A COMPARISON OF TWO METHODS TO ESTABLISH DRUG-REACTION RELATIONSHIPS IN THE ADRAC DATABASE

M. A. Mamedov School of Information Technology
and Mathematical Sciences
University of Ballarat
Victoria 3353 Australia
Email: m.mammadov@ballarat.edu.au

G.W. Saunders School of Information Technology
and Mathematical Sciences
University of Ballarat
Victoria 3353 Australia
Email: g.saunders@ballarat.edu.au

ABSTRACT

In this paper we present drug-reaction relations in the form of weights which indicates the “probability” of occurrence of reactions. The comparison of two different methods for establishing such representations are made: one uses all drugs involved and the other method uses only suspected drugs reported.

INTRODUCTION

An Adverse Drug Reaction (ADR) is defined by the WHO as: “a response to a drug that is noxious and unintended and occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease, or for modification of physiological function” [10]. The authors of this paper are developing an alternative approach to the ADR problem [3, 4]. Some of the problems concerning ADRs are discussed in the research report [3].

One of the main problems of ADR is the following: given a patient (the set of drugs and reactions occurred) to identify drug(s) which are responsible for these reactions. Such drugs are termed, in the ADRAC database, as “suspected” drugs. The accurate definition of suspected drugs for each report has a very important effect on the quality of the database for future study of drug-reaction relationships.

The information collected in the ADRAC database consist of mainly two sets of information: Individual patient information, including “reason for use”, “history” and so on, and information about drug(s) and reactions observed. In this paper we will use only the second set of information, because the former requires more of a data mining approach to extract useful information.

Therefore, we consider drug-reaction relationships not involving any other patient information. In other words we define for each drug a vector of weights which indicate the probability of occurrence of each reaction. This problem can be considered as a text categorization problem, where each patient is considered as one document, and the set of drug(s) taken by this patient is considered as a text related to this document; that is, each drug is considered as a word. For a review of some of the issues in textcategorization (ee ([7, 11]).

ADRAC Data

TABLE 1: The Five Card2 Classes

	1010	Cardiovascular general	28	13512
	1020	Myo endo card. & valve	37	674
	1030	Heart rate & rhythm	21	810
	1040	Vascular (extra cardiac)	24	5901
	Non-cardiovascular reaction(s) ‡		6874	
	Total	677	36394	

* Number of reaction terms in class
† Total occurrence of reaction class -- see text
‡ Given any 10xx SOC + any non-cardiovascular reaction terms(s) – see text

The Australian Adverse Drug Reaction Advisory Committee (ADRAC)database has been developed and maintained by the Therapeutic Goods Administration (TGA) with the aim to detect signals from adverse drug reactions as early as possible. The ADRAC data contains 137,297 records collected from 1971 to 2001. A more detailed account of the ADRAC database is given in [3].

In ADRAC there are 18 Systems Organ Class (SOC) reaction term classes, one of which is

Cardiovascular. The Cardiovascular class consists of four sub-classes. In this paper we will consider the part of ADRAC data related to the cardiovascular type of reactions. All records having at least one reaction from these four sub-groups were collected. We call this dataset **Card20**. In this dataset some records may have a reaction from outside Cardiovascular group. In order to group reactions into reaction classes we define four classes according to these four sub-groups and plus a fifth class that contains reactions belonging to the other 17 SOCs. For some details seeTable 1 (for more details refer to [3]).

Statement of the problem

Let X denote the set of all patients and D denote the set of all drugs used by these patients. Let *c* be a finite number of possible reactions (classes). Given patient $x \in X$ we denote by $D(x)$ the set of drugs taken by this patient. In ADRAC data the number of drugs reported for a patient is restricted to 10. We also denote by

$Y(x) = (Y_1, Y_2, \ldots, Y_c)$ a c -dimensional vector of reactions observed for the patient x; where

$Y_i = 1$ if the reaction *i* has occurred, and

$Y_i = 0$ if it has not.

The goal of the study of drug-reaction relationships is to find a classifier

where given drug $d \in D$ the components h_i of the vector $h(d)=(h_1, h_2, \ldots, h_c)$ associate the “probabilities” of the occurrence of the reactions i = 1, 2, …, *c*. Here

is the set of all c -dimensional vectors with non-negative coordinates.

