Scientific
Research
Publishing

Identifying Association Rules among Drugs in Prescription of a Single Drugstore Using Apriori Method

Ahmad Yoosofan[1], Fatemeh Ghovanlooy Ghajar[2], Sima Ayat[1], Somayeh Hamidi[1], Farshad Mahini[3]

[1]Electrical and Computer Engineering Department, University of Kashan, Kashan, Iran
[2]Electrical and Computer Engineering Department, Isfahan University of Technology, Isfahan, Iran
[3]Department of Mechanical Engineering, Villanova University, Villanova, PA, USA
Email: yoosofan@kashanu.ac.ir

Abstract

These days, health care systems such as pharmacies and drugstores normally produce high volumes of data. Consequently, utilizing data mining methods in health care systems has become a conventional process. In this research, Apriori algorithm has been applied to perform data mining using the data obtained from the prescriptions ordered within a pharmacy. Ten association rules were achieved from the assigned pharmaceutical drugs in those prescriptions using the aforementioned Apriori algorithm. The accuracy of these rules is also manually studied and reviewed by a physician. Among these association rules, Vitamin D and Calcium pills are the most interrelated medications, and Omeprazole and Metronidazole rankd second in terms of association. The results of this study provide useful feedback information about associations among drugs.

1. Introduction

Increasing growth of data in different fields and the need for data analysis to obtain useful information and results have made researchers face diverse difficulties. Data collection, by itself, will not simply lead to obtaining useful results. Therefore, it is necessary to treat the data as a raw material in order to extract the useful information by performing an analytical methodology. Also, it is difficult to obtain useful results from databases ma-

nually or visually without the aid of computers or powerful analytical tools [1]. Knowledge discovery in database (KDD) is a general approach for the analysis and extraction of useful information from databases using fully automatic methods [2]. KDD is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results. Major KDD application areas include but are not limited to marketing, telecommunication and manufacturing.

Ultimately, the concept of KDD has been expanded to include all the stages of knowledge discovery. Data mining refers to a stage of KDD, in which learning algorithms are applied to data. In fact, this concept has been expanded to the entire stages of knowledge discovery over time and after utilizing data mining; therefore, data mining may refer to the entire process of knowledge discovery as well [3] [4].

Data mining has been applied within health care systems in different aspects, and its use has been also emphasized [5]-[8]. In particular, some studies have been conducted on pharmaceutical data [9]. In general, medical data are unique in comparison with the data generated in other fields. Thus, they require more precision because of fact that prescriptions contain individuals' personal and private information, in which simple mistakes can cause considerable ethical issues [10]. It has been observed that some problems have been encountered as a result of the increasing use of technology, particularly Internet based prescriptions [11]; therefore, its consequences must be considered to reduce the scope of these problems. World Health Organization has stated that many of the prescriptions in developing countries are not properly filled [12]. Moreover, incorrect or excessive drug consumption is related to the individuals' mentality who simply assumes that more drug consumption promises faster recovery or medical treatment.

In this paper, attempts have been made to clarify what drug items within the medical prescriptions brought to the drugstores are inter-dependent. To achieve this goal, a completely real database, from which valid information can be extracted, has to be available. In this regard, no such database that can be easily available for researchers is found in Iran. As a result, the information of a drugstore is manually collected, which is an extremely time consuming task.

2. Background

During the recent years, many papers have been published on purchase portfolio analysis, and newer methods for the optimal sale of drug items in pharmacies have been proposed [13] [14]. There are also methods which express selected maximum profit with sale difference [15]. For example, ChEMBL is a large database of information which has medical properties and has been manually collected [16]. To extract useful information, the hypotheses should be tested in databases and then error existence should be tracked [17]. The cases which should be solved can be considered as data mining issues, in which makes the pattern detection a very important tool in this regard [18]. In most of the mentioned papers [13]-[17], purchase portfolio analysis has been more popular. However, Lv *et al.* [19] dealt with a special software system, in which the symptoms given by users/ patients were the inputs. The system could then prescribe drugs based on the symptoms. With the volunteer assistance, users could also perform the primary evaluation of new drugs recommend by others [19].

Food and Drug Administration (FDA) in the United States has also used information about the side effects of drugs in its database using data mining methods. To this end, multi-item gamma poisson shrinker (MGPS) was applied, which could identify the side effects of drugs five years earlier than the traditional method with 67% success rate [20].

