

# Constructing Drug Ingredient Interaction Network to Ensure Medication Security

Zhiren Mao, Pingyi Zhou

State Key Lab of Software Engineering  
Computer School, Wuhan University  
Wuhan, China

chyanmao@whu.edu.cn  
zhou\_pinyi@whu.edu.cn

Jiaxiang Zhong, LuoJia Jiang

Hubei Shuiguohu High School  
Hubei Wuchang Experimental High  
School

Wuhan, China  
zjx001202@gmail.com  
zw87652016@gmail.com

Chongzhi Deng, \*Jin Liu

Wuhan No.6 High School  
Computer School, Wuhan University  
Wuhan, China

dczdcz2001@yahoo.com  
\*Corresponding author  
jinliu@whu.edu.cn

**Abstract**—With the fast development of big data, we can do better in modern digital health. The rational use of drugs is main threat of medication security. With big data mining, we analyze the drug instructions to automatically detect the adverse drug interaction in drug combination. The proposed method synthetically employs the NLP and complex network to automatically construct drug ingredient interaction network for mining and inference adverse drug interaction in drug combination. First, many drug instructions are collected by crawler from professional medicine websites. Next, NLP is utilized to automatically extract effective drug ingredient for building drug ingredient library. Then, drug ingredient interactions are extracted from drug instructions. The last, build drug ingredient interaction network. Because the drug ingredient interaction network is kind of complex, the principle of complex network is used to analyze the drug ingredient interaction network. Then we automatically generate a report to make knowledge visualize. The constructed drug ingredient interaction network is verified in experiment. The experiment results indicate that the validity and effectiveness of our proposed method.

**Keywords**—Medication security; Big data mining; Complex network; Drug ingredient interaction; Knowledge visualize

## I. INTRODUCTION

With the fast development of big data, we can do better in health [1] [2] [3]. According to the World Health Organization (WHO), more than 50% of drugs are prescribed and deployed in unreasonable ways all over the world. Meanwhile, more than 50% patients use drugs in the wrong way. A third of death in patients is attributed to the drug abuse and the irrational use of drug. In American, the irrational use of drugs has been the fourth reason that cause the death and more than 10000 fatal adverse drug events were reported. The problem is more serious in china.

To illustrate this problem, we give a common application scenario. When a person is sick, prescription drugs and the over-the-counter (OTC) drugs can be get easily from online stores (e.g. taobao, JD, etc.). Since potential adverse drug reactions are difficult to recognize by the users themselves. These online platform should generates usage report for its users. For an ordinary user, potential adverse drug reactions are difficult to recognize through the drug instruction. The knowledge from these drug instruction should be extracted and shown to users in an easy way.

In this paper, we mainly focus on how to find irrational drug

combination. Drug is made up of auxiliary components and effective components [4]. When two or more drugs are used commonly, effective components may cause changes of pesticide effect which include synergistic effect, antagonism, etc [4] [5]. Reasonable drug interactions can enhance curative effect or reduce the adverse drug reactions. Unreasonable drug interaction in a prescription can reduce curative effect or increase toxicity [6]. So, we construct drug ingredient interaction network to discover and inference irrational drug combination.

This paper firstly utilizes nature language processing (NLP) technology to extract and recognize the drug effective ingredient and the drug ingredient interaction. Next, these drug ingredient interactions are well grouped and marked in type codes. Then build the drug ingredient interaction network. Last, the networks is used to discover and infer adverse drug reactions and the medical report is generated for users.

ADR-Miner approach contains several steps. First, preprocess these crawled drug instructions data into well-structured format. Second step contains two functions: the first function is to recognize and extract drug's effective ingredient and the second function is to recognize and extract drug's ingredient interaction from drug introductions. The third step groups and codes these extracted drug's ingredient interactions. The fourth step constructs the drug ingredient interaction network. The fifth step is to discover and infer adverse drug reactions based on given drug combination and the drug ingredient interaction network. Last, we generate medical report in an easy way for users to understand.

In short, the main contributions of this paper are summarized as follow:

- We formulate a problem that aims to discover the adverse drug reactions from the drug ingredient interaction network for users to avoid irrational drug combination.
- We present an ADR-Miner as a novel analytic framework to tackle this problem, which includes the recognition and extraction of the drug effective ingredient and the drug ingredient interaction and the method to discover and infer adverse drug reaction (ADR).
- We evaluate ADR-Miner based on some irrational drug combination case, in which result shows DI-Miner is effective and promising.

