

テキストマイニングを使用した 新市場の探索について

会員 川上 成年

要 約

自社技術シーズを活用できる新市場の検討手法に、MFT フレームがある。MFT フレームとは、市場ニーズと企業サイドの技術をうまく接続し、商品開発のヒントを得るためのフレームワークである。MFT フレームを活用すると、ある技術シーズがあったとき、その技術特性の中で、いままで着目されなかった機能を生かして商品化するという発想が得られる。本稿では、特許情報のテキストマイニングにより、市場用語、機能用語、及び、技術用語の抽出と構造化を実行し、MFT モデルの作成を試みる。さらに、テキストマイニングにより生成された MFT モデルを用いて新市場の探索を試みる。

目次

- はじめに
 - 1.1 MFT フレームについて
- 本稿の目的
 - 2.1 テキストマイニングによる MFT モデルの作成
 - 2.2 MFT モデルを用いた新市場の探索
- テキストマイニングを使用した MFT モデルの作成について
 - 3.1 考え方と作成手順
 - 3.2 用語の抽出
 - 3.3 用語の分類
 - 3.4 用語の構造化
- MFT モデルを使用した新市場の探索について
 - 4.1 考え方と作成手順
 - 4.2 市場用語の抽出
 - 4.3 市場用語の構造化
- おわりに

場に届けられなければ意味がないと考える。MFT フレームを活用すると、ある技術シーズがあったとき、その技術特性の中で、いままで着目されなかった機能を生かして商品化するという発想が得られる。

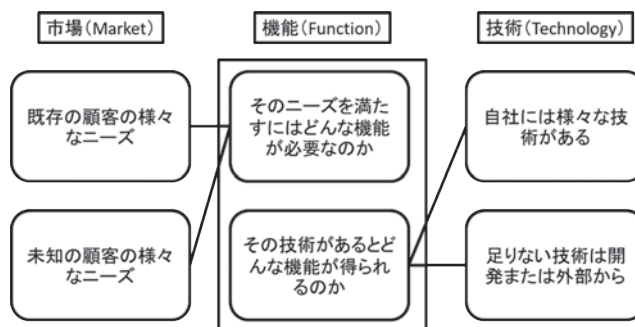


図 1 MFT フレーム

2. 本稿の目的

2.1 テキストマイニングによる MFT モデルの作成

本稿では、特許情報のテキストマイニングにより、市場、機能、及び、技術用語の抽出と構造化を実行し、MFT モデルの作成を試みる。本稿では仮想事例として、特定の掃除機メーカーである D 社を取り上げ、D 社の MFT モデルを作成する。

2.2 MFT モデルを用いた新市場の探索

テキストマイニングにより生成された MFT モデルを用いて新市場の探索を試みる。本稿では仮想事例として、D 社の技術シーズについて、その機能を生かすことができる新たな市場の探索を試みる。

1. はじめに

1.1 MFT フレームについて

MFT フレームとは、市場ニーズと企業サイドの技術をうまく接続し、商品開発のヒントを得るためのフレームワークである⁽¹⁾。MFT フレームの M は市場 (Market) を意味し、F は機能 (Function) を意味し、T は技術 (Technology) を意味する。

図 1 に示すように、MFT フレームでは、要素技術と市場ニーズとの間に機能という概念を置くことで、商品化や事業化のイメージを容易にする。

MFT フレームでは、技術は、提供機能を通じて市

3. テキストマイニングを使用した MFT モデルの作成について

3. 1 考え方と作成手順

① 本稿の MFT モデルについて

図 2 は、本稿の MFT モデルの概念図である。本稿の MFT モデルは、多空間デザインモデル⁽³⁾を MFT フレームに適合するモデルに変形したものである。

多空間デザインモデルとは、使用される知識や手法も異なる様々な領域におけるデザイン行為を包括的視点により表現するモデルである。多空間デザインモデルを使用した発想法と分析法を活用することで、的確なデザイン思考、及び、新規なデザイン解の導出を行うことが可能となる。

多空間デザインモデルでは、デザイン思考を心理要素群から物理要素群への変換ととらえ、心理的なデザイン要素が表現される心理空間と、物理的なデザイン要素が表現される物理空間と、に空間を大きく分割する。心理空間は価値空間と意味空間から構成され、物理空間は状態空間と属性空間から構成される。

