分散剤	51	0.08	0.21	0.08	0.12
染料	45	0.23	0.79	0.23	0.34
着色剤	17	0.00	0.00	0.00	0.00
ホットメルトインク	38	0.53	0.87	0.53	0.61
非水性溶媒	38	0.08	0.21	0.08	0.10
多色インクセット	37	0.44	0.38	0.44	0.39
非高分子添加剤	76	0.04	0.17	0.04	0.07
記録方法	239	0.60	0.66	0.60	0.63
その他	1283	0.83	0.75	0.83	0.79

件数が少ないカテゴリーは正解率が低い！

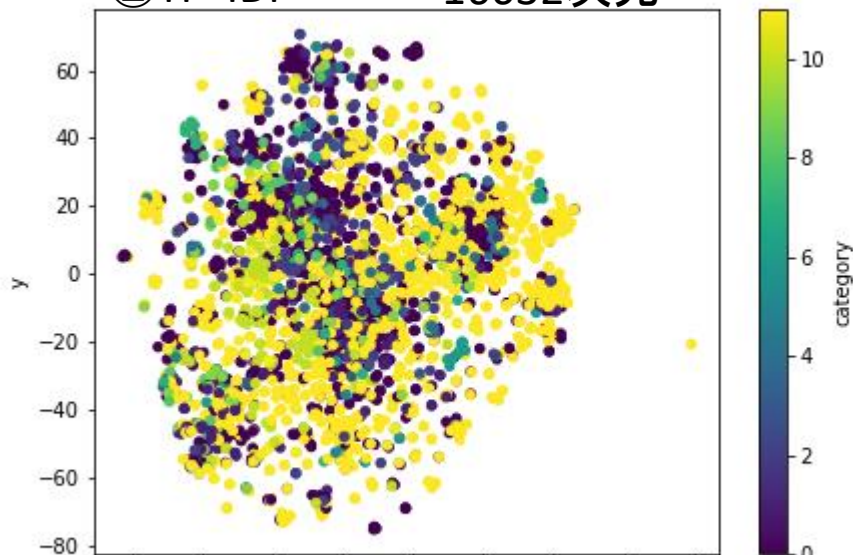
各種チューニングが必要

t-SNEによる各文書ベクトルの2次元可視化

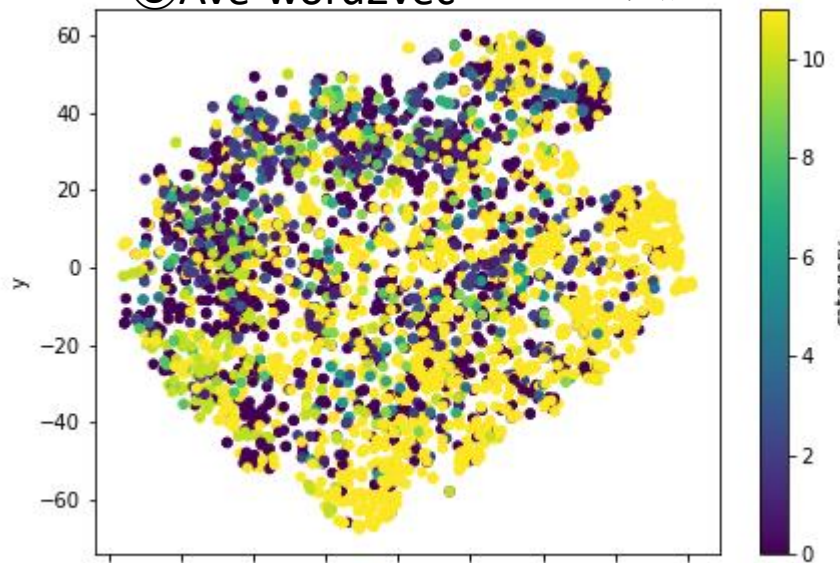
①BoW 16632次元



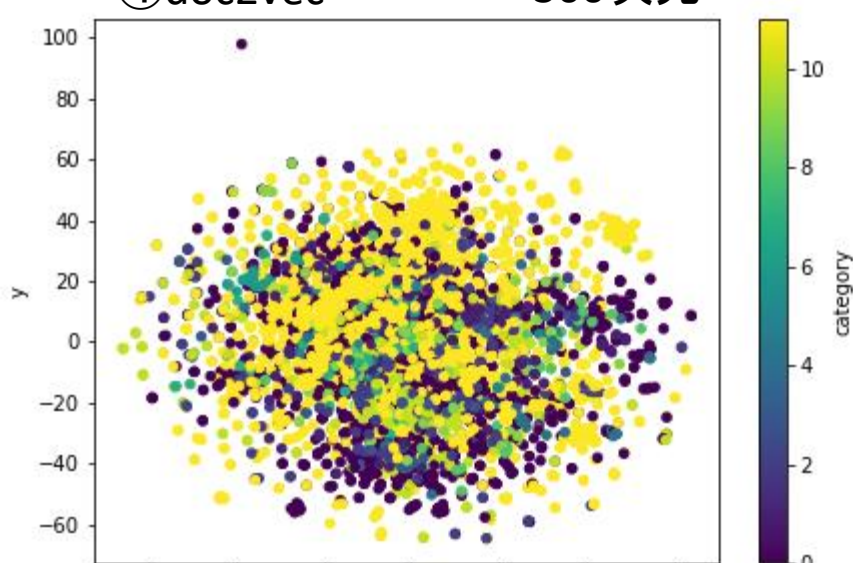
②TF-IDF 16632次元



③Ave-word2vec 300次元

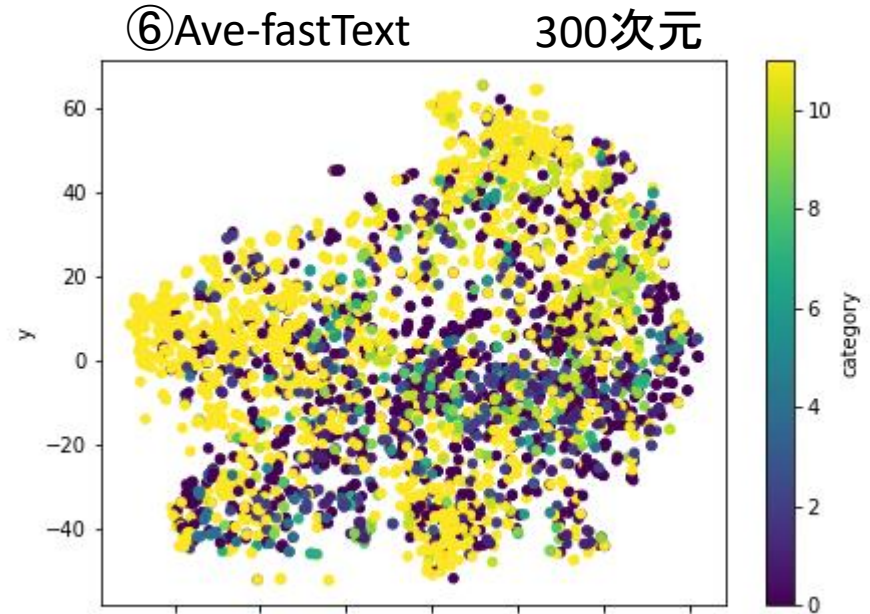
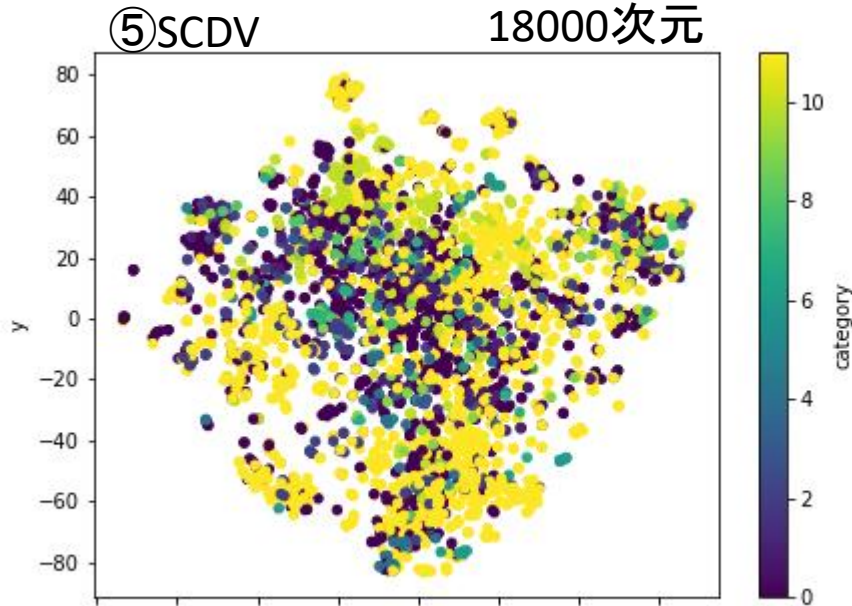


