The study of drug–reaction relationships using global optimization techniques

M. A. MAMMADOV*, A. M. RUBINOV and J. YEARWOOD

University of Ballarat, Ballarat, Victoria, 3353 Australia

(Received 1 June 2005; revised 2 February 2006; in final form 9 March 2006)

In this paper we develop an optimization approach for the study of adverse drug reaction (ADR) problems. 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 it can be used for solving many ADR problems, we concentrate on two of them here: the accurate identification of drugs that are responsible for reactions that have occurred, and drug–drug interactions. Based on drug–reaction relationships, we formulate these problems as an optimization problem. The approach is applied to cardiovascular-type reactions from the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database. Software based on this approach has been developed and could have beneficial use in prescribing.

Keywords: Global optimization; Adverse drug reaction; Multi-label classification; Suspected drugs; Drug–drug interaction

1. Introduction

In general the problem of classification is the determination of the classes from a set of predefined categories that an object belongs to, based on a set of descriptors of the object. For example the text categorization task is to label an incoming message (document) with the label of one or more of the predefined classes. There have been a number of approaches to solving categorization problems by finding linear discriminant functions. In these approaches there are assumptions that each class has a Gaussian distribution. Least squares fit has also been used. Without any distributional assumptions a linear separator can be found by using a perceptron with minimization of the training error. Another approach that has been used in text categorization and information retrieval is logistic regression and this is closely related to support vector machines, which have recently had much success in text categorization.

In text categorization problems, although techniques have been developed for feature selection, interest has been primarily in classification and there has not been much interest in determining the features (words) that are responsible for assigning a particular document to a particular class.

*Corresponding author. Email: m.mammadov@ballarat.edu.au
In studies of adverse drug reactions (ADRs) on patients, a patient record consists of a list of the drugs that have been taken and the reactions that have been experienced. The question that is of interest is ‘Which drugs are responsible for each reaction?’ Certainly this is a classification problem, but the interest is focused on determining the features (drugs) that are most important in determining the class (reaction), rather than simply determining the class based on the set of drugs that the patient took.

In this paper we consider a situation in which we have $n$ records. In the case of text categorization these would be documents and in the case of ADRs each record would represent a patient with a number of drugs taken and various reactions that were observed or reported. Let the records or objects be $x_1, \ldots, x_n$, where $n$ is the number of records (documents/patients): each record (document/patient) is a vector of terms (words/drugs): $x_i = (x_{i1}, \ldots, x_{im})$, $i = 1, \ldots, n$. So $m$ is the number of terms (words/drugs). $x_{ij} = 1$ if the word (drug) $j$ is used in record $i$, $x_{ij} = 0$ if not.

In a classification task there may be two disjoint classes (binary classification), many disjoint classes (multi-class classification) or multiple overlapping classes (multi-label classification). In many cases multi-class classification problems are reduced to many binary classification problems. Below we look at multi-label classification, which is suited to ADR problems.

Let $c$ be the number of classes (reactions). Each record (document/patient) can belong to a number of these $c$ classes. We denote by $y_i = (y_{i1}, \ldots, y_{ic})$ the vector of classes for $x_i$ ($i = 1, \ldots, n$), where $y_{ij} = 1$, if $x_i$ belongs to the class $j$, and $y_{ij} = 0$ if not.

The problem of feature weighting is to find a weight matrix $W = (w_1, \ldots, w_m)$; that is, a weight vector $w_j = (w_{1j}, \ldots, w_{cj})$ for each feature (word/drug) $j$; such that, the vectors $Wx_i$ and $y_i$ ($i = 1, \ldots, n$) are close overall. This problem can be formulated in the form of different optimization problems. In this paper we use the following objective function introduced in [1]

$$w_p = \arg \inf_w \frac{1}{n} \sum_{i=1}^{n} (\|y_i\|)^{-p} \cdot \sum_{j=1}^{c} \left( \frac{\|y_i\|}{H_{ij}} - H_{ij} - y_{ij} \right)^2$$

Here $p = 0, 1, 2$, $H_{ij} = \sum_{q=1}^{m} w_{jq}x_{iq}$, and $\|y_i\|$ is the number of positive coordinates in the vector $(y_{i1}, \ldots, y_{ic})$.

In the application of this approach to ADR problems, $\|y_i\|$ is the number of reactions that occurred for patient $i$. To explain the advantage of this formula we refer to the example given in [1]. This example showed that, in terms of ADRs, equation (1) is more preferable than formulae used, for instance, by linear least squares fit (LLSF) and logistic regression (LR) (see [2] and references therein). At the same time, using functions such as equation (1) requires the solution of a significantly harder high-dimensional global optimization problem where the objective function is non-convex. In this paper we present a method that allows us to find quite ‘deep’ local minima of function (1).

The choice of parameter $p$ leads to different weight matrices $W$ (drug–reaction relationships in ADR problems). Which one is better? Of course, we cannot answer this question; for different situations different weight matrices could be better. That is why, it is very useful to consider different choices, $p = 0, 1, 2$ (algorithms $A(p)$).

Drug–reaction relationships, described by matrix $W$, can be used for the study of many ADR problems. In this paper, we consider two of them: the accurate identification of drugs that are responsible for reactions that have occurred, and drug–drug interactions. Based on drug–reaction relationships, we formulate these problems as an optimization problem. The numerical experiments are carried out on the cardiovascular-type reactions from the Australian Adverse Drug Reaction Advisory Committee (ADRAC) database. Based on the approach
described here, software has been developed for predicting reactions given a set of drugs and for identifying the most likely drugs responsible for given reactions.

2. Adverse drug reactions

An 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’ [3]. ADRs are estimated to be the fourth leading cause of death in the USA [4], and the amount of published literature on the subject is vast [5]. Some of the problems concerning ADRs are discussed in our research report [6]. 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) [7], 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 USA and approved by the food and drug administration. A common approach to the assessment of ADRs uses the Bayesian method [8]. For example, the Bayesian confidence propagation neural network (BCPNN) [9], an empirical Bayesian statistical data mining program, called a gamma Poisson shrinker (GPS) [10], and the multi-item gamma Poisson shrinker (MGPS) [11], which have been applied to the United States 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 [6,12] 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 this work the ADR problem was formulated as a text categorization problem having some peculiarities. This approach was applied to the ADRAC database.

One of the main problems of ADR is the following: given a patient (that is, the sets of drugs and reactions) identify the drug(s) that are responsible for the adverse reactions experienced. In the ADRAC database drugs thought to be responsible for the reactions are labelled as ‘suspected’ drugs. The accurate definition of suspected drugs for each report has a very significant impact on the quality of the database for the future study of drug–reaction relationships. In this paper, we develop the approach introduced in [1,6,12]. In particular, we consider drug interactions and discuss the possibility of using this approach in drug–drug interaction problems.

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. A more detailed account of the ADRAC database is given in [6].

In ADRAC there are 18 system organ class (SOC) reaction term classes, one of which is the cardiovascular SOC. The cardiovascular class consists of four sub-classes. In the present paper we will consider the part of the ADRAC data related to the cardiovascular, type of reactions. We collect all records having at least one reaction from these four sub-groups. We call this dataset Card20. In this dataset some records may have a reaction from outside the cardiovascular group. We define four classes according to these four sub-groups and additionally a fifth class that contains reactions belonging to the other 17 SOCs. For the number of records see table 1 in section 9.

The information about each patient consists of mainly two sets of information: individual patient information and information about drug(s) and reaction(s). In the current 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.

Not using patient information has the disadvantage of not studying the more complex relationship between patient characteristics, drugs and reactions. So, for example, our approach will not be used to identify groups of patients that are more likely to have classes of reactions.

Another reason for focusing 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 attributed to the voluntary nature of the ADRAC reporting system. Some of the problems of such a reporting system are discussed in [13–17].

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 [2,18,19]. In the current paper, together with the algorithm A(p), described below, we applied the algorithm BoosTexter (version AdaBoost.MH with real-valued predictions, [18]), which seems to be suitable for drug–reaction representations purposes.

3. 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 reactions (classes). Given patient \( x \in \mathcal{X} \), we denote by \( \mathcal{Y}(x) = (Y_1(x), Y_2(x), \ldots, Y_c(x)) \) a \( c \)-dimensional vector of reactions observed for this patient; where \( Y_i(x) = 1 \) if the reaction \( i \) has occurred, and \( Y_i(x) = 0 \) if it has not. Let \( D(x) \) be the set of all drugs taken by the patient \( x \). In the ADRAC data, the number of drugs reported for a patient is restricted to ten. Some of these drugs are reported as suspected drugs responsible in the reactions \( \mathcal{Y}(x) \). Therefore, we divide the set \( D(x) \) into two parts: \( DS(x) \) – the set of suspected drugs, and \( DN(x) \) – the set of non-suspected drugs. Clearly \( D(x) = DS(x) \cup DN(x) \), and it may be \( DN(x) = \emptyset \). We also note that, in the ADRAC data, for some patients, suspected drugs are reported in the form of interaction.

The goal of the study of drug–reaction relationships is to find a function \( h : \mathcal{D} \to R^c_+ \), where given drug \( d \in \mathcal{D} \) the components \( h_i \) of the vector \( h(d) = (h_1, h_2, \ldots, h_c) \) are the weights (‘probabilities’) of the occurrence of the reactions \( i = 1, 2, \ldots, c \). Here \( R^c_+ \) is the set of all \( c \)-dimensional vectors with non-negative coordinates.