価値空間では、文化的価値や機能的価値など様々な視点からの価値を表現するデザイン要素とそれらの関係が表現される。意味空間では、デザイン対象の持つ機能性や品質やイメージなどを表現するデザイン要素とそれらの関係が表現される。状態空間では、デザイン対象が置かれる時空間に関する環境や条件を表現する要素とそれらの関係が表現される。属性空間では、寸法、材料等、技術的内容を表現するデザイン要素とそれらの関係が表現される。

本稿の MFT モデルは、市場空間、機能空間、及び、技術空間の 3 つの空間から構成された多空間デザインモデルである。

市場空間では、商品、サービスを表現するデザイン要素とそれらの関係が表現される。機能空間では、機能性や品質などを表現するデザイン要素とそれらの関係が表現される。技術空間では、機構、部品、材料等、技術的内容を表現するデザイン要素とそれらの関係が表現される。

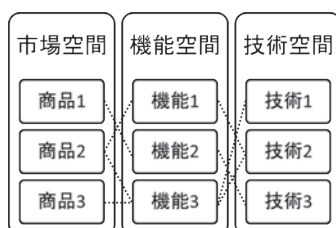


図 2 本稿の MFT モデル

本稿の MFT モデルは、市場、機能、技術に分類される要素の関係性に着目し、新規なデザイン解（すなわち新市場）の導出を行うことを目的とする。

② 特許情報からの市場、機能、及び、技術用語の抽出

市場に関する用語は、発明の名称として、特許情報から抽出可能である。また、機能に関する用語、及び、技術に関する用語は、発明として、要約から抽出可能である。

本稿では、特許情報から市場に関する用語、機能に関する用語、及び、技術に関する用語、並びに、それらの関係性を抽出することにより、MFT モデルを作成する。

③ テキストマイニングを用いたモデル化

テキストマイニングとは、文書形式のデータを定量的な方法で分析することをいう。テキストマイニングでは、コンピュータの処理によってデータの中から自動的に言葉を取り出し、様々な統計手法を用いた探索的な分析を行う。これにより、パターンやルール、ひいては知識の発見を目指すものである。

従来、特許情報から市場用語、機能用語、及び、技術用語を抽出するためには、人間が 1 件 1 件明細書を読み込んで、その意味内容を理解したのち、市場用語、機能用語、及び、技術用語を明細書から抽出する作業を行うことになる。

人間が明細書を読み込んで、市場用語、機能用語、及び、技術用語を抽出する場合には、熟練者でも明細書 1 件当たり、5~10 分程度の時間が必要である。そうすると明細書 100 件処理するには 9~18 時間程度の時間が必要となる。これでは、多忙な知財担当者等が、業務として行うのは難しい。作業時間に制限がある場合には、分析できる明細書の件数も少なくせざるを得ない。

これらキーワード抽出やコーディングの処理をテキストマイニングにより処理する方法がある⁽²⁾。そこで、本稿では、テキストマイニングを利用して、市場用語、機能用語、及び、技術用語の抽出作業を自動化することにより作業を効率化する。これにより、大量の特許情報を処理することができ、人間が気がつかない新市場を発見できる可能性も高まる。

本稿では、KH Coder⁽⁴⁾を使用してテキストマイニ

ングを行う。KH Coder は、テキストマイニング用のフリーソフトウェアである。フリーソフトであることから、学術的な目的であれ商業的な目的であれ、無料で自由に使用できる。

KH Coder は、<http://khcoder.net> からダウンロードできる。KH Coder については、KH Coder に添付されるマニュアルや、解説書等が存在する。これらを参照することにより、容易に操作を理解できる。

なお、本稿では、KH Coder を使用することを前提に記載されているため、操作について不明な点があれば、マニュアルを適宜参照いただきたい。

④ 作成手順

図3にMFTモデルの作成手順を示す。

用語の抽出 (STEP1) では、特許情報を特許データベースから取得し、KH Coder を使用して、形態素解析を行う。

用語の分類 (STEP2) では、KH Coder を使用して、トピック抽出及びコーディングを行う。

用語の構造化 (STEP3) では、KH Coder を使用して共起ネットワークを出力し、表計算ソフト等を使用してMFTモデルを作成する。

以下、STEP1~3について、詳細に説明する。



図3 MFTモデルの作成手順

3. 2 用語の抽出 (STEP1)