The rest of this paper is organized as follow: Section 2 discusses the related work. Section 3 formulates problems. Section 4 covers the technical details of our framework. Section 5 evaluate the performance of our approach. Section 7 concludes.

## II. MINING AND ANALYZING DRUG DATA

Our work is related to the studies which focus on mining and analyzing drug information in other kinds of information platform (e.g. social media twitter [7], adverse event reporting system (AERS) [8] [9], clinical data [10], professional open data base [11] [12] [13]). However, their methods cannot be directly applied to our problem. In some literatures, there is a lot of work that apply semantic text analysis in different data source [10] [11][14]. At present, all these work are just on drug-to-drug level. This paper firstly utilizes the ingredient-to-ingredient level relations to discover adverse drug reactions.

## III. PROBLEM FORMULATION

In this paper, we formally formulate it as a research problem.

A drug introduction includes an attribute set: the drug name (dn), rating, effective components (ec), known drug interactions (da) and other information. Without loss of generality, in this paper, we choose  $d = \{dn, ec, da\}$ , since we can extract informative information from the common attributes. We mainly focus on the drug name (dn), effective components (ec) and known drug interactions (da). Known drug interactions are described in some sentences sequence  $S = \{s_1, s_2, \dots, s_m\}$ . Through semantic analysis technology, each sentence  $s_i$  describing the known drug interactions is processed into a list which includes elements  $adr = \langle ec_i, ec_j, t_k \rangle$  ( $i \neq j, t_k$  is the drug ingredient interaction type).  $adr = \langle ec_i, ec_j, t_k \rangle$  denotes the  $t_k$  type drug ingredient interaction between effective component  $ec_i$  and  $ec_j$  effective component.

In our work, we consider five problems, defined as follow:

**Problem 1.** How to construct a drug effective ingredient lib  $EC = \{ec_1, ec_2, ec_3, \dots, ec_m\}$  from all crawled drug instructions data  $D = \{d_1, d_2, d_3, \dots, d_n\}$ .

**Problem 2.** How to construct a drug ingredient interaction list  $ADR = \{adr_1, adr_2, adr_3, \dots, adr_l\}$  from all known drug interactions attributes.

**Problem 3.** How much categories these recognized drug ingredient interaction  $ADR = \{adr_1, adr_2, adr_3, \dots, adr_l\}$  can be grouped into.

**Problem 4.** How to construct the drug ingredient interaction network and whether the drug ingredient interaction network is complex network.

**Problem 5.** What kind of complex network the drug ingredient interaction network is and how to discover or infer adverse drug reactions based on drug combination  $\langle d_i, d_j \rangle$  ( $i \neq j$ ) and the drug ingredient interaction network.

## IV. OUR FRAMEWORK

Section 1 shows the overview of ADR-Miner framework. In this section, we present each step of our framework in detail.

### A. Preprocessing and semantic analyzing

We firstly collect large amount of raw drug instructions from professional medicine website (e.g. <http://www.dxy.cn/>) by web crawler. Then, we collect more than 32000 drug introductions.

As mentioned in section 3, we choose  $d = \{\text{drug name (dn), effective components (ec), known drug interactions (da)}\}$  as research data. The known drug interactions (da) often consist of more than one sentences. We first split da into several sentence  $\{s_1, s_2, s_3, \dots, s_h\}$  in this paper.

The raw drug instruction should be converted into sentences. The attribute known drug interactions (da) is preprocessed into a list of sentences. For each da, we get  $da = \{s_1, s_2, s_3, \dots, s_h\}$ . In this step, each drug instruction is preprocessed into a tuple  $d = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ . If a drug includes more than two effective ingredient, we will not be sure which effective ingredient is involved in known drug interactions. We filter out about 13000 drug introductions without stating its effective ingredient or including more than two or more of the effective ingredient from more than 32000 drug introductions.

### B. Extracting

This subsection attempts to address problem 1 and problem 2. The preprocessing step generates a tuple set  $D = \{d_1, d_2, \dots, d_n\}$  ( $d_i = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ ). In this step, our goal is to realize two functions: the first function is to recognize and extract drug's effective ingredient from drug introductions and the second function is recognize and extract drug's ingredient interaction from drug introductions.

