

意図どおりの技術マップを作成するための特許自動分類技術

—大量の特許を対話的に独自の観点で分ける支援技術—

A Clustering Technique for Creating an Intended Patent Technology Map

—Interactive Clustering Based on Individually Specified Classification for a Large Amount of Patent Data—

株式会社日立製作所 情報・通信システム社 公共システム事業部 **佐藤 祐介**

PROFILE

2004年株式会社日立製作所入社。2009年度まで同社中央研究所にて特許文献を対象とした検索支援、文献重要性評価、半自動分類などマイニング技術の研究に従事。2010年度より特許情報提供サービス Sharesearch のシステム開発を担当。

株式会社日立製作所中央研究所／東京工業大学精密工学研究所 **岩山 真**

PROFILE

1992年株式会社日立製作所入社。文書検索、自然言語処理等の研究に従事。また、NTCIRにおいて特許検索用テストコレクションの作成に携わる。2009年度より特許版産業日本語委員会委員。

1 はじめに

特許戦略を立てる際は、競合他社の出願特許を調査し、各社の強み、弱み技術を明らかにし、自社の現状と比較する。そのためには、比較したい技術観点で各社の特許を分類整理する必要がある。例えば、表1は、ロボットに関する各社の特許を技術(目的)で分類した例である。本稿ではこのような表のことを技術マップと呼ぶ。

表1 技術マップの例

	エンターテインメント性 向上	歩行性能 向上	経路計画 性能向上	実用性 向上	作業性能 向上
A社	110	35	15	25	15
B社	0	75	50	10	7
C社	0	30	10	12	10
自社	5	25	7	18	3

この技術マップから、A社はエンターテインメント性向上に注力しているなど、ロボット技術に関する各社の傾向がわかる。また、自社の注力技術はC社と似ていることも読み取れるため、例えば、新規の技術開発を進める、他社と協業し弱点を補う、といった戦略を立てることもできる。技術マップの作成には人手コストがかかるため、その自動化が強く望まれている。本稿では、技術マップを効率良く作成するための方式を紹介する。

2 技術マップ作成におけるニーズと課題

技術マップを作成する際は、比較のための分類基準を決める必要がある。IPCなど既存の分類体系がそのまま使えれば良いが、IPCの分類体系が必ずしも自社の製品の特徴を反映しているとは限らない。その場合、独自の分類基準で、各社を比較したいというニーズが出てくる。

独自の分類基準を作成する方式には、文書集合を与えるだけで、文書を自動分類するものもあるが、結果を見ても、どのような基準で文書が分類されたのかが分からないことも多い。結果として、想定どおりに分類されない場合は、人手で分類することになってしまう。一方で、分類基準を自動的に学習し、学習した分類基準に従って文書を自動分類する方式もあるが、学習の際に多数の事例(明細書と分類結果の対)を正解として与えないと十分な分類精度が得られない。

本稿では、上記2つの方式を組み合わせた対話型の技術マップ作成方式を紹介する。本方式により、思いどおりの技術マップが低コストで作成できるようになる。

3 技術マップ作成の流れ

本稿で提案する技術マップ作成の流れは図1のとおりである。

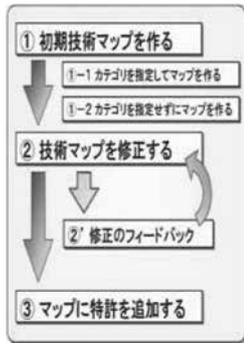


図1 技術マップ作成の流れ

まず、システムが初期技術マップを自動作成する(①)。その後、作成した技術マップに対して、ユーザがシステムに正解を与えると(②)、システムはその正解に基づいて技術マップを更新する(②')。必要であれば、更新された技術マップに対して新たな正解をシステムに与えることもできる。この操作を繰り返すことで、次第に求める技術マップに近づいていく。技術マップに新たな特許を追加する(③)こともできる。

②において、ユーザは一度に全ての正解をシステムに与えなくてもよい。一部の正解をシステムに与えることで、システムは残りの特許についても学習結果を反映させるため、以前は誤って分類されていた特許が更新後に正しく分類される場合もある。つまり、②と②'の繰り返しの中で、ユーザは段階的に効率良く意図した技術マップを作成することができる。

3.1 初期技術マップの作成

まずは、初期技術マップを作成する。この技術マップは逐次修正されるため、完璧なものでもなくてもよい。

こういった技術観点で分類したいかが既に分かっている場合は、分類基準(カテゴリと呼ぶ)をキーワードや文章で指定する(①-1)(図2参照)。システムは入力された分類基準に基づいて初期技術マップを作成する。

