

Mining Medical Causality for Diagnosis Assistance

Sendong Zhao, Bing Qin, Ting Liu
Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, Harbin, China
{sdzhao, bqin, tliu}@ir.hit.edu.cn

ABSTRACT

In this paper, we formally define the problem of mining medical domain causality for diagnosis assistance. We present four specific tasks to solve this problem, i.e., mining medical causality from text, medical treatment effectiveness measurement, disease prediction via causal representation and explicable medical treatment recommendation. Mining medical causality and medical treatment effectiveness measurement are fundamental problems with downstream applications of disease prediction and medical treatment recommendation. However, these problems have never been systematically studied before. There are existing studies on mining medical causality, effectiveness measurement and disease prediction, which both have their own drawbacks making them not suitable for our problem of mining medical domain causality for diagnosis assistance. We have studied the problem of mining medical causality and disease prediction. Specifically, we present a unified framework combining different resource and inference rules in a factor graph for causality mining, and present a symptom-disease network embedding method to solve the sparsity of symptom-disease links for disease prediction. Both methods are general method and can be easily applied to related domains.

CCS Concepts

•Social and professional topics → Medical records;
•Computing methodologies → Causal reasoning and diagnostics; •Applied computing → Health care information systems; Health informatics;

Keywords

Medical Causal Mining; Effectiveness Measurement; Disease Prediction; Medical Treatment Recommendation

1. INTRODUCTION

In the medical context, causal knowledge usually refers to causal relations between diseases and symptoms, living

habits and diseases, symptoms which get better and therapy, drugs and side-effects, etc [11]. All these causal relations are usually in medical literature, forum and clinical cases and compose the core part of medical diagnosis. Therefore, mining these causal knowledge to predict disease and recommend therapy is of great value for assisting patients and professionals.

The task of mining these causal knowledge for diagnosis assistance can be decomposed into four constitutes: (1) mining medical causality from text; (2) medical treatment effectiveness measurement; (3) disease prediction and (4) explicable medical treatment recommendation. However, these tasks have never been systematically studied before. For my PhD thesis, I plan to formally define the problem of mining medical domain causality for diagnosis assistance and propose methods to solve this problem.

1. Mining these textual causalities can be very useful for discovering new knowledge and making decisions. Many studies have been done for causal extraction from the text [1, 12, 13]. However, all these studies are based on pattern or causal triggers, which greatly limit their power to extract causality and rarely consider the frequency of co-occurrence and contextual semantic features. Besides, none of them take the transitivity rules of causality leading to reject those causalities which can be easily get by simple inference. Therefore, we formally define the task of mining causality via frequency of event co-occurrence, semantic distance between event pairs and transitivity rules of causality, and present a factor graph to combine these three resources for causality mining.

2. Treatment effectiveness analysis is usually taken as a subset of causal analysis on observational data. For such real observational data, PSM and RCM are two dominant methods. On one hand, it is usually difficult for PSM to find the matched cases due to the sparsity of symptom. On the other hand, we should check every possible (symptom, treatment) pair by exploiting RCM, leading to make the characteristic of exploding up, especially when we want to check the causal relation between a combination of symptoms and a combination of drugs. Besides, the larger number of symptom or treatment in the combination the less number of patient case retrieved, which lead to the lack of statistical significance. Specifically, patients tend to take tens of herbs as the treatment each time in Traditional Chinese Medicine (TCM). Therefore, how to evaluate the effectiveness of herbs separately and jointly is really a big challenge. This is also a very fundamental research topic supporting many downstream applications.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM 2017 February 06-10, 2017, Cambridge, United Kingdom

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4675-7/17/02.

DOI: <http://dx.doi.org/10.1145/3018661.3022752>

3. Both hospitals and on-line forums have accumulated sheer amount of records, such as clinical text data and on-line diagnosis Q&A pairs. The availability of such data in large volume enables automatic disease prediction. There are some papers on disease prediction with electronic health record (EHR) [9], but the research on disease prediction with raw symptoms is still necessary and challenging. Therefore, we propose a general new idea of using the rich contextual information of diseases and symptoms to bridge the gap of disease candidates and symptoms, and detach it from the specific way of implementing the idea using network embedding.

