

Few-shot 学習を用いた特許分類

Patent Classification Using Few-shot Learning

中央大学理工学部ビジネスデータサイエンス学科教授

難波 英嗣

2001年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。博士（情報科学）。東京工業大学精密工学研究所助手。広島市立大学大学院情報科学研究科准教授等を経て、2019年より中央大学理工学部教授。自然言語処理、テキストマイニングの研究に従事。

✉ nanba@kc.chuo-u.ac.jp

☎ 03-3817-1883

1 はじめに

本研究では、少量の教師データで学習が可能な Few-shot 学習を用い、特定領域の特許を効率的に収集するための分類器を構築する。一般に特許検索では、まず、検索キーワードと国際特許分類や F タームなどの分類コードを組み合わせて複雑な検索式を構築し、この検索式を用いて目的の特許を検索する。次に、得られた検索結果を一件ずつ人手で確認し、目的の特許を収集する。ここで、企業が長期間にわたって製品を開発する場合、この一連の検索作業は開発期間中に繰り返し実施する必要があるが、最終ステップである検索結果の確認を毎回人手で行うには非常にコストがかかるため、この確認作業の自動化が求められている。

検索結果の中から必要な特許を見つけ出す作業は、文書分類の一種と考えられる。近年では、文書分類を行う際に深層学習が使われるのが一般的であるが、これには大量の訓練データが必要となり、データの作成コストなどが理由でデータを用意できないことが少なくない。これらの問題に対して本研究では、少量の教師データでも学習をすることができる Few-shot 学習を使い、特許分類器を構築する。

Few-shot 学習では、事前学習時と fine-tuning 時に分かれており、大量の訓練データで事前学習をしていれば、fine-tuning 時とのタスクやデータのドメインが近い場合、fine-tuning 時に少量の訓練データで学習をすることができる。そこで本研究では、特許のテキストと F タームと呼ばれる特許に割り当てられる分類体系をラ

ベルとして使い、「自分たちが求めている特定分野の特許」か「それ以外か」に分類する少量のデータで学習が可能な特許の分類器を作成する。本研究により、大量のデータを必要とせず、少量のデータで特許の分類器を学習することが可能となり、データの作成コストの問題が解消される。また、特許調査において、必要となる特許を効率的に調査することも可能となる。

2 関連研究

2.1 Few-shot 学習

Few-shot 学習とは、転移学習 [1] の一種で、少量の学習データでも学習することができる学習手法である。転移学習とは、ある領域において豊富にある学習データで学習した知識を、別の領域の学習に転移させる学習手法である。Few-shot 学習は、事前学習時と Few-shot 時に異なるクラスのデータを使用する問題設定で、事前学習時に得たパラメータを初期値として、Few-shot 時に少量のデータを使って事前学習時にはなかったクラスの Few-shot 用のデータの予測を行う。Few-shot 学習は主に、画像分類などのコンピュータービジョンの分野で使用されてきた。Snell ら [2] は、Few-shot 学習において、図 2 のように各クラスのプロトタイプ表現と、クエリとの距離を計算することで、画像などを分類することができる距離学習を提案している。また Few-shot 学習には、メタ学習と呼ばれる学習方法を学習することを目的としている手法があり、目的のタスクに関連するデータから学習方法を学習し、目的のタスクの性

能を向上させることを目指している。しかし Chen[3] による、Few-shot 学習を使った分類タスクの評価方法の問題を指摘した比較実験によると、メタ学習 [4] よりも、最適化された Fine-tuning ベースの Baseline による Few-shot 学習が最も精度等が高いと発表している。そのため、本研究では、上記で紹介した Fine-tuning を基にした Few-shot 学習をベースとして利用する。

2.2 文書分類

F タームを使った特許分類

SVM を用いた特許分類が行われており、Li ら [5] は一般的な SVM をベースに F タームを使った特許分類を提案している。入力データに、特許文書の全ての内容と、F タームの説明文を使用している。それらの文書から、名詞、独立動詞、独立形容詞、未知のいずれかの単語を特徴語として選択し、それらを tf-idf で特徴ベクトルを計算、正規化を行い、SVM に入力を行なっている。富永ら [6] は、ニューラルアテンションモデルをベースに、付与根拠箇所を利用した特許分類を提案している。このモデルは、付与根拠箇所推定と F ターム分類の 2 つのモデルが合わさっており、前者では人手による付与根拠箇所データを教師データとして付与根拠箇所を推定している。後者は、推定した付与根拠箇所を使い、F タームの分類を行う。本研究と先行研究の異なる点は、Few-shot 学習を使うかどうかであり、本研究は少量のデータで特許分類を行える点である。