Chinese word segmentation system ICTCLAS2015 provided by Zhang Huaping (<http://ictclas.nlpir.org/>) help us divide a sentence  $s_i$  into a word sequence  $= \{w_1, w_2, \dots, w_g\}$ . Each word  $w_i$  in the sequence  $W$  is marked the part of speech [15]. We firstly extract drug's effective ingredient term to construct a drug's effective ingredient term lib for recognizing the effective ingredient in text written in natural language. For a tuple  $d = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ , we can easily extract to construct a term lib  $EC = \{ec_1, ec_2, ec_3, \dots, ec_m\}$ . Because many drugs contain the same effective ingredient and some drug introductions show they consist more than two effective ingredients that we cannot distinguish which ingredient participates the response of pharmacodynamics described in da. So, we extract 2747 kind of effective ingredient from more than 18000 tuple.

The detail effective component lib  $EC$  is added to the word lib of ICTCLAS system. All the terms in the lib  $EC$  are set to noun. For each tuple  $d = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ , we utilize the ICTCLAS system to divide sentence  $s_1$  into a word sentence. For example, the first sentence of da in the drug Benzylpenicillin Sodium for Injection is "Chloramphenicol, erythromycin, tetracycline, sulfa can interfere with the activity of this product". These pharmaceutical ingredients are recognized and converted into four sentence: Chloramphenicol can interfere with the activity of this product, erythromycin can interfere with the activity of this product, tetracycline can interfere with the activity of this product and sulfa can interfere with the activity of this product. So, for each  $d = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ , the da is updated by the method.

### C. Grouping

In this subsection, we attempt to solve problem 3. For a drug induction  $d_i = \{dn, ec, da = \{s_1, s_2, \dots, s_h\}\}$ , each sentence  $s_i$  in  $d_a$  is segmented and token by ICTCLAS system. Because of the specification of sentence  $s_i$ , we group all these sentences with tool WEKA [16]. By grouping, we summarize 43 kinds of drug ingredient interaction by analyzing the grouping result. All these types are coded. These types can be used as the edge of network in section 4.4 and the template to generate medical report in section 4.6. Fig 1 gives the proportion of each drug reaction types. So, for each sentence  $s_i$  in  $da$ , we generate a tuple  $adr = \langle ec_i, ec_j, t_k \rangle$  ( $i \neq j$ ,  $t_k$  is the drug ingredient interaction type). In the tuple  $adr$ ,  $ec_i$  is the  $ec$  attribute of  $d_i$  and  $ec_j$  is recognized from sentence  $s_i$ . By processing, we recognizes more than 11000 drug ingredient interaction  $ADR = \{adr_1, adr_2, adr_3, \dots, adr_l\}$ .

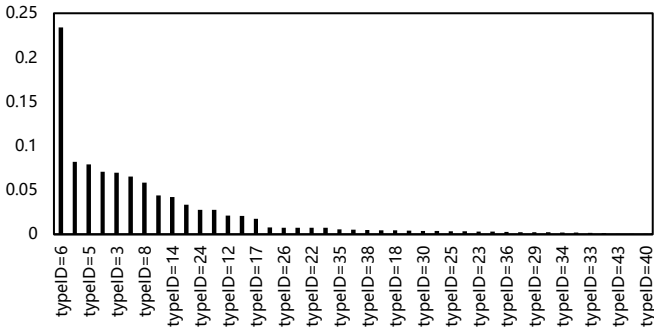


Fig. 1. The proportion of each adverse drug action types

### D. Constructing network

This subsection attempts to address problem 4. We get more than 2700 effective ingredients  $EC = \{ec_1, ec_2, \dots, ec_m\}$  in section 4.2 and recognized more than 11000 drug ingredient interaction  $ADR = \{adr_1, adr_2, \dots, adr_l\}$  in section 4.3. This section utilize these information to construct ingredient-to-ingredient level drug ingredient interaction network.

To describe the relations among effective ingredients, we defined a directed network  $DIN = (V, E)$ . Each effective ingredient  $ec_i$  in  $EC$  is regarded as a node  $v$  in network  $DIN$ . So,  $EC$  can be regarded as the node set  $V$  of  $DIN$ . For each  $adr = \langle ec_i, ec_j, t_k \rangle$ ,  $adr$  is converted to an edge  $e$  in which  $ec_i$  is the tail and  $ec_j$  is the head.  $t_k$  in  $adr$  denotes the type of the edge  $e$ . A matrix  $A$  is used to denote the network  $DIN$ . If there is a drug ingredient interaction between node  $v_i$  and  $v_j$ , the element  $a_{ij}$  of the matrix  $A$  denotes the type  $t_k$ . If not,  $a_{ij}$  is equal to 0.

