Personalised Drug Prescription for Dental Clinics Using Word Embedding

Wee Pheng Goh[1,2], Xiaohui Tao[1], Ji Zhang[1], Jianming Yong[1], XueLing Oh[2], and Elizabeth Zhixin Goh[2,3]

[1] University of Southern Queensland, Australia
{weepheng.goh,xtao,ji.zhang,jianming.yong}@usq.edu.au
[2] Glory Dental Surgery Pte Ltd, Singapore
xueling.oh@glorydental.com
[3] University Of Queensland, Australia
elizabeth.goh@uqconnect.edu.au

Abstract. The number of drugs in drug databases is constantly expanding with novel drugs appearing on the market each year. A dentist cannot be expected to recall all the drugs available, let alone potential drug-drug interactions (DDI). This can be problematic when dispensing drugs to patients especially those with multiple medical conditions who often take a multiple medications. Any new medication prescribed must be checked against the patient's medical history, in order to avoid drug allergies and reduce the risk of adverse reactions. Current drug databases allowing the dentist to check for DDI are limited by the lack of integration of the patient's medical profile with the drug to be prescribed. Hence, this paper introduces a software which predicts the possible DDI of a new medication against the patient's medical profile, based on previous findings that associate similarity ratio with DDI. This system is based conceptually on a three-tier framework consisting of a knowledge layer, prediction layer and presentation layer. The novel approach of this system in applying feature vectors for drug prescription will be demonstrated during the conference (http://r.glory.sg/main.php). By engaging with the interactive demonstration, participants will gain first-hand experience in the process from research idea to implementation. Future work includes the extension of use from dental to medical institutions, and it is currently being enhanced to serve as a training tool for medical students.

Keywords: Feature vector · Similarity ratio · Drug interaction.

1 Introduction

The increasing number of patients taking multiple drugs for a multitude of medical problems emphasis-es the importance for dentists to take precautions during drug prescription. Remembering all the potential DDI can be a significant burden [6], hence our decision support tool aims to assist the prescription process through integration with the patient's medical profile and DDI prediction at point-of-care in order to reduce prescription error and subsequent adverse drug reactions. A common cause of hospital admissions worldwide is adverse drug reactions, with incidence reaching 24% [7]. Many such admissions could have been avoided if more care was taken in drug prescription, such as by considering the patient's drug allergies. The design of the demonstration system which predicts a drug-pair's DDI potential is based on previous findings that the more similar the drug-pair, the higher the similarity ratio. The drug-pair's similarity ratio is obtained from word embedding associated with the neighbouring terms of each drug in the drug-pair. This prediction process within the prediction layer combines with a knowledge layer and presentation layer to form a three-tier conceptual framework.

The rest of the paper is as follows: Section 2 outlines the related work in DDI and shows how our model differs in the way the drug-drug relationship is detected and utilised. Section 3 discusses the previous findings in terms of the conceptual framework and how word embedding is used to determine a drug-pair's similarity ratio to determine if it is appropriate for prescription. Section 4 then illustrates use of the demonstration system in the dental setting to assist the dentist in safe drug prescription. The demonstration plan of the system is outlined in Section 5 with additional functions to the system being reported in Section 6.

1.1 Why is such a system needed

There is increasing use of technology in healthcare to store and retrieve patient records, and other applications such as treatment planning and drug prescription should also be considered. A system is required not just to advise if two drugs adversely react with one another, but whether a drug to be prescribed is safe for a particular patient. In other words, the system should be integrated with the medications the patient is currently taking and their drug allergies. The statutory requirements of having electronic treatment plans, electronic medical records, and the daily reporting of diagnoses and drugs dispensed compound the need for such a decision support system.

1.2 Features of the demonstration system

Participants during the conference will have first-hand opportunity to experience the novel features of the system. This includes:

- Predicting if a drug is safe for prescription
- Dynamic presentation of safe drugs for prescription in relation to the patient's medications and drug allergies by varying the threshold value
- Security features assigning different levels of privilege to different user categories
- Standardised clinical terminology such as the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) employed for the drug codes to ensure portability with existing systems and adherence to international clinical standards

2 Related Work

A common technique to extract drug information from bio-medical corpus is to extract feature vectors and build a predictive model [2] to determine the adverse relationship of the drug-pair. By identifying neutral candidates, negation cues and scopes from bio-medical text, a similar work to determine the confidence level of a drug interaction was also carried out by Bokh *et.al* [1]. A recent approach by Sahu and Anand based on long short-term-memory claimed to outperform traditional methods that explicitly relies on feature engineering [9].

