

## テキストマイニング技術の特許データへの適用

Application of Text mining Technology to Patent Data

林 田 英 雄

**要 約** 最近特許に関する話題が新聞上でよく見られるようになってきた。従来からの特許の話題は特許侵害と呼ばれる、他の出願人の特許を侵害したケースがほとんどであった。ところが最近では、発明報酬に絡んだ特許の価値評価に関する話題、特定領域の自社・他社の特許を分析して研究開発戦略や特許戦略の方向性を意思決定したり、特許ポートフォリオの競争力や優位性を把握し技術経営を行うといった話題が増えてきている。

SBI インテストラ株式会社では、これらの市場のニーズに応えるために知財経営システムソリューション StraVision を開発した。日本ユニシスはこの StraVision の開発に参加し、データマイニング、テキストマイニングの技術とソリューション MiningPro 21 文書マイニング・システムを適用した。

本稿では、StraVision による特許分析を紹介し、その開発における日本ユニシスのソリューション MiningPro 21 文書マイニング・システムの適用について述べる。

**Abstract** Recently, a lot of articles about the patent have often come to be seen on the newspaper. The previous articles about the patent consisted mostly of cases called patent infringement where a particular patent is infringed by other applicants.

However, the articles about patents are getting more and more, including one about the value evaluation of patent involving the compensation for the invention of the patent, and one about performance of the decision making about the direction of R&D and patent strategies by analyzing one's and other company's patents in the specific domain, and the technology management resulting from understanding both of the competitive power and advantage of the patent portfolio.

In order to meet the above marketing needs, SBI Intechstra Co., Ltd. has developed an intellectual property management system solution StraVision. Nihon Unisys participated in development of this StraVision, and applied the data mining and text mining technologies, and the solution of MiningPro 21 text mining system.

This paper introduces the patent analysis by StraVision, and discusses the application of MiningPro 21 text mining system of Nihon Unisys performed in the development processes.

### 1. はじめに

最近特許に関する話題が新聞上でよく見られるようになってきた。従来からの特許の話題は、特許侵害と呼ばれる他の出願人の特許を侵害したケースがほとんどであった。ところが最近では、発明報酬に絡んだ特許の価値評価に関する話題、企業の技術戦略立案のための特許を通じた技術分析に関する話題が増えてきている。特許の価値評価は、発明者の属する企業が発明者に対して払う発明報酬の算定や、特許を企業の無形資産として価値評価を行い信託する際に用いられる。もう一方の特許を通じた技術分析は、特定領域の技術を持つ自社と競合他社について、保有する特許を複数の切り口で分類し評価することで、社員や開発資金など限られた資源を投

下する際の技術開発戦略の立案において非常に重要である。

SBI インテクストラ株式会社（以下 SBI インテクストラ社）では、これらの市場のニーズに応えるために特許データベースの提供から、管理、分析、経済価値評価までのすべてをサービスする知財経営システムソリューション StraVision を開発し、2005 年 6 月サービスを開始した。日本ユニシスはこの StraVision の開発全般を担当しており、そこにデータマイニング、テキストマイニングの技術とソリューション MiningPro 21 文書マイニング・システムを適用した。

本稿では、StraVision による特許分析を紹介し、その開発における MiningPro 21 文書マイニング・システムの適用について述べる。

## 2. 知財経営システムソリューション StraVision

### 2.1 StraVision とは

StraVision は、自社と競合他社の特許を整理、分類して企業の技術を正確に把握、比較することで技術戦略立案を支援し、更に特許を資産価値評価することで、企業的意思決定を支援するためのシステムソリューションである。その主たる機能には、最新特許データの提供、分析対象の特許データを整理し分類するポートフォリオ管理、分類された特許データを通して自社と競合他社の技術を分析するポートフォリオ分析、競合他社における発明者のグループを特定し、その研究課題を通して競合他社の技術動向を分析する発明者グループ分析、各特許についてその価値を金額で算定する経済価値評価がある。

StraVision は 1985 年以降の日米の特許データを提供している。特許庁から提供されている特許情報には、公開、公告、登録、公表、再公表という特許公報のデータ、特許庁内の特許データの発生、更新情報である整理標準化データがある。一般の特許検索システムでは、各公報や整理標準化データの情報が個別に提供される。StraVision の特許データベースは分析データとして活用するためのもので、各公報や整理標準化データの必要な項目を一元管理しており、分析で利用する情報を特許単位に統合して提供している。このためユーザは必要となる情報を寄せ集めて分析用データを作成する必要はない。

ポートフォリオ管理は、StraVision の特許データベースから分析の対象となる特許データを抽出し、整理、分類する。ポートフォリオとは、分析対象の特許データを階層的に管理する仕組みである。データを分類する視点によって分析対象となる特許データの分類方法は異なる。このようなそれぞれの分類視点を分類軸という。ポートフォリオ管理では、特許が解決しようとしている課題（解決課題）と解決するための手段（解決手段）というような異なる切り口の分類軸を作ることでもできる。

ポートフォリオ分析は、ポートフォリオ管理機能を用いて整理、分類した自社や競合他社の特許を対象に、解決課題と解決手段というような分類軸で、分析を行う機能である。この分析の際には各社の保有する特許を件数だけでなく、特許の重要度（質）を数値化した指標（Patent Competency Index, 以下 PCI<sup>\*1</sup>）値を基に分析する。このように複数の切り口で分析することで、競合他社の研究動向を分析し、重要特許の絞り込みなどが行える。