4. Recommendation in medical domain is usually a decision-making issue, which requires the ability of explaining “why”. The ability of explaining “why” are basically from two paths. Consider the recommendation suggest you eat more vegetables. You probably do not believe it if there is nothing attached. But if the recommendation gives the literally reasons why eating more vegetables is good you might like to take this suggestion. Consider another scenario, if the recommendation gives you the data of the contrast which show that people who eat more vegetables are healthier than those eat less, it is certain that you also want to take this recommendation. Based on these two intuitions, we present a recommendation model based on proofs which are either literally reasons or difference from contrast.

2. THE PROPOSED RESEARCH

In this research, we argue that medical domain causality mining is vital for assisting diagnosis.

2.1 Mining Medical Domain Causality from Text

A considerable amount of causal relations are emerging in medical literature and medical forums every day. Mining these textual causalities can be very useful for discovering new knowledge and making decisions. One can find an unknown adverse-effect of a drug by mining causal relation from medical forum data, which is very difficult to get from medical experiments. In another scenario, mining “what kind of habit can cause diabetes” from medical literature can be very useful for doctors giving their patients good suggestions. Table 1 presents examples of medical causality in text, in which example A and B are similar explicit ways to express medical causal relations. However, example C and D alternatively show similar implicit ways. Note that example B and C express the same causality which can be taken as the clue to bridge the gap between explicit way and implicit way. These examples show some other hints for mining causality, i.e., 1) causal pairs tend to co-occur frequently; 2) the same causal knowledge can be expressed in both explicit and implicit ways; 3) different causal pairs are usually expressed in the same way. Obviously, from A and B we can see that the explicit causality is quite easy to recognize. Therefore, it will be interesting to study how we can leverage explicit causality for mining implicit causality.

2.2 Treatment Effectiveness Measurement

Treatment effectiveness analysis is usually taken as a subset of causal analysis on observational data. For such real observational data, PSM and RCM are two dominant methods. On one hand, it is usually difficult for PSM to find the matched cases due to the sparsity of symptom. On the other

A: Some medical associations report that smoking can <i>[cause]</i> poor blood flow in arms and legs. $\Rightarrow [\text{smoking}]_{\text{cause}} \xrightarrow{\times 11} [\text{poor blood flow}]_{\text{effect}}$
B: Some medical institutes report that drinking alcohol can <i>[cause]</i> diabetes . $\Rightarrow [\text{drinking alcohol}]_{\text{cause}} \xrightarrow{\times 47} [\text{diabetes}]_{\text{effect}}$
C: The American Diabetes Association explains what drinking alcohol can do to the conditions that cause diabetes . $\Rightarrow [\text{drinking alcohol}]_{\text{cause}} \xrightarrow{\times 47} [\text{diabetes}]_{\text{effect}}$
D: The American Heart Association explains what cigarette smoking can do to the conditions of forming thrombus . $\Rightarrow [\text{cigarette smoking}]_{\text{cause}} \xrightarrow{\times 23} [\text{forming thrombus}]_{\text{effect}}$

Table 1: **A** is an explicit cause-effect example. **B** is an explicit cause-effect example taking similar expression with **A**. **C** is the same cause-effect with **B** but expressed in an implicit way. **D** is an implicit cause-effect example taking similar expression with **C**.

hand, we should check every possible (symptom, treatment) pair by exploiting RCM, leading to make the characteristic of exploding up, especially when we want to check the causal relation between a combination of symptoms and a combination of drugs. Besides, the larger number of symptom or treatment in the combination the less number of patient case retrieved, which lead to the lack of statistical significance. Specifically, patients tend to take tens of herbs as the treatment each time in Traditional Chinese Medicine (TCM). Therefore, how to evaluate the effectiveness of herbs separately and jointly is really a big challenge. This is also a very fundamental research topic supporting many downstream applications, such as explicable medical treatment recommendation.