In the next step, given a set of drugs $\Delta \subset D$, we need to define a vector $H=(H_1, H_2, \ldots, H_c)$, where the component H_i indicates the probability of occurrence of the reaction *i* after taking the drugs). In other words, we need to define a function

where

$S(D)$ is the set of all subsets of D. The function *H* can be defined in different ways and it is an interesting problem in terms of ADR(s). We will discuss this problem below.

Given patient $x \in X$ and a set of drugs $D(x)$, we will use the notation $H(x) = H(D(x))$.

In this statement, this problem is a multi-class, multi-label text categorization problem, but there are some interesting points that should be mentioned in relation to ADRs. One of the main characteristics is that the number of drugs (i.e. words in the context of text categorization) for each patient is restricted to 10, and for majority of patients just one drug was used. This sparseness of data complicates learning and classification, but on the other hand, this allows us to introduce simple and fast algorithms..

Potential Reactions

The vectors *h*(*d*) show what kind of reactions are caused by the drugs $d \in D(x)$. Therefore the vector *H*(*x*) can be considered as potential reactions which could occur with the patient *x*. But what kind of reactions will occur? This will depend upon the individual characteristics of the patient as well as external factors. Different patients can have different predispositions for differentreactions. Some reactions which have potentially high degrees may not be observed because of the strong resistance of the patient to developing these reactions. The function *H* can be defined in different ways. The study of more sensible definitions of the function *H* is an interesting problem for future investigations. This problem is also related to the study of Interaction of Drugs [3]. In this paper we will not study this problem.

In the calculation below we will use the following linear function *H* [3]: $H = (H_1,\ldots,H_c)$; where, for each subset) ⊂ *D* the components H_i are defined as follows:

In this case, for each patient $x \in X$, we have $H(x) = (H_1(x),\ldots,H_c(x))$, where

(1.1)

The use of this function means that, we accumulate the effects from different drugs. For example, if $h_i(d_n) = 0.2$ (n=1,2) for some reaction *i*, then there exists a potential of 0.4 for this reaction; that is, the two small effects (i.e., 0.2) become a geater effect (i.e., 0.4). This method seems a more natural way because physically both drugs are taken by the patient, and could even be worse if there are interactions.

Evaluation measure

To evaluate the accuracy of established drug-reaction relations by a given classifier (*h*,*H*); that is, to evaluate the closest of the two vectors *H*(*x*) (predicted reactions) and *Y*(*x*) (observed reactions) we will use the following two measures considered in [7].

1. Coverage. This measure evaluates the performance of a classifier for all the reactions that have been observed.
Given $x \in X$, we denote by $T(x)$ the set of all ordered reactions $\tau = \{i_1,\ldots,i_c\} \subset \{1,\ldots,c\}$ satisfying $H_{i1}(x) \exists\ldots\exists H_{ic}(x)$. Then according to a reaction vector $(Y_1(x),\ldots,Y_c(x))$,we define rank and the error as:

Obviously, the number $rank_\vartheta$,(x) and $error_\vartheta(x)$ depend on the order ϑ One way to avoid the dependence on ordering is to take the middle value of maximal and minimal ranks. In this paper we will use this way. We define the rank as

where

and

The numbers $\mathrm{rank}_{\mathrm{max}}(\mathrm{x})$ and $\mathrm{rank}_{\mathrm{min}}(\mathrm{x})$ associated to the “worst” and “best” ordering, respectively.
To define the average error – coverage, we will use the formula:

Note that, $E_{cov} = 0$ if a classifier makes a prediction such that for all $x \in X$ the ranks for observed reactions are placed in the top of the ordering list of weights $H_i(x)$. The smaller the value of E_{cov} the better.

2. Average Precision. One-error and coverage do not completely describe multi-label classification problems. In [7] the average precision was used to achieve more completely evaluation. We also will use this measure. Similar to the average error, the average precision depends on a given order $\vartheta = \{\vartheta_1\ldots, \vartheta_c\} \in T(x)$. So we define the average precision as a midpoint of average precisions obtained by the “worst” and “best” ordering. Let $Y(x) = \{\mathrm{l} \in \{1, \ldots, c\} : Y_\mathrm{l}(x)= 1\}$ be a set of reactions that have been observed for the patient x. Given order $\vartheta = \{\vartheta_1, \ldots, \vartheta_c\} \in T(x)$, (that is, $H_{\tau 1}(\mathrm{x}) \exists \ldots \exists H_{\vartheta c}(\mathrm{x})$), we define the rank for each reaction $\mathrm{l} \in Y(x)$ as $rank_\vartheta(\mathrm{x};\mathrm{l}) = k$, where $\vartheta_k = \mathrm{l}$. Then, Average Precision will be defined as:

where

P_{av} is expressed as a percentage; the larger the value of P_{av}, the better.

