

傾向スコアを用いた除籍資料予測の試み

橋本郷史

東邦大学医学メディアセンター

【目的】 除籍のための資料選定には多くの労力がかかる。除籍すべき資料をある程度機械的に選定する方法があれば業務上の負荷が軽減できる。発表者は 2021 年の JMLA の学術集会の発表で、蔵書データに含まれる資料の要素情報から、除籍のされやすさが分かることを報告した。例えば、ある資料が改訂版であれば改訂版でない資料と比較して、同期間内における除籍の可能性が X 倍になる、といった具合である。一方で、実際に各資料が除籍の対象となるかを蔵書データから予測できるかどうかについては不明であったため、今回の研究ではその点について確認を行った。

【方法】 分析対象は、2011 年度から 2020 年度に本学の医学部図書館で購入した資料の内、2011 年以降に出版された図書資料とした。2020 年度末までの除籍実績でロジスティック回帰分析を行い、資料の持つ要素が除籍にどのように影響するかを数値化し、各資料の除籍されやすさ(傾向スコア)を算出した。続いて、2020 年度末までの除籍とこのスコアとの関係を ROC 曲線によって確認し、除籍判断のためのカットオフ値を決めた。最後に、このカットオフ値を使って 2021 年度の除籍を予測した場合の的中率を確認した。ロジスティック回帰分析は除籍をアウトカムとし、説明変数には、資料の登録から 2021 年 3 月 31 日まであるいは除籍されるまでの日数(生存時間)に加え、前回の研究の結果を考慮し次の要素を用いた。「改訂版である」「複本である」「本学教職員の著作である」「複数人による著作である」「NDC491, 496, 499 番台の資料である(それぞれ)」。

【結果】 分析の対象は 14,442 件で、この内 2020 年度末までに除籍された資料は 1,611 件、2021 年度に除籍された資料は 524 件だった。前述のモデルで多変量のロジスティック回帰分析を行った結果は左下の表のようになった。ここから計算した傾向スコアと 2020 年度末までの除籍との関係を ROC 曲線で表現したものが右下の図である。曲線下面積は 0.741(95%CI 0.729-0.754)となった。感度と特異度の和が最大となる傾向スコア 0.124 をカットオフ値として 2021 年度の除籍を予測した場合、この値以下となる資料は分析対象の約 7 割で、その陰性的中率は 98%だった。また、この値より大きい場合の陽性的中率は 9%だった。

この結果をどのように解釈すればよいのかについて、また、この手法の限界と応用可能性について、講演内で報告する。

要素	オッズ比 (95%信頼区間)	
(切片)	0.05 (0.04-0.06)	
生存時間(1年)	0.96 (0.94-0.98)	
改訂版である	4.05 (3.62-4.53)	
複本である	2.65 (2.32-3.03)	
東邦教職員著作	0.58 (0.46-0.72)	
複数人著者	1.54 (1.28-1.84)	
分類	491(基礎医学)	0.98 (0.82-1.17)
	496(眼科・耳鼻)	0.68 (0.48-0.98)
	499(薬学)	3.41 (2.31-5.03)

