EXPLORING HOSTILE PSYCHIATRIC PILLS REACTION SIGNAL PAIRS

G.Gomathi[1], J.Ganesh[2]

[1,2] *PG scholar, Assistant Professor, Dr.Sivanthi Aditanar College of Engineering(India)*

ABSTRACT

Mining basic relationships between events with less probability results is really important in various real-world problems. Judging the basic relations can prevent the negative results sourced from their predecessor. Here an interesting data mining skeleton is applied to extract basic relations from the digital data sets. Here the pill-related measures of interest happen rarely. Particularly, a new interestingness measure, special basic-leverage is used which is stand on the fuzzy recognition-primed decision (RPD) model for decision-making. Based on this founded new measure, a mining algorithm was generated to extract basic relationships among pills and their related Hostile pill reactions (HDRs).Psychiatric pills for depression and its associated symptoms are taken to find out the severity of each pill- symptom pair. The result shows the usefulness in finding the severity of each pair for further analysis.

Index Terms: Hostile Pill Reactions, Association Rules, Data Mining Algorithms, Interestingness Measure, and Recognition Primed Decision Model

I. INTRODUCTION

Judging basic relations between various events with relatively minimum frequency is very useful for several real-world applications. For example, an appropriate dose for a pill may cause one or more Hostile pill reactions (HDRs), even though the probability is minimum. Finding this kind of basic relationships can help us prevent or correct negative outcomes caused by its antecedents. However, mining these relationships is challenging due to the difficulty of capturing fundamentality among events and the infrequent nature of the events of interest in these applications.

In this paper, a knowledge-based approach is employed to capture the degree of fundamentality of an event pair within each sequence since the determination of fundamentality is often ultimately application or domain dependent. An interestingness measure that incorporates the fundamentality across all the sequences in a database is developed. The study was motivated by the need of finding HDR signals in postmarketing surveillance, even though the proposed framework can be applied to many different applications. HDRs represent a serious world- wide problem [1], [2]. They can complicate a patient's medical condition or contribute to increased morbidity, even death. HDRs contribute to about 5 percent of all hospital admissions and represent the fifth commonest cause of death in hospitals [3].

Even though premarketing clinical trials are required for all new pills before they are approved for marketing, these trials are necessarily limited in sample- size and duration, and thus are not capable of detecting rare HDRs. In general, an HDR cannot be recognized by these trials if its occurrence rate is less than 0.1 percent [1]. Therefore, pill safety depends heavily on postmarketing surveillance; that is, the monitoring of impacts

of medicines once they have been made available to consumers.

As electronic patient records become more and more easily accessible in various health organizations such as hospitals, medical centers, and insurance companies, they provide a new source of information that has great potential to generate HDR signals much earlier [4], [5].

In order to mine infrequent basic relations efficiently, it is necessary to develop a new data mining framework. The framework includes 4 steps

1. An exclusion mechanism that can effectively reduce the undesirable effects caused by frequent events. A new measure is named special basic- leverage measure is used.

2. A data mining algorithm to mine HDR signal pairs from electronic patient database based on the new measure is proposed. The algorithm's computational complexity is analyzed.

3. The new special basic-leverage measure with previously proposed basic-leverage measure as well as two traditional measures in the literature: leverage and risk ratio.

4. How the length of hazard period T affects the performance of the special basic-leverage measure is also found to show the superiority of this measure.

II. RELATED WORK

Finding the association between two event sets is an active and important area of data mining research [6], [7], [8], [9]. Many measures have been proposed to do mining association rules in the form of X⟶Y, where X and Y are two event sets and they are disjoint. For instance, the support and confidence measures were the original interestingness measures proposed for association rules [10]. The support of an association rule supp (X⟶Y) is the proportion of sequences in which both X and Y occur at least once, among all the event sequences. The confidence of an association rule is defined as conf (X⟶ Y) = supp (X⟶Y) /supp (X ⟶), where supp(X ⟶) is the proportion of sequences that contain X. Given these two measures, the association rule mining problem can be formalized as Judging those rules whose support and confidence are greater than prespecified thresholds minimumsupport and minimumconfidence, respectively.