発明者グループ分析は、ポートフォリオ中の特定領域の特許を基に、その発明者の情報から競合他社の発明者のグループを特定し、グループごとに出願された特許の分類を行う。分類された特許を基に競合他社の発明者グループの研究課題を把握し、早期に他社の研究開発動向を

分析し、その対応策を立案することが可能になる。

経済価値評価は、企業における研究プロジェクトの絞り込み、事業計画の検証、ライセンス技術の価値把握、資産価値評価、知財活動評価などの目的に応じて、個々の特許について経済価値を金額で評価する。StraVision では三つの視点で経済価値を評価することができる。一つ目は、ディスカунテッド・キャッシュ・フロー (DCF) 法<sup>\*2</sup> とリリース・フロム・ロイヤリティ (RFR) 法<sup>\*3</sup> を用い、将来、特許から得られる収益を基準に特許の価値を算定するインカム・アプローチである。二つ目は、ブラックショールズ・オプション・プライシング・モデル (BSOPM)<sup>\*4</sup> を用い、類似特許のライセンス取引データから特許の価値を類推するマーケット・アプローチである。三つ目は特許の取得、開発にかかる費用から特許の価値を算定するコスト・アプローチである。

StraVision はこれらの機能を、日本で初めて一つのサービスとして提供したものであり、ユーザは特許に関する分析や評価をすべて行うことができる。

### 3. 特許データ分析と分析のための特許データ管理

#### 3.1 特許データの分析手法

特許データ分析には大別すると2種類の分析手法がある。一つ目は対象となる特許データ中に現れる特徴的なキーワードに着目して、関連の高いキーワードを近くに配置したマップを表示し、その特定領域の技術動向を分析する手法である。マップには図1のようにキーワードの出現頻度の高い領域を高度に見立てて表現した等高線マップや、高度の代わりに温度に見立てて表現したヒートマップと呼ばれるマップがある。

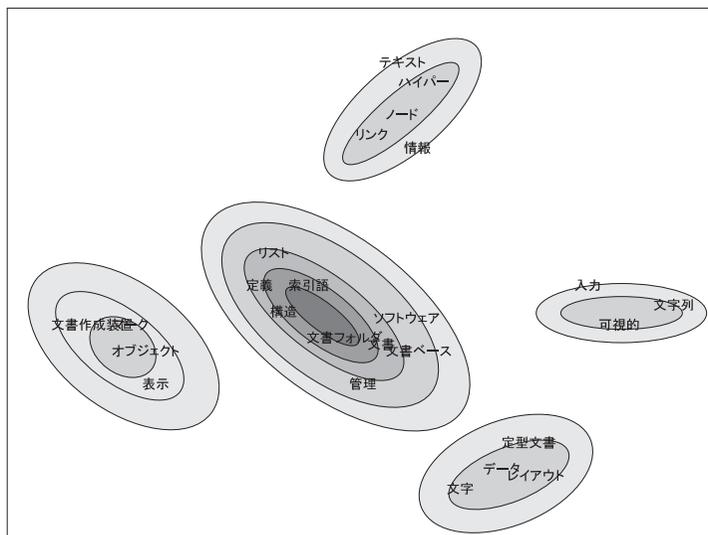


図1 等高線マップの例

二つ目は特許データを様々な基準で分類し、分類した特許の件数や PCI を用いて図2のようなバブルチャートなどに表示して分析する手法である。

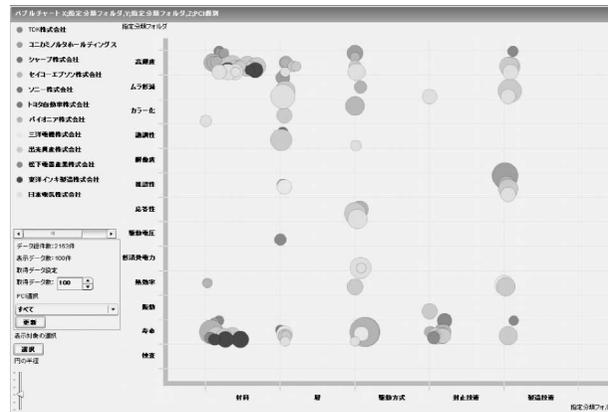


図2 技術分類と企業固有の分類で作成したバブルチャート

StraVision における特許データ分析は，出願人ごとの特徴的なキーワードを一覧して分析することもできるが，後者のように分類された特許の件数や PCI をバブルチャート上で分析する手法が中心となっている．

### 3.2 StraVision における特許データの分析

StraVision は，自社と競合他社の保有特許を対象に，複数の分類軸で整理された特許データを抽出して，特許の件数や PCI で出願人ごとのシェアを比較し，更に細かい技術分類ごとに，自社と競合他社の件数や PCI で定量的に分析していく．

特許は，一般的に広く権利を主張しているものほど重要であり，また他の特許で多く引用されているものほど重要である．このため特許を使った分析では単純に件数だけを比較することはできない．PCI は，特許の重要度（質）を数値化して，客観的にその技術の有効性を表すために導入された指標である．PCI の算出には図 3 のように，特許の権利範囲を表す項目，外からの注目度を表す項目，自社の注目度を表す項目，そして特許権利の状態を表す項目を用いる．特許の権利を表す項目は，特許の権利を主張する請求項の文字数や国際特許分類 (International Patent Code 以下 IPC)<sup>\*5</sup> のクラスレベルの種類である IPC クラス数がある．また外からの注目度には，他の特許にどれほど引用されているかを表す被引用数や実際にその特許の閲覧請求がされた回数である閲覧請求回数などがある．更に，自社の注目度には，他の特許より優先して早期に審査を行うよう求めた早期審査の有無などがある．そして特許権利の状態を表す項目には，登録，公開，審査請求などがある．PCI は，各ユーザがこれらの項目をどれほど重要視するかを表すウェイトを加味して算出される．StraVision は，特許公報，整理標準化データか