① 特許情報取得

特許情報を特許データベースから取得する。本稿では、検索対象企業名を、「D社」として特許検索し取得した938件の特許出願を分析の対象とした。特許情報は、J-PlatPatなどの特許情報データベースから取得すればよい。

特許情報としては、課題及び解決手段の内容が短くまとめられている要約書の情報をダウンロードした。

要約書は、課題と解決手段が項目立てて記載されているため、技術の内容を把握するために都合が良い。

② 形態素解析

938件の要約データに対して形態素解析を行った。形態素解析とは、文法的な情報の注記の無い自然言語のテキストデータ(文)から、対象言語の文法や、辞書と呼ばれる単語の品詞等の情報に基づき、形態素(言語で意味を持つ最小単位)の列に分割し、それぞれの形態素の品詞等を判別する処理のことをいう。これにより、分析対象となる市場用語、機能用語、及び、技術用語を抽出した。

3. 3 用語の分類 (STEP2)

① コードの設定

まず、コーディング処理用のコードを設定する。コーディングとは、仮説やテーマに基づいて、抽出語を組み合わせる処理のことをいう。コーディング処理を行う前には、いくつかのカテゴリー(以下、コードという)を設定し、「特定の記述がデータ中にあればそのデータを特定のコードに分類すること」といった基準を設定する。この基準のことをコーディングルールという。

本稿では、D社の938件の特許出願に対し、潜在的ディリクレ配分法(Latent Dirichlet Allocation:以下LDAとする)による処理を行い、機能トピックモデルと技術トピックモデルを生成し、コードの設定を行った。

LDAとは、1つの文書が複数の潜在的トピックから確率的に生成されると仮定した言語モデルの一種である。文書内の各単語はあるトピックが持つ確率分布に従って出現すると仮定される。LDAでは、トピックごとに単語の出現頻度分布を想定することで、トピック間の類似性やその意味を解析できる。

LDAでは、自然言語からなる複数の文章(ここでは、要約書)を形態素解析して複数の用語を抽出し、複数の特許情報データの各々における複数の用語の出現頻度(出現回数)からベクトル空間モデル(文書における単語の出現頻度を行列の形で表現したもの)を求め、このベクトル空間モデルを潜在的ディリクレ配分法で処理することにより、複数のトピックと当該複数のトピックの各々における出現確率に応じた複数の用語とを抽出する。

なお、本稿のLDA処理には、KH Coderに実装された機能を使用しているため、上記のようなLDAの原理を知らずとも、簡単にLDA処理を行うことができる。

本稿では、機能トピックモデルと技術トピックモデルとを別個に生成した。そして、これら機能トピックモデルと技術トピックモデルとから、機能コードと技術コードとを設定した。

なお、市場コードについては、トピックモデルを使用せず、商品の名称に関する用語を強制抽出し、これを市場コードとした。

図4は、LDAにより生成された機能トピックモデルを示す。図に示すように、機能トピックは14個となり、各機能トピックには単語の集合が割り当てられている(図4の左側の「トピック」と「単語の集合」の列)。

各機能トピックに割り当てられた単語の集合から、各トピックの意味するところを人間が解釈して、機能トピックにおける各トピックの名称(分類)を機能コードとして図4に示すように設定した(図4の右側の「機能コード」の列)。

トピック	単語の集合	機能コード
#1	制御, 供給, 励起, 応答, 測定	制御性
#2	位置, 直立, 格納, 移動, 開放	格納性
#3	清浄, 通過, 具備, 再生, 貫通	清浄性
#4	収容, 保持, 開示, 離間, 隣接	収容性
#5	回転, 付属, 加熱, 結合, 除去	加熱性
#6	駆動, 移動, 検出, 自律, 装着	移動性
#7	支持, 処理, 排出, 吸入, 制限	排出性
#8	分離, 収集, 含有, 有, クリーニング	分離性
#9	使用, 清掃, 動作, 改善, 接触	使用性
#10	接続, 作動, 出力, 入力, 選択	接続性
#11	吸引, 掃除, 連結, 操作, 防止	吸引力
#12	配置, 固定, 圧縮, 振動, 低減	圧縮性
#13	放出, 生成, ロック, 解除, キャッチ	放出性
#14	形成, 発生, 加湿, 噴出, 堆積	加湿性

