Using association and overlapping time window approach to detect drug reaction signals

S. Ivkovic
Email: s.ivkovic@ballarat.edu.au

G. Saunders
Email: g.saunders@ballarat.edu.au

R. Ghosh
Email: r.ghosh@ballarat.edu.au

J. Yearwood
Email: j.yearwood@ballarat.edu.au

School of Information Technology and Mathematical Sciences
University of Ballarat, Victoria 3353 Australia

Abstract

The problem with detecting adverse drug reactions (ADRs) from drugs is that they may not be obvious until long after they are widely prescribed. Part of the problem is these events are rare. This work describes an approach to signal detection of ADRs based on association rules (AR) in Australian drug safety data. This work was carried out using the Australian Adverse Drug Reactions Advisory Committee (ADRAC) database, which contains a hundred and thirty seven thousand records collected in 1972-2001 period.

Many signal detection methods have been developed for drug safety data, most of which use a classical statistical approach. Some of these stratify the data using an ontology for reactions, but the application of drug ontologies to ADR signal detection methods has not been reported.

We propose a novel approach for detecting various signal levels by using an overlapped windowing approach. The overlapping windows help to detect smooth transition of signal. We use association rules for measuring significant change over time for different hierarchical levels of drugs (using the Anatomical-Therapeutic-Chemical (ATC) system of drug classification ontology) and their reactions based on the System Organ Classes (SOC) ontology. Using association rules and their strength for different levels in the drug and reaction hierarchy, helps in the detection of signals at particular levels in higher order using a bottom up approach.

The results of a preliminary investigation of ADRAC data using our method demonstrate that this approach could produce a powerful and robust ADR signal detection method.

I. INTRODUCTION

Post-marketing safety databases are large, sparse, full of biased, uncontrolled and incomplete data [1]. The detection of unknown and unexpected connections between drug exposure and adverse events (AEs) is one of the major challenges of pharmacovigilance. For the identification of possible connections in large databases, automated statistical systems have been introduced with promising results. From the large numbers of associations so produced, the human mind has to identify signals that are likely to be important, in need of further assessment and follow-up and that may require regulatory action [2]. 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) [3], Reporting Odds Ratio (ROR), and Yule's Q [4]. Recently, regulatory authorities, drug monitoring centers, and pharmaceutical companies, have directed more attention to the development, implementation , and deployment of computer-assisted signal detection systems.

The three most commonly used methods with the greatest published experience are: proportional reporting ratios (PRRs) [5], [6], an application of empiric Baysean screening (EBS) known as the mulit-item gamma Poisson shrinker (MGPS) [6], [7], [8], and the Bayesian confidence propagation neural network (BCPNN) [6], [9].

The Australian Adverse Drug Reaction Advisory Committee (ADRAC) database has been developed and maintained by the Therapeutic Goods Administration (TGA) with the aim to detect signals from adverse drug reactions (ADRs) as early as possible. The ADRAC data contain 137,297 voluntarily reported adverse drug reaction records involving 5057 different drugs, based on the 'drug dictionary' used by ADRAC of 7416 available drug terms, and 868 different reactions, based on 1392 available reaction terms. There are many fields in the ADRAC database, including report date and fields containing drug and reaction terms. There are two groups of ten fields, which are allocated for drug terms and reaction terms. A more detailed account of the ADRAC database is given in [10].

A. Data Consolidation

The biggest challenge in summarizing safety data is the need to consolidate the massive amount of data into a manageable format. One way is to group the safety data into K classes characterized by body systems and determined in conjunction with underlying disease and treatments involved. Such pooling of data through coding is especially helpful for rare events [11], [12]. The World Health Organization (WHO) maintains such a classification of reaction terms [13]. These are groups of adverse reaction preferred terms pertaining to the same systems organ class (SOC). The Australian drug safety data maintained by ADRAC uses WHO ADR terms and SOC grouping; see [10], [14], [15]. For more details of SOC information in ADRAC see [16].

There has been a need for a similar kind of grouping to apply to drugs in the ADRAC data, which are currently listed by trade name [15]. In these data, some of these trade names were assigned anatomical-therapeutic-chemical classification (ATC) codes by ADRAC, but in cases where the trade name had more than one ingredient, ATC codes were not assigned. This paper applies an *ad hoc* classification made by us [16] where each trade name code was assigned a corresponding

ATC code. The classification system implemented was the WHO Collaborating Centre for Drug Statistics Methodology ATC codes in order to enable: (i) the grouping of the same substance into one code, and (ii) the grouping of related substances into more general categories. To this purpose the ATC codes for drugs have been purchased by us from Ms Kirsten Myhr, RELIS Øst, Ullevál University Hospital, 0407 OSLO (http://www.whocc.no/).

B. ATC embedded code

The ATC encoding system implements an an embedded encoding system, which employs a seven character coding system. As an example, the first member of the system will be used. The level we call level 0, the character is 'A', the level 1 is 'A01', level 2 is 'A01A', level 3 is 'A01AA' and level 4 is 'A01AA01'. The embedded coding can be utilized to simplify database queries and algorithm coding. Employing this embedded coding, which resides in the ATC code string, can be now utilized to aggregate the data at the required level.

C. Classification of ADRAC data for this analysis

For the purposes of this present investigation, in cases where the original drug code had more than one ingredient, a unique code was assigned from ATC codes for combinations of ingredients. This was done in order to preserve a one-to-one correspondence between original drug codes and the ATC codes facilitating comparison of the two drug classification schema. This resulted in 1806 ATC drug terms from the 5081 drug trade name terms. There are five levels of classes in the ATC system: level 0 had 14 classes, level 1 had 95, level 2 had 217, level 3 had 581 and level 4 had 1785.