There exist some works on rare eventset (or item set) mining in the literature [11], [12], [13], [14]. A straightforward approach to finding infrequent eventsets is to relax the uniform minimal support criterion (i.e., minimumsupport) as indicated in [11]. One drawback of this mining approach is the extremely high computational cost [12], [13]. In addition, if this approach was used to discover rare events like HDRs, thresholds minimumsupport and minimumconfidence would have to be set very small, possibly leading to a lot of false relations [5].

A couple of studies attempted to use a less uniform support criterion [12], [13], [14]. Yun et al. adopted a relative support (a rate against the relative frequency of the data existing in a database) approach [13], while Liu and his colleagues proposed a method that relied on multiple minimal support thresholds specified by users [12]. These approaches output all frequent eventsets and association rules together with a subset of all infrequent ones.

The recognition-primed decision (RPD) [16] model is a primary naturalistic decision-making approach which seeks to explicitly recognize how human decision makers handle complex tasks and environment based on their experience.

III. PROPOSED FRAMEWORK

3.1 A New Special Basic-Leverage Measure

<X,Y > and C<X,Y > is used to represent a pair of events and the degree of fundamentality of the pair in a sequence, respectively. C<X,Y > is calculated by matching the token values extracted from a sequence with the token values of the defined experiences, each of which corresponds to a unique category of fundamentality. Specifically, a vector V =(c_1,c_2, . . .,c_i, . . . ,c_m) is used to represent the set of token values extracted from an event sequence, where c_i represents the value of *i*th token and m is the total number tokens. Similarly, V'= (c_1',c_2',. . . ,c_i,. . . ,c_m) is used to represent the set of token values in an experience. The similarity between a pair of token values c_i and c_i' i as local similarity $S_L(c_i,c_i')$whose calculation depend on the type of the token (i.e.,quantitative, nominal or fuzzy); that is,

$S_L(c_i,c_i') = 0$,

if c_i or c_i' is unknown

overlap(c_i,c_i'),

if token i is nominal

1-normalized_diff(c_i,c_i'),

if token i is quantitative

Fuzzy_sim(c_i,c_i'),

if token i is fuzzy

overlap(c_i,c_i')=1,if c_i=c_i',

0,otherwise

normalized_diff(c_i,c_i')=|c_i-c_i'|/delta$_i$ where

delta$_i$=a_i-b_i

a_i,b_i are the maximum and minimum values for token i respectively

Fuzzy_sim(c_i,c_i')= poss(c_i,c_i'),if poss(c_i,c_i') <0.5

Where poss(c_i,c_i')=max(min($\mu_{Ci}(x)$,$\mu_{Ci'}(x)$)) for all xϵX

Here $\mu_{Ci}(x)$,$\mu_{Ci'}(x)$ are membership functions of fuzzy set c_i,c_i' respectively.

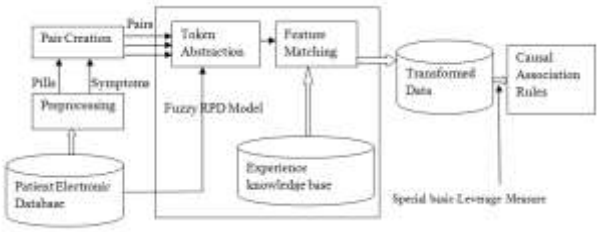

Figure 1: Algorithm For Mining Basic Association Rules.

