

日本 MOT 学会による査読論文 (2016-1)

人の意味理解能力とテキストマイニングの大量データ処理技術 とを融合した特許文献のハイブリッドクラスタリング手法

Hybrid clustering method of patent literatures that was obtained by using the human understanding capabilities and the text mining

上野 覚、山田 充
Satoru Ueno, Mitsuru Yamada

要 旨

大量の特許文献をクラスタリングする手法として、人の意味理解能力とテキストマイニングの大量データ処理技術との融合によるハイブリッドクラスタリング手法を提案した。これはテキストマイニングを活用して特許文献群中に基準文献を設け、この文献の技術内容を目視により把握し、これらに基づき設定したクラスターの技術内容を用いて特許文献群のクラスタリングを行う手法である。この手法をデジタルカメラの特許文献群に適用した結果、類似性の高い特許文献群では有効であることが分かった。

ABSTRACT

As clustering method of a large amount of patent literatures, we proposed the hybrid clustering method that was obtained by using the human understanding capabilities and the text mining. A reference literature was provided in the patent literatures using a text mining. Technical content of the cluster was estimated from the reference literatures and keywords extracted from the patent literature. Based on the technical content of the cluster, a clustering of patent literatures was performed. We examined this technique to the patent literatures of the digital camera. It was found that this technique was effective to the patent literatures having a high similarity.

キーワード：テキストマイニング、特許情報、内容推定、技術動向分析、ハイブリッドクラスタリング手法

投稿区分：研究論文

1. はじめに

従来製品やサービスの陳腐化による収益低下が問題となっており、各企業には持続的な発展に向け新たな施策が求められている。名和は 1990 年～2010 年までの 20 年間、上場企業の売上高、営業利益及び、時価総額の伸び率から日本の勝ち組企業 100 社を選定

し、成功要因を分析している [1]。その結果、市場開拓力、事業モデル構築力、経営変革力及びオペレーション力を成功要因に挙げている。このうち市場開拓力、事業モデル構築力を成長のツインエンジンとしている [1] ことから、今後の成長戦略として、各企業は自社のみならず自社を取り巻く周辺の技術動向及び市場動

向を的確に把握し、それに基づいた技術・経営戦略の構築が重要になっている。

技術動向・市場動向の把握について、特許情報は年間 30 万件以上出願され、客観性が高く、「技術情報」、「権利情報」及び「経営情報」として貴重な情報源となっており、特許情報を活用した技術・市場動向の把握が望まれる。しかし、特許情報から技術動向を把握するには、大量の特許文献を人が読み込みまとめているため大変な作業となっている。人の目視を行うことなく、特許分類を活用して作成される特許マップもあるが、これは発明を様々な観点から捉え統計処理を行っているので、発明の主題を把握し纏めていくことは困難である。大量の特許文献を簡便な方法で分析、要約して技術動向を作成する方法が開発されれば、研究開発効率の向上に寄与すると共に企業戦略の構築・意志決定にも資するものと期待される。

近年、ICT (Information and Communication Technology) の進展に伴いテキスト自動要約やテキストマイニングなど人の処理能力を超えた膨大な文献の分析を可能とするテキスト処理技術が進展している。前者のテキスト自動要約技術については、奥村、難波らは、テキスト自動要約技術を概観しており、その中で複数テキストを対象にした自動要約技術について報告している [2]。また、渡邊、太田、片山、石川らは、文章から意味情報を抽出しこれを類似自然言語に融合することで分野に依存しない複数文書要約方法を提案している [3]。このように複数文書を要約する方法は種々検討されているが、人の意味理解能力を用いてクラスタリングや曖昧性の解消を行いながら技術動向を作成する方法はあまり報告されていない。

また、後者のテキストマイニングは、大量の文献に記述されている内容を数値化し統計的に解析することで、その相関や傾向を分析すると共に文書の類似性に基づいてクラスタリングも行われている。特許文献分析への応用では、豊田、菰田らが、膨大な特許情報を、一つ一つ目視する方法では技術動向が把握できないため大規模なデータから知見を抽出する方法として、テキストマイニングを用いている [4]。また、那須川は、テキストマイニングによる特許情報の分析について、「人手のみのレベルから一歩も二歩も進んだ分析に踏み込み、戦略的な分析に注力できる可能性が高くなる。」 [5] とテキストマイニングの技術戦略分析への有効性について指摘しており、今後具体的な事例への展開が望まれる。しかし、テキストマイニングは、解析過程で文章を形態素 (キーワード) に切り分け、元の文書の意味情報が喪失しているため、さまざまな観点からの内容解釈が容

認されている探索的データ解析手法で、曖昧性解消が課題となっている。また、テキストマイニングによる文書クラスタリングでは、抽出されたキーワードを用いて「文書×抽出語」表が作成され様々なクラスタリングが行われているものの、クラスタリング精度が問題となっている。より精度良くクラスタリングを行うため、クラスターの技術内容に基づいたクラスタリング手法の開発が望まれている。

そこで本研究では、テキストマイニングに係るこれら 2 つの問題点を解決し、大量の特許文献をクラスタリングして技術動向を作成するための方法論として、人の意味理解能力とテキストマイニングの大量データ処理技術とを融合する手法 (ハイブリッドクラスタリング手法とする。) を提案すると共に当該手法の有効性について実証を試みる。具体的には、特許文献のテキストマイニングにより得られる類似度を活用して特許文献群中に基準文献を設け、この文献の技術内容を目視により把握し、これを基に他の特許文献群のキーワード等も併用することでクラスターの技術内容を設定し、これに基づいて特許文献群のクラスタリングを行う手法である。このように人の目視により把握した技術内容に基づいて特許文献群のクラスタリングを行うことで、より精度良いクラスタリングが行えるものと期待される。

