

テキストマイニングを用いた 論文・社内報告書の分析事例

テキストマイニングを用いたナレッジの活用

株式会社村田製作所

藤田 聡



1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて

1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて

近未来では機械学習が使えて当たり前時代に

過去

表計算ソフト
文書作成ソフト
どう使う？
必要？



現在

表計算ソフト
文書作成ソフト
業務に不可欠
使えるのが当たり前

近未来

AI
機械学習
どう使う？
必要？



AI
機械学習
使えるのが当たり前

近未来では、こうなるのでは



- 専門スタッフによる機械学習活用（専門組織）
- 既に機械学習を活用している（社内に点在）

教育プログラム

各論

- 期間：約2ヶ月 概論1回（2h） + 各論4回（各半日）

演習

- 期間：約2.5ヶ月 演習問題にTry（グループワーク、研修）

追加

- 希望者に追加でテキストマイニング講座

課題

- 期間：約6ヶ月 実課題への取り組み

1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて

分析依頼システム

依頼書No.〇〇〇-□□□

(件名)

(分析目的)

製品Aの現象△△について走査型電子顕微鏡で観察し、表面組成を分析する

(分析結果概要)

試料表面を観察した結果...

表面組成を分析した結果...

走査型電子顕微鏡果報告

表面組成分析結果報告書

分析依頼書
分析試料

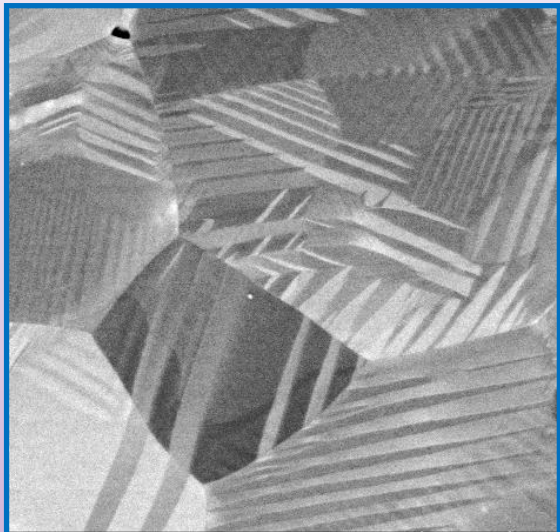
分析依頼者

分析担当者

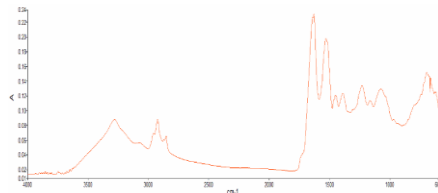
分析結果

社内に対して様々な分析サービスの提供

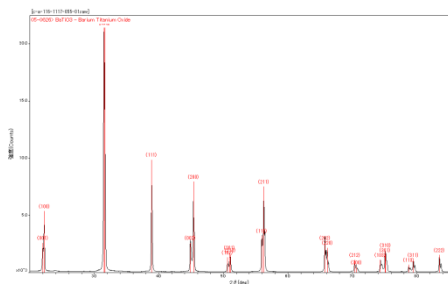
結晶粒子観察 (FE-SEM)



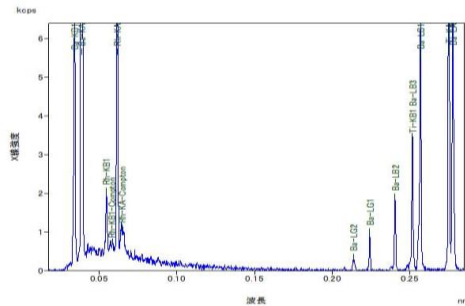
有機分析 (FT-IR)



結晶構造分析(XRD)



組成分析(XRF)



分析依頼システム

依頼書No.〇〇〇-□□□

(件名)

(分析目的)

製品Aの現象△△について走査型電子顕微鏡で観察し、表面組成を分析する

(分析結果概要)

試料表面を観察した結果...

表面組成を分析した結果...