D. Critical Terms

From the third quarter of 1998, a new field has been added at the end of the WHO-ART file which indicates Critical Terms. Critical terms are a subset of adverse reaction terms referring to, or possibly being indicative of, serious disease states, which have been regarded as particularly important to follow up. [13], [17] This is a very important component of expert knowledge, which can be incorporated into the ADR signal detection. In this report we weight critical reaction terms to highlight associations which are likely to be of greater interest and particularly those that are rare.

E. Motivation

Automated signal detection and data mining techniques are evolving rapidly, but several outstanding issues remain problematic. These include the absence of a gold standard against which to evaluate signal detection methods, lack of validation and comprehensive head-to-head comparisons of the various methods, lack of transparency of the Bayesian-based methods, lack of best practices or recommended data mining protocols, the influence of AE dictionary structure on performance characteristics, limited specificity, and potential for misapplication of these techniques [18]. In one study there were 136 relevant DECs that triggered safety-related labelling changes for 39 drugs during a 7-month period. PRRs generated a signal of disproportionate reporting with almost twice as many drug-event combinations (DECs) as MGPS (77 *versus* 40). DECs generating signals of disproportionate reporting with only PRRs were both medically serious and non-serious. No medically important events were signalled by MGPS only [19]. Other difficulties with MGPS are "signal leakage", in which a signal truly associated with drug A may manifest as an apparent signal with drug B due to the fact that the two drugs are frequently co-prescribed, and "signal masking", whereby a signal for drug A can be diminished in magnitude if drug B has a very dominant signal in the database [20]. Some of the logic (for example triage logic) for signal selection and follow-up are discussed in relation to UMC data by Meyboom *et al.* [2] and Ståhl *et al.* [21], [22]. Therefore there is still more work to be done in the ADR signalling methodology.

II. METHOD AND RESULTS

A. Aim

The aim of this paper is to present the developments in our approach to developing an ADR signalling method. We use the SOC reaction term and ATC drug term hierarchies to establish drug-reaction associations, by traversal of the ATC tree where reaction frequencies are significant (based on Chi-Square – χ^2). The algorithm then surveys the temporal reporting of these reactions using a varying time window, which smooths monthly variation and seeks to find frequent associations (also using the χ^2 test) of a critical nature. The algorithm then generates a report on the signal status of this reaction set.

B. Research Methodology

1) The association between drug/class and reactions: Among many different techniques used to extract useful knowledge from databases, Association Rule mining, has in recent years, attracted the attention of data mining communities [23]. Association rule mining is a form of data mining used to discover interesting relationships amongst two or more attributes in data. Association rules were introduced by Agrawal et al. [23] and originated with the problem of supermarket basket analysis. In basket analysis association rules were used to find associations between the items bought by a customer in order to find which items were frequently bought together. The findings can be used to understand customers' buying habits and preferences in order to increase profits.

An association rule is an expression of the form $X \Rightarrow Y$ *[support, confidence]*, where X and Y are sets of items that are often found together in a given collection of data. The attribute group on the left hand side of the arrow is called the antecedent or "left hand side" (LHS) and the group of attributes on the right hand side of the arrow is called the consequent or "right hand side" (RHS). While association rules have proven to be useful in practical applications, AR algorithms tend to generate large numbers of rules, most of which are of little interest. In order to filter uninteresting rules the interesetingness and usefulness of an association rule has been based on user defined support and confidence thresholds.

IEEE COMPUTER SOCIETY

If both support and confidence values are greater than the threshold, the association rule is considered interesting. The support is the percentage of rows in the database containing both *X* and *Y*. In this study we do not use support as a measure of interestingness because higher support will filter out rare associations. In many real life applications some items appear frequently in the data, while others rarely appear. If we use support as a measure of interesetingness those rules that involve rare items will not be found. The confidence is the conditional probability of *Y* given *X*, for example, *confidence* = $P(Y \mid X)$.

Our method seeks to find if there exists any association between any particular drug or drug class and a set of critical reactions using a tree structure. For example an association rule R1 *drug = M01AH01 ⇒ reaction = 636 [confidence 35%]* means that 35% of patients who took drug M01AH01 (Celecoxib) had reaction with code 636 (hypertension). This association can also be represented as conditional probability $P(reaction = 636 \mid drug = M01AH01)$

In this study the traditional AR approach is not directly used. We organize generated association rules by separating the discovered rules into rule sets. Each rule set contains association rules that share a common consequent. For example association rules *drug = M01 ⇒ cardiovascular reactions* and *drug = M02 ⇒ cardiovascular reactions* represent a rule set because they share the same consequent.

We generate the tree using a breadth first search. The root of the tree denotes the top drug class level (for example, M) and children nodes denote a subclass of the parent node (for example, M01, M02 and M03) and leaves denote a particular drug (for example, M01AH01). During traversal of the tree we find if another level of the tree needs to be generated according to our convergence criteria. The details of our algorithm are as follows:

Step 1. We select the top drug class and and a reaction class of interest. For example we select drug class M and reaction class – cardiovascular reactions (SOC1000) .

Step 2. We generate the next level of the drug class as child nodes of the tree. We find the set of children for the drug classes that are associated with the selected reaction class. For example for drug class M we find children M01, M02, M03 and M04 and then for each child we generate group of association rules for class – cardiovascular reactions (SOC1000).