2. ハイブリッドクラスタリング手法によるクラスタリング及び技術内容記述

2.1 特許データ

本研究で検討する特許データは一例として、日本が技術的にも事業的にも優位性を持つ「デジタルカメラ」を取り上げた。デジタルカメラの中で、近年、ミラーレスカメラの分野が成長している。ミラーレスカメラは、デジタル一眼レフより小型かつ安価であるものの、画像の質が高い特徴がある。ミラーレスタイプの販売台数シェアは、オリンパス 29.1%、ソニー 26.4%、パナソニック 14.2% であり (集計期間: 2013 年 1 月 1 日 ~ 12 月 23 日)、本研究ではシェアの最も大きいオリンパス(株) (以下、O 社) を選定した。

検索結果の特許データは、パテント・インテグレーション(株)の ASP サービスから入手した [6]。検索は、特許庁が 2008 年度にデジタルカメラ装置の出願動向分析に用いた方法と同様に、IPC (国際特許: H04N 5/222, H04N 5/225, H04N 5/228, H04N 5/232, H04N 5/235, H04N 5/238, H04N 5/243, H04N 5/247, H04N 5/253, H04N 5/257) の範囲をキーワード等 (デジタル+電子+スチル、出願年及び企業名 (オリンパスイメージング(株)を含む)) で絞った [7]。出願年は、2004 年 10 月以降とし、登録された特許

発明を抽出した。検索結果について、パテント・インテグレーション(有)の ASP サービスを利用して個々の特許データを入手した [6]。

2.2 特許文献の類似度算出

テキストマイニングは、大量の特許文献を分析することで個々の特許文献の分析では得られなかった有用な知見を得ることができるが、その一つに、ベクトル空間モデルによる文献の類似度がある。ベクトル空間モデルは、文献から抽出されたキーワードの重要度に基づいて文献を多次元空間上のベクトルとして表現し、ベクトルの向きを比較することにより文献の類似度を調べる方法である。本研究では、2.1 で入手した特許データについて、テキストマイニング用ソフトウェア termmi を用いて類似度を算出した。termmi は、東京大学と横浜国立大学により共同で開発されたフリーウェアで、キーワードを重要度付きで抽出し、これを用いてベクトル空間モデルにより文献の類似度を算出することができる。また termmi は、文献群全体を基準に個々の文献の似ている度合いを数値化した値を算出することができ、本研究ではこれを用いて特許文献群中の各文献の類似度を算出した。算出された類似度により、分析対象文献は以下のように推測される [8]。

- ・類似度が高い文献は、文献群中では一般的なトピックスを扱っている。
- ・類似度が低い文献は、文献群中ではユニークなトピックスを扱っている。

2.3 技術内容記述に係る要因について人とテキストマイニングの比較

特許文献のクラスタリング及び技術内容記述に係る要因について、人の目視による場合とテキストマイ

ニングとの比較を行ったが、その結果を表 1 に示す。人の目視による場合、特許文献の内容理解は優れており精度良く分析を行うことができるものの、膨大な量の特許文献の処理は苦手としている。一方、テキストマイニングの場合、膨大な文献処理は可能であるが、特許文献の内容把握は現状では不可能である。単語の集まりで、ある程度技術内容は把握できるものの、テキストマイニングは様々な観点からの解釈が容認されていることからの確信や信頼性の問題がある。ここで表 1 の色づけを行った部分に示しているように、人が得意とする機能(意味理解能力)と、テキストマイニングが得意とする機能(大量データ処理技術)を組み合わせることで、大量の特許文献群の技術内容を把握し、クラスタリングを行い、技術内容を記述することも可能と期待され、この手法をハイブリッドクラスタリング手法とする。

2.4 ハイブリッドクラスタリング手法のフローチャート

図 1 に本研究で提案するハイブリッドクラスタリング手法のフローチャートを示す。

Step-1: 2.1 で検索により得られた特許文献群について、パテント・インテグレーション(有)の ASP サービスを利用して予め機械的にクラスタリング(予備クラスタリングとする)を行った。

Step-2: Step-1 で得られた予備クラスターについて、クラスター内の全ての特許文献についてテキストマイニングを行い、キーワードを抽出する。このクラスター内の文献群全体に対する各文献の類似度を算出し、類似度の高い順に番号(1~n)を付ける。類似度は、このクラスター中の全文献に対して個々の文献の類似度を算出しているため、類似度の高い文献 1,2 はこのクラスター中で最も一般的なトピックスを扱っている特許文献と考えられ、これを基準

表 1 特許文献群のクラスタリング及び技術内容記述における「人」と「テキストマイニング」の比較

	概要	特許文献群のクラスタリング及び技術内容記述に係る要素		
		大量文書の抽出・クラスタリング	大量文書の類似度算出	個々の文献の技術内容記述
人の目視	人が文献を読み込み技術内容を分析	×	×	○
テキストマイニング	文章を何らかの単位(形態素等)に分解し、記述されている内容の相関や傾向を図表化	○	○	×※

※単語の集まりである程度技術内容は把握できるものの、様々な観点からの解釈が容認されている。→ 人との協働により、特許文献群の技術内容を精度良く把握することが可能となる。