The use of neural networks in extracting information on drug interaction has become a recent trend. By using word embedding technique to build word vectors, neural network is employed to learn the context features for extracting information on drug interaction [11].In a recent work, contextual features were obtained and used to extract drug information from bio-medical literature. This was made possible from semantic embeddings and position embeddings employed within a neural network [10]

Although these studies use data mining methods to extract relevant information to predict DDI, they do not take into account the drug profile of the patient.

Although there are a wide variety of decision support tools to assist the dentist in drug prescription, these tools only provide information on whether a drug-pair is in an adverse relationship. Even with Micromedex [1], one of the common packages which obtained the highest scores in terms of completeness and consistency in an evaluation of drug resources for drug-ethanol and drug-tobacco interactions, there is no integration with the medical profile of the patient, not to mention the availability of daily reports for the patients which are commonly needed by the clinics for effective patient care and inventory control.

Such tools are only as helpful as they claimed to be provided that the search terms are entered correctly. In addition, most of these systems do not take into account the allergy that the patient has, not to mention that the system is not integrated with the patient's medical profile.

The crucial need to integrate the patient's medical profile with the knowledge obtained from data mining motivated us to embark on this study. Although our system is similar to that proposed by [3] in terms of using information from the patient, the unique approach adopted in this paper goes one step further in using such information to support the decision-making process for the dentist at point-of-care within the clinical work-flow. Moreover, the word embedding method is also adopted which uses features that relate the similarity of a drug-pair in terms of how closely the words are related to each drug of the drug-pair. This approach distinguishes from our earlier work where feature vectors were constructed based on term similarities within the drug corpus [5]. In the application that is demonstrated here, the prescription of the drug will take into account the current drug that the patient is taking and the drug that the patient is allergic to. The framework of the customised system will be explained in the next section.

3 Research Findings

In view of the need for the professional user to dispense the drug efficiently and effectively, an approach is needed to ensure that drugs prescribed are not in an adverse relationship with the drugs that patient is currently taking. At the same time, it must not belong to the same class of drugs which the patient is allergic to.

To solve this problem, we have discovered a novel approach in predicting the probability of an interaction of a drug-pair. Based on this approach, a system has also been built and ready for deployment in a dental clinic, which will be described in the demonstration plan in this paper. This section describes the conceptual framework behind the deployment model and the details of the prediction layer of the conceptual framework.

3.1 Conceptual framework

The conceptual model behind such a system is based on the innovative 3-tier framework shown in Figure 1.

Knowledge Layer The knowledge layer consists of the bio-medical text which described the properties of the drugs. It comprises the domain knowledge in raw data form. Drugbank is used in this study where relevant information is extracted to construct a new taxonomy relating to interactions and side-effects. With this knowledge base, information is enhanced from the raw data available at Drugbank. As this is being constructed and made available in a structured format, different models can be developed and used to determine and predict if the drug is in an adverse relationship with another drug. Features relating to each drug like their potential side-effects can also be easily retrieved.

Prediction layer From the drug taxonomy, text mining was conducted to extract relevant information, where the text for each drug was extracted, cleaned and stored in order to provide information on the underlying properties of a drug-pair, enabling the similarity of the drug-pair to be computed. The flexibility

[1] https://www.micromedexsolutions.com

and robustness of the three-tier framework allowed the calculation of drug-pair similarity to proceed with various approaches. Many approaches can be used within this layer but for this paper, the word embedding approach is described and demonstrated.

This word embedding approach uses word vectors to compute the feature vectors which are then used to predict the similarity of a drug-pair. With such information, it can be readily applied to a decision support system to assist dentist in their drug-prescription at point-of-care.

In order to compute feature vectors, a trained model of embedded vectors is needed. One of the tools which has the text corpus trained is the Gensim implementation of word2Vec[2] where similar words can be efficiently obtained. Given a target word, Word2Vec predicts the occurrence of context words neighbouring the target word.

Suppose v_t is a target embedding vector for the target word t and v_c is a context embedding vector for the context word c, then P, the conditional probability in the neural probabilistic language model can be defined as:

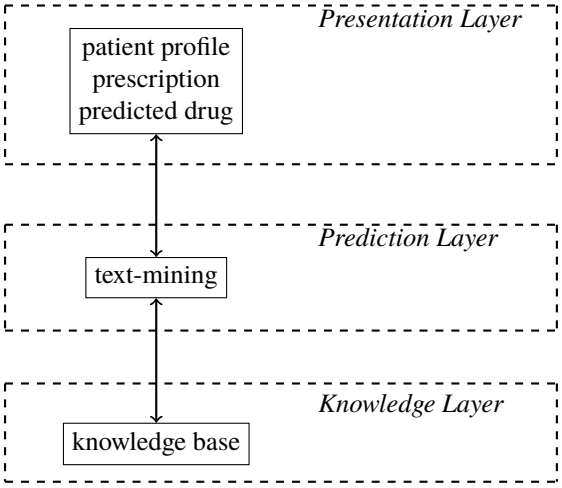

Fig. 1: Three-tier conceptual framework

$$P(c|t) = \frac{exp(v_c^T v_t)}{\sum_{i=1}^{k} exp(v_c^T v_{t_i})} \tag{1}$$

where k is the size of the vocabulary of the corpus.