R1 *drug = M01 ⇒ cardiovascular reactions [conf]*
R2 *drug = M02 ⇒ cardiovascular reactions [conf]*
R3 *drug = M03 ⇒ cardiovascular reactions [conf]*
R4 *drug = M04 ⇒ cardiovascular reactions [conf]*
If the discrepancy between the confidence values of these association rules is substantially high, this group is considered different on the basis of cardiovascular reactions, otherwise the group is considered similar. If the group is considered different by using the χ^2 test we find if there is a significant difference in frequencies. We generate a contingency table containing frequencies and calculate the chi-square value If we find
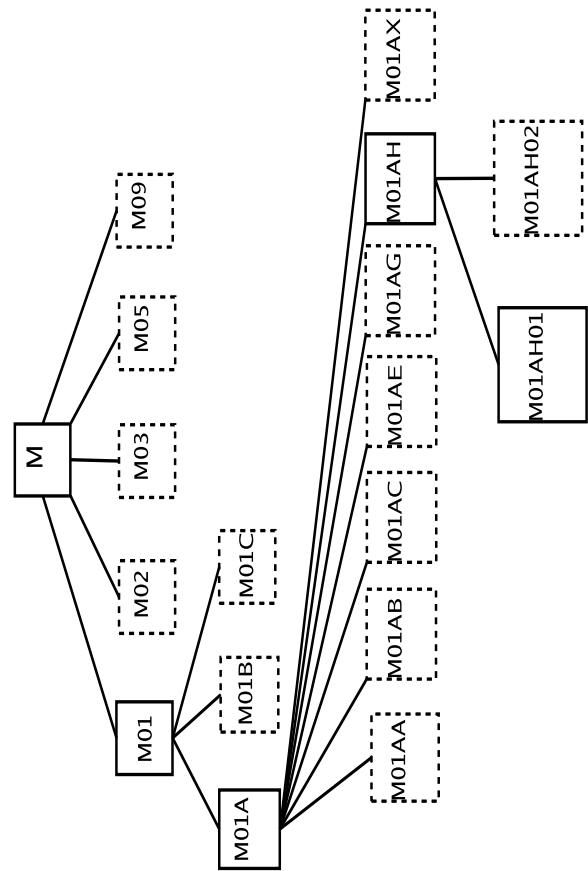

Fig. 1.   Tree traversal of signal algorithm.

significant differences we choose child(ren) with maximum frequency. Figure 1 illustrates breadth search steps.

Figure 1 depicts such a traversal. For example, if drug class M is chosen, the children M01, M02, M03, M04, M05, M09 are examined. M01 is found to be significant, so its children, M01A, M01B, M01C, are examined. M01A is found to be significant, so its children, M01AA, M01AB, M01AC, M01AE, M01AG, M01AH, M011AX are examined. M01AH is found to be significant, so its children, M01AH01, M01AH02 are examined. M01AH01 (see Figure 2, where reaction class SOC1000 was chosen) is found to be significant. The algorithm terminates at this leaf.

Step 3. We continue the above step 2 until:

- the lowest level of drug class is found or,
- χ^2 test fails. This means that this particular drug class level is responsible for the reactions

Once search process terminates we generate association rules for drug(s) and adverse reactions. A drug can be either drug class (for example, M01AH) or drug (for example, M01AH01). Adverse reactions (ADR) are reactions belonging to the same reaction class. For example ADR 67, 68,104 belong to cardiovascular reaction class (SOC1000) as illustrated in Figure 2. Association rules displayed in Figure 2 are used for descriptive purposes. For examplethe highest bar

COMPUTER
SOCIETY

in this figure shows that almost 15% of patients taking drug M01AH01 with cardiovascular reactions had experienced ADR 636 (hypertension). This finding could be represented as an association rule *drug = M01 AND cardiovascular reactions ⇒ hypertension [conf 14%]*

2) Generating signals using overlapping time windowing approach: Once an association between a particular drug and a set of significant reaction classes has been found from above, we plot the frequency of the significant reaction classes over time. A time window length of 3 has been considered with a phase shift of 1 time interval. This overlapping time window with a sift helps to smooth the signals. For each time frame we use χ^2 test to find the signals which indicates sudden increase in number of reactions. Depending on the percentage of critical terms reported from time of signal back to the start time, we label the signal. Six different types of signals – OK (15% or below), notice (between 15% and 30%), strong notice (between 30% and 40%), warning (between 40% and 50%), strong warning (between 50% and 60%), Alert(60% or above). Furthermore we look at the support of frequencies of all reactions from time of start to time of signal.

C. Experimental Results

1) Drugs giving strong signals: **Celebrex and cardiovascular (SOC 1000)**

The output if drug M01AH01 and cardiovascular reactions are selected, the histogram in Figure 2 is displayed and the following output is made:

"Cardiovascular system reactions for drug M01AH01 Celecoxib (total of 282 reactions)

1) Celecoxib (M01AH01) makes 1.952) 22.18% (593 cases) of patients taking Celecoxib had CRITICAL reactions overall 3) 77.82% (2080 cases) of patients taking Celecoxib had NON CRITICAL reactions overall 4) 43.62% (123 cases) of patients taking Celecoxib with Cardiovascular system reactions - CRITICAL and they make 4.60 of all patients taking Celecoxib 5) 56.38% (159 cases) of patients taking Celecoxib with Cardiovascular system reactions - NON CRITICAL and they make 5.95 of all patients taking Celecoxib"

"Cardiovascular system - reaction monthly counts for drug M01AH01 - Celecoxib (total:282 for 22 months)"

Monthly counts are given in Figure 3.