Few-shot 学習を用いた文書分類

Few-shot 学習ではドメインに関する問題が発生することが多い。その理由は事前学習時と Few-shot 学習時においてデータの領域が異なったり、クラス同士の関連性が弱かったりすること原因で、あるタスクでは有益な語彙的特徴が、別のタスクでは重要ではない場合があるからである。Yu ら [7] による研究では、タスク同士の関連度を数値化し、関連度の高いタスク動詞をそれぞれクラスに割り当て、各クラスで学習を行うことでモデルの性能を向上させている。また Brown ら [8] が開発した GPT-3 による Few-shot 学習を用いた文書分類もある。GPT-3 は事前学習をした自然言語処理のモデルの 1 つであり、膨大な量のテキストで事前学習し、1750 億個のパラメータを持った言語モデルである。

ただ、事前学習時のデータには特許が含まれているのが分からないという問題と、GPT-3 の事前学習のデータを変えて特許に特化したモデルにすることができないという問題がある。本研究はテキスト情報を扱うが、特許という専門性に富んだテキストを事前学習し、それを分類することを目的としている。

3 Few-shot 学習を用いた特許分類

(1) 検索式を用いて特許検索を行い、(2) 検索結果に文書分類を適用することで目的の特許を収集する、という一連の検索作業の中で、手順 (2) で用いる文書分類器を構築することが本研究の目的であり、おおよそ 100 ~ 200 件程度の訓練事例が利用できるという状況を想定している。この目的を達成するためのデータセットと Few-shot 学習による分類器の構築について本設で述べる。

Few-shot 学習は、解きたい問題と類似する問題を事前学習時に大量に学習させておき、学習済みモデルに解きたい問題のデータを使って Fine-tuning させることで、Few-shot 学習を実現する。今回は、事前学習時の似たような問題として「F ターム」を使い、モデルの有効性を検証する。

F タームとは、日本の特許文献に記載された発明の技術的特徴による分類体系、また分類記号である。特許中の技術観点、「目的、用途、構造、材料等」を付した F タームリストに照らして特許ごとに付与している。F タームの構造は、「テーマ」という技術分野と、その技術分野を「観点」と呼ばれる発明の目的や利用分野、材料などで細分化した、9 桁の英数字で表される。

本研究で扱うデータセットは、事前学習用と Few-shot 用の 2 つがある。それぞれのデータセットの特許には重複がなく、また Few-shot 用のデータセットの特許に付与されている F ターム (正例となる F ターム) を含む特許を事前学習用のデータセットから排除している。また、F タームを使ったデータセットの中で、主に「ランダムカテゴリ」と「近接カテゴリ」の 2 種類を Few-shot 用のデータセットとして扱う。ランダムカテゴリとは、正解ラベルにある 1 つのテーマや F タームを決め、負例をランダムなテーマや F タームにすることである。例えば、ある 1 つの F タームを正解ラベルとし、この

F タームが付与されている特許 100 件を正例として抽出して、正解ラベルとしている 1 つの F ターム以外の F タームが付与されている特許から、負例として 100 件を抽出する。近接カテゴリとは、図 1 の F タームの「FA01」と「FA08」のように、階層が同じで隣り合うカテゴリのことである。正解ラベルをある 1 つのテーマや F タームにし、不正解ラベルを隣接するテーマや F タームにする。

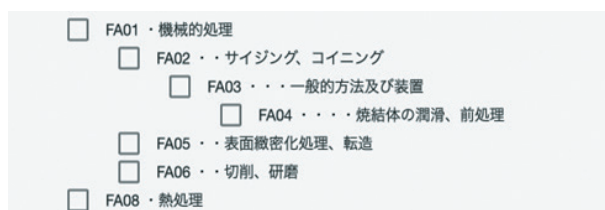


図 1 隣接する F タームの例

この 2 種類のデータを扱い、複数の異なる条件のデータセットを作成する理由としては、実際に解く問題が、問題ごとに分類の粒度が異なるためである。そのため、F タームを使った場合、どの条件でどのくらいのパフォーマンスが出るのかを調べる。