図4 機能トピックモデル

図5は、LDAにより生成された技術トピックモデルを示す。図に示すように、技術トピックは17個となり、各技術トピックには単語の集合が割り当てられている(図5の左側の「トピック」と「単語の集合」の列)。

各技術トピックに割り当てられた単語の集合から、各トピックの意味するところを人間が解釈して、技術トピックにおける各トピックの名称(分類)を技術コードとして図5に示すように設定した(図5の右側の「技術コード」の列)。

なお、意味を解釈不能なトピックについては、設定不能として、技術コードを割り当てていない。

トピック	単語の集合	技術コード
#1	真空, チャンバ, 塵埃, 汚れ, コネクタ	チャンバ構成
#2	部分, 外側, 内側, アーム, フレーム	アーム構成
#3	立体, ファン, 手段, ベース, ロボット	ファン構成
#4	モータ, ロータ, 磁石, ホイール, 方法	モーター構成
#5	システム, 機械, 電圧, 段階, 信号	制御構成
#6	本体, 部材, ホース, 手段, 部品	ホース構成
#7	空気, 出口, 入口, ケーシング, 羽根車	ケーシング構成
#8	フィルタ, アセンブリ, 材料, 構造, 容器	フィルタ構成
#9	ノズル, 内部, 通路, 組立, ポア	ノズル構成
#10	サイクロン, ユニット, 複数, 下流, 上流	サイクロン構成
#11	流体, 入口, 出口, 流れ, 機器	設定不能
#12	ハンドル, ユーザ, 機構, 状態, 機器	ハンドル構成
#13	ダクト, 1つ, 気流, エア, ゴミ	ダクト構成
#14	開口, 攪拌, 導管, ツール	攪拌構成
#15	ヘッド, 表面, ハウジング, ブラシ, 剛毛	ヘッド構成
#16	方向, 周り, 外壁, 長手, 内壁	設定不能
#17	電気, 器具, シャーシ, 手持ち, チューブ	シャーシ構成

図5 技術トピックモデル

市場コード	コーディングルール
掃除機	掃除機 or 清掃器具 or 掃除器具
ヘアケア機器	ヘアケア機器 or ヘアドライヤー or ヘアケア電気器具
送風機	送風機
乾燥装置	乾燥装置

図6 市場コードとコーディングルール

機能コード	コーディングルール
制御性	制御 and (供給 or 励起 or 応答 or 測定)
格納性	格納 and (位置 or 直立 or 移動 or 開放)
清浄性	清浄 and (通過 or 具備 or 再生 or 貫通)
収容性	収容 and (保持 or 開示 or 離間 or 隣接)
加熱性	加熱 and (回転 or 付属 or 結合 or 除去)
移動性	移動 and (駆動 or 検出 or 自律 or 装着)
排出性	排出 and (支持 or 処理 or 吸入 or 制限)
分離性	分離 and (収集 or 含有 or 有 or クリーニング)
使用性	使用 and (清掃 or 動作 or 改善 or 接触)
接続性	接続 and (作動 or 出力 or 入力 or 選択)
吸引力	吸引 and (掃除 or 連結 or 操作 or 防止)
圧縮性	圧縮 and (配置 or 固定 or 振動 or 低減)
放出性	放出 and (生成 or ロック or 解除 or キャッチ)
加湿性	加湿 and (形成 or 発生 or 噴出 or 堆積)

図7 機能コードとコーディングルール

② コーディング

図6~8は、市場、機能、及び、技術コードと、それらに対応するコーディングルールを示す。コーディング処理では、特定の記述がデータ中にあればそのデータを特定のコードに分類すること、といった基準を設定する。この基準のことをコーディングルールという。

本稿では、LDAにより生成された機能トピックモデルと技術トピックモデルの各トピックに割り当てられた単語の集合から、コーディングルールを設定した。

具体的には、各トピックに割り当てられた単語の集合から、特徴となる単語を1つ選び、他の単語同士は“or”で接続し、それらと、特徴となる単語とを“and”で接続した。