項目	名称	内容
1	名称: 駆動装置の駆動協調制御 代表公開: [不明] キーワード: 自動車の駆動協調制御において所望の駆動力から出力される駆動力のエネルギーを排出し、駆動装置の異常(故障)発生に伴う異常を速やかに検出して車両に通知する。エンジンモーター/ルックアップ/クラッチ/変速機、最終減速ギア等	
2	名称: 駆動装置の制御 代表公開: [不明] キーワード: 複数のバッテリーまたは電力蓄積デバイスを用いてシステムの要求に応じた電力的な電力収支を管理する技術	

図2 カテゴリ指定例

分類基準が明確でない場合は、カテゴリの数のみを指定して初期技術マップを作成することもできる(①-2)。いずれの作成方法においても、「クラスタリング」と呼ばれる方式を利用する。各文書に現れる単語の頻度分布に基づいて文書間の類似度を計算し、類似する文書を自動的にまとめる。ただし、全てシステム任せになるため、どのような基準で文書がまとめられたかが分かりにくいケースも多い。いずれの場合もカテゴリの名前は人手で入力する。

図3は初期技術マップの例である。

項目	出題人	駆動装置の駆動協調制御	駆動装置の駆動協調制御	駆動装置の駆動協調制御	駆動装置の駆動協調制御	駆動装置の駆動協調制御	駆動装置の駆動協調制御
1	出題人	1000	1000	1000	1000	1000	1000
2	出題人	1000	1000	1000	1000	1000	1000
3	出題人	1000	1000	1000	1000	1000	1000
4	出題人	1000	1000	1000	1000	1000	1000
5	出題人	1000	1000	1000	1000	1000	1000
6	出題人	1000	1000	1000	1000	1000	1000
7	出題人	1000	1000	1000	1000	1000	1000
8	出題人	1000	1000	1000	1000	1000	1000
9	出題人	1000	1000	1000	1000	1000	1000
10	出題人	1000	1000	1000	1000	1000	1000
11	出題人	1000	1000	1000	1000	1000	1000
12	出題人	1000	1000	1000	1000	1000	1000
13	出題人	1000	1000	1000	1000	1000	1000
14	出題人	1000	1000	1000	1000	1000	1000

図3 作成された初期技術マップの例

3.2 技術マップの修正

初期技術マップを作成した後は、このマップに対して正解をシステムに与えながら技術マップを更新していく。正解は以下のいずれかの方法で指定する(②)。

- ・正しく分類された特許を指定する。
- ・誤って分類された特許の正しい分類先を指定する。

技術マップの各セルをクリックすると、そのセルに分類されている特許がリストアップされるので、そのリスト中で上記の2種類の方法のいずれかを選択して実行する。

図4のように、プルダウンメニューから正しい分類先が選択できる。

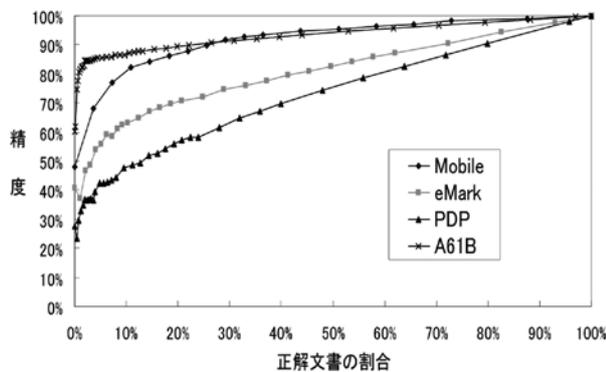
発明の名称	出題人	IPC	発行日	公開日	審査・権利状況	【ハイブリッド技術】協調・全システム制御技術
内燃機関の出力制御装置	トヨタ自動車株式会社	B60W10/06	2012/04/05	2012/04/05	登録	10 駆動装置の駆動協調制御
ハイブリッド車両	トヨタ自動車株式会社	B60W10/06	2011/10/27	2011/10/27	登録	10 駆動装置の駆動協調制御
車両の高電圧ケーブルの配線構造	トヨタ自動車株式会社	B60R15/02	2011/05/31	2011/05/31	登録	10 駆動装置の駆動協調制御
ハイブリッド車両の制御装置	トヨタ自動車株式会社	B60W10/06	2007/11/15	2007/11/15	登録	10 駆動装置の駆動協調制御
車両の制御装置	トヨタ自動車株式会社	B60W10/06	2004/02/12	2004/02/12	登録	10 駆動装置の駆動協調制御