④doc2vec 300次元



↑カラーマッピングはカテゴリーID

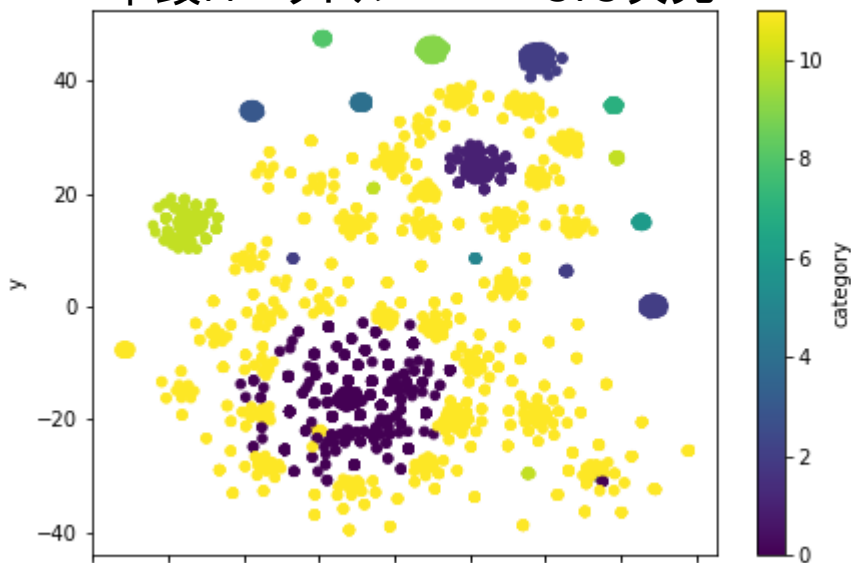
t-SNEによる各文書ベクトルの2次元可視化



ID	カテゴリー	件数
11	その他	1283
10	記録方法	239
9	非高分子添加剤	76
8	多色インクセット	37
7	非水性溶媒	38
6	ホットメルトインク	38
5	着色剤	17
4	染料	45
3	顔料分散剤	51
2	顔料インク	198
1	インクジェットインク	150
0	インク	926
	計	3098

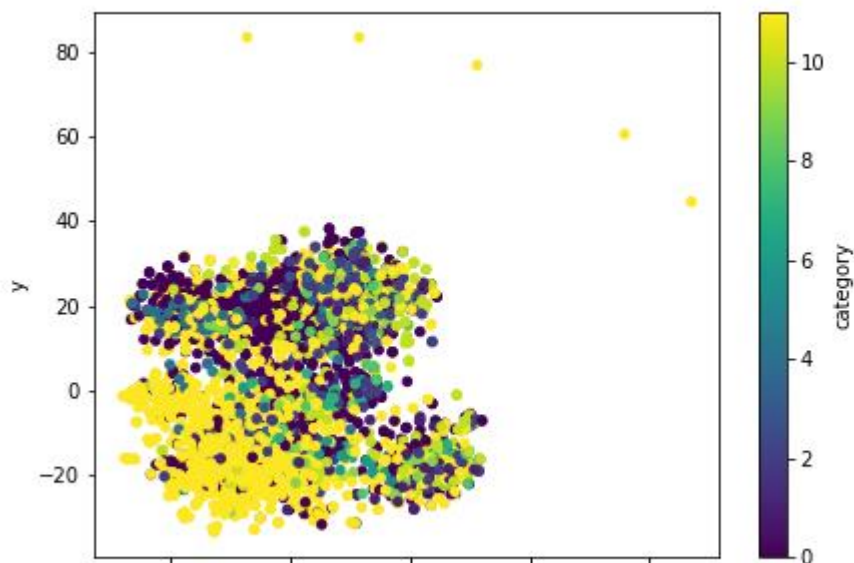
特許分類を用いた文書のベクトル化例 (t-SNE)