技術コード	コーディングルール
チャンバ構成	チャンバ and (真空 or 塵埃 or 汚れ or コネクタ)
アーム構成	アーム and (部分 or 外側 or 内側 or フレーム)
ファン構成	ファン and (立体 or 手段 or ベース or ロボット)
モーター構成	モータ and (ロータ or 磁石 or ホイール or 方法)
制御構成	システム and (機械 or 電圧 or 段階 or 信号)
ホース構成	ホース and (本体 or 部材 or 手段 or 部品)
ケーシング構成	ケーシング and (空気 or 出口 or 入口 or 羽根車)
フィルタ構成	フィルタ and (アセンブリ or 材料 or 構造 or 容器)
ノズル構成	ノズル and (内部 or 通路 or 組立 or ポア)
サイクロン構成	サイクロン and (ユニット or 複数 or 下流 or 上流)
ハンドル構成	ハンドル and (ユーザ or 機構 or 状態 or 機器)
ダクト構成	ダクト and (1つ or 気流 or エア or ゴミ)
攪拌構成	攪 and (開口 or 拌 or 導管 or ツール)
ヘッド構成	ヘッド and (表面 orハウジング or ブラシ or 剛毛)
シャーシ構成	シャーシ and (電気 or 器具 or 手持ち or チューブ)

図8 技術コードとコーディングルール

3. 4 用語の構造化 (STEP3)

① 共起ネットワーク出力

図9は、市場コード、機能コード、技術コードの共起ネットワークを示す。図9の共起ネットワークは、特定のコード（例えば、「分離性」）と他のコード（例えば「清浄性」）とのJaccard係数を計算して生成されたものである。共起度が大きいほど高い共起関係ある、すなわち、同じ文書中に共に出現することが多いといえる。

コード間の共起度が所定値以上（例えば、0.2以上）の場合に、コード間に関連性があるとみなす。これにより、用語を構造化する。

図9に示すように、共起ネットワークでは、市場コード、機能コード、及び、技術コードの各用語は、市場空間、機能空間、及び、技術空間の区別なく配置される。このため、市場コード、機能コード、及び、技術コードの関係がわかりにくいものとなる。

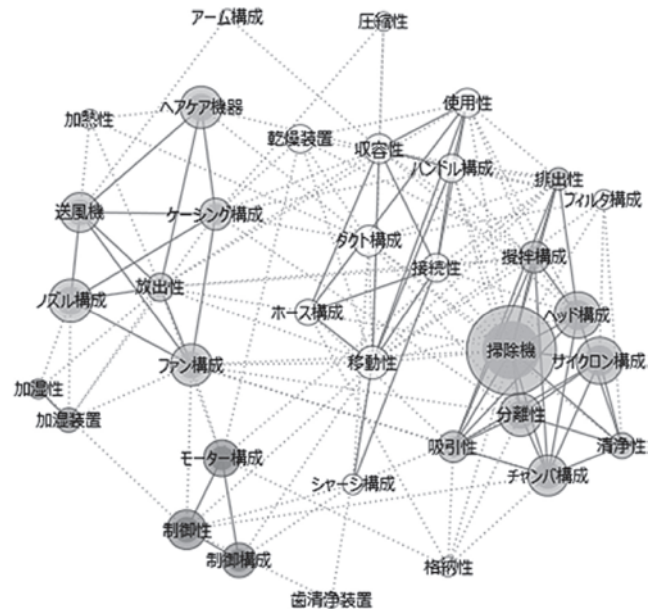


図9 共起ネットワーク

② MFT モデル作成

図10は、作成されたMFTモデルを示す。共起ネットワーク中の市場コード、機能コード、及び、技術コードを、市場空間、機能空間、及び、技術空間に分けて配置して、MFTモデルを作成する。（なお、図の見やすさを向上するために各コードを上位概念化したグループを新たに加えている。）

MFTモデルでは、市場コード、機能コード、及び、技術コードが、市場空間、機能空間、及び、技術空間に分けて整理される。このため、市場コード、機能コード、及び、技術コードの関係がわかりやすくなる。