2.3 Disease Prediction via Causal Representation

Both hospitals and on-line forums have accumulated sheer amount of records, such as clinical text data and online diagnosis Q&A pairs. The availability of such data in large volume enables automatic disease prediction. There are some papers on disease prediction with electronic health record (EHR) [10, 9], but the research on disease prediction with raw symptoms is still necessary, which reflecting on three aspects. 1) Predicting diseases with raw symptoms is a necessary complement to EHR-based disease prediction because EHR is really expensive for both patients and researchers. Besides, a lot of diseases are unnecessary for a formal medical checkup. 2) Online medical forums are in urgent need of answers about diseases with raw symptoms as questions. 3) Disease prediction with raw symptoms is more feasible to incorporate online medical forum Q&A data.

However, exploiting clinical text data or online diagnosis Q&A pairs for disease prediction is nontrivial because of the serious issue of sparseness, that is, 1) the number of symptoms and diseases is in thousand-level due to the variety of symptoms and diseases and the diversity of natural language expressions, 2) the number of symptom-disease links is *relatively small* (see Figure 1). To address the sparsity problem, we propose a general new idea of using the rich contextual information of diseases and symptoms to bridge the gap of disease candidates and symptoms, and detach it from the specific way of implementing the idea using network embedding. Specifically, we create a two-layer network which

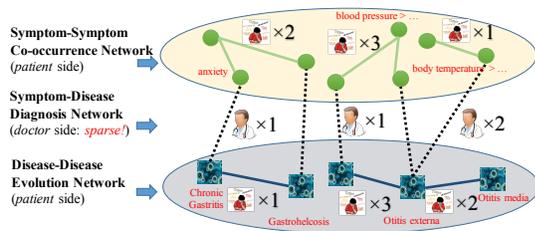


Figure 1: *Three information networks we have for disease prediction.* Traditional methods rely on the symptom-disease diagnosis network with information provided by doctors and experts. However, this network seriously suffers from the sparseness. Fortunately, we have the symptom co-occurrence network and disease evolution network that carry rich contextual information from the patient side. How to integrate these three networks to bridge the symptoms and diseases is important and challenging.

consists of a bipartite diagnosis network and the following two homogeneous networks. (see Figure 1).

2.4 Explicable Medical Treatment Recommendation

Recommendation in medical domain is usually a decision-making issue, which requires the ability of explaining “why”. The ability of explaining “why” are basically from two paths. Consider the recommendation suggest you eat more vegetables. You probably do not believe it if there is nothing attached. But if the recommendation gives you the following information (“Eating vegetables provides health benefits — people who eat more vegetables and fruits as part of an overall healthy diet are likely to have a reduced risk of some chronic diseases. Vegetables provide nutrients vital for health and maintenance of your body”) you might like to take this suggestion. Consider another scenario, if the recommendation gives you the data of the contrast which show that people who eat more vegetables are healthier than those eat less, it is certain that you also want to take this recommendation. All existing recommendation systems use a number of different technologies but they can be classified into two broad groups, i.e., content-based systems and collaborative filtering systems. For medical treatment recommendation, content-based systems examine properties of the treatment recommended and collaborative filtering systems recommend treatments based on similarity measures between patients and treatments. However, all these systems don’t consider the ability of explaining “why” seriously. Above all, developing explicable medical treatment recommendation system is necessary and valuable.

3. METHODOLOGY

We introduce our algorithms for causality mining for assisting diagnosis.