データセットに使う特許は、日本の特許データである。本研究では、特許に記載されている「要約文」と特許に付与されている「F ターム」を利用する。そのため、要約文がない特許に関しては、データセットに含めていない。

4 実験

4.1 実験条件

■実験データ

実験データは大きく分けて 2 つある。1 つ目は、2004 年から 2014 年の特許を事前学習用に、2018 年の特許を Few-shot 用に使用した。事前学習用のデータと Few-shot 用のデータには重複がなく、テーマと F タームの重複もない。モデルに入力する文書は特許の要約分を利用している。事前学習は、学習データ 287,971 件、検証データ 191,981 件、テストデータ 36,679 件に分けて学習を行なった。

表 1 事前学習用と Few-shot 用のデータの内訳

年度	特許総数	テーマ数	F ターム数
2004 ~ 2014 (事前学習用)	479,952	349	2,664
2018(Few-shot 用)	141,093	264	2,209

また、Few-shot 用のデータの中から、正解ラベルに対して 100 件の特許、不正解ラベルに対して 100 件の特許を抽出した 200 件のデータ群を 1 つのトピック群としたデータセットを、以下のような 5 つの種類を作成している。

①~②ランダムカテゴリのデータセット：正解、不正解ラベルを F ターム、テーマにした、データセット。

③~⑤近接カテゴリのデータセット：正解、不正解ラベルを階層が同じの近接 F ターム、7 桁 F ターム、テーマにしたデータセット。2 つ目は、カテゴリごとの特許を使ったデータセットである。このデータセットに使用するデータは、商用の特許分類サービス PatentNoiseFilter(http://corp.ird-pat.com/pdf/Pamphlet_PNF.pdf) で、機械学習による分類器の構築の際に実際に使われたデータである。以下の 5 種類のデータセットを作成した。

⑥ PatentNoiseFilter データセット：PatentNoiseFilter データセットとは、発明通信社が提供する特許分類システム PatentNoiseFilter (<https://www.hatsumei.co.jp/patentnoisefilter/>) において、実際の業務で使用された小規模な特許分類用のデータセットである。以下の 3 種類のトピックが存在する。

- ・使い捨てマスク：正解ラベルとして使い捨てマスクに関する特許を 150 件、負例にそれ以外の特許を 150 件使用した 1 つのトピックに関するデータセットを使用する。
- ・釣具：リール、ルアー、釣竿、釣具その他に関する特許を集め、「①正例をリール、負例をそれ以外」、「②正例をリールとルアー、負例をそれ以外」、「③正例をリールとルアーと釣竿、負例をそれ以外」とした 3 種類のデータを、正例 100 件、負例 100 件をデータセットとして作成した。
- ・口につけるマスク：正例に口に付けるマスクに関する特許を 5 件、負例にそれ以外の特許を 5 件抽出し、

学習・評価・テスト(3:1:1)に分け、10通り(5C3)のデータセットを作成。

■実験手法

本研究ではFew-shotモデルの他に、比較手法として、BERTとSVM(Support Vector Machine)、ランダムフォレスト(以降RF)を分類器として使用する。BERTはTransformerによる双方向のエンコード表現を持つ大規模コーパスを事前学習データとした事前学習済みモデルで、文書分類などのタスクで高い精度を出しているモデルである。Few-shotモデルは、BERTを特許データで事前学習させたモデルである。また、SVMは機械学習モデルの一種で、データを直線に分割するデータであるサポートベクターと、直線の距離が最大となる直線を決定し、分類タスクなどを行うアルゴリズムである。RFは、木構造を用いて分類などを行う機械学習の手法である決定木を複数使うアンサンブル学習のアルゴリズムである。本研究では、SVMとRFで用いる文書は、tf-idfを用いて特徴量抽出を行っている。評価は精度、再現率、F1値により行う。各ラベルの閾値は0.5とする。

4.2 実験結果

実験結果を表2～表7に示す。ランダムFタームおよびランダムテーマについては、Few-shotが再現率、精度、F1値において最も高い結果となった。一方、近接データを使った実験では、近接7桁Fタームおよび近接Fタームにおいて1部、SVMを下回った。PatentNoiseFilterデータセットでは、Few-shotが再現率、精度、F1値において最も高い結果となった。