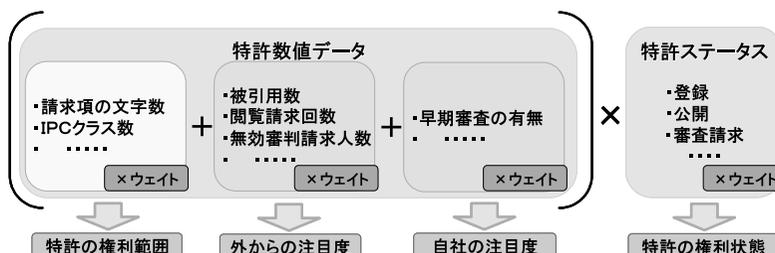


図3 PCI の評価項目

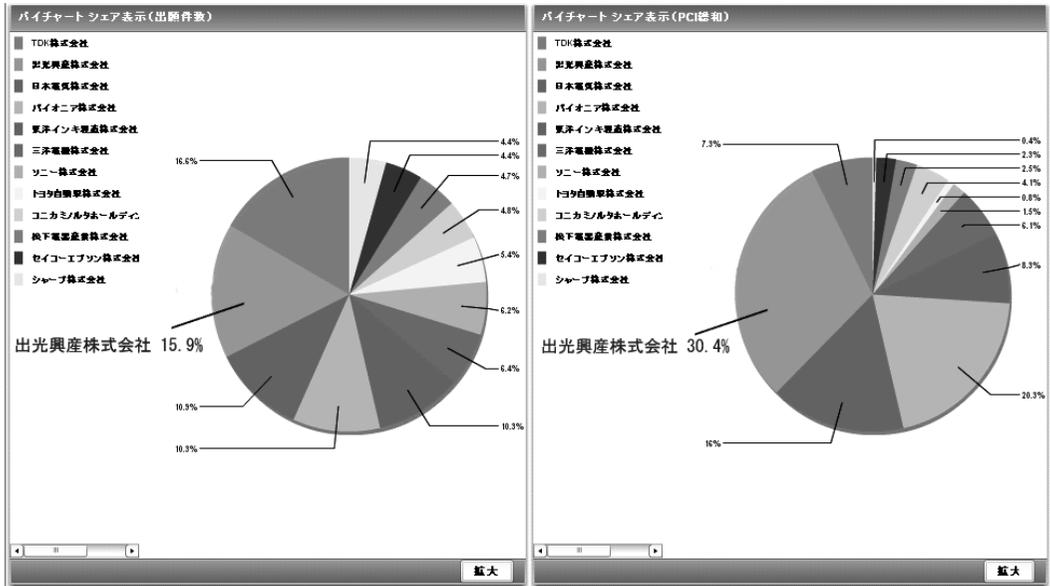


図4 特許件数・PCIによる企業別の相対シェアマップ

ら取れるこれらの有効な数値情報をデータベース上で管理した初めてのシステムである。

分析では、目的にあわせて出願人別や出願人別分類別に特許のPCIを合算した値を用いてマップを表示する。PCIを用いると、単なる件数の比較ではなく、有用な特許がどれほど出願されているかを把握できる。図4のように円グラフの左側2番目の出光興産は左の特許件数によるシェアでは15.9%であるが、右のPCIによるシェアでは30.4%とシェアが拡大していることがわかる。

次に出願人別技術分類別の出願件数とPCIによる比較を行う。シェアマップでは出願人単位で比較を行ったが、次に個別の技術分類で自社と競合他社の特許を比較する。図5のように右側には、ポートフォリオ管理機能を用いて作成した特許データの分類ごとに棒グラフを表示している。棒グラフは、横軸に出願人、縦軸に特許件数やPCIを表示しており、左側ではその中の特定技術分類を拡大表示している。

また、図6のバブルチャートでは解決課題と解決手段別に、横軸に出願件数によるシェアを、縦軸に出願件数の増加率をとり、出願人ごとの該当特許のPCIの合算値をバブルの大きさとして表示している。特定の解決課題と解決手段において、件数でのシェアが少なく増加率の多い出願人は、今後その分野への傾注が予想され、逆に件数でのシェアは多いものの増加率が低い出願人は、その分野からの撤退が予想される。このような結果から各企業の傾注している技術について分析することができる。

その他にもStraVisionには出願日、キーワード、IPCに着目した分析用グラフが用意されている。

### 3.3 ポートフォリオ管理機能による特許データの管理

StraVisionは自社、競合他社の特許データを分類し、その結果をバブルチャートなどに表示して分析する。このため特許データを、階層的に分類するポートフォリオという考え方を導入