Optimization problems

The algorithm A(p), that described below, aims to minimize the distance between predicted reactions H(x) and observed reactions Y(x). We will consider the following distance functions:

where

the number of reactions for the patient x, and the sign “bar” indicates a normalization:

In the distance dist_0 a normalization is made such that the sums

are equal to the number of reactions. In dist_2, after multiplying by

we get the corresponding sums are equal to 1. dist_1 can be considered as a middle version. These distance functions are slightly different from the Linear Least Squares Fit (LLSF) mapping function [11,12].

It would be interesting to consider the Euclidian distance. But some preliminary analysis showed that this distance does not provide us a reasonable evaluation.

Given a classifier (h,H), the average distance error will be calculated as

Here $|X|$ stands for the cardinality of the set X.
Therefore, we have the following optimization problem:

subject to:

In this paper we will describe algorithm $A(p)$ which aims to minimize the average distance error

. This aim changes by taking different numbers p = 0,1,2, which provides different classifiers $A(p)$, p = 0,1,2.

A solution to the optimization problem (1.4),(1.5)

The function in (1.3) is non-convex and non-linear, and therefore may have many local minimum points. We need to find the global optimum point. The number of variables is $|D|$Ac. For the data Card20, that we will consider, $|D|$ = 3001 and c =5. Thus we have a global optimization problem with 15005 variables, which is very hard to handle using existing global optimization methods. Note that, we also tried to use local minimization methods which were unsuccessful. This means that there is a clear need to develop new optimization algorithms for solving problem (1.4),(1.5), taking into account some peculiarities of the problem.

In this paper we suggest one heuristic method for finding a “good” solution to the problem (1.4),(1.5). This method is based on the proposition given below. We denote by S the unit simplex in R^c; that is,

In this case for each $h(d) \in S$ the component $h_i(d)$ indicates simply the probability of the occurrence of the reaction i.

Given drug d we denote by $X(d)$ the set of all records in X, which used just one drug – d. Simply, the set $X(d)$ combines all records where the drug d was used alone.

Consider the problem:

Proposition 1. *A point* $h^*(d) = (h^*_1(d), \ldots, h^*\_c(d))$ *where*

(1.6),(1.7).

Now, given drug d, we consider the set $X_{all}(d)$ which combines all records that used the drug d. Clearly $X(d) \subset X_{all}(d)$. The involvement of other drugs makes it impossible to solve the corresponding optimization problem similar to (1.6), (1.7). In this case, we will use the following heuristic approach to find a “good” solution.

We denote by $N_{drug}(x)$ the number of drugs taken by the patient x. Then, we set:

This formula has the following meaning. If $N_{drug}(x) = 1$ for all $x \in X_{all}(d)$, then (1.9) provides global minimum solution. Let $N_{drug}(x) > 1$ for some record $x \in X_{all}(d)$. In this case, we assume that all drugs are responsible to the same degree; so we associate only the part $1/N_{drug}(x)$ of the reactions $Y_j(x)$ to this drug.

Algorithm *A*(*p*)

We will consider three versions of the algorithm $A(p)$, corresponding to the distance functions dist_p, p = 0,1,2, respectively. Each of these versions tends to minimize the average distance calculated by its own distance measure.

For each drug d we define the sets $X(d)$ – the set of all cases where drug d was used alone and $X_{all}(d)$ – the set of all cases where drug d was used. The set $X(d)$ carries very important information, because here the drug d and reactions are observed in a pure relationship.

Therefore, if the set $X(d)$ contains a "sufficiently large" number of records, then it will be reasonable to define the weights $h_j(d)$, (j = 1,…,c) by this set.

We consider two numbers: $|X(d)|$ – the number of cases where the drug is used alone, and $P(d) = 100\ |X(d)|\ /\ |X_{all}(d)|$ – the percentage of these cases. To determine whether the set $X(d)$ contains enough records we need to use the both numbers. We will consider a function $\phi(d) = a\ |X(d)| + b\ P(d)$ to describe how large the set $X(d)$ is.

