

# 階層構造に対応した Attention 付ニューラルネットワークによる特許文書の FI 予測

東京工業大学	*星野雄毅	HOSHINO Yuki
楽天株式会社	内海祥雅	UTSUMI Yoshimasa
楽天株式会社	松田義郎	MATSUDA Yoshiro
楽天株式会社	齋藤歩美	SAITO Ayumi
東京工業大学	田中義敏	TANAKA Yoshitoshi
01405430 東京工業大学	中田和秀	NAKATA Kazuhide

## 1. はじめに

現在、特許は毎年 30 万件をこえる出願があり、それらをすべて人手で処理している。そのため、特許の権利化までには多くの時間がかかっており、平均 14.3ヶ月となっている [1]。この期間は特許として認められないことから、審査時間の短縮は課題とされている。

特許文書には FI (File Index) という内容に基づく分類が与えられる。FI は国際特許分類 (IPC) をさらに細分化された形で与えられる日本独自の分類方法で、現在約 19 万種類存在する。FI は大きく 2 つの特徴があり、一つ目としては階層構造を持っていることがあげられる。図 1 のようにはじめはセクションと呼ばれる A の「生活必需品」のような大まかな分類から徐々に細分化されていき、最終的には A41D1/00 101 J の「和服の製造」のような細かい分類にまで分けられる。二つ目の特徴として、FI は一つの特許に複数与えられることがあげられる。このように階層構造を持っていることとマルチラベルであるということが一般的な分類問題と異なる。現在 FI はすべて審査官が付与していることから、これらを自動化ないしは半自動化できると権利化までの時間短縮につながると考えられる。

## 2. 既存研究

特許文書に対する FI 予測というタスクに関する既存研究として、階層マルチラベルニューラルネットワーク (Hierarchical Multi-Label Neural Network: 以下 HMLN) [2] が提案されている。しかし HMLN は文章をそのまま入力としており、コンピュータのメモリの都合上、特許文章全体を一括して処理することができない。よって、請求項ごと

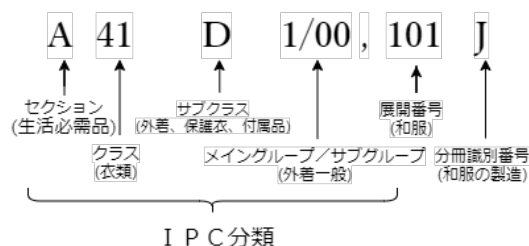


図 1: FI の例

に分割して処理を行っている。このため、請求項を横断した学習をすることが不可能となる。さらに、データ数が増えるため大量の特許文章による学習が難しくなることが加わり、精度が落ちるといった問題が生じる。また、予測対象分野ごとに特有の処理を行わないため、精度が低くなるという問題もある。本研究では、これらの問題を解決することで、精度の高い FI 予測モデルを構築する。

## 3. 提案モデル

今回提案するモデルでは、HMLN に対して次の 4 点の改良を行った。

### 3.1. 名詞割合入力

2 節で述べたように、HMLN では請求項ごとに分割して別々に学習するため、精度が落ちるといった問題があった。そこで、今回のモデルでは名詞のみが FI に影響を与えているという仮定のもと、文書を分かち書きした後、出てきた名詞とその割合を入力とした。これにより一つの特許を一つのデータとして学習することが可能になり、精度の向上が期待される。

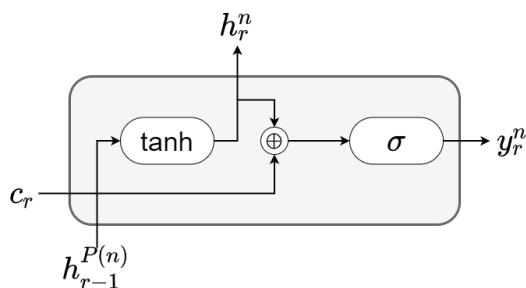


図 2: Attention 付き分割処理モデルのユニット

### 3.2. 割合つき Set Transformer

前述したとおり今回のモデルの入力は抽出してきた名詞の集合とそれらの出現割合である。今回エンコーダは集合情報の入力処理できる Set Transformer[3] の入力部分を変え、元の特徴量にあらかじめ割合を掛けたものを入力とした。これにより、文章にその単語が含まれていたかどうかだけでなく、文章においてどの単語が中心的役割を果たしているのかを自然に特徴量に組み込むことができるようになった。

### 3.3. Attention 付き分割処理モデル

HMLN のデコーダはすべての分野に対して同一の処理を行っていたため、分野の範囲が広がると学習の質が落ちてしまう可能性があった。そこで、分野ごとに異なる階層的な処理を行えるように、予測対象 FI に合わせた階層構造を持つようにデコーダを作成した。しかし Attention 機構については計算時間がかかり、すべてのユニットに導入するとメモリ不足や計算時間が問題になるため、各階層において上位階層の全出力  $y_r$  を受けて、以下の式のように定めた  $q_r$  をクエリとしてグローバルに導入した。

$$q_r = W_r y_{r-1} + b_r$$

モデルのユニット構造は図 2 のように、親ノードの出力  $h_{r-1}^{P(n)}$  と Attention による文脈ベクトル  $c_r$  を入力とし 2 つの出力  $h_r^n$  と  $y_r^n$  を得る。ここで、 $h_r^n$  は子ノードに引き継がれ、 $y_r^n$  はモデル全体の出力となる。また、 $h_r^n$  の次元は、対象ノード以下の FI 数に応じてそれぞれ定めた。このように、各ノード固有の特徴量を  $h$  で与え、前の階層の出力をもとにしたグローバルな文脈ベクトルを  $c$  で与えることで、予測精度の向上させた。

### 3.4. 損失関数

HMLN の損失関数のポイントとなっていたのは階層損失であった。これは「A41」が付与される確率が「A」が付与される確率よりも大きい時に与えられる損失である。しかし、階層損失は階層構造を加味できる一方で、「A41」が実際に付与されたとしても付与しないように学習してしまうという問題点があった。これに対応するために、損失関数を以下のように定めた。

$$L = t_M^k \sum_{i=1}^M \log(y_i^{K_{i,k}}) + (1 - t_M^k) \sum_{i=1}^M \log(1 - y_i^{K_{i,k}})$$

これは各層の出力を各ノードにおいて親ノードが付与されていたと仮定したときの子ノードが付与される確率とみなしたことに相当する。これにより、階層損失を入れずに階層構造を加味できるため、学習がスムーズに進むことが期待される。

そのほか、過学習を抑制するためエンコーダ、デコーダの入力層と出力層に dropout[4] を導入した。

## 4. 数値実験

数値実験として 250531 件の実際の特許文章を用いた FI 予測を行った。実験結果については発表会にて報告する。

### 参考文献

- [1] 特許庁. 特許行政年次報告書 2020 年版. サンワ, 2020.
- [2] 市瀬将也, 今村剛, 中田和秀, 田中善敏. 階層マルチラベルニューラルネットワークを用いた特許文書における FI の予測, 日本オペレーションズ・リサーチ学会 2019 年春季研究発表会, 2-F-4, 2019.
- [3] Lee, J., Lee, Y., Kim, J., Kosiosek, A., Choi, S., and Teh, Y.W.. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, 3744–3753, 2019.
- [4] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, No. 56, 1929–1958, 2014.