

知的財産法を論理式に自動変換するための TF-IDF を用いた特許法の特徴分析

富田 晟央[†] 加納 徹^{††} 赤倉 貴子^{††}[†] 東京理科大学大学院工学研究科情報工学専攻 ^{††} 東京理科大学工学部情報工学科

1. はじめに

情報通信技術の発展に伴い、知的財産と知的財産教育の重要性が高まっている。しかし、大学において、特に工学系学部では、その教育が十分ではない[1]という背景から、工学系の学生を対象とした学習システムに関する研究が行われている。小松らは、特許法について、法令文の論理構造をもとに論理回路を構築し、インタフェースに反映した学習システムを作成した[2]。このシステムでは、問題作成過程において、手作業で条文を論理回路に変換しなければならないことが課題とされていた。著者らは、この課題に着目し、知的財産法を自動で論理回路に変換する手法の提案を目指す。

2. 先行研究と目的

著者らはこれまで、特許法を対象とする形態素解析・係り受け解析の結果を用いて、要件効果構造[3]に基づく条文の構成素識別と、それらの分割をパターンマッチで行った。識別ルールを表1に示す。

表1のルールに基づく構成素の識別は、対象とした34条文101項のうち、約半数を意図通りに行うことができたものの、識別パターンの定義が不十分であり、精度に課題があった。そこで、本稿では、新たな識別パターンを抽出するために、TF-IDF (Term Frequency-Inverse Document Frequency) による特許法の分析結果について有用性を検討する。

表1: 構成素識別パターンと適合数 (一部抜粋)

構成素	識別パターン	適合数
主題部	普通名詞+(と)は, 普通名詞+が,	42
条件部	~場合は, ~ときは, ~において など	73
内容部	~を	117
対象部	普通名詞+(に(対し),	5
規定部	~とする, ~する, ~できる など	125

3. 分析手法・結果

TF-IDF とは、単語の重要性を定量的に評価する手法であり、以下の式で表される。

$$tfidf_{t,d} = tf(t_i, d_j) \times idf(t_i)$$

$$tf(t_i, d_j) = \frac{f(t_i, d_j)}{\sum_{t_k \in d_j} f(t_k, d_j)}, idf(t_i) = \log\left(\frac{N}{dt(t_i) + 1}\right)$$

$tf(t_i, d_j)$ は、 f を出現回数関数として、文書 d_j 内の単語 t_i の出現回数を、文書 d_j 内の全ての単語の出現回数で割った商である。 $idf(t_i)$ は、全文書数を単語 t_i が出現する文書数で割った商の対数である。このとき、分母がゼ

ロになることを防ぐために1を加算する。

$tfidf$ は、ある文書内での出現回数は多いが、他の文書には出現しない単語ほど大きな値となる。本稿では、附則と括弧書きを除く特許法の条文全1,114文書を文書集合としてTF-IDFを適用した。特許法第2条第2項『この法律で「特許発明」とは、特許を受けている発明をいう。』の各単語のTF-IDFの値を表2に示す。

表2より、TF-IDFの値が最も高い単語は「いう」である。「いう」は、全文書中において6回しか出現せず、「特許発明」のような用語の定義において「Aとは、Bをいう。」という形式で使用されるとわかった。一方、「は」や「を」のように、多くの文書に出現する単語は、TF-IDFの値が低くなる傾向があるとわかった。

表2: 特許法第2条第2項の分析結果

単語	TF-IDF	単語	TF-IDF
いう	0.4589	特許	0.2017
受けて	0.4498	で	0.1919
発明	0.4173	を	0.1785
法律	0.3307	と	0.1662
いる	0.2994	は	0.1026
この	0.2577		

4. 考察

結果より、TF-IDFは特許法において用いられる特徴的な単語を明らかにできることが示唆された。したがって、この結果は、コンピュータに定義文を判定させるアルゴリズムに利用できると考えられる。また、TF-IDFの値が低い単語は、汎用的な単語として、パターンマッチの識別規則に応用できると考えられる。

5. まとめ

本稿では、知的財産法の学習支援システムに組み込む、条文の論理回路を自動作成するために、条文の特徴分析を行った。TF-IDFに基づく分析結果から、条文中の固有表現と汎用的表現を、定量的に判別できることが示された。今後の方針として、今回の結果を構成素の識別規則に組み込み、精度の向上を図ることがあげられる。

参考文献

- [1] 赤倉貴子, 中村修也, 加藤浩一郎, “国立大学工学系学部における知的財産教育の現状分析,” 日本教育工学会論文誌, vol.42(Suppl.), pp.81-84, 2018.
- [2] 小松昌史, 加納徹, 赤倉貴子, “知的財産法の問題演習におけるError-based Simulationを用いた学習支援システムの開発と評価,” 信学技報, ET2019-64, pp.33-36, 2019.
- [3] 田中規久雄, “法律効果規定部の意味機能について,” 情報処理学会研究報告自然言語処理, vol.124, no.21, pp.1-8, 1998.