

T5 を用いた技術課題・解決手段推定による特許マップ自動生成

小堀 佑樹 酒井 浩之 永並 健吾
成蹊大学 理工学部 情報科学科

us202062@cc.seikei.ac.jp, {h-sakai, kengo-enami}@st.seikei.ac.jp

概要

本研究では、T5 モデルを使用して特許文書から特許マップを生成するための技術課題・解決手段の端的な表現を生成する。技術課題の学習データは人手で作成し、解決手段の学習データはタイトルデータをターゲット文として機械的に作成する。推定された技術課題・解決手段のカテゴリを Word2Vec により統合することで直感的な特許マップを生成する。推定された技術課題は人手評価、解決手段はタイトルデータとの一致度合いを機械的に評価する。また、人手で作成された特許マップのカテゴリと自動生成された特許マップのカテゴリの類似度を評価する。

1 はじめに

近年、自社企業や競合企業の特許文書を用いた分析手法が数多く行われている。その手法の一つとして特許マップの作成があげられる。図1は特許流通支援チャートに掲載されたプラスチック光ファイバの特許マップ[1]である。図1の技術課題は発明が解決しようとする課題、解決手段はこの課題を解決するための手段である。特許マップを作成することで、自社の知財状況、他社の技術動向の視覚的な理解が可能になる。しかしながら、技術課題と解決手段を軸に取った特許マップは主に人手で推定・作成が行われており、企業の分析・比較には多くの手間や時間が必要になる。

そこで本研究では、特許文書から抽出した技術課題・解決手段のデータから T5[2]を使用して技術課題のカテゴリ・解決手段を推測する。そして推定したカテゴリを軸として特許マップを自動的に生成する。また、Word2Vec[3]を用いて人手で作成された特許マップとの類似度を求め、本手法によって自動的に生成した特許マップの評価を行う。

本研究の関連研究として、酒井らの特許文書から技術課題を示す表現を自動抽出する手法[4]や特許

図 1.4.1-2 プラスチック光ファイバの課題と解決手段の分布

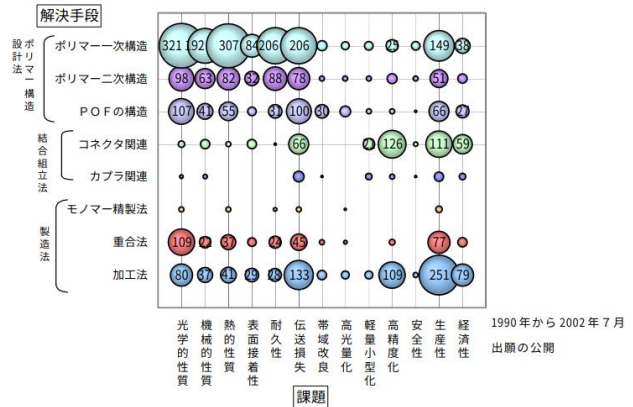


図 1 人手で作成された特許マップの例

文書からの技術課題情報の抽出手法[5]があり、抽出した手がかり表現から共通頻出表現を抽出することで技術課題情報の抽出を試みていた。また、特許の請求項を利用した手法[6]では、TF・IDF 値から特許文書間の類似度を計算することで企業分析を試みている。本研究では請求項ではなく、技術課題・解決手段に該当する記述を利用して特許マップを自動生成する。さらに、Word2Vec を使用して人手で作成した特許マップとの類似度を求め評価している点が異なる。Xiang らの手法[7]では BERT を用いて技術課題・解決手段のフレーズを抽出している。これらの研究に対し、本研究は T5 を用いている点で異なる。

2 技術課題・解決手段の推定

本研究では 2020 年度に公開された特許データ 207,169 件を使用する。技術課題・解決手段に該当する記述として、特許明細書の【発明の名称】(invention-title), 【発明の概要】内の【発明が解決しようとする課題】(tech-problem), および【課題を解決するための手段】(tech-solution)を使用する。

