

医薬品需要の効率的時系列クラスタリング —医薬品の需要予測に基づく、在庫量最適化と 流通の非効率解消に向け—

大友 翔一, 金本 祥平, 山口 洋介, 橋上 英宜, 赤林 英夫

薬局の生産性向上のためには、医薬品の適正在庫量の把握が重要となる。そのためには、高精度な医薬品の需要予測が肝要である。しかし、一概に需要予測と言っても、調剤薬局で取り扱う医薬品の点数は多岐に及び、また店舗数も多いため、すべての医薬品に関して薬局ごとの需要予測モデルを構築することは、現実的ではない。そこで本稿では、クラスタリングの手法を医薬品の需要に関する時系列データに応用し、その消費パターンごとに複数のクラスターとしてまとめる手法として TSclust を利用した結果を述べる。このクラスターを用いることで、医薬品の需要予測モデルの簡素化が期待される。

キーワード：調剤データ、需要傾向、分類、クラスタリング、時系列クラスタリング、TSclust

1. 背景

国による地域包括ケア推進の中で、薬局は薬中心から患者中心への業務への変革が求められている。一方現場では人的リソースが不足しており、この変革に対応するためには、業務効率の改善が必須である。他方、昨今では人工知能や AI などの計算機科学的な技術や、日々蓄積される大量のデータ処理技術や、データ解析の手法に関する進歩が目覚ましい。そこで今、医療の現場では、こうしたデータ処理や、データ解析の技術を導入し、人材不足の解消に対応することは必要不可欠であると言えるであろう。

その中で、株式会社ファーマクラウドでは、医療現場の業務効率の改善や薬局に特化した薬剤師アシスタント AI の開発を目的として、複数の WEB サービスを提供している。

なかでも、医薬品在庫シェアリングサービスであるメドシェアからは、薬局における調剤データが日々蓄積される。このデータを活用することで、医薬品の需給予測モデルを構築することは、一見すると容易であ

るように思われる。しかし、調剤薬局で取り扱う医薬品の商品点数は多岐に及び、また薬局の店舗数も多い。さらに医薬品の消費傾向に関しても、都市部や地方、駅前、オフィス街など立地条件などにも大きく影響を受ける。そのため、それらの特色を踏まえたうえで、各薬局ですべての医薬品に関して個別の需要予測モデルを構築することは現実的ではない。

そこで、本稿では、医薬品の需要に関して、クラスタリング手法を時系列データに適用し、各医薬品ごとの予測モデルではなく、複数の医薬品の需要傾向を一つのクラスター単位にまとめる方法に関して考察したい。なお本稿で言及する医薬品の消費に関するデータは、2014年12月1日から2018年6月30日におけるメドシェア導入店舗の、各医薬品についての消費量を、1日単位で集計および合算を行ったものとする。

2. 需要予測における各種手法と現在

2.1 機械学習・AI分野の現状

機械学習・AIの分野に関しては、日々新たな論文が投稿されており、そのことに関して、本稿で改めて説明するまでもないであろう。また、それら論文の成果は、Google や Microsoft をはじめとする IT 企業では、すでに製品化あるいはサービス化を行っている。そのため、平易な予測モデルなどであれば、プログラミングなどの専門知識を必要とせずとも、データさえあれば、容易に機械学習のモデルが作成できるようになりつつある。

2.2 薬局における調剤業務への適応と問題点

こうした機械学習や AI に関するサービスは数多あ

おおとも しょういち
慶應義塾大学産業研究所
〒108-8345 東京都港区三田 2-15-45
s.otomo@sanken.keio.ac.jp

かねもと しょうへい, やまぐち ようすけ,
はしかみ ひでのぶ
株式会社ファーマクラウド

〒101-0051 東京都千代田区神田神保町 1-2-3
あかばやし ひでお
慶應義塾大学経済学部, 慶應義塾大学産業研究所
〒108-8345 東京都港区三田 2-15-45

り、AI や人工知能の民主化が喧伝されている。一方で、それらを医薬品の需要予測に適応する場合には、いくつかの課題がある。たとえば、必要な予測モデルの数が挙げられる。医薬品は、現状流通している商品数で約 7 万種類程度になる。また、ファーマクラウドのサービスを導入している店舗数が 2019 年 3 月末の段階で 500 店舗以上ある。このときに、7 万種類の医薬品の需要に関して、各店舗ごとの予測モデルを作成する場合、3500 万以上のパターンのモデルを作成する必要がある。今後、サービス導入を行う店舗を増加させる際に、同様のアプローチを採用することは現実的ではないであろう。