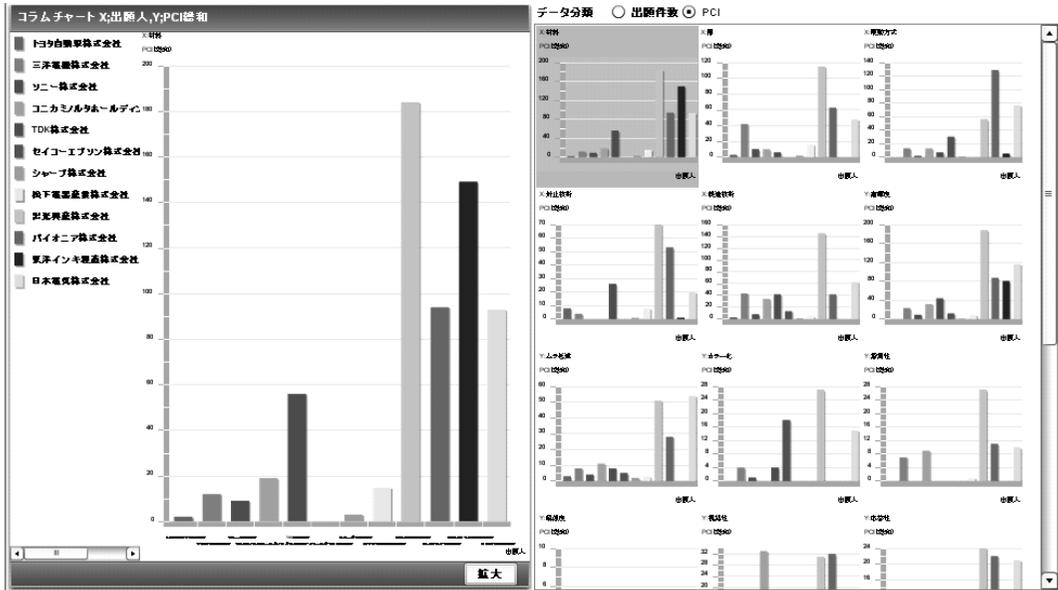


図5 特許件数, PCI による技術分類ごとの企業比較 (コラムチャート)

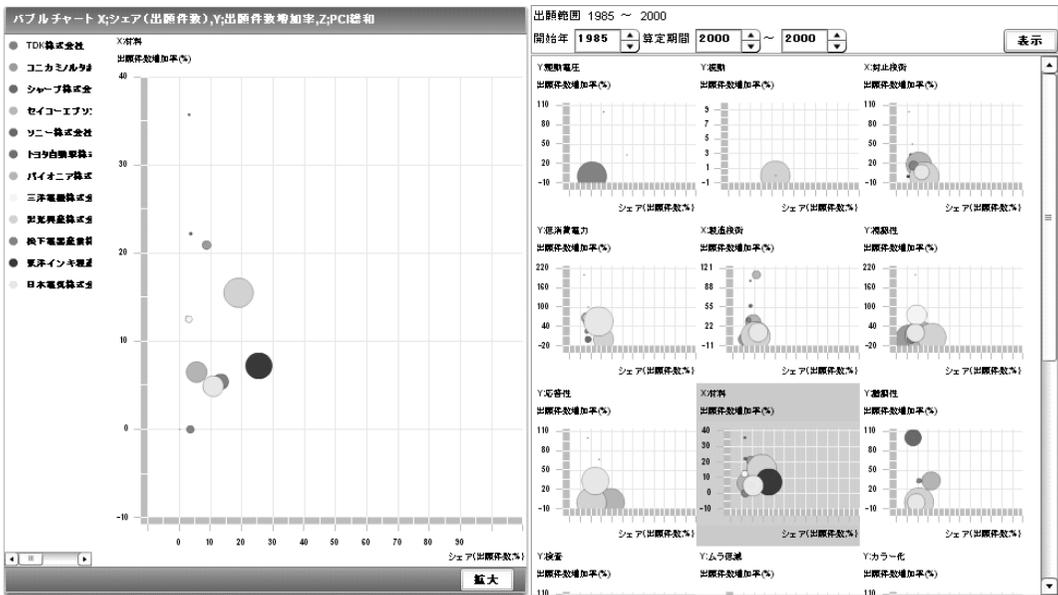


図6 解決課題・解決手段ごとの特許件数と伸び率に着目したマップ

している．ポートフォリオとは、分析対象の特許データを Windows の Explorer のように階層的なフォルダを用いて管理する仕組みである．ユーザはポートフォリオ上で企業独自の視点で対象データを分類する．ポートフォリオには図7のように一つのゴミ箱と階層的な複数のフォルダが存在する．フォルダの下に更にフォルダがある場合、そのうち一つは未分類フォルダというどこにも割り当てられなかった特許データを格納するフォルダがある．ゴミ箱や未分類フォルダ以外のフォルダは、それぞれが一つの特許分類に対応している．

ポートフォリオでは, StraVision データベースから検索条件を指定して対象特許を定義する．次に自社の分類基準でフォルダを作成して、自社の特許データを振り分ける．更に作成したフ