"Null hypothesis: There are no ADR signals in this time span is REJECTED WARNING (46.27% critical and 53.73% non critical reactions (out of 67) before the signal on 200002) (position 4)"

Overlapping window counts are given in Figure 4.

"Null hypothesis: There are no ADR signals in this time span is REJECTED High Frequency WARNING (42.42% critical and 57.58% non critical reactions (out of 231) before the signal on 200011)"

Heparin and haemic and lymphatic system (SOC 1200)

The output if drug B01AB01 and haemic and lymphatic system reactions are selected, the histogram in Figure 5 is displayed and the following output is made:
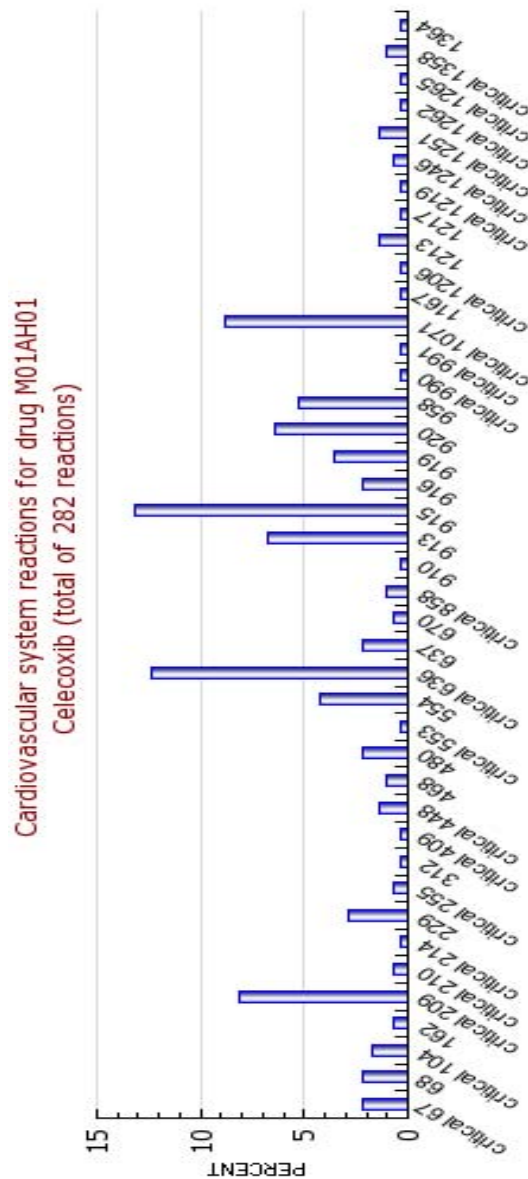

Fig. 2.   Celecoxib (M01AH01) cardiovascular reactions (SOC1000).

"Haemic and lymphatic systems reactions for drug B01AB01 Heparin (total of 220 reactions)

Heparin (B01AB01) makes 0.53% (cases) of all records (out of 137279) 2) 37.15% (269 cases) of patients taking Heparin had CRITICAL reactions overall 3) 62.85% (455 cases) of patients taking Heparin had NON CRITICAL reactions overall 4) 87.27% (192 cases) of patients taking Heparin with Haemic and lymphatic systems reactions - CRITICAL and they make 26.52 of all patients taking Heparin 5) 12.73% (28 cases) of patients taking Heparin with Haemic and lymphatic systems reactions - NON CRITICAL and they make 3.87 of all patients taking Heparin"
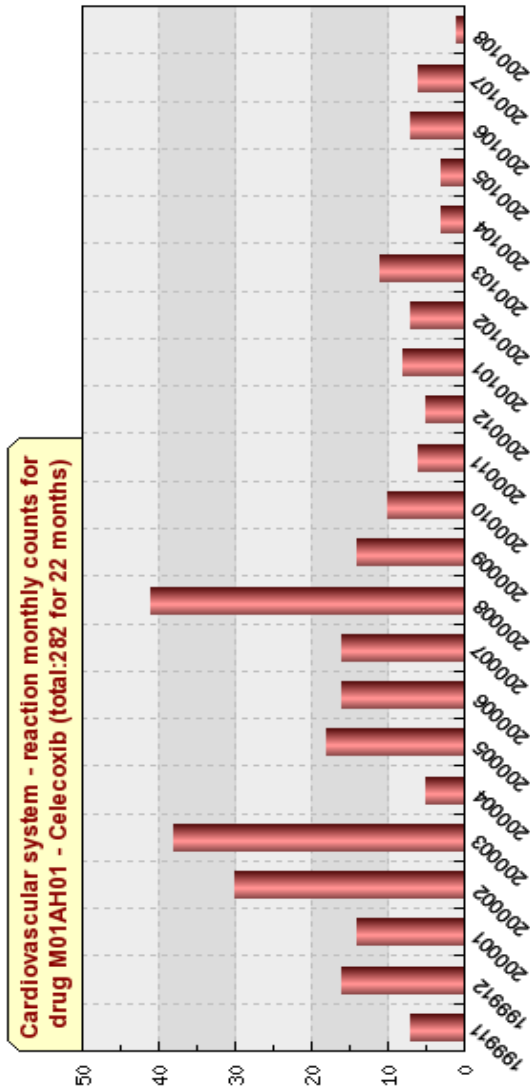
Fig. 3. Celecoxib (M01AH01) cardiovascular reactions (SOC1000) – monthly reactions.
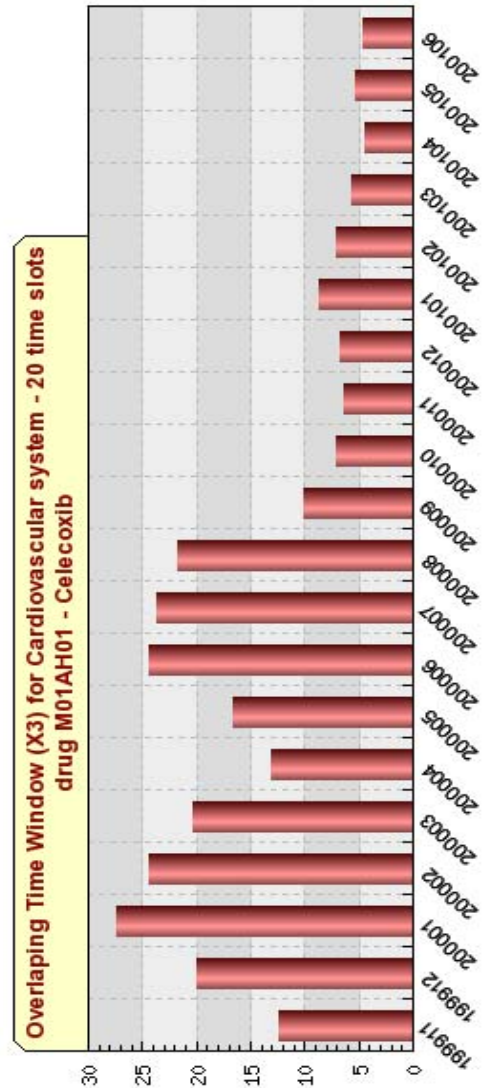