実際にファーマクラウドでは、GCP、BigQuery など Google のサービスを使用しているが、たとえば BQML では現時点における 1 日当たりで作成可能なモデル数などについての制約がある。

2.3 提案手法

本稿では、モデル数が爆発的に増加することを避けるため、各医薬品の調剤データに対して TSclust [1] を活用する方法を述べる。また、その各薬品の時系列間の類似性検索を行った結果である、同一クラスター内における近い需要傾向の医薬品に共通する性質およびその分割されたクラスターの特徴に関する論述する。

3. 時系列クラスタリング

3.1 TSclust

文献 [1] によれば、時系列クラスタリングは、幅広い分野で応用されている活発な研究分野である。事実、経済学、金融学、医学、生態学、環境学、工学、そのほか多くの分野で自然に発生し、活発な改善が日々行われており、特に近年、目覚ましい発展をした手法である。TSclust は、時系列クラスタリングを実装したライブラリとして CRAN¹[2] にて公開されている。また、実装方法に関しては文献 [3] に詳細に記述されている。

3.2 医薬品需要に対する TSclust の利用

2014 年 12 月 1 日から 2018 年 6 月 30 日における、メドシェアの導入店舗における処方回数合計の上位 100 種の医薬品に関して、時系列の消費量をグラフ化すると図 1 のようになる。この医薬品の需要を表す調剤データに対して TSclust を使用する。クラスタリングは、非類似度、距離尺度の測定、あるいは従来のクラスタ

リング手法により実行される。メソッドとして実装されている分類の手法は 20 種以上ある。ここではその中から ACF (Autocorrelation-based Dissimilarity)、DTW (Dynamic Time Warping Distance)、EUCL (Euclidean Distance) の三つのメソッドに関しての実行結果を記載する。

3.3 クラスタリングの際の距離尺度

3.3.1 ACF

単純な自己相関 ACF または PACF (Partial Auto-Correlation) の重み付きユークリッド距離を計算する。自己相関 (dist.PACF) 係数 p も Ω も指定されていない場合は、一様重み付けが用いられる。 p が明示されている場合は、以下の式で計算される [2]。

$$D(x, y) = \{(\hat{p}_x - \hat{p}_y)^t \Omega (\hat{p}_x - \hat{p}_y)\}^{\frac{1}{2}}$$

医薬品の消費傾向には医薬品ごとの特徴があり、たとえば花粉症や風邪などの季節と関連の強い急性疾患に対して処方される医薬品の場合は、毎年同じ時期に需要が急増する。高血圧症や糖尿病に対しての処方薬のように恒常的に服用することの多い薬品などに関しては、1 週間や 1 カ月を単位として医薬品が処方されるため、基本的に患者の来局する日や必要とされる数量はある程度定まる。結果として、それぞれの医薬品の消費に関しては、自己相関を繰り返すことが多い。

3.3.2 DTW

医薬品の需要は、薬局の立地や規模などの条件によって大きく左右され、1 日当たりの来局患者数もそれらに影響を受ける。また、商品点数に対して、いわゆる売れ筋商品が特定の医薬品に集中しているため、データがスパースになりやすい。結果として、異なる医薬品においてはデータ長が異なることが多い。文献 [4, 5] によれば、こうした長さの等しくない時系列データに関しては、pdc が有効であることが知られている。また、文献 [2] によれば、DTW は pdc を基本とする関数として実装されている。そのため、こうした医薬品の需要データの性質上からは DTW が有効に作用することが予想される。

3.3.3 EUCL

ユークリッド距離は計算方法が単純であり、計算量も少ないことから、最も一般的に使用される距離尺度である。 x, y を比較する時系列データとし、 x の i 番目のデータを x_i とすると、以下の式で計算される。

$$D(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

¹ Comprehensive R Archive Network は、R 本体や各種パッケージをダウンロードするための Web サイトである。

一方で、長さの異なる時系列データの比較や時間軸方向にずれがある時系列データ同士の距離の計算を正しく行えないことが問題点とされている [6]。今回の場合は、医学あるいは薬学的な、薬効分類や投与経路などの観点から医薬品を正しく分類することよりも、医薬品の消費傾向をクラスタリングし、予測モデルを簡素

