Научная статья
Original article
УДК 004

EFFICIENT TIME-SERIES CLUSTERING OF DRUGS DEMAND TO INVENTORY OPTIMIZATION IN HEALTH CENTERS

ЭФФЕКТИВНАЯ КЛАСТЕРИЗАЦИЯ ВРЕМЕННЫХ РЯДОВ СПРОСА НА ЛЕКАРСТВА ДЛЯ ОПТИМИЗАЦИИ ЗАПАСОВ В МЕДИЦИНСКИХ ЦЕНТРАХ

Wedyan Habeeb Hameed, Institute of Information Technology And Intelligent Systems (ITIS) Kazan Federal University, Kazan, Russia, wedy.hameed@gmail.com.
Enas Mahmoud Jassim, University of Diyala, Diyala, Iraq, enasmahmoud@uodiyala.edu.iq.

Abstract: To eliminate distribution flaws and boost pharmacy productivity, it is critical to understand the appropriate amount of drugs to keep on hand. For this purpose, it is important to forecast the demand for pharmaceutical products with high accuracy. However, even if demand can be predicted, it is not realistic to build a single demand-forecasting model for all pharmacies and for all drugs because the number of drugs handled by pharmacies is diverse, the number of stores is large, and the location of the pharmacy affects the amount of drug consumption. Therefore, in this paper, we apply the aggregation method of time series data to drug demand, and describe the results of using TSclust as a method for integrating multiple cohorts for each consumption pattern. By using this combination, it is expected that the demand forecasting model for

pharmaceutical products will be simplified and made more efficient, thus reducing the waste of public money.

Аннотация: Чтобы устранить недостатки в распределении и повысить производительность аптек, крайне важно понимать, какое количество лекарств нужно иметь под рукой. Для этой цели важно с высокой точностью прогнозировать спрос на фармацевтическую продукцию. Однако, даже если спрос можно спрогнозировать, нереально построить единую модель прогнозирования спроса для всех аптек и для всех лекарств, поскольку количество лекарств, обрабатываемых аптеками, разнообразно, количество магазинов велико, а местоположение аптеки влияет на объем потребления лекарств. Поэтому в этой статье мы применяем метод агрегирования данных временных рядов к спросу на наркотики и описываем результаты использования TSclust в качестве метода интеграции нескольких когорт для каждой модели потребления. Ожидается, что при использовании этой комбинации модель прогнозирования спроса на фармацевтическую продукцию будет упрощена и станет более эффективной, что сократит трату государственных средств.

Key words: Forecasting, Demand, Pharmaceuticals, Model, Medicine, Tsclust.

Ключевые слова: Прогнозирование, Спрос, Фармпродукция, Модель, Медицина, Tsclust.

1. Introduction

In the context of promoting comprehensive community care by the government, pharmacies are required to bring about a change in operations that are drug-centric or patient-centered [8]. Due to the lack of human resources in the health field, it is necessary to improve work efficiency in order to respond to this change. On the other hand, in recent years, remarkable progress has been made in computer science technologies such as artificial intelligence, techniques for collecting data accumulated from multiple websites, and methods for processing and analyzing that data. Therefore,

it can said that it is indispensable for the medical field to introduce data processing and data analysis techniques to solve the shortage of human resources. Among the websites, the site of (Pharma Cloud Inc.), which it offers many of its services for the purpose of improving the efficiency of the work of medical websites and developing the work of pharmacies. In particular, big data has collected in pharmacies daily from ClinCalc, a drug stock-sharing service. At first glance, it seems easy to build a supply and demand forecasting model for pharmaceutical products by taking advantage of this data, but the number of commodity products that pharmacies deal with is diverse, and the number of pharmacies is also large. Moreover, the drug consumption trend has greatly influenced by site conditions such as urban areas, rural areas, train stations, and office areas. Therefore, based on these characteristics, each pharmacy has a separate demand forecast model for drugs, and it differs on the others models. Therefore, in this paper, we would like to look at how to apply the forecasting method to time-series data on drugs demand and summarize multi-drug demand trends in a single cluster unit rather than a prediction model for each drug separately [5]. The drug consumption data used in this paper had collected from ClinCalc (2019) on a daily basis.