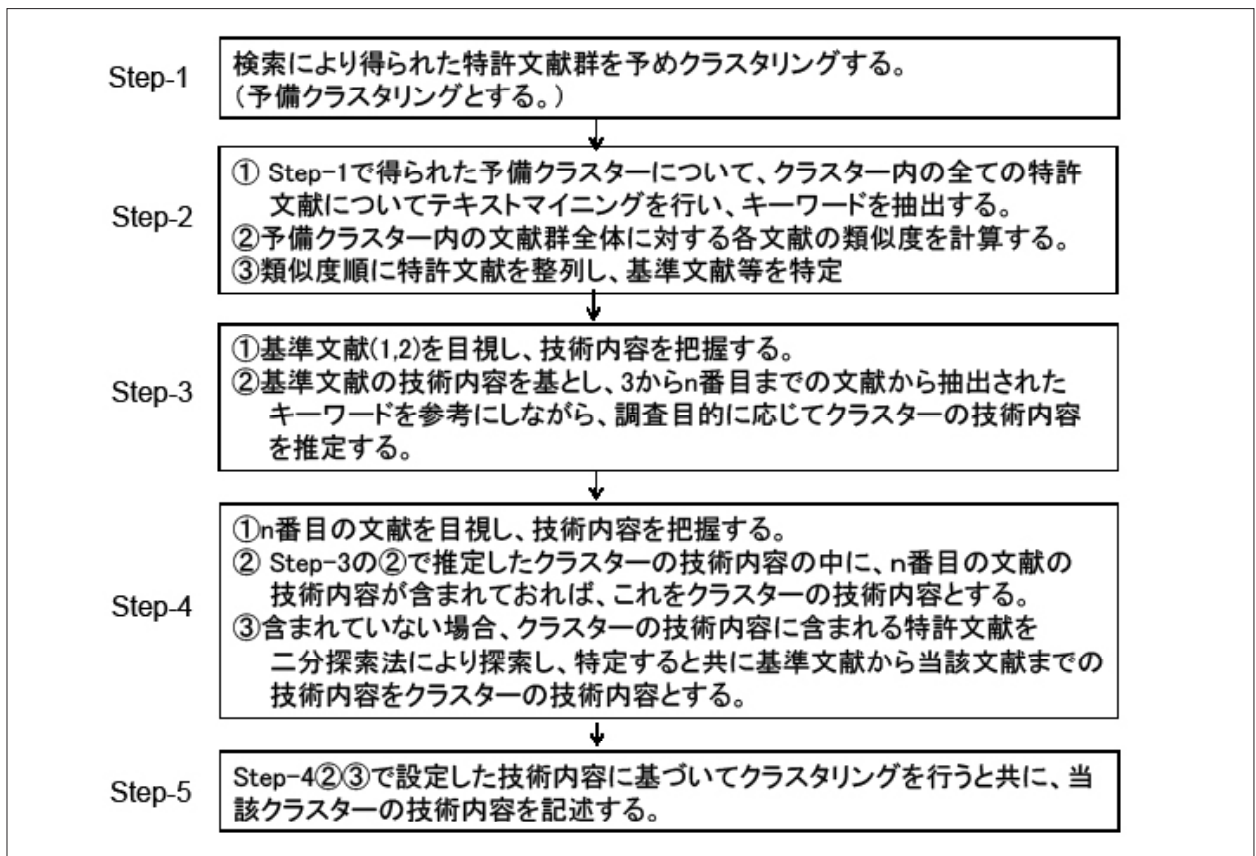


図1 ハイブリッドクラスタリング手法のフローチャート

文献 1,2 とする。基準文献を 2 つ選定したのは、これら文献はクラスターの技術内容の基準となることから、重要度が高いと考えたためである。一方、類似度の低い文献 n は、クラスター中でユニークな内容を扱っている文献と考えられる [8]。

Step-3：目視により基準文献 1,2 の技術内容を把握する。本研究では、第 1 請求項、発明が解決しようとする課題（目的）及び課題を解決するための手段などから把握した。基準文献の技術内容を基とし、3 番目から n 番目までの文献から抽出されたキーワードを参考にしながら、調査目的に応じてクラスターの技術内容を推定する。

Step-4：類似度が最も低い特許文献 n の技術内容を目視により把握する。Step-3 で推定したクラスターの技術内容の中に特許文献 n の技術内容が含まれておれば、これをクラスターの技術内容とする。含まれていなければ、クラスターの技術内容に含まれる特許文献を二分探索法（バイナリサーチ）により探索し、特定すると共に基準文献から当該文献までの技術内容をクラスターの技術内容とする。二分探索法は、あらかじめ整列している文献を対象とし、検索範囲を 2 分割することで、検索対象を一気に絞り込む効率的な検索手法である。

Step-5：Step-4 で設定したクラスターの技術内容に

基づいてクラスタリングを行うとともに当該クラスターの技術内容を記述する。

3. 結果及び考察

3.1 デジタルカメラの登録状況と

予備クラスタリング結果

O 社の 2004 年 10 月以降に出願された発明について、特許・インテグレーション(株)の ASP サービスにより検索した結果、251 件の登録特許が抽出された。その 251 件の発明について、予備クラスタリングを行った結果を図 2 に示す。図中の 1 点が 1 件の特許文献を表しており、類似性が高い特許文献は近い距離に、類似性が低い特許文献は離れた距離にそれぞれプロットされ、特許検索結果の全体像や特許文献相互の関係を把握し易い。図 2 に示しているように、類似した特許文献が集まったクラスターのうち他のクラスターと重なりが少なく左下部分に見られる 30 件の特許が集積したクラスター（ピンククラスターとする）、左上部分に見られる 10 件の特許が集積したクラスター（黒色クラスターとする。）及び中央やや左よりの 20 件の特許が集積したクラスター（グリーンクラスターとする）を分析対象とした。

