Detect Adverse Drug Reactions for Drug Aspirin

Yihui Liu and Uwe Aickelin

Abstract—Adverse drug reaction (ADR) is widely concerned for public health issue. In this study we propose an original approach to detect the ADRs using feature matrix and feature selection. The experiments are carried out on the drug Aspirin. Major side effects for the drug are detected and better performance is achieved compared to other computerized methods. The detected ADRs are based on the computerized method, further investigation is needed.

1. INTRODUCTION

ADR—Adverse drug reaction (ADR) is widely concerned for public health issue. ADRs are one of most common causes to withdraw some drugs from market [1]. Now two major methods for detecting ADRs are spontaneous reporting system (SRS) [2], [3], and prescription event monitoring (PEM)[4],[5]. The World Health Organization (WHO) defines a signal in pharmacovigilance as "any reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously" [6]. For spontaneous reporting system, many machine learning methods are used to detect ADRs, such as Bayesian confidence propagation neural network (BCPNN) [7], decision support method [8], genetic algorithm [9], knowledge based approach [10], etc. One limitation is the reporting mechanism to submit ADR reports, which has serious underreporting and is not able to accurately quantify the corresponding risk. Another limitation is hard to detect ADRs with small number of occurrences of each drug-event association in the database.

In this paper we propose feature selection approach to detect ADRs from The Health Improvement Network (THIN) database. First feature matrix, which represents the medical events for the patients before and after taking drugs, is created by linking patients' prescriptions and medical events for the patients before and after taking drugs. The time period of observation is set to 60 days. Matrix $X$ and $Y$ are feature matrix after patients are divided into groups. We set 100 patients as one group.

II. FEATURE MATRIX AND FEATURE SELECTION

A. The Extraction Of Feature Matrix

To detect the ADRs of drugs, first feature matrix is extracted from THIN database, which describes the medical events that patients occur before or after taking drugs. Then feature selection method of Student's t-test is performed to select the significant features from feature matrix containing thousands of medical events. Figure 1 shows the process to detect the ADRs using feature matrix. Feature matrix $A$ describes the medical events for each patient during 60 days before they take drugs. Feature matrix $B$ reflects the medical events during 60 days after patients take drugs. In order to reduce the effect of the small events, and save the computation time and space, we set 100 patients as a group. Matrix $X$ and $Y$ are feature matrix after patients are divided into groups.

B. Medical Events And Readcodes

Medical events or symptoms are represented by medical codes or Readcodes. There are 103387 types of medical events in "Readcodes" database. The Read Codes used in general practice (GP), were invented and developed by Dr James Read in 1982. The NHS (National Health Service) has expanded the codes to cover all areas of clinical practice. The code is hierarchical from left to right or from level 1 to level 5. It means that it gives more detailed information from level 1 to level 5. Table 1 shows the medical symptoms based on Readcodes at level 3 and at level 5. 'Other soft tissue disorders' is general term using Readcodes at level 3. 'Foot pain', 'Heel pain', etc., give more details using Readcodes at level 5.
C. Feature Selection Based On Student's t-test

Feature extraction and feature selection are widely used in biomedical data processing [11]-[18]. In our research we use feature selection method of Student's t-test [19] to detect the significant ADRs from thousands of medical events. Student's t-test is a kind of statistical hypothesis test based on a normal distribution, and is used to measure the difference between two kinds of samples.

Based on Readcodes at level 1-5, totally 14546 medical events are obtained. Removing the patients registered in GP less than 12 months. The variable of ratio \( R \) is defined to evaluate significant changes of the medical events, using ratio of the patient number after taking the drug to one before taking the drug. The variable \( R \) represents the ratio of patient number after taking the drug to the number of whole population having one particular medical symptom.

The ratio variables \( R_1 \) and \( R_2 \) are defined as follows:

\[
R_1 = \frac{N_1}{N_2} \quad \text{if } N_2 \neq 0;
\]
\[
R_2 = \frac{N_1}{N_2} \quad \text{if } N_2 = 0;
\]

\[
R_2 = \frac{100 \times N_1}{N_2}, \quad \text{I N}
\]

where \( N_1 \) and \( N_2 \) represent the numbers of patients before or after they take drugs for having one particular medical event respectively. The variable \( N \) represents the number of whole population who take drugs.