In the next step, given a set of drugs \( \Delta \subset \mathcal{D} \), we need to define a vector

\[
H(\Delta) = (H_1(\Delta), H_2(\Delta), \ldots, H_c(\Delta))
\]  

(2)

where the component \( H_i(\Delta) \) indicates the probability of occurrence of the reaction \( i \) after taking the drugs \( \Delta \). In other words, we need to define a function \( H : S(\mathcal{D}) \to R^c_+ \), where \( S(\mathcal{D}) \) is the set of all subsets of \( \mathcal{D} \).

Let \( \Delta \subset \mathcal{D} \). The vectors \( h(d) \) show what kind of reactions are caused by the drugs \( d \in \Delta \). Therefore the vector \( H(\Delta) \) can be considered as potential reactions that could occur after taking the drugs \( \Delta \). 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 different reactions. Some reactions, which have potentially high degrees of occurrence, may not be observed because of the strong resistance of the patient to developing these reactions. But the existence of these potential reactions could have an influence on
the patient somehow. The results obtained in [6] have shown that the information about the existence of potential reactions (but that were not reported to ADRAC) helps to make prediction of reaction outcomes (bad and good) more precise.

The function $H$ can be defined in different ways and it is an interesting problem in terms of ADR(s). We will use the linear (sum) function $H(\Delta)$ (see [6]), where the components $H_i(\Delta)$ are defined as follows

$$H_i(\Delta) = \sum_{d \in \Delta} h_i(d), \quad i = 1, \ldots, c.$$  

(3)

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 greater effect (i.e. 0.4). This method seems more natural, because physically both drugs are taken by the patient, and the outcome could even be worse if there were drug–drug interaction(s).

Given patient $x \in \mathcal{X}$, we can define potential reactions $\mathcal{H}(x) = H(\Delta)$ corresponding to the set of drugs $\Delta \subset D(x)$. If $\Delta = D(x)$, then we have $\mathcal{H}(x) = \mathcal{H}^A(x) \doteq H(D(x))$, which means that all the drugs taken by the patient $x$ are used in the definition of potential reactions; whereas, if $\Delta = DS(x)$, then $\mathcal{H}(x) = \mathcal{H}^S(x) \doteq H(DS(x))$, which means that we use only suspected drugs neglecting all the others. We can also consider the potential reactions $\mathcal{H}(x) = \mathcal{H}^N(x) \doteq H(DN(x))$.

Therefore, drug–reaction relationships will be represented by vectors $h(d), d \in D$. The definition of these vectors depends on the drugs that are used in the calculations: we can use either all drugs or only suspected drugs. The evaluation of different drug–reaction representations can be defined by the closeness of two vectors: $\mathcal{H}(x)$, the vectors of potential (predicted) reactions, and $\mathcal{Y}(x)$, the vectors of observed reactions. We will use the evaluation measure average precision, presented in section 4, to describe the closeness of these reaction vectors.

The results obtained in [1] have shown the efficiency of using suspected (only) drugs for the calculation of drug–reaction relationships. Therefore, in the present paper all calculations will be carried out using drug–reaction relationships obtained from suspected drugs.

4. Evaluation measure: average precision

To evaluate the accuracy of established drug–reaction relations by a given classifier $(h, H)$; that is, to evaluate the closeness of the two vectors $\mathcal{H}(x)$ (predicted reactions) and $\mathcal{Y}(x)$ (observed reactions) we will use the average precision measure considered in [18]. Note that, this measure allows us to achieve more complete evaluation in multi-label classification problems.

Let $\mathcal{Y}(x) = \{l \in \{1, \ldots, c\} : \mathcal{Y}_l(x) = 1\}$ be the set of reactions that have been observed for the patient $x$ and $\mathcal{H}(x) = \{\mathcal{H}_1(x), \ldots, \mathcal{H}_c(x)\}$ 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

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

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

In the case, when the numbers $\mathcal{H}_i(x), i = 1, \ldots, c,$ are different, there is just one order $\tau$ satisfying this condition. But if there are reactions having the same potential reactions 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 $\text{rank}_\tau(x; l) = k$, where the number $k$ satisfies $\tau_k = l$. Then precision is defined as

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

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 $\tau$ then $P_\tau(x) = 1$. Clearly the number $P_\tau(x)$ depends on order $\tau$. We define

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

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

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

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

5. Formulation of the optimization problem

Given a vector $V = (V_1, \ldots, V_c)$, with non-negative coordinates, we will use the notation

$$\|V\| = \sum_{i=1}^{c} V_i \quad (4)$$

Let $x \in \mathcal{X}$. 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

$$\text{dist}(\mathcal{H}(x), \mathcal{Y}(x)) = \sum_{i=1}^{c} (\|\mathcal{Y}(x)\| \cdot \bar{\mathcal{H}}_i(x) - \mathcal{Y}_i(x))^2 \quad (5)$$

where the sign ‘bar’ stands for a normalization:

$$\bar{\mathcal{H}}_i(x) = \begin{cases} \frac{1}{\|\mathcal{H}(x)\|} \mathcal{H}_i(x) & \text{if } \|\mathcal{H}(x)\| > 0 \\ 0 & \text{if } \|\mathcal{H}(x)\| = 0 \end{cases} \quad (6)$$

The algorithm $A(p)$ uses the following distance measure (we assume that $\|\mathcal{Y}(x)\| > 0$)

$$\text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)) = \|\mathcal{Y}(x)\|^{-p} \cdot \text{dist}(\mathcal{H}(x), \mathcal{Y}(x)), \quad p = 0, 1, 2 \quad (7)$$

Note that, these distance functions are slightly different from the LLSF mapping function considered in [3,19].

To explain the difference between the distances $\text{dist}_p$, $p = 0, 1, 2$, consider the case $\|\mathcal{H}(x)\| > 0$. The following representation is true

$$\text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)) = \sum_{i=1}^{c} (a_i - b_i)^2$$
where \( a_i = (\|\mathcal{Y}(x)\|^{1-p/2}/\|\mathcal{H}(x)\|)\mathcal{H}_i(x), \ b_i = \|\mathcal{Y}(x)\|^{-p/2}\mathcal{Y}_i(x) \), and clearly
\[
\sum_{i=1}^{c} a_i = \sum_{i=1}^{c} b_i = \|\mathcal{Y}(x)\|^{1-(p/2)}
\]

In the distance \( \text{dist}_0 \) (that is, \( p = 0 \)) the sums \( \sum_i a_i \) and \( \sum_i b_i \) are equal to the number of reactions \( \|\mathcal{Y}(x)\| \). For \( \text{dist}_1 \) and \( \text{dist}_2 \) the corresponding sums are equal to \( \sqrt{\|\mathcal{Y}(x)\|} \) and 1, respectively. \( \text{dist}_1 \) can be considered as a middle version, because the number of reactions \( \|\mathcal{Y}(x)\| \geq 1 \) and therefore
\[
1 \leq \sqrt{\|\mathcal{Y}(x)\|} \leq \|\mathcal{Y}(x)\|
\]

From equations (5) to (7), it is not difficult to observe that the following property holds
\[
\text{dist}_p(\lambda \mathcal{H}(x), \mathcal{Y}(x)) = \text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)), \quad \text{for all} \ \lambda > 0 \quad (8)
\]

The algorithm \( A(p) \) aims to define drug–reaction relations \( h(d) \) minimizing the average distance \( \text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)) \) over all training examples. In other words, we consider the following optimization problem
\[
\begin{align*}
\text{minimize:} \quad & E_{av}^p = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)) \\
\text{subject to:} \quad & h_i(d) \geq 0, \ i = 1, \ldots, c, \ d \in \mathcal{D}
\end{align*} \quad (9)
\]

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

### 6. A solution to the optimization problem [equations (9), (10)]

The function in equation (9) is non-convex and has a large number of local minimum points. The number of variables is \( |\mathcal{D}| \cdot c \). For the data Card20, that we will consider, \( |\mathcal{D}| = 3001 \) and \( c = 5 \). Thus we have a global optimization problem with 15 005 variables, which is very hard to handle using existing global optimization methods. Taking into account some peculiarities of the problem, we suggest an algorithm that allows us to find sufficiently ‘deep’ local minimum point of the objective function in equation (9). We solve this problem in three steps.

**Step 1.** First we find some ‘good’ initial point for the problem equations (9), (10). In this stage we use a method developed in [1] (see section 6.1). Let this initial point be
\[
h^0(d) = (h^0_1(d), \ldots, h^0_c(d)), \quad d \in \mathcal{D}. \quad (11)
\]

**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
\[
h(d) = \lambda(d)(h^0_1(d), \ldots, h^0_c(d)), \quad d \in \mathcal{D}. \quad (12)
\]

Then we consider the global optimization problem with box constraints
\[
\begin{align*}
\text{minimize:} \quad & \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \text{dist}_p(\mathcal{H}(x : \lambda), \mathcal{Y}(x)) \\
\text{subject to:} \quad & 0 \leq \lambda(d) \leq \lambda_{\text{max}}, \quad d \in \mathcal{D}
\end{align*} \quad (13)
\]
where \( \mathcal{H}(x : \lambda) = (\mathcal{H}_1(x : \lambda), \ldots, \mathcal{H}_c(x : \lambda)) \) and

\[
\mathcal{H}_i(x : \lambda) = \sum_{d \in \Delta(x)} \lambda(d) h^0_i(d)
\]

The number \( \lambda_{\text{max}} \) could be chosen any positive number. We set \( \lambda_{\text{max}} = 2 \) letting the values \( \lambda(d) = 1 \), which correspond to the initial point \( h^0(d) \), to be in the centre of the box, equation (14).