筆頭FIベクトル 575次元

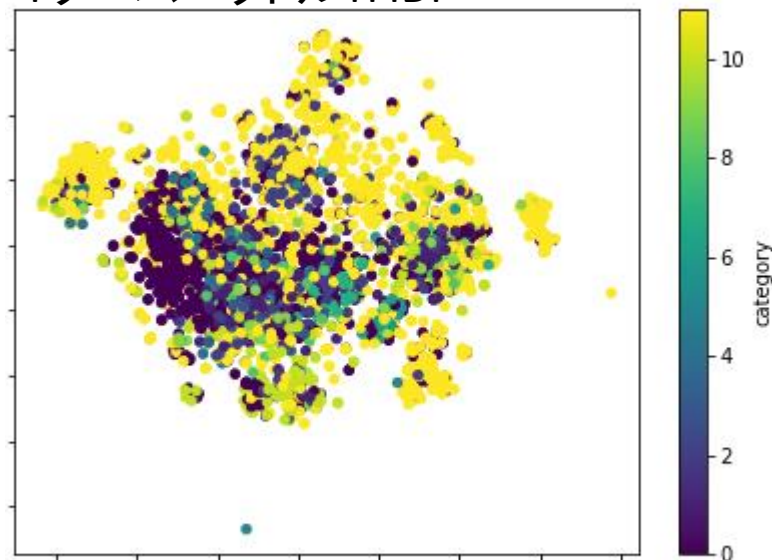


ID	カテゴリー	件数
11	その他	1283
10	記録方法	239
9	非高分子添加剤	76
8	多色インクセット	37
7	非水性溶媒	38
6	ホットメルトインク	38
5	着色剤	17
4	染料	45
3	顔料分散剤	51
2	顔料インク	198
1	インクジェットインク	150
0	インク	926
	計	3098

FタームベクトルBoW 2839次元



FタームベクトルTFIDF 2839次元



処理時間(計算コスト)

ベクトル化手法	次元	ベクトル化	t-SNE	XGB	SGD	GNB	SVC	SVCrbf	RFg	AdaBoost	MLP	合計(秒)	(時間)
BoW	16632	0.41	226.80	4090.27	28.72	53.71	2090.64	2262.17	47.84	610.81	メモリエラー	9411.4	2.6
tfidf	16632	0.41	226.58	4190.11	27.54	51.37	1975.358[1917.530[44.92	607.59	メモリエラー	5148.5	1.4
Word2Vec average	300	12.39	23.83	227.99	0.59	0.86	31.67	36.42	21.04	60.75	メモリエラー	415.5	0.1
Doc2Vec	300	21.99	33.29	233.18	0.60	0.81	42.60	36.70	20.86	61.43	メモリエラー	451.5	0.1
SCDV	18000	155.56	246.55	5838.03	31.86	54.32	1954.29	2060.65	64.83	1115.41	メモリエラー	11521.5	3.2
fasttext average	300	34.11	25.08	227.23	0.63	0.83	31.909[36.12	21.16	60.89	メモリエラー	406.0	0.1
SCDV-fasttext	18000	255.05	246.63	5817.18	31.91	54.59	1945.49	2064.68	65.13	1113.92	メモリエラー	11594.6	3.2
処理時間合計(秒)		479.92	1028.76	20623.99	121.84	216.48	6064.68	6496.74	285.77	3630.79		38949.0	10.8
(時間)		0.13	0.29	5.73	0.03	0.06	1.68	1.80	0.08	1.01		10.8	

使用CPU Intel Core i7-8700K 3.70GHz コア数:6 論理プロセッサ数:12

1論理プロセッサを使用して計測

タスクマネージャーによる負荷計測:CPU 4.35GHz 約11%

ベクトル化共通パラメータ(一部)

features_num = 300

min_word_count = 1

context = 5

downsampling = 1e-3

epoch_num = 10

分類精度重視の観点からは**XGB**がお奨め、
ただし処理時間はかかる

分類精度、使いやすさ、処理時間、
マルチプロセッサ対応の観点からは
Random Forestがお奨め
ただしTF-IDFは精度が悪い

word2vecによる「粘土」の類似語抽出

分散表現ベクトル

One hotベクトル

word2vec「粘土」の類似語

形態素 専門用語抽出

順位	類似語	類似度	順位	頻度	専門用語	順位	頻度
1	スメクタイト	0.774	555	26	スメクタイト	1655	7
4	サポナイト	0.646	2101	4	サポナイト	4655	2
5	ヘクト	0.637	2099	2	ヘクトライト	4656	2
7	スチーブン	0.630	2100	2	スチーブンサイト	4703	2
8	ナイト	0.615	1448	4	カオリナイト	2669	4
9	マイカ	0.614	1449	4	マイカ	3441	3
11	モンモリロナイト	0.599	359	53	モンモリロナイト	246	52
12	カオリ	0.597	1635	3	カオリナイト	2669	4
14	タルク	0.587	1446	4	タルク	2691	4
16	ゼオライト	0.561	1175	7	ゼオライト	1652	7
17	セリ	0.554	2184	4	セリサイト	5112	2

黄色セルは形態素解析による分かち書きに失敗しているが類似語として上位に存在している

主な粘土鉱物(Wikipedia)

カオリナイト(高陵石)
スメクタイト
モンモリロン石(モンモリロナイト)
絹雲母(セリサイト)
イライト
海緑石(グローコナイト)
緑泥石(クロライト)
滑石(タルク)
沸石(ゼオライト)