Therefore, we define $h(d)$=($h_1(d)$, … $h_c(d)$) as follows:.

where $h^*(d)$ and $h^{**}(d)$ are defined by (1.8) and (1.9),respectively.

New drugs and new events

We define *a new drug* (in the test set) as a case when this drug either is a new drug which has not occurred in the training set or has never been considered as a suspected drug in the training set. For all such a new drug d, we set $h_i(d) = 0$, I = 1, …,c. It is possible that for some new (test) example all suspected drugs are new. We call this case as *a new event*. This situation is, mainly, related to the fact that, new drugs are constantly appearing on the market. Obviously, to make prediction for such examples does not make sense. Therefore, in the calculations below, we will remove all new events from test sets. For details see Table 2

TABLE 2: Card20. *The training and test sets.*

Year	Number of Records		
	Training	Test	Removed*
1996	12600	1049	98
1997	13747	1091	163
1998	15001	1418	265
1999	16684	1746	169
2000	18599	1988	158
2001	20745	1054	65

**'Removed' means how many records were removed from test set. For example, in 1996 there are 1147 records and 98 of them are new events. Then, the number of records in the test set for this year is 1049 (=1147-98)*

We make calculations in two versions: in the first version we consider all drugs that have been taken by patients as suspected drugs, in the second we use only those drugs which are reported as suspected in the ADRAC data.

In the calculations below we take as a test set records sequentially from each year, starting from 1996 until 2001. For example, if records from 1999 are taken as a test set, then all records from years 1971-1998 form a training set. In the Table 2 we summarized the number of records in test and training sets, and, also, the number of new events removed (new events are defined by suspected drugs only).

The results of numerical experiments

TABLE 3 The results obtained by the algorithm $A(p)$, p = 0,1,2.*

Year	Evaluation Measure		$A(0)$		$A(1)$		$A(2)$	
			Training	Test	Training	Test	Training	Test
1996	E_{cov}	All drugs	0.553	0.605	0.510	0.586	0.450	0.562
		Suspected drugs	0.507	0.588	0.464	0.571	0.415	0.556
	P_{av}	All drugs	81.36	79.81	82.61	80.05	83.72	79.79
		Suspected drugs	82.86	80.21	83.94	80.03	84.49	79.75
					1997	E_{cov}	All drugs	0.552
		Suspected drugs	0.509	0.613	0.464	0.589	0.417	0.570

	P_{av}	All drugs	81.41	78.94	82.63	79.42	83.78	79.24
		Suspected drugs	82.80	79.69	83.95	80.23	84.51	79.95
1998	E_{cov}	All drugs	0.551	0.679	0.512	0.650	0.454	0.621
		Suspected drugs	0.509	0.672	0.465	0.641	0.417	0.625
	P_{av}	All drugs	81.45	77.83	82.58	78.68	83.68	78.40
		Suspected drugs	82.79	78.01	83.89	78.58	84.47	78.41
1999	E_{cov}	All drugs	0.556	0.585	0.517	0.573	0.461	0.545
		Suspected drugs	0.515	0.569	0.471	0.562	0.424	0.543
	P_{av}	All drugs	81.39	80.46	82.46	80.63	83.48	81.03
		Suspected drugs	82.66	80.92	83.75	80.92	84.35	80.88
2000	E_{cov}	All drugs	0.549	0.691	0.512	0.668	0.457	0.647
		Suspected drugs	0.512	0.682	0.470	0.651	0.423	0.633
	P_{av}	All drugs	81.57	77.54	82.60	77.88	83.58	77.23
		Suspected drugs	82.70	77.77	83.73	78.23	84.37	77.60
2001	E_{cov}	All drugs	0.561	0.712	0.520	0.685	0.467	0.672
		Suspected drugs	0.525	0.713	0.480	0.683	0.435	0.667
	P_{av}	All drugs	81.22	76.89	82.39	77.50	83.35	77.55
		Suspected drugs	82.29	76.92	83.49	77.38	84.07	77.58

*"All drugs" means that the drug-reaction weights are calculated assuming all drug(s) suspected, "Suspected drugs" means that we use only suspected drug(s) reported in ADRAC data

We apply the algorithm $A(p)$ using a function $\phi(d) = |X(d)| + P(d)$ to describe the informativeness of the set $X(d)$. We also need to set a number p^* The calculations show that the results are not essentially changed for different values of p^* in the region $p^* \exists 30$. We set $p^* = 80$ in the calculations. The results are presented in Table 3.

