An optimization approach to the study of drug-drug interactions

M.A.MAMMADOV and A.BANERJEE

Center for Informatics and Applied Optimization
University of Ballarat, Victoria, 3353, Australia

m.mammadov@ballarat.edu.au
abanerjee@students.ballarat.edu.au

Abstract

Drug-drug interaction is one of the important problems of Adverse Drug Reaction (ADR). In this paper we develop an optimization approach for the study of this problem. This approach is based on drug-reaction relationships represented in the form of a vector of weights, which can be defined as a solution to some global optimization problem. Although this approach can be used for solving many ADR problems, we concentrate here only on drug-drug interactions. Based on drug-reaction relationships, we formulate this problem as an optimization problem. The approach is applied to different classes of reactions from the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database.

Keywords: global optimization; adverse drug reaction; multi-label classification; drug-drug interaction

1 Introduction

An Adverse Drug Reaction (ADR) is defined by 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" [20]. ADRs are estimated to be the fourth leading cause of death in the USA [14], and the amount of published literature on the subject is vast [1]. Some of the problems concerning ADRs are discussed in our research report [7]. Many approaches have been tried for the analysis of adverse reaction data, such as: Fisher's Exact Test and matched pair designs (McNemar's test) [18], Reporting Odds Ratio (ROR). One approach that has had some success is the Proportional Reporting Ratios (PRR) for generating signals from data in the United Kingdom. The Norwood-Sampson Model has been applied to data in the United States of America and approved by the Food and Drug Administration. A common approach to the assessment of ADRs uses the Bayesian method [4]. For example, the Bayesian confidence propagation neural network (BCPNN) [2], an empirical Bayesian statistical data mining program, called a Gamma Poisson Shrinker (GPS) [5], and the Multi-item Gamma Poisson Shrinker (MGPS) [16], which have been applied to the United Sates Food and Drug Administration Spontaneous Reporting System database.

Each method has its own advantages and disadvantages with respect to applicability in different situations and possibilities for implementation. In [7], [10], [11] a new approach was developed where the main goal was to study, for each drug, the possible reactions that can occur; that is, to establish drug-reaction relationships. In these studies the ADR problem was formulated as a text categorization problem and the drug-reaction relationships defined as a solution to some global optimization problem. This approach was applied to the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database.

The approach developed in [7], [10], [11] allows us to study many ADR problems. In this paper we mainly concentrate on drug-drug interactions. We discuss the possibility of using this approach to the study of drug-drug interactions. The numerical experiments have been carried out on the basis of different classes of reactions in ADRAC database.

2 ADRAC database

The 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. It contains 137,297 records collected from 1972 to 2001. In the initial preprocessing stage we remove the records with missing information in fields related to reactions and drugs, and, further we consider 137,172 records having a complete information in these fields. A more detailed account of the ADRAC database is given in [7].

To define reaction classes, we will use a reaction tree presented in ADRAC database. This tree contains two levels. The first level, that will be referred to as *All Data,* consists of 18 different type of reaction classes called System Organ Class (SOC). Four of these classes contain subclasses of reactions that defines the second level. These classes are: *Blood*, *Body*, *Cardiovascular* and *Neurological*. To generate datasets related to these four reaction classes we collect all records having at least one reaction from these four classes. The *Cardiovascular* class contains four subclasses and *Blood*, *Body* and *Neurological* classes contain three subclasses.

In fact, records may have reactions from different classes. We need to take into account this situation when generating datasets corresponding to each of these four classes. Thus, we define one extra subclass, for each class, that contains reactions belonging to the other 17 SOCs. For the number of records in each subclass see Table 2 presented in Section 3.

The information about each patient consists of mainly two sets of information: individual patient information and information about drug(s) and reaction(s). In this paper we will use only the second set of information. By understanding the drug-reaction relationship in the absence of information about other factors influencing this relationship, we expect to be able to establish a clearer relationship between drugs and reactions. Another reason for focussing primarily on drugs and reactions relates to the inconsistent quality and quantity of relevant data on factors which also play a role in the drug-reaction association. This is largely due to the voluntary nature of the ADRAC reporting system. Some of the problems of such a reporting system are discussed in ([3],[6], [13], [17], [19]).

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 text categorization see [15], [21], [22].

2.1 Drug-reaction representations

We denote by $\mathcal{X}$ the set of all patients and by $\mathcal{D}$ the set of all drugs used by these patients. Let c be a finite number of possible reaction classes. As mentioned above, we will consider two level reaction classes. In the first level (that is, for *All Data*), we have $c = 18$; in the second level we have $c = 5$ for *Cardiovascular* class and $c = 4$ for *Blood*, *Body* and *Neurological* class of reactions.

Given drug $d \in \mathcal{D}$ drug-reaction relationships will be represented by a vector

$$h(d) = (h_1(d), h_2(d), \ldots, h_c(d)),$$

where non-negative numbers $h_i(d)$ stand the weights ("probabilities") of the occurrence of the reactions $i = 1, 2, \ldots, c$. The goal of the study of drug-reaction relationships is to find these relationships, that is, the function $h : \mathcal{D} \to R_+^c$, in an optimal way. Here R_+^c is the set of all c-dimensional vectors with non-negative coordinates.

Given a set of drugs $\Delta \subset \mathcal{D}$, we define a vector

$$H(\Delta) = (H_1(\Delta), H_2(\Delta), \ldots, H_c(\Delta)), \quad H_i(\Delta) = \sum_{d \in \Delta} h_i(d), \quad i = 1, \ldots, c; \tag{2.1}$$

where the component $H_i(\Delta)$ indicates the probability of occurrence of the reaction i after taking the drugs Δ. We call $H(\Delta)$ a vector of *Potential Reactions* related to the set Δ (see [7]).

3 Drug-drug Interactions

Drug-drug Interaction is one of the main problems of ADR. The proposed approach of drug-reaction presentations allows us to consider this problem from a mathematical point of view.

There is not an explicitly formulated and commonly used definition for this problem. For future discussions, we will use a definition presented in [12], where the drug-drug interaction problem is described as follows:

Definition 3.1 ([12]) *"A drug-drug interaction occurs when the effect of one drug is altered by the presence of another drug in the Body. For example:*

D1. *One drug might reduce or increase the effects of another drug.*

D2. *Two drugs taken together may produce a new and dangerous interaction.*

D3. *Two similar drugs taken together may produce an effect that is greater than would be expected from taking just one drug."*

In this definition, two different effects from drugs are considered: - the effect of drugs in terms of recovery from some diseases, and, - the effect of drugs in terms of producing some aside reactions (ADRs). For the sake of definiteness, we have to consider these effects separately.