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,N} \\ \dots & \dots & \dots \\ a_{N,1} & \dots & a_{N,N} \end{bmatrix} \quad (1)$$

Fig.2 shows the drug ingredient interaction network which includes more than 2700 nodes and more than 11000 edges. With complex networks theory we identify what kind of complex network the drug ingredient interaction network is and analyze some feature of the network. Some important index to evaluate a network are graph density [17], degree distribution [18], average path length [19], clustering coefficient [20] and etc.

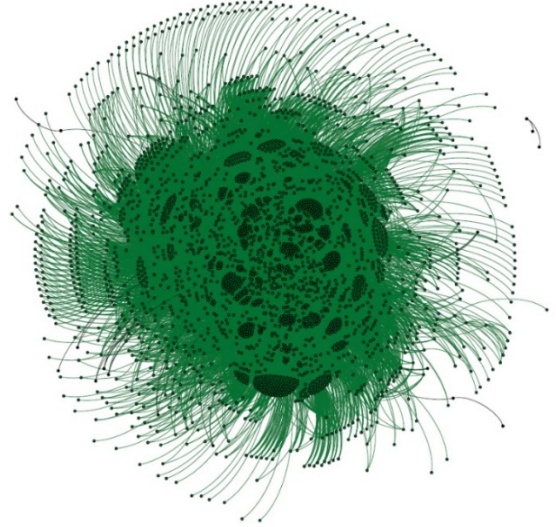


Fig. 2. The drug ingredient interaction network

Graph density  $GD$  is used to evaluate the closeness among nodes.  $GD$  is defined in equation (2). For a complete graph, the graph density of the complete graph is 1. By computing, the graph density  $GD$  of the network is 0.014. So, the drug ingredient interaction network is a sparse one. Because the known drug reaction came from report and drug test, there are a lot of unknown drug reactions in more than 98% the remaining space.

$$GD = \frac{|E|}{2 * N(N-1)} \quad (2)$$

The degree of a node  $v$  in  $DIN$  is the number of the edges incident to the node. The degree of a node  $v$  is denoted  $\text{deg}(v)$  or  $d(v)$ . The in-degree of a node  $v$  in  $DIN$  is the number of the edges in which the node is the head. The in-degree of a node is denoted  $d^-(v)$ . The out-degree of a node  $v$  is the tail, which is denoted  $d^+(v)$ . In the network, the distribution function  $P(k)$  is used to denote the number of nodes whose degree is equal to  $k$ .  $P(k)$  is defined in equation (3). Similarly, the in-distribution function  $P^-(k)$  and the out-distribution function  $P^+(k)$  is defined in equation (4) and (5).

$$P(k) = \sum_{i=1}^n \prod(\text{deg}(v_i) == k) \quad (3)$$

$$P^-(k) = \sum_{i=1}^n \prod(d^-(v_i) == k) \quad (4)$$

$$P^+(k) = \sum_{i=1}^n \prod(d^+(v_i) == k) \quad (5)$$

$\prod(x)$  is an indicator function. When the  $x$  is true, the value of the indicator function is 1. When false, it is 0.

Generally speaking, the degree of a node  $v$  measures the degree of important of  $v$  in  $DIN$ . The average degree  $\langle k \rangle$  is generated by a simple formula (6). By computing, the average degree  $\langle k \rangle$  is 85.121.

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N \text{deg}(v_i) \quad (6)$$

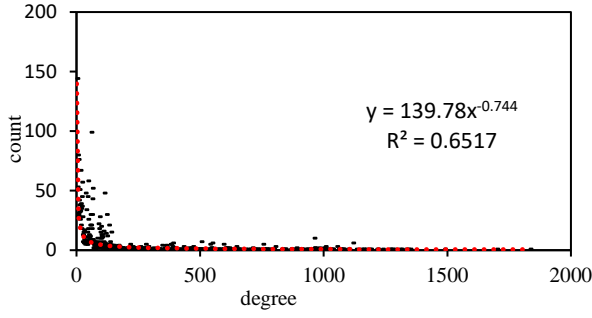


Fig. 3. The degree distribution of the network

Fig.3 shows the degree distribution of the network *DIN*. Fig.4 and fig. 5 show the in-degree and out-degree distribution of the network. We found the in-degree distribution contributing many low-degree value and out-degree distribution contributing many high-degree value. Because the more large degree means the more important role in the network. In the drug ingredient interaction network, most of the nodes have only a few links

