

Web データに基づく医療用医薬品の有用性評価

石川 祥子[†] 加藤 誠[†]

[†] 筑波大学 〒305-8550 茨城県つくば市春日 1-2
E-mail: †s2121604@s.tsukuba.ac.jp, ††mpkato@acm.org

あらまし 本論文では、Web 上で公開されている複数の指標に基づいて、機械学習によって医療用医薬品の有用性を判定する手法を提案する。提案手法では、医療用医薬品の情報から有効性、安全性、経済性などの指標を抽出し、これらを教師あり学習モデルにおける特徴量とした。実験では、専門家によって付与された有用性を Web 上で公開されている指標から判定するタスクに取り組んだ。有用性は複数の医療機関が公開している情報を収集し、特徴量は公的医療情報データベースから取得した。実験結果に基づいて有用性を判定するための手法を検証し今後の課題を検討した。

キーワード 医療・ヘルスケア、評価・データセット、情報抽出、医療経済、医薬品

1 はじめに

最近では、日本の医療費で大きな割合を占める医薬品に関して、効果に対する経済性が重視され始めている。すでに日本と類似した国民皆保険制度を維持する国では、医療費の効率的運用のために、患者の臨床情報に加え医薬品情報も活用している。例えば、イギリスでは NICE: National Institute for Health and Care Excellence が医薬品情報を収集し、医薬品や治療法に対して、効果と経済的価値の観点から、保険適用の可否を判断している [1] [2] [3]。

しかし、医療費の効率的運用に資する医療経済性の判定には複数のハードルがある。まず、大量の臨床情報や医薬品情報を解析する必要があるが、統計データが政府や民間企業に散在しており、集約したデータセットでの解析が困難である。しかも情報量が多く判断自体に専門領域の知識が必要なため、ヒューリスティックな解析となり、時間・労力がかかる。また医療費支出という公共性の高さから、有用性の判断に公平性が欠かせない [4]。

そのため、本論文では医療用医薬品において、医療経済的視点での有用性を判定する手法を提案する。本提案では、保険対象となる医療用医薬品の有用性判定を行う。医薬品情報は届出義務があり、PMDA (独立行政法人医薬品医療機器総合機構) で公開されている添付文書情報を用いる [5]。有用性の付与については、日本フォーミュラ学会による有用性が高い医薬品のリストを参考にし、有用性自体の判断に関しては、同じく日本フォーミュラ学会による医療経済性評価の定義する項目を参考にした [4]。その項目は安全性・有効性・経済性で構成され、それぞれ添付文書で対応する記載項目として副作用数・臨床試験数・薬価を設定した。判定方法は機械学習における教師あり学習を使用し、実験によってモデル間での性能を比較した。実験結果から二値分類での有用性判定は困難であることが示された。

本論文では、第 2 節で関連研究について述べる。第 3 節では提案手法と有用性判定のための指標やデータセット作成および

学習方法について説明する。第 4 節では、実験の詳細と結果、第 5 節では、本論文の結論とともに今後の検討課題について説明する。

2 関連研究

関連研究は、医療・ヘルスケア情報の種類別に分けられ、画像・音声・テキスト・数値と多くのデータ種類がある。またその情報は公開情報 (論文・添付文書・教科書・ガイドラインなど)、臨床情報 (カルテ・薬歴・健康診断結果・遺伝情報など)、患者による情報 (Web 上や SNS に公開された日記・口コミなど) に大別される [6]。

公開情報については、主に医学・薬学研究者と共同で行った関連研究が存在する。例えば、添付文書から深層学習で特定項目を情報抽出する研究などがある [7]。また、添付文書や他の医学情報データベースから情報量が少ない小児に関する医薬品情報データベースを構築する研究が進められている [8]。これらはいずれも医学的には重要な公開情報ではあるものの、データ形式が標準化されていない添付文書から情報抽出を図り、データベース化やデータセット構築を目指している。

臨床情報においては、電子カルテなどある程度の構造化・標準化がなされており、医療者による記録であることから、機械学習手法、特に自然言語処理による関連研究が多く存在する [9] [10] [11] [12]。また画像診断に関しては、教師データが付与されたデータセットが公開され、臨床現場ですでに実用化されている [13]。また罹患情報などプライバシー保護が重要な情報では分散型機械学習などの研究がある [14]。患者による情報は、非文法的で表現ゆれが多いことから、大規模なデータセット構築が途上である。

