

# 非線形 SVM 用いた医薬特許 パテントビリティ予測モデルの構築

宮岡 慧<sup>†</sup>赤倉 貴子<sup>††</sup><sup>†</sup> 東京理科大学工学部経営工学科<sup>††</sup> 東京理科大学工学部情報工学科

## 1. はじめに

日本国の特許法は先願主義である。そのため、重要特許は他者より早く出願する必要がある。特に、医薬特許は莫大な研究開発期間と開発費がかかるため、慎重かつ素早く出願する必要がある[1]。正確な明細書を出すために、パテントビリティ予測モデルの研究が行われている[2]。本研究では、非線形 SVM(Support Vector Machine)[3]を用いてパテントビリティ予測モデルを構築し、精度向上を目的とする。

## 2. パテントビリティ

本研究では、パテントビリティと呼ぶ基準を扱う[2]。これは特許出願に対して最終的に特許処分され、特許が成立する確率である。このパテントビリティによって、特許出願が新規性、進歩性、明細書の記載要件を満たしているかを判断することができる。

## 3. モデル構築

本研究では、医薬特許に関する公開公報 5,326 件とその審査経過データを扱う。このデータから、明細書統計量とパテントビリティの正解ラベルを抽出し、学習データとする。表 1 に、明細書統計量を構成する特徴量を示す[2][4]。また、パテントビリティの正解ラベルは、特許成立を 1、特許不成立を-1とする。

モデル構築の概要を図 1 に示す。学習データを訓練データとテストデータに分割し、SVM に入力する。ここで、RBF(Radial Basis Function)カーネルを用いて非線形識別を行う。

表1 明細書統計量

明細書全体の文字数	請求項の数	請求項全体の文字数
請求項1の文字数	図表の枚数	IPC分類の数
発明者の数	発明の詳細な文字数	先行技術文献の数
薬理試験の実施回数	国内優先権の主張の有無	パリ条約優先権の主張の有無

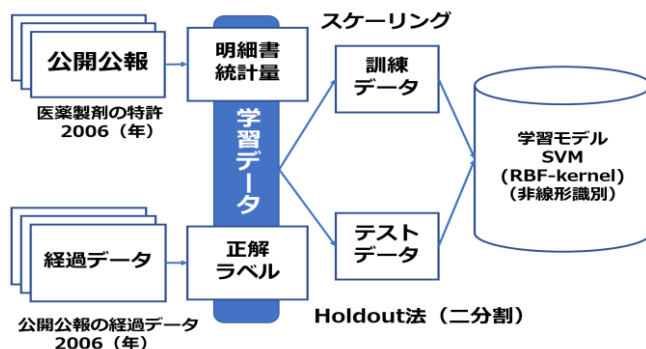


図1. モデル構築

## 5. 評価指標

モデル精度は AUC(Area Under the Curve)で評価する。この値は 0~1 までの値をとり、1に近いほど高精度であり、0.5に近いほど低精度である。

## 6. 実験結果

SVM を用いた実験結果を図 2 に示す。実験結果から、SVM モデルの AUC の値は 0.702 となった。これにより、Hido ら[2]の 0.594 と比較すると 0.108 向上したことが分かる。また、医薬特許に関するロジスティック回帰モデルの 0.566 と比較すると、SVM の AUC の値は 0.136 向上した。

以上より、医薬特許に関するパテントビリティ予測モデルでは、非線形識別が可能である SVM が有効であるといえる。

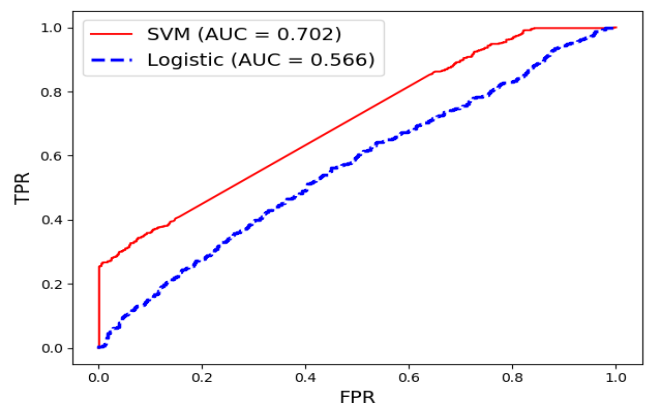


図 2. 実験結果

## 7. まとめと今後の課題

本研究では、医薬特許に関するパテントビリティ予測モデルの構築を行った。非線形 SVM モデルを導入することで、Hido ら[2]の予測モデルより 0.108 精度が向上した。

今後の課題として、未知データを用いてモデルを検証すること、ランダムフォレストなどの分析手法を検討することが必要となる。

## 参考文献

- [1] 當麻 博文, “医薬特許関連特許とジェネリックに医薬について,” 高島国際特許事務所, [http://www.phs.osaka-u.ac.jp/homepage/yaku/sotugo/pdf/h22\\_04\\_0/pdf](http://www.phs.osaka-u.ac.jp/homepage/yaku/sotugo/pdf/h22_04_0/pdf), 参照 2019 年 1 月 30 日
- [2] S. Hido, S. Suzuki, and R. Nishiyama, et al., “Modeling Patent Quality: A System for Large-scale Patentability Analysis using Text Mining,” *Journal of Information Processing*, vol. 20, no. 3, pp. 1180-1191, 2012.
- [3] V. N. Vapnik, “Statistical Learning Theory,” John Wiley & Sons, New York, 1998.
- [4] 永田健太郎, 渡部俊也, “日本特許の質に関する実証分析,” 日本知財学会第 6 回年次学術研究発表会, pp. 326-329, 2008.