#### In-Degree Distribution

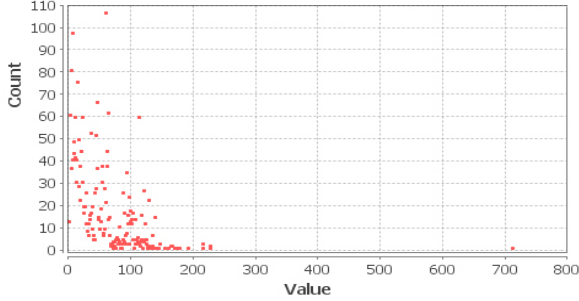


Fig. 4. The in-degree distribution of the network

while a few nodes have a large number of links to other nodes. By curve fitting, we get the distribution function  $P(k) = 139.78 \cdot k^{-0.744}$ . The multiple correlation coefficients  $R^2$  which is an index to evaluate the degree of linear correlation between the variables is 0.6517. Because the degree distribution is power law distribution, the drug ingredient interaction network is a scale-free network [21][22].

#### Out-Degree Distribution

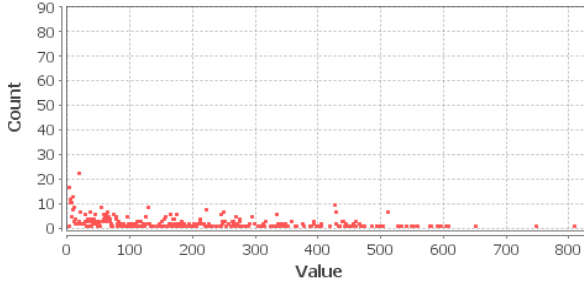


Fig. 5. The out-degree distribution of the network

Centrality is an important index to evaluate the relative importance of nodes of a network. There are many method to measure centrality. This paper utilizes eigenvector centrality. For each node  $v_i$ , a relative score  $x_i$  is assigned to the node. In the relative score of a node  $v_i$ , these links to high relative score nodes have larger contribution than these links to low relative score nodes. The formula of node relative score  $x_i$  is given as follow:

$$x_i = \frac{1}{\lambda} \sum_j^N \prod(a_{ij}) x_j \quad (7).$$

In the formula (7),  $N$  is the number of nodes and  $\lambda$  is a constant. All these nodes relative score in the network is regarded as a vectory  $x = \{x_1, x_2, \dots, x_N\}$ . So, the formula (7) is converted into formula (8). In the formula (8), each characteristic vector solution corresponds to different characteristic constant. Each node  $v_i$  in the network is assigned a relative score  $x_i$  by solving the formula (8). Fig.6 shows the

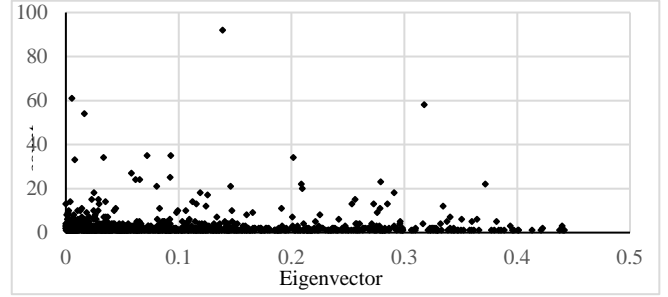


Fig. 6. The eigenvector centrality distribution of the network

eigenvector centrality distruction of the network. The the eigenvector centrality distruction is shown in equation (9).