表2 ランダムFタームのデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.944	0.940	0.942
BERT	0.888	0.860	0.874
SVM	0.914	0.925	0.919
RF	0.913	0.910	0.911

表3 ランダムテーマのデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.932	0.924	0.932
BERT	0.851	0.824	0.851
SVM	0.898	0.899	0.898
RF	0.894	0.881	0.894

表4 近接Fタームのデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.693	0.680	0.686
BERT	0.601	0.592	0.596
SVM	0.668	0.688	0.678
RF	0.672	0.663	0.667

表5 近接7桁Fタームのデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.745	0.737	0.741
BERT	0.665	0.643	0.654
SVM	0.790	0.724	0.756
RF	0.763	0.726	0.744

表6 近接テーマのデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.908	0.905	0.906
BERT	0.781	0.749	0.765
SVM	0.908	0.900	0.904
RF	0.847	0.838	0.842

表7 PatentNoiseFilterデータセットの分類結果

手法	精度	再現率	F1値
Few-shot	0.787	0.792	0.783
BERT	0.655	0.715	0.672
SVM	0.758	0.680	0.660
RF	0.756	0.710	0.716

5 考察

実験結果全体について

表2,3のランダムカテゴリのデータセットの場合と表4,5,6の近接カテゴリのデータセットの場合の両方

と言えることは、Few-shotの方がBERTより全ての評価指標が高い結果となっている。これは特許データの事前学習によって特許に関する専門用語のベクトルを上手く獲得し、特許を分類する際に重要な役割を果たしていると考えられる。一方で事前学習がされていないBERTは、特許に関する専門用語のベクトルが上手く獲得できておらず、Few-shotと比べて分類精度等が落ちていると考えられる。

ランダムカテゴリのデータセットについて

表2.3より、Few-shotの方がSVMとRFよりも全ての評価指標が高い結果となっている。これは、Few-shotの特許データの事前学習による特許単語のベクトル獲得と、BERT特有の文章の文脈理解が精度向上に役立っていると考えられる。一方でSVMやRFの場合、単語ベクトル獲得のために、文脈を考慮しないtf-idfを使っているため、単語ベクトルが上手く獲得できておらず、分類精度がFew-shotと比べて落ちていると考えられる。

近接カテゴリのデータセットについて

表4.5,6より、近接テーマ、近接7桁のFターム、近接Fタームのデータセットの順で、精度等の数値が小さくなっている。これは特許分類の粒度が細くなるにつれて、カテゴリの区別が難しくなっていくからである。

表6より、Few-shotとSVM、RFの精度等が大きく変わらないがBERTのみ下がっている現象が起きている。SVMとRFで使用している、特徴量ベクトルを抽出するtf-idfは単語の語順を考慮していない。しかしテーマの分類の場合、例えば特許が機械翻訳かそうではないかという問題になるため、単語の語順を考慮せずとも、機械翻訳に関する単語があれば分類できる。そのため、SVMやRFでもテーマの場合、単語の語順を考慮しなくても精度等が高くなる。一方で、BERTの場合、単語の順序を考慮するが、パラメータ数が多いわりに精度等の数値が高くない。このことから、特許に関する事前学習をしない、そしてテーマによる特許分類の場合、単語の語順を考慮しない方が分類結果の向上が見込めることができる。ただ、7桁のFターム、9桁のFタームのように分類難易度が高まるにつれて、BERTとSVM、RFの精度等が変わらなくなってくる。これは、隣り合うFタームを分ける分類基準として、「ある成分

の配合が〇〇以上か未満か」などように分類の粒度が細かくなり、単語の語順が重要となる。そのため、分類難易度が高くなる場合、BERTのように単語の語順を考慮する必要があると言える。

カテゴリごとの特許を使ったデータセット

SVMは、近接Fターム、近接7桁Fターム、近接テーマでそれぞれ優れた分類性能が確認された。一方で、PatentNoiseFilterデータセットを使った結果において、トピックによって、極端に再現率が低下することが分かった。ユーザの立場からすると、安定性が重要。平均値が高くても、検索課題によって極端に悪くなる場合があると、システムを使用してもらえなくなる可能性がある。そこで以下に示すGMAPという評価尺度を用いてPatentNoiseFilterデータセットの実験結果を評価した。

$$GMAP = \exp\left(\frac{1}{n} \sum_{i=0}^n \ln x_i\right)$$

これはシステムの頑健性を測るための尺度で、検索課題ごとの再現率、精度、F1値の平均を、算術平均ではなく幾何平均で計算する。検索課題ごとの再現率、精度、F1値を掛け算するので、ひとつでも極端に低い値が含まれていると、GMAPの値が急激に下がる。