3. Goals and Advantages of the Proposed System

The aim of the research conducted in this paper is to design and implement a system, which could find the association among drugs in a pharmacy. Finding such an association among the drugs has the following advantages:
- Helping with the internal design and suitable arrangement of drugs in pharmacies;
- Considering discount by insurance companies for the drugs of a cluster;
- Preparing related drugs in order to minimize or to avoid missing drug prescriptions; because, most clients would normally purchase drugs from the pharmacies if all the drugs in the prescription are available there.

In this paper, the data mining algorithm and methodology that was applied in this research is explained in the first step. Then, the implementation of this algorithm on the data set is explained, and finally, the obtained rules and results are investigated.

4. Research Method

To find the associations among the drugs, there should be a complete database in order to extract the necessary information. This database should contain the name, type, code, dosage, and conditions of each drug. In a summary, the database of this study includes 3444 drug items (medications). To construct such a database, prescriptions were manually collected in a drugstore. Then, the software code was written to be used for entering these prescriptions into the database. In the next step, the data inside this database were converted into a single format in order for the data mining software applied in this research to be able to easily work on it. Then, association rules were extracted from these data sets. For this purpose, SPSS Clementine Software, which included different data mining algorithms, was applied. Apriori algorithm which is one of the algorithms for finding association rules, was then applied in this research. Advantage of this algorithm is that it reduces searching problems to a controllable and manageable size. It is also extremely useful for reducing searching space [21].

4.1. Apriori Algorithm

This algorithm is one of the conventional algorithms to find association rules among the data inside a database or dataset. These rules are mostly found based on transactions and items inside a database. In this discussion, item refers to a set of interrelated data, which conveys a concept (object or entity), among which some associations are supposed to be found. In fact, an item can be single member and only include one piece of data. A set of items which are put beside each other and construct a work unit with a record is called transaction. For example, in a store, a purchase portfolio of a customer from the store is a transaction and the purchased items inside the purchase portfolio are its items. Each of these items contain one or more pieces of data, which can be item number, item name/identity, and item price for the merchandise inside the shop.

This algorithm works in the following two steps:
- Finding conventional item sets;
- Constructing association rules based on the found sets;

The two following hypotheses are considered in this algorithm:
- Each subset of an iterative item set is iterative. If set $\{a, b, c\}$ is assumed to be iterative, then set $\{a, b\}$ is also iterative;
- Each hyperset of a non-iterative item set is non-iterative. If set $\{a, b\}$ is assumed to be non-iterative, then set $\{a, b, c\}$ is also non-iterative.

Apriori algorithm constructs a series of large item-set with length of $K + 1$ from the selected item-sets with length of K in each time and continues until an item-set with the longest length is achieved, provided that its support exceeds the required threshold.

4.1.1. Support

Support refers to the probability of the existence of both antecedent and subsequent in a transaction. In this discussion, transaction means a prescription. The term "tuple" is sometimes used instead of transaction. In other words, rule "support" indicates the ratio of the transactions including both sets A and B to total transactions available in set D (1),

$$\text{Support}(A \Rightarrow B) = \frac{\text{Number of transactions containing both } A \text{ and } B}{\text{Total number of transactions}}$$

4.1.2. Confidence

Confidence indicates accuracy or truth of rule (2),

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Number of transactions containing both } A \text{ and } B}{\text{Number of transactions containing } A}$$

5. Applying Apriori Algorithm

Input data of this research were converted into a textual file, which was then applied as the input source in Clementine. The stored data was read from the file and sent to a stream. Ten rules were obtained after the implementation of our proposed Apriori algorithm, which were shown as antecedent and subsequent. Association

among the drugs is graphically shown in **Figure 1**. In this figure, stronger associations are shown by bolder lines and weaker ones are indicated using dotted lines.

6. Results

Table 1 shows the obtained association rules. Then, the obtained rules are studied.

Rules 1 and 2:

(1) If Vitamin D3 *then* Calcium-D tablet;

(2) If Calcium-D *then* Vitamin D3.