$$Ax = \Phi x \quad (x = \{x_1, x_2, \dots, x_N\}, \Phi = \{\lambda_1, \lambda_2, \dots, \lambda_N\}) \quad (8)$$

$$EP(k) = \sum_{i=0}^N \prod(x_i == k) \quad (9)$$

In the network *DIN*, the distance between  $v_i$  and  $v_j$  is defined as the shortest path length of the two nodes and denoted as  $d_{ij}$ . If the node  $v_i$  and  $v_j$  is disconnected,  $dis_{ij} = \infty$ . The maximum distance of any two nodes in a network is the diameter of the network. The network diameter is get by the formula:

$$Dia = \max_{1 \leq i < j \leq N} dis_{ij} \quad (10)$$

The diameter of the network *DIN* is 6. The average distance of any two nodes in a network is the average path length of a network. The formal definition is given as equation (11). By computing, the average path length of the network *DIN* is 2.46. So, the network *DIN* has some characteristics of small-world network [23]. For a small-world network, the network not only has short average path distance, but also has large clustering coefficient.

$$L = \frac{1}{C_N^2} \sum_{1 \leq i < j \leq N} dis_{ij} \quad (11)$$

For a node  $v_i$  in a network  $G$ , there are  $\deg(v_i)$  edges connecting with other  $\deg(v_i)$  nodes. These  $\deg(v_i)$  nodes is called the neighbor of the node  $v_i$ . There are at most  $C_{\deg(v_i)}^2$  edges between these  $\deg(v_i)$  nodes. The clustering coefficient  $C_i$  of the node  $v_i$  is the ratio of the actual edges  $AE_i$  between these  $\deg(v_i)$  nodes to  $C_{\deg(v_i)}^2$ . Formally,

$$C_i = \frac{AE_i}{C_{\deg(v_i)}^2} \quad (12)$$

The clustering coefficient of the network is the average clustering coefficient of all node in the network. Formally,

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (13)$$

By computing, the average clustering coefficient of the network is 0.201.

Through the analysis of the above, we evaluate some characteristics of the network  $DIN$ . We find that the network  $DIN$  not only is a scale-free network but also a small-world network.

### E. Finding and inferencing

We construct the drug ingredient interaction network above. In this subsection, we attempt to solve problem 5. So, we give a method to automatically discover and infer adverse drug reactions based on given drug combination and the drug ingredient interaction network.

When a user purchase some drug in online store (e.g. taobao, JD, etc.). Because of the limited cognitive ability, some potential adverse drug reactions are very difficult to find by the user itself. A drug in an online store contains the drug's introduction. We can recognize the drug effective ingredients from the drug introduction. Based on this principle, we can recognize all these effective ingredients of these drugs which are in user shopping cart. These effective ingredients can have some adverse drug reactions. Utilize the constructed drug ingredient interaction network, these potential adverse drug reactions can be detected.

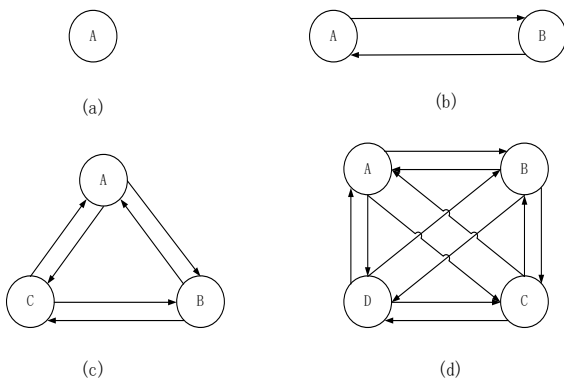


Fig. 7. The candidate set of potential adverse drug reactions

Fig. 7 shows the generating candidates principle. When the effective ingredient set  $Eu$  recognized from  $Du$  only contains a kind of effective ingredient, the fig.7 (a) shows the circumstance. In this circumstance, there is no potential adverse drug reactions. When the effective ingredient set  $Eu$  contains two kinds of effective ingredient, the fig.7 (b) shows the circumstance. In this circumstance, we first generate a candidate set =  $\{A \rightarrow B, B \rightarrow A\}$ . In the candidate set  $H$ , candidate element  $h_i$  is a binary sequence. For each candidate element  $h_i$  in the candidate set  $H$ , if  $h_i$  can be found in the drug ingredient interaction network, the candidate element  $h_i$  and corresponding drug ingredient interaction type  $t_k$  which can be used to generate re-port are add to the result set  $R$ . So, the element of  $R$  is two-tuple  $r = \langle h_i, t_k \rangle$ . Fig.7. (c) and fig.7. (d) Separately shows circumstance of  $|Eu| = 3$  and  $|Eu| = 4$ . If the effective ingredient set  $|Eu| = m$ ,

the number of the element of the candidate set  $H$  is equal to  $m(m-1)*2$ .