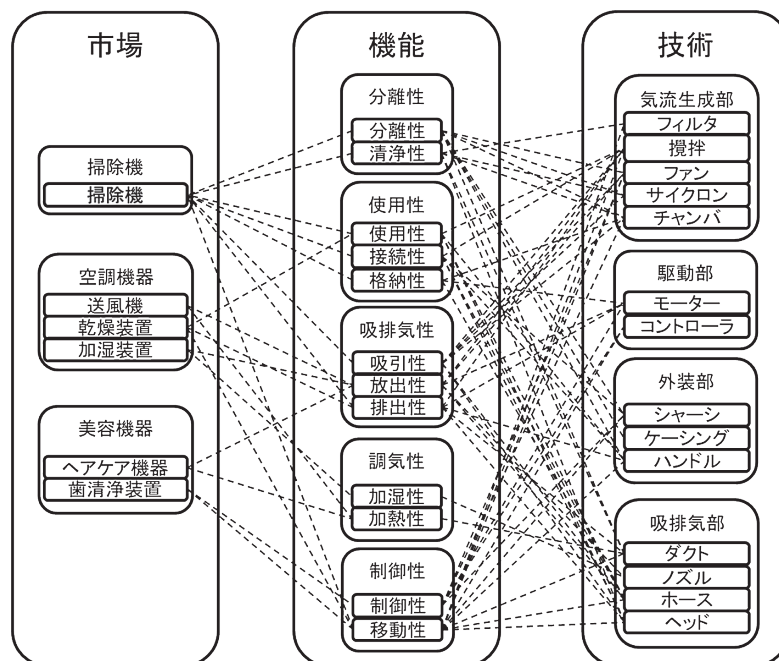


図10 MFTモデル

4. MFT モデルを使用した新市場の探索について

ここでは、D 社が有する「サイクロン構成」という技術シーズに着目し、その主な機能である「分離性」を生かした新市場の探索を試みる。

4. 1 考え方と作成手順

① 新市場の探索の考え方

多空間発想法⁽⁵⁾では、デザイン要素（本稿でいう市場用語、機能用語、技術用語）の分解や追加を行うことより、新規デザイン解を発想する。市場用語の分解や追加には、アイデアを詳細化・再検討し、新たな市場用語を発見する必要がある。

本稿では、テキストマイニングを使用して、市場用語の追加を行い、新市場の探索を試みる。以下、本稿における新市場の探索の考え方について述べる。

図 11 は、本稿における新市場の探索の考え方を示す。本稿では、大量の特許情報を収集し、図 11 に示すように、D 社について作成した MFT モデルと共起するノードを特許情報から抽出し、新市場の探索を試みる。

また、本稿では、市場用語の収集は、D 社の事業分野外の特許情報からも収集することとする。異なる事業分野から市場用語を収集することにより、意外性のある新市場を発見できる可能性がある。

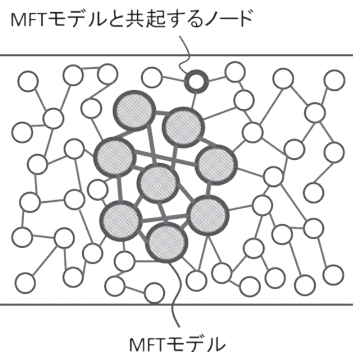


図 11 MFT モデルとの共起

② 作成手順

図 12 に市場用語の追加の手順を示す。

市場用語の抽出 (STEP4) では、特許情報を特許データベースから収集し、KH Coder を使用して、形態素解析及びコーディング処理を実行する。

市場用語の構造化 (STEP5) では、KH Coder を使用して、共起ネットワークを出力し、共起ネットワークから市場用語を抽出して MFT モデルを改変する。

以下、STEP4, 5 について、詳細に説明する。

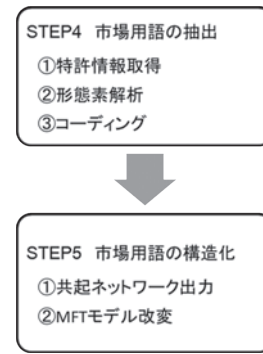


図 12 新市場探索の手順

4. 2 市場用語の抽出 (STEP4)

① 特許情報取得

特許情報を特許データベースから取得する。

本稿では、「分離性」という機能を適用することができる新市場の探索を目的とするため、機能コード「分離 and (収集 or 含有 or 有 or クリーニング)」or「清浄 and (通過 or 具備 or 再生 or 貫通)」を特許検索の検索式に用いた。