One aspect of Definition 3.1 is related to recovery from taking the drugs. The main conclusion we can make from this definition is that:

(a) the effects are directed to increase or decrease of some features (functions) in the *Body*; and,

(b) for the normal recovery, these effects should be at some optimal levels.

Then, it is clear that, normal recovery will be impossible if the level of these effects is greater or less than the optimal level, and, this will produce situations (in particular, adverse drug reactions) described in Definition 3.1. Such situations may occur using only one drug (for example, in *Cardiovascular* class there are 3 records, having just one drug used, reported as drug interaction).

We note that the introduction of vectors $h(d)$ has been inspired by a (data mining) approach where the goal was to describe the possible side reactions that can occur after taking the drug d. This requires an assumption that, all the components of these vectors are non-negative, and, a large number in these components show a high probability of the corresponding reactions. In particular, this large number can be considered as a large deviation (divergency) from the optimal levels described in **(b).**

The study of the effects of drugs in terms of recovery requires considering quite different kinds of drug-reaction relationships describing the influence of drugs on some functions (features). In this case, negative numbers could be used to describe the effects directed to the decrease of these features. The study of this kind of drug-reaction relationships will be very important, although this probably will require explicit description of drugs and more medical investigations.

In this paper, we consider drug-reaction relationships in the form of vectors $h(d)$ defined in Section 2.1. Now we discuss how this approach matches with Definition 3.1 in terms of side reactions; in other words, we aim to discuss how drug-reaction presentations, in the form of vector of weights, could be used for description of drug-drug interactions.

Consider, for example, two drugs d_1, d_2, with weights $h(d_1)$, $h(d_2) \in R^c_+$, and assume that $H(x) = h(d_1) + h(d_2)$ is a vector of potential reactions. Clearly, $H(x) \geq h(d_i)$, $(i = 1, 2)$. This means that, the effect from two drugs taken together will be stronger than the effect expected from taking just one drug. In other words, potential reactions, defined above, can be used to study part **D1** (except for the part related to reduction of effect) and part **D3** of Definition 3.1. The more interesting case is **D2**. We now consider this case in detail.

3.1 Case D2

The reality is that, one drug causes not one but many different reactions. The grouping of similar reactions under one class (using by a reaction tree) also does not help: almost every drug causes

many different types of reaction classes. This requires that we consider the following cases separately:

D2.1. Two (or more) drugs taken together may produce a *new* type of reaction(s) that has not been observed with these drugs when they used alone.
D2.2. Two (or more) drugs taken together *mainly* produce a *new* type of reaction(s) that is different from the reaction(s) *mainly* observed with these drugs when they used alone.

The second case, **D2.2**, can be well explained by drug-reaction relationships in the form of vectors $h(d)$. Consider an example. Let there be 5 reaction classes and

$$h(d_1) = (0.6, 0.4, 0.0, 0.0, 0.0),\ h(d_2) = (0.0, 0.4, 0.6, 0.0, 0.0).$$

The vector of potential reactions related to these two drugs is

$$H = (0.6, 0.8, 0.6, 0.0, 0.0).$$

This means that the first and the third reactions mainly occurred when the drugs were used alone, but when they used together, the second reaction becomes the main reaction.

The interesting case is **D2.1** which coincides with **D2** of Definition 3.1. This case arises if, say for the example above, the fourth reaction occurs for some patient after taking the two drugs. This is the case, that cannot be studied by the approach of drug-reaction relationships $h(d)$. We note two important issues related to this case.

Note 1: Drug Combinations

How to study the case **D2.1**? To be able to study this case, we have to consider each combination of drugs in interaction separately. Moreover, we need to have enough records with the same combination of drugs, in order, to derive statistically significant results.

Table 1: The repeating combinations of drugs in interaction. N_{comb} is the number of different combinations; N_{freq} is the number of frequency of each combination

All Data	N_{comb}	1	1	2	2	3	5	7	29	102
	N_{freq}	28	19	10	9	6	5	4	3	2
Blood	N_{comb}	1	1	1	1	2	7	9	42	
	N_{freq}	22	18	9	8	5	4	3	2	
Body	N_{comb}	1	1	1	2	15				
	N_{freq}	10	7	4	3	2				
Neurological	N_{comb}	1	1	9	31					
	N_{freq}	8	5	3	2					
Cardiovascular	N_{comb}	1	4	16						
	N_{freq}	9	3	2						

The calculation of the number of repeating combinations and their frequencies provided the results presented in Table 1. From this table we see that only a few number of combinations repeated "many" times. For example, in the *Cardiovascular* type of reactions, there are 364 drug-drug interaction cases, where each record uses 2 to 7 drugs under interaction. The counting of repeating combinations provided the following results: one combination occurred 9 times, 4 different combinations occurred 3 times and 16 different combinations occurred only 2 times. Moreover, it is most likely that, the majority of each of these repeating cases related to one person having multiple records in the data. All the other records use different combinations of drugs in interactions! Similar situations we observe for the other classes and also for *All Data* (that is, 18 SOC). For example, in *All Data* there are only 6 different combinations of drugs that occurred more than 9 times and the records corresponding to these repeating cases is less than 4.5 percent of all drug-drug interaction cases. Therefore, the approach of considering different drug combinations cannot be effective in the study of drug-drug interactions in general.

To have "enough" repeating combinations, we have to use drug classes (generated from similar drugs) instead of drugs (defined by trade names in this paper). Therefore, the case **D2.1** needs substantial investigations including generating drug classes.

Note 2: The frequency of records classified as D2.1

Now we consider the problem how frequently the case **D2.1** occurs. In other words, we want to know, how many records, out of all drug-drug interactions, can be classified as a case **D2.1**. This is an interesting question. It turns out that even to count the number of these records is not easy.

The difficulty is related to defining the statement that "a drug is *used alone*". We consider here three different definitions:

- drug d is the only drug that has been taken (in this case, of course, it is also a suspected drug);
- drug d is the only drug reported as a suspected drug out of all the drugs (more than two) that have been taken; and
- drug d is one of the drugs (more than two) reported as suspected drugs (but not interacted).