Given the ingredients term lib  $EC$ , the drug set  $Du$  and the drug ingredient interaction network  $DIN$ , a method to discover and infer adverse drug reactions is shown in algorithm 1.

---

#### Algorithm 1. Find and inference algorithm for detecting adverse drug reactions

---

Input:

$EC$  : the ingredients term lib.

$Du$  : the drug combination.

$DIN$  : the drug ingredient interaction network.

Output:

$R$  : these discovered adverse drug reactions.

1. For the drug set  $Du = \{d_1, \dots, d_n\}$ , we utilize the ingredients term lib  $EC$  to recognize all these effective ingredients  $Eu = \{e_1, \dots, e_m\}$  form the drug set  $Du = \{d_1, \dots, d_n\}$ .
  2. For these effective ingredients  $Eu = \{e_1, \dots, e_m\}$ , we generate the candidate set  $H = \{h_1, h_2, \dots, h_{m(m-1)*2}\}$ .
  3. For each element  $h_i$  in the candidate set  $H$ , we utilize the drug ingredient interaction network  $G$  to detect whether the element  $h_i$  is a kind of adverse drug reactions. If the element  $h_i$  is a kind of adverse drug reactions, a tuple  $\langle h_i, t_k \rangle$  is add to the result set  $R$ .
- 

### F. Generating interpretation and knowledge visualization

For a given drug combination  $Du$ , we utilize algorithm 1 to get a result set  $R$ . For the result set  $R$ , if the  $R$  is an empty set. It means that no adverse drug reactions is found in the given drug combination  $Du$ . If the  $R$  is not an empty set, for each element  $r = \langle h_i, t_k \rangle$  in the set  $R$ , we utilize the corresponding drug ingredient interaction type  $t_k$  to illustrate the relation in  $h_i = \langle e \rightarrow e \rangle$ . A report is generated for user to illustrate the result.

## V. EVALUATION

To evaluate if ADR-Miner can really help users to recognize adverse drug reactions, we conduct the evaluation. We first collect reported medical cases to form the test set. Then we introduce the metric used in our evaluation. Last, based on the test case and the metric, we evaluate the performance.

### A. Test set

In this subsection, we collect some common adverse drug reactions to form test set. Without loss of generality, a large number of reported adverse drug reaction cases are not selected as the test sample. Some common adverse drug reactions which have been recorded in the common drug incompatibility table are selected. We select 260 adverse drug reactions from the common drug incompatibility table which have been verified.

### B. Big data Performance analysis

In this section, we introduce the metric used in our evaluation. These metrics include precision, recall and F-measure which are defined as follow:

$$precision = \frac{TP}{TP + FP} \quad (14)$$

$$recall = \frac{TP}{TP + FN} \quad (15)$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \quad (16)$$

Where  $TP$ ,  $FP$ ,  $FN$  separately denote the number of the true positive, false positive, and false negative.

### C. Evaluation of framework performance

In this section, we utilize the test set constructed above to evaluate the performance of the framework. For all samples in the test set, every sample is the true adverse drug interaction sample. If a sample is recognized by the framework, it is a true positive (TP). If a sample is not recognized by the framework, it is a false negative (FN). Utilizing the metrics, the precision is 1, the recall is 0.51 and the F-measure is 0.68.

## VI. CONCLUSION

The adverse drug interaction is a serious threat for medicine security. This paper proposed DI-Miner framework for ordinary users to avoid adverse drug interactions. The approach employs NLP, big data mining, knowledge visualization, complex network technology to create a drug ingredient interaction network. The network is used to tell whether a drug combination has adverse drug interactions. The proposed method can not only eliminate the gap between the human's limited cognitive ability and the medicine knowledge but also make health service more intelligent. The experiment and analysis proved the validity and effectiveness of our approach.

As we all know, it is the first time that an ingredient-level method has been used to detect adverse drug interaction. And we are now planning for further expansion. First, since we haven't utilized all the information of drug introductions, the method only recognize generalization adverse drug interaction and the patients' physical information is not considered in our method. Second, a detecting report will be generated for users, the report does divide the danger level of recognized adverse drug reactions. We will explore an automatic method for dividing the danger level in our future work [24][25].