Fig. 4. Celecoxib (M01AH01) cardiovascular reactions (SOC1000) – overlapping window reactions.

The most frequent reaction, THROMBOCYTOPENIA (1243) 160 cases severity 10.

"Haemic and lymphatic systems - reaction monthly counts for drug B01AB01 - Heparin (total:220 for 123 months)

Null hypothesis: There are no ADR signals in this time span is REJECTED ALERT (88.64% critical and 11.36% non critical reactions (out of 44) before the signal on 198612) (position 24)"

Overlapping window counts are given in Figure 6.

"Overlapping Time Window (X3) for Haemic and lymphatic systems - 121 time slots drug B01AB01 - Heparin

Null hypothesis: There are no ADR signals in this time span is REJECTED Very low frequency ALERT (86.67% critical and 13.33% non critical reactions (out of 180) before the signal on 199807)"

2) Drug giving weak signals: **Paracetamol and cardio-vascular (SOC 1000)**

The output if drug N02BE01 and cardiovascular reactions are selected, the histogram in Figure 7 is displayed and the following output is made:

"Cardiovascular system reactions for drug N02BE01 Parac-etamol (total of 20 reactions)

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
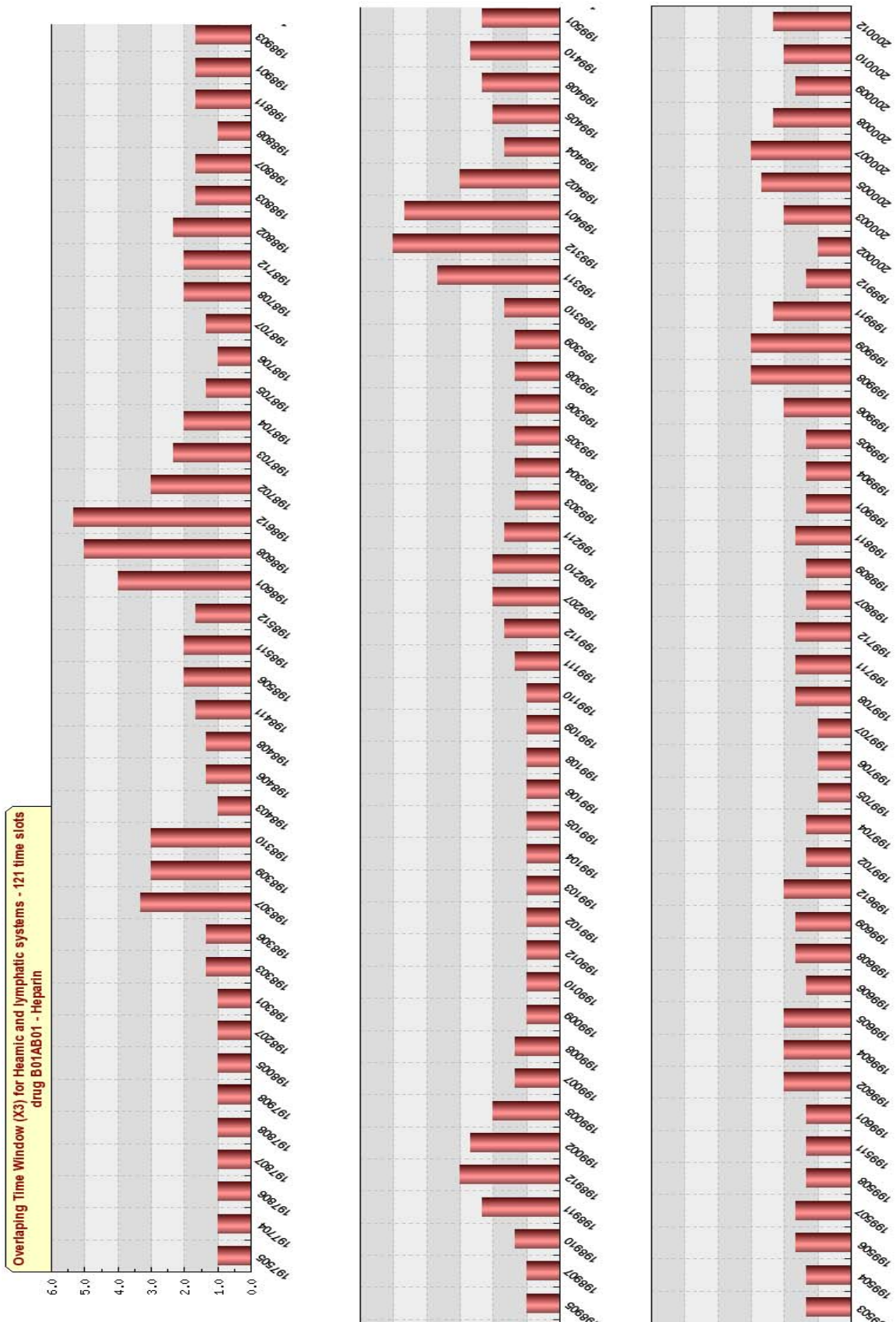
0-7695-2504-0/05 $20.00 © 2005 **IEEE**