化することが目的である。そのため、計算量が少ない点を重視するならば有効に作用することが予想される。

4. 医薬品需要への TSclust の応用

3.3 節で述べた三つのメソッドを実行して、医薬品の需要を四つのクラスターに分類し、その結果を容易

処方回数上位100種医薬品

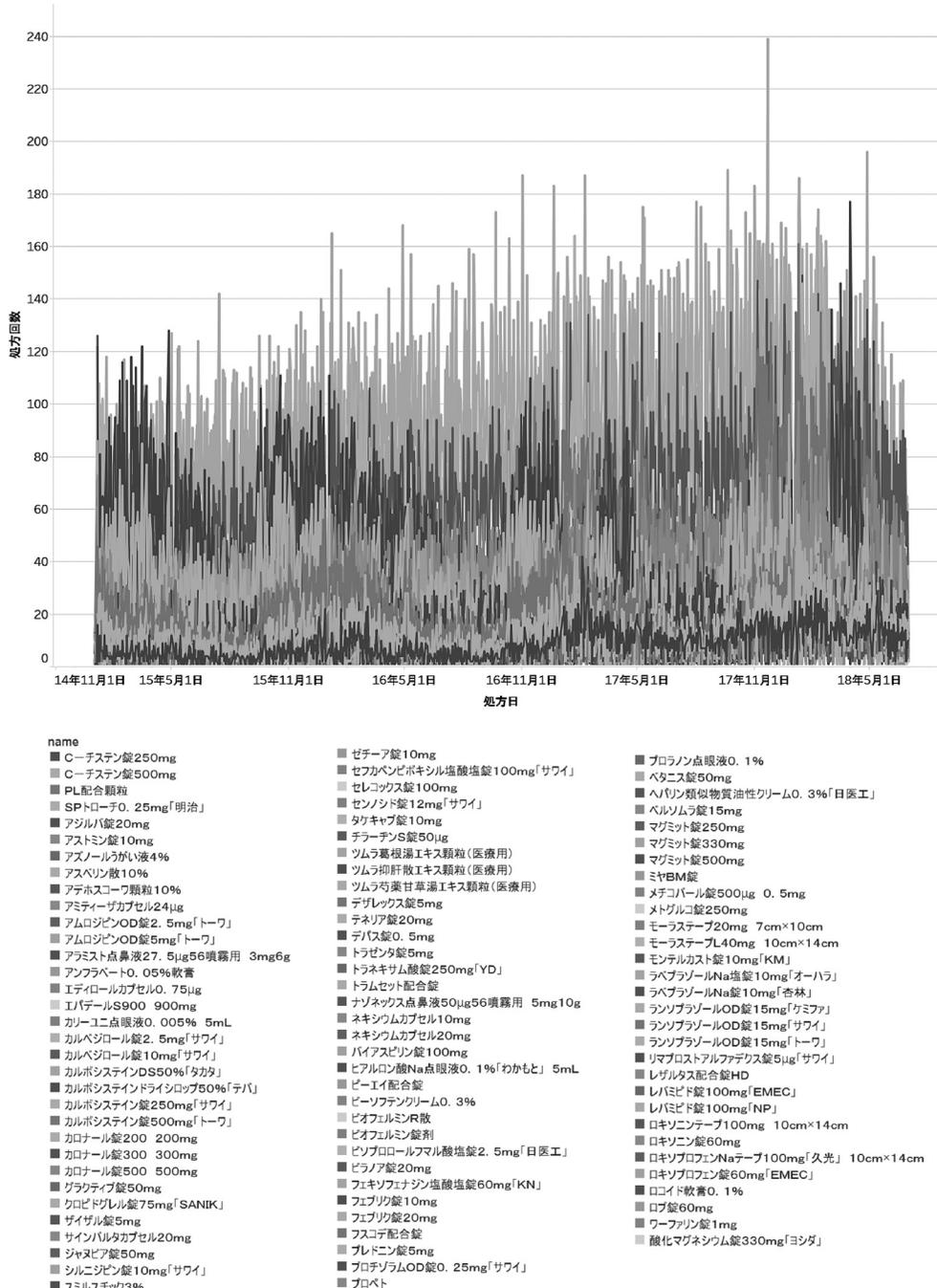


図1 処方回数上位100種の日ごとの処方回数

に可視的に理解可能となるようにしたものが、図2である。

ファーマクラウドの目的とするところに、調剤薬局の在庫管理における意思決定支援あるいは薬剤師のアシスタント AI の作成がある。よって本稿では、各クラスターに関して一定数以上の要素がある ACF を用いた分類に関して後述する。ACF に基づき分類する場合は図3のようなデンドログラムが作成される。

また、分類の階層やクラスターの数は変更可能であり、それに応じて計算機への負荷も変動する。この点

に関しては別の機会に詳述する。

4.1 TSclust を用いた医薬品需要の分類結果

既存の医薬品の分類体系も実にさまざまな分類の規則がある。たとえば、YJ コードのように薬効分類番号や投与経路と成分、剤形などにより分類される。これらは文献 [7] に詳しい。

一方、本稿で紹介する TSclust を用いた分類は、医薬品の消費パターンから導かれる分類であり、これまでの薬効分類番号や投与経路と成分などとは直接的な関係をもたない。しかしクラスタリング結果は、前述の四つのそれぞれのクラスターに関して、後述するような特徴をもつことがわかった。

4.2 各クラスターにおける医薬品の特徴

既述の医薬品に関して、TSclust の ACF を用いて四つのクラスターに分類した場合、下記のように分類された。

1. 高齢者や介護施設、老人ホームなどで処方されることが多い医薬品
2. 胃腸、整腸関連のもので、神経性のものが想定される医薬品
3. 急性疾患に対して処方されることが多い医薬品
4. 中年の生活習慣病に対して処方されることが多い医薬品

ここで分類した医薬品および市場に流通するすべての医薬品が、この分類に完全に当てはまるものではない。だが、この大まかな傾向から、クラスター番号の1番に従属する医薬品は、常用される医薬品が多く含まれる結果となった。また、3番に従属する医薬品は急性疾患に対して処方されるため、同一の患者ではない場合が多いことや、需要に対して季節性の影響の強いものが多い医薬品となった。これらの結果から、医薬品の需要に固有の特徴によって分類されたと言える