The number of variables in this problem is equal to the number of drugs; that is, 3001 in the calculations below.

The use of equation (12) 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% of all records), the scaling (12) does not affect the classification (that is, the value \( P_{av} \)). The effect of this scaling works when two or more drugs were involved together. The large numbers \( \lambda(d) \) obtained as a solution to equations (13) and (14) will indicate the importance of these drugs in terms of causing reactions.

To solve this problem we apply the algorithm AGOP (algorithm for global optimization problems), developed in [20,21] which is designed to solve global optimization problems with box constraints (see section 6.3). Let the optimal solution to the problem equations (13) and (14), be \( \lambda^0(d), d \in \mathcal{D} \), which provides the weight vectors

\[
h^i(d) = (h^1_1(d), \ldots, h^1_c(d)), \quad h^0_i(d) = \lambda^0(d) h^0_i(d), \quad d \in \mathcal{D}, \quad i = 1, \ldots, c.
\]

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

\[
\min_{h(d)} E_{av} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x))
\]

s.t. \( \theta_1 h^1_i(d) \leq h_i(d) \leq \theta_2 h^1_i(d), \quad d \in \mathcal{D}, \quad i = 1, \ldots, c \) (17)

In the calculations below, we set \( \theta_1 = 0.01, \theta_2 = 2 \). The number of variables in this problem is equal to \( |\mathcal{D}|c \); that is, 15 005. To solve this problem we apply the algorithm AGOP. We denote the solution found by

\[
h^{GO}(d) = (h^{GO}_1(d), \ldots, h^{GO}_c(d)), \quad d \in \mathcal{D}.
\]

### 6.1 Step 1: Calculation of initial points

We suggest one heuristic method for finding a ‘good’ solution to the problem, equations (9) and (10). This method is based on the proposition given below.

We denote by \( S \) the unit simplex in \( R^c \); that is

\[
S = \{ h = (h_1, \ldots, h_c): h_i \geq 0, \quad h_1 + \cdots h_c = 1 \}.
\]

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 \( \mathcal{X} \), which used just one drug, \( d \). Simply, the set \( X(d) \) combines all records where the drug \( d \) was used alone.

Consider the problem

\[
\min_{h(d)} f(h(d)) = \sum_{x \in \mathcal{X}(d)} \| \mathcal{Y}(x) \|^{-p} \cdot \sum_{j=1}^c (\| \mathcal{Y}(x) \| h_j(d) - \mathcal{Y}_j(x) )^2
\]

s.t. \( h(d) = (h_1(d), \ldots, h_c(d)) \in S \) (19)
**Proposition 6.1** A point $h^*(d) = (h^*_1(d), \ldots, h^*_c(d))$, where

$$h^*_j(d) = \left( \sum_{x \in X(d)} \|Y(x)\|^{2-p} \right)^{-1} \cdot \sum_{x \in X(d)} \|Y(x)\|^{1-p} Y_j(x), \quad j = 1, \ldots, c, \quad (21)$$

is the global minimum point for the problem equations (19) and (20).

**Proof** For each $j \in \{1, \ldots, c\}$, from equation (19) we have

$$\frac{\partial f}{\partial h_j(d)} = 2 \sum_{x \in X(d)} \|Y(x)\|^{1-p}(\|Y(x)\| h_j(d) - Y_j(x)) = 0.$$

Solving this equation with respect to $h_j(d)$, we obtain equation (21). We also note that,

$$\sum_{j=1}^c Y_j(x) = \|Y(x)\|,$$

and therefore

$$\sum_{j=1}^c h^*_j(d) = \left( \sum_{x \in X(d)} \|Y(x)\|^{2-p} \right)^{-1} \cdot \sum_{x \in X(d)} \|Y(x)\|^{1-p} \sum_{j=1}^c Y_j(x) = 1.$$

This means that solution $h^*(d)$ satisfies condition (20).

Proposition is proved. ■

Now, given drug $d$, we consider the set $X_{all}(d)$, which consists of 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 equations (19) and (20). In this case, we will use the following heuristic approach to find a ‘good’ solution.

(S) (single). 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 $h_j(d), (j = 1, \ldots, c)$ only by this set neglecting all the mixed cases.

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 both numbers. We will consider a function $\phi(d) = a|X(d)| + bP(d)$ to describe how large the set $X(d)$ is.

Therefore, if the number $\phi(d) \geq p^*$, where $p^*$ is an a priori given number, then we use only the set $X(d)$ to calculate weights $h(d)$; in other words, we use equation (21), which provides a global minimum $h(d) = h^*(d)$ for the part of data $X(d) \subset \mathcal{X}$.

We denote by $\mathcal{D}'$ the set of all drugs from $\mathcal{D}$ for which the weights are calculated in this way.

(M) (mixed). If the set $X(d)$ is not ‘sufficiently large’; that is, $\phi(d) < p^*$, then we have to use the set $X_{all}(d)$ which contains patients $x \in \mathcal{X}$ having more than one drug taken. In this case we use $h(d) = h^{**}(d)$; where

$$h^{**}_j(d) = \left( \sum_{x \in X_{all}(d)} \|Y(x)\|^{2-p} \right)^{-1} \cdot \sum_{x \in X_{all}(d)} \|Y(x)\|^{1-p} \frac{\text{rem}(Y_j(x))}{|\Delta'(x)|}, \quad j = 1, \ldots, c, \quad (22)$$

Here, given a patient $x$, the set $\Delta''(x) = \Delta(x) \setminus \mathcal{D}'$ combines all drugs, the weights for which are not calculated in the first step. Note that, $\Delta(x)$ is the set of drugs corresponding to the patient $x$, and we will consider either $\Delta(x) = D(x)$ (all drugs) or $\Delta(x) = DS(x)$ (suspected drugs).
We denote by \( \phi(Z_{\text{max}}) \), \( j \in \{1, \ldots, c\} \), stands for the ‘remaining’ part of the reaction \( \mathcal{Y}_j(x) \), associated with the drugs \( \Delta''(x) \). For the calculation of \( \phi(Z_{\text{max}}) \) see section 6.2.

This formula has the following meaning. If \( |\Delta''(x)| = 1 \) for all \( x \in X_{\text{all}}(d) \), then, given \( \phi(Z_{\text{max}}) \), \( j = 1, \ldots, c \), equation (22) provides a global minimum solution (similar to equation (21)). If \( |\Delta''(x)| > 1 \), for some patient \( x \in X_{\text{all}}(d) \), then we use the assumption that all suspected drugs are responsible to the same degree; that is, for this patient, we associate only the part \( 1/|\Delta''(x)| \) of the reactions \( \phi(Z_{\text{max}}) \) to this drug.

Therefore, a ‘good’ initial point for the problem equations (9) and (10), that is, the vectors \( h(d) = (h_1(d), \ldots, h_c(d)) \) in section (6.1) will be defined as follows

\[
h(d) = \begin{cases} 
    h^*(d) & \text{if } \phi(d) \geq p^* \\
    h^{**}(d) & \text{otherwise}
\end{cases}
\]  

(23)

where \( h^*(d) \) and \( h^{**}(d) \) are defined by equations (21) and (22), respectively.

Remark 6.2 We note that the weight \( h_i(d) \) is not exactly a probability of the occurrence of the reaction \( i \); that is, the sum \( \sum_{i=1}^c h_i(d) \) does not need to be equal 1.

6.2 Calculation of \( \phi(Z(x)) \)

Consider a particular patient \( x \in \mathcal{X} \). We divide the set of drugs \( \Delta(x) \) (which could be all drugs or suspected drugs used by this patient) into two parts

\[
\Delta(x) = \Delta'(x) \cup \Delta''(x)
\]

where for each drug \( d \), in the set \( \Delta'(x) \), the vector \( h(d) \) has already been defined by equation (21), and the set \( \Delta''(x) \) combines all the other drugs. Note that it may be \( \Delta'(x) = \emptyset \).

We set \( \mathcal{H}(x) = G(x) + Z(x) \) where the sum \( G(x) = \sum_{d \in \Delta'(x)} h(d) \) defines a part of potential reactions associated with the drugs \( \Delta'(x) \), and \( Z(x) \) the other (unknown) part which will be defined by the drugs \( \Delta''(x) : Z(x) = \sum_{d \in \Delta''(x)} h(d) \).

We will use a reasonable assumption that all drugs in \( \Delta''(x) \) are responsible in the observed reactions \( \mathcal{Y}(x) \) to the same degree; that is, we associate equal parts \( 1/|\Delta''(x)| \) of \( Z(x) \) to each drug in \( \Delta''(x) \). Therefore, after accepting such an assumption, we need only to find \( Z(x) \) trying to minimize the distance \( \text{dist}_p(\mathcal{H}(x), \mathcal{Y}(x)) \), \( p = 0, 1, 2 \). This is equivalent [see equation (7)] to minimizing the distance \( \text{dist}(\mathcal{H}(x), \mathcal{Y}(x)) \).