Fig. 1 shows an overall picture of the data mining algorithm. First, preprocessing is needed to get two types of information: 1) a list of all pills P (p_1; p_2; . . . ; p_m) in the database and the support count for each pill; 2) a list of all symptoms S(s_1; s_2; . . . ; s_n) in the database and the support count for each symptom. The lists of pills and symptoms are needed to form all possible pill-symptom pairs whose basic strengths will be assessed. Since a patient database normally only contains a subset of all pills on the market and a subset of all symptoms, it is necessary to search the Patient Pill Table and the Patient Symptom Table to get the pills and symptoms covered by the database. Using the discovered pills and symptoms in the database (instead of all

pills on the market and all available symptoms) can avoid forming unnecessary pairs, which will reduce the computational complexity of the data mining algorithm. In addition, the support count for each pill or symptom will be used to calculate the special basic-leverage value for related pairs. Hence, their values also need to be computed.

Algorithm 1.Pills and Support Count searching for each pill

1: pillHashTable = null

2: for each patient $P_k \epsilon$ Database do

3: retrieve all the pills D_k taken by the patient

4: for each pill $p_{kl} \epsilon D_k$ do

5: σ=pillHashTable.getValue (p_{kl}) +1{update support count}

6: pillHashTable: putValue (p_{kl}, σ)

7: end for

8: end for

9: return (pillHashTable)

Algorithm 2. Pair Creation and Estimation

1: for each pill $p_m \epsilon$ D do

2: for each symptom $s_n \epsilon$ S do

3: retrieve PIDs that support pair <p_m, s_n> from database

4: if (count (PIDs) >= mincount) then

5: x= basiclev(p_m, s_n, PIDs)

6: y= reverse basiclev (p_m,s_n) = basic- leverage(s_n,p_m,PIDs)

7: special basiclev value = x- y

8: output pair <p_m, s_n> and its special basiclev value

9: end if

10: end for

11: end for

Algorithm 3. Method basiclev(X,Y,PIDs)

1: search pill/symptom hash table to get support count for
Y-σ_Y

2: for each PID that supports the pair do

3: V = token-abstraction(PID)

4: SV = {sv|sv = S_G(V ,V') ^ V' ϵ EKB}
{calculate similarity values}

5: SV' = normalization(SV)

6: C<X;Y > = weightedSum(SV',W){W: weights calculated by w_i=(i-1)/(m-1) }

7: if ($C_{<X,Y>}$ > 0) then

8: accumulatedVotes+=$C_{<X,Y>}$

9: contributionCases++{number of cases whose votes are greater than 0}

10: end if

11: end for

12: supp(X $\xrightarrow{c}$ Y) = accumulatedVotes/N

13: supp(X $\xrightarrow{c}$)= contributionCases/N

14: supp($\xrightarrow{c}$ Y)=σ_Y/N

15: return supp ((X $\xrightarrow{c}$ Y) - supp(X $\xrightarrow{c}$) * supp($\xrightarrow{c}$ Y)

Algorithm 1 shows how to search a Database for the list of pills and the support count for each pill. The discovered pills and their support counts are stored in a hash table. Initially, the hash table is empty. P_k is used to represent the set of pills taken by the kth patient p in the database. For each patient, first retrieve all the pills taken the patient. For each of these pills, then check whether the hash table contains the pill, then the pill's support count is increased by 1. After all the patient cases are searched, the hash table is returned. Note that, when calculating the support count for a pill, it is counted only once for one patient case even if the pill appears several times in that patient case. One can see that the computational complexity of this searching process can be affected by the total number of patients N in the database and the average number of pills taken by a patient. The later is normally determined by the type of patients in the database. For instance, old patients often have multiple diseases and thus may take multiple pills either at the same time or different times.