## ACKNOWLEDGMENT

Jin Liu is the corresponding author. We would like to thank Pingyi Zhou for his constructive comments which help a lot. This work is partly supported by National Natural Science Foundation of China (NSFC) (grant No. 61572374, U163620068, U1135005), the Academic Team Building Plan from Wuhan University and National Science Foundation (NSF) (grant No. DGE-1522883).

## REFERENCES

- [1] Xu Z, Yan Z, Mei L, et al. The Big Data Analysis of the Next Generation Video Surveillance System for Public Security[C]//Workshop on E-Business. Springer International Publishing, 2015: 171-175.
- [2] Luo X, Liu Y, Li Q. Big data - related technologies and applications[J]. Concurrency and Computation: Practice and Experience, 2016, 28(15): 4036-4037.
- [3] Xu Z, Mei L, Hu C, et al. The big data analytics and applications of the surveillance system using video structured description technology[J]. Cluster Computing, 2016, 19(3): 1283-1292.
- [4] Brown S D, Nativo P, Smith J A, et al. Gold nanoparticles for the improved anticancer drug delivery of the active component of oxaliplatin. Journal of the American Chemical Society, 2010, 132(13): 4678-4684.
- [5] Azizi M, Ménard J, Bissery A, et al. Pharmacologic demonstration of the synergistic effects of a combination of the renin inhibitor aliskiren and the AT1 receptor antagonist valsartan on the angiotensin II-renin feedback interruption. Journal of the American Society of Nephrology, 2004, 15(12): 3126-3133.
- [6] Drug-drug interactions. CRC Press, 2008.
- [7] Bian J, Topaloglu U, Yu F. Towards large-scale twitter mining for drug-related adverse events. Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM, 2012: 25-32.
- [8] Harpaz R, DuMouchel W, Shah N H, et al. Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. Clinical Pharmacology & Therapeutics, 2012, 91(6): 1010-1021.
- [9] Harpaz R, Chase H S, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. BMC bioinformatics, 2010, 11(9): 1.
- [10] Pathak J, Kiefer R C, Chute C G. Using linked data for mining drug-drug interactions in electronic health records. Studies in health technology and informatics, 2013, 192: 682.
- [11] Lindquist M, Ståhl M, Bate A, et al. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. Drug Safety, 2000, 23(6): 533-542.
- [12] Wishart D S, Knox C, Guo A C, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research, 2006, 34(suppl 1): D668-D672.
- [13] Wong C M, Ko Y, Chan A. Clinically significant drug-drug interactions between oral anticancer agents and nonanticancer agents: profiling and comparison of two drug compendia. Annals of Pharmacotherapy, 2008, 42(12): 1737-1748.
- [14] Xie F, Xu Z. Semantic based annotation for surveillance big data using domain knowledge[J]. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 2015, 9(1): 16-29.
- [15] Xia F. The part-of-speech tagging guidelines for the Penn Chinese Treebank (3.0). 2000.
- [16] Zhang K, Zan H, Chai Y, et al. Survey of the Chinese Function Word Usage Knowledge Base. Journal of Chinese Information Processing, 2015.
- [17] Scott J. Social network analysis. Sage, 2012.
- [18] Dorogovtsev S N, Mendes J F F, Samukhin A N. Size-dependent degree distribution of a scale-free growing network. Physical Review E, 2001, 63(6): 062101.
- [19] Fronczak A, Fronczak P, Hołyst J A. Average path length in random networks. Physical Review E, 2004, 70(5): 056110.
- [20] Zhou T, Yan G, Wang B H. Maximal planar networks with large clustering coefficient and power-law degree distribution. Physical Review E, 2005, 71(4): 046141.
- [21] Holme P, Kim B J. Growing scale-free networks with tunable clustering. Physical review E, 2002, 65(2): 026107.
- [22] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Physical review letters, 2001, 86(14): 3200.
- [23] Newman M E J, Watts D J. Renormalization group analysis of the small-world network model. Physics Letters A, 1999, 263(4): 341-346.
- [24] Liu Z, Wei C, Ma Y, et al. UCOR: an unequally clustering-based hierarchical opportunistic routing protocol for WSNs[C]//International Conference on Wireless Algorithms, Systems, and Applications. Springer Berlin Heidelberg, 2013: 175-185.
- [25] Liu Z, Niu X, Lin X, et al. A Task-Centric Cooperative Sensing Scheme for Mobile Crowdsourcing Systems[J]. Sensors, 2016, 16(5): 746.