As we consider a particular patient \( x \), for the sake of simplicity, we will drop the variable \( x \). Therefore, to find the vector \( Z = (Z_1, \ldots, Z_c) \), we need to solve the following problem

\[
\begin{align*}
\text{minimize:} & \quad \sum_{i=1}^c \left( \frac{\|\mathcal{Y}\| (G_i + Z_i)}{\sum_{j=1}^c (G_j + Z_j)} - \mathcal{Y}_i \right)^2 \\
\text{subject to:} & \quad 0 \leq Z_i, \quad G_i + Z_i \leq Z_{\text{max}}, \quad i = 1, \ldots, c
\end{align*}
\]  

(24) (25)

We denote by \( \phi(Z_{\text{max}}) \) the optimal value of the objective function in problem (24), (25). Denote \( I_0 = \{i \in \{1, \ldots, c\} : \mathcal{Y}_i = 0\} \); \( I_1 = \{1, \ldots, c\} \setminus I_0 = \{i \in \{1, \ldots, c\} : \mathcal{Y}_i = 1\} \). The cardinality of the set \( I_1 \) is \( \|\mathcal{Y}\| \). Let \( G^0 = \max\{G_i : i \in I_0\} \); \( G^1 = \max\{G_i : i \in I_1\} \).
Therefore, it is not difficult to see that, at the point $Z$, the value of objective function is zero; that is, this point is optimal. Here

PROPOSITION 6.3

Proof

If $G^0$ is the optimal solution to the problem

1. First we fix the variables $\xi_i$, $i \in I_0$, taking $\xi_i = \xi_i^*$, $i \in I_0$. Denote $a = \sum_{i \in I_0} \xi_i^*$. Consider the following functions with respect to variables $(\xi_i)_{i \in I_1}$

$$g_0((\xi_i)_{i \in I_1}) = \sum_{i \in I_0} \left( \frac{\|Y\| \xi_i^*}{a + \sum_{j \in I_1} \xi_j} \right)^2, \quad g_1((\xi_i)_{i \in I_1}) = \sum_{i \in I_1} \left( \frac{\|Y\| \xi_i}{a + \sum_{j \in I_1} \xi_j} - 1 \right)^2$$

By assumption, the point $(\xi_i^*)_{i \in I_1}$ is an optimal solution to the problem

\begin{align*}
\text{minimize:} & \quad g_0((\xi_i)_{i \in I_1}) + g_1((\xi_i)_{i \in I_1}) \quad \text{s.t.} \quad G_i \leq \xi_i \leq Z_{max}, \quad i \in I_1
\end{align*}

Consider the following problem

\begin{align*}
\text{minimize:} & \quad g_1((\xi_i)_{i \in I_1}) \quad \text{s.t.} \quad \xi_i \leq Z_{max}, \quad i \in I_1
\end{align*}

Assume that $(\xi_i^0)_{i \in I_1}$ is an optimal solution to equation (31). We show that all the coordinates $\xi_i^0$, $i \in I_1$, are equal.
Let $\xi_k^0 < \xi_m^0$ for some $k, m \in I_1$. We construct a new point $\xi_i^\prime, i \in I_1$, taking $\xi_i^\prime = \xi_i^0, i \in I_1, i \neq k, m$, and $\xi_k^\prime = \xi_m^\prime = \xi''$, where $\xi''$ is the middle point of $\xi_k^0$ and $\xi_m^0$; that is, $\xi'' = \xi_k^0 + \delta, \delta = (\xi_m^0 - \xi_k^0)/2$. Clearly $\sum_{j \in I_1} \xi_j = \sum_{j \in I_1} \xi_j^0$. Denote $b = \|Y\|/(a + \sum_{j \in I_1} \xi_j^0)$.

Then, we have
\[
g_1((\xi_i^\prime)_{i \in I_1}) - g_1((\xi_i^0)_{i \in I_1}) = [b\xi'' - 1]^2 - [b\xi'' - 1]^2 - [b(\xi'' - \delta) - 1]^2
\]
\[-[b(\xi'' + \delta) - 1]^2 = -2b^2\delta^2 < 0
\]
which means that $(\xi_i^0)_{i \in I_1}$ is not optimal in problem (31).

This contradiction shows that all the coordinates $\xi_i^0, i \in I_1, are equal. Let $\xi_i^0 = \xi^0, i \in I_1$.

Now we show that $\xi^0 = Z_{\text{max}}$. Let $\xi^0 < Z_{\text{max}}$. Then
\[
\frac{\|Y\|\xi^0}{a + \|Y\|\xi^0} < \frac{\|Y\|Z_{\text{max}}}{a + \|Y\|Z_{\text{max}}} < 1
\]
(32)

Consider a new point $\xi_i^{\text{max}} = Z_{\text{max}}, i \in I_1$. As $|I_1| = \|Y\|$, from equation (32) we obtain
\[
g_1((\xi_i^{\text{max}})_{i \in I_1}) < g_1((\xi_i^0)_{i \in I_1})
\]
which is a contradiction.

Therefore, the point $\xi_i^{\text{max}}, i \in I_1$ is the optimal solution to problem (31). Then, it is not difficult to see that this point is also optimal in problem (30), because function $g_0((\xi_i)_{i \in I_1})$ is decreasing with respect to each variable $\xi_i$. Thus $\xi_i^* = Z_{\text{max}}, i \in I_1$.

2. Letting $\xi_i = \xi_i^{\text{max}} = Z_{\text{max}}, i \in I_1$ in equation (28), we have the following problem with respect to variables $\xi_j, j \in I_0$

\[
\text{minimize: } \sum_{i \in I_0} f_i((\xi_j)_{j \in I_0}) + \sum_{i \in I_1} F((\xi_j)_{j \in I_0})
\]
\[
\text{subject to: } G_j \leq \xi_j \leq Z_{\text{max}}, j \in I_0.
\]
(34)

Here
\[
f_i((\xi_j)_{j \in I_0}) = \left(\frac{\|Y\|\xi_i}{\sum_{j \in I_0} \xi_j + \|Y\|Z_{\text{max}}}\right)^2,
\]
\[
F((\xi_j)_{j \in I_0}) = \left(\frac{\|Y\|Z_{\text{max}}}{\sum_{j \in I_0} \xi_j + \|Y\|Z_{\text{max}}} - 1\right)^2
\]

By assumption, the point $\xi_j = \xi^*, j \in I_0$, is an optimal solution to this problem. It is not difficult to see that $F((\xi_j)_{j \in I_0})$ is an increasing function with respect to each variable $\xi_j$. We have the same for each function $f_i((\xi_j)_{j \in I_0})$. Therefore, we obtain that
\[
\xi_j^* = G_j, \quad \text{for all } j \in I_0
\]
3. Therefore we have shown that the optimal solution $\xi^*$ to problem (28), (29) satisfies
\[
\xi_i^* = G_i, \quad \forall i \in I_0 \quad \text{and} \quad \xi_i^* = Z_{\text{max}}, \quad \forall i \in I_1
\]

This means that, the point $Z_i^*, i = 1, \ldots, c$, defined by equation (27) is the unique optimal solution to problem (24), (25) in the case $G^0 > 0$. 

\[
M. A. Mammadov et al.
\]
The study of drug–reaction relationships

The value of the objective function (24) at this solution is

\[ \phi(Z^\text{max}) = \sum_{i \in I_0} \left( \frac{\|Y\| G_i}{\sum_{j \in I_0} G_j + \|Y\| Z^\text{max}} \right)^2 + \sum_{i \in I_1} \left( \frac{\|Y\| Z^\text{max}}{\sum_{j \in I_0} G_j + \|Y\| Z^\text{max} - 1} \right)^2 \]

From this formula we obtain \( \phi(Z^\text{max}) \to 0 \) as \( Z^\text{max} \to \infty \).

The proposition is proved.

This proposition shows that for the optimal solution \( Z^* \) the sums \( Z_i^* + G_i, i \in I_1 \), are constant being equal to \( Z^\text{max} \).

We can decrease the distance \( \phi(Z^\text{max}) \) by increasing the number \( Z^\text{max} \). Note that, \( Z^\text{max} \) serves to restrict the values \( Z_i \) in order to obtain

\[ \max_{i=1,...,c} (G_i + Z_i) = 1, \] (35)

which means that the patient \( x \) would be taken into account with the weight 1 (like the patients in \( X(d) \)). Therefore, we need to chose a number \( Z^\text{max} \) close to 1.

We will define the number \( Z^\text{max} \) as follows

\[ Z^\text{max} = \max\{1, G^0 + \varepsilon, G^1\}, \quad \text{where } \varepsilon > 0. \]

The choice of such a number \( Z^\text{max} \) has the following meaning. First we note that, if \( G^0 < 1 \) and \( G^1 \leq 1 \) then there is a number \( \varepsilon > 0 \) such that \( Z^\text{max} = 1 \); that is, (35) holds. On the other hand, if \( G^0 \geq 1 \) and \( G^0 \geq G^1 \), then the weights \( Z_i^* + G_i = Z^\text{max} \), corresponding to the occurred reactions \( i \) (that is, \( Y_i = 1 \)) will be greater than the weights \( Z_i^* + G_i = G_i \), corresponding to the non-occurred reactions. In this case, choosing the number \( \varepsilon > 0 \) smaller, we get more closer approximation to equation (35).