The aim in the training of the corpus is then to maximise J, the log function of P. Hence from equation (1),

$$J = logP(c|t) = logexp(v_c^T v_t) - log \sum_{i=1}^{k} exp(v_c^T v_{t_i}) \tag{2}$$

Once the text corpus has been trained by word2Vec, the output vector for any name of a drug can be conveniently obtained through built-in Java methods included in word2Vec.

Presentation layer This layer is important as it serves as an interface between the prediction layer and the user. A well-designed user-friendly interface will help dentists adopt such a system in their clinical work-flow. As highlighted in Table 1 for the three layers in the framework, user requirements in the user layer need to be efficiently mapped onto the prediction layer to enable useful and relevant information to be extracted for further computing of the similarity ratio.

Presentation Layer	**Prediction Layer**	**Knowledge Layer**
• Efficient mapping of user requirements • User friendly interface	• Efficient choice of programming approach • Implementation of data mining • Algorithm design	• Bio medical data sources, drug taxonomy • Drug properties

Table 1: Features of conceptual framework

[2] https://radimrehurek.com/gensim/models/word2vec.html

The user layer also distinguish our system from many other systems as it contains personalized information of the patient.

Besides, the presentation layer also presents the results from the prediction layer. Hence this layer is important as a supporting tool to the dentist in deciding whether the drug to be prescribed is safe for the patient.

The drug which the dentist is going to prescribe is also stored in this layer. Such information is needed in the predictive layer for extraction of feature vectors. In order to maintain user-friendliness, which is crucial for clinical adoption of the system,it is important for this layer to present the results such that the user can easily understand.

Based on the result transmitted from the predictive layer, the service at this layer will then advise the user if the drug to be prescribed is safe for prescription. This approach allows the presentation layer to crystallize the results in a meaningful and friendly manner. More details on this layer is described in demonstration plan.

The drug taxonomy used in this research is taken from DrugBank, an open source bio-medical corpus. DrugBank is used just for convenience, though any other sources of corpus can be used. In order to build the model for subsequent knowledge discovery and decision making pertaining to drug prescription, a knowledge base has to be built.

Many methods exists to predict the interaction of a drug pair. To make use of the plethora of drug information within the bio-medical domain, textual data related to drug interactions are used in this demonstration. It has been found that the more similar the text between the pair of drugs, the higher the probability that the drug pair is similar. The novel approach in using textual data for computing similarity ratio is explained in this section. A survey of existing packages in drug interaction is also described.

3.2 Evolution of discovered drug interactions

Along with the conceptual framework, the research findings on the approach that is used in predicting the similarity is proposed and developed in this section.

Based on research findings that similar drugs has similar patterns from word embeddings associated with the textual description of the drugs [4], a prototype can be constructed which aligns with the presentation layer of the conceptual framework.

Once the data is transformed from the unstructured textual data to structured patterns, data mining is performed to extract appropriate information. The remaining stages are evaluation, visualisation and decision making based on the knowledge obtained. These stages can be iterative where the output of one stage may indicate the need for the previous stage to be refined. For example in the evaluation stage, if the results are not satisfactory, the process of transforming the data or data mining methods can be amended.

This is achieved from data mining and evaluation which aims to discover patterns and meanings from the knowledge base. These two stages leads to the knowledge needed for the user to decide if a drug is safe to be prescribed to the patient.

Research work has been done to observe the relationship between drug interactions and word embeddings from the textual data that describe the drugs [4, 5]. As cited by Nguyen (2019) from linguist JR Firth that "you shall know a word by the company it keeps" [8], related words and hence similar drugs can be known by finding similar words that describe the drugs. From word embeddings which is a vector representation of the words, similarity ratio of each drug can be obtained for comparison.

Within the predictive layer, a model is developed to compute the feature vectors from the textual description of the drug. These feature vectors are an indication of the probability of an adverse interaction

of a drug-pair. With such information, it can be readily applied to various applications like the deployment of the drug prescription system proposed in this paper.