2. Existing methods of forecasting demand

There are many such services related to machine learning and artificial intelligence that have being promoted every day, but these services face many challenges when dealing with the expectations of demand for medicines. For example, among those challenges is the number of predictive models required. There are currently about 70,000 types of pharmaceutical products on the market, and more than 500 stores had added to Pharma-Cloud at the end of March 2019. When creating a forecast model for each store regarding the demand for 70,000 types of pharmaceutical products, it is necessary to create a model with more than 35 million patterns. It may not be realistic to adopt a similar approach when increasing the number of stores offering services in the future. In fact, Pharma-Cloud uses Google services like GCP and Big Query, but BQML has a limit on the number of forms that can created per day.

2.1 Method Suggestion

This article describes how TSclust can used to distribute the data for each drug to avoid the explosive growth in the number of models. Additionally, because of searching for similarities between time series for each drug, common characteristics of drugs with similar demand trends within the same group are included. We will also discuss the characteristics of divided groups [1].

3. Time-series clustering.

3.1 TSclust

Time series clustering is an active research field that is applied in a wide range of fields. In fact, it occurs naturally in economics, finance, medicine, ecology, environmental studies, engineering, and many other fields, with active improvements every day. TSclust is published in CRAN1 [2] as a library that implements time series clustering. In addition, the implementation method has described in detail in [3].

3.2. Utilization of TSclust for drug demand

TSclust has used for dispensing data that represents the demand for this drug. Clustering has performed by measuring dissimilarity, distance scale, or by conventional clustering techniques. There are more than 20 classification methods implemented as methods. Here, ACF (Autocorrelation-based Dissimilarity), DTW (Dynamic Time Warping Distance), EUCL are among them. The execution results for the three methods of (Euclidean Distance) have described. Table 1 shows the top 100 drugs with the total number of prescriptions at ClinCalc store in (2019).

Table 1. The Top 100 Drugs from ClinCalc (2019)

Rank	Drug Name	Total Prescriptions (2019)	Rank	Drug Name	Total Prescriptions (2019)
1	Atorvastatin	112,104,359	51	Methylphenidate	14,233,405
2	Levothyroxine	102,595,103	52	Apixaban	14,042,889
3	Lisinopril	91,862,708	53	Ranitidine	13,586,751
4	Metformin	85,739,443	54	Glipizide	13,424,610
5	Metoprolol	74,578,817	55	Ergocalciferol	13,273,652

6	Amlodipine	73,542,114	56	Quetiapine	13,114,560
7	Albuterol	60,679,987	57	Budesonide;	12,473,902
8	Omeprazole	52,546,641	58	Estradiol	12,393,425
9	Losartan	51,773,869	59	Acetaminophen;	11,962,650
10	Gabapentin	47,149,505	60	Ondansetron	11,856,066
11	Hydrochlorothiazide	38,609,803	61	Naproxen	11,762,233
12	Sertraline	37,157,933	62	Glimepiride	11,504,531
13	Simvastatin	36,812,966	63	Spironolactone	11,432,027
14	Montelukast	32,154,358	64	Clonidine	11,418,367
15	Acetaminophen;	30,355,778	65	Insulin Lispro	11,389,229
16	Pantoprazole	28,880,217	66	Loratadine	11,374,226
17	Furosemide	28,352,226	67	Cetirizine	11,110,560
18	Fluticasone	27,893,102	68	Topiramate	10,927,224
19	Escitalopram	27,510,958	69	Lorazepam	10,875,212
20	Fluoxetine	27,110,302	70	Ethinyl Estradiol;	10,860,083
21	Rosuvastatin	27,041,319	71	Lamotrigine	10,690,317
22	Bupropion	25,722,873	72	Diltiazem	10,604,813
23	Amoxicillin	25,702,634	73	Hydrochlorothiazide;	10,268,139
24	Dextroamphetamine;	24,600,698	74	Diclofenac	10,115,975
25	Trazodone	23,934,213	75	Hydroxyzine	9,898,263
26	Duloxetine	23,821,965	76	Buspirone	9,881,603
27	Prednisone	22,889,929	77	Latanoprost	9,800,569
28	Tamsulosin	21,934,065	78	Paroxetine	9,783,755
29	Ibuprofen	21,746,702	79	Lisdexamfetamine	9,775,262
30	Citalopram	21,546,700	80	Fluticasone;	9,762,036
31	Meloxicam	21,459,849	81	Pregabalin	9,625,189
32	Pravastatin	20,683,277	82	Propranolol	9,277,061
33	Carvedilol	20,602,256	83	Cephalexin	9,246,463
34	Potassium	20,001,670	84	Cholecalciferol	9,068,152
35	Tramadol	19,838,715	85	Insulin Aspart	9,067,406
36	Clopidogrel	19,447,746	86	Finasteride	8,986,897
37	Insulin Glargine	19,211,653	87	Fenofibrate	8,970,219
38	Aspirin	18,143,138	88	Sitagliptin	8,866,811
39	Atenolol	18,091,488	89	Folic Acid	8,860,645
40	Venlafaxine	17,713,653	90	Doxycycline	8,809,374
41	Alprazolam	17,533,262	91	Rivaroxaban	8,799,404
42	Ethinyl Estradiol;	16,505,642	92	Tizanidine	8,729,694
43	Allopurinol	15,900,788	93	Amoxicillin;	8,372,244