The obtained results show that both Vitamin D and Calcium had maximum confidence coefficient and were given on the topmost row of the table. Both of these drugs were mentioned as antecedent and subsequent. Vitamin D and Calcium are two vital substances in the human body. Calcium plays an effective role in the strength
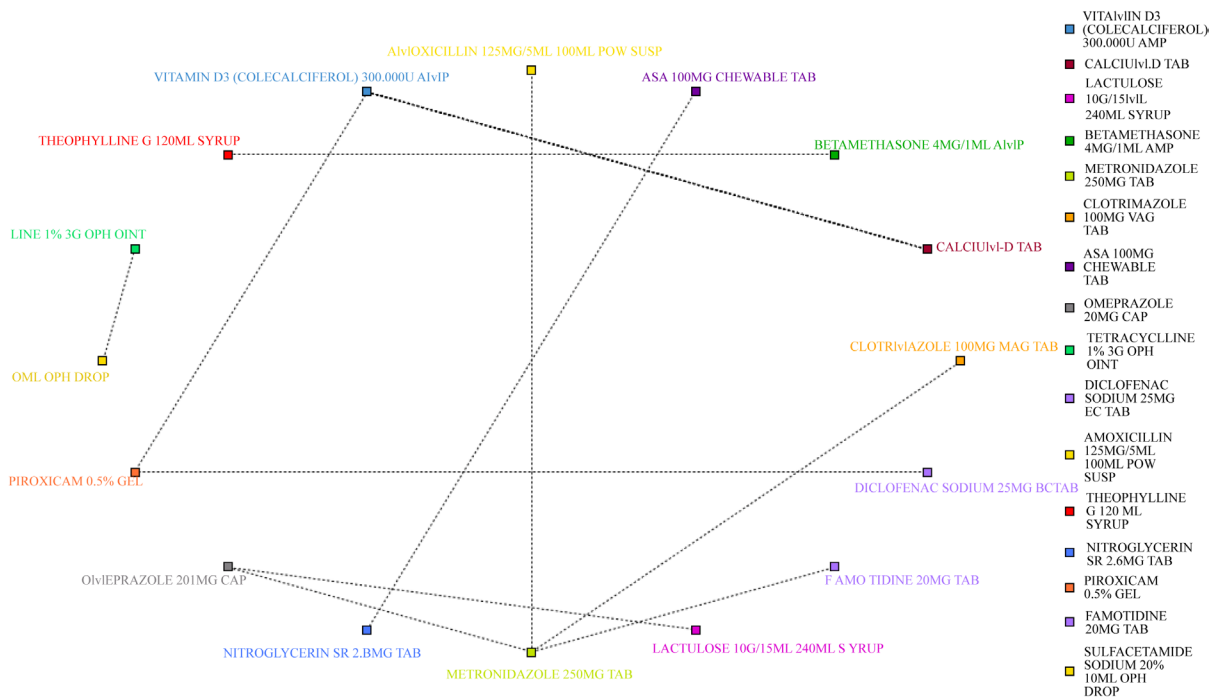

Figure 1. Association graph of drugs.

Table 1. Results of Apriori algorithm.

	support	Confidence	Antecedent	Consequent
1	4.0	33.3	Vitamin D3 (Colecalciferol)	**Calcium-D tablet**
2	4.3	30.8	Calcium-D tablet	**Vitamin D3 (Colecalciferol)**
3	4.7	28.6	Omeprazole 20 mg cap	**Metronidazole 250 mg tablet**
4	4.0	25.0	Vitamin D3 (Colecalciferol)	**Piroxicam 0.5% gel**
5	4.0	25.0	Betamethasone 4 mg/1ml amp	**Theophylline 120 g/ml syrup**
6	4.0	25.0	Amoxicillin 125 mg/5ml 100 ml	**Metronidazole 250 mg tablet**
7	5.7	23.0	ASA 100 mg chewable tablet	**Nitroglycerin sustained release 2.6 mg tablet**
8	4.7	21.0	Omeprazole 20 mg Cap	**Lactulose 10 g/15ml syrup**
9	6.7	20.0	Metronidazole 250 mg tablet	**Clotrimazole 100 mg vaginal tablet**
10	6.7	20.0	Metronidazole 250 mg tablet	**Omeprazole 20 mg cap**

of bones, teeth, blood coagulation, neural contractions, and muscles. Vitamin D also plays an effective role in the strength of bones and teeth as well as regulation of calcium balance in human body. Therefore, these two drugs are mostly prescribed together, because Vitamin D increases Calcium absorption through intestine and regulates Calcium required for body absorption.