3.2 ハイブリッドクラスタリング手法によるクラスタリング及び技術内容記述

(1) ピンククラスターのクラスタリング及び技術内容記述

図2の予備クラスタリングにより得られたピンククラスター中の特許発明30件について、termmiを使ってキーワードを抽出し、またベクトル空間法により類似度を算出した。表2に特許発明の一覧を類似度の高い順に示し、それぞれの特許発明から抽出したキーワードを重要度の高い順に第3位まで示す。

最も類似度の高い特許発明は、特許第4472662号

号で、次いで特許第4900923号、最も類似度が低い特許発明は特許第5501839号であった。抽出されたキーワードは、「レンズ」又は「レンズ群」が全ての特許発明で最も重要度が高くなっていることから、このピンククラスターはレンズから構成される光学系に関する発明であることが分かる。次に抽出されたキーワードの重要度についてみると、第1位、第2位はレンズ又はレンズ群となっており、この傾向は20番目の特許第5432472号まで続いている。また、群の記述がない特許発明は、30番目の特許第5501839号のみであり、これは、明細書を見ると4

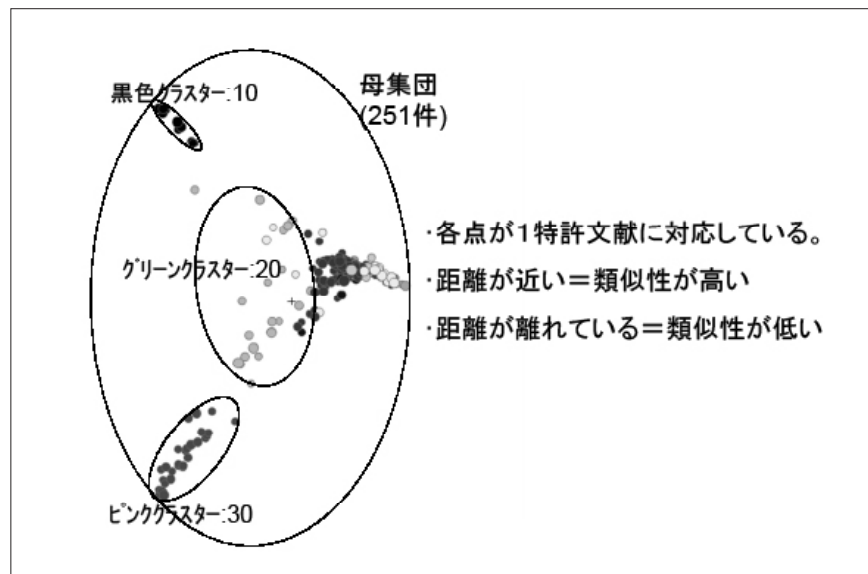


図2 オリンパスGのデジタルカメラに関する登録特許（251件）の予備クラスタリング結果

表2 ピンククラスター中の登録特許とキーワード

特許登録番号	キーワード(重要度順)			特許登録番号	キーワード(重要度順)		
1 特許第4472662号	レンズ群	レンズ	接合レンズ	16 特許第4919360号	レンズ	レンズ群	群
2 特許第4900923号	レンズ群	レンズ	屈折力	17 特許第4766908号	レンズ群	レンズ	レンズ面
3 特許第4758134号	レンズ群	レンズ	電子撮像装置	18 特許第5117521号	レンズ群	レンズ	電子撮像装置
4 特許第5408864号	レンズ群	レンズ	結像光学系	19 特許第4925281号	レンズ群	レンズ	電子撮像装置
5 特許第4947990号	レンズ群	レンズ	変倍光学系	20 特許第5432472号	レンズ	レンズ群	結像光学系
6 特許第4906439号	レンズ群	レンズ	屈折力	21 特許第4863733号	レンズ群	物体側	屈折力
7 特許第4509164号	レンズ群	レンズ	群	22 特許第4813132号	レンズ群	像側	レンズ
8 特許第5075644号	レンズ群	レンズ	撮像装置	23 特許第4884054号	レンズ群	屈折力	広角端
9 特許第4558056号	レンズ群	レンズ	電子撮像装置	24 特許第5105837号	レンズ群	ズーム光学系	屈折力
10 特許第5275718号	レンズ群	レンズ	結像光学系	25 特許第4790320号	レンズ群	電子撮像装置	屈折力
11 特許第4722993号	レンズ	レンズ群	物体側	26 特許第4666669号	レンズ群	像側	像側レンズ群
12 特許第4900924号	レンズ群	レンズ	屈折力	27 特許第4047886号	レンズ	レンズ群	結像光学系
13 特許第4839160号	レンズ群	レンズ	結像光学系	28 特許第4790309号	レンズ	レンズ群	光学系
14 特許第4839158号	レンズ群	レンズ	結像光学系	29 特許第5096517号	レンズ群	像側	撮像装置
15 特許第5302381号	レンズ	レンズ群	物体側	30 特許第5501839号	レンズ	撮像光学系	物体側