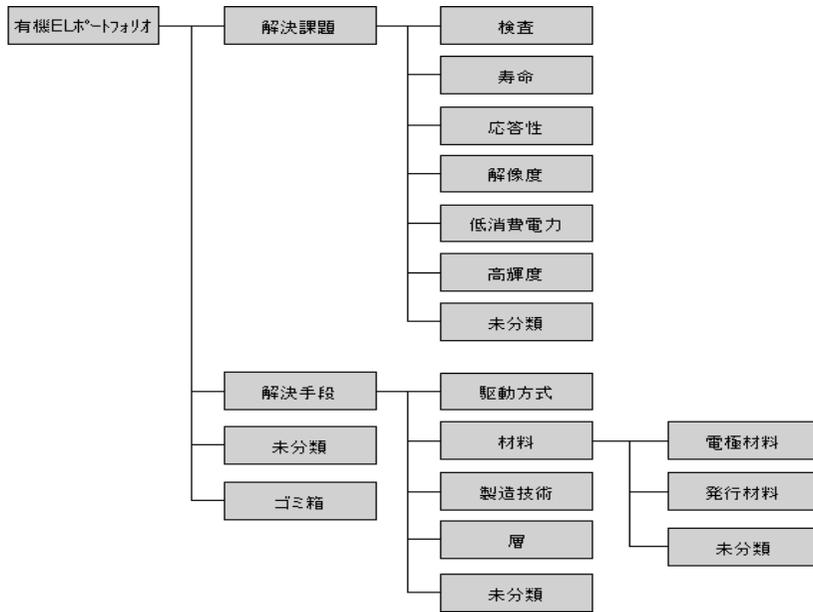


図7 有機ELポートフォリオ例

フォルダに競合他社の特許を振り分けることで、自社の分類基準で自社と競合他社の特許を分類でき、その結果、自社の基準で件数やPCIの分析が行える。

#### 4. ポートフォリオ管理におけるテキストマイニングの適用

##### 4.1 ポートフォリオ管理に必要とされるテキストマイニングの機能

ポートフォリオ管理では自社と競合他社の特許を整理、分類する。しかし対象の特許データは数千から数万に及ぶため、人手によるポートフォリオの作成は手間を要する。更に自社特許をポートフォリオに振り分けした後、他社特許を同様の基準で振り分ける作業も手間を要する。また、対象特許データを検索するときも、条件で指定するキーワードによっては、本来対象とすべき特許が抽出されないことがある。

これらの課題を解決するために、StraVisionにMiningPro 21 文書マイニング・システム<sup>\*6</sup>を適用した。ポートフォリオの対象特許を検索する際のキーワード指定を支援するため、関連するキーワードを提示する機能、特許データの分類基準を持たないユーザでも特許データを分類できる自動分類機能、他社特許や今後出願される特許を自社特許と同じ基準でフォルダに振り分ける自動仕分の機能を実現した。

##### 4.2 関連キーワードの抽出

StraVisionのデータベースからポートフォリオの対象特許データを抽出するには、検索条件を指定する。検索条件にはIPCなどのコード類、出願日などの期間以外に、発明の名称、解決課題、解決手段のキーワードを指定することもできる。そこでStraVisionでは漏れなく特許データを抽出するために、ユーザが指定したキーワードと関連性の高いキーワードを表示する機能を提供している。

関連キーワードの抽出は、全特許の1%に当たる特許データを用いて行う。まず、対象デー



### 4.3 特許データの自動分類

ポートフォリオ管理では自社の基準に従って特許データを階層的に整理、分類する。しかし従来から自社特許の棚卸しを行って行けば特許分類の基準が存在するが、これから特許の管理と活用を行う企業には自社の基準は存在しないことが多い。このような状態で数万の特許データを分類することは、かなりの手間を要する。この手間を軽減するために StraVision では MiningPro 21 文書マイニング・システムの分類の機能を用いた特許データの自動分類機能を備えている。

特許データの自動分類は、前述の潜在的意味インデクシングをキーワードだけでなく、IPC や発明者にも適用できるように拡張して作成された空間上でクラスター分析を行う方法で実現している。自動分類のために潜在的意味インデクシングに用いる項目を分類基準と呼ぶ。分類基準には、日本特許では IPC、発明者以外に解決課題や解決手段、請求項、件名中のキーワードを、米国特許では IPC、発明者だけでなく要約や請求項、件名中のキーワードが指定できる。特許データの要約、解決課題、解決手段、請求項からは複数のキーワードが抽出されるだけでなく、複数の IPC や複数の発明者も抽出される。IPC や発明者を前述の潜在的意味インデクシングの対象に加えることで、同時に利用される関連の高いキーワードだけでなく、同時に出現する IPC が空間上で近傍に配置される。

次にこの空間に特許データを配置する。各特許データは、そこに出現するキーワード、IPC、発明者を空間上で特定し、その重心の位置に配置する。その結果、同じようなキーワードや IPC が指定されている特許データは空間上で近傍に配置される。このようにして配置された特許データを  $k$  means 法と呼ばれるクラスター分析の手法を用いてユーザが指定した数のグループに分類する。分類した特許データのグループに、そのグループで頻度が多いキーワードを五つ取り出して、各グループの仮のタイトルとする。でき上がったグループは、図 10 のように一つのフォルダに対応付けてポートフォリオ上に配置する。

The screenshot shows the StraVision software interface. On the left is a sidebar with a tree view of categories under 'ポートフォリオ' (Portfolio), including '有機ELデモ(自動分類後)' (Organic EL Demo (after automatic classification)), '材料' (Materials), '層' (Layers), '駆動方式' (Driving methods), '封止技術' (Sealing technology), '製造技術' (Manufacturing technology), '未分類' (Unclassified), and 'ごみ箱' (Trash). The main area displays a search result table for '有機ELデモ(自動分類後) / 材料 / 基板:透明電極;誘電;フィルム'. The table has columns for '選択' (Select), '出願番号' (Application No.), '出願日' (Application Date), and '発明の名称' (Invention Name). The right side shows a detailed view of a selected patent, including its abstract (要約).