Therefore, \( \text{rem}(Y_j(x)) \) will be defined as

\[ \text{rem}(Y_j(x)) = Z_j^*, \quad j = 1, \ldots, c \] (36)

### 6.3 Steps 2 and 3: global optimization

Now we consider the global optimization problem (13), (14), with respect to variables \( \lambda(d), d \in D \). As mentioned before, the number of variables in these problems is too large (3001 and 15 005) to handle in terms of global optimization. We also note that the calculation of values of the objective function is time consuming, and, therefore the use of existing global optimization methods is extremely restricted. In [20] we introduced a new global optimization algorithm named AGOP. In terms of both the time consumption and the accuracy of results obtained, this algorithm provided very good results in comparison with other global optimization algorithms. In this paper we will apply the algorithm AGOP for solving problem (13), (14).

AGOP is designed for solving unconstrained continuous optimization problems. It uses a line search mechanism where the descent direction is defined via a dynamical systems approach.

It can be applied to a wide range of functions, requiring only function evaluations to work. In particular it does not require gradient information and can be used to find minima of non-differentiable functions.
Briefly, AGOP works as follows (for more detailed description see [20,21]). Consider the problem

$$\text{minimize } f(x) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \text{s.t. } x \in B$$

where $B$ is a given box constraints. AGOP must first be given a set of points, say $\Omega_1 = \{x_1, \ldots, x_q\} \subset \mathbb{R}^n$. Generally, a suitable choice for an initial set of points is generated from the vertices of the box $B$.

Suppose that $x_\ast \in \Omega_1$ has the smallest cost of the points in $\Omega_1$, that is, that $f(x_\ast) \leq f(x)$ for all $x \in \Omega_1$. The set $\Omega_1$ and the values of $f$ at each of the points in $\Omega_1$ allow us to generate a dynamical system. This dynamical system determines a possible descent direction $v$ at the point $x_\ast$ (see [20,21] for details). An inexact line search along this direction provides a new point $\hat{x}_{q+1}$. A local search about $\hat{x}_{q+1}$ is then carried out. This is done using a direct search method called local variation. This is an efficient local optimization technique that does not explicitly use derivatives and can be applied to nonsmooth functions. A good survey of direct search methods can be found in [22]. Letting $x_{q+1}$ denote the optimal solution of this local search, the set $\Omega_1$ is augmented to include $x_{q+1}$. Starting with this updated $\Omega_1$, the whole process can be repeated. The process is terminated when $v$ is approximately 0 (or a prescribed bound on the number of iterations is reached). The solution returned is the current $x_\ast$, that is, the point in $\Omega_1$ with the smallest cost.

### 7. Evaluation of correctness of suspected drugs reported

Drug–reaction representations in the form of a vector of weights allow us to evaluate the correctness of suspected drugs reported.

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 suspected drugs $DS(x)$ and non-suspected drugs $DN(x)$. Our aim in this section is to evaluate how correctly suspected drugs are identified.

The method of evaluation is based on distance measure (5). Assume that for each drug $d \in D$ the vector of weights $h(d)$ are calculated. Then we can define potential reactions $\mathcal{H}^S(x)$ and $\mathcal{H}^N(x)$, corresponding to the sets of suspected drugs and non-suspected drugs, respectively. We have

$$\mathcal{H}^S_i(x) = H_i(DS(x)) = \sum_{d \in DS(x)} h_i(d), \quad i = 1, \ldots, c;$$

$$\mathcal{H}^N_i(x) = H_i(DN(x)) = \sum_{d \in DN(x)} h_i(d), \quad i = 1, \ldots, c.$$
The study of drug–reaction relationships

113

The number \( \mu \) indicates the proportion of suspected and non-suspected drugs in the definition of potential reactions. Clearly, \( \bar{H}(x, 1) = \bar{H}^S(x) \) and \( \bar{H}(x, 0) = \bar{H}^N(x) \), which implies

\[
\text{dist}(\bar{H}(x, 1), \mathcal{Y}(x)) = \text{dist}(\bar{H}^S(x), \mathcal{Y}(x))
\]

\[
\text{dist}(\bar{H}(x, 0), \mathcal{Y}(x)) = \text{dist}(\bar{H}^N(x), \mathcal{Y}(x))
\]

It is important to note that, the combination of all drugs with equal weights; that is, the vector \( \bar{H}_A(x) = H(D(x)) = \bar{H}^S(x) + \bar{H}^N(x) \) is also considered in equation (37). To confirm this, it is sufficient to consider the case \( \|H^S(x)\| > 0 \) and \( \|H^N(x)\| > 0 \). Then we take \( \mu' = \|H^S(x)\| / \|H_A(x)\| = (0, 1) \) and get [see equation (6)]

\[
\bar{H}(x, \mu') = \mu' \frac{H^S(x)}{\|H^S(x)\|} + (1 - \mu') \frac{H^N(x)}{\|H^N(x)\|} = \bar{H}^A(x)
\]

which implies

\[
\text{dist}(\bar{H}(x, \mu'), \mathcal{Y}(x)) = \text{dist}(\bar{H}^A(x), \mathcal{Y}(x)) = \text{dist}(\bar{H}^A(x), \mathcal{Y}(x))
\]

Consider the following minimization problem with respect to \( \mu \)

\[
\text{minimize: } f(\mu) = \text{dist}(\bar{H}(x, \mu), \mathcal{Y}(x)), \quad \text{s.t. } 0 \leq \mu \leq 1
\] (38)

The optimal solution \( \mu^* \) to problem (38) gives an information about the correctness of definition of suspected drugs. For instance, if \( \mu^* = 1 \) then we see that the suspected drugs provide the better approximation to the observed reactions than if we involve the other drugs. We refer this situation as 100% correctness. Whereas, if \( \mu^* = 0 \) then non-suspected drugs provide better approximation to the observed reactions and we can conclude that in this case suspected drugs are defined completely wrong. Therefore, the optimal value \( \mu^* \) can be considered as an evaluation measure for the correctness of suspected drugs.

From equation (7) we obtain the following:

**Proposition 7.1** The optimal solution \( \mu^* \) to the problem (38) is optimal with respect to the all distance measures \( \text{dist}_p \), \( p = 0, 1, 2 \); that is, given vectors of weights \( h(d), d \in D(x) \), for all \( p = 0, 1, 2 \) the following inequality holds

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

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

It is clear that, problem (38) can have many optimal solutions \( \mu^* \); that is, different proportions of suspected and non-suspected drugs can provide the same closeness to the observed reactions. In this case we will define the correctness of suspected drugs, as the maximal value among all optimal solutions \( \mu^* \)

\[
\mu^*(x) = \max\{\mu^*: \mu^* \text{ is an optimal solution to equation (38)}\}.
\] (39)

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

Problem (38) can be easily solved. Let

\[ A = \|Y(x)\|^2 \sum_{i=1}^{c} (\tilde{H}^S_i(x) - \tilde{H}^N_i(x))^2 \]

\[ B = \|Y(x)\| \sum_{i=1}^{c} (\tilde{H}^S_i(x) - \tilde{H}^N_i(x)) (\|Y(x)\| \tilde{H}^N_i(x) - Y_i(x)) \]

Then, we find the derivative of the function \( f(\mu) \), defined by equation (38), in the following
form

\[ f'(\mu) = 2(A\mu + B) \] (40)

We note that \( A \geq 0 \). Moreover, if \( A = 0 \) then \( B = 0 \). Therefore, from equation (40) we
have

**Proposition 7.2** The optimal solution \( \mu^*(x) \) to the problem (38), satisfying equation (39),
can be found as follows.

1. If \( A = 0 \) then \( \mu^*(x) = 1 \).
2. If \( A > 0 \) then

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

Therefore, we have defined the correctness of suspected drugs for a particular patient \( x \).
Given set of patients \( \mathcal{X} \), average correctness of suspected drugs will be calculated as

\[ P_{sus} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \mu^*(x) \] (41)

**7.1 Remark**

The above definition of correctness of suspected drugs can be considered as a method where
the group of suspected drugs (already defined) are taken versus the group of non-suspected
drugs. In fact, drug–reaction representations in the form of vectors of weights allow us to
consider more general statements of this problem. Here, we formulate some of them.

Consider a patient \( x \in \mathcal{X} \). Let \( D(x) = \{d_1, d_2, \ldots, d_{|D(x)|}\} \) be the set of drugs that have
been taken, and \( h(d), d \in D(x) \), be the vectors of weights for these drugs. We introduce a
\( |D(x)| \)-dimensional vector \( v = (v_1, v_2, \ldots, v_{|D(x)|}) \), where the \( i \)th component \( v_i \) indicates
the degree of responsibility of the drug \( d_i \) in the observed reactions. In this case, potential
reactions can be defined as follows

\[ \mathcal{H}(x, v) = \sum_{i=1}^{|D(x)|} v_i h(d_i). \]

Then we can consider the following optimization problem

minimize: \( \text{dist}(\mathcal{H}(x, v), Y(x)) \) s.t. \( 0 \leq v_i \leq 1, \quad i = 1, \ldots, |D(x)| \) (42)

If, for instance, the optimal value for a particular drug \( i \) is 0.6 (that is, \( v_i = 0.6 \)) then
we say that the degree of responsibility of this drug in the observed reactions is 0.6. Such
information about drugs is more complete than just saying ‘suspected’ or ‘non-suspected’. But the application of this method encountered the absence of any such kind of classification of suspected drugs in the ADRAC data.