Fig. 6. Heparin (B01AB01) Haemic and lymphatic systems reactions (SOC1200) – overlapping window reactions.

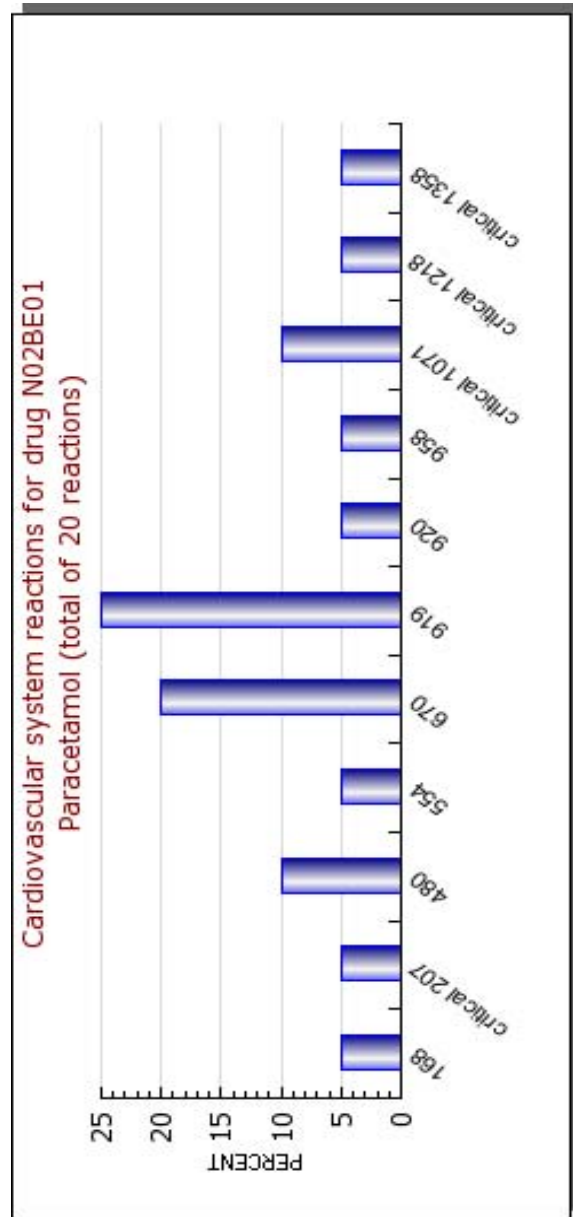Fig. 5. Heparin (B01AB01) Haemic and lymphatic systems reactions (SOC1200).

Fig. 7. Paracetamol (N02BE01) cardiovascular reactions (SOC1000).

1) Paracetamol (N02BE01) makes 0.25% (cases) of all records (out of 137279) 2) 22.81% (78 cases) of patients taking Paracetamol had CRITICAL reactions overall 3) 77.19% (264 cases)of patients taking Paracetamol had NON CRITICAL reactions overall 4) 25.00% (5 cases)of patients taking Paracetamol with Cardiovascular system reactions - CRITICAL and they make 1.46 of all patients taking Paracetamol 5) 75.00% (15 cases) of patients taking Paracetamol with Cardiovascular system reactions - NON CRITICAL and they make 4.39 of all patients taking Paracetamol

Null hypothesis: There are no ADR signals in this time span

is ACCEPTED No SIGNAL"

Paracetamol and Haemic and lymphatic systems (SOC 1200)

The output if drug N02BE01 and haemic and lymphatic systems reactions are selected, the histogram in Figure 8 is displayed and the following output is made:

"Haemic and lymphatic systems reactions for drug N02BE01 Paracetamol (total of 11 reactions)

1) Paracetamol (N02BE01) makes 0.25% (cases) of all records (out of 137279) 2) 22.81% (78 cases) of patients taking Paracetamol had CRITICAL reactions overall 3) 77.19% (264 cases) of patients taking Paracetamol had NON CRITICAL

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)