The comparison of the results obtained for training sets show that using information about suspected drugs reported in the ADRAC data provides much better results. We also see that the drug-reaction relations established only by suspected drugs provide more precise prediction: in almost all cases the results obtained for test sets are better if we use suspected drugs only ("Susp. Drug" rows in Table 3).

One more important fact should also be noted. In all cases above the results obtained are much better than the default values (we define default values assuming that for each record all reactions can occur with the same weight). This emphasizes that it possible to study drug-reaction relations, not involving other information about patients. The drug-reaction relationships could then be used, together with the patient information, to enhance the prediction of reactions that could occur.

Conclusion

In this paper we have studied drug-reaction relations in thedomain of the Cardiovascular group of reactions from ADRAC data. These relations are presented in the form of a vector of weights. The results show the possibility of studying drug-reaction relations, not involving other information about patients. In all cases above the results obtained are much better than the default values. We demonstrated that by removing drugs not "suspected" of causing reactions, that an improvement in accuracy was obtained.

To develop new algorithms taking into account the peculiarities of ADRs is an important problem. The development of these algorithms should help us to extract more useful information from ADRAC data. In particular, the study of drug-reaction associations, drug-drug interactions and the influence of other data fields contained in the ADRAC data are interesting problems for future investigation.

ACKNOWLEDGMENTS

The work reported in this paper has been supported by The Australian Research Council, the TGA (Therapeutic Goods Administration), the Ballarat Division of General Practice, the Medical Software Industry Association, the Therapeutic Guidelines Pty. Ltd. We are indebted to Patrick Purcell from the TGA for providing expert comments. We also thank Rob Schapire for his helpful advice in using BoosTexter.

REFERENCES

[1] Brown Jr., Stephen D., Landry, Frank J.: Recognizing, Reporting, and Reducing Adverse Drug Reactions. Southern Medical Journal, Vol. 94 (2001) 370–374

[2] Heely, Emma, Riley, Jane, Layton, Deborah, Wilton, Lynda V., Shakir, Saad A. W.: Prescription-event monitoring and reporting of adverse drug reactions. The Lancet. Vol. 358 (2001) 182–184

[3] Mamedov, M.A., Saunders, G. W.: An Analysis of Adverse Drug Reactions from the ADRAC Database. Part 1: Cardiovascular group. University of Ballarat School of Information Technology and Mathematical Sciences, Research Report 02/01, Ballarat, Australia, February (2002) 1–48 http://www.ballarat.edu.au/itms/researchpapers/paper2002.shtml

[4] Mamedov M, Saunders G. A Fuzzy Derivative Approach to Classification of outcomes from the ADRAC database. International Trans Operational Research 2003. *In press*

[5] Mamedov M, Saunders G. Analysis of Cardiovascular Adverse Drug Reactions from the ADRAC Database. Proceedings of the APAC Conference and Exibition on Advanced Computing, Grid Applications and eResearch 2003. Royal Pines Resort, Gold Coast, Queensland, 29 September – 2 October 2003 http://www.apac.edu.au/apac03/

[6] Pirmohamed, Munir, Breckenridge, Alasdair M., Kitteringham, Neil R., Park, B. Kevin: Adverse drug reactions. British Medical Journal, Vol. 316 (1998) 1294–1299

[7] Schapire, Robert E., Singer, Yoram: Boostexter: A boosting-based system for text categorization. Machine Learning, Vol. 39 (2000) 135–168

[8] Troutman, William G., Doherty, Kevin M.: Comparison of voluntary adverse drug reaction reports and corresponding medical records. Am. J Health-Syst Pharm, Vol. 60 (2003) 572–575

[9] van Puijenbrock, Eugène P., Diemount, Willem L., van Grootheest, Kees: Application of Quantitative Signal Detection in the Dutch Spontaneous Reporting System for Adverse Drug Reactions. Drug Safety, Vol. 26 (2003) 293–301

[10] World Health Organization. WHO Technical Report No 498, 1972 and Note for Guidance on Clinical Safety Data Management: Definitions and Standards for Expedited Reporting (CPMP/ICH/377/95)

[11] Yang Yiming and Liu Xin: A Re-examination of Text Categorization Methods. Proceedings of SIGIR-99, 22nd ACM International Conference on Research and