

誰が病院の悪評を立てるのか-決定木を用いた病院口コミデータの分析-

Who spread a bad reputation on a hospital?

中井 豊

Yutaka NAKAI

関西大学・ソシオネットワーク戦略研究機構

要旨: 悪評の流布は中長期に見て病院経営における重要なリスク要因とされる。本研究は、代表的な病院の口コミサイト「病院なび」を対象にして、主に精神科を運営する全国の医療機関に対する口コミデータをスクレーピングし、「誰が低い評価点を付けるのか」を、機械学習を用いて探索した。具体的には、5点満点中3点あるいは4点を悪評とみなし、家族・本人、年齢、性別の属性を説明因子として、決定木を構成した。

キーワード: 精神科, 悪評, 決定木

Abstract: The spread of bad reviews is considered an important risk factor in hospital management from a medium- to long-term perspective. This study scrapes reputation data on a typical hospital review site, Hospital Navi, for nationwide medical institutions that operate mainly in psychiatry, and explored "who gives low evaluation scores" using machine learning. Specifically, a score of 3 or 4 out of 5 was considered a bad review, and decision trees were constructed using the attributes of family/person, age, and gender as explanatory factors.

Keywords: psychiatry, bad reputation, decision tree

1. 本研究の目的

医療機関に対する悪評の流布は中長期に見て病院経営における重要なリスク要因とされる(折戸 2017, 和田 2001, 井上・富田 2000)。そして、「誰が」「何を気にして」悪評を立てるのかを理解することがリスクマネジメントの基本となる。そこで、本研究では、医療機関に対する口コミサイトのオープンデータ、特に医療機関に対する利用者の総合コメント(テキストデータ)に注目し、機械学習を用いて、「誰が」「何を気にして」悪評を立てるのかを明らかにすることを目的とする。本報告は、「誰が」に関する分析結果を中間報告する。

2. 医療機関の評判情報のスクレーピング

医療機関に対する代表的な口コミサイトである「病院なび」のオープンデータ(全国 13216 施設 @2022.10.1-10.5)の内、診療科リストのトップに精神科を掲げる医療機関のデータ(評価スコア(5点満点), 年齢, 性別, 本人・家族別, 総合コメント文)をスクレーピングし, 280 レコードを収集した。280 レコード中, 5 点満点が 212 (75.7%), 4 点が 49 (17.5%), 3 点が 19 レコード (3.2%) であり, 1 点と 2 点評価はなかった。本研究では, 3 点と 4 点評価を悪評と見做した。

3. 決定木による分析

評価スコアを被説明変数, 3 種の利用者属性(年齢, 年齢, 性別, 本人・家族別)を説明変数として, 決定木を用いて, 悪評が立つケースを調べた。

データのダミー化については、評価スコアに対して悪評:1・普通:0 とし、利用者属性に対して男:1・女:0, 本人:1・家族:0 とした。年齢は、10代から70代までの7水準である。また、データを訓練データ（70%）とテストデータ（30%）に分割した。

決定木のアルゴリズムはGini不純度の最小化とし、決定後の葉の数とテストデータ作成のシード値の2つをハイパーパラメーターとし、テストデータに対する正解率とF値の両者の最大化を実現する決定木を探索した（Optunaを利用）。探索するハイパーパラメータの空間は、葉の数2から10の9ケース、シード値150ケースで、計1350ケースである。また、葉の最小サイズを6と固定し、探索回数を2000回とした。正解率とF値の2値に注目するのは、悪評の数が少なく正解率だけでは無意味な解（全て普通と判定する解）が析出する可能性があることから、これを防ぐためF値を加えた（F値は混同行列の再現率と適合率の調和平均であり、学習器のバランスの良さを示す）。従って、パラメーターの探索は多目的最適化に当たっており、パレートフロントに位置するパラメータを最適解と解釈する。

4. 分析結果—析出した決定木の構造と性能

ハイパーパラメーターの探索結果を図1に、析出した決定木を図2に、決定木の性能と変数の重要度を図3に示す。パレートフロントに位置する3ケースが最適解の候補となるが、無意味な解でないこと、また同時に、混同行列（図3）が有意（後述）であることから、葉の数が7、テストデータ作成の乱数シード値が86のケースが最適解と判断した。一般的に、葉の数を増やすほど学習データに対する精度が上がるが、過学習によりテストデータに対する精度が落ちてゆく。葉の数を10まで探索した結果、パレートフロントでの葉の数が10未満の7となったことから、過学習に陥っていないと思われる。

trial_no	テスト_F値	テスト_正解率	葉の数	Test_data_シード数
1999	0.414	0.798	4	58
1945	0.478	0.714	7	164
1977	0.491	0.679	7	86