| 選択                       | 出願番号          | 出願日        | 発明の名称                        |
|--------------------------|---------------|------------|------------------------------|
| <input type="checkbox"/> | 特許2002-021419 | 2001-07-30 | 有機EL表示装置                     |
| <input type="checkbox"/> | 特許2000-047898 | 2000-02-24 | 有機EL発光素子                     |
| <input type="checkbox"/> | 特許2002-063570 | 2002-03-08 | 有機エレクトロルミネッセンス素子の製造方法        |
| <input type="checkbox"/> | 特許平10-350335  | 1998-12-09 | 有機エレクトロルミネッセンス表示装置及び有機EL発光素子 |
| <input type="checkbox"/> | 特許2000-129129 | 2000-04-28 | 有機エレクトロルミネッセンス素子、それを         |
| <input type="checkbox"/> | 特許2000-200362 | 2000-06-30 | 有機エレクトロルミネッセンス素子             |
| <input type="checkbox"/> | 特許2000-374994 | 2000-12-08 | 色変換フィルタ基板、該色変換フィルタ基          |
| <input type="checkbox"/> | 特許2000-108163 | 2000-04-10 | ディスプレイおよびその製造方法              |
| <input type="checkbox"/> | 特許平11-301284  | 1999-10-22 | EL表示装置およびEL表示装置の製造方法         |
| <input type="checkbox"/> | 特許2001-284443 | 2001-09-19 | 高電流用異方導電性フィルム及び有機EL          |
| <input type="checkbox"/> | 特許2000-226468 | 2000-07-27 | 感敏斜線性樹脂組成物および有機EL素子          |
| <input type="checkbox"/> | 特許2001-003410 | 2001-01-11 | 有機ELディスプレイ                   |
| <input type="checkbox"/> | 特許2002-146328 | 2002-05-21 | ITO膜付基板、およびその製造方法、な          |
| <input type="checkbox"/> | 特許平11-302147  | 1999-10-25 | 有機エレクトロルミネッセンス素子             |
| <input type="checkbox"/> | 特許平06-064292  | 1994-04-01 | 薄膜エレクトロルミネッセンス素子             |
| <input type="checkbox"/> | 特許2000-200353 | 2000-06-30 | 有機エレクトロルミネッセンス素子及び有機         |
| <input type="checkbox"/> | 特許2002-051358 | 2002-02-27 | 有機発光ダイオードデバイス                |
| <input type="checkbox"/> | 特許2002-062644 | 2002-03-07 | 有機EL表示装置およびその製造方法            |
| <input type="checkbox"/> | 平04-239145    | 1992-09-08 | 透明導電性膜層体の製造方法                |
| <input type="checkbox"/> | 特許平08-258996  | 1996-09-30 | エレクトロルミネッセンス表示装置             |
| <input type="checkbox"/> | 特許2001-146910 | 2001-05-16 | 有機電界発光素子及びその製造装置             |

図 10 有機 EL ディスプレイの材料に関する特許を自動分類した例

このように StraVision における特許データの自動分類では、潜在的意味インデクシングを行うことで、同義語を指定しなくても、類似性の高い特許を同じグループに分類できる。更に、利用する企業ごとのユーザ単語や同義語の辞書を使用することができるため、業種業態ごとに異なる専門用語を考慮した自動分類を行うことができる。

#### 4.4 特許データの自動仕分

ポートフォリオ管理で自社の特許を分類した後、他社の特許を同じ基準で振り分ける作業や、新たに出願された特許を同様の基準で振り分ける作業は、自社の特許データを仕分ける時と同様に手間を要する。この作業を支援するため、StraVision ではメールの自動仕分のような検索による自動仕分とテキストマイニングによる自動仕分の機能を提供している。

自動仕分のために、最初にポートフォリオの各フォルダに検索条件式による仕分ルールかマイニングによる仕分ルールかを設定する。自動仕分はポートフォリオの第一階層にある未分類フォルダの特許データを対象に、各フォルダに設定された検索条件式やマイニングを用いて作られた仕分ルールを評価して、特許データを仕分けるべきかどうかを判断する。ポートフォリオのフォルダは階層的になっており、最初に第一階層の各フォルダに特許データが仕分けられ、更にその下にフォルダがあれば、そのフォルダの仕分ルールで評価して仕分ける。

検索条件式による自動仕分では、ポートフォリオの対象特許を指定した時と同様に IPC などのコードや出願日などの期間以外に、発明の名称、解決課題、解決手段などのキーワードを検索条件にして仕分ルールを設定できる。自動仕分は条件に合致した特許データを対象のフォルダに仕分ける。

マイニングによる自動仕分では、仕分ける基準となる IPC、解決課題や解決手段、請求項のキーワードを一覧からユーザが選択する。ここで選択する IPC やキーワードを判別基準と呼ぶ。判別基準の選択は、それが出現する特許数と F 値<sup>\*\*</sup>を参考にして行う。判別基準の値を基に各フォルダに属するかどうかの 0/1 の二値を重回帰モデルで推定して、各判別基準に重み付けを行う仕分ルールの作成を行う。作成されたルールの評価は、まず、各特許データに仕分ルールを適用してフォルダへ仕分けられる確率を 0~100 で算出する。その結果を 0 10,10 20, ... , 80 90,90 100 というように 10 等分してグループを作成し、特許をその適用結果に従って該当するグループに割り当てる。このグループをランクと呼ぶ。作成したランクは確率の高いものから順に評価していく。各ランクの特許でそのフォルダに割り当てられるとユーザが判断した特許の割合（正答率）が要求する精度の範囲であれば、そのランクの確率値より高い特許はフォルダに割り当てるというルールとなる。このようにして求めたランクの確率値を閾値と呼ぶ。