Clearly, the last definition leads to a large number of reaction classes associated to this drug when it is *used alone.* The fact is that we cannot ignore this definition, as it states that, this drug was a suspected drug (the presence of other suspected drugs does not provide any additional information) in reactions observed.

Therefore, considering these three definitions, we aim to count the number of records that can be classified as **D2.1.** For this, we first generate three sets of records, having no drug interactions, named *Data1*, *Data2* and *Data3. Data1* combines all the records having only one drug used, *Data2* combines all the records having only one drug reported as suspected, and *Data3* combines all records (having of course, no drug interactions). Then, we define that, a record, having an interaction of drugs d_i , ..., d_m, belongs to the case **D2.1**, if some reaction, observed for this record, is not observed for all these drugs in the corresponding data (*Data1*, *Data2* or *Data3*). The results obtained are presented in Table 2.

Table 2: The number of records classified as **D2.1** corresponding to data *Data1*, *Data2* and *Data3*: n_1 corresponds to *Data1*, n_2 corresponds to *Data2*, n_3 corresponds to *Data3*. c is the number of classes/subclasses, N is the total number of records in each class of reactions, N_{inter} is the number of records having drug-drug interactions

Reaction tree	c	N	N_{inter}	n_1	n_2	n_3
All Data	18	137172	1668	28	15	6
Blood	4	8574	460	7	1	0
Body	4	34456	420	6	4	1
Cardiovascular	5	21871	364	7	4	1
Neurological	4	50591	683	6	3	1

As we can see from Table 2 the total number of drug-drug interactions in the ADRAC database is 1668 out of 137172 records. In this table we also present the number of drug-drug interactions in each of four classes of reactions. We note that some drug-drug interactions, having reactions from different classes, are present in different classes of reactions; that is, there is overlapping between these four classes.

The main conclusion that we can draw from Table 2, is that the number of drug-drug interactions classified as **D2.1** is sufficiently small. For example, in the class of *Blood* there are 7 records according to *Data1*, only 1 record according to *Data2* and none record according to *Data3*. Even, considering all ADRAC data, that is, 1668 drug-drug interaction cases, we have 28 records according to *Data1*, 15 records according to *Data2* and only 6 records according to *Data3*.

Thus, there are only a few records corresponding to the case **D2.1**. Moreover, the analysis of these records shows that, it is quite possible, these numbers could be much smaller, if the drugs

in interactions were indicated more "correctly". To explain this, consider the record from the year 1987 (we call it *Record*(1987)), which used 4 drugs: ID numbers 460, 782, 3498, 4714. This record is classified as **D2.1** according to *Data3* in both *Cardiovascular* class and *All Data*. First consider this record in *Cardiovascular* class that consists of 5 subclasses. The vector of observed reactions is (1, 0, 0, 0, 1), which means that the first and fifth (that is, others) reactions have been observed. The frequency of reactions related to these drugs are presented in Table 3.

Table 3: The frequency of reactions related to the drugs with ID numbers 460, 782, 3498, 4714, corresponding to data *Data3*. F_k stands for the frequency of k-th reaction subclass in the *Cardiovascular* class (k = 1, ..., 5). (i) indicates the drugs in interaction, "+" in row R stands for the reactions observed in the patient *Record*(1987) who took these four drugs

R	+				+
Drug ID	F_1	F_2	F_3	F_4	F_5
460 (i)	0	0	0	0	0
4714 (i)	0	0	2	1	0
782	35	8	33	48	117
3498	50	5	21	17	83

For this record, it was reported that the drugs 460 and 4714 had interacted. As we can see, drugs 460 and 4714 are not associated with the observed reactions (the first and the fifth), as a result, this case is considered as **D2.1**. In this case, the identification of drugs 782 and 3498 as non-suspected might be an error (or misprint), because these drugs frequently caused the first and the fifth reactions. If, it was an error, then the record *Record*(1987) would not be classified as **D2.1**.

The same situation for this record, is found while considering all the 18 SOC; that is, *All Data* (see Table 4). In this case this record was classified as **D2.1** because of the occurrence of the second class of reactions that was not observed with the drugs 460 and 4714 when they used alone according to data *Data3*. The conclusion that these drugs interacted might be an error as the other drugs, 782 and 3498, have a high frequency of causing this class of reactions when used alone.

Table 4: The frequency of reactions related to the drugs with ID numbers 460, 782, 3498, 4714, corresponding to data *Data3*. F_k stands for the frequency of k-th reaction class in *All Data* (k = 1, ..., 18). (i) indicates the drugs in interaction, "+" in row R stands for the reactions observed in the patient *Record*(1987) who took these four drugs

R		+		+	+	+				+			+					+
Drug ID	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}
460 (i)	6			3	2	2	2		1			1	1		1			1
4714 (i)	1		1	3		1	2	1		3		1	2					
782	149	39	6	205	114	66	24	30	13	105	68	25	19	26	2		2	131
3498	54	3	4	111	16	40	21	13	3	68	18	22	11	1				87

Therefore, according to **Notes 1** and **2**, we will only concentrate on the case **D2.2,** which, as mentioned above, can be well described by our approach.

We also mention one very important issue. The records with interaction can be divided into two parts: the first part that combines all records where all the drugs taken are reported as interacting, and the second part that only some drugs, out of all drugs taken, are reported as interacting. To evaluate the accuracy for the first part of the records is quite difficult. This needs to develop new methods for evaluation. The method described in Section 5 can be used for the second part of interactions. We will consider only this part aiming to check the possibility of using drug-reaction relationships for the study of drug-drug interactions. Good results obtained in this way will indicate the reasonableness and efficiency of the proposed approach.

Therefore, in this paper, we aim to study the possibility of using vectors of weights $h(d)$, calculated for each drug, for drug-drug interactions. In other words, we aim to check the closeness of potential reactions to the observed reactions for patients having interactions of drugs. In this way,

we can establish the accuracy with which the potential reactions could be used for the prediction of reactions in drug-drug interaction cases.

4 An optimization approach to determine drug-reaction representations

In this section we describe the Algorithm $A(p)$, that uses an optimization approach to determine drug-reaction representations.

Given a vector $V = (V_1, \cdots, V_c)$, with nonnegative coordinates, we will use the notation

$$\|V\| = \sum_{i=1}^{c} V_i. \tag{4.2}$$

Let $x \in \mathcal{X}$. We denote by $\mathcal{Y}(x) = (\mathcal{Y}_1(x), \mathcal{Y}_2(x), \cdots, \mathcal{Y}_c(x))$ a c-dimensional vector of reactions observed for this patient; where $\mathcal{Y}_i(x) = 1$ if the reaction i has occurred, and $\mathcal{Y}_i(x) = 0$ if it has not. Let $D(x)$ be the set of all drugs taken by the patient x and $\mathcal{H}(x) = \mathcal{H}(D(x))$ is a vector of potential reactions (see (2.1)).

