Exploring the Use of a Network Model in Drug Prescription Support for Dental Clinics

Wee Pheng Goh, Xiaohui Tao, Ji Zhang
Faculty of Health, Engineering and Sciences
University of Southern Queensland
Australia
{*wee.goh, xiaohui.tao, ji.zhang*}*@usq.edu.au*

Jianming Yong
Faculty of Business, Education, Law and Arts
University of Southern Queensland
Australia
jianming.yong@usq.edu.au

Yongrui Qin
University of Huddersfield
United Kingdom
y.qin2@hud.ac.uk

Elizabeth Zhixin Goh
Glory Dental Surgery Pte Ltd
Singapore
eg@glory.sg

Aimin Hu
Guilin Tourism University
China
763523955@qq.com

***Abstract*—With more patients taking multiple medications and the increasing digital availability of diagnostic data such as treatment notes and x-ray images, the importance of decision support systems to help dentists in their treatment planning cannot be over emphasised. Based on the hypothesis that a higher similarity ratio between drugs in a drug-pair indicates that the combination of the drug-pair has a higher chance of an adverse interaction, this paper describes an efficient approach in extracting feature vectors from the drugs in a drug-pair to compute the similarity ratio between them. The feature vectors are obtained through a network model where the information of the drugs are represented as nodes and the relationships between them represented as edges. Experimental evaluation of our model yielded a superior F score of 74%.**

The use of a network model will drive research efforts into more efficient data-mining algorithms for information retrieval, similarity search and machine learning. Since it is important to avoid drug allergies when prescribing drugs, our work when integrated within the clinical work-flow will reduce prescription errors thereby increasing health outcomes for patients.

***Keywords*-drug adverse interaction; clinical decision support; network model; drug prescription;**

I. Introduction

The amount of data available to dentists has increased tremendously in recent years. This has led to increasing dependence on a clinical decision support system (CDSS) to aid the treatment planning process especially for drug prescription. Drugs such as painkillers and antibiotics are very commonly prescribed in the dental setting to relieve pain and resolve infections, in primary and supplementary roles. Reducing and relieving pain with analgesic medications are naturally important procedures within the clinical work-flow of a dental clinic [7]. For patients with an abscessed tooth, common drugs such as *amoxicillin* and *clindamycin* are also used to treat the infections as a supplement to a root canal treatment or removing the tooth. Antibiotics are also used after surgical procedures such as placement of dental implants and gum treatment [10].

Though there are systems which examine the interactions within a pair of drugs, they are not associated with the medical profile of the patient. Hence, we aim to pursue a framework which obtains content-specific data from publicly available data sources, performs data-mining on the data and presents the results to the dentist with consideration of the individual patient's drug allergies and the drugs they are currently taking. Although the data-mining layer of our three-tier framework can include many approaches for extracting information, this paper examines the use of a network model which is based on the hypothesis that a higher similarity ratio indicates a greater chance of an adverse interaction between drugs in a drug-pair. The network model provides an effective platform for obtaining feature vectors and hence their similarity ratio based on the number of common paths connecting the set of common interacting drugs between the drug-pair. In our model, we represent all drugs as nodes in a network to enable us to compute their proximity in terms of the number of shared entities between the drug-pair as reflected in the number of connecting paths between them. The primary theoretical contribution of this research is the use of the network model in hierarchically representing the drug-pairs within the context of a CDSS considering the individual patient's medical profile. Examining the paths linking the common drugs within the set of interacting drugs for each drug in the drug-pair allows the system to arrive at a similarity ratio. This then allows the dentist to decide if the drug is safe for the patient. Such a system containing information on interacting drug-pairs based on the patient's personal profile will also be useful in clinical education relating to drug dispensing, such as in medicine, nursing and pharmacy.

The rest of paper is organised as follows: Section II

discusses the related work in data-mining and detection of interactive drug-pairs; Section III explains the framework and Section IV describes the experiment while the results are discussed in V. Section VI describes a clinical scenario to demonstrate if a drug is safe based on a specific patient's profile and Section VII presents the conclusions obtained with an outline of potential future work.

II. Related work

Many works on drug interactions have been performed in recent years, notwithstanding the lack of a single dataset with information on drug interactions from all publicly available sources [1]. Hence, knowledge representation as per the source of the dataset is used in the design of many CDSS for making diagnostic inferences ([6], [5])

By using data-mining techniques, semantic meanings and context-specific knowledge based on medical profiles of patients can be obtained from the knowledge base. Wright *et al.* used a data-mining method to identify relationships between diabetic medications [12]. By identifying patterns within the drug database, the system was able to predict the subsequent medication to be prescribed with significant accuracy.