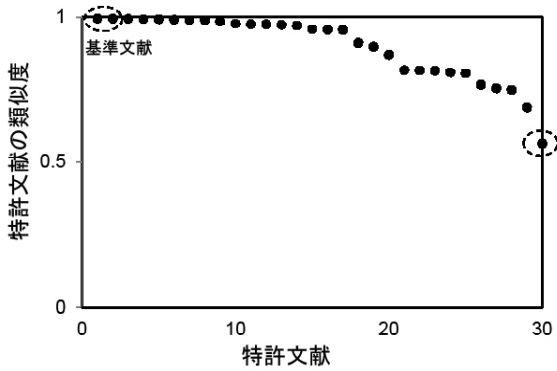


図3 ピンククラスター中の特許文献の類似度

枚のレンズから光学系が構成されている。

図3に、特許文献を類似度順に整理させたグラフを示す。この図で、類似度が高く一般的なトピックスを扱っている特許発明（1番目の特許第4472662号、2番目の特許第4900923号）について、第1請求項、明細書中の発明が解決しようとしている課題（発明の目的）などから技術内容を検討した。1番目の特許発明は、ズームレンズと電子撮像素子を備えたカメラに関する発明で、技術内容は「変倍比も約3倍程度で、広角側の画角が広く、Fナンバーも明るいズームレンズであって、良好な結像性能を得ることのできるズームレンズのレンズレイアウトとそのようなズームレンズを備えてカメラを提供すること」[9]等である。2番目は、ズームレンズやそれを用いた電子撮像装置に関する発明で、特に小型化を志向した電子撮像装置に関するもので、技術内容は「正屈折力の第1レンズ群、負屈折力の第2レンズ群、正屈折力の第3レンズ群、正屈折力の第4レンズ群を有するズームレンズで、第2レンズ群の工夫により、小型化、収差性能の確保を行いやすくしたズームレンズを提供すること、またこのようなズームレンズを備えた小型電子撮像装置の提供等」[10]である。

以上のことから、1番目の特許第4472662号及び2番目の特許第4900923号の技術内容はいずれもデジタルカメラの光学系に関する発明で、レンズ構成やズーム時の動きなど細かいところに違いが見られる。

また、3番目から30番目までの特許文献から抽出されたキーワードは、「レンズ群」「レンズ」「接合レンズ」「屈折力」「電子撮像装置」「結像光学系」「変倍光学系」で、概ねレンズ等の光学系に係るものであった。27番目の特許第4047886号では、特徴語として「ミラースキャン機能」が抽出されていた。

基準文献1.2の技術内容を基準として、更に3番

目からn番目までの特許文献から抽出されたキーワードを参照しながら、ピンククラスターの技術内容について「レンズで構成される高機能な光学系及びそれを備えた撮像装置に関する発明」（技術内容A）と推定した。

以上のことから、ピンククラスターの技術内容として「レンズで構成される高機能な光学系及びそれを備えた撮像装置に関する発明」と推定した。

類似度の最も低い30番目の特許第5501839号の技術内容は、撮像光学系やそれを用いた撮像装置に関する発明で、技術内容は「小型・高画角でありながら諸収差を良好に補正した撮像光学系及びそれを用いた撮像装置を提供すること」[11]であるが、群の記述はない。1番目、2番目では光学系を構成するレンズの枚数が多く「レンズ群」であり、ズーム機構を備えた光学系に関する発明であるが、特許第5501839号では、光学系を形成するレンズは4枚であり、またズーム機構の記述もないものの、撮像光学系やそれを用いた撮像装置に関する発明であるため、技術内容Aに属するとした。

以上のことから、ピンククラスターの技術内容は「レンズで構成される高機能な光学系及びそれを備えた撮像装置に関する発明」とした。ここで「ミラースキャン機能」が抽出された27番目の特許第4047886号の技術内容を明細書により確認した結果、ミラースキャン機能を有する電子撮像装置に係るもので技術内容Aに属しないと除外した。よってピンククラスターは、1～26番目、28～30番目の文献から構成されるとした。

この技術内容の妥当性を実証するため、3～29番目の特許文献について目視により技術内容を確認した。その結果、3～26番目、28番目、29番目の発明の技術内容は、クラスターの技術内容（技術内容A）に属しており、上記の推定は妥当と考えられた。これらのことからピンククラスターでは、ハイブリッドクラスタリング手法の有効性は実証された。

(2) 黒色クラスターのクラスタリング及び技術内容記述

図2の予備クラスタリングにより得られた黒色クラスター中の特許発明10件について、表3に特許発明の一覧を類似度の高い順に示し、抽出したキーワードを重要度の高い順に第3位までを示す。最も類似度が高い特許発明は、特許第5049104号で、次いで特許第4728313号、最も類似度が低い特許発明は、特許第5485428号であった。キーワードは、「光学筐体」「カメラ本体」「光軸」「収納部」等が抽出されており特許文献間で明瞭な違いは認められないが、9番目の特許第5259439号と、10番目の特許第5485428号では「衝撃」が抽出されている。

表 3 黒色クラスター中の登録特許とキーワード

登録番号	キーワード(重要度順)		
	1	2	3
1 特許第5049104号	光学筐体	カメラ本体	光軸
2 特許第4728313号	光学筐体	光軸	収納部
3 特許第4782896号	カメラ本体	光学筐体	収納部
4 特許第4764532号	カメラ本体	光学筐体	収納部
5 特許第4847623号	光学筐体	カメラ本体	収納部
6 特許第4847624号	カメラ本体	光学筐体	収納部
7 特許第4759658号	光学筐体	カメラ本体	カメラ本体収納部
8 特許第4954033号	光軸	光学筐体	収納部
9 特許第5259439号	収納部	衝撃	カメラ本体収納部
10 特許第5485428号	収納部	衝撃	カメラ本体収納部