図 11 の例では、60 70,70 80,80 90,90 100 の各ランクは正答率がすべて 80% を超えている。このため特許データに判別ルールを適用した結果、60 以上の確率である特許はフォルダに割り当てると判断しても誤る確率は 20% 以下である。また、00 10,10 20,20 30 の各ランクは正答率が何れも 15% 以下である。このため適用結果が 30 未満の確率である特許はフォルダに割り当てないと判断しても誤る確率は 15% 以下である。問題になるのは 30 40,40 50,50 60 の各ランクである。30 60 の確率の特許データは、どちらに判断しても精度は確保されない。これらのランクをグレーゾーンと呼ぶ。StraVision では、このグレーゾーンの特許データだけを対象に自動仕分ルールを再度作成し、結果を評価することで、グレーゾーン用のルールを別途作

評価順序 ↑

| 確率     | 正答率 | 件数 | 備考     |
|--------|-----|----|--------|
| 00-10  | 5   | 30 | ×      |
| 10-20  | 10  | 30 | ×      |
| 20-30  | 15  | 50 | ×      |
| 30-40  | 30  | 60 | グレーゾーン |
| 40-50  | 50  | 80 | グレーゾーン |
| 50-60  | 60  | 60 | グレーゾーン |
| 60-70  | 80  | 50 | ○      |
| 70-80  | 85  | 30 | ○      |
| 80-90  | 90  | 30 | ○      |
| 90-100 | 95  | 20 | ○      |

図 11 判別ルールの評価例

成することができる。

特許データの自動仕分を行う時には、対象データに最初のルールを適用し、その確率が閾値以上であればそのフォルダに割当て、グレーゾーンであればグレーゾーン用のルールを適用して割り当てるかどうかを判断し、それ以外であれば割り当てないようにして、単一のフォルダへの自動仕分を行う。これを全フォルダに対して実施することで、ポートフォリオの自動仕分を実現している。

### 5. StraVision の評価

StraVision を利用したユーザからは、StraVision の効果についてのレポートが上がってきている。そこでは StraVision で行われた新しい試みが評価されている。

まず PCI を用いて特許を分析できる点が評価されている。日本の特許制度では、権利化するつもりが無くても他社が特許を取得することを阻む目的で防衛的特許出願をすることができる。このような特許は審査請求されることもなく、出願から 3 年が経過すると出願そのものが取り下げられたものとしてみなされる。1985 年以降の特許件数は日本特許が約 760 万件、米国特許が約 240 万件であるが、実際に登録されて権利化されている日本の特許は約 200 万件である。このように各社の特許を単純に件数だけで比較すると、注力している技術が見極められなくなる。PCI を利用すれば、特許の重要性に関連のある定量的な項目を用いることができる。また他の特許検索サービスと異なり、これらの項目が StraVision のデータベースには格納されており、ユーザは各項目間のウェイトを指定するだけで、PCI が利用できる。

更にこれらのデータが特許庁の提供する情報を常に反映して更新されていることも評価されている。経過情報である整理標準化データを利用せず、特許公報のみを用いたデータでは、出願時に審査請求が行われていなければ、その後に行われる審査請求などの審査、審判、登録の情報は反映されない。StraVision では、整理標準化データをデータベースに取り込んでいるため、分析で用いるデータも最新に保たれている。このため StraVision では過去に行った分析手順を呼び出すだけで、最新の特許の状態を反映した分析が行える。

また、特許データの整理、分類を行うポートフォリオ管理にテキストマイニングの技術を用いたことも評価されている。従来の特許データの整理・分類は、特定条件に合致するデータをメールの仕分ルールのように分類してフォルダに仕分していた。この方式では特許データの約 50% がユーザの期待どおり仕分けされたにすぎなかった。StraVision では、従来からの検索条件式による仕分けに加えて、テキストマイニング・データマイニングを用いたことで、特許

データの約 70～80% が期待どおりに仕分できるようになった。

このように StraVision は特許データベースの構造、特許データの整理、分類、そして分析と経済価値評価まで、すべてが意思決定支援のために構築されており、従来の特許検索中心のシステムや、公報データの分析システムとは大きく異なっている点が評価されている。

## 6. お わ り に

2003年3月には知的財産基本法が施行され、内閣に知的財産戦略本部が設置された。日本は今、国を挙げて知的財産立国を目指し、知的財産の創造、保護及び活用に関する施策を集中的かつ計画的に推進しようとしている。

このような環境の変化の中で、今回開発した StraVision は、特許データの提供、整理、分類、分析、価値評価までのすべてを日本ではじめて一つのソリューションとして提供したものである。データを整理、分類するテキストマイニング、データマイニングの技術をうまく活用することで、ポートフォリオの作成を支援し、PCI などの特許データに付随する数値項目を利用して特許データの分析が可能となった。特許などの大量のデータを整理、分類することは非常に手間がかかるが、データマイニングの技術を用いることでこの手間を軽減し、出願される新しい特許データをポートフォリオの適正なフォルダに配置することができる。このため、従来は分析の度に行っていたデータ抽出、整理、分類という事前準備作業が減り、本質的な特許データの分析をいつでも最新のデータを使って行うことが可能となった。