Rules 3 and 10:

(3) If Omeprazole then Metronidazole

(10) If Metronidazole then Omeprazole

Metronidazole and Omeprazole are among the drugs which are prescribed for treatment of digestive problems. Metronidazole is an antibacterial drug and Omeprazole has an antiulcer effect. To cure digestive ulcers, these two drugs are mostly prescribed together. More specifically, if Metronidazole is prescribed, Omeprazole will be prescribed with the probability of 28.6%.

Rule 4:

(4) If Vitamin D3 then Piroxicam G

Piroxicam gel is an anti-inflammatory drug which is very effective for the alleviation of pain and muscular stiffness and inflammation. Vitamin D is very useful for strengthening bones, reinforcing nerves, curing contraction and cramp of muscles, and curing muscular weakness. This rule states that if Vitamin D is prescribed for a patient, then Piroxicam gel will be also prescribed with the confidence coefficient of 25%.

Rule 5:

(5) If Betamethasone then Thentheophylline G

Theophylline G is used for treatment of spasm of bronchi and prescribed together with Betamethasone, as an antihistamine drug to cure seasonal allergies, asthma, and bronchitis.

Rule 6:

(6) If Amoxicillin then Metronidazole

Amoxicillin is an antibiotic drug from penicillin group and is applied for curing many infections, particularly those in ear or respiratory tract. Metronidazole is also one of the antibacterial drugs, which is prescribed to cure infections.

Rule 7:

(7) If Aspirin (ASA) then Nitroglycerin

Aspirin (ASA) is applied for pain alleviation and fever reduction and is also effective in the prevention of blood coagulation. Nitroglycerine is a vessel dilating medication and is regarded as a member of antianginal drugs. Prescription of these two drugs reduces hypertension and is applied in the treatment of cardiovascular diseases.

Rule 8:

(8) If Omeprazole then Lactulose

Lactulose is one of the Laxatives, which is used together with Omeprazole. It has an anti-ulcer property to cure digestive diseases.

Rule 9:

(9) If Metronidazole then Clotrimazole

Clotrimazole is one of the classified antifungal drugs and is applied to treat yeast and fungal infections. This drug is also prescribed along with metronidazole which has an antibacterial effect to cure digestive diseases.

7. Conclusions

Obtaining association rules among drugs based on the prescriptions of a drugstore is the objective of this paper. Since drugs in a prescription are not put beside each other randomly but based on patient's illness and symptoms, the association between the drugs which are prescribed together can be found using the drugs inside prescriptions. To do so, a database of prescriptions and their drugs is collected from the sales orders in a drugstore. For this purpose, a program is written. Then, the format of these data is converted into a suitable format for Clementine software. Some of the association rules are identified by applying Apriori algorithm in this software. It is identified that Vitamin D and Calcium are the most interrelated drugs and are applied beside each other in most prescriptions. Also, Omeprazole and Metronidazole rank second in terms of association. Other association rules are found and discussed in the paper.

Application of other algorithms to find association rules of these data is one of the proposed ideas for contin-

uing the work on this paper. Dealing with other aspects of association among these drugs in prescriptions is also another recommendation. The most important problem addressed in this research is the data collection which is done with difficulty, because the participating drugstore in our tests is used to keep prescriptions manually. They have to be converted into computerized prescriptions. Therefore, it is suggested to collect the data of those drugstores which receive their prescriptions through computers and, if there are several drugstores, a comparison can be made among them. If data samples from drugstores in different geographical regions are collected, a better comparative result can be obtained. Since a comprehensive result cannot be obtained from the prescriptions of a drugstore.

Acknowledgements

References

[1] Lin, C., Lin, C.-M., Li, S.-T. and Kuo, S.-C. (2008) Intelligent Physician Segmentation and Management Based on KDD Approach. *Expert Systems with Applications*, **34**, 1963-1973. http://dx.doi.org/10.1016/j.eswa.2007.02.038

[2] Ryu, T.-W. and Eick, C.F. (2005) A Database Clustering Methodology and Tool. *Information Sciences*, **171**, 29-59. http://dx.doi.org/10.1016/j.ins.2004.03.016

[3] Giudici, P. (2003) Applied Data Mining: Statistical Methods for Business and Industry. Wiley, Hoboken.