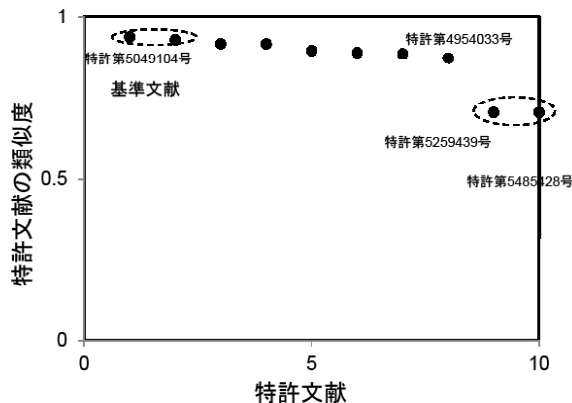


図 4 黒色クラスター中の特許文献の類似度

図 4 に、特許文献を類似度順に整列させたグラフを示す。類似度は、8 番目の特許第 4954033 号まで 0.8 以上と高く、9 番目、10 番目で 0.7 まで低下している。また、表 3 のキーワードには示されていないが、9 番目、10 番目の特許文献では、「ブレ補正機構」が抽出されており、他の 8 つの特許文献からはこのキーワードは抽出されておらず、このような違いが類似度の違いに繋がったものと考えられる。

類似度が高く一般的なトピックスを扱っている 1 番目と 2 番目の特許文献について技術内容を検討した。類似度の最も高い特許第 5049104 号は、折り曲げ光学系を有する光学筐体をカメラ本体内部に収納しているデジタルカメラに関する分野であり、技術内容は「緩衝部材によって生じる反発力等の影響を受けずにレンズの移動方向にかかる衝撃力を吸収し得ると共に、衝撃部材を設けても機器本体のレンズの移動方向の大型化を抑えることのできるデジタ

ルカメラを提供すること」[12]であった(技術内容 B とする)。また、類似度が 2 番目に高い特許第 4728313 号の技術内容を見ると、技術内容 B と同じであった。

また、3 番目から 10 番目までの特許文献から抽出されたキーワードは、「光学筐体」「カメラ本体」「光軸」「収納部」等で特許文献間で明瞭な違いは認められないが、9 番目の特許第 5259439 号と、10 番目の特許第 5485428 号では「衝撃」が抽出されている。

1 番目と 2 番目の特許文献の技術内容を基に、更に 3 番目から 10 番目までの特許文献のキーワードを参考にしながら黒色クラスターの技術内容について「緩衝部材によって生じる反発力等の影響を受けずにレンズの移動方向にかかる衝撃力を吸収し得ると共に、衝撃部材を設けても機器本体のレンズの移動方向の大型化を抑えることのできるデジタルカメラを提供すること」と推定した。

10 番目(特許第 5485428 号)の文献は、衝撃を受けた際に鏡枠に取り付けた衝撃吸収部材により鏡枠への衝撃を吸収するブレ補正機能付きのデジタルカメラに関する技術内容で、「ブレ補正機能を有する鏡枠がカメラ本体収納部内で緩衝移動するデジタルカメラにおいて、非撮影時におけるカメラ本体収納部と鏡枠との間のガタ(空間)をなくし、また、外部からの衝撃によりカメラ本体収納部内を鏡枠が移動する際のスムーズな緩衝移動や緩衝移動後の復帰を正確に行うことのできるデジタルカメラを提供する」[13]となっている。この技術内容は、黒色クラスターの技術内容 B 内に属しないと判断され、技術内容 C とする。

基準文献と 10 番目の文献で、技術内容が異なっていたことから、二分探索により類似度が 6 番目の特許文献(特許第 4847624 号)を目視した結果、技術内容は、「衝撃吸収部材によって生じる反発力等の影響を受けないようにしてレンズを含む光学筐体にかかる衝撃力を吸収し得ると共に、衝撃吸収部材を設けても大型化を抑えることのできるデジタルカメラの提供」[14]で、これは技術内容 B であった。更に類似度が 8 番目(特許第 4954033 号)の技術内容は、「緩衝部材によって生じる反発力等の影響を受けずにレンズの移動方向にかかる衝撃力を吸収し得ると共に、衝撃部材を設けても機器本体のレンズの移動方向の大型化を抑えることのできるデジタルカメラを提供すること」[15]であり、技術内容 B に属するものと判断された。しかし、類似度が 9 番目(特許第 5259439 号)の技術内容は、10 番目の技術内容と同じ記述であり、技術内容 C に属すると判断された。1 番目(特許第 5049104 号)から 8 番目(特許第 4954033 号)までの技術内容は、同じ技術内容

表4 グリーンクラスター中の登録特許とキーワード

特許登録番号	キーワード(重要度順)			特許登録番号	キーワード(重要度順)		
1 特許第4891441号	レンズ群	光軸	像側	11 特許第4790775号	レンズ群	レンズ装置	対物側
2 特許第4891440号	レンズ群	光軸	像側	12 特許第5085023号	カメラボディ	カメラ	撮像面
3 特許第4633752号	レンズ群	レンズ	電子撮像装置	13 特許第5355974号	光像	電子カメラ	撮像光学系
4 特許第5047762号	光軸	光学筐体	光軸方向	14 特許第576295号	撮像レンズ	交換レンズ	デジタルカメラ
5 特許第5363063号	デジタルカメラ	交換レンズ	撮像素子	15 特許第5291892号	撮像素子モジュール	撮像素子	レンズユニット
6 特許第4790250号	光軸	反射面	レフレックスカメラ	16 特許第5450851号	フォーカスレンズ	表示倍率指示部	ライブビュー表示倍率
7 特許第5129172号	駆動部	レンズ鏡筒	撮像素子	17 特許第4536593号	カメラ	防塵フィルタ	光
8 特許第4648477号	圧電素子	防塵フィルタ	振動	18 特許第4624065号	光学ズーム	電子ズーム	撮像装置
9 特許第5060935号	撮像素子	機器筐体	熱反射部材	19 特許第4774452号	振動	防塵フィルタ	振動装置
10 特許第5197454号	フォーカスレンズ	デジタルカメラ	表示倍率指示部	20 特許第5124493号	電子機器	外装部材	金属