2.1 T5 を用いた技術課題推定

特許文書の技術課題記述からの特許マップに使用する技術課題推定は以下の3つの Step で構成される。

Step 1: 特許文書から【発明が解決しようとする課題】を技術課題記述として抽出

Step 2: Step1 で抽出した技術課題記述に対応する技術課題を手で作成

Step 3: Step1, Step2 で得られた学習データを用いて、T5 で特許文書から技術課題を推定

T5 の学習データは無作為に選んだ計 220 件の特許データから技術課題記述を抽出し、人手で対応する技術課題を作成する。技術課題は特許流通チャートに記載されている技術課題を参考に人手で作成した。T5 の学習データの例を表 1 に示す。

表 1 技術課題推定の学習データ

技術課題	技術課題記述
耐久性	特許文献1のように基材を加工して凸部を形成する方法では、加工上の制約で一定以上の高さでしか凸部を形成することができないため、凸部の高さが高くなりすぎて、摺動抵抗の増加による定着部材等の部材の破損が生じやすくなるという課題があった。

作成した学習データでファインチューニングした T5 v1.1 事前学習済みモデルにより技術課題を推定する。技術分類の特許は国際的に統一された分類体系である IPC 第 4 版をベースとしてさらに細分化された日本独自の分類であるファイルインデックスと、キーワードを組み合わせてアクセスする。表 2 にテストデータとして用意した技術分類とその特許数を示す。

表 2 技術分類と特許データ数

技術分類	対象の特許数
プラスチック光ファイバ	569 件
ハイブリッド電気自動車制御	632 件
食品乾燥加工技術	434 件

各技術分類に該当する特許文書に学習済み T5 v1.1 モデルを使用して推定された技術課題を表 3 に示す。

表 3 推定された特許の技術課題の例

技術分類	技術課題
プラスチック光ファイバ	耐久性
ハイブリッド電気自動車制御	効率化
食品乾燥加工技術	対真空強度

2.2 T5 を用いた解決手段推定

特許文書の解決手段記述からの特許マップに使用する解決手段推定は以下の3つの Step で構成される。

Step1: 特許文書から【課題を解決するための手段】を解決手段記述として抽出

Step2: 特許タイトルに対応する解決手段とみなし、学習データを作成

Step3: Step2 で得られた学習データを用いて、T5 で特許文書から解決手段を推定

発明の部品名や機能名などが特許タイトルになる。課題を解決する手段として、特許タイトルは適していると考えられる。

特許マップを生成する技術分類の特許データから無作為に選んだ 220 件の特許文書のタイトルと解決手段記述を抽出して T5 の学習データを自動作成する。学習データとして使用した文の一部を表 4 に示す。

表 4 解決手段推定の学習データ

解決手段	解決手段記述
光コネクタ	上記目的を達成するための主たる発明は、メカニカルスプライス部と、前記メカニカルスプライス部を収容するとともに、光ファイバを口出した光ケーブルを把持する把持部材を固定するためのハウジングとを備え、前記ハウジングは、後部の幅方向の寸法を狭めた形状をしており、上側の開放した半錐体面状の下案内内部を有するベースと、下側の開放した半錐体面状の上案内内部を有するカバーとを有しており、前記下案内内部と前記上案内内部とによって、前記光ファイバを前記メカニカルスプライス部に案内するための錐体面状の案内内部が構成されていることを特徴とする光コネクタである。

作成した学習データでファインチューニングした T5 v1.1 事前学習済みモデルとテストデータから解決手段を推定する。各技術分野に該当する特許文書に学習済み T5 v1.1 モデルを使用して推定された解決手段の一部を表 5 に記す。

ⁱ sonoisa/t5-base-japanese-v1.1

表 5 推定された特許の解決手段

技術分類	解決手段
プラスチック光ファイバ	樹脂成形体及び樹脂成形体の製造方法
ハイブリッド電気自動車制御	電池制御装置
食品乾燥加工技術	重量測定装置

3 特許マップ自動生成

推定された技術課題，解決手段は，ほぼ同じものが別の表現で出力されることがある（効率化，効率，効率向上など）．特許マップの見やすさのために Word2vec で推定されたカテゴリ（技術課題，解決手段）の平均ベクトル（分散表現）を求め，統合する．Word2Vec モデルは 2017 年の全ての特許文書を使ったモデルを使用した．

- Step1: カテゴリに含まれる各単語の分散表現からカテゴリの平均ベクトルを求める．
- Step2: 頻出数が 2 以下のカテゴリ語は統合対象とし，Step1 で求めたカテゴリ語の平均ベクトルから，頻出数 3 以上のものとの \cos 類似度を求め，最も類似度の高いものに統合する．