この検索式による特許検索により、特許情報を 2989 件取得した。特許情報としては、要約書の情報を取得した。

② 形態素解析

特許情報 2989 件の情報に対して形態素解析を行う。これにより、特許情報 2989 件から、分析対象となる用語を抽出する。

③ コーディング

図 6~8 に示す市場コード、機能コード、及び、技術コードと各コードに対応するコーディングルールにより、形態素解析で得られた用語のコーディング処理を実行する。

4. 3 市場用語の構造化 (STEP5)

① 共起ネットワーク出力

図 13 は、市場用語と機能コード「分離性」との共起ネットワークを示す。本稿では、MFT モデルの一部である機能コード「分離性」との共起ネットワークを生成している。

図 13 に示すように、機能コード「分離性」を構成する機能用語（図中の四角で囲われた文字）と共起する市場用語（図中の丸で囲われた文字）は大量に存在する。これら多数の市場用語は、機能コード「分離

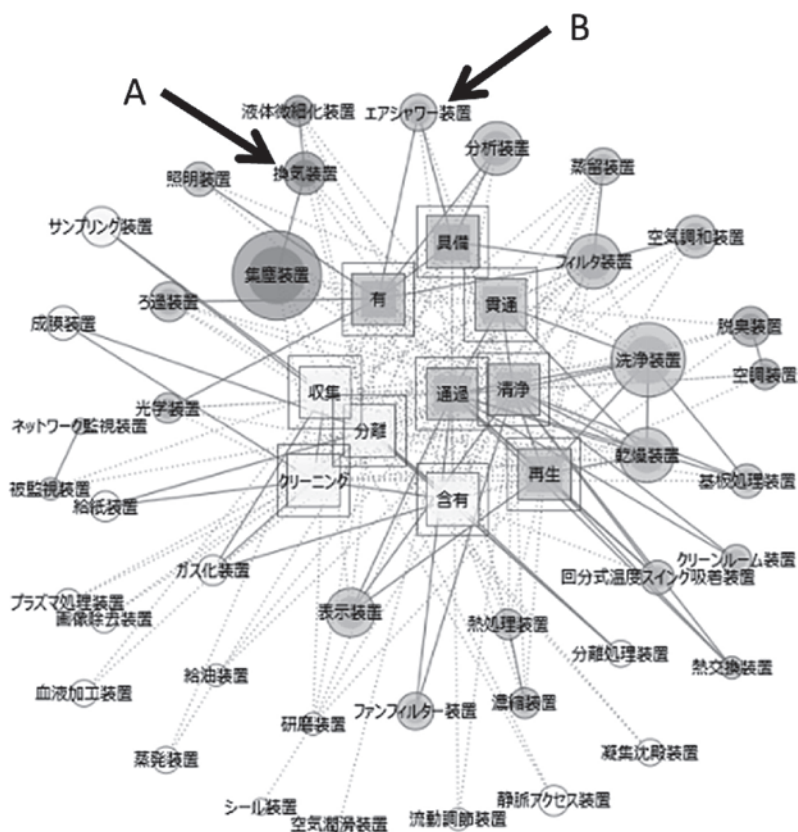


図 13 共起ネットワーク

性」と何らかの共起関係を有する市場用語といえる。

収集した特許情報は D 社の業務範囲以外の対象を含むことから、D 社の業務範囲に限定されることなく幅広い市場用語が抽出されていることがわかる。

図 13 に示すように、機能コード「分離性」に関連する市場用語として、「エアシャワー装置」や「換気装置」があることがわかる（図 13 中の矢印 A, B）。

図 14 は、図 13 に示す共起ネットワークに存在する市場用語のリスト（一部）を示す。

ここでは、この「エアシャワー装置」（矢印 B）や「換気装置」（矢印 A）という市場用語を、機能コード

「分離性」を適用できる新市場の候補として選択した。

② MFT モデルの改変

図 10 の MFT モデルに、上記選択された市場用語を追加して、MFT モデルを改変する。

図 15 は、改変された MFT モデルである。図 15 に示すように、市場空間に、住設機器（「エアシャワー装置」と「換気装置」の上位概念）が追加されている。

市場空間に「住設機器」を追加することにより、機能用語「分離性」及び技術用語「サイクロン」との関係性が構築できると推測される。

このようにして、新たな市場用語を追加して構造化することにより、技術シーズ「サイクロン」を活用可能な、新市場を探索することができる。

また、この MFT モデルは特許情報から構成されているため、「エアシャワー装置」と「換気装置」の用語の背後には、該当する複数の特許出願がバックデータとして存在する。これら特許出願の内容を分析することにより、競合相手や協業先の検討もでき、また、技術的な実現性も検討することが可能である。

