International Journal of Innovative Research in Medical Science (IJIRMS)
Volume 03 Issue 01 January 2018, ISSN No. - 2455-8737
Available online at - www.ijirms.in

Open Access Journal | Research Article | DOI: 10.23958/ijirms/vol03-i01/01

Clustering Patients with Adverse Drug Reactions

Andy W. Chen

University of British Columbia, Vancouver, BC, Canada
Email id - *andywchenca99@gmail.com*

Abstract:

Adverse drug reactions are serious unintended effects from drug usage. It affects millions of people worldwide each year, resulting in numerous deaths and hospitalizations. Because millions of people are affected, it would be beneficial to know the frequent cases so health care practitioners can use more caution when prescribing drugs to patients. In this paper, I present a clustering model that identifies the significant groups of patients with adverse drug reactions. Clustering model is an unsupervised machine model that finds the best number of groups or classes for the instances in the data set. Using the clustering approach, I find that the optimal number of groups of patients in the data is five, meaning that there are five groups that can be identified from the data. Knowing this, when health care practitioners prescribe drugs for treatment, they would be more knowledgeable about the kinds of patients that belong to a certain group of frequent cases and use more caution to avoid high risk prescriptions.

Keywords: ***Adverse Drug Reaction Reporting Systems, Drug Interactions, Adverse Drug Effects.***

Introduction

Adverse drug reactions are harmful side effects from taking drugs. These effects can be serious or lethal and can result in hospitalizations or deaths. Past research on adverse drug reactions focus on finding factors that contribute to the occurrences of adverse drug reactions. Ferner and Aronson1 studied how countries differ in attitudes to adverse drug reactions and found that many local cultural factors influence reporting on adverse drug reactions. Alomar2 explored the effect of certain factors such as smoking or alcohol on the occurrence of adverse drug reactions and found significant associations. Maria et al.3 identified predisposing and enabling factors on adverse drug reactions and found that knowledge about adverse drug reactions contribute to the occurrences and reporting of them. Scholl et al.4 studied adverse drug reactions based on proportionality analysis and found ways to improve the efficiency of data screening using predictive models. Thakker et al.5 analyzed the adverse drug reactions and found significant associations between the incidence, causative drugs, and other patient characteristics in India.

Materials and Methods

I use the three data sets from the FAERS database which is a collection of reports on adverse drug reactions by FDA. The three data sets record patient demographics, drug information related to the reported adverse drug reaction, and outcome of the reaction. The data ranges from 2012 to 2017. I merge the three data sets to create a feature set for each patient tracked by a primary id. There are 12,228 patients, and 47% of them are male.

I then build unsupervised learning model using k-means clustering. This model does not predict the ADR outcome, but rather finds whether there exist distinctive clusters of patients based on patient demographics and drug prescription. The data set used is the same, covering years 2012 to 2017. I find the optimal number of clusters k using the elbow method. The idea is to build multiple clustering models by starting with k =1 and incrementing k each time. Each clustering model is evaluated by calculating variation within each cluster using the sum of squared errors. The within-cluster variance decreases as k increases. The optimal k is chosen to be the one after which there is no significant reduction in variation. Once the groups are found, I look for special characteristics of each group that distinguish it from other groups. The number of patients in each cluster is also known. The clusters with high number of occurrences should received more attention.

Results

For the clustering model, the optimal number of clusters was found to be 5. This is shown in Figure 1, which is a plot of the variance among all instances in each cluster, when k clusters are defined. Starting by assuming there is 1 cluster, the variance within that cluster across all features is calculated. The same is repeated assuming there are 2 clusters, 3 clusters, and so on. 5 was found to be the optimal because it is the lasts number k which brings significant reduction in the variance. In other words, beyond k = 5, the improvement is not too significant. The table below shows the values of the features within each cluster. The results help characterize each cluster by looking for distinctive features.