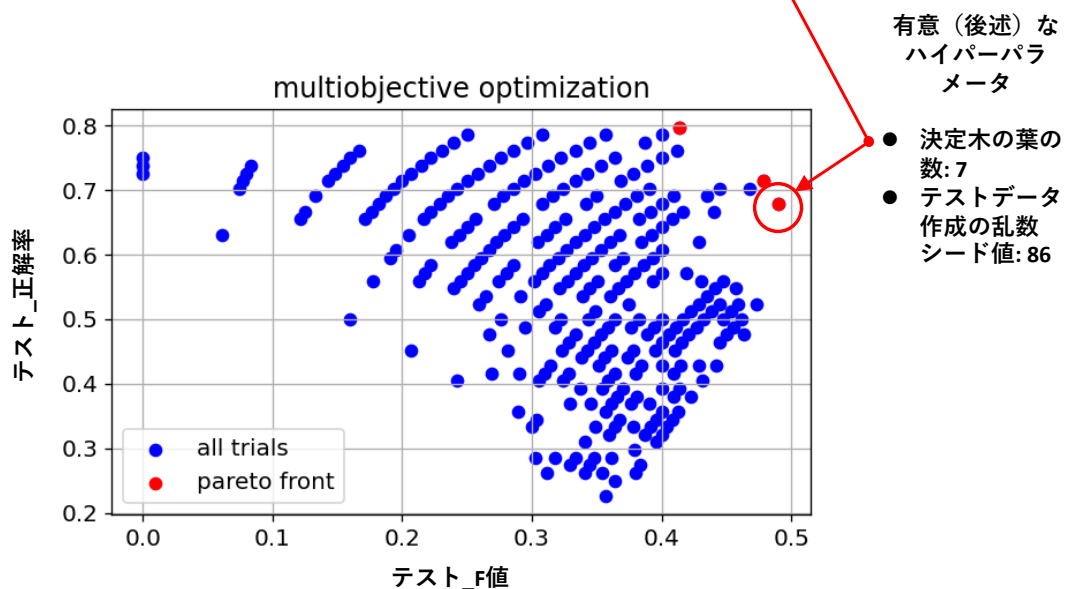


図1. ハイパーパラメーターの探索結果

図3によれば、決定木のテストデータに対する正解率が68%、再現率が65%であるが、適合率が39%と低くなっている。また、混同行列は、危険率5%の χ 二乗検定で有意となった。析出した決定木の適合率は低いですが、悪評に注目する本研究では再現率の高さが重要であるため、本決定木は相応の性能を有していると考えます。

ただし、本研究は、「誰が（どんな人）」が悪評を立てるのかを明らかにすることを目的としているため、決定木の子測性能より変数の重要度の方が重要であることに注意しよう。図3によれば、年齢、性別、本人・家族のいずれもが悪評判定に貢献することが分かった。

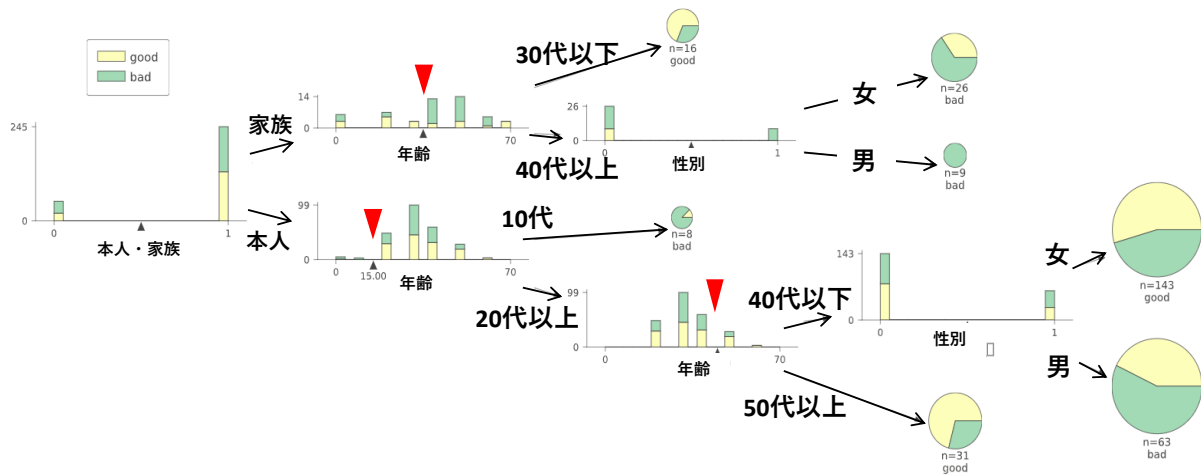


図2 決定木の構造（葉の数：7）

訓練_正解率	テスト_正解率	テスト_適合率	テスト_再現率	テスト_F値
0.608	0.679	0.394	0.650	0.491
混同行列	カイ二乗値	P値	自由度	
下記参照	5.931	0.015	1	