We define the distance between predicted potential reactions $\mathcal{H}(x) = (\mathcal{H}_1(x), \ldots, \mathcal{H}_c(x))$ and observed reactions $\mathcal{Y}(x) = (\mathcal{Y}_1(x), \ldots, \mathcal{Y}_c(x))$ as:

$$dist\,(\mathcal{H}(x), \mathcal{Y}(x)) = \sum_{i=1}^{c} (\bar{\mathcal{H}}_i(x) - \mathcal{Y}_i(x))^2; \tag{4.3}$$

where the sign "bar" stands for a normalization with respect to the number of observed reactions $\|\mathcal{Y}(x)\|$:

$$\bar{\mathcal{H}}_i(x) = \begin{cases} \frac{\|\mathcal{Y}(x)\|}{\|\mathcal{H}(x)\|}\,\mathcal{H}_i(x) & \text{if } \|\mathcal{H}(x)\| > 0; \\ 0 & \text{if } \|\mathcal{H}(x)\| = 0. \end{cases} \tag{4.4}$$

Given $p = 0, 1, 2$, we will use the following distance measure (we assume that $\|\mathcal{Y}(x)\| > 0$):

$$dist_p\,(\mathcal{H}(x), \mathcal{Y}(x)) = \|\mathcal{Y}(x)\|^{-p} \cdot dist\,(\mathcal{H}(x), \mathcal{Y}(x)), \quad p = 0, 1, 2. \tag{4.5}$$

Note that, these distance functions are slightly different from the Linear Least Squares Fit (LLSF) mapping function used in text categorization (see, for example, [21], [22]).

The algorithm $A(p)$ aims to define drug-reaction relations $h(d)$ minimizing the average distance $dist_p\,(\mathcal{H}(x), \mathcal{Y}(x))$ over all training examples. In other words, we consider the following optimization problem:

$$E^p_{av} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} dist_p\,(\mathcal{H}(x), \mathcal{Y}(x)) \;\; \rightarrow \;\; \min; \tag{4.6}$$

$$\text{subject to :} \qquad h_i(d) \geq 0, \quad i = 1, ..., c, \;\; d \in \mathcal{D}. \tag{4.7}$$

Here $|\mathcal{X}|$ stands for the cardinality of the set $\mathcal{X}$. Note that by taking different numbers $p = 0, 1, 2$, we get different versions of $A(p)$, $p = 0, 1, 2$, which generate different drug-reaction representations $h(d)$.

Therefore drug-reaction representations will be defined as a solution to the optimization problem (4.6)-(4.7). The function in (4.6) is non-convex and has a large number of local minimum points. The number of variables is $|\mathcal{D}| \cdot c$. For the *All Data*, we have $|\mathcal{D}| = 5057$ and $c = 18$. Thus we have a global optimization problem with extremely large number of variables, which is very hard to handle using existing global optimization methods. Taking into account some peculiarities of the problem, we suggest an algorithm which allows us to find sufficiently "deep" local minimum point of the objective function in (4.6). We solve this problem in three steps.

Step 1. First we find some "good" initial point for the problem (4.6), (4.7). In this stage we use a method developed in [11]. Let this initial point be

$$h^0(d) = (h^0_1(d), ..., h^0_1(d)), \quad d \in \mathcal{D}. \tag{4.8}$$

Step 2. In the second step, we introduce new variables $\lambda(d)$, $d \in \mathcal{D}$, and represent (scale) drug reaction relationships in the form:

$$h(d) = \lambda(d)\,(h_1^0(d), ..., h_1^0(d)), \quad d \in \mathcal{D}. \tag{4.9}$$

Then we consider the global optimization problem

$$\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} dist_p\,(\mathcal{H}(x:\lambda), \mathcal{Y}(x)) \;\; \rightarrow \;\; \min; \tag{4.10}$$

$$\text{subject to :} \quad 0 \leq \lambda(d) \leq 2, \quad d \in \mathcal{D}; \tag{4.11}$$

where $\mathcal{H}(x:\lambda) = (\mathcal{H}_1(x:\lambda), ..., \mathcal{H}_c(x:\lambda))$ and

$$\mathcal{H}_i(x:\lambda) = \sum_{d \in \Delta(x)} \lambda(d)\, h_i^0(d).$$

The number of variables in this problem is equal to the number of drugs. The use of formula (4.9) means that we take as stable the proportions between weights, for each drug, calculated in the first step. For the patients having only one drug (this is, about 50 percent of all records), the scaling (4.9) does not affect the classification. The effect of this scaling works when two or more drugs were involved together. The large numbers $\lambda(d)$ obtained as a solution to (4.10), (4.11) will indicate the importance of these drugs in terms of causing reactions.

To solve this problem we apply the algorithm AGOP which is developed in [8], [9]. Let the optimal solution to the problem (4.10), (4.11) be $\lambda^0(d), \;\; d \in \mathcal{D}, \;\;$ which provides the weight vectors

$$h^1(d) = (h_1^1(d), ..., h_c^1(d)), \quad h_i^1(d) = \lambda^0(d) h_i^0(d), \quad d \in \mathcal{D},\; i = 1, ..., c. \tag{4.12}$$

Step 3. In the last step, we consider the problem

$$E_{av}^p = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} dist_p\,(\mathcal{H}(x), \mathcal{Y}(x)) \;\; \rightarrow \;\; \min; \tag{4.13}$$

$$\text{subject to :} \quad \theta_1\, h_i^1(d) \leq h_i(d) \leq \theta_2\, h_i^1(d), \quad d \in \mathcal{D},\; i = 1, ..., c. \tag{4.14}$$

In the calculations below, we set $\theta_1 = 0.01$, $\theta_2 = 2$. The number of variables in this problem is equal to $|\mathcal{D}|\, c$. To solve this problem we apply the algorithm AGOP. We denote the solution found by:

$$h^{GO}(d) = (h_1^{GO}(d), ..., h_c^{GO}(d)), \quad d \in \mathcal{D}. \tag{4.15}$$

5 Evaluation Measures

In this section we describe two evaluations measures that will be used in numerical experiments. First we aim to evaluate the accuracy of established drug-reaction relations by a given classifier; that is, to evaluate the closeness of the two vectors $\mathcal{H}(x)$ (predicted reactions) and $\mathcal{Y}(x)$ (observed reactions). For this aim, we will use the *Average Precision* measure considered in [15]. Note that, this measure is based on the ordering of weights in $\mathcal{H}(x)$, and it allows us to achieve more complete evaluation in multi-label classification problems. In the second evaluation measure we consider two sets of drugs for each patient: the drugs in interaction and the others. The aim here is to evaluate the "responsibility (weight)" of drugs in interaction, compared with the others, in the reactions observed for this patient. The second evaluation measure is based on distance between $\mathcal{H}(x)$ and $\mathcal{Y}(x)$.