3.1 Mining Medical Domain Causality

In this task, we take both the power of frequency of cause-effect co-occurrence, semantic distance between causality pairs and the transitivity rules of causality for mining medical domain causality. Specifically, we present a factor graph to combine these three resources for causality mining. The model make it possible to model three kinds of important but entirely different resources as factors in a unified frame-

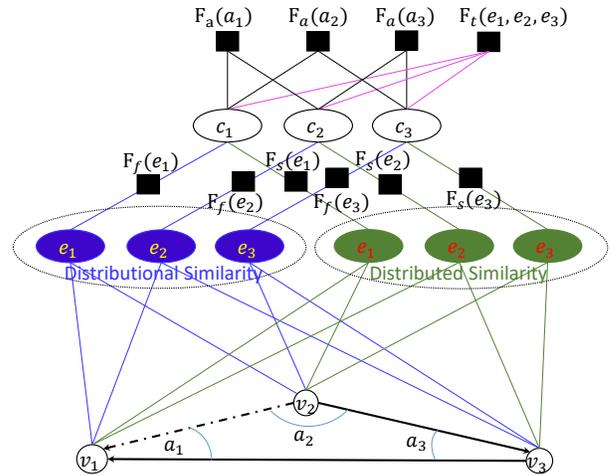


Figure 2: *Our CAUSALSIM for causality mining.* It is a factor graph which leverage the power of both the co-occurred frequency between events and the semantic distance between event pairs. It also takes the transitivity of causality into consideration.

work. Specifically, we use co-occurrence counts of events collected automatically from an unannotated corpus, which is the component based on distributional similarity. To capture the semantic distance between pairs, we use word2vec [6] to predict embeddings concerning distributed representation of event pairs.

A factor graph provides an explicit way to factorize the global joint probability by introducing local variable and factor nodes. The introduction of factor nodes allows the optimization algorithms to be derived in a simple and general form, because the factor graph unifies the graph with the same representation. In this study, to represent the three kinds of resources for causality mining, we present a unified framework CAUSALSIM to take distributed similarity, distributional similarity and causal transitivity as different local factors shown in Figure 2. Therefore, the joint probability can be factorized as distributional factor $Pr(c_{e_i}|f_{e_i})$, distributional factor $Pr(c_{e_i}|s_{e_i})$, angle rule factor $Pr(C_{a_i})$ and triangle rule factor $Pr(C_{t_i})$.

$$Pr(C|G, F, S) = \prod_{e_i} Pr(c_{e_i}|f_{e_i}) \prod_{e_i} Pr(c_{e_i}|s_{e_i}) \prod_{a_i} Pr(C_{a_i}) \prod_{t_i} Pr(C_{t_i}) \quad (1)$$

where e_i represents each causal relation, a_i represent an angle structure composed by a vertex v_i and two edge e_m and e_n , t_i represents a triangle structure composed by three vertices and three edges.

All these four factors can be instantiated by exponential-linear functions:

$$Pr(c_{e_i}|f_{e_i}) = \frac{1}{Z_{e_i=(v_m, v_n)}} \exp\{\alpha\Gamma \cdot I(v_m, v_n)\} \quad (2)$$

$$Pr(c_{e_i}|s_{e_i}) = \frac{1}{Z_{e_i}} \exp\{\beta \min Dist(e_i, R_k)\} \quad (3)$$

$$Pr(C_{a_i}) = \frac{1}{Z_{a_i}} \exp\{\delta A(a_i)\} \quad (4)$$

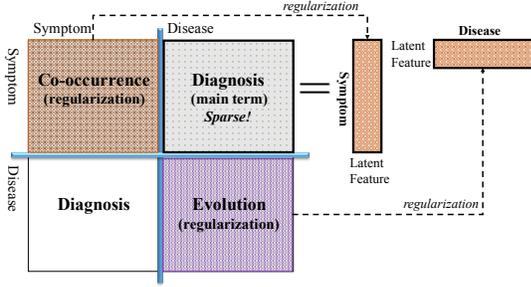


Figure 3: Our CONTEXTCARE method that learns effective latent features of symptoms and diseases. We adopt the diagnosis network to learn the representations as the main term, and we use the symptom co-occurrence network and disease evolution network to regularize the representations (as lasso constraints) in order to alleviate the sparseness of the symptom-disease diagnosis network.