When predicting the relationship of a drug-pair, relevant information from knowledge bases are usually extracted and examined. For example, Tari *et al*. developed a method combining text mining and automated reasoning to predict enzyme-specific drug-drug interactions (DDI) [11]. Yan *et al*. also used text mining techniques to create features based on relevant information such as genes and disease names extracted from drug databases to augment limited domain knowledge [13]. These features were then used to build a logistic regression model to predict DDI.

Though these systems predict the extent of drug interaction within the drug-pair, they are not customised to the medical profile of the patient. Of the many methods used to create feature vectors to compute the similarity of a drug-pair to predict DDI, this paper will explore the use of a network model, where drugs are represented on each node with their relationship represented as edges. This model fits into the data-mining layer of our framework as elaborated in the next section.

III. Framework

As mentioned earlier, knowledge representation is a typical technique used in CDSS development. The knowledge base needs to be updated regularly to represent knowledge in a timely manner, which is imperative for CDSS user acceptance [3]. The separation of the knowledge base from the main information systems allows the former to be updated and managed effectively, for each layer to be developed and maintained independently while simultaneously ensuring standardised interfacing between layers and cheaper integration with existing systems [4].

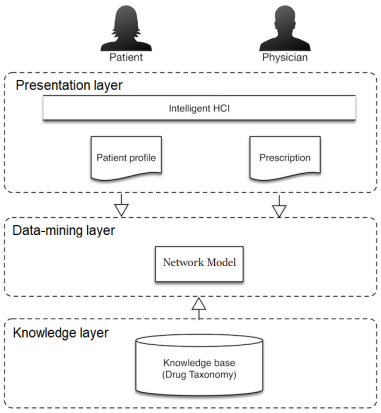

Figure 1: Three-tier architecture of the proposed model

Hence our framework has a three-tier architecture (Figure 1) which separates the knowledge layer containing the knowledge base from the data-mining layer and the presentation layer.

The network model, residing in the data-mining layer aims to compute the similarity ratio of a drug-pair based on the feature vectors gathered from the knowledge layer. Our approach adapts from [2] which describes the theoretical foundations of measuring similarity within a network.

To obtain the feature vectors and the similarity ratio within this model, the logarithm begins by formally defining the set of interactive drugs as $A^r = \{a_1, a_2...a_k\}$ where *r* is the attribute of the relationship with the vertex drug-pair, which can be minor, moderate or major.

The items in A is also the subset of out-neighbours of drug d_1, denoted as $O(d_1)$. Individual drugs in A which interact with d_1 is then denoted by $O_i(d_1)(1 \leq i \leq |O(d_1)|)$. Hence, $O_i(d_1) \in \{A^r\}$

This is followed by getting the path and the vectors

Get Path Using notation from [9], the path from drug d_1 to set of interactive drugs A with ratings r is denoted by: $d_1 \xrightarrow{r} A$, which can also be written as $d_1(r)A$ where r is the relationship between d_1 and A. For our model, the relationship $r = \{1, 2, 3\}$.

Thus, $d_1(1)A$ shows drug d_1 has a minor interaction with the drugs in set A.

Get Vectors Create single row matrix M^r for drugs d_1 and d_2 to indicate the positional match in the adverse drugs for d_1 and d_2 with interaction rating r where the number of columns in M^r = $|A^r|$.

If $O_i(d_1)$ is found in A^r at position u, then the u^{th} column in matrix M^r will be updated as r, $M^r(1, u) \longleftarrow r$.

IV. Experiment

We chose an experimental approach to assess the accuracy and efficiency of the proposed method, testing the hypothesis that similar drug-pairs have a higher similarity ratio compared to that of dissimilar pairs.

The comprehensive dataset on DDI required for this project was sourced from drugbank.ca, a richly annotated database popularly used to build drug repositories on drug indications [8].

From the set of drugs that interacts adversely with each drug in the drug-pair, paths and feature vectors were built. This data-mining process aimed to discover the number of common paths for each drug-pair. As explained in Section III, vectors were then built so that a similarity ratio can be computed for the drug-pair.