5.1 Average Precision

Let $Y(x) = \{l \in \{1, \ldots, c\} : \; \mathcal{Y}_l(x) = 1\}$ be the set of reactions that have been observed for the patient x and $\mathcal{H}(x) = \{\mathcal{H}_1(x), \cdots, \mathcal{H}_c(x)\}$ be potential reactions calculated for this patient. We denote by $\mathcal{T}(x)$ the set of all ordered reactions $\tau = \{i_1, \ldots, i_c\}$ satisfying the condition

$$\mathcal{H}_{i_1}(x) \;\geq\; \ldots \;\geq\; \mathcal{H}_{i_c}(x);$$

where $i_k \in \{1, \ldots, c\}$ and $i_k \neq i_m$ if $k \neq m$.

In the case, when the numbers $\mathcal{H}_i(x), \quad i = 1, \cdots, c,$ are different, there is just one order satisfying this condition. But if there are reactions having the same weights then we can order potential reactions in different ways; that is, in this case the set $\mathcal{T}(x)$ contains more than one order.

Given order $\tau = \{\tau_1, \ldots, \tau_c\} \in \mathcal{T}(x),$ we define the rank for each reaction $l \in Y(x)$ as $rank_\tau(x; l) = k$, where the number k satisfies $\tau_k = l$. Then *Precision* is defined as:

$$P_\tau(x) = \frac{1}{|Y(x)|} \sum_{l \in Y(x)} \frac{|\{k \in Y(x) : \ rank_\tau(x; k) \leq rank_\tau(x; l)\}|}{rank_\tau(x; l)}.$$

Here, we use the notation $|S|$ for the cardinality of the set S. This measure has the following meaning. For instance, if all observed reactions $Y(x)$ have occurred on the top of ordering τ then $P_\tau(x) = 1$. Clearly the number $P_\tau(x)$ depends on order τ. We define

$$P_{best}(x) = \max_{\tau \in \mathcal{T}(x)} P_\tau(x) \quad \text{and} \quad P_{worst}(x) = \min_{\tau \in \mathcal{T}(x)} P_\tau(x),$$

which are related to the "best" and "worst" ordering. Therefore, it is sensible to define the *Precision* as the midpoint of these two versions: $P(x) = (P_{best}(x) + P_{worst}(x))/2$.

Average Precision over all records $\mathcal{X}$ will be defined as:

$$P_{av} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P(x). \tag{5.16}$$

5.2 The evaluation of responsibility of drugs in interaction

Consider a particular patient x and let $D(x)$ be the set of drugs used by this patient and $\mathcal{Y}(x)$ be the set of observed reactions. The set $D(x)$ consists of two parts: $DI(x)$ - drugs under interaction and $DO(x)$ - the other drugs.

The method of evaluation is based on distance measure (4.3). Assume that for each drug $d \in \mathcal{D}$ the vector of weights $h(d)$ are calculated. Then, by formula 2.1, we can define potential reactions $\mathcal{H}^I(x)$ and $\mathcal{H}^O(x)$, corresponding to the sets of drugs interacted - $DI(x)$, and the other drugs - $DO(x)$, respectively:

$$\mathcal{H}^I(x) = H(DI(x)), \quad \mathcal{H}^O(x) = H(DO(x)).$$

The method, used in this paper for the evaluation of responsibility of drugs in interaction, can be identified as "all drugs interacted versus all the other drugs taken". For this aim we consider convex combinations of these two group of drugs and try to find the optimal combination which provides the maximal closeness to the reactions observed. In other words we are looking for a combination of interacted and non-interacted drugs which is optimal in the sense of distance (4.3). Before considering convex combinations we need to be careful about the "comparability" of the vectors $\mathcal{H}^I(x)$ and $\mathcal{H}^O(x)$ in the sense of scaling. For this reason, it is meaningful to consider convex combinations of normalized (see (4.4)) vectors $\bar{\mathcal{H}}^I(x)$ and $\bar{\mathcal{H}}^O(x)$. Therefore we define

$$\bar{\mathcal{H}}(x, \mu) = \mu\, \bar{\mathcal{H}}^I(x) + (1 - \mu)\, \bar{\mathcal{H}}^O(x), \quad 0 \leq \mu \leq 1. \tag{5.17}$$

Note that, $\|\bar{\mathcal{H}}^I(x)\| = \|\bar{\mathcal{H}}^O(x)\| = \|\mathcal{Y}(x)\|$ and, therefore, $\|\bar{\mathcal{H}}(x, \mu)\| = \|\mathcal{Y}(x)\|$ for all $\mu \in [0, 1]$.

The number μ indicates the proportion of the drugs interacted and the other drugs in the definition of potential reactions. Clearly, $\bar{\mathcal{H}}(x, 1) = \bar{\mathcal{H}}^I(x)$ and $\bar{\mathcal{H}}(x, 0) = \bar{\mathcal{H}}^O(x)$, which implies

$$dist(\bar{\mathcal{H}}(x, 1), \mathcal{Y}(x)) = dist(\mathcal{H}^I(x), \mathcal{Y}(x)),$$

$$dist(\bar{\mathcal{H}}(x, 0), \mathcal{Y}(x)) = dist(\mathcal{H}^O(x), \mathcal{Y}(x)).$$