混同行列は、危険率5%で有意

テストデータの混同行列		予測値	
		普通	悪評
真の値	普通	44	20
	悪評	7	13

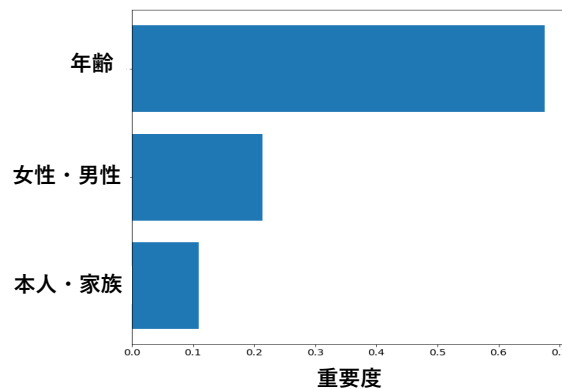


図3 決定木の性能（混同行列）と因子の重要度

5. 分析結果—誰が悪評を立てるのか

この決定木から、以下の属性の利用者が悪評を立てる可能性があることが示唆された。

- A) 10代の患者本人
- B) 20代から40代で、どちらかと言えば男性の患者本人
- C) 40代以上で、特に男性の患者家族

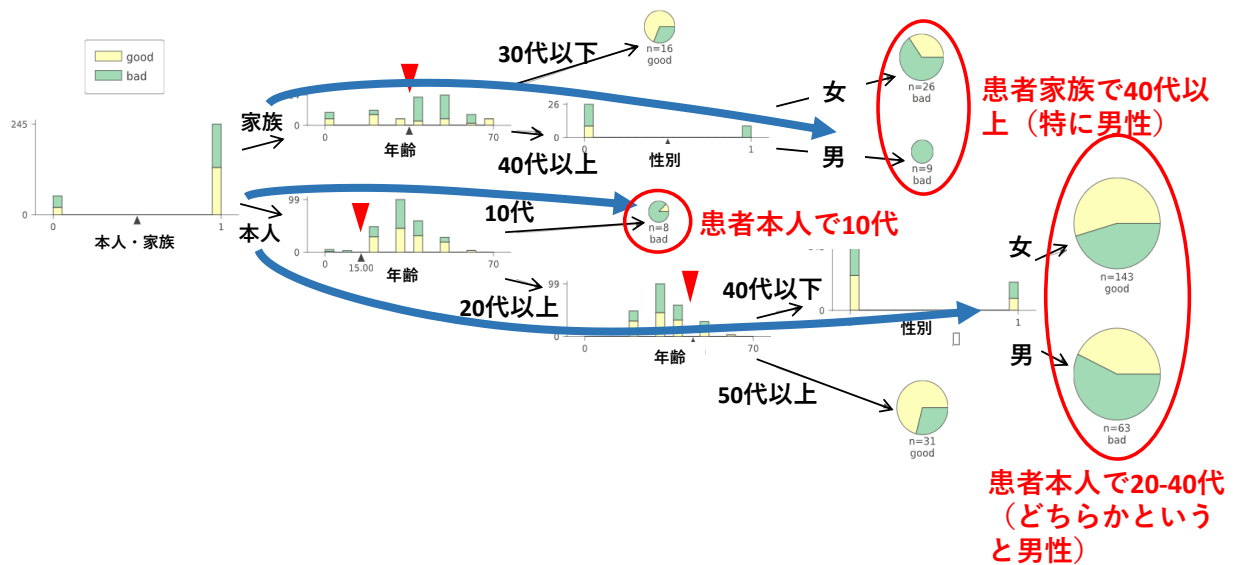


図4 誰が悪評を立てるのか

6. まとめと今後

上述の B) のケースは最も該当数が多いにも関わらず、悪評と普通の分離が明確でなく、属性だけによる決定木の精度の限界が現れており、よりリアルな（そして予測精度を上げる）決定木を構成するには、属性以外の説明変数が必要である。本研究では、利用者毎に評価コメントを取得しているの、個々人が何を気にしているのか、トピック分析を行うことが可能である。今後、誰が「何を気にして」悪評を立てるのかを検討してゆく。

文 献

- 折戸洋子, 2017, 「医師および患者によるクチコミ情報の発信・共有と共感型コミュニティ形成」『日本情報経営学会誌』37巻2号, P64-80.
- 和田ちひろ, 2001, 「離反患者の原因分析とその影響」『医療マネジメント学会雑誌』2巻2号, P153-157.
- 井上淳子・富田健司, 2000, 「医療機関における患者との関係性構築 リレーションシップ・マーケティング概念を用いて」『医療と社会』10巻3号, P97-108.