With the similarity ratio results from the experiment, a threshold of θ = 0.5 was used to predict if the drug-pair was similar. A value of 0.5 or higher meant the drug-pair was considered similar, and a value lower than 0.5 meant the drug-pair was considered dissimilar. The performance of the model was measured by counting the number of correct predictions (true positives). For convenience and ease of computation, the standard metrics recall, precision and F score were used to gauge how well the prediction was being made. Precision indicated how accurately the model predicted drug-pairs as similar, while recall indicated how accurately similar drug-pairs were predicted. Thus,

$$precision = \frac{TP}{TP + FP} \qquad recall = \frac{TP}{TP + FN} \tag{1}$$

where TP is True Positive, FN is False Negative, and FP False Positive.

V. Results and Discussion

Cut off	r=1 (Major)	r=1 (Combined)	r=2 (Major)	r=2 (Combined)
0.1	0.55	0.61	0.52	0.70
0.2	0.51	0.55	0.45	0.59
0.3	0.57	0.42	0.41	0.49
0.4	0.68	0.44	0.40	0.44
0.5	0.74	0.43	0.47	0.36
0.6	0.74	0.43	-	0.35
0.7	0.71	0.41	-	0.38
0.8	0.68	0.39	-	0.34
0.9	0.67	0.39	-	-

Table I: F score distribution

Table I shows the F score when the experiment is run with a dataset of positive pairs and negative pairs. When the threshold is 0.5, the model produces a F score of 0.74 with a recall rate of 0.61 and precision of 0.94. r=1 indicates only common paths in the direct neighbourhood of the drug is considered. When the distance is extended at r=2, where paths connecting the drugs that interact with the drug-pair are also considered, it can be noticed that the F score drops drastically. As expected, the performance deteriorated when additional attributes of adverse interactions were introduced such as minor and moderate interactions. However, due to fewer possible paths when only major interactions were considered, the cut-off occurs sooner, where beyond that, there were no true positives obtained in the experiment. This explains the unavailability of F score when the threshold was over 0.6.

Since both positive and negative outcomes are important within the clinical workflow, we also examined the performance of our model to predict true negatives. To do so, a plot of true positive rate *tpr* against true negative rate *tnr* (Equation 2) was used to obtain the area under the receiver operating characteristic curve (AUC).

$$tpr = \frac{TP}{TP + FN} \qquad tnr = \frac{FP}{FP + TN} \tag{2}$$

With this plot, at r=1 with paths connecting only the major interactions considered, the AUC yielded 0.61 compared to 0.36 when minor and moderate interactions were included. This was expected due to the noise introduced by the additional paths.

VI. Adapting the model to prescription support for dental clinics

Our three-tier framework can be easily applied to a clinical environment to support the prescription of drugs to patients. This section describes a typical CDSS to help decide if the drug is safe for prescription.

Given the set of prescription drugs, an algorithm can be developed to adapt the framework to a clinical scenario to ensure the drug is not in adverse relationship with what the patient is taking and is dissimilar to the drugs that the patient is allergic to.

From Algorithm 1, there are two tests that consider the adverse relationship and the similarity within drug-pairs. To facilitate suggestion of an alternative drug by the system, a set of candidate drugs, similar to the drug to be prescribed, was created. Based on the relationship within drug-pairs in the drug taxonomy located in the knowledge layer, the system searched for any adverse relationship between the drug to be prescribed and each drug that the patient is currently taking. If there was no adverse relationship detected, the system proceeded to test for similarity between the drug to be prescribed and each drug in the set of drugs that the patient is allergic to. Each drug in the prescription set of drugs had to clear both tests. If either test was not cleared, an alternative drug in the candidate set was suggested. If the candidate set was exhausted, the system would then be unable to recommend an alternative drug.

Consider a patient who is allergic to *penicillin* and is currently taking *warfarin*. During a dental appointment, it is decided that antibiotics are required. The dentist may consider the commonly prescribed *amoxicillin*. Use of our model will note the similarity of *amoxicillin* to *penicillin*, which the patient is allergic to, and will thus recommend an alternative drug. As indicated in Algorithm 1, not only should this drug be dissimilar to *penicillin*, it should also not be in adverse relationship with *warfarin*. One such drug