It is important to note that, the combination of all drugs with equal weights; that is, the vector $\mathcal{H}^A(x) = H(D(x)) = \mathcal{H}^I(x) + \mathcal{H}^O(x)$ is also considered in (5.17). To confirm this, it is sufficient to consider the case $\|H^I(x)\| > 0$ and $\|H^O(x)\| > 0$. In this case $\|H^A(x)\| = \|H^I(x)\| + \|H^O(x)\| > 0$. Then we take $\mu' = \|\mathcal{H}^I(x)\| / \|\mathcal{H}^A(x)\| \in (0, 1)$ and get (see (4.4))

$$\bar{\mathcal{H}}(x, \mu') = \mu' \frac{\|\mathcal{Y}(x)\|}{\|\mathcal{H}^I(x)\|} \mathcal{H}^I(x) + (1 - \mu') \frac{\|\mathcal{Y}(x)\|}{\|\mathcal{H}^O(x)\|} \mathcal{H}^O(x)$$

$$= \frac{\|\mathcal{Y}(x)\|}{\|H^A(x)\|}\mathcal{H}^I(x) + \frac{\|\mathcal{Y}(x)\|}{\|\mathcal{H}^A(x)\|}\mathcal{H}^O(x) = \bar{\mathcal{H}}^A(x);$$

which implies

$$dist(\bar{\mathcal{H}}(x,\mu'),\mathcal{Y}(x)) = dist(\bar{\mathcal{H}}^A(x),\mathcal{Y}(x)) = dist(\mathcal{H}^A(x),\mathcal{Y}(x)).$$

Consider the following minimization problem with respect to μ;

$$f(\mu) \doteq dist\,(\bar{\mathcal{H}}(x,\mu),\mathcal{Y}(x)) = \sum_{i=1}^{c}(\bar{\mathcal{H}}_i(x,\mu) - \mathcal{Y}_i(x))^2 \;\rightarrow\; \min; \quad 0 \le \mu \le 1. \tag{5.18}$$

The optimal solution μ^* to problem (5.18) gives an information about the responsibility of drugs interacted. For instance, if $\mu^* = 1$ then we see that the drugs interacted provide the better approximation to the observed reactions than if we involve the other drugs. We refer this situation as 100 percent responsibility. Whereas, if $\mu^* = 0$ then the other drugs provide better approximation to the observed reactions and we can conclude that in this case drugs interacted are defined completely non-responsible. Therefore, the optimal value μ^* can be considered as an evaluation measure for the responsibility of drugs interacted.

From (4.5) we obtain the following:

Proposition 5.1 *The optimal solution μ^* to the problem (5.18) is optimal with respect to the all distance measures $dist_p$, $p = 0,1,2$; that is, given vectors of weights $h(d)$, $d \in D(x)$, for all $p = 0,1,2$ the following inequality holds:*

$$dist_p(\bar{\mathcal{H}}(x,\mu^*),\mathcal{Y}(x)) \le dist_p(\bar{\mathcal{H}}(x,\mu),\mathcal{Y}(x)), \quad \text{for all} \quad \mu \in [0,1].$$

This proposition shows that, given patient $x \in \mathcal{X}$ and given vectors of weights $h(d)$, the definition of responsibility of drugs interacted, as an optimal value μ^*, does not depend on choice of distance functions $dist$ and $dist_p$, $p = 0,1,2$.

It is clear that, problem (5.18) can have many optimal solutions μ^*; that is, different proportions of interacted and other drugs can provide the same closeness to the observed reactions. In this case we will define the responsibility of drugs interacted, as the maximal value among all optimal solutions μ^* :

$$\mu^*(x) = \max\{\mu^* : \quad \mu^* \;\text{ is an optimal solution to (5.18)}\}. \tag{5.19}$$

The reason for such a definition can be explained; for instance, if $\mu^* = 1$ (only interacted drugs) and $\mu^* = 0$ (only the other drugs) are the two different optimal solutions, giving the closest approximation to the observed reactions, then there would be no reason to doubt the responsibility of drugs interacted.

Problem (5.18) can be easily solved. Let

$$A = \sum_{j=1}^{c} \left(z_i \,\|\mathcal{Y}(x)\| - z\,\mathcal{Y}_i(x)\right)\left(z_i \,\|\bar{\mathcal{H}}^O(x)\| - z\,\bar{\mathcal{H}}_i^O(x)\right);$$

$$B = \sum_{j=1}^{c} \left(\bar{\mathcal{H}}_i^O(x)\,\|\mathcal{Y}(x)\| - \|\bar{\mathcal{H}}^O(x)\|\mathcal{Y}_i(x)\right)\left(z_i \,\|\bar{\mathcal{H}}^O(x)\| - z\,\bar{\mathcal{H}}_i^O(x)\right);$$

where $z_i = \bar{\mathcal{H}}_i^I(x) - \bar{\mathcal{H}}_i^O(x)$, $z = \|\bar{\mathcal{H}}^I(x)\| - \|\bar{\mathcal{H}}^O(x)\|$. Then, we find the derivative of the function $f(\mu)$, defined by (5.18), in the following form:

$$f'(\mu) = \frac{2}{(z\mu + \|\bar{\mathcal{H}}^O(x)\|)^4}\,(A\mu + B). \tag{5.20}$$

From (5.20) we have

Proposition 5.2 *The optimal solution $\mu^*(x)$ to the problem (5.18) can be found as follows.*

1) Let $A = 0$. Then

$$\mu^*(x) = \begin{cases} 0 & \text{if} \quad B > 0; \\ 1 & \text{otherwise.} \end{cases}$$

2) Let $A > 0$. Then

$$\mu^*(x) = \begin{cases} 0 & \text{if } B > 0; \\ \min\{1, -B/A\} & \text{otherwise.} \end{cases}$$

3) Let $A < 0$. Then

$$\mu^*(x) = \begin{cases} 0 & \text{if } f(0) < f(1); \\ 1 & \text{otherwise.} \end{cases}$$

Therefore, we have defined the responsibility of drugs interacted for a particular patient x. Given set of patients $\mathcal{X}$, *Average Responsibility* of drugs in interaction will be calculated as

$$P_{int} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \mu^*(x). \tag{5.21}$$

The numbers $P(x)$ and $\mu^*(x)$, in formulae 5.16 and 5.21, give some information about each interaction case. For instance, if $P(x) = 1$ (that is, 100 percent) and $\mu^*(x) = 1$ then we can conclude that the potential reactions defined by the drugs in interaction provide 100 percent correct prediction of reactions in both evaluation measures. Therefore, in this case, we can say that the potential reactions could be used for reaction predictions in the case of interactions.

6 The results of numerical experiments

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.

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, having drug-drug interactions, are taken as a test set, then all records from years 1972-1998, including also drug interaction cases, form a training set. Training sets are used to determine drug-reaction relationships.