With the use of word2Vec, vectors related to a given drug can be obtained. Given the word embedding obtained for each neighbouring terms of the drugs d_i and d_j, feature vectors $\overrightarrow{f_i}$ and $\overrightarrow{f_j}$ can be obtained as follows:

$$\overrightarrow{f_i} = \left[w_{i1}, w_{i2} \ldots w_{ik}\right] \tag{3}$$

$$\overrightarrow{f_j} = \left[w_{j1}, w_{j2} \ldots w_{jk}\right] \tag{4}$$

where w_{ik} and w_{jk} are the respective word embeddings of the neighboring words of drugs d_i and d_j and k is the number of neighboring words for each drug in the drug pair. It also follows that $|\overrightarrow{f_i}|$ and $|\overrightarrow{f_j}|$ is k.

The similarity ratios $sim(d_i, d_j)$ for the drug pair d_i and d_j can then be computed as:

$$sim(d_i, d_j) = \frac{\sum_{p=1}^{k} w_{ip} \times w_{jp}}{\sqrt{\sum_{p=1}^{k} w_{ip}^2} \times \sqrt{\sum_{p=1}^{k} w_{jp}^2}} \tag{5}$$

Word2vec is used to obtain the vectors as it is versatile in predicting related words.

As an illustration, Figure 2 shows a plot where related words like *painkillers* and *Tylenol* are clustered together. Similar entities like *Android* and *Note 5* are displayed towards the right.

While word2Vec is not strictly a deep neural network, the output vector that it produces in numerical format within the deep learning models can be easily understood by other deep networks making it very suitable for use in such works.

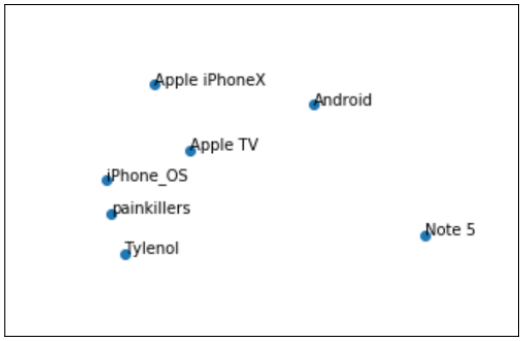

Fig. 2: Grouping of similar words

4 Use Case Scenario

This section illustrates how the prototype can be used at point-of-care to advise if the drug prescribed by the dentist is safe for the patient, and if not, to suggest alternatives, taking into account the drug profile of the patients.

4.1 Patient Registration

Once the user is signed into the system, the user can either register a new patient into the system or search for the record of an existing patient. The drug profile of the patient is shown beneath the patient's demographic record. The user can either edit them or click on the **Prescribe Drug** button to begin prescribing a drug for the patient.

4.2 Drug Prescription

At the screen on prescribing drug (Figure 3), the user may enter any part of the name of the drug and the system will display all the drugs with names that contains the pattern of the string of words entered by the user. For example, if user enter *rfar*, the drug *warfarin* will appear.

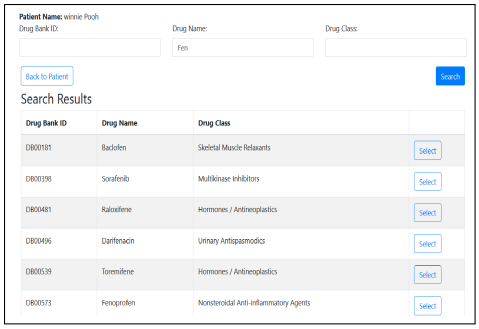

Fig. 3: Prescribing drug

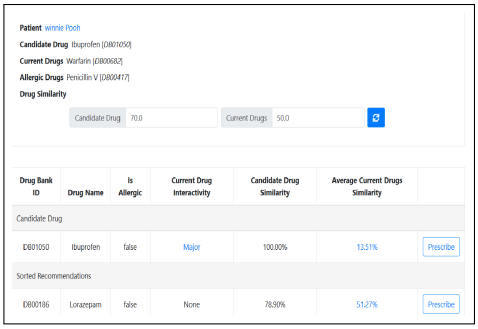

Fig. 4: Alternative drugs for selection

Figure 4 shows the list of alternative drugs for the user's consideration since the prescribed drug *Ibuprofen* interacts with the existing drug *Warfarin* which patient is currently taking. The size of the list of drugs varies depending on the threshold set by the user. In this case, the suggested candidate drug is at least 70% similar to *Ibuprofen*. There is also at least a 50% chance that the drug is similar to the current drug that patient is taking.