多くの企業が StraVision を用いて、特許データの分析を行い、自社の技術戦略を的確に立案し、企業の競争力の源泉とすることを望みたい。

- 
- \* 1 特許の公報・経過情報における特許の量・質を表す客観的データを抽出し、産業・技術分野別の特性を加味した上で特許の総合的な競争力を表すようにウェイト付けを行って算出される指標。質・量を表すデータには他の特許にどれほど引用されているかを表す被引用数やその特許の権利を主張する請求項の数などなどがある。詳細は参考文献<sup>2)</sup>カバーストーリを参照。
  - \* 2 特許の寄与度を将来のフリーキャッシュフローに乗じることにより特許の価値を算出し、現在価値を算定する方法。詳細は参考文献<sup>1)</sup>6 6 を参照。
  - \* 3 特許を利用している製品売上の予測値とライセンス料率（ロイヤリティレート）により算出し、現在価値を算定する方法。参考文献<sup>1)</sup>6 5 を参照。
  - \* 4 オプション価格算出モデルの一つで、市場で比較的容易に観察、推測できる五つのパラメータ（変数）を用いる。StraVision ではこの手法を応用した TRUU が用いられている。詳細は参考文献<sup>1)</sup>6 22.6 24 を参照。
  - \* 5 国際的に統一された特許文献のための分類体系で、全ての技術に分類が付与されている。国際特許分類はセクション（1桁）、クラス（2桁）、サブクラス（1桁）、メイングループ（3桁）、セパレータ（1桁）、サブグループ（5桁）からなる。例えば H 04 M 1/02（水工一般についての設備または装置あるいは方法）では、セクション（H）までは電気を、クラス（H 04）まででは電気通信技術を、サブクラス（H 04 M）まででは電話通信を、メイングループ（H 04 M 1）までではサブステーション装置を、サブグループ（H 04 M 1/02）では電話機の構造的態様を表し、多くの桁を使えば使うほどより詳細な内容を表している。
  - \* 6 日本ユニシスが開発したテキストマイニングソリューション。テキストデータから抽出したキーワードを基にテキストデータをグルーピングすることができる。テキストデータのグルーピングには、テキストデータ中のキーワードから、キーワード間の関連度とテキストデータ間の関連度を算出して関連度の高いデータを一つのグループにする分類（クラスタリング）と、事前にユーザが判断したグルーピング結果と個々のテキストデータ中のキーワードから、

キーワードに自動的にウェイト付けを行い、ユーザの基準でテキストデータをグルーピングする判別(クラシファイ)がある。また、キーワードやテキストデータの関連度を算出する分類の機能を利用することで概念検索を実現することもできる。テキストデータに対して精度の高いグルーピングを行うために、同義語辞書や正規表現を用いた文字列パターンマッチングを行うワードパターンに対応し、潜在的意味インデクシング<sup>8)</sup>や判別結果の信頼性評価などの技術が適用している。詳細は参考文献<sup>9)</sup>を参照。

- \* 7 キーワードが特許データに出現する頻度が高いほど大きくなり、また全体の特許に出現するほど小さくなる値で、キーワードの量と局所性を掛け合わせた値。一般的には次のように定義されている。

$$tfidf(T) = f(T) \times \log\left(\frac{N_0}{N(T)}\right) \quad (T): \text{キーワード } T \text{ の対象特許中の頻度}$$

$N_0$  : 対象特許数  
 $N(T)$ : キーワード T を含む特許数

StraVision では参考文献<sup>4)</sup>にある文中に出現する単語数を考慮に入れた TF・IDF を採用している。

- \* 8 多変量データの持つ情報を少数個の総合指標に要約する手法。詳細は参考文献<sup>7)</sup>の第 2 章を参照。
- \* 9 判別結果および全ての判別基準を標準化(平均を 0, 分散を 1 に変換)したときの重回帰式を求めたときの、偏回帰係数(標準偏回帰係数)と標準誤差を用いて F 値 = (偏回帰係数/標準誤差)<sup>2</sup>のように求めたものである。

**参考文献** [ 1 ] 松本 公明, 「知財評価の基本と仕組みがよくわかる本」, 秀和システム, 2004 年 10 月。

[ 2 ] フジサンケイビジネスアイ, 「知財情報 & 戦略システムカバーストーリー: 知的資産情報を戦略的に活用せよ」, フジサンケイビジネスアイ, 2005 年 10 月。

[ 3 ] 蓑谷千鳳彦, 「ブラック・ショールズモデル」, 東洋経済新報社 2000 年 3 月。

[ 4 ] Akiko Aizawa, “ The Feature Quantity: An Information Theoretic Perspective of Tfidf like Measures ” proc. of ACM SIGIR 2000, pp.104 111 ( 2000 ) 。

[ 5 ] Richard Razgaitis ( 原著 ), 菊池純一 ( 翻訳 ), 石井康之 ( 翻訳 ), IPTT グループ ( 翻訳 ), 「アーリーステージ知財の価値評価と価格設定」, 中央経済社, 2004 年 3 月。

[ 6 ] 寺本義也, 山本尚利, 山本大輔, 「最新 技術評価法」, 日経 BP 社, 2003 年 7 月。

[ 7 ] 田中豊, 脇本和昌, 「多変量解析法」, 現代数学社, 1983 年 5 月。

[ 8 ] 金明哲, 村上征勝, 永田昌明, 大津起夫, 山西健司, 「統計科学のフロンティア 10 言語と心理の統計」, 岩波書店, 2003 年 3 月。

[ 9 ] 林田英雄, 脇森浩志, 「技報 84 号一般号テキストマイニング技術とその応用」, 日本ユニシス, 2005 年 2 月。

**執筆者紹介** 林 田 英 雄 ( Hideo Hayashida )

1987 年神戸商科大学管理科学科卒業。同年日本ユニシス(株)入社。意思決定支援システム, 情報検索システムの開発を行い, 現在日本ユニシス・ソリューション(株)データサイエンスビジネスで MiningPro 21 文書マイニング・システムの開発, 適用を担当。OR 学会会員。