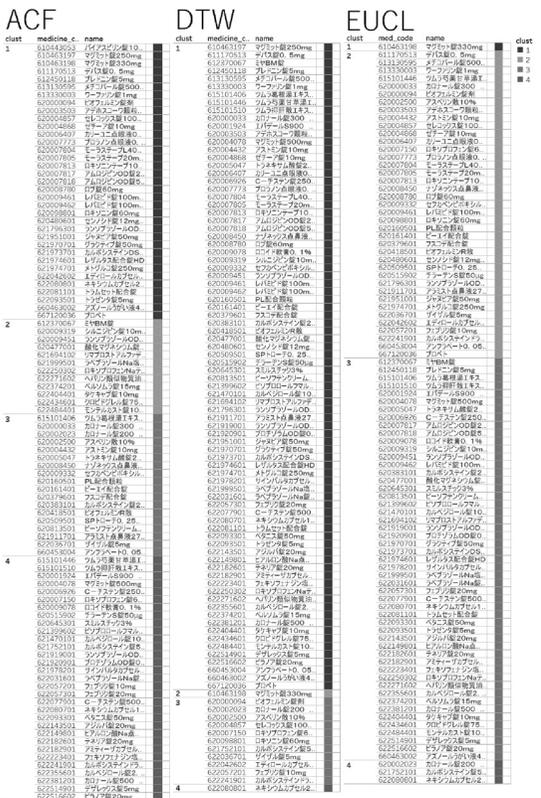


図2 処方回数上位 100 種の TSclust による分類結果

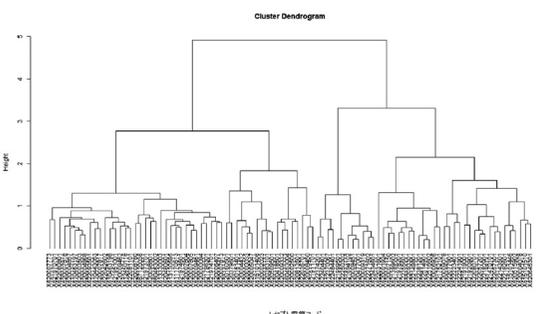


図3 処方回数上位 100 種の TSclust (ACF) によるデンドログラム

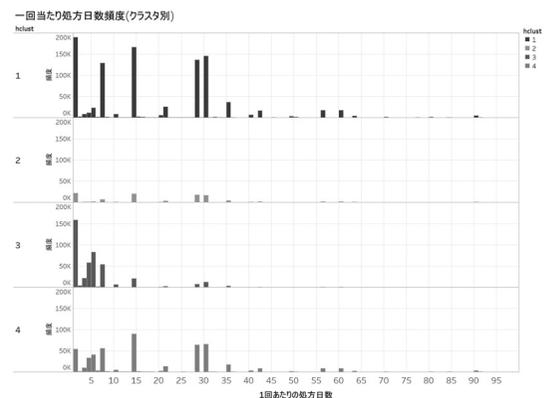


図4 1回当たり処方日数頻度 (クラスター別)

であろう。

また、各クラスターの医薬品の処方日数の分布にも、これらの特徴は表れている。図4を参照すると、たとえば、クラスター番号の1番に従属する医薬品は、1日分の処方のほかに、7日、14日、28日、30日というように、1週間を単位とする7の倍数もしくは1カ月を単位とする30の倍数が多くなる。これは、常用される医薬品に多くみられる特徴であるが、2番に従属する医薬品は、56日や60日といった、2カ月以上分を一度に処方される長期処方ほとんど見受けられない。また3番に従属する医薬品は、1日の処方以外も多いが、ほぼすべての処方が1週間以内の処方日数に収まっていることが見受けられる。

5. おわりに

本稿では、医薬品の需要予測に関して時系列の消費パターンからクラスタリングを行う手法を提示した。本稿で取り上げた三つの分類手法の中では、各クラスターに関して一定以上の数があり、最もうまく分類されている「薬剤師の直感に比較的近い」ものはACFによる分類であることがわかった。

また、ACFを用いた分類は、結果的にクラスターには医薬品ごとの特徴が含まれており、既存の分類体系とは異なり、実処方に際しての用途や処方日数ごとに近いクラスターに分類されることがわかった。

今後は精度を下げることなく、計算量を減らすことができるようなアルゴリズムの改良や、新たなアルゴリズムの開発が望まれる。また、医薬品の需要予測に

関するアルゴリズムなどの理論研究のみならず、社会実装も課題である。研究成果の予測モデルに基づき医薬品の受発注を行うことで、突発的な大量発注や、大量発注の結果として過剰になった医薬品の廃棄などの削減が期待される。結果、各薬局における、医薬品在庫量の最適化に貢献できるであろう。

謝辞 本研究は、慶應義塾大学産業研究所が株式会社ファーマクラウドからの助成を受け、共同で実施したものである。

参考文献

- [1] P. Montero and J. A. Vilar, “TSclust: An R package for time series clustering,” *Journal of Statistical Software*, **62**, pp. 1–43, 2014.
- [2] P. M. Manso and J. A. Vilar, “The Comprehensive R Archive Network,” <https://cran.r-project.org/web/packages/TSclust/TSclust.pdf> (2019年03月25日閲覧)
- [3] P. Montero and J. A. Vilar, “TSclust, Rdocumentation,” <https://www.rdocumentation.org/packages/TSclust/versions/1.2.4/topics/TSclust> (2019年03月25日閲覧)
- [4] A. M. Brandmaier, “Permutation Distribution Clustering,” <https://cran.r-project.org/web/packages/pdc/pdc.pdf> (2019年04月02日閲覧)
- [5] A. M. Brandmaier, “pdc: An R package for coplexity-based clustering of time series,” *Journal of Statistical Software*, **67**, pp. 1–23, 2015.
- [6] 吉田将, チャクラボルティ バサビ, “時系列データにおける距離尺度の評価と分析,” 平成27年度電気関係学会東北支部連合大会講演論文集, pp. 124–124, 2015.
- [7] 山中理, “知っているが業務効率化! 医薬品コードの種類と活用のコツ,” 月刊薬事, **60**, pp. 9–12, 2018.