本論文では、公開情報である添付文書からデータ収集し、有用性の判定を提案している。添付文書を扱った関連研究では、記述やフォーマットの混在した添付文書から標準化や構造化可能な情報抽出をタスクとしていることが多い。本提案では、添付文書情報から構成された医療経済的有用性という新たな指標を判定するタスクを設定しており、問題設定が異なっている。

3 提案手法

本節では提案である有用性を判定する手法について説明する。この有用性判定は、教師あり機械学習によって行う。3.1節で有用性を判定するための指標について説明する。3.2節で対象としたデータと3.1節で示した指標に対応する特徴量について述べる。3.3節では、有用性判定に使用する教師あり機械学習を検討した。

3.1 有用性判定のための指標

本研究において、有用性の定義は日本フォーミュラリ学会が示すフォーミュラリ選定指針を参考にした[4]。フォーミュラリとは、患者に対してEBM (Evidence-Based Medicine: 根拠に基づく医療) をもとに有効性、安全性、経済性などの観点から推奨される医薬品リストである。このフォーミュラリは医療経済的な視点から費用対効果も考慮される。現在、日本では全国統一基準としてのフォーミュラリは存在せず、先進的な医療圏の医療関係者によって作成された地域フォーミュラリや、高度医療を担う大学病院などで作成・公開された医療機関フォーミュラリが散発的に存在している。

本提案では、主要なフォーミュラリ選定基準である有効性、安全性、経済性に対して、以下を指標として設定した。有効性は、添付文書に記載されている臨床試験の数とした。臨床試験は医薬品に対して最初に発売される先発品の承認過程で実施され、有意差が示された試験が記載されている。そのため実施回数が多いほど有効性および安全性が担保されているとした。また安全性は有害性と置き換え、添付文書記載の副作用の数が少ないほど有害事象がなく安全性に優れているとした。経済性に関しては費用負担性で代替し、薬価(製剤の価格)が安価であるほど経済性に優れ有用性が高いとした。

3.2 データセット

本節では判定対象としたデータについて、入手方法と実験用データセットの作成方法を述べる。本提案では有用性判定の学習データとして、日本フォーミュラリ学会によるモデルフォーミュラリを使用した。またテストデータとして、複数の医療機関によって公開された医療機関フォーミュラリを利用した。これらのフォーミュラリは、疾病に対する効果ごとに「薬効分類」としてグループ化されており、薬効分類には推奨される医薬品が示されている。医薬品は日本薬局方で規定された医薬品名で記載され、慣習的に一般名と呼ばれることが多い。それら医薬品が各社で製品化・承認販売されたものが製剤となり、別途製剤名が付与される。各製剤には、承認販売時に添付文書の届出・公開義務がある。正例はフォーミュラリで示された薬効分類の中で推奨されている医薬品(一般名)集合とし、推奨する順番として第1位から第3位程度まで選択順位が示されている。逆に同じ薬効分類でもフォーミュラリで推奨・記載されていない医薬品集合は負例として収集した。その後、フォーミュラリに記載のある製剤については、PMDAで公開されている添付文書の全文HTMLをクローリングし、記載内容から臨床試験数と副

作用数を収集した[5]。また別途、経済性の指標として、厚生労働省による薬価基準収載品目リストから、フォーミュラリに存在する製剤の薬価を抽出した[15]。薬価の抽出方法は、クローリングした添付文書の全文HTMLから薬価基準収載品目リストで使われるコードを収集し、対応する薬価を探索した。構築したデータセット項目は以下に示す。

本提案で評価対象としたのは、医薬品集合: $M = \{m_1, m_2, \dots, m_n\}$ であり、それらには、本実験で対象とした15個の薬効分類: $C = \{c_1, c_2, \dots, c_{15}\}$ のいずれかが付与されている。また、ある薬効分類 $c \in C$ が付与された医薬品集合は $M_c \subseteq M$ であり、ある薬効分類 c において、 M_c 中の医薬品に対し、関数 f_c によって選択順位が与えられている。ただし、 $f_c: M_c \rightarrow \{0, 1, 2, 3\}$ であり、 $f_c(m)$ が高い方がフォーミュラリで示されるような有用性の高い医薬品として優れていることを意味する。そして、各医薬品 m から製造販売された製剤集合 P_m に対して、各製剤には文書 d_p が存在する。