44	Hydrochlorothiazide;	15,709,833	94	Amitriptyline	8,178,156
45	Cyclobenzaprine	15,597,385	95	Lovastatin	8,091,735
46	Clonazepam	15,578,495	96	Alendronate	7,811,899
47	Zolpidem	15,419,648	97	Levetiracetam	7,560,850
48	Azithromycin	15,300,433	98	Sumatriptan	7,050,329
49	Oxycodone	14,669,103	99	Hydralazine	6,655,156
50	Warfarin	14,632,370	100	Sulfamethoxazole;	6,630,866

Source: The Top 200 Drugs of 2019. (2021, September 12). Retrieved from ClinCalc: https://clincalc.com/DrugStats/Top200Drugs.aspx

3.3 Distance scale for clustering

3.3.1 The autocorrelation function (ACF)

Calculate the weighted Euclidean distance of a simple autocorrelation ACF or PACF (Partial Auto-Correlation). If neither the autocorrelation coefficient p nor Ω has specified, uniform weighting is used. If p has specified, it has calculated by the following formula [2].

$$D\,(x, y) = \{(\hat{}P_x - \hat{}P_y)^t \Omega(\,(\hat{}P_x - \hat{}P_y)^{1/2}$$

The consumption tendency of medicines has its own characteristics. For example, in the case of medicines prescribed for acute diseases such as pollinosis and colds, which have strongly related to the season, the demand for medicines increases rapidly at the same time every year. For drugs that have often taken regularly, such as prescription drugs for hypertension and diabetes, the drugs have prescribed on a weekly or monthly basis, so the day the patient visits the doctor. Moreover, the required number is determined to some extent. As a result, autocorrelation has often repeated for the consumption of each drug.

3.3.2 Dynamic Time Warping (DTW)

Demand for medicines has greatly influenced by conditions such as the location and scale of pharmacies, and the number of patients visiting the pharmacy per day has affected by them. In addition, data tends to be sparse because so-called top-selling

products are concentrated on specific drugs with respect to the number of products. As a result, different drugs often have different data lengths. According to the literature [4, 5], it is known that DTW is used to collect time series of similar shapes. Cluster centroids are computed with respect to DTW. A barycenter is the average sequence from a group of time series in DTW space. It is also effective for such time series data with unequal lengths. In addition, according to [2], DTW has implemented as a function based on DTW. Therefore, due to the nature of the demand data for these drugs, he expects DTW to work effectively.

3.3.3 EUCL

The EUCL scale is the most commonly used distance scale because of its simple calculation method and small amount of calculation. If x and y are time-series data to be compared, and the i^{th} data of x is xi, it is calculated by the following formula.

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

On the other hand, the problem is that it is not possible to correctly compare time-series data with different lengths and calculate the distance between time-series data with deviations in the time axis direction [6]. In this case, the purpose is to cluster the consumption tendency of pharmaceutical products and simplify the prediction model rather than correctly classify pharmaceutical products from the viewpoint of medical or pharmaceutical efficacy classification and administration route. Therefore, it has expected that it will work effectively if the point that the amount of calculation is small has emphasized [4].

4. Application of TSclust to pharmaceutical demand

Figure 1 shows the execution of the three methods described in Section 3.3 to classify the demand for medicines into four clusters so that the results can easily and visually understood. The purpose of Pharma-Cloud is to support decision-making in the inventory management of dispensing pharmacies or pharmacists. In this paper, we will discuss the classification using his ACF, which has more than a certain number of elements for each cluster. The classification hierarchy and the number of clusters can changed accordingly.