Bに属し、9番目(特許第5259439号)と10番目(特許第5485428号)は技術内容Cに属するものと判断された。類似度が4番目の特許第4764532号では、光学的な技術内容に係る「被写体光」が抽出されていた。技術内容Bは機械的な技術内容のため、この特許発明の技術内容を明細書により確認した結果、技術内容Bに属していた。

以上のことから、黒色クラスターは、1番目(特許第5049104号)から8番目(特許第4954033号)までの文献から構成され、その技術内容として「緩衝部材によって生じる反発力等の影響を受けずにレンズの移動方向にかかる衝撃力を吸収し得ると共に、衝撃部材を設けても機器本体のレンズの移動方向の大型化を抑えることのできるデジタルカメラを提供すること」とした。

黒色クラスターの技術内容に係る妥当性を実証するため、3番目、5番目、7番目の特許文献について目視により技術内容を確認した。その結果これらの特許発明の技術内容はいずれも黒色クラスターの技術内容である技術内容Bに属していた。これらのことから黒色クラスターでは、ハイブリッドクラスタリング手法の有効性は実証された。

(3) グリーンクラスターの分析結果

図2の予備クラスタリングにより得られたグリーンクラスター中の特許発明20件について、最も類似度が高い特許発明は、特許第4891441号、次いで特許第4891440号、最も類似度が低い特許発明は、特許第5124493号であった。表4に特許発明の一覧を類似度の高い順に示し、抽出したキーワードを重要

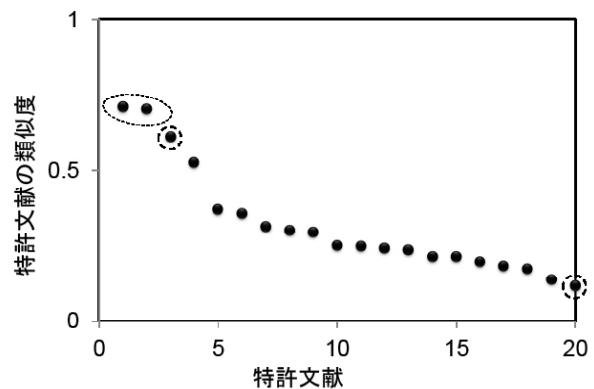


図5 グリーンクラスター中の特許文献の類似度

度の高い順に第3位までを示す。キーワードは、第1位と第2位で共に同じキーワード(レンズ群、光軸、像側)が抽出されていたが、それ以外の特許文献では、キーワードの出現には傾向は認められなかった。図5に、特許文献と類似度の関係を示す。グリーンクラスターでは、類似度が3番目から急激に低下しており、かつ大部分の類似度が0.5以下と低く、前2者とは異なっていた。

このクラスター中で類似度の高い2つの特許発明(1番目の特許第4891441号、2番目の特許第4891440号)について技術内容を検討した。1番目の技術内容は、ズームレンズ及び撮像装置に関するもので、技術内容は「広角高変倍比のズームレンズで、しかも広角端でのF値が2.0乃至2.8程度と明るいズームレンズとそれを用いた撮像装置を提供すること」[16]である(技術内容Dとする)。2番目も

同じ技術内容 D であった。これは、両者で抽出されたキーワードが同じであったことと良く一致していた。しかし表 4 に見られるようにキーワードの出現には傾向が認められないことから、このクラスターの技術内容を文章の形で推定することは、困難と考えられた。因みに 3 番目と 20 番目の特許文献の技術内容は、3 番目の技術内容は、光路折り曲げズーム光学系を有する薄型電子撮像装置に関する分野で、技術内容は「光学系の光路（光軸）をミラー等の反射光学素子で折り曲げたリアフォーカスのズームレンズを採用し、高い仕様、性能を維持しつつ変倍時の移動群の移動に関する制約条件を克服できる光学系のタイプを採用して、奥行方向が極めて薄い電子撮像装置を提供すること」[17]であり、また、20 番目の特許文献の技術内容は、電磁輻射抑制のための複数種類の金属からなる外装部材をそなえた電子機器に関する分野で、技術内容は「複数種類の金属により外装部材を構成した電子機器であって、機器内部からの電磁輻射を抑止しつつ、異種金属間で生じる電食作用の抑止をすることが可能な電子機器の提供」[18]であり、これら何れも技術内容 D とは異なる技術内容である。

このピンククラスターでは、図 5 から、類似度が 3 番目から急激に低下しており、かつ大部分の類似度が 0.5 以下と低いことから、文章の形で技術内容を記述することは困難と考えられた。この様な低い類似度のクラスターについては、上位概念化したキーワードを用いて箇条書きのように記述するなど、技術範囲を広く記載する手法の検討が今後望まれる。

3.3 考察

上記 3 クラスターの技術内容分析結果から、種々の類似度の特許文献を有するクラスターに対するハイブリッドクラスタリング手法の適用について検討を行った。図 2 の「予備クラスタリング」を行った結果で、「密集領域」を形成しているピンク及び黒色クラスターについては共通の技術内容が設定でき、