0-7695-2504-0/05 $20.00 © 2005 **IEEE**
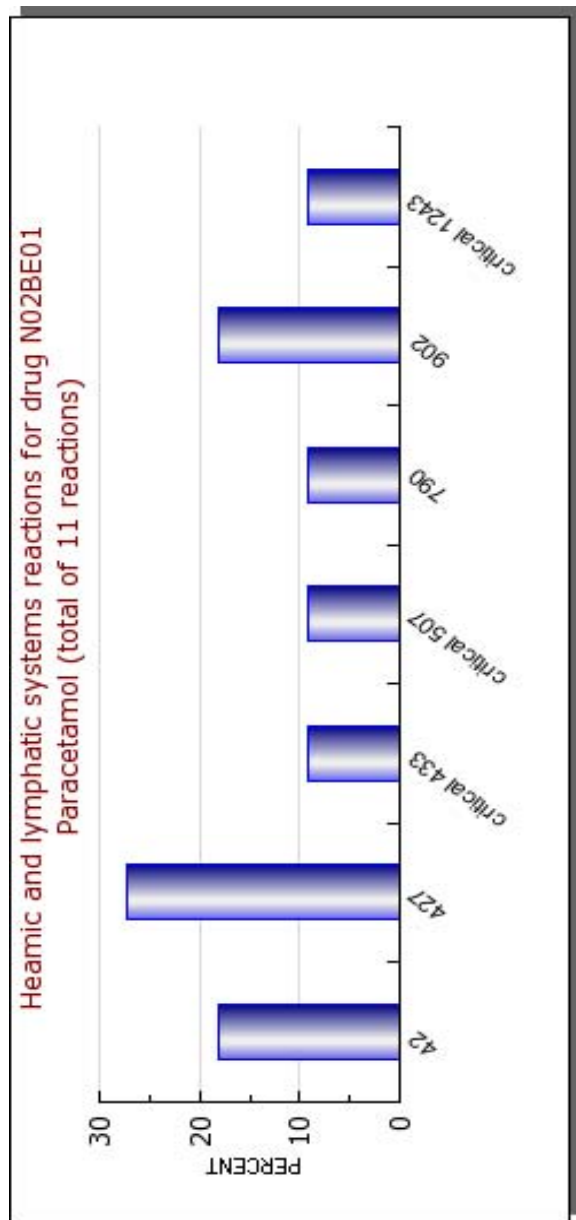
COMPUTER SOCIETY

Fig. 8. Paracetamol (N02BE01) haemic and lymphatic systems reactions (SOC1200).

reactions overall 4) 27.27% (3 cases) of patients taking Paracetamol with Haemic and lymphatic systems reactions - CRITICAL and they make 0.88 of all patients taking Paracetamol 5) 72.73% (8 cases) of patients taking Paracetamol with Haemic and lymphatic systems reactions - NON CRITICAL and they make 2.34 of all patients taking Paracetamol

Non-sliding windows graph Haemic and lymphatic systems - reaction monthly counts for drug N02BE01 - Paracetamol (total:11 for 11 months)

Null hypothesis: There are no ADR signals in this time span is ACCEPTED

NO SIGNAL"

III. DISCUSSION

In this paper we have presented case studies that illustrate the way in which the ontologies and the ADR signal detection approach work hand in hand to allow a user to understand at each level of the tree what may be involved in an adverse reaction to a drug or drug class. We chose examples of drugs that illustrated where a known signal is detected in particular reaction classes and one drug were a signal should not be detected and indeed is not, for the reaction classes in which the other two drugs gave signals. These examples show that the method behaves as expected. More extensive testing and refinement of this method is planned.

IV. CONCLUSION

The ATC ontology enables pooling of all data relating to a single ingredient, rather than having it split between different trade names. without this, the straight application of association rules would not be useful. Our approach provides a means for reducing the computation of all association rules and trying to filter them. This approach is stepwise meaningful because the ontologies provide meaning. The fact that there are several levels of granularity in this classification allows analysis at different levels. We have demonstrated that, using the ATC and SOC ontologies our association rule methods can explore even very rare associations in the ADRAC data. The tree structure of the ATC ontology was exploited to develop an adverse event signalling method that can 'drill down' the ATC tree to find individual drugs that have significantly more reactions in a given SOC. Then, using the critical term ontology, monthly reaction frequencies are examined to produce warnings when reactions rise to a significant level. A time windowing approach was used to smooth reaction frequencies. Thus the application of ontologies to drug safety data enables a significant advance in adverse event signalling methodology.

Acknowledgements

REFERENCES

[1] Manfred Hauben, Data Mining in Pharmacovigilance: Mining a Low Grade Ore?, *Abstracts of the 8th World Congress on Clinical and Experimental Pharmacology and Physiology*, Vol. 31, No. s1, i–x, August 2004. url: http://www.blackwell-synergy.com/links/doi/10.1111/j.1440-1681.2004.04042.x/abs/

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)

0-7695-2504-0/05 $20.00 © 2005 IEEE

[2] Ronald H.B. Meyboom, Marie Lindqvist, Antoine C.G. Egberts & I. Ralph Edwards, Signal Selection and Follow-Up in Pharmacovigilance, *Drug Safety*, Vol. 25, No. 6, 459–465, 2002.

[3] P. Tubert-Bitter & B. Begaud, Comparing Safety of Drugs, *Post Marketing Surveillance*, Vol. 7, 119–137, 1993.

[4] Eugène P. van Puijenbroek, Andrew Bate, Hubert G. M. Leufkens, Marie Lindqvist, Roland Orre & Antoine C. G. Egberts, A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions, *Pharmacoepidemiology and Drug Safety*, Vol. 11, No. 1, 3–10, 2002.