走査型電子顕微鏡果報告

表面組成分析結果報告書

分析依頼書
分析試料

分析依頼者

分析担当者

分析結果

1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて

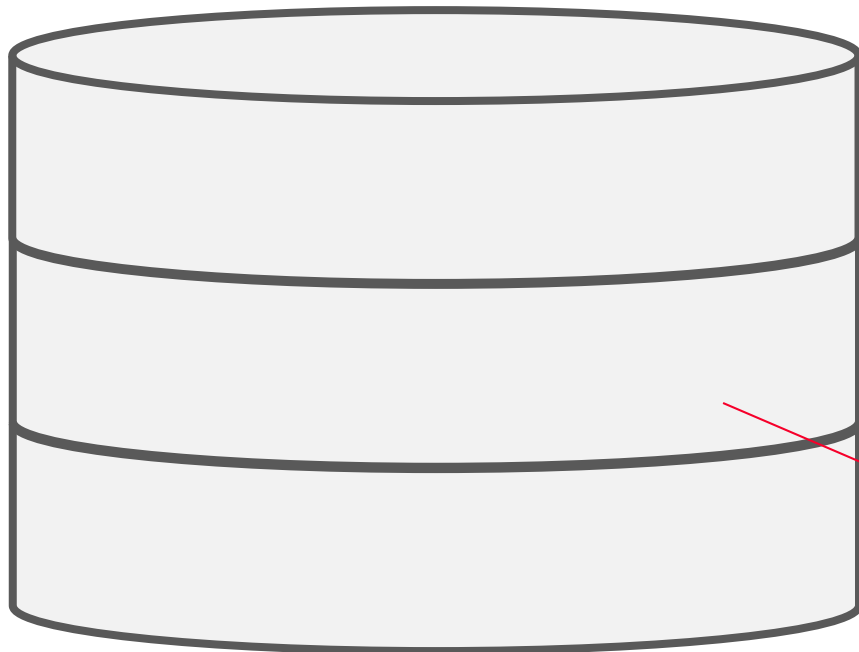
蓄積されたナレッジを活用し業務を効率化したい

- 分析技術者が自力で類似の案件を探し、分析指針を立案し、分析結果を考察する業務を効率化したい。
- 手がかりが少なく難しい分析依頼へのとっかかりが欲しい。

ただ、実際は、探すのが難しい...

情報量が多い
ファイルサーバー

情報が限定的な
分析依頼システムサーバー



依頼書No.○○○-□□□

(件名)

(分析目的)
商品Aの現象 (1) について
●●で部品組成を分析し、
□□で観察を行う

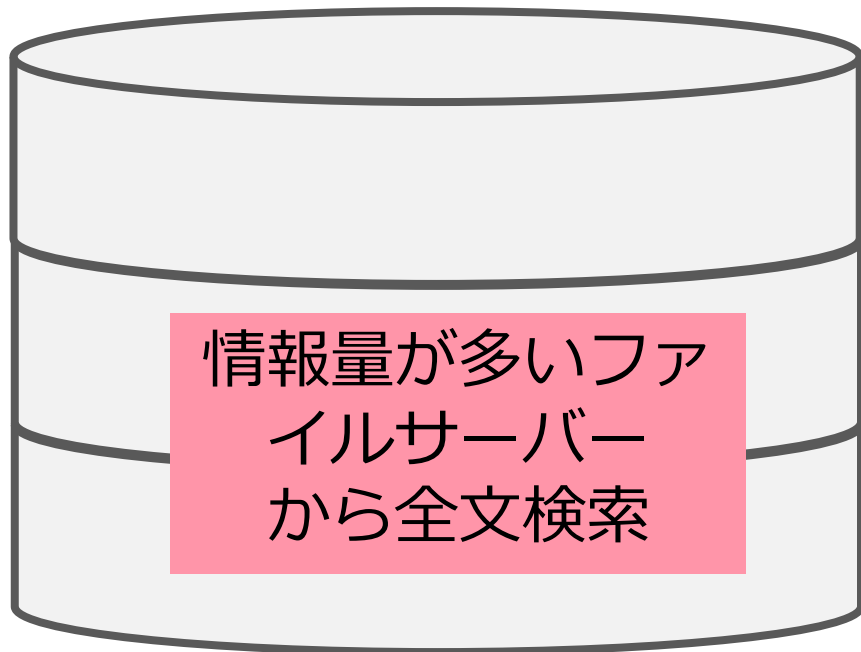
(分析結果概要)
●●分析結果から...
□□では...