D. Other Parameters

The variable of ratio \( R \) is defined to evaluate significant changes of the medical events, using ratio of the patient number after taking the drug to one before taking the drug. The variable \( R \) represents the ratio of patient number after taking the drug to the number of whole population having one particular medical symptom.

The ratio variables \( R_1 \) and \( R_2 \) are defined as follows:

\[
R_1 = \frac{N_1}{N_2} \quad \text{if } N_2 \neq 0;
\]
\[
R_2 = \frac{N_1}{N_2} \quad \text{if } N_2 = 0;
\]

\[
R_2 = \frac{100 \times N_1}{N_2}, \quad \text{I N}
\]

where \( N_1 \) and \( N_2 \) represent the numbers of patients before or after they take drugs for having one particular medical event respectively. The variable \( N \) represents the number of whole population who take drugs.

| Level Readcodes | Medical events |
|-----------------|----------------|
| Level 3         | Muscle Level 3 |
|                 | N24.00 Other soft tissue disorders pain |
|                 | N245.16 Leg pain |
|                 | N245111 Toe pain |
| Level 5         | N245.13 Foot pain |
|                 | N245700 Shoulder pain |
|                 | N245.15 Heel pain |

TABLE 1. MEDICAL EVENTS BASED ON READCODES AT LEVEL 3 AND LEVEL 5.

In this study we propose a novel method to successfully detect the ADRs using feature matrix and feature selection. A feature matrix, which characterizes the medical events before patients take drugs or after patients take drugs, is created from THIN database. The feature selection method of Student's t-test is used to detect the significant features from thousands of medical events. The significant ADRs, which are corresponding to significant features, are detected. Experiments are performed on the drug Aspirin. The detected ADRs are based on the computerized method, further investigation is needed.

IV. CONCLUSIONS

In this study we propose a novel method to successfully detect the ADRs using feature matrix and feature selection. A feature matrix, which characterizes the medical events before patients take drugs or after patients take drugs, is created from THIN database. The feature selection method of Student's t-test is used to detect the significant features from thousands of medical events. The significant ADRs, which are corresponding to significant features, are detected. Experiments are performed on the drug Aspirin. The detected ADRs are based on the computerized method, further investigation is needed.

REFERENCES

[1] G. Severino and M.D. Zompo, "Adverse drug reactions: role of pharmacogenomics," Pharmacological Research, vol. 49, pp. 363-373, 2004.
[2] Y. Qian, X Ye, W. Du, J. Ren, Y. Sun, H. Wang, B. Luo, Q. Gao, M. Wu, and J. He, "A computerized system for signal detection in spontaneous reporting system of Shanghai China," Pharmacoeconomics and Drug Safety, vol. 18, pp. 154-158, 2009.
[3] K.S. Park and 0. Kwon, "The state of adverse event reporting and signal generation of dietary supplements in Korea," Regulatory Toxicology and Pharmacology, vol. 57, pp. 74-77, 2010.
[4] R. Kasliwal, L.V. Wilton, V. Cornelius, B. Aurich-Barrera, and S.A.W. Shakir, "Safety profile of Rosuvastatin-results of a prescription-event monitoring study of 11680 patients," Drug Safety, vol. 30, pp. 157-170, 2007.
[5] R.D. Mann, K. Kubota, G. Pearce, and L. Wilton, "Salmeterol: a study by prescription-event monitoring in a UK cohort of 15,407 patients," J Clin Epidemiol, vol. 49, pp. 247-250, 1996.
[6] R.H. Meyboom, M. Lindquist, A.C. Egberts, and I.R. Edwards, "Signal detection and follow-up in pharmacovigilance," Drug Saf, vol. 25, pp. 459-465, 2002.
[7] A. Bate, M. Lindquist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner, and R.M. De Freitas, "A Bayesian neural network method for adverse drug reaction signal generation," Eur J Clin Pharmacol, vol. 54, pp. 315-321, 1998.
[8] H. Hauben and A. Bate, "Decision support methods for the detection of adverse events in post-marketing data," Drug Discovery Today, vol. 14, pp. 343-357, 2009.
[9] Y. Koh, C.W. Yap, and S.C. Li, "A qualitative approach of using genetic algorithm in designing a probability scoring system of an adverse drug reaction assessment system," International Journal of Medical Informatics, vol. 77, pp. 421-430, 2008.
[10] C. Henegar, C. Bousquet, A.L. Lillo-Le, P. Degoulet, and M.C. Jaunet, "A knowledge based approach for automated signal generation in pharmacovigilance," Stud Health Technol Inform, vol. 107, pp. 626-630, 2004.
[11] Y. Liu, "Feature extraction and dimensionality reduction for mass spectrometry data," Computers in Biology and Medicine, vol. 39, pp. 818-823, 2009.
TABLE II THE TOP POTENTIAL ADRS FOR ASPIRIN BASED ON P-VALUE OF STUDENT'S T-TEST.