また各医薬品 m において、フォーミュラリの判定指標である費用負担性(経済性)は製剤集合 P_m における文書 d_p に記載された製剤 p の薬価の平均値とした。平均値としたのは、製剤集合 P_m には、最初に発売され比較的高価な先発品に加え、先発品の特許権が切れた後に発売される安価な後発品が存在する場合があります。また、 P_m の処方時にどちらを選択するかは任意であることや、後発品の存在自体が製剤集合 P_m の経済性に資する点を考慮し、平均値を採用した。そして薬価は安価なほど費用負担性(経済性)として優れている。

$$\text{費用負担性(経済性)}(m) = \frac{1}{|P_m|} \sum_{p \in P_m} \text{薬価}(p)$$

各医薬品 m の有害性(安全性)は製剤集合 P_m における文書 d_p に記載された製剤 p の副作用数の平均値とした。本実験では副作用数の収集対象を、添付文書に必須記載の「重大な副作用」と比較的小さい「その他の副作用」の項目に記載されている副作用用語としている。副作用用語はデファクトスタンダードとして利用されている米国 National Cancer Institute (NCI) が発行する「有害事象共通用語規準: Common Terminology Criteria for Adverse Events (CTCAE)」の日本語版を辞書として使用している[16]。しかしCTCAEは同一の有害事象を重症度によって分けているなど複雑なことから、「その他の副作用」に記載する副作用の用語の範囲は製剤 p の発売元の製薬会社によって幅がある。例えば「その他の副作用」項目で、同一の有害事象でもCTCAEでの重症度のグレードが低い場合は記載を省略している製薬企業もあり、同一の有害事象でも添付文書への副作用語の記載数には変動がある。そのため、副作用数は製剤集合 P_m における平均値を採用することで有害性(安全性)の指標とし、少ないほど無害で安全性が高い医薬品とした。

$$\text{有害性(安全性)}(m) = \frac{1}{|P_m|} \sum_{p \in P_m} \text{副作用数}(p)$$

そして有効性を示す臨床試験数は、製剤集合 P_m における文書 d_p に記載された製剤 p の臨床試験数の最大値とした。臨床

表 1 データセットの詳細

学会モデルフォーミュラリ		医療機関フォーミュラリ	
正例	負例	正例	負例
医薬品	48	18	41
製剤	167	56	180

試験は、製剤集合 P_m で最初に発売された先発品や後発品のうち臨床試験が追加された製剤以外は、添付文書への記載が流用として省略されている。流用として記載が省略されているのは、臨床試験の実施にはコストがかかるため、同じ医薬品において先行して承認されている製剤の臨床試験の内容を継承することが許容されているためである。また追加で臨床試験が実施された場合は、先行した臨床試験の内容に試験項目や試験デザインが追補となり、以降販売される製剤にも追加された臨床試験の内容は継承される。そのため、製剤集合 P_m の有効性を示す臨床試験数は最大値を採用した。臨床試験数はプラセボを使用するなどした対照群に対する有意差が示されているため、臨床試験数が多いほど有効性が高いとした。

$$\text{有効性}(m) = \max_{p \in P_m} \text{臨床試験数}(p)$$

上記のデータセットの統計情報は表 1 で示す。学習データである学会モデルフォーミュラリでは、有用性が高いとしてフォーミュラリと認められた正例が 48 医薬品 167 製剤であり、非フォーミュラリである負例が 18 医薬品 56 製剤となる。またテストデータである医療機関によって選定されたフォーミュラリでは、正例が 41 医薬品 196 製剤、負例が 180 医薬品 692 製剤である。

3.3 学習方法

有用性判定モデルは教師あり学習を使用した。3.2 節で述べたデータセットでは、推奨する順位が付与されていたが、まず二値分類として、副作用数、臨床試験数、薬価を特徴量とし、ラベルはフォーミュラリか否とする実験を行った。教師あり学習モデルは、SVM, RandomForest, 決定木 (DT) の 3 モデルで実施した。

4 実験

本節では、3.2 節で述べたデータセットに対し、3.1 節の指標を特徴量として、3.3 節の学習方法を実施した。タスクは、有用性の有無についての分類とした。各節では、実験設定の説明と結果について述べる。

4.1 実験設定

実験では、有用性の有無を分類するための手法として、教師あり機械学習の SVM, RandomForest, DT の 3 モデルを実施・比較する。また、特徴量として使用する指標は、副作用数、臨床試験数、薬価とする。ラベルの設定はフォーミュラリか否かとする二値とする。実験の評価指標は、分類タスクでよく用いられる ACC: accuracy (正解率), PRE: precision (適合率), REC: recall (再現率), F1: F1-score とした。