図4 技術マップ更新操作例

システムは、与えられた正解特許を使い、技術マップを自動的に更新する(②')。ここでは、「制約付きクラスタリング」という技術を使う。

通常のクラスタリング技術は、クラスタの重心(重心とは、そのクラスタに属している文書により計算した平均値で、おおそクラスタの中心に相当する)が安定するまで再分類と重心の更新を繰り返す。一方、制約付きクラスタリング技術では、与えられた正解情報を常に満たしながら更新を繰り返す。正解を与えることで、クラスタリングの挙動を制御できる点が特徴である。

正解の指定(②)とそれに基づく技術マップの修正(②')というサイクルを繰り返しながら技術マップを更新していく。グラフ1に、システムに与えた正解の総数とクラスタリング精度との関係を示す。ここでは、特許庁が公開している技術分類マップから4テーマを選び、初期技術マップをカテゴリ数のみとして(①-2)、ランダムに5個ずつ正解をシステムに与えながら技術マップを逐次更新していった。



グラフ1 正解文書の割合と分類精度の関係

テーマ毎に精度曲線は異なるが、10~20%程度の正解をシステムに与えることで精度が大きく向上していることがわかる。最大で(Mobileの場合)、20%の正解をシステムに与えることで90%の精度が得られた。

しかしながら、クラスタリングを活用した方式では、分類精度が100%の技術マップを作成することが困難という課題がある。特に、ある程度の数の正解をシステムに与えると、それ以降は与えた正解の分しか精度が向上しない。つまり、制約付きクラスタリングによる精度向上効果が見込めなくなってしまう。グラフ1では、精度向上曲線が直線になっている範囲がそれに相当する。この状態に達してしまうと、残りは人手で分類するか、

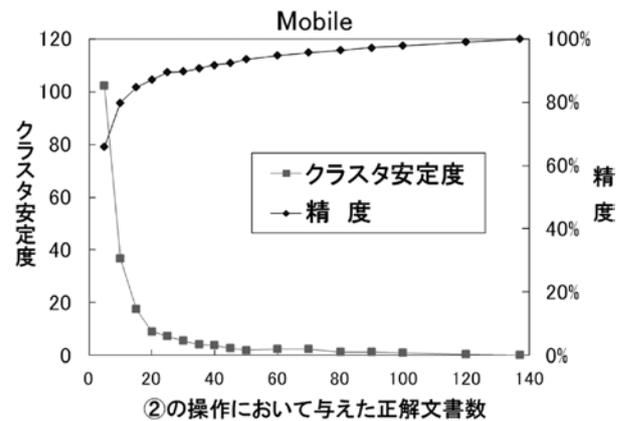
別の手法で自動分類するしかない。

ところが、精度曲線は正解がわかっていないと描けないため、上記のようなシステムに正解を与える止め時を事前に知ることはできない。そのため、我々は、精度曲線を近似するクラスタ安定度という指標を提案している^[2]。

この提案では、制約付きクラスタリングの内部処理において、所属先クラスタが変わった文書の総数をクラスタ安定度と定義している。グラフ2は、Mobileというテーマでの実験において、以下の2つの関係をプロットしたものである。

- ・クラスタ安定度(左縦軸)と②の操作において与えた正解文書数(横軸)の関係
- ・制約付きクラスタリングの精度(右縦軸)と②の操作において与えた正解文書数(横軸)の関係

これら2つのプロットを見ると、クラスタ安定度と精度には関連があることがわかる。この場合の相関係数は90%である。他のテーマでも相関係数は85%~92%と高かった。よって、ユーザは、クラスタ安定度の変化を見ることで、正解を与えた効果を予測できる。



グラフ2 クラスタ安定度と精度の関係

4 おわりに

本稿では、技術マップを効率良く作成するための方式を紹介した。ユーザは正解をシステムに与えることで、自分の意図に合った技術マップを作成することができる。また、正解を与える操作とシステムによるクラスタリングを繰り返しながら対話的に技術マップを更新していくため、最終的にシステムに与える正解の数も少量で済む。更に、クラスタリングの効果が無くなるタイミングもわかる。

本方式の一部は、特許情報提供サービス Sharesearch の分析オプションとして提供されている。分析オプションは、技術マップの作成機能（クラスタリング機能）に加え、統計分析機能、被引用分析機能、分析ポータル機能等を提供し、特許分析業務の効率化を総合的にサポートする。

参考文献

- [1] Sato, Y., Iwayama, M., "Interactive Constrained Clustering for Patent Document Set", Proceedings of the CIKM 2009 Patent Information Retrieval Workshop, pp. 17-20, 2009.
- [2] 佐藤佑介, 岩山真, "半教師有りクラスタリングを適用した対話的文書分類技術の提案", 情報処理学会情報学基礎研究会, 2009-FI-95, No. 7, 2009.