表8 PatentNoiseFilterデータセットの実験のGMAPによる評価結果

手法	精度	再現率	F1値
Few-shot	0.768	0.779	0.768
BERT	0.600	0.699	0.636
SVM	0.747	0.597	0.626
RF	0.742	0.694	0.705

評価結果

この結果から、Few-shotは表7に示す平均値だけでなく、頑健性の点でも他の手法より優れていることがわかる。

ランダムテーマとランダムFタームのデータセットにおける、Few-shotとSVMのラベル予測の誤りについて

誤りは大きく2つのパターンに分けられる。1つ目は不正解ラベルに対して正解ラベルを予測する時に多くある、「不正解を正解とした特許の内容と正解ラベルのFタームの特許の内容が似ているため、Few-shotが不

正解ラベルを正解ラベルと予測してしまっている」誤りである。2つ目は、正解ラベルに対して不正解ラベルを予測しているパターンに多くあり、「正解ラベルに関する単語などが少なく、正解ラベルなのに不正解ラベル予測をしている」誤り方である。

近接 F タームのデータセットにおいて、Few-shot のラベル予測の誤りについて

誤りの共通項として、特許の要約文の中に、属する F タームを特徴づける単語等が見られないことがあったことである。隣り合う F タームの場合、ある部分的な単語などだけが異なることが多く、その他の内容はほとんど似ている。そのため、属する F タームを特徴づける単語等がない場合、正しく分類することが難しくなる。実際に、正解ラベルを正解、不正解ラベルを不正解と予測している事例を見ると、特許の要約文の中には、属する F タームを特徴づける単語等が入っていることが多く見られた。このことから特許の要約文だけを使うのではなく、請求項やその他の情報を使うことによって、情報量を増やす必要があると考えられる。そうすることによって、要約文だけでは分類することが難しい特許でも、正しく分類できる可能性がある。

6 おわりに

本研究では、約 60 万の特許を使ってデータセットを作成し、作成データセットを基に Few-shot 学習を用いた特許分類器を構築した。実験の結果、F タームを基にした 5 つのデータセットの内、近接 F タームのデータセットの再現率と近接 F タームのデータセットの再現率と F1 値以外は、全て提案手法である Few-shot 学習を基にした特許分類器が最も良い結果となった。また、カテゴリごとのデータセットを使用した実験では、マクロ平均、GMAP においては Few-shot が最も良い結果となった。

参考文献

- [1] Pan, S. J. and Yang, Q.: A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering, Vol. 22, Issue 10, pp. 1345-1359, 2010.
- [2] Snell, J., Swersky, K., and Zemel, R.: Prototypical Networks for Few-shot Learning, Proceedings of the 31st Conference on Neural Information Processing Systems, 2017.
- [3] Chen, W., Liu, Y., Kira Z., Wand, Y. F., Huang, J.: A Closer Look at Few-shot Classification, Proceedings of the 7th International Conference on Learning Representations, 2019.
- [4] Finn, C., Abbeel, P., and Levine, S.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, Proceedings of the 34th International Conference on Machine Learning, pp. 1126-1135, 2017.
- [5] Li, X., Chen, H., Zhang, Z., and Li, J.: Automatic Patent Classification using Citation Network Information: An Experimental Study in Nanotechnology, Proceedings of the 7th ACM/IEEE- CS Joint Conference on Digital Libraries, pp 419-427, 2007.
- [6] 富永 泰規, 久々宇 篤志: 特許文献への分類付与における付与根拠箇所推定, 情報の科学と技術, 68 巻・7 号, pp. 338-342, 2018.
- [7] Yu, M., Guo, X., Yi, J., Chang, S., Potdar, S., Cheng, Y., Tesauro, G., Wang, H., and Zhuo B.: Diverse Few-Shot Text Classification with Multiple Metrics, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1206-1215, 2018.
- [8] Brown T. et al.: Language Models are Few-Shot Learners, Proceedings of the 34th Conference on Neural Information Processing Systems, 2020.