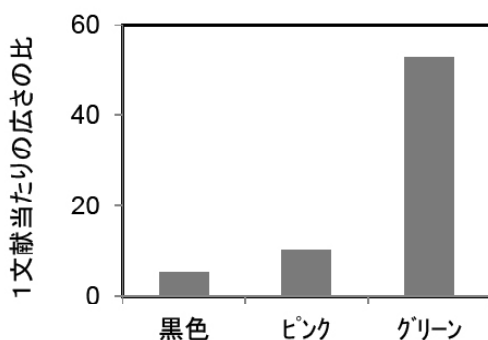


図 6 各クラスターの 1 文献当たりの広さの関係

この設定を基にクラスターの技術内容を推定・記述することができた。一方、「疎な領域」を形成しているグリーンクラスターについては共通の技術内容の設定が困難であった。これらの結果から、クラスターの疎・密状態の定量化について、図 2 に示すように、それぞれのクラスターを囲む楕円を設定し楕円の中の特許文献数から検討し、その値を図 6 に示す。ここで、「1 特許文献当たりの広さ」は、クラスター内の「文献相互の類似性」を表していると考えられる。試料数が 3 つしかないため十分ではないが、大まかに、ピンクと黒色クラスターのように 1 特許文献当たりの領域が狭く文献間の類似性が高いクラスターでは共通の技術内容を記述できるが、グリーンクラスターのように 1 特許文献当たりの領域が広く、文献間の類似性の低いクラスターでは技術内容を設定することは困難であった。このようなクラスターについては、上位概念化したキーワードを用いて箇条書き的に記述する方法が広い技術領域を設定するため有効と考えられる。また、ハイブリッドクラスタリング手法は、類似性が高い文献群のクラスタリングでは有効であったことから、予備クラスタリングが重要であると考えられた。

4. まとめ

膨大な特許文献群から技術動向を作成する場合、従来、目視による作業が行われており多くの時間・労力を要し問題となっている。近年 ICT の進展により、大量の文書を解析するテキストマイニング手法が進展してきたものの、テキストマイニングは探索的データ解析手法で、さまざまな観点からの解釈が容認されており、技術内容分析における曖昧性解消の点で課題が残されている。また、テキストマイニングによる文書クラスタリングでは、抽出されたキーワードを用いて「文書×抽出語」表を作成し様々なクラスタリングが行われているものの、クラスタリング精度に問題を有している。より精度良くクラスタリングを行うため、特許文献群を技術内容に基づいてクラスタリング手法の開発が望まれている。

そこで本研究では、テキストマイニングに係るこれら 2 つの問題点を解決し、大量の特許文献から技術動向を作成するための方法論として、人の意味理解能力とテキストマイニングの大量データ処理技術とを融合するハイブリッドクラスタリング手法を提案した。具体的には、特許文献群のテキストマイニングによる得られる類似度を活用して特許文献群中に基準文献を設け、この文献の技術内容を目視により把握し、これを基に他の特許文献群のキーワード等も活用することでクラスターの技術内容を推定し、この技術内容に基づいてクラスタリングを行う手法

である。

当該手法の有効性についてデジタルカメラに係る特許文献を用いて検討した結果、2次元平面への投影図が密集領域を形成し類似度が高い特許文献群では、技術内容を推定しクラスタリングを行うことが可能で、ハイブリッドクラスタリング手法の有効性が実証された。例えば、今回用いたデジタルカメラに係るクラスター中で類似度が高いピンク及び黒色クラスターでは、共通の技術内容を一意に設定でき、これをクラスターの技術内容として記述することができた。黒色クラスターでは70%、ピンククラスターでは13%の特許文献の目視で良いような結果が得られている。このようにハイブリッドクラスタリング手法は、類似性が高い特許文献群のクラスタリングでは有効であったことから、予備クラスタリングが重要であると考えられた。

謝辞

本論文を査読頂いた先生方には、貴重なご意見を頂き感謝いたします。

本論文の中に、先生方のご指摘事項を取り入れさせて頂きました。ありがとうございました。

(うえの さとる、やまだ みつる)

《参考文献》

1. 名和高司 (2013) 『失われた20年の勝ち組企業100社の成功法則』 PHP 研究所
2. 奥村学、難波英嗣 (2005) 『知の科学 テキスト自動要約』 オーム社
3. 渡邊拓也、太田学、片山薫、石川博 (2004) 『分野に依存しない複数文書要約方法の提案』 第15回データ工学ワークショップ
4. 豊田裕貴、菰田文男編著 (2011) 『特許情報のテキストマイニング』 ミネルヴァ書房
5. 那須川哲哉 (2006) 『テキストマイニングを使う技術 / 作る技術』 東京電機大学出版局
6. パテント・インテグレーションのweb場所：
<https://patent-i.com/ja/index/>
7. 特許庁 (2009) 『平成20年度特許出願技術動向調査報告書 デジタルカメラ装置 (要約版)』
8. termmi の web 場所：<http://gensen.dl.itc.u-tokyo.ac.jp/termmi.html>

9. オリンパス(株) (2010.3.12) 「ズームレンズ及びズームレンズと撮像素子を備えたカメラ」『特許第4472662号』
10. オリンパスイメージング(株) (2012.1.13) 「ズームレンズおよびそれを用いた電子撮像装置」『特許第4900923号』
11. オリンパス(株) (2014.3.20) 「撮像光学系及びそれを用いた撮像装置」『特許第5501839号』
12. オリンパスイメージング(株)、オリンパス(株) (2012.7.27) 「デジタルカメラ」『特許第5049104号』
13. オリンパスイメージング(株) (2014.2.28) 「デジタルカメラ」『特許第5485428号』
14. オリンパスイメージング(株) (2011.10.21) 「デジタルカメラ」『特許第4847624号』
15. オリンパスイメージング(株) (2012.3.23) 「デジタルカメラ」『特許第4954033号』
16. オリンパス(株) (2011.12.22) 「ズームレンズ及び撮像装置」『特許第4891441号』
17. オリンパス(株) (2010.11.26) 「電子撮像装置」『特許第4633752号』
18. オリンパスイメージング(株) (2012.11.2) 「電子機器」『特許第5124493号』