We can also consider a special case when each variable $v_i$ takes only two values: 1 (which means that $i$ is a suspected drug) and 0 (which means that $i$ is a non-suspected drug). In this case, we obtain a combinatorial optimization problem which is to find an optimal subset of drugs that provides the closest approximation to the observed reactions.

The application of problem (42) allows us to study suspected drugs in each report more precisely. It is our opinion that, the determining of responsibility of each drug from the set of drugs that have been taken, is a very important problem in terms of ADRs. But there are some issues that should be mentioned.

Such a precise statement of the problem should be accomplished with more precise definitions of function $H(x)$ [in this paper we use equation (3)] and weights $h(d)$, which have even more impact on the results. First of all, the times of starting and withdrawing drugs should be taking into account. Such information is presented in the ADRAC data but more research needs to be done in this area. The other factor that could be helpful for more precise definitions of weights $h(d)$, relates to the amount of general use of each drug, and the difficulty of getting such information is the major factor in ADR problems.

8. 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 [23], where the drug–drug interaction problem is described as follows:

**Definition 8.1 ([23])** ‘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 side effects (ADRs). For the sake of definiteness, we have to consider these effects separately.

One aspect of definition 8.1 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;

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 8.1. Such situations may occur using only one drug (in the part of data ADRAC related to the cardiovascular-type of reactions there are three 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 (reactions, features in the body). 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 3. Now we discuss how this approach matches with definition 8.1 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.

For the sake of simplicity, consider two drugs \( d_1, d_2 \), with weights \( h(d_1), h(d_2) \in \mathbb{R}_+^c \), and assume that \( H(x) \geq h(d_1) + h(d_2) \) 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 \( D_1 \) (except for the part related to reduction of effect) and part \( D_3 \) of definition 8.1. The more interesting case is \( D_2 \). We now consider this case in detail.

**The case \( D_2 \).**

The reality is that, one drug causes not one but many different reactions. The grouping of similar reactions under one class (such as the 4 reaction classes used in this paper for the cardiovascular-type of reactions) 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 are 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, \( D_2.2 \), can be explained by drug–reaction relationships in the form of vectors \( h(d) \). Consider an example. Let there be five reaction classes and

\[
\begin{align*}
  h(d_1) &= (0.6, 0.4, 0.0, 0.0, 0.0), \\
  h(d_2) &= (0.0, 0.4, 0.6, 0.0, 0.0).
\end{align*}
\]

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 \( D_2.1 \) which coincides with \( D_2 \) of definition 8.1. 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. How to study the case D2.1  To be able to study this case, we have to consider different combinations of drugs, under interaction. Moreover, we need to have enough records with the same combination of drugs, in order, to derive statistically significant results.

The calculations have shown that, this is not the case when some combination of drugs has been observed many times. For example, in the data considered here (which combines all 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 2 times (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 under interactions! Therefore, this 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 (say defined by trade names as in this paper). Therefore, the case D2.1 needs substantial investigations including generating drug classes.

Note 2. How frequently the case D2.1 occur  In other words, we want to know, how many records out of 364 drug–drug interactions (in this paper, we concentrate on cardiovascular type of reactions), 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 related to the explicit definition of what it means to say a drug, \( d \), is used alone. There may be three different definitions:

a. drug \( d \) is the only drug that has been taken (in this case, of course, it is also a suspected drug);

b. drug \( d \) is the only drug reported as a suspected drug out of all the drugs (more than two) that have been taken; and

c. 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 types associated to this drug according to its used alone cases. 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 for the used alone cases, 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 the records (having of course, no drug interactions). Then, we define that, a record, having the drugs \( d_1, \ldots, d_m \), as interacted, 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 calculations have provided the following results for the number of records corresponding to the case D2.1:

a. there are seven records according to Data1;

b. there are four records according to Data2;

c. there is only one record 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 under interactions were indicated more correctly. To explain this, consider an example. The only record in Data3 (dated 1987) uses 4 drugs: ID numbers 460, 782, 3498, 4714 and the vector of observed reactions is (1, 0, 0, 0, 1), which means that the first and fifth reactions have been observed. According to Data3, the frequency of reactions related to these drugs are: drug 460 ⇒ (0, 0, 0, 0), drug 782 ⇒ (35, 8, 33, 48, 117), drug 3498 ⇒ (50, 5, 21, 17, 83) and drug 4714 ⇒ (0, 0, 2, 1, 0). In this record, drugs 460 and 4714 were reported as a drug–drug interaction. 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 a mistake (or misprint), because these drugs frequently caused the first and the fifth reactions. If, it was a mistake, then this record would not be classified as D2.1.

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.

In this paper, we are dealing with drugs under interaction as suspected drugs, and define the vectors \( h(d) \) for each drug. Then we are trying to consider how this knowledge could be used for the prediction of reactions reported as a case of drug–drug interaction. The high accuracy in such prediction will indicate the usefulness of the proposed approach.

Here we have to mention one very important issue. The records with interaction can be divided into two parts: the first part (210 records) combines all records where all the drugs taken are reported as interacting, and the second part (154 records), where only some of 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 the previous section 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.

We will use the following two methods for the evaluation of correctness.

First we will use the methodology developed in the previous section. We divide the set of drugs \( D(x) \), taken by a particular patient \( x \), into two subsets: \( I(x) \) is the set of drugs which are reported as interaction, \( O(x) \) is the set of all other drugs. We will consider only records \( x \) where the set \( I(x) \) contains at least two drugs and the set \( O(x) \) is not empty.

As in the previous section, we define potential reactions \( \hat{H}^I(x) \) and \( \hat{H}^O(x) \), corresponding to the drugs \( I(x) \) and \( O(x) \), respectively. Then we consider convex combinations of these vectors [see equation (37)]

\[
\hat{H}(x, \mu) = \mu \hat{H}^I(x) + (1 - \mu) \hat{H}^O(x), \quad 0 \leq \mu \leq 1.
\]

Similar to equations (38) and (39), the maximal optimal value \( \mu = \mu^{**}(x) \) which minimizes the distance \( \text{dist}(\hat{H}(x, \mu), \mathcal{Y}(x)) \), will be taken as the degree of responsibility of the drugs \( I(x) \) in the observed reactions \( \mathcal{Y}(x) \).

Given a set of patients \( \mathcal{X} \), average responsibility of drugs in interaction will be calculated as

\[
P_{int} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \mu^{**}(x)
\]
We also apply the evaluation measure presented in section 4. This will provide a precision $P(x)$ calculated for each record $x$ having interaction effects of drugs. Then we calculate the average precision $P_{av}$ for all these records.

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

9. The results of numerical experiments

9.1 Algorithms

In the analysis below, together with the algorithm $A(p)$, we also use the algorithm BoosTexter which provides information about the possible accuracy that could be achieved in the experiments below. We will use BoosTexter, version AdaBoost.MH with real-valued predictions ([18]). Given a patient $x$, this algorithm produces a prediction vector that will be denoted by $B(x) = (B_1(x), \ldots, B_c(x))$, where the numbers $B_i(x)$ are real values which can be positive or negative. In other words, it defines potential reactions that we are interested in.

Therefore, for each patient $x$, these two algorithms produce potential (predicted) reactions in the form of vectors $H(x)$ and $B(x)$. The methods of calculating these vectors are quite different: $A(p)$ uses only drugs have been taken by the patient $x$, while BoosTexter uses all the drugs that are included in the list of ordered ‘weak hypotheses’ generated (that is, the drugs that have not been taken by the patient $x$, are used for the calculation of the vector $B(x)$). Moreover, $A(p)$ defines drug–reaction relationships for each drug, while BoosTexter defines drug–reaction relationships for a subset of drugs and some of these drugs included in the list of ‘weak hypotheses’ multiple times with different weights. In other words, for some drugs (say $d$), BoosTexter generates different vectors (say $h^k(d)$, $k = 1, \ldots, K$), depending on the order in the list of ‘weak hypotheses’. This feature makes it impossible to use BoosTexter for solving some other interesting problems like evaluation of suspected drugs and drug–drug interactions.

In the calculations below, we ran BoosTexter with the number of rounds set at 3000. Note that this algorithm defines a weak hypothesis using one drug at each round. In the data considered here, there are 2896 suspected drugs. Therefore, choosing the number of rounds as 3000, gives a possibility of using all suspected drugs.

**Versions** $A(p)$, $A^{GO}(p)$, $p = 0, 1, 2, \ldots$. 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. We used 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 $p^* \geq 30$. We set $p^* = 80$ in the calculations.

On the other hand, for each $A(p)$, we will consider two versions.

1. In the first version, which will be denoted by $A(p)$, we use the weights $h^0(d)$, $d \in \mathcal{D}$, in (6.1), which are calculated by (23) as a good initial point. In this case we do not use the second and third steps in the solving of problem (9), (10).

2. In the second version, which will be denoted by $A^{GO}(p)$, the weights $h(d)$, $d \in \mathcal{D}$, will be calculated as a solution, obtained by algorithm AGOP, to the global optimization problem (13), (14); that is, we use $h^{GO}(d)$, $d \in \mathcal{D}$, presented in (18).
The use of these two versions allows us to make very important observations. For example, we observe that the ‘responsibility’ of some drugs, in terms of causing reactions, has essentially changed by years. Another observation is that, the decrease in the value of the average distance measure $E_{av}^P$ [see equation (9)], in fact leads to an increase in average precision $P_{av}$ (section 4) which is the main evaluation measure used for multi-label classification problems. This emphasizes the effectiveness of using the distance measure (7).