4.1 Results of the classification of drug demand using TSclust

The existing classification system for pharmaceutical products also has various classification rules. For example, like the YJ code, it has classified according to the drug efficacy classification number, administration route and component, dosage form [7]. On the other hand, the classification using TSclust introduced in this paper is a classification derived from the consumption pattern of medicines and has no direct relationship with the drug efficacy classification number, administration route, or ingredients so far. However, the clustering results showed that each of the above four clusters had unique characteristics.

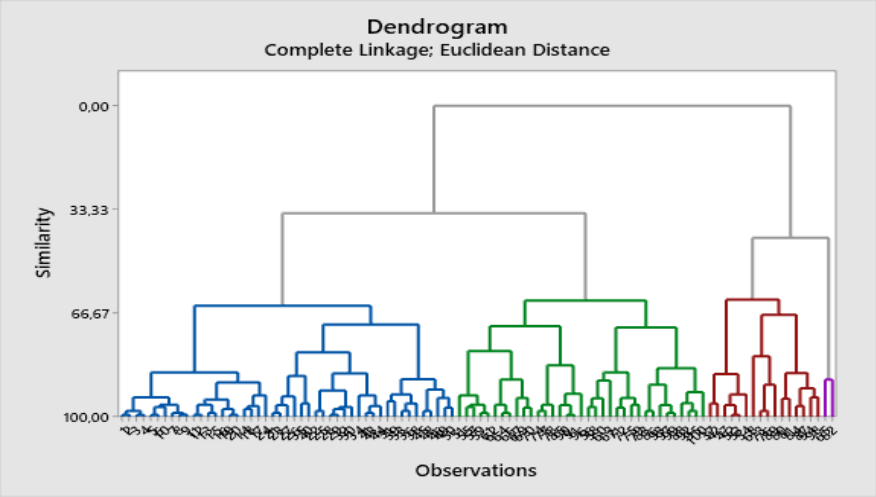
4.2 Characteristics of medicines in each cluster

When the above-mentioned drugs had classified into four clusters using TSclust's ACF, they have classified as follows: (1) Medicines that have often prescribed for the elderly, long-term care facilities, and elderly homes. (2) Drugs related to gastrointestinal and intestinal regulation that have expected to be neurological. (3) Drugs often prescribed for acute illness. (4) Medicines often prescribed for middle-aged lifestyle-related diseases, figure 1 shows Cluster analysis of observations by using TSclust (AFC), and figure 2 shows the number of observations in each cluster.

The medicines classified here and not all medicines on the market do not completely fall into this classification. However, this general tendency resulted in the fact that the drugs subordinate to the cluster number 1 contained many commonly used

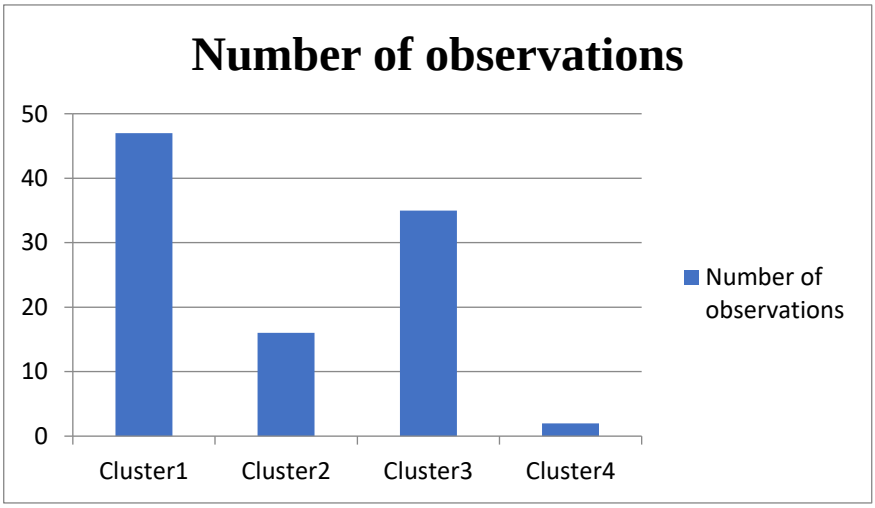
drugs. In addition, since the drugs subordinate to No. 3 have prescribed for acute diseases, they are often not the same patient, and many of them have a strong seasonal influence on demand. From these results, it can said that it had classified according to the characteristics specific to the demand for medicines.

Figure 1. Cluster analysis of observations using TSclust (AFC)



Source: Prepared by the researchers based on the information in table 1.