<https://ja.wikipedia.org/wiki/粘土鉱物>

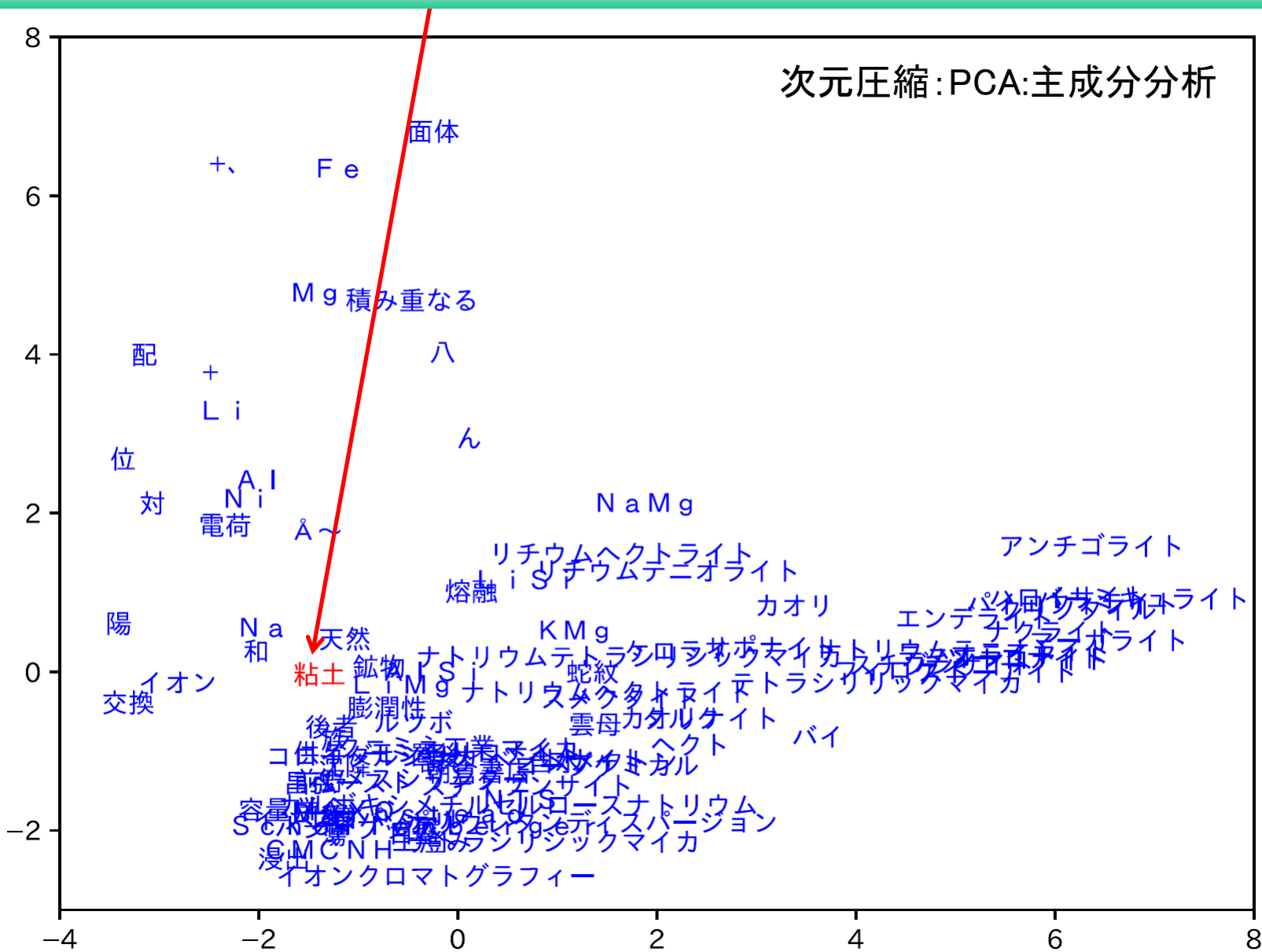
word2vecを使用すると文脈に「粘土」の記載のない文からも具体的な粘土鉱物を学習しており検索クエリの拡張支援ツールとして有用である

専門用語抽出(続き)

専門用語	順位	頻度
水素型スメクタイト	1657	7
合成スメクタイト	1979	6
スメクタイト族	3864	2
スメクタイト群粘土鉱物	4002	2
スメクタイト粘土鉱物	4740	2
合成マイカ	7890	1
カオリン	7203	1