| Rank | Reacodes | Medical events | NB   | NA  | R1   | R2   |
|------|----------|----------------|------|-----|------|------|
| 1    | 127.50   | Chronic kidney disease stage 3 | 177  | 1169 | 6.60 | 6.83 |
| 2    | M03.0000 | Cellulitis NOS | 178  | 711  | 3.99 | 4.15 |
| 3    | F4C.00    | Acute conjunctivitis | 200  | 702  | 3.51 | 4.10 |
| 4    | H06.0000  | Chest infection NOS | 339  | 1392 | 4.11 | 8.13 |
| 5    | N13.10    | Cervicalgia - pain in neck | 206  | 660  | 3.20 | 3.85 |

Level 1-5

| Rank | Reacodes | Medical events | NB   | NA  | R1   | R2   |
|------|----------|----------------|------|-----|------|------|
| 6    | K197.00  | Haematuria     | 55   | 304  | 5.53 | 1.78 |
| 7    | N143.00  | Scintatica     | 126  | 409  | 3.25 | 2.39 |
| 8    | K190.00  | Urinary tract infection, site not specified | 298  | 1010 | 3.39 | 5.90 |
| 9    | F46.00   | Cataract       | 115  | 475  | 4.13 | 2.77 |
| 10   | C34.00   | Gout           | 134  | 421  | 3.14 | 2.46 |
| 11   | 1M10.00  | Knee pain      | 225  | 752  | 3.34 | 4.39 |
| 12   | 1992.00  | Vomiting       | 87   | 419  | 4.82 | 2.45 |
| 13   | H060.00  | Acute bronchitis | 222  | 791  | 3.56 | 4.62 |
| 14   | N245.17  | Shoulder pain  | 212  | 710  | 3.35 | 4.15 |
| 15   | M0..00   | Skin and subcutaneous tissue infections | 63   | 287  | 4.56 | 1.68 |
| 16   | D00..00  | Iron deficiency anaemia | 55   | 273  | 4.96 | 1.59 |
| 17   | F45..00  | Glaucoma       | 64   | 236  | 3.69 | 1.38 |
| 18   | F4D.00   | Blepharitis    | 59   | 249  | 4.22 | 1.45 |
| 19   | F501.00  | Infective otitis externa | 105  | 357  | 3.40 | 2.60 |
| 20   | J510.00  | Diverticulosis | 38   | 195  | 5.13 | 1.14 |

Level 1-3

| Rank | Reacodes | Medical events | NB   | NA  | R1   | R2   |
|------|----------|----------------|------|-----|------|------|
| 1    | 121.00   | Chronic renal impairment | 206  | 1389 | 6.74 | 8.11 |
| 2    | H06.00   | Acute bronchitis and bronchiolitis | 944  | 2930 | 3.10 | 17.11 |

TABLE III THE TOP POTENTIAL ADRS FOR ASPIRIN BASED ON DESCENDING ORDER OF R1 VALUE.