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. It is possible that in a new (test) example all drugs taken are new. We call this case a *new event*. This situation mainly relates to the fact that, new drugs are constantly appearing on the market. Obviously, to make analysis for such examples does not make sense. Therefore, in the calculations below, we will remove all new events from test sets.

As mentioned in Section 5, for the evaluation, we will use two measures: *Average Responsibility* - P_{int} and *Average Precision* P_{av}.

For our analysis we consider the records having more than 3 drugs, where some of drugs were reported as interacted (in the ADRAC data the value 2 was associated with these drugs) and the others were reported as non-suspected (the value 0 was used in this case). To make the problem of evaluation of drug-drug interactions meaningful, we need to consider the records for which both parts are non-empty sets.

The results obtained are presented in Tables 5-9. Training sets are used for calculating of weights for each drug. As in [11], the weights are calculated by using suspected drugs. Then the evaluation of interaction of drugs is done only for test sets, because, in training sets, drugs in interaction (as suspected drugs) are used for the calculation of weights. The number of cases in the test sets are also presented in these tables.

In the last row of Tables 5-9, we present the average results obtained by all test sets in a particular class of reactions. To have some idea about the accuracy achieved, let us consider the results obtained by the algorithm $A(2)$ for the *Cardiovascular* class of reactions. We have $P_{int} = 71.6$ and $P_{av} = 82.8$. The first number means that, in the observed reactions, the "degree of responsibility" of the drugs in interaction versus the all other drugs, is more than 70 percent which is sufficiently high. The second number indicates high accuracy in the prediction of these reactions in terms of text categorization measure *Average Precision*.

Considering the all ADRAC database (Table 5) we see that the degree of responsibility of the drugs in interaction versus the all other drugs, is sufficiently high for all versions $A(p),\ p = 1, 2, 3$, being around 75 percent. The *Average Precision* of approximately 68 percent is also quite high taking into account the large number of classes (that is, 18 classes). This emphasizes that, drug-drug interaction cases could be successfully explained and predicted by the weights calculated for each drug.

In fact the accuracy of this method could be much higher if we could calculate weights more "correctly". To show this, we did the following.

First we note that, the numbers P_{int} and P_{av} are the average values of $\mu^*(x)$ and $P(x)$ calculated for each patient x (see Section 5). Different versions $A(p)$ provide different values $\mu^*(x)$ and $P(x)$. We take the corresponding maximal values obtained by different versions $A(p),\ p = 1, 2, 3$, and then calculate the average responsibility and precision. The results obtained are presented in the columns "max" in Tables 5-9. These results are much better than the results obtained by a particular version of $A(p)$.

For the class *Body* (Table 7) the the degree of responsibility of the drugs in interaction is around 60 percent, but the *Average Precision* is very high - 98.7. This means that drugs in interaction, along with the drug-reaction representations, can be used to predict the reactions that occurred with high accuracy.

The highest accuracy is obtained for the class *Blood* (Table 6), where both *Average Responsibility* of drugs in interaction and *Average Precision* is sufficiently high.

For the class *Neurological* (Table 9) the results presented in the columns "max" are sufficiently high in both evaluation measures. The results obtained by a particular version of the algorithm $A(p)$ are low compared to "max". This indicates the necessity for future investigations for more efficient methods to calculate drug-reaction representations.

Table 5: Evaluation of drug-drug interactions: *All Data;* that is 18 SOC

Test	N	P_{int}				P_{av}			
Year		$A(0)$	$A(1)$	$A(2)$	max	$A(0)$	$A(1)$	$A(2)$	max
1996	44	82.1	81.9	86.1	92.2	71.7	76.8	74.9	80.8
1997	45	74.0	67.1	63.6	79.4	66.2	65.3	62.5	71.9
1998	71	68.4	68.9	71.0	75.4	63.6	64.5	63.5	70.7
1999	18	70.3	58.0	65.1	80.6	71.0	68.0	68.6	72.7
2000	11	87.2	93.7	85.5	96.3	62.9	57.3	54.7	71.1
2001	53	79.4	81.1	82.0	87.4	75.0	73.1	68.4	78.6
total	242	75.3	74.0	75.0	83.2	68.6	68.7	66.5	74.7

Table 6: Evaluation of drug-drug interactions: class *Blood*

Test	N	P_{int}				P_{av}			
Year		$A(0)$	$A(1)$	$A(2)$	max	$A(0)$	$A(1)$	$A(2)$	max
1996	22	79.1	77.7	75.5	79.6	97.7	97.0	99.2	100
1997	14	81.8	81.8	89.4	89.4	96.4	96.4	96.4	96.4
1998	16	83.7	83.0	82.4	83.9	96.9	96.9	95.8	96.9
1999	6	66.7	50.0	50.0	66.7	88.9	88.9	91.7	91.7
2000	1	100	100	100	100	100	100	100	100
2001	14	86.9	83.5	84.5	87.0	96.4	92.9	91.7	96.4
total	73	81.4	78.8	79.7	83.0	96.3	95.4	95.9	97.3

Table 7: Evaluation of drug-drug interactions: class *Body*

Test	N	P_{int}				P_{av}			
Year		$A(0)$	$A(1)$	$A(2)$	max	$A(0)$	$A(1)$	$A(2)$	max
1996	8	59.3	50.0	50.3	61.3	94.8	94.8	94.8	94.8
1997	10	60.0	70.0	70.0	70.0	100	100	100	100
1998	9	55.6	55.6	67.2	67.2	98.1	98.1	98.1	98.1
1999	6	50.0	50.0	48.1	59.1	100	100	100	100
2000	2	50.0	50.0	50.0	50.0	100	100	100	100
2001	11	72.7	63.6	63.4	81.8	100	100	100	100
total	46	60.3	58.7	60.7	68.5	98.7	98.7	98.7	98.7

Table 8: Evaluation of drug-drug interactions: class *Cardiovascular*

Test	N	P_{int}				P_{av}			
Year		$A(0)$	$A(1)$	$A(2)$	max	$A(0)$	$A(1)$	$A(2)$	max
1996	4	48.1	78.7	81.0	92.0	87.5	89.6	87.5	89.6
1997	13	79.8	80.4	68.7	83.5	79.5	79.5	84.0	84.6
1998	8	87.5	82.7	88.8	88.8	61.5	68.8	79.2	79.2
1999	2	33.3	33.3	33.3	33.3	58.3	58.3	39.2	58.3
2000	2	50.0	50.0	98.1	100	75.0	75.0	100	100
2001	14	63.7	59.7	68.1	78.8	87.7	84.8	86.3	91.3
total	43	67.9	68.8	71.6	80.6	78.4	79.0	82.8	85.7