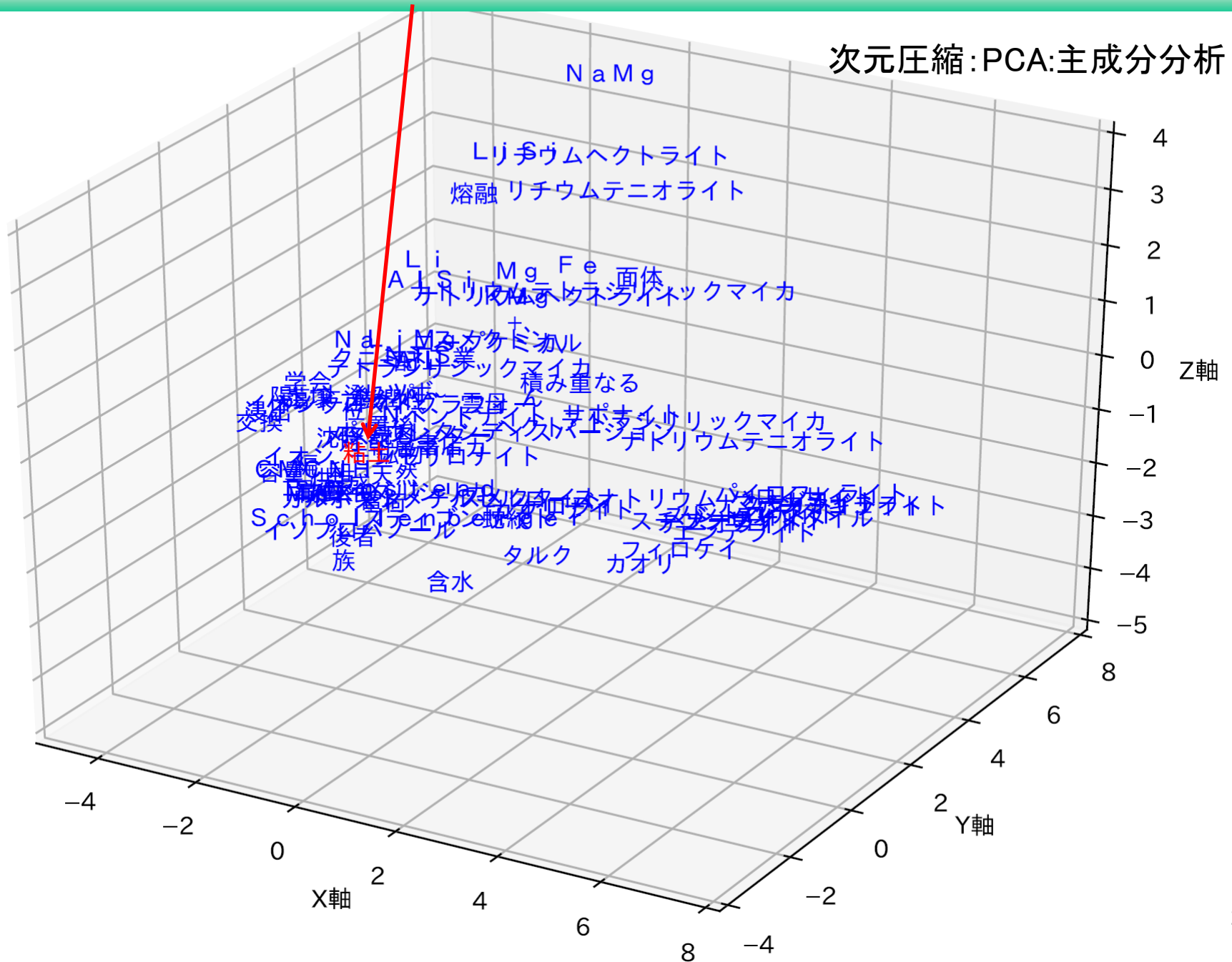
※専門用語抽出システムを使用して形態素解析辞書を強化、あるいは「日化辞辞書」の使用が有効

word2vecによる粘土の類似100ワードの2次元可視化



word2vecによる粘土の類似100ワードの3次元可視化

次元圧縮: PCA:主成分分析



商用の分散表現学習ツール

(株)NTTデータ数理システム Text Mining Studio 類義語アドオン
https://www.msi.co.jp/tmstudio/TMSSynonymAddon.pdf

検索

顔料

	選択	単語	類似度
▶ 1	<input checked="" type="checkbox"/>	媒染剤	0.875
2	<input type="checkbox"/>	目止め	0.868
3	<input type="checkbox"/>	にかわ	0.860
4	<input type="checkbox"/>	シート状	0.860
5	<input type="checkbox"/>	膠	0.858