After getting the list of pills P(p_1, p_2, . . . ,p_m) and the list of symptoms S(s_1, s_2,. . . , s_n), the next step is to generate all the possible pairs, each of which represents a CAR. Algorithm 2 shows the process for pair generation and evaluation. Please note that most existing data mining methods mine all interesting association rules that combine all possible events or items in a database. Pill- symptom pairs can be easily generated, given D and S. The complexity of this pair generation process is O(m * n) where m and n are the number of pills and symptoms, respectively. In addition, mining infrequent patterns is done, it is inappropriate to prune pairs using the support measure (i.e., support > minimumsupport). However, in postmarketing surveillance, a signal pair generated by a data mining method is generally not considered as valid if only one or two patient cases contain the pair [9]. Therefore, utilize a minimum support count mincount = 5 to further reduce the number pairs that will be evaluated. Specifically, retrieve the PIDs of those patient cases that contain both the pill and the symptom of each pair. This is done by sending a SQL statement to query the database. Modern database management systems can utilize optimization techniques like index to speed up this process. If the number of PIDs that support a pair is greater than or equal to mincount, the pair is evaluated. To evaluate the strength of the basic association of the pair, its basic-leverage value is first computed. Then, the pair is reversed. That is, the pair <pm,s_n> becomes <s_n,d_n>. The reverse basic-leverage value of pair <pm,s_n> is equal to the basic-leverage value of its reversed pair <s_n,pm>.After that, the special basic-leverage value of the pair is computed by subtracting its reverse basic-leverage value from its normal basic-leverage value.

Algorithm 3 shows how to compute the basic- leverage value of a general pair between event X and Y. Both X and Y could be either pill event or symptom event. First, the pill or symptom hash table is searched in order to get the support count for event Y. Then, for each PID that supports the pair, a process called token abstraction is used to extract a set token values V from the related patient case. Specifically, a list of pill start dates and a list of symptom dates are retrieved from the Patient Pill Table and the Patient Symptom Table, respectively. Note that, since each patient case records the patient's history for a long period of time, the same pill may be prescribed many times and thus multiple pill start dates may exist within one patient case. Similarly,

there may exist multiple symptom dates for the symptom within the same patient case. Therefore, comparing each start date of the pill with each symptom date of the symptom in order to obtain interested temporal patterns for the pair within the case is done. The complexity of this process is $O(L_d * L_s)$, where L_d and L_s are the length of the list of pill start dates and the length of the symptom dates, respectively. L_d and L_s often depend on the characteristics of the patient. For example, a patient with chronic diseases tends to take the same pill for many times and have repeated symptoms. After getting the temporal patterns, token values for temporal association, rechallenge, and dechallenge are derived from these patterns using fuzzy rules. Note that, in order to make this algorithm more generic, the token other explanations was not utilized in this study since it is pill dependent. Readers are referred to the previous paper for specific fuzzy rules that were used in this process [16]. After the set of token values V of the pair are extracted from the related patient case, similarity values are computed between V and each set of token values V' in the experience knowledge base. After that, the degree of fundamentality C<X, Y> within the current patient is computed. If C<X, Y> is greater than 0, it is added to the accumulated votes and the number of cases whose votes are greater than 0 is increased by one. After the above computation is done for all the supporting cases of the pair, the basic-leverage value of the pair is computed and returned. Finally, rank all the pairs in a decreasing order according to their special basic-leverage values after all these values are computed. The higher thevalue, the more likely a basic association exists in the pair or CAR.

IV. RESULTS

The experimental results showed that this special basic-leverage measure outperformed traditional interestingness measures like risk ratio and leverage because of its ability to capture suspect basic relationships and exclude undesirable effects caused by frequent events. However, due to the complexity, incompleteness, and potential bias of the data, some HDR may still be falsely ranked high based on the special basic-leverage. That is, false relations cannot be completely avoided using this method. Please note that, in postmarketing surveillance, data mining is primarily used to shorten the long list of potential pill-HDR pairs. Like the other data mining methods, the signal pairs found by this approach will be subject to further analysis (e.g., epidemiology study), case review, and interpretation by pill safety professionals experienced in the nuances of pharmacoepidemiology and clinical medicine [15].