表 2 実験の結果：二値分類

	ACC	PRE	REC	F1
SVM	0.19	0.19	1.00	0.31
RandomForest	0.19	0.19	1.00	0.31
DT	0.71	0.25	0.27	0.26

4.2 実験結果

実験結果は表 2 に示す。分類結果は、DT が最も ACC が高く、SVM, RandomForest は低い結果となった。DT では、図 1 で優先度の高い特徴量が青で示されており、特徴量の優先度の高さは薬価、臨床試験数、最後に副作用数の順となった。またフォーミュラリか否かを臨床試験数と薬価でプロットした図 2 や図 3 からは、学習データとテストデータ双方で薬価が高くなるほど「0: 非フォーミュラリ」が増えていることから、薬価がフォーミュラリの判定には影響を及ぼしていることが示唆される。実験の結果から、典型的な機械学習モデルでは、有用性判定のための分類は精度が低く、分類問題には適さないと考えられる。また有用性の判定には特徴量の中で薬価が最も影響していると推測される。

5 まとめ

本論文では、医療用医薬品の有用性を判定する手法を提案した。判定手法として機械学習での分類を用い、フォーミュラリから作成したデータセットに対する実験を行った。実験の結果、副作用数と臨床試験数および薬価を利用した分類では、3 モデルでは DT の ACC がやや高かったものの、全般的に評価指標の値は低く、分類問題としては適さないと考えられる。今後の課題としては、選択順位を活用したランキング学習として、予測性能の向上を図る予定である。

文献

- [1] Ogden Joy, “QALYs and their role in the NICE decision-making process”, Prescriber (London, England), vol.28, no.4, pp.41-43 2017.
- [2] Smith Richard, “Measuring utility values for QALYs: Two methodological issues”, Health Econ., vol.2, no.4, pp.349-355 1993.
- [3] McClellan Mark, “Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality?”, Journal of Economic Perspectives, vol.25, no.2, pp.69-92 2011.
- [4] 今井博久, “病院フォーミュラリの策定に係る標準的手法開発および地域医療への影響の調査研究”, 厚生労働科学研究費補助金 行政政策研究分野 厚生労働科学特別研究 2020.
- [5] 独立行政法人医薬品医療機器総合機構, 医療用医薬品情報検索, <https://www.pmda.go.jp/PmdaSearch/iyakuSearch/>
- [6] 荒牧英治, “特集 AI 時代の医薬品情報のあり方を考える (2) 人工知能による医薬品副作用情報の自動抽出に関する動向”, 医薬品情報学, vol.20, no.1, pp.N5 2018.
- [7] 小島謙介, 岩田浩明, 中津井雅彦, 奥野恭史, “医薬品添付文書からの薬剤情報抽出システム”, 言語処理学会年次大会発表論文集 (Web); 言語処理学会年次大会発表論文集 (Web), vol.26th 2020.
- [8] 栗山猛, 中村秀文, 森田英明, 森川和彦, 石川洋一, 加藤省吾, 荒牧英治, 井上永介, “小児領域の医薬品の適正使用推進のため