DOI: 10.23958/ijirms/vol03-i01/01

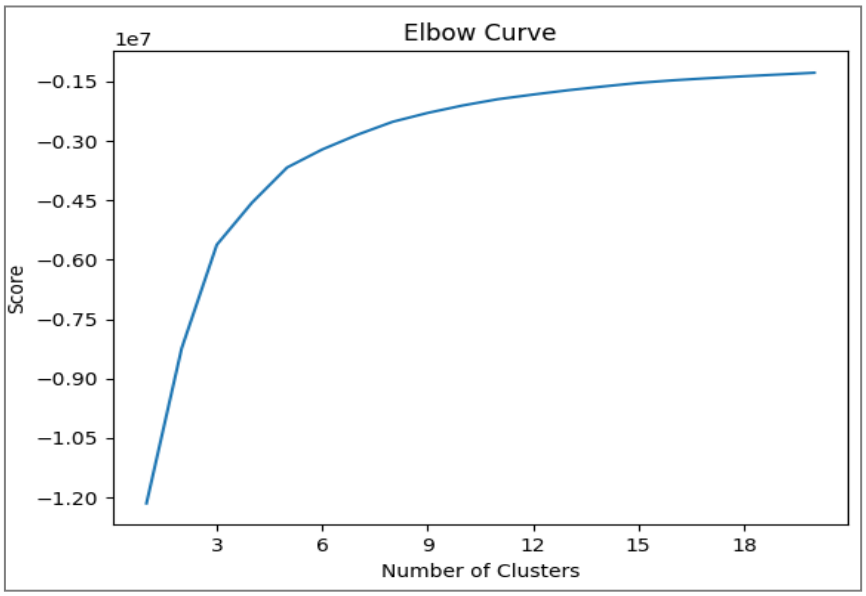

Figure 1: Within-cluster variance for different number of clusters

Table 1: Features within each cluster

Feature	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Count	2644	1366	389	3975	3854
Average Age	39.4	54.0	5.7	73.0	66.0
Average Weight (kg)	65.3	115.9	17.7	56.7	83.4
Sex Percentage (M/F)	36%/64%	62%/38%	54%/46%	35%/65%	59%/41%
Most Common Drug 1	Gilenya(19%)	Nexium(11%)	Singulair(15%)	Nexium(9%)	Nexium(9%)
Most Common Drug 2	Prednisolone(6%)	Metformin(7%)	Soliris(10%)	Xarelto(8%)	Xarelto(7%)
Most Common Drug 3	Ribavirin(5%)	Gilenya(6%)	Citalopram(8%)	Eliquis(8%)	Omeprazole(7%)
Most Common Drug 4	Omeprazole (5%)	Xarelto(6%)	Prednisolone(7%)	Omeprazole(6%)	Crestor(7%)
Most Common Drug 5	Nexium(4%)	Omeprazole(5%)	Methotrexate(6%)	Ribavirin(6%)	Atorvastatin(7%)
Most Common Country1	USA (22%)	USA (47%)	USA (25%)	Japan (40%)	USA (33%)
Most Common Country2	GB (20%)	GB (31%)	Germany (20%)	Germany(18%)	GB (30%)
Most Common Country3	Japan (18%)	Germany (4%)	Japan (17%)	USA (14%)	Germany (6%)

Discussion

From the results, the most distinctive feature is the combination age and weight. Cluster 1 consists of mostly middle-aged females. The country of origin is even between United States, Great Britain, and Japan. This group suffers predominantly from Gilenya, which is often used to treat sclerosis, or stiffening of body tissues. The other drugs have similar percentage of occurrences in this group. Cluster 2 consists of mostly overweight and predominantly male patients using Nexium, an antacid for treating stomach problems such as ulcers. They are mostly from the United States and Great Britain. Cluster 3 is the youngest group with an average age of only 5.7 years. They suffer mostly from ADRs due to Singulair (a drug for asthma) and Soliris (a drug for blood disorders). Cluster 4 is mainly female Japanese elders who use a variety of drugs with equal occurrences such as Nexium, Xarelto, and Eliquis. Both Xarelto and Eliquis are drugs for preventing blood clots. Cluster 5 is also mostly elders who suffer from ADRs related to Nexium and Xarelto most. The difference from cluster 4 is that cluster 5 is predominantly male from United States and Great Britain.

Conclusion

This paper presents the results of a clustering model which groups patients with adverse drug reactions into 5 groups. Each group can be characterized by a distinguishable set of features. The more frequent cases require more attention from health care practitioners when treating the patients and prescribing drugs. The cluster model can also be updated when more data become available to give the most updated results.

References

[1] Ferner RE, Akronson JK. National differences in publishing papers on adverse drug reactions. Br J Clin Pharmacol. 2005 Jan; 59(1): 108–111.
[2] Alomar MJ. Factors affecting the development of adverse drug reactions (Review article). Saudi Pharm J. 2014 Apr; 22(2):83-94.

DOI: 10.23958/ijirms/vol03-i01/01

[3] Maria K, Kaye KI, Davis S, Shenfield GM. Factors Influencing Adverse Drug Reaction Reporting in New South Wales Teaching Hospitals. J Pharm Pract Res 2004; 34: 32–35.
[4] Scholl JHG, van Hunsel FPAM, Hak E, van Puijenbroek EP. A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. Pharmacoepidemiol Drug Saf. 2017 Dec 21.
[5] Thakkar S, Patel TK, Vahora R, Bhabhor P, Patel R. Cutaneous Adverse Drug Reactions in a Tertiary Care Teaching Hospital in India: An Intensive Monitoring Study. Indian J Dermatol. 2017 Nov-Dec; 62(6):618-625.

DOI: 10.23958/ijirms/vol03-i01/01