辞書登録 バックファへ

タグクラウド 樹形図

媒染剤 目止め にかわ シート状 膠 釉薬
ポスターカラー ラッカー 装飾用 支持体
アクリル系 象嵌 平織り 水性ペン 蒸着 染料 輪郭線
緑釉 油絵具 朱漆 漂白 媒剤 油粘土 水性 はり合わせる
離型剤 油性 水彩絵の具 上絵 シヤモット アクリル絵の具 淡彩 油性マジック 塗り重ねる

閉じる

検索履歴

- 顔料 名詞 一般
BCCWJ-L<C:#Program
表示件数 50
分かち書き以外の単語
- 顔料 名詞 一般
Wikipedia<C:#Program
表示件数 50
分かち書き以外の単語
- 顔料 名詞 一般
Wikipedia-L<C:#Progra
表示件数 50
分かち書き以外の単語

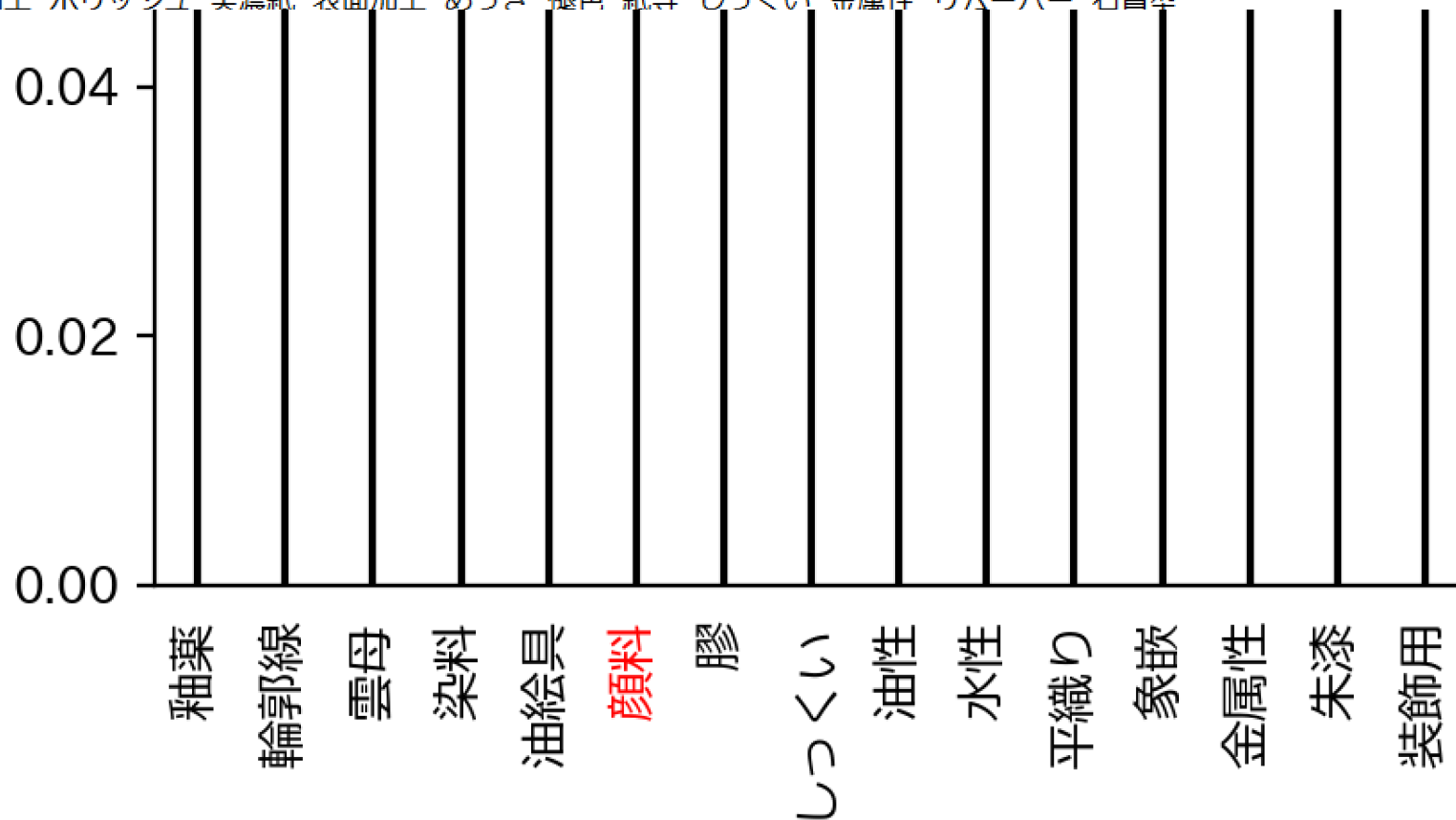
商用の分散表現学習ツール

媒染剤 目止めにかわ シート状 膠 釉薬 ポスターカラー

ラッカー 装飾用 支持体 アクリル系 象嵌 平織り 水性ペン 蒸着

染料 輪郭線 緑釉 油絵具 朱漆 漂白 媒剤 油粘土 水性 はり合わせる 離型剤

油性 水彩絵の具 上絵 シヤモット アクリル絵の具 淡彩 油性マジック 塗り重ねる 絵絹 ワニス マスキングテープ サンドペーパー 雲母
エンボス加工 ポリッシュ 美濃紙 表面加工 めっき 褪色 紙等 しっくい 金属性 リムーバー 石膏型



まとめと今後の検討課題

まとめ

- 特許文書のベクトル化方法7種類と文書分類8種類の組み合わせを検討した
- ・分類性能は⑤SCDV × ①XGBが良かった
 - ・計算時間を加味するとベクトル化方法では②TF-IDF、文書分類では④SVC、⑥RFgも選択肢である
 - ・パラメータチューニング、集団学習等により更なる分類性能の向上が見込める
- 実世界より解くべき問題(課題)を抽出して定式化し、機械学習を実行して、結果を吟味してアクションに繋げることが重要である

今後の検討課題 →続きはJapio Yearbook2019で

1. 特許文書のベクトル化手法

- ・文書、文、単語のラベル付きベクトル化
- ・トークナイズ(分かち書き)の影響(英語: subword、日本語: SentencePiece)
- ・deep-learning系の文書のベクトル化 1次元CNN,RNN,LSTM,BERT

2. 機械学習による文書分類

- ・発明の構成要素に対応した特徴量抽出、特徴量選択
- ・粒度(例: FI: 発明の主題、Fターム: 発明の記載要素、専門用語)の影響
- ・学習用教師データの質と量

3. 特許調査への応用検討

- ・多言語対応: 日本語、英語(トークナイザー対応済)、中国語(別途対応予定)
- ・特許分類(FI、Fターム)、統制語(PATOLISフリーキーワード)等One hotベクトル活用
