journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A similarity measure for case based reasoning modeling with temporal abstraction based on cross-correlation

Florian Hartge[a,*], *Thomas Wetter*[a], *Walter E. Haefeli*[b]

[a] *Institute for Medical Biometry and Informatics, Department Medical Informatics, Im Neuenheimer Feld 400, D-69120 Heidelberg, Germany*
[b] *Department Internal Medicine VI, Clinical Pharmacology and Pharmacoepidemiology, University of Heidelberg, Im Neuenheimer Feld 410, D-69120 Heidelberg, Germany*

ARTICLE INFO

Article history:
Received 9 November 2004
Received in revised form 19 October 2005
Accepted 20 October 2005

Keywords:
Case based reasoning
Temporal abstraction
Similarity measure
Adverse drug events

ABSTRACT

Adverse drug events (ADEs) are a major limitation of drug safety. They are often caused by inappropriate selection of dose and the concurrent use of drugs modulating each other (drug interaction). Risk assessment and prevention strategies must therefore consider co-administered drugs, individual doses, and their timing. In a new approach we evaluated the performance of cross correlation, commonly used in signal processing, to determine similarities in patient treatments. To achieve this, patient treatments were modeled as groups of vectors representing discrete time intervals. These vectors were cross-correlated and the results evaluated to find clusters in time courses indicating similarity in treatment of different patients.

To evaluate our algorithm, we then created a number of test cases. The focus of this article is on each treatment, and its pattern in time and dosage. The algorithm successfully produces a relatively low similarity score for cases that are completely different with respect to their pattern of time and dosage but high scores when they are equal (score of 0.699) or similar (score of 0.528) in their therapies, and thus succeeds in having a relatively high specificity (27/30). Such an approach might help to considerably reduce the problem of false alarms which hampers most existing alerting systems for medication errors or impending ADEs.

1. Introduction

Adverse drug events (ADEs) are a major factor limiting safety and effectiveness of drug therapy in modern health care systems. For example, in the USA, an estimated 770,000 people are subject to an ADE during their stay in a hospital and about 140,000 of them die because of it [1]. The estimated annual costs are about $4 billion, of which about half may be avoidable [2].

Although not studied thus far, there is no evidence to suggest that the problem is much different or substantially smaller in Germany. However, there have been some efforts in the last years to overcome this drawback [3–5]. Computerized physician order entry systems (CPOE), barcode identification, and automated dosing systems have been introduced to reduce the errors inherent to the process of administering medication (from the prescription process to dispensing) like misspelling, misreading, and confusion of medication or patient data. Another effective option is to support the physician during the prescription process with pertinent information and decision support on medical therapy [6]. This is the focus of this paper.

* *Corresponding author*. Tel.: +49 6221 56 7461; fax: +49 6221 56 4997.
E-mail address: florian.hartge@med.uni-heidelberg.de (F. Hartge).
0169-2607/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2005.10.005

It is a common approach to warn the practitioner while he is entering the intended medication for a patient if there is an assumed adverse interaction with another drug or with the patient's state of health. This alert process is normally aided by the use of databases containing lists of interactions or other known risk situations [7]. Standard practice entails the computer to issue warnings to the user if the patient has drug X or condition Y (e.g. allergy) in the presence of which drug Z is known to likely cause an ADE. However, the specificity of such alert systems is limited because potentially harmful drug combinations do not result in an ADE in all instances. This leads conventional warning systems to issue many false alarms. These warning systems have a rather good sensitivity but their poor specificity prompts practitioners to ignore them [8,9] or to override alerts [10,11]. The aim of this ongoing research is to develop an electronic drug prescription system, which more selectively identifies risks associated with drug therapy, with more specific alerts, and therefore better acceptance than the current systems. The proposed algorithm may be a key element in such a system since it will directly influence its user-friendliness.

2. Background

In most medical specialties, the occurrence of ADEs is a complex process and the circumstances leading to a particular event are not always known in detail. While it is often possible to retrospectively identify an ADE, it is much more difficult to point out exactly what leads to it, especially considering individual variability of important pharmacologic processes between different patients [12,13]. In such a relatively uncertain domain it is challenging to form sets of clear rules that identify ADEs. Therefore, case similarity suggests itself as an approach that may circumvent the problems of characterizing ADEs through sets of rules.

2.1. *Introduction to case based reasoning (CBR)*

CBR is a sub-discipline of artificial intelligence that is based on the assumption of analogy. The main theorem of CBR is that if two problems are similar, their solutions also are. The mechanics are therefore that a pool of problems and their respective solutions within a certain domain is collected. If then a new problem arises, the pool is searched for the most similar problem. The solution of this already solved problem is then applied to the new problem at hand. The solution of the old problem may have to be slightly modified to the new needs (Fig. 1).

CBR has some intriguing advantages. First it works well in domains where the knowledge is relatively rudimentary. Second, in order to build a system it is not necessary to query experts about their way of reasoning. Rather records of old cases such as progress notes or standardized charts can be used. Third, it is easy to add new knowledge in the form of new cases or to adapt existing ones and, therefore, fourth, it is simple to manage and maintain an up-to-date base of knowledge [14].

In our system we want to collect a reference base of detailed medical cases with verified ADEs. If a new case arises in practice with striking similarity to a case in the database, then the system should alert the practitioner. One problem with this approach is the complexity of modeling time elapsed and temporal interrelationships between exposure with different drugs and concurrently developing events. Most drug effects do not only depend on pharmacodynamics, but are time dependent and often related to pharmacokinetics. Therefore, the temporal relationships between drug exposure of a patient and events are essential to assume similarity in different cases.

An earlier approach using CBR and temporal abstraction [15] made it possible to cover the feature space of a problem with a fixed set of predictable states this problem can reach. Therefore, a case modeling of problems was developed which consists of a distribution of different states over time. This is only possible if a predefined set of attributes with conceivable value ranges can directly be mapped to the set of medical states. Obviously it is not possible to predefine every possibly occurring medical state during treatment. Therefore we aimed to find a different case modeling and similarity measure.

2.2. *Introduction to cross correlation*

We took a new approach in using cross-correlation to determine similarities in treatments. In signal processing, the cross-correlation function is commonly used to determine similarities between two time-dependent signals. To determine cross-correlation between two signals we first have to define the correlation between data in general. Correlation between two groups of data implies that they move or change with respect to each other in a structured way. The correlation coefficient is an indicator for the strength and sense or direction of correlation. For N pairs of data (x, y) the correlation coefficient is calculated thus

$$r_{xy} = \frac{1/(N-1)\sum_{n=1}^{N}(x(n)-\bar{x})(y(n)-\bar{y})}{\left(\left(1/(N-1)\sum_{n=1}^{N}(x(n)-\bar{x})^2\right)\left(1/(N-1)\sum_{n=1}^{N}(y(n)-\bar{y})^2\right)\right)^{1/2}} \quad (1)$$

The 'active part' of Eq. (1) is the enumerator summation which tends to zero if there is little common movement between x and y, and approaches high positive and negative values depending on whether x and y tend to move together or in opposite senses. The denominator terms merely have a normalizing effect which delimits the range of the correlation coefficient to $[-1,1]$ [16].

The correlation coefficient can be used to determine correlation between two sets of data. When measuring the correlation between two signals from two time series of discrete numbers, it is possible that two signals may have common components but different timing. The cross-correlation function (Eq. (2)) is that function which is formed from successive values of the correlation coefficient taken at time shifts $k = 1$,