統合したカテゴリを用いて特許マップを生成する．図 2 に本手法にて自動的に生成した技術分類「プラスチック光ファイバ」の特許マップを示す．また，T5 によって推定された解決手段カテゴリと比較するため，FI ファセット末端のサブグループやタイトルを解決手段カテゴリとして，Word2Vec を用いて上述の方法で統合した特許マップを生成する．FI ファセット末端のサブグループを用いて生成した特許マップを付録 A に，タイトルを用いて生成した特許マップを付録 B に示す．

4 評価

4.1 技術課題・解決手段推定の評価

T5 v1.1 モデルで推定された技術課題が特許マップの技術課題カテゴリとして正しいか，人手で判定する．比較のため，T5 モデルⁱⁱで推定された技術課題の判定結果も併記する．

ⁱⁱ sonoisa/t5-base-japanese

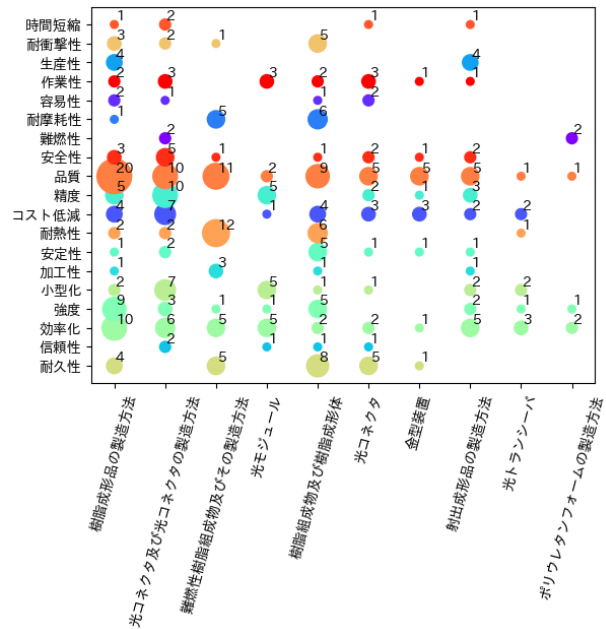


図 2 自動生成したプラスチック光ファイバの特許マップ

表 6 技術課題の正誤判定

技術分類	プラスチック光ファイバ	ハイブリッド電気自動車制御	食品乾燥加工技術
T5	18 / 30	19 / 30	24 / 30
T5 v1.1	22 / 30	19 / 30	26 / 30

(正答数 / 総判定数)

解決手段の評価として，タイトルを正解とし，T5 v1.1 モデル，および，T5 モデルで推定された解決手段の結果の一致度合いを 4.2 特許マップとの比較による評価に示す．

4.2 特許マップとの比較による評価

人手で作成された特許マップと本手法によって自動生成された特許マップとを比較し，評価する．評価方法を以下に示す．

- Step1: Word2Vec でカテゴリ語の平均ベクトルを算出し，さらに特許マップに使用したカテゴリ群で平均をとり全体の平均ベクトルを算出する．
- Step2: 特許流通チャートの手による特許マップから同様に全体の平均ベクトルを算出し， \cos 類似度を計算する．

人手による特許マップとの類似度が高いほど，適した特許マップを生成できたことになる．

表 7 解決手段の正誤判定

技術分類		プラスチック光ファイバ	ハイブリッド電気自動車制御	食品乾燥加工技術
T5	完全一致	96 / 349	233 / 412	58 / 214
	部分一致 1	116 / 349	106 / 412	74 / 214
	部分一致 2	20 / 349	6 / 412	7 / 214
	一致しない	117 / 349	67 / 412	75 / 214
T5 v1.1	完全一致	88 / 349	228 / 412	85 / 214
	部分一致 1	80 / 349	101 / 412	47 / 214
	部分一致 2	32 / 349	11 / 412	8 / 214
	一致しない	149 / 349	72 / 412	74 / 214

(部分一致 1 はタイトルが解決手段を含む, 部分一致 2 は解決手段がタイトルを含む)

(正答数 / 総判定数)

比較手法として, 技術課題では Cabocha を用いて「～性」の単語を抽出し, その単語群平均ベクトルと特許流通チャートのカテゴリ群平均ベクトルの cos 類似度を比較手法として計算する. また, 解決手段ではタイトルを使用した場合, FI ファセットを使用した場合の特許マップを評価する. 結果を表 7, 表 8 に示す.