[5] S.J.W. Evans , P.C. Waller & S. Davis, Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports, *Pharmacoepidemiology and Drug Safety*, Vol. 10, No. 6, 483–486, November 2002.

[6] Manfred Hauben & Xiaofeng Zhou, Quantitative methods in pharmacovigilance: focus on signal detection, *Drug Safety*, Vol. 26, No. 3, 159–186, 2003.

[7] A. Bate , Marie Lindqvist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner & R. De Freitas, A Bayesian neural network method for adverse drug reaction signal generation, *European Journal of Clinical Pharmacology*, Vol. 54, No. 4, 459–465, 1998.

[8] Ana Szarfman, Stella G. Machado & Robert T. O'Neill, Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than-Expected Combinations of Drugs and Events in the US FDA's Spontaneous Reports Database, *Drug Safety*, Vol. 25, No. 6, 381–392, 2002.

[9] A. Bate , Marie Lindqvist, R. Orre, I.R. Edwards & R.H.B. Meyboom, Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs, *European Journal of Clinical Pharmacology*, Vol. 58, 483–490, 2002.

[10] M.A. Mamedov & G.W. Saunders, An Analysis of Adverse Drug Reactions from the ADRAC Database. Part1: Cardiovascular group, *University of Ballarat School of Information Technology and Mathematical Sciences Research Report*, No. 02/01, 1–48, February 2002. url: http://www.ballarat.edu.au/ard/itms/publications/researchPapers/papers2002.shtml

[11] V. Pinkston & E.J. Swain, Management of Adverse Drug Reaction and Adverse Event Data through Collection, Storage and Retrieval, *Detection of New Adverse Drug Reactions*, M.D.B. Stephens, J.C.C. Talbot,& ,P.A. Routledge (Ed), Macmillan Reference Ltd, 281–296, 1998.

[12] C. Chuang-Stein, Statistics for Safety Data, *Detection of New Adverse Drug Reactions*, M.D.B. Stephens, J.C.C. Talbot,& ,P.A. Routledge (Ed), Macmillan Reference Ltd, 271–279, 1998.

[13] World Health Organization , The WHO Adverse Reaction Terminology, *Technical Report on Reaction Terminology*, 2002. url: http://www.umc-products.com/graphics/3149.pdf

[14] M. Mamedov & G. Saunders, A Fuzzy Derivative Approach to Classification of outcomes from the ADRAC database, *International Transactions in Operational Research*, Vol. 11, No. 12, 169–180, March 2004.

[15] M.A. Mamedov, G.W. Saunders & E. Dekker, A Study of Drug-Reaction Relationships in Australian Drug Safety Data, *Proceedings of the 2nd Australian Data Mining Workshop*, S.J. Simoff, G.J. Williams& M. Hegland, (Ed), University of Technology Sydney, 143–162, December 20003. url: http://datamining.csiro.au/adm03/

[16] Gary W. Saunders, Draft anatomical therapeutic chemical ATC classification of Australian drug safety data, *University of Ballarat School of Information Technology and Mathematical Sciences Research Report*, No. 04/22, 1–10, February 2002. url: http://www.ballarat.edu.au/ard/itms/publications/researchPapers_2004/04-22.pdf

[17] World Health Organization , WHO-ART Critical Term List, *Current Critical Term List*, World Health Organization, Uppsala Monitoring Centre (UMC), 2002.

[18] M. Hauben & Lester Reich, Safety Related Drug-Labelling Changes: Findings from Two Data Mining Algorithms, *Drug Safety*, Vol. 27, No. 10, 735–744, 2004.

[19] M. Hauben, A brief primer on automated signal detection, *The Annals of Pharmacotherapy*, Vol. 37, No. 7, 1117–1123, 2004.

[20] Ana Szarfman, W. DuMouchel, D. Fram, J. M. Tonning, J. Almenoff, R. D. Fleischer & J. G. Levine, Lactic acidosis: Unraveling the Individual Toxicities of Drugs Used in HIV and Diabetes Polytherapy by Hierchical Baysean Logistic Regression Data Mining, *11th Annual FDA Science Forum*, U. S. Food and Drug Administration, April 2005.

[21] M. Ståhl, M. Lindqvist, I. R. Edwards & E. G. Brown, Introducing triage logic as a new strategy for the detection of signals in the WHO Drug Monitoring Database, *Pharmacoepidemiology and Drug Safety*, Vol. 13, 355–363, 2004.

[22] M. Ståhl, I. R. Edwards, Geoffrey Bowring, Anne Kiuru & Marie Lindqvist, Assessing the impact of drug safety signals from the WHO database presented in 'SIGNAL': results from a questionnaire of National pharmacovigilance Centres, *Drug Safety*, Vol. 26, No. 10, 721–727, 2003.

[23] Rakesh Agrawal, Tomasz Imielinski & Arun Swami, Mining Association Rules between Sets of Items in Large Datasets, *Proceedings of the ACM SIGMOD Conference on Management of Data*, Washington, D.C., May 1993.

[24] Saunders, Gary, Ivkovic, Sasa, Ghosh, Ranadhir, and Yearwood, John, Applying Anatomical Therapeutic Chemical (ATC) and Critical Term Ontologies to Australian Drug Safety Data for Association Rules and Adverse Event Signalling. Australasian Ontology Workshop (AOW 2005) in: The 18th Australian Joint Conference on Artificial Intelligence, December 2005, Sydney, Australia. http://www.cse.unsw.edu.au/ aow05/

IEEE
COMPUTER
SOCIETY