9.2 New drugs and 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 new drugs $d$, we set $h_i(d) = 0$, $i = 1, \ldots, c$. It is possible that for some new (test) example all suspected drugs 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 predictions for such examples does not make sense. Therefore, in the calculations below, we will remove all new events from test sets.

9.3 Training and test sets

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 table 1 we summarize the number of records in the test and training sets, and, also, the number of new events removed. In the second part of this table we present the number of records in the training and test sets having at least two drugs that have been used.

9.4 Predictions of reactions

The results obtained by the algorithms BoosTexter and $A(p)$ in [1] have shown that, in terms of prediction of reactions, the use of only suspected calculating drugs, for calculating drug–reaction relationships and then prediction of reactions, provided better results than if we use all drugs (suspected and non-suspected). Therefore, in the calculations below we will use only suspected drugs for calculating weight vectors $h(d), d \in D$. The results obtained are presented in table 2 (by BoosTexter) and in tables 3, 4 (by $A(p)$ and $A^{GO}(p), p = 0, 1, 2$).

The algorithms BoosTexter and $A(p)$ define potential reactions in quite different ways. There are some important points that make using the algorithm $A(p)$ preferable for the study of drug–reaction associations.

<table>
<thead>
<tr>
<th>Test year</th>
<th>Number of records</th>
<th>Records with ≥2 drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td>1996</td>
<td>12600</td>
<td>1049</td>
</tr>
<tr>
<td>1997</td>
<td>13747</td>
<td>1091</td>
</tr>
<tr>
<td>1998</td>
<td>15001</td>
<td>1418</td>
</tr>
<tr>
<td>1999</td>
<td>16684</td>
<td>1746</td>
</tr>
<tr>
<td>2000</td>
<td>18599</td>
<td>1988</td>
</tr>
<tr>
<td>2001</td>
<td>20749</td>
<td>1054</td>
</tr>
</tbody>
</table>
The study of drug–reaction relationships

Table 2. The results obtained by the algorithm BoosTexter2_1 [18] for average precision \((P_{av})\) by using suspected drugs. The algorithm BoosTexter2_1 [18] was set to run 3000 training rounds. Average precision is presented in percent.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BoosTexter</td>
<td>(P_{av})</td>
<td>84.15</td>
<td>80.42</td>
<td>84.17</td>
<td>80.66</td>
<td>84.10</td>
<td>83.90</td>
</tr>
</tbody>
</table>

Table 3. The results obtained by the algorithms \(A(p)\) and \(AGO(p)\), \((p = 0, 1, 2)\), for average distance measure \((E_{av}^0)\) and average precision \((P_{av})\) by using suspected drugs: test years 1996, 1997, 1998. Average precision is presented in percent.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Measures</th>
<th>Test year 1996</th>
<th>Test year 1997</th>
<th>Test year 1998</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A(0))</td>
<td>(E_{av}^0)</td>
<td>0.587</td>
<td>0.810</td>
<td>0.590</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>82.91</td>
<td>80.25</td>
<td>82.80</td>
</tr>
<tr>
<td>(AGO(0))</td>
<td>(E_{av}^0)</td>
<td>0.566</td>
<td>0.838</td>
<td>0.570</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>83.54</td>
<td>79.82</td>
<td>83.47</td>
</tr>
<tr>
<td>(A(1))</td>
<td>(E_{av}^1)</td>
<td>0.369</td>
<td>0.472</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>83.96</td>
<td>80.07</td>
<td>83.97</td>
</tr>
<tr>
<td>(AGO(1))</td>
<td>(E_{av}^1)</td>
<td>0.356</td>
<td>0.488</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>84.78</td>
<td>80.25</td>
<td>84.76</td>
</tr>
<tr>
<td>(A(2))</td>
<td>(E_{av}^2)</td>
<td>0.258</td>
<td>0.314</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>84.51</td>
<td>79.61</td>
<td>84.51</td>
</tr>
<tr>
<td>(AGO(2))</td>
<td>(E_{av}^2)</td>
<td>0.248</td>
<td>0.321</td>
<td>0.248</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>85.31</td>
<td>79.73</td>
<td>85.29</td>
</tr>
</tbody>
</table>

First we note that, the algorithm \(A(p)\) calculates weights for each drug, which is very important because in this case we establish drug–reaction relations for all drugs. BoosTexter does not calculate weights for each drug. Moreover, BoosTexter classifies examples so that drugs that are not used are still assigned weights in the function \(B(x)\). In the other words,

Table 4. The results obtained by the algorithms \(A(p)\) and \(AGO(p)\), \((p = 0, 1, 2)\), for average distance measure \((E_{av}^0)\) and average precision \((P_{av})\) by using suspected drugs: test years 1999, 2000, 2001. Average precision is presented in percent.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Measures</th>
<th>Test year 1999</th>
<th>Test year 2000</th>
<th>Test year 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A(0))</td>
<td>(E_{av}^0)</td>
<td>0.594</td>
<td>0.743</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>82.72</td>
<td>80.91</td>
<td>82.72</td>
</tr>
<tr>
<td>(AGO(0))</td>
<td>(E_{av}^0)</td>
<td>0.576</td>
<td>0.753</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>83.38</td>
<td>81.01</td>
<td>83.34</td>
</tr>
<tr>
<td>(A(1))</td>
<td>(E_{av}^1)</td>
<td>0.374</td>
<td>0.468</td>
<td>0.376</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>83.79</td>
<td>80.94</td>
<td>83.79</td>
</tr>
<tr>
<td>(AGO(1))</td>
<td>(E_{av}^1)</td>
<td>0.363</td>
<td>0.474</td>
<td>0.366</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>84.57</td>
<td>80.88</td>
<td>84.51</td>
</tr>
<tr>
<td>(A(2))</td>
<td>(E_{av}^2)</td>
<td>0.262</td>
<td>0.331</td>
<td>0.264</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>84.37</td>
<td>80.89</td>
<td>84.35</td>
</tr>
<tr>
<td>(AGO(2))</td>
<td>(E_{av}^2)</td>
<td>0.254</td>
<td>0.337</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>(P_{av})</td>
<td>85.05</td>
<td>80.64</td>
<td>85.06</td>
</tr>
</tbody>
</table>
reactions are predicted not only by drugs actually used, but also, drugs which were not taken. But anyhow, the application of the algorithm BoosTexter is very useful for having some idea about the possible ‘maximal’ accuracy that could be achieved in reaction predictions.

The results presented in tables 2, 3 and 4 show that the algorithms $A(p)$, provide approximately the same accuracy for reaction predictions. It is interesting to compare the results obtained by $A(p)$ and $A^{GO}(p)$. We have the following important observations.

1. As the algorithm $A^{GO}(p)$ uses a global optimization method (the second and third steps) to solve problem (13), (14), the average distance $E_{av}^p$ for training sets is always better (that is, less) than for $A(p)$. An important fact here is that, the decrease in average distance leads to an improvement (that is, increase) of the other evaluation measure – average precision $P_{av}$ (see results for training sets in tables 3 and 4). This emphasizes the efficiency of using the distance measure (7).

2. Another important observation is related to the results obtained for test sets. Comparing the results obtained for $A(p)$ and $A^{GO}(p)$, we see that the decrease in $E_{av}^p$ for training sets (which is good) results in an increase in $E_{av}^p$ for test sets (which is not good), while $P_{av}$ may increase or decrease. How can we explain this situation?

First of all we note that the evaluation measure average distance is much more stable than average precision, which depends on the order of weights (a small change in the weights, that are close, can change the order of the weights). Therefore, the increase in $E_{av}^p$ for test sets (while for training sets it becomes better) indicates some changes in the definition of importance of the drugs (described by the optimal values $\lambda(d)$, see [equation (12)]) in terms of causing adverse reactions. For example, a drug (say $d$) that was important in the previous period (that is, $\lambda(d)$ was a large number) may not be important for the next period (that is, the number $\lambda(d)$ becomes smaller). This situation is mainly related to the newly introduced drugs. As example, in table 5 we present the optimal values $\lambda(d)$ of 5 drugs for which the first adverse reaction reported in 1997 (as suspected drugs).

From table 5 we observe that, for all drugs in this table, the $\lambda$ values increase until the period 1972–1999 and then decrease. Such behaviour is common for most drugs after they came to market. This means that most newly introduced drugs have not been correctly reported as suspected drugs in many cases. This can be explained as follows: when a new drug is used together with some other drugs then this drug was mainly reported as ‘suspected’, although, it is quite possible that, the other drugs were ‘more responsible’ (we note that, the value $\lambda$ could be changed due to the records having more than one drug taken). This observation shows a clear need for more detailed analysis of the responsibility of newly introduced drugs in observed reactions.