表 7 技術課題の類似度

技術分類	プラスチック光ファイバ	ハイブリッド電気自動車制御	食品乾燥加工技術
Cabocha	0.900	0.783	0.787
T5	0.917	0.788	0.882
T5 v1.1	0.922	0.863	0.847

表 8 解決手段の類似度

技術分類	プラスチック光ファイバ	ハイブリッド電気自動車制御	食品乾燥加工技術
タイトル	0.748	0.832	0.590
FI	0.817	0.763	0.808
T5	0.766	0.833	0.691
T5 v1.1	0.770	0.836	0.721

5 考察

特許文書に対して, 提案手法である「T5 による技術課題・解決手段推定」と「Word2Vec によるカテゴリ統合」を行うことで, 人手で作成された特許流通チャートとのカテゴリ群の類似度は, 技術課題は最大で 0.922, 解決手段は最大で 0.836 を示した. タ

イトルデータから Word2Vec のみで生成された特許マップよりも T5 で生成されたものの方が解決手段の類似度の高い結果となり, T5 によって元のタイトルデータとは違う出力をしたものが類似度を上昇させたか, もしくは Word2Vec によってカテゴリを統合したことが結果の向上につながったと考えられる.

FI ファセットを解決手段カテゴリに用いると, 類似度が高くなった. しかし, FI は文言が長いため, 特許マップのカテゴリとして直感的でない. その点において本手法による特許マップのほうに実用性があると考えられる.

6 むすび

本研究では, 特許文書から技術課題・解決手段を軸に取る特許マップを自動生成する手法を提案した. T5 を用いて特許文書から技術課題・解決手段をそれぞれ推定し, Word2Vec を用いてカテゴリの統合を行うことで, 自動的にかつ視認性の高い特許マップを生成することができた.

評価の結果, 技術課題・解決手段の双方について「T5 による技術課題・解決手段推定」と「Word2Vec によるカテゴリ統合」を行うことで, 人手で作成された特許流通チャートとの類似度の高い特許マップを得ることができた. しかしながら, 本手法では技術分野の特許文書の半数近くを学習データとして使用してしまうために, 特許マップに反映される特許数が少ない. 今後は二分割交差検証を行うことで, 特許マップの有効特許数を多く保ったまま有用に活用できるようにしていくことが求められる.

参考文献

- [1] 特許流通支援チャート 平成14年度 化学
11 プラスティック光ファイバ,
<https://www.inpit.go.jp/blob/katsuyo/pdf/chart/fkagaku11.pdf>
- [2] Colin Raffel, Noam Shazeer, Adam Roberts,
Katherine Lee, Sharan Narang, Michael Matena,
Yanqi Zhou, Wei Li, Peter J. Liu, “Exploring the
Limits of Transfer Learning with a Unified
Text-to-Text Transformer”, *Journal of Machine
Learning Research* 21, pp.1-67, 2020
- [3] Mikolov, T., Chen, K., Corrado, G., & Dean,
J. (2013). Efficient estimation of word
representations in vector space, *arXiv preprint
arXiv:1301.3781*.
- [4] 坂地 泰紀,野中 尋史,酒井 浩之,増山 繁:
“特許文書からの課題・効果表現対の自動抽出”,
電子情報通信学会論文誌 D, Vol.J93-D, No.6,
pp.742-755, 2010.
- [5] 酒井 浩之,野中 尋史,増山 繁:“特許明細
書からの技術課題情報の抽出”, *人工知能学会論
文誌* 24巻6号I, pp.531-540, 2009.
- [6] 金地 しおり,酒井 浩之,北島 良三, :“特
許の請求項を使用した企業分析手法”, 平成30
年度, 成蹊大学理工学部情報科学科 卒業論文.
- [7] Xiang Shi, Zikun Feng, Jiawei Liu, Qikai Cheng,
Wei Lu: “Automatic Construction of technology
Function Matrix”, *3rd Workshop on Extraction of
Knowledge Entities from Scientific
Documents(EEKE2022)*, p74-79, 2022.