Algorithm 1: Applying Model for personalised prescription support

input : Let D^p be the set of prescription drugs;
Medicine to be prescribed by the dentist
Medicine the patient is currently taking
Medicine the patient is allergic to

output: recommended prescription

```
Let δ be the flag for adverse relationship of drug-pair
Let θ be the threshold of similarity of drug class of drug-pair
Initialise δ to false and θ to false
while there are more medicine in D^p do
    get drug ID drugID in prescription set
    create default candidate set that belongs to same class as drugID
    set recommended medicine d^r be medicine from prescription set
    while δ is false or θ is false do
        while there are more medicine that patient is taking do
            if adverse relationship exists between d^r and medicine that patient is taking then
                δ = false;
                break from loop;
            end
        end
        while there are more medicine the patient is allergic to do
            if d^r belongs to same class as medicine that patient is allergic to then
                θ = false;
                break from loop
            end
        end
        if (δ == false ) or (θ == false) then
            get d^r from next item in candidate set
        end
        if there are no more items in candidate set then
            break and exit from testing for δ,θ
        end
    end
    display recommended medicine
end
```

that fits both criteria is *troleandomycin*. On receiving the suggestion of this alternative drug, the dentist can then decide whether it is appropriate to be prescribed after further consideration of the duration and dosage of the patient's current medications.

VII. Conclusion

This paper presents the use of a network model in the data-mining of bio-medical data within a three-tier framework, allowing easy updating of the knowledge base and effective presentation of results to the user. Experimental results show that this model performs better when only nodes within the direct neighbourhood of the drug-pair are used to find the common paths within the network. Besides the network model, we plan to compare and contrast this approach with the word embedding approach, where feature vectors are based on the nearest terms to the drug-pair. Not only will this study aid implementation of a personalised CDSS to benefit dentists in the prescription of drugs at point-of-care, our work will also provide potential for more efficient algorithms with the use of a network model in information retrieval, similarity search and machine learning.

Acknowledgments

This research has received support from Glory Dental Surgery Pte Ltd, Singapore (*http://glory.sg*) and was undertaken collaboratively with their panel of dentists.

References

[1] Serkan Ayvaz, John Horn, Oktie Hassanzadeh, Qian Zhu, Johann Stan, Nicholas P. Tatonetti, Santiago Vilar, Mathias Brochhausen, Matthias Samwald, Majid Rastegar-Mojarad, Michel Dumontier, and Richard D. Boyce. Toward a complete dataset of drug-drug interaction information from publicly available sources. *Biomedical Informatics*, 55:206–217, 2015.

[2] Glen Jeh and Jennifer Widom. Simrank: A measure of structural-context similarity. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 538–543, New York, NY, USA, 2002. ACM.

[3] Mohammed Khalilfa. Clinical decision support: Strategies for success. *Procedia Computer Science*, 37:422–427, 2014.

[4] Jeong Ah Kim, Insook Cho, and NamSun Kim. CDSS Service Integration Reference Model. *Advanced Science and Technology Letters*, 24:9–12, 2013.

[5] Vijay Kumar Mago, Nitin Bhatia, Ajay Bhatia, and Anjali Mago. Clinical decision support system for dental treatment. *Journal of Computational Science*, 3(5):254–261, 2012.

[6] Seon Gyu Park, Sungin Lee, Myeng-Ki Kim, and Hong-Gee Kim. Shared decision support system on dental restoration. *Expert Systems with Applications*, 39(14):11775–11781, 2012.

[7] A Pozzi and L Gallelli. Pain management for dentists: the role of ibuprofen. *Annali di Stomatologia*, 2(3-4 Suppl):3–24, 2011.

[8] Khare R, Wei CH, and Lu Z. Automatic extraction of drug indications from fda drug labels. *AMIA Annual Symposium Proceedings*, pages 787–794, 2014.

[9] Chuan Shi, Zhiqiang Zhang, Ping Luo, Philip S. Yu, Yading Yue, and Bin Wu. Semantic path based personalized recommendation on weighted heterogeneous information networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 453–462, New York, NY, USA, 2015. ACM.

[10] Kathie Suda, Heather Henschel, and Ursula Patel. Use of antibiotic prophylaxis for tooth extractions, dental implants, and periodontal surgical procedures. *Open Forum Infectious Diseases*, 5(1):250–254, 2018.

[11] L Tari, S Anwar, S Liang, J, and C Baral. Discovering drugdrug interactions: a text-mining and reasoning approach based on properties of drug metabolism. *Bioinformatics*, 26(18):547–553, 2010.

[12] Aileen P. Wright, Adam T. Wright, Allison B. McCoy, and Dean F. Sittig. The use of sequential pattern mining to predict next prescribed medications. *Journal of Biomedical Informatics*, 53:73–80, 2015.

[13] S. Yan, X. Jiang, and Y. Chen. Text mining driven drug-drug interaction detection. *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pages 349–354, 2013.