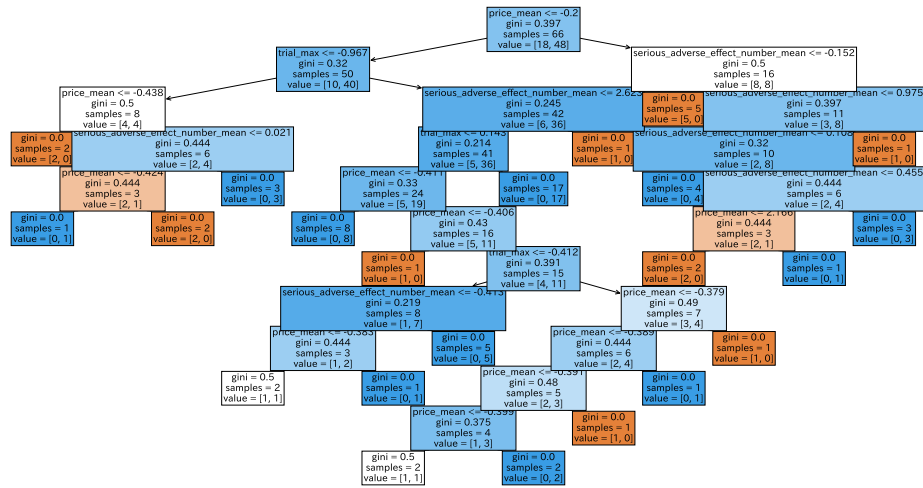


図 1 決定木 (DT) の Tree 図

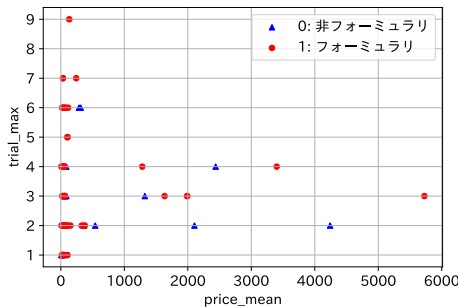


図 2 学習データ: 学会フォーミュラリにおける薬価および臨床試験数のプロット図

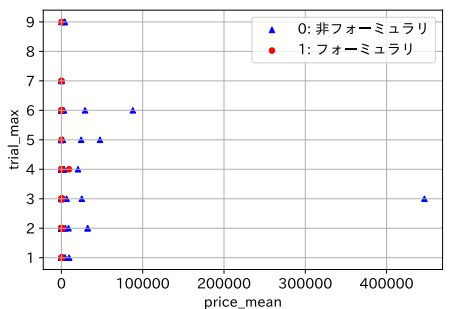


図 3 テストデータ: 医療機関フォーミュラリにおける薬価および臨床試験数のプロット図

の人工知能を用いた医療情報データベースの利活用に関する研究”, 医療情報学, vol.41, no.2, pp.70-71 2021.

[9] Osawa Itsuki, Goto Tadahiro, Yamamoto Yuji, Tsugawa Yusuke, “Machine-learning-based prediction models for high-need high-cost patients using nationwide clinical and claims data”, npj Digital Medicine, vol.3, no.1, pp.148 2020.
 [10] Yang Xi, Chen Aokun, PourNejatian Nima, Shin Hoo

Chang, Smith Kaleb E., Parisien Christopher, Compas Colin, Martin Cheryl, Costa Anthony B., Flores Mona G., Zhang Ying, Magoc Tanja, Harle Christopher A., Lipori Gloria, Mitchell Duane A., Hogan William R., Shenkman Elizabeth A., Bian Jiang, Wu Yonghui, “A large language model for electronic health records”, npj Digital Medicine, vol.5, no.1, pp.194 2022. [10] [11]
 [11] Vimont Alexandre, Leleu Henri, Durand-Zaleski Isabelle, “Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France”, The European Journal of Health Economics, vol.23, no.2, pp.211-223 2022.
 [12] Chengliang Yang, Chris Delcher, Elizabeth Shenkman and Sanjay Ranka, “Machine learning approaches for predicting high cost high need patient expenditures in health care”, BioMedical Engineering OnLine, vol.17, no. Suppl 1, pp.131, Nov 2018.
 [13] Tharsanee R. M., Soundariya R. S., Kumar A. Saran, Karthiga M., Sountharajan S., “Deep convolutional neural network-based image classification for COVID-19 diagnosis”, Data Science for COVID-19, pp.117-145, Elsevier, 2021.
 [14] Dayan Ittai, Roth Holger R., Zhong Aoxiao, Harouni Ahmed, Gentili Amilcare, Abidin Anas Z., Liu Andrew, Costa Anthony Beardsworth, Wood Bradford J., Tsai Chien-Sung, Wang Chih-Hung, Hsu Chun-Nan, Lee C. K., Ruan Peiyang, Xu Daguang, Wu Dufan, Huang Eddie, Kitamura Felipe Campos, Lacey Griffin, de Ant6nio Corradi Gustavo C6sar, Nino Gustavo, Shin Hao-Hsin, Obinata Hirofumi, Ren Hui, Crane Jason C., Tetreault Jesse, Guan Jiahui, Garrett John W., Kaggie Joshua D., Park Jung Gil, “Federated learning for predicting clinical outcomes in patients with COVID-19”, Nat.Med., vol.27, no.10, pp.1735-1743, Oct 2021.
 [15] 厚生労働省, <https://www.mhlw.go.jp/topics/2022/04/tp2022-0401-01.html>, 2022/12/20
 [16] 日本臨床腫瘍研究グループ (Japan Clinical Oncology Group : JCOG), Common Terminology Criteria for Adverse Events (CTCAE) version 5.0, <http://www.jcog.jp/doctor/tool/ctcae5.html>, 2022/09/05