- ・データソース(PATOLIS抄録、DWPI抄録)の影響

謝辞

免責

本報告は2019年の「アジア特許情報研究会」のワーキングの一環として報告するものである。本報の内容は筆者の私見であり所属機関の見解ではない。

謝辞

最後に機械学習の初心者である筆者を様々な形でサポートしていただいた研究会の研究活動で出会った多くの皆様に感謝申し上げます。

用語

fastText

Facebookが2016年8月18日に公開した自然言語処理ライブラリ
Googleが発表した「Word2Vec」の分散表現ベクトルを基に改良したもの
fastTextは、単語をsubword(部分語)に分解したものを学習に組み込んでいることが
最大のポイント

BERT

BERT(Bidirectional Encoder Representations from Transformers)は、広い範囲の
自然言語処理タスクにおいて最先端(state-of-the-art)の結果を得る言語表現事前
学習の新しい方法

XGBoost

XGBoost (eXtreme Gradient Boosting) は、勾配ブースティング木を使ったアルゴリズムを
オープンソース使って実装するソフトウェア。

Boosted treesは、Gradient BoostingとRandom Forestのアルゴリズムを組み合わせた
アンサンブル学習を行う。

XGBoostの主な特徴と理論の概要

<https://qiita.com/yh0sh/items/1df89b12a8dcd15bd5aa>

参考文献

- 1) 安藤俊幸, 機械学習を用いた効率的な特許調査方法
ーディープラーニングの特許調査への適用に関する基礎検討ー
Japio YEAR BOOK 2018, 2018, p. 238-249.
http://www.japio.or.jp/00yearbook/files/2018book/18_3_05.pdf

- 2) 安藤俊幸, 機械学習を用いた効率的な特許調査方法
ーニューラルネットワークの特許調査への適用に関する基礎検討ー
Japio YEAR BOOK 2017, 2017, p. 230-240.
http://www.japio.or.jp/00yearbook/files/2017book/17_3_04.pdf

- 3) 安藤俊幸, テキストマイニングと機械学習による効率的な特許調査
数理システムユーザーコンファレンス2017(2017年11月2日)
http://www.msi.co.jp/userconf/2017/pdf/muc17_501_2.pdf

- 4) 安藤俊幸, 機械学習を用いた効率的な特許調査
アジア特許情報研究会における研究活動紹介
「特技懇」誌, 2018.11.26. no.291
https://tokugikon.smartcore.jp/tokugikon_shi