Table 9: Evaluation of drug-drug interactions: class *Neurological*

Test	N	P_{int}				P_{av}			
Year		$A(0)$	$A(1)$	$A(2)$	max	$A(0)$	$A(1)$	$A(2)$	max
1996	11	36.4	36.4	36.4	45.5	62.1	59.8	59.8	65.2
1997	15	56.6	64.8	49.4	68.2	78.9	81.1	77.8	85.6
1998	16	70.9	64.3	68.1	81.0	69.8	85.4	87.0	90.6
1999	5	80.0	78.3	68.8	88.8	80.0	90.0	83.3	90.0
2000	5	0.0	50.64	60.0	60.0	51.7	65.0	75.0	75.0
2001	24	77.5	71.7	57.9	81.7	79.5	81.2	83.0	88.5
total	76	61.1	62.7	56.1	72.6	73.0	78.5	78.9	84.2

7 Conclusion

In this paper we have presented a new optimization approach to study drug-drug interactions. Our focus was comprehensive, considering the fact that all drug-reaction relations was taken into account. We chose to traverse down the reaction tree from all ADRAC called *All Data* level down to the level of individual class of reactions: *Blood*, *Body*, *Cardiovascular* and *Neurological*. The suggested method of representation for drug-reaction relations in the form of a vector of weights is examined for the prediction of reactions in drug-drug interaction cases. The results obtained have shown that the reactions that occurred in the cases of interaction of drugs, reported in the ADRAC data, could be predicted by this method with sufficiently high accuracy.

At the end we note that, the approach of drug-reaction representations has been implemented in software. This software can be used to solve the following two problems:

- given a set of drugs to predict the reactions that are most likely to occur; and
- given a set of drugs and a set of reactions occurred, to determine the drugs that are the most likely cause these reactions.

This software has potential application in prescribing activities by GPs and also in pharmacy and dispensing.

References

[1] Aronson, J.K., Derry. S. and Loke, Y.K., 2002, Adverse drug reactions: keeping up to date. *Fundam Clin Pharmacol,* **16,** 49-56.

[2] Bate, A., Lindquist, M., Edwards, I.R., Olsson, S., Orre, O., Lansner, A. and De Freitas, R., 1998, A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol,* **54,** 315-321.

[3] Brown, SD Jr. and Landry, F.J., 2001, Recognizing, Reporting, and Reducing Adverse Drug Reactions. *Southern Medical Journal,* Vol. **94(4),** 370-374.

[4] Coulter, D.M., Bate, A., Meyboom, R., Lindquist, M. and Edwards, I.R., 2001, Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining. *BMJ,* **322(7296),** 1207-1209.

[5] DuMouchel, W., 1999, Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *American Statistician,* **53(3),** 177-202.

[6] Heely, E., Riley, J., Layton, D., Wilton, L.V. and Shakir, S.A.W., 2001, Prescription-event monitoring and reporting of adverse drug reactions. *The Lancet,* Vol. **358,** 182-184.

[7] Mamedov, M.A., Saunders, G.W., 2002, 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/research-papers/paper2002.shtml

[8] Mammadov, M.A., 2004, A new global optimization algorithm based on a dynamical systems approach. In Proc. *International Conference on Optimization: Techniques and Applications - ICOTA6*, Ballarat, December, Also in: *Research Report 04/04, University of Ballarat, 2004.* http://www.ballarat.edu.au/ard/itms/publications/researchPapers.shtml

[9] Mammadov, M.A., Rubinov, A.M. and and Yearwood, J., Dynamical systems described by relational elasticities with applications. To appear in *Continuous Optimization: Current trends and Applications*, V.Jeyakumar and A.Rubinov (eds), Springer.

[10] Mammadov, M.A., Saunders, G. and Yearwood, J., 2004, A Fuzzy Derivative Approach to Classification of Outcomes from the ADRAC Database. *International Transactions in Operational Research,* **11,** No. 2, 169-179.

[11] Mammadov, M.A., Rubinov, A.M. and Yearwood, J., An Optimization Approach to Identify the Relationship between Features and Output of a Multi-label Classifier. *In press*

[12] http://ohioline.osu.edu/ss-fact/0129.html

[13] Munir, P., Breckenridge, A.M., Neil R.K. and B. Kevin Park, 1998, Fortnightly review: Adverse drug reactions. *British Medical Journal,* Vol. **316,** 1295-1298.

[14] Redfern, W.S., Wakefield, I.D., Prior, H., Pollard, C.E., Hammond, T.G. and Valentin, J-P., 2002, Safety pharmacology - a progressive approach. *Fundam Clin Pharmacol,* **16,** 161-173.

[15] Schapire, R.E. and Singer, Y., 2000, Boostexter: A boosting-based system for text categorization. *Machine Learning,* Vol. **39,** 135-168.

[16] Szarfmann, A., Machado, S.G. and O'Neill, R.T., 2002, Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than-Expected Combinations of Drugs and Events in the US FDA's Spontaneous Reports Database. *Drug Safety,* **25(6),** 381-392.

[17] Troutman, W.G. and Doherty, K.M., 2003, Comparison of voluntary adverse drug reaction reports and corresponding medical records. *Am. J Health-Syst Pharm,* Vol. **60,** 572-575,

[18] Tubert-Bitter, P. and Begaud, B., 1993, Comparing Safety of Drugs. *Post Marketing Surveillance,* **7,** 119-137.

[19] van Puijenbrock, E.P., Diemount, W.L. and van Grootheest, K., 2003, Application of Quanitative Signal Detection in the Dutch Spontaneous Reporting System for Adverse Drug Reactions. *Drug Safety,* Vol. **26,** (5), 293-301.

[20] 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) (http://www.emea.eu.int/pdfs/human/ich/037795en.pdf)

[21] Yang, Y. and Liu, X., 1999, A Re-examination of Text Categorization mMethods. In Proceedings of *SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval,* Vol. **39,** 42-49.

[22] Yang, Y., Zhang, J. and Kisiel, B. 2003, A Scalability Analysis of Classifiers in Text Categorization. *SIGIR'03*, July 28-August 1, 2003, Toronto, Canada.