●●分析結果報告 (ファイル)
□□分析結果報告 (ファイル)

情報量が多いファイル
サーバーから全文検索
で情報を得たい。

情報量が多い
ファイルサーバー

情報が限定的な
分析依頼システムサーバー



情報量が多いファ
イルサーバー
から全文検索

依頼書No.〇〇〇-□□□

(件名)

(分析目的)
商品Aの現象 (1) について
●●で表面組成を分析し、
□□で観察を行う

(分析結果概要)
●●分析結果から...
□□では...

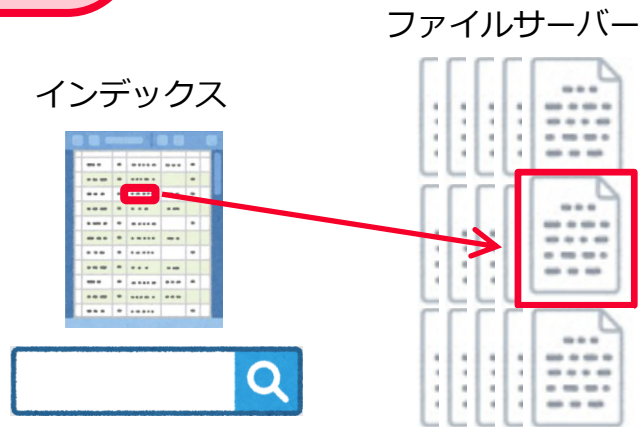
●●分析結果報告 (ファイル)
□□分析結果報告 (ファイル)



分析依頼システム
で学習

- テキストマイニング
- クラスタリングによる文書分類

検索エンジンの中で、自然言語処理による単語認識を行い、文書分類し、作成されたインデックスを参照して検索されます。





商品Aの分析依頼書を使った小規模な検討

- 探したい現象を検出できるか（テキストマイニング）
- 検出した結果から文書分類できるか（クラスタリング）

<取り組み課題の設定>

- 商品Aの分析依頼書（合計200件）
- 件名 分析目的 分析結果概要 を抽出

依頼書No.〇〇〇-□□□

(件名)

(分析目的)
商品Aの現象(1)について
●●で表面組成を分析し、
□□で観察を行う

(分析結果概要)
●●分析結果から...

□□では...現象(1)が---

●●分析結果報告 (ファイル)

□□分析結果報告 (ファイル)

The screenshot shows a table with three main columns: '件名' (Item Name), '分析目的' (Analysis Purpose), and '分析結果概要' (Analysis Result Summary). The text in the table is mostly blurred, but the structure matches the form shown on the left. A red arrow points from the '分析目的' section of the form to the corresponding column in the table.

<取り組み課題の設定>

7種類に分類

現象 (1)

現象 (2)

現象 (3)

現象 (4)

現象 (5)

現象 (6)

その他 (7)

- ・商品Aの分析依頼書 (合計200件)
- ・件名 分析目的 分析結果概要 を抽出
- ・依頼が多い**6つの現象**を検出できるか
- ・検出した結果から文書分類できるか



検討の流れ

①

分かち書きによる単語抽出

②

グルーピング機能を使ったキーワード（現象）検出

③

クラスタリングによる文書分類

②グルーピング機能を使ったキーワード（現象）検出

1つの現象に対して、異なる名称／略語、類似の現象があるが、同じ現象として認識されない。これらを1つのグループとして扱うことで認識可能になる

1つのグループ

現象（1）
-名称1
-名称1の略語
-名称2
-名称2の略語
-名称3
-類似現象(1)'

②グルーピング機能を使ったキーワード（現象）検出

カテゴリデータ → 検索インデックス

	A	B	C	D	E	F	G	H	I
1	文書	現象 (1)	現象 (2)	現象 (3)	現象 (4)	現象 (5)	現象 (6)	原文	
2	1	0	0	0	0	1	0	【相談管理No. 20	
3	2	0	0	0	0	0	0	【相談管理No. 20	
4	3	0	0	0	0	1	1	【相談管理No. 20	
5	4	0	0	0	0	1	0	【相談管理No. 20	
6	5	0	0	0	0	1	0	【相談管理No. 20	
7	6	0	0	0	0	1	0	【相談管理No. 20	
8	7	0	0	0	0	1	0	【相談管理No. 20	
9	8	0	0	0	0	1	0	【相談管理No. 20	