V. CONCLUSION

Mining the basic association between two events is very vital and useful in many real applications. It can help people discover the fundamentality of a type of events and avoid its potential Hostile effects. However, mining these relations is very difficult especially when events of interest occur infrequently. A new measure, special basic-leverage, is developed based on an experience-based fuzzy RPD model. This measure can be used to quantify the degree of association of a CAR. Moreover, the measure was designed to mask the undesirable effects caused by high-frequency events. A data mining algorithm was developed to search a real electronic patient database for potential HDR signals. Results showed that this algorithm could effectively make known HDRs rank high among all the symptoms in the database.

REFERENCES

[1] J. Talbot and P. Waller, Stephens' Detection of New Hostile Pill Reactions, fifth ed. John Wiley & Sons, 2004.

[2] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, To Error is Human: Building a Safer Health System. Nat'l Academy Press, 2000.

[3] J. Lazarou, B.H. Pomeranz, and P.N. Corey, ''Incidence of Hostile Pill Reactions in Hospitalized Patients: A Meta-Analysis of Prospective Studies,'' J. Am. Medical Assoc., vol. 279, pp. 1200-1205, Apr. 1998.

[4] Y. Ji, H. Ying, M.S. Farber, J. Yen, P. Dews, R.E. Miller, and R.M. Massanari, ''A Distributed, Collaborative Intelligent Agent System Approach for Proactive Postmarketing Pill Safety Surveillance,'' IEEE Trans. Information Technology in Biomedicine, vol. 14, no. 3, pp. 826-837, Dec. 2010.

[5] H. Jin, J. Chen, H. He, G. Williams, C. Kelman, and C. O'Keefe, ''Mining Unexpected Temporal Relations: Applications in Detecting Hostile Pill Reactions,'' IEEE Trans. Information Technology in Biomedicine, vol. 12, no. 4, pp. 488-500, July 2008.

[6] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Addison-Wesley, 2005.

[7] L. Geng and H.J. Hamilton, ''Interestingness Measures for Data Mining: A Survey,'' ACM Computing Surverys, vol. 38, no. 3, article 9, 2006.

[8] A.K.H. Tung, H. Lu, J. Han, and L. Feng, ''Efficient Mining of Intertransaction Association Rules,'' IEEE Trans. Knowledge and Data Eng., vol. 15, no. 1, pp. 43- 56, Jan./Feb. 2003.

[9] C. Marinica and F. Guillet, ''Knowledge-Based Interactive Postmining of Association Rules Using Ontologies,'' IEEE Trans. Knowledge and Data Eng., vol. 22, no. 6, pp. 784-797, June 2010.

[10] R. Agrawal and R. Sreikant, ''Fast Algorithms for Mining Association Rules,'' Proc. 20th Int'l Conf. Very Large Databases, 1994.

[11] G.M. Weiss, ''Mining with Rarity: A Unifying Framework,'' ACM SIGKDD Explorations Newsletter, vol. 6, pp. 7-19, 2004.

[12] B. Liu, W. Hsu, and Y. Ma, ''Mining Association Rules with Multiple Minimum Supports,'' Proc. Fifth ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining, 1999.

[13] H. Yun, D. Ha, B. Hwang, and K.H. Ryu, ''Mining Association Rules on Significant Rare Data Using Relative Support,'' J. Systems Software, vol. 67, pp. 181-191, 2003.

[14] Y.-H. Hu and Y.-L. Chen, ''Mining Association Rules with Multiple Minimum Supports: A New Mining Algorithm and a Support Tuning Mechanism,'' Decision Support Systems, vol. 42, pp. 1-24, 2006.

[15] M. Hauben and X. Zhou, ''Quantitative Methods in Pharmacovigilance: Focus on Signal Detection,'' Pill Safety, vol. 26, pp. 159-186, 2003.

[16] Yanqing Ji, Hao Ying, R. Michael Massanari, John Yen, ''A Fuzzy Logic-Based Computational Recognition- Primed Decision Model''.