[4] Giudici, P., Figini, S. and Giudici, P. (2009) Applied Data Mining for Business and Industry. Wiley, Hoboken. http://dx.doi.org/10.1002/9780470745830

[5] Koh, H.C. and Tan, G. (2011) Data Mining Applications in Healthcare. *Journal of Healthcare Information Management*, **19**, 65.

[6] Ogwueleka, F. (2009) Data Mining Applications in Healthcare. *International Journal of Natural and Applies Sciences*, **5**(1).

[7] Rosset, S., Perlich, C., Świrszcz, G., Melville, P. and Liu, Y. (2010) Medical Data Mining: Insights from Winning Two Competitions. *Data Mining and Knowledge Discovery*, **20**, 439-468. http://dx.doi.org/10.1007/s10618-009-0158-x

[8] Moghaddassi, H., Hoseini, A., Asadi, F. and Jahanbakhsh, M. (2012) Application of Data Mining. *Health Information Management*, **9**, 297-304.

[9] Almenoff, J., Tonning, J., Gould, A.L., Szarfman, A., Hauben, M., Ouellet-Hellstrom, R., *et al*. (2005) Perspectives on the Use of Data Mining in Pharmaco-Vigilance. *Drug Safety*, **28**, 981-1007. http://dx.doi.org/10.2165/00002018-200528110-00002

[10] Cios, K.J. and William Moore, G. (2002) Uniqueness of Medical Data Mining. *Artificial Intelligence in Medicine*, **26**, 1-24. http://dx.doi.org/10.1016/S0933-3657(02)00049-0

[11] Jena, A.B. and Goldman, D.P. (2011) Growing Internet Use May Help Explain the Rise in Prescription Drug Abuse in The United States. *Health Affairs*, **30**, 1192-1199. http://dx.doi.org/10.1377/hlthaff.2011.0155

[12] Adebayo, E. and Hussain, N. (2010) Pattern of Prescription Drug Use in Nigerian Army Hospitals. *Annals of African Medicine*, **9**, 152-158. http://dx.doi.org/10.4103/1596-3519.68366

[13] Hamuro, Y., Kawata, H., Katoh, N. and Yada, K. (2002) A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History. In: Arikawa, S. and Shinohara, A., Eds., *Progress in Discovery Science*, Vol. 2285, Springer Berlin Heidelberg, Berlin, 565-575.

[14] Wong, R.C.-W., Fu, A.W.-C. and Wang, K. (2003) MPIS: Maximal-Profit Item Selection with Cross-Selling Considerations. 3*rd IEEE International Conference on Data Mining*, *ICDM*, Melbourne, 19-22 November 2003, 371-378. http://dx.doi.org/10.1109/ICDM.2003.1250942

[15] Breiman, L. (1996) Bias, Variance, and Arcing Classifiers. Technical Report No. 460, University of California, Berkeley.

[16] Warr, W.A. (2009) ChEMBL. An Interview with John Overington, Team Leader Chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). *Journal of Computer-Aided Molecular Design*, **23**, 195-198. http://dx.doi.org/10.1007/s10822-009-9260-9

[17] Hamuro, Y., Katoh, N., Matsuda, Y. and Yada, K. (1998) Mining Pharmacy Data Helps to Make Profits. *Data Mining and Knowledge Discovery*, **2**, 391-398. http://dx.doi.org/10.1023/A:1009748731133

[18] Reinhold, D. (2005) Market Basket Analysis by Means of a Growing Neural Network. *The International Review of*

Retail, *Distribution and Consumer Research*, **15**, 151-169. http://dx.doi.org/10.1080/09593960500049332

[19] Lv, S.Y. and Wang, F.D. (2011) Review: Recommender System in E-Commercial Environment. *International Conference on E-Business and E-Government* (*ICEE*), Shanghai, 6-8 May 2011, 1-4. http://dx.doi.org/10.1109/icebeg.2011.5881715

[20] Ruben, D. and Canlas, R.D. (2009) Data Mining in Healthcare: Current Applications and Issues. Master's Thesis, Carnegie Mellon University, Pittsburgh.

[21] Larose, D.T. (2004) Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Hoboken. http://dx.doi.org/10.1002/0471687545