現象 (j) を

含む = 1

含まない = 0

②グルーピング機能を使ったキーワード（現象）検出

クラスタリングによる文書分類 → 文書分類の正解率

	A	B	C	D	E	F	G	H	I	J	K	L
1	文書	現象 (1)	現象 (2)	現象 (3)	現象 (4)	現象 (5)	現象 (6)	その他	ClusterID	答え	原文	
2	1	0	0	0	0	1	0	1	1	現象 (5)	【相談管理No. 210	
3	2	0	0	0	0	0	0	0	19	現象 (2)	【相談管理No. 200	
4	3	0	0	0	0	1	1	0	12	現象 (6)	【相談管理No. 200	
5	4	0	0	0	0	1	0	0	11	現象 (5)	【相談管理No. 200	
6	5	0	0	0	0	1	0	0	11	現象 (5)	【相談管理No. 200	
7	6	0	0	0	0	1	0	0	11	現象 (5)	【相談管理No. 200	
8	7	0	0	0	0	1	0	0	11	現象 (5)	【相談管理No. 200	
9	8	0	0	0	0	1	0	0	11	現象 (5)	【相談管理No. 200	
10	9	0	0	0	0	1	0	0	11	現象 (5)	Metro 工程不良 (

③ クラスタリングによる文書分類

- ハードクラスタリング結果

→ 文書分類の正解率は、現象によってバラツキ

文書分類	正解率 %
現象 (1)	80
現象 (2)	64
現象 (3)	67
現象 (4)	42
現象 (5)	100
現象 (6)	93
その他	93

表現が多様な現象や
複数の現象を含む文書は
正解率が低くなる



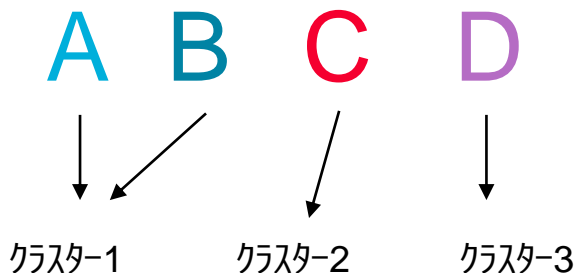
商品Aの分析依頼書を使った小規模な検討

- 探したい現象を検出できるか（テキストマイニング）
→ **現象は十分検出できた**
- 検出した結果から文書分類できるか（クラスタリング）
→ **現象（4）の正解率に課題（42% ⇒ 60%以上を目指したい）**

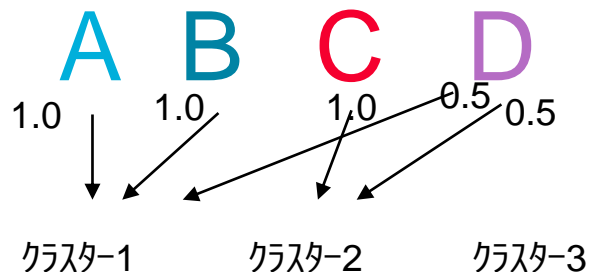
③ クラスタリングによる文書分類

- ソフトクラスタリングへの期待有り
 - クラスタへの所属確率で表現 (0~1)
 - テキストデータのクラスタリングが可能

ハードクラスタリング (イメージ)



ソフトクラスタリング (イメージ)



1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて



一連の検討から得られた課題

- 文書分類の正解率の要因詳細解析
- ソフトクラスタリングによるきめ細かい分類の検討
- グループピング漏れ対応が必要
- 小規模な検討でも異なる表現が非常に多く同義語も増える可能性大
- 実用化後のメンテナンスが不可欠

これらの課題に対応するとともに、自然言語処理と検索エンジンを組み合わせたシステムの構築を行っていきたい。

1. 当社の機械学習への取り組み
2. テキストマイニングの対象
3. 課題・取り組みについてご紹介
4. 今後の課題
5. 教育プログラム、コンサルティングについて

- レベルに応じて、わかるまで、懇切丁寧
- 専門業務では無く、短時間で使いこなしたい場合に有効

TMS、VAPを使ってみて...

- ユーザー辞書、同義語辞書の作成が非常に効率的
- グループ機能では原文に元の言葉を残したまま検出できる
- 初心者でもコーディング無しで利用可能
- ソフトクラスタリング（新しいクラスタリング法）への期待有り



ご清聴ありがとうございました