5 Demonstration Plan

The use case scenario described in Section 4 will be demonstrated interactively on the desktop with the aim of illustrating how the prototype can be used at point-of-care to advise if the drug prescribed by the dentist is safe for the patient, and if not, to suggest alternatives, taking into account the drug profile of the patients. The audience will be able to get first hand experience on the way the system is designed and implemented based on the research findings that similar drugs have similar feature vectors.

The demonstration plan is outlined below:

- Illustrate the versatility of word2Vec in relation to the ability to find similar words.
- Introduce the research problem and explain the main motivation behind this study and the creation of the system. The framework of the system is also presented to the conference participants.
- Demonstrate on the use of the system with a scenario of how a dentist can utilise this to prescribe a painkiller to a patient using the patient's medical profile. Next, a search on a typical painkiller will be conducted with the system displaying a list of recommended drugs.
- Show the various features of the system, including the daily reporting system as well as the security features of the system. The daily reporting system allows the user to generate a daily report on drugs that are dispensed for the day, for good clinical record keeping.

From the demonstration and participation in the interactive sessions, conference attendees will be engaged with first hand experience and appreciate the way research findings related to word embedding can be deployed for practical purposes.

6 Conclusions

This paper presents a novel approach in advising the suitability of a drug for prescription by predicting the similarity of a drug-pair and in practical terms, integrating this with the patient's medical status by considering their drug allergies to avoid allergic reactions, and the drugs they are currently taking to avoid adverse DDI. The demonstration system's algorithm is based on our previous findings that the similarity ratio of a drug-pair is an indicator of the probability of an adverse DDI within the drug-pair. The use of such a prototype can be extended from the dental clinic to medical institutions. The system is currently being enhanced so that it can be used on mobile devices, not only as a decision support tool, but also as a training tool for medical students.

Acknowledgment

This paper is undertaken collaboratively with the panel of dentists from Glory Dental Surgery Pte Ltd. The authors would like to thank the panel of dentists for their technical contribution and enriching the author's understanding of the requirements in drug dispensing. Feedback from the dental clinic is taken into consideration so that the deployment of the system aligns with the clinical work flow of a typical dental clinic.

References

1. Bokharaeian, B., Diaz, A., Chitsaz, H.: Enhancing extraction of drug-drug interaction from literature using neutral candidates, negation, and clause dependency. PLoS ONE **11**(10), 1–20 (10 2016)
2. Bui, Q., Sloot, P., vanMulligen, E., Kors, J.: A novel feature-based approach to extract drug-drug interactions from biomedical text. BioInformatics **30**(23), 3365–3371 (2014)
3. Casillas, A., Perez, A., Oronoz, M., Gojenola, K., Santiso, S.: Learning to extract adverse drug reaction events from electronic health records in spanish. Expert Systems with Applications **61**, 235–245 (2016)
4. Goh, W.P., Tao, X., Zhang, J., Yong, J., Qin, Y., Goh, E.Z., Hu, A.: Exploring the use of a network model in drug prescription support for dental clinics. In: 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing. pp. 168 – 172 (2018)
5. Goh, W., Tao, X., Zhang, J., Yong, J.: Mining drug properties for decision support in dental clinics. In: Kim, J., Shim, K., Cao, L. (eds.) PAKDD 2017:Advances in Knowledge Discovery and Data Mining. pp. 375 – 387. Springer International Publishing (2017)
6. Kheshti, R., Aalipour, M., Namazi, S.: A comparison of five common drug-drug interaction software programs regarding accuracy and comprehensiveness. Journal of research in pharmacy practice **5**, 257 – 263 (2016)
7. Lieber, N.S.R., Ribeiro, E.: Adverse drug reactions leading children to the Emergency Department. Revista Brasileira de Epidemiologia **15**, 265–274 (2012)
8. Nguyen, T., Le, N., Ho, Q., Phan, D., Ou, Y.: Using word embedding technique to efficiently represent protein sequences for identifying substrate specificities of transporters. Analytical Biochemistry **577**, 73 – 81 (2019)
9. Sahu, S.K., Anand, A.: Drug-drug interaction extraction from biomedical texts using long short-term memory network. Journal of Biomedical Informatics **86**, 15 – 24 (2018)
10. Sun, X., Dong, K., Ma, L.: Drug-drug interaction extraction via recurrent hybrid convolutional neural networks with an improved focal loss. Journal of Biomedical Informatics **21**, 37 (2019)
11. Victor Sudrez-Paniagua, Isabel Segura-Bedmar, P.M.: Exploring convolutional neural networks for drugdrug interaction extraction. Database **2017**, bax019 (2017)