### 9.5 Evaluation of correctness of suspected drugs reported

One of the advantages of algorithm $A(p)$ includes the determination of weights for each drug, and, then the classification of reactions, observed for each patient, on the basis of drugs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>455</td>
<td>0.5125</td>
<td>0.7129</td>
<td>0.7116</td>
<td>0.4842</td>
</tr>
<tr>
<td>905</td>
<td>0.9062</td>
<td>1.0014</td>
<td>1.3078</td>
<td>1.0398</td>
</tr>
<tr>
<td>2018</td>
<td>1.1732</td>
<td>1.3837</td>
<td>1.6054</td>
<td>1.1217</td>
</tr>
<tr>
<td>3853</td>
<td>0.7755</td>
<td>1.0314</td>
<td>1.2478</td>
<td>1.0477</td>
</tr>
<tr>
<td>3045</td>
<td>0.5863</td>
<td>0.6670</td>
<td>0.9612</td>
<td>0.5910</td>
</tr>
</tbody>
</table>
The study of drug–reaction relationships

123

actually used by this patient. This advantage allows us to use the algorithm \( A(p) \) to study the
identification of suspected drugs and of drug–drug interactions.

In this section we will evaluate the correctness of suspected drugs reported. The methodology is described in section 7. As mentioned above, BoosTexter cannot be used for this task, so only algorithm \( A(p) \) is used.

The case when a patient uses only one drug, is not interesting to consider, because in this case there is no doubt that the drug used should be a suspected drug. That is why, we consider records having two or more drugs that have been taken. The number of patients in the training and test sets are presented in table 1.

The results obtained are presented in table 6. We see that, the suspected drugs reported in the ADRAC data are determined with sufficiently high accuracy. For instance, the accuracy 78.0 means that, in the optimal combination of suspected and non-suspected drugs which provides the closest approximation to the observed reactions, the suspected drugs are used with weight 0.78 (non-suspected: 0.22). This could be considered as a high degree of ‘responsibility’.

We also see that using global optimization methods makes this prediction more accurate. The results obtained by algorithms \( A^{GO}(p) \) are better than the results obtained by \( A(p) \), \( p = 1, 2, 3 \).

Note that, the correct identification of suspected drugs in each new report has very important practical applications. The method described here provides us with an alternative method to the method based on expert opinions that is mainly using in practice.

9.6 Interaction of drugs

As mentioned before the study of interactions of drugs is an interesting problem. We consider here the possibility of using vectors of weights in drug–drug interactions. In other words, we aim to evaluate the closeness of potential reactions (calculated by a vector of weights) to the observed reactions in drug–drug interaction cases. For the evaluation of closeness we will use two measures: average responsibility \( P_{int} \) and average precision \( P_{av} \).

For our analysis we consider the records having more than three drugs, where some of drugs were reported as an interaction (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). Of course, 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 7 and 8. Training sets are used for calculating weights for each drug. As in the previous section, the weights are calculated by using

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Train</td>
<td>Test</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>( A(0) )</td>
<td>78.1</td>
<td>72.0</td>
<td>78.0</td>
<td>72.2</td>
<td>77.7</td>
<td>71.3</td>
</tr>
<tr>
<td>( A^{GO}(0) )</td>
<td>79.4</td>
<td>72.5</td>
<td>79.4</td>
<td>72.8</td>
<td>79.1</td>
<td>71.7</td>
</tr>
<tr>
<td>( A(1) )</td>
<td>78.7</td>
<td>71.9</td>
<td>78.6</td>
<td>71.3</td>
<td>78.2</td>
<td>71.8</td>
</tr>
<tr>
<td>( A^{GO}(1) )</td>
<td>79.8</td>
<td>73.5</td>
<td>79.8</td>
<td>73.0</td>
<td>79.4</td>
<td>72.2</td>
</tr>
<tr>
<td>( A(2) )</td>
<td>78.4</td>
<td>72.6</td>
<td>78.2</td>
<td>71.3</td>
<td>77.8</td>
<td>71.0</td>
</tr>
<tr>
<td>( A^{GO}(2) )</td>
<td>79.5</td>
<td>74.5</td>
<td>79.4</td>
<td>71.9</td>
<td>78.9</td>
<td>73.3</td>
</tr>
</tbody>
</table>
suspected drugs. Then the evaluation of interaction of drugs is done only using the test sets, because, in the training sets, the interaction of drugs (as suspected) 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 7 and 8, we present the average results obtained by all test sets which consists of 55 cases. To have some idea about the accuracy achieved, let us consider the results obtained by the algorithm $A^{GO}(2)$. We have $P_{int} = 69.5$ and $P_{av} = 81.2$. The first number means that, in the observed reactions, the ‘degree of responsibility’ of the drugs under interaction versus the all other drugs, is approximately 70%. The second number indicates high accuracy in the prediction of these reactions. 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 $x$. Different versions $A(p)$ provide different values $\mu^*(x)$ and $P(x)$. We take the corresponding maximal values obtained by different versions, and then calculate the average responsibility and precision. The results obtained are presented in the columns ‘max’ in tables 7 and 8. These results are much better than the results obtained by a particular version.

Therefore, we could achieve higher accuracy in the prediction if we could calculate weights $h(d)$ more ‘correctly’. The comparison of results presented in tables 7 and 8 that the application of global optimization techniques provides better drug–reaction relationships described by vectors $h(d)$.

<table>
<thead>
<tr>
<th>Test year</th>
<th>$N$</th>
<th>$P_{int}$</th>
<th>$A(0)$</th>
<th>$A(1)$</th>
<th>$A(2)$</th>
<th>max</th>
<th>$A^{GO}(0)$</th>
<th>$A^{GO}(1)$</th>
<th>$A^{GO}(2)$</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>4</td>
<td>71.8</td>
<td>83.3</td>
<td>78.2</td>
<td>90.6</td>
<td>63.1</td>
<td>80.6</td>
<td>92.4</td>
<td>95.7</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>16</td>
<td>69.0</td>
<td>72.3</td>
<td>74.0</td>
<td>79.7</td>
<td>77.6</td>
<td>77.7</td>
<td>67.9</td>
<td>80.8</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>15</td>
<td>70.0</td>
<td>69.4</td>
<td>62.8</td>
<td>70.1</td>
<td>71.3</td>
<td>60.9</td>
<td>69.5</td>
<td>78.4</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>3</td>
<td>38.9</td>
<td>50.7</td>
<td>76.0</td>
<td>76.0</td>
<td>50.8</td>
<td>56.4</td>
<td>91.2</td>
<td>92.7</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>14</td>
<td>77.1</td>
<td>73.1</td>
<td>63.0</td>
<td>77.1</td>
<td>64.2</td>
<td>58.4</td>
<td>67.9</td>
<td>79.8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>67.9</td>
<td>69.2</td>
<td>66.4</td>
<td>74.5</td>
<td>67.6</td>
<td>64.9</td>
<td>69.5</td>
<td>79.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Evaluation of drug–drug interactions by using a vector of weights obtained by algorithms $A(p)$ and $A^{GO}(p)$, $(p = 0, 1, 2)$, for $P_{int}$. $N$ is the number of records.

<table>
<thead>
<tr>
<th>Test year</th>
<th>$N$</th>
<th>$P_{av}$</th>
<th>$A(0)$</th>
<th>$A(1)$</th>
<th>$A(2)$</th>
<th>max</th>
<th>$A^{GO}(0)$</th>
<th>$A^{GO}(1)$</th>
<th>$A^{GO}(2)$</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>4</td>
<td>89.6</td>
<td>89.6</td>
<td>89.6</td>
<td>89.6</td>
<td>89.6</td>
<td>87.5</td>
<td>89.6</td>
<td>89.6</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>16</td>
<td>71.9</td>
<td>80.2</td>
<td>84.4</td>
<td>84.4</td>
<td>79.2</td>
<td>85.9</td>
<td>84.4</td>
<td>87.5</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>15</td>
<td>61.1</td>
<td>64.5</td>
<td>70.6</td>
<td>71.1</td>
<td>61.1</td>
<td>67.8</td>
<td>70.0</td>
<td>70.6</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>3</td>
<td>55.6</td>
<td>72.2</td>
<td>69.4</td>
<td>72.2</td>
<td>55.6</td>
<td>72.2</td>
<td>61.1</td>
<td>72.2</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>3</td>
<td>77.8</td>
<td>77.8</td>
<td>100.0</td>
<td>100.0</td>
<td>77.8</td>
<td>77.8</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>14</td>
<td>85.6</td>
<td>83.8</td>
<td>89.1</td>
<td>92.7</td>
<td>87.7</td>
<td>86.5</td>
<td>87.5</td>
<td>91.3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>73.2</td>
<td>76.9</td>
<td>82.2</td>
<td>83.5</td>
<td>75.8</td>
<td>80.1</td>
<td>81.2</td>
<td>83.9</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Evaluation of drug–drug interactions by using a vector of weights obtained by algorithms $A(p)$ and $A^{GO}(p)$, $(p = 0, 1, 2)$, for $P_{av}$. $N$ is the number of records.
10. Conclusion

In this paper we have used a new optimization approach to study multi-label classification problems. In particular we have focussed on drug–reaction relations in the domain of the cardiovascular group of reactions from the ADRAC data. The suggested method of representation for drug-reaction relations in the form of a vector of weights is examined in the prediction of reactions. The suggested method was applied for the evaluation of correctness of suspected drugs and for the study of drug–drug interactions. 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.

In this paper we presented the results obtained for the cardiovascular group of reactions. Similar calculations have been carried out for other groups of reactions defined in the ADRAC data.

This approach for determining which drugs are responsible for reactions has been implemented in software. This software can be used to solve the following two problems:

a. given a set of drugs to predict the reactions that are most likely to occur; and
b. 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