| Rank | Reacodes | Medical events | NB   | NA  | R1   | R2   |
|------|----------|----------------|------|-----|------|------|
| 1    | 1Z1E.00  | Chronic kidney disease stage 3A without proteinuria | 0    | 38  | 38.00 | 0.22 |
| 2    | F4E2.00  | Epiphora       | 1    | 30  | 30.00 | 0.18 |
| 3    | F591.00  | Sensorineural hearing loss | 1    | 27  | 27.00 | 0.16 |
| 4    | SK160000 | Other hip injuries | 1    | 22  | 22.00 | 0.13 |
| 5    | 1A45.12  | Haematuria - symptom | 1    | 20  | 20.00 | 0.12 |
| 6    | F4A4.00  | Corneal ulcer. | 0    | 19  | 19.00 | 0.11 |
| 7    | 165.05   | Facial swelling | 1    | 19  | 19.00 | 0.11 |
| 8    | M12.12   | Contact eczema | 0    | 18  | 18.00 | 0.11 |
| Rank | Readcodes | Medical events | NB | NA | R1  | R2   |
|------|-----------|----------------|----|----|-----|------|
| 1    | B33..00   | Other malignant neoplasm of skin | 85 | 334 | 3.93 | 1.95 |
| 2    | B76..00   | Benign neoplasm of skin | 76 | 224 | 2.95 | 1.31 |
| 3    | B34..00   | Malignant neoplasm of female breast | 19 | 99 | 5.21 | 0.58 |
| 4    | B22..00   | Malignant neoplasm of trachea, bronchus and lung | 14 | 80 | 5.71 | 0.47 |
| 5    | B46..00   | Malignant neoplasm of prostate | 57 | 163 | 2.86 | 0.95 |
| 6    | BB2..00   | [M]Papillary and squamous cell neoplasms | 15 | 86 | 5.73 | 0.50 |
| 7    | B8..00    | Carcinoma in situ | 16 | 58 | 3.63 | 0.34 |
| 8    | B13..00   | Malignant neoplasm of colon | 12 | 45 | 3.75 | 0.26 |
| 9    | B32..00   | Malignant melanoma of skin | 4 | 29 | 7.25 | 0.17 |
| 10   | J01..00   | Suspected malignancy | 11 | 51 | 6.64 | 0.30 |
| 11   | B14..00   | Malignant neoplasm of rectum, rectosigmoid junction and anus | 5 | 31 | 6.20 | 0.18 |
| 12   | BB3..00   | [M]Basal cell neoplasms | 7 | 37 | 5.29 | 0.22 |
| 13   | B49..00   | Malignant neoplasm of urinary bladder | 8 | 31 | 3.88 | 0.18 |
| 14   | B10..00   | Malignant neoplasm of oesophagus | 4 | 22 | 5.50 | 0.13 |
| 15   | B59..00   | Malignant neoplasm of unspecified site | 3 | 26 | 8.67 | 0.15 |
| 16   | BB5..00   | [M]Adenomas and adenocarcinomas | 19 | 54 | 2.84 | 0.32 |
| 17   | B17..00   | Malignant neoplasm of pancreas | 3 | 19 | 6.33 | 0.11 |
| 18   | B7F..00   | Benign neoplasm of brain and other parts of nervous system | 0 | 10 | 10.00 | 0.06 |
| 19   | B58..00   | Secondary malignant neoplasm of other specified sites | 3 | 18 | 6.00 | 0.11 |
| 20   | B43..00   | Malignant neoplasm of body of uterus | 2 | 15 | 7.50 | 0.09 |
| 21   | B57..00   | Secondary malig neop of respiratory and digestive systems | 3 | 17 | 5.67 | 0.10 |
| 22   | B81..00   | Carcinoma in situ of respiratory system | 1 | 10 | 10.00 | 0.06 |
| 23   | B82..00   | Carcinoma in situ of skin | 0 | 7 | 7.00 | 0.04 |
| 24   | BB4..00   | [M]Transitional cell papillomas and carcinomas | 6 | 19 | 3.17 | 0.11 |
| 25   | B44..00   | Malignant neoplasm of ovary and other uterine adnexa | 2 | 10 | 5.00 | 0.06 |
| 26   | B83..00   | Carcinoma in situ of breast and genitourinary system | 15 | 32 | 2.13 | 0.19 |
| 27   | B80..00   | Carcinoma in situ of digestive organs | 3 | 11 | 3.67 | 0.32 |
| 28   | B51..00   | Malignant neoplasm of brain | 1 | 7 | 7.00 | 0.04 |
| 29   | B11..00   | Malignant neoplasm of stomach | 5 | 14 | 2.80 | 0.08 |
| 30   | BBG..00   | [M]Fibromatous neoplasms | 0 | 4 | 4.00 | 0.02 |