	抽出語	共起	Jaccard
	集塵装置	38	0.0138
	洗浄装置	26	0.0094
	乾燥装置	14	0.0051
	フィルタ装置	14	0.0051
	分析装置	13	0.0047
	表示装置	11	0.0040
A →	換気装置	8	0.0029
	空気調和装置	8	0.0029
	ろ過装置	7	0.0025
	サンプリング装置	7	0.0025
	脱臭装置	7	0.0025
B →	エアシャワー装置	6	0.0022
	ファンフィルター装置	6	0.0022
	蒸留装置	6	0.0022
	回分式温度スイング吸着装置	5	0.0018
	基盤処理装置	5	0.0018
	照明装置	5	0.0018
	成膜装置	5	0.0018

図 14 市場用語のリスト（一部）

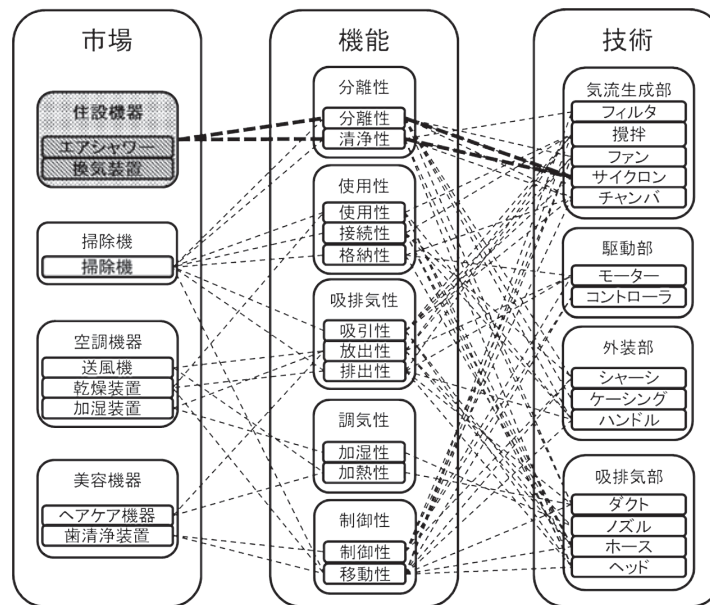


図 15 新規 MFT モデル

5. おわりに

5. 1 テキストマイニングを使用した MFT モデルの作成について

本稿では、特許情報から MFT モデルを作成した。MFT モデルの作成には、LDA による処理を使用した。これにより、MFT モデルの作成作業を省力化できた。一方、LDA のさらなる有効な使用法については今後の課題となる。

5. 2 MFT モデルを使用した新市場の探索について

新市場の探索は、MFT モデルと共起する市場用語を抽出することにより行い、新市場の探索の一助とできることが確認できた。一方、新市場探索の抽出基準

については今後の課題となる。

(参考文献)

- (1) 嶋田毅 執筆, 図解 基本フレームワーク 50, ダイヤモンド社 (2015)
- (2) 川上成年, テキストマイニングを使用した多空間デザインモデルの作成について, パテント, Vol.74, No.3 (2021)
- (3) 松岡由幸 監修, デザイン科学概論, 慶應義塾大学出版会, p.46-53, (2018)
- (4) 樋口耕一, 社会調査のための計量テキスト分析 内容分析の継承と発展を目指して, 株式会社ナカニシヤ出版 (2014)
- (5) 高野修治, 佐藤浩一郎, 松岡由幸, 多様なユーザへの対応に向けた多空間デザイン法の事例適用と思考分析, 日本デザイン学会, デザイン学研究 (2013)

(原稿受領 2021.12.14)