Figure 2. The number of observations in each cluster



Source: prepared by the researchers based on the results of TSclust (AFC).

Conclusion

In this paper, we presented a method for clustering from time-series consumption patterns for drug demand forecasting. Of the three classification methods discussed in this article, the most well classified and "relatively close to the pharmacist's intuition" is the ACF classification. In addition, classification-using ACF results in clusters containing characteristics for each drug, and unlike the existing classification system, they have classified into clusters that are close to each other according to the purpose of the actual prescription and the number of prescription days. In the future, it is desired to improve the algorithm and develop a new algorithm that can reduce the amount of calculation without lowering the accuracy. Theoretical research on algorithms related to drug demand forecasting, social implementation is also an issue. By placing and receiving orders for medicines based on the prediction model of research results, it has expected to reduce sudden large-scale orders and the disposal of excess medicines. As a result, it will contribute to the optimization of the amount of medicine in stock at each pharmacy.

ACKNOWLEDGEMENTS. The work has performed according to the Russian Government Program of Competitive Growth of Kazan Federal University and Diyala University.

Literature

1. P. Montero and J. A. Vilar, 2014. “TSclust: An R package for time series clustering,” Journal of Statistical Software, 62, p. (1- 43).
2. P.M. Manso and J.A. Vilar, 2019. “The Comprehensive R Archive Network,” https://cran.r-project.org/web/ packages / TSclust / TSclust.pdf.
3. P. Montero and J. A. Vilar, 2019. “TSclust, R- documentation,”https: //www.rdocumentation.org/packages/ TSclust / versions / 1.2.4 / topics / TSclust.
4. A. M. Brandmaier, 2019. “Permutation Distribution Clustering,” https://cran.r-project.org/web/packages/pdc/pdc.pdf
5. A. M. Brandmaier, 2015. “pdc: An R package for complexity-based clustering of time series,” Journal of Statistical Software, 67, pp.(1-23).

6. M. Yoshida, C. V. Basabi, "Evaluation and Analysis of Distance Scales in Time Series Data," Proceedings of the 2015 Tohoku Chapter Joint Conference of the Institute of Electrical and Related Engineers, pp. (124-124).
7. O. Yamanaka, 2018. "If you know it, work efficiency will be improved Types of drug codes and tips for their use," Monthly Pharmaceutical Affairs, 60, pp. (9–12).
8. Hameed, W. H. (2021). Forecasting Drug Needs And Quantification Tools In Healthcare Institutions. Quantum Center, pp. (3-12).

Литература

1. П. Монтеро и Дж. А. Вилар, 2014. «TSclust: пакет R для кластеризации временных рядов», Журнал статистического программного обеспечения, 62, с. (1- 43).
2. П.М. Мансо и Дж.А. Вилар, 2019 г. «Комплексная сеть архивов R», https://cran.r-project.org/web/packages/TSclust/TSclust.pdf.
3. П. Монтеро и Дж. А. Вилар, 2019. «TSclust, R-документация», https://www.rdocumentation.org/packages/TSclust/versions/1.2.4/topics/TSclust.
4. А. М. Брандмайер, 2019. «Кластеризация распределения перестановок», https://cran.r-project.org/web/packages/pdc/pdc.pdf.
5. А. М. Брандмайер, 2015. «PDC: пакет R для кластеризации временных рядов на основе сложности», Журнал статистического программного обеспечения, 67, стр. (1-23).
6. М. Ёсида, К. В. Басаби, «Оценка и анализ масштабов расстояний в данных временных рядов», Материалы совместной конференции главы Тохоку 2015 г. Института инженеров-электриков и смежников, стр. (124-124).
7. О. Яманака, 2018 г. «Если вы это знаете, эффективность работы повысится. Типы кодов лекарств и советы по их использованию», Ежемесячные фармацевтические дела, 60, стр. (9–12).

8. Хамид, В. Х. (2021). Прогнозирование потребности в лекарствах и инструменты количественного определения в медицинских учреждениях. Квантовый центр, стр. (3-12).

© Wedyan Habeeb Hameed, Enas Mahmoud Jassim, 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.

Для цитирования: Wedyan Habeeb Hameed, Enas Mahmoud Jassim EFFICIENT TIME-SERIES CLUSTERING OF DRUGS DEMAND TO INVENTORY OPTIMIZATION IN HEALTH CENTERS// Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022.