$$Pr(C_{t_i}) = \frac{1}{Z_{t_i}} \exp\{\gamma T(t_i)\} \quad (5)$$

where $\Gamma \cdot x = \frac{e^{bx} - 1}{e^x - 1}$ is for rewarding high frequently co-occurrence cause-effect pair. $I(\cdot)$ is the mutual information function. $Dist(e_i, R_k)$ is the context based semantic distance between a causal candidate e_i and the k th cluster center R_k of known causalities, where R_k is the closest domain-sensitive center with e_i .

$$A(c_1, v_1, c_2) = \begin{cases} -1, & c_1 + c_2 = 2 \text{ and } Ca(v_1, c_1) \text{ and } Ef(v_1, c_2) \\ 1, & c_1 + c_2 = 2 \text{ and } Ca(v_1, c_2) \text{ and } Ef(v_1, c_1) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $Ca(v_1, c_1)$ represents that v_1 is a cause in relation e_1 with label c_1 . $Ef(v_1, c_1)$ represents v_1 is an effect in relation e_1 with label c_1 .

$$T(c_1, c_2, c_3) = \begin{cases} -1, & c_1 + c_2 + c_3 = 2 \\ 1, & c_1 + c_2 + c_3 = 3 \\ 0, & c_1 + c_2 + c_3 = 0 \text{ or } 1 \end{cases} \quad (7)$$

3.2 Disease Prediction via Causal Representation

To address the sparsity problem, we propose a general new idea of using the rich contextual information of diseases and symptoms to bridge the gap of disease candidates and symptoms, and detach it from the specific way of implementing the idea using network embedding. Specifically, we create a two-layer network which consists of a bipartite diagnosis network and the following two homogeneous networks i.e., *Symptom co-occurrence network* and *Disease evolution network*. (see Figure 1).

To make use of those two networks as contextual information for accurate disease prediction, we propose a novel method, CONTEXTCARE, to *integrate the three networks of different aspects but rich information*. The basic ideas of CONTEXTCARE are 1) the linked disease and symptom in the bipartite diagnosis network should be closer in a certain way; 2) frequently co-occurred symptoms should be closer in representation; 3) evolutionary diseases should be closer in representation. Compacting symptoms in 2) and diseases in 3) both benefit alleviating the sparsity of the bipartite diagnosis network. By encoding all these aspects, our CONTEXTCARE utilizes information from the three networks to

facilitate the predictive model. CONTEXTCARE takes the *s-pars*e symptom-disease diagnosis network as the main term in the objective function and adopts the *contextual information* including the symptom co-occurrence and disease evolution as regularization terms (as lasso constraints), as shown in Figure 3.

Specifically, We define a main loss based on simple energy function $f(S^p, d)$ on each symptoms-disease pairs (S^p, d) and two lasso constraints loss. The final objective loss of our model combines main loss and two constraints loss with weight α and β .

$$f(S^p, d) = \left\| \frac{1}{|S^p|} \sum_{s \in S^p} \mathbf{s} + \mathbf{t} - \mathbf{d} \right\|_1 \quad (8)$$

We are using an ℓ_1 norm in the latent space, but other metrics could be used as well. We take the symptom-disease as a transition vector \mathbf{t} . True symptoms-disease pairs (S^p, d) are assumed to have low energies in the energy function.

Two lasso constraints on frequently co-occurred symptoms and evolutionary diseases are as follows.

$$R_1(G^{SS}) = \sum_{(s_i, s_j) \in E^{co}} w_{ij} \|\mathbf{s}_i - \mathbf{s}_j\|_1 \quad (9)$$

where $w_{ij} = \frac{|\Gamma(s_i) \cap \Gamma(s_j)|}{|\Gamma(s_i) \cup \Gamma(s_j)|}$ is the Jaccard similarity between symptoms s_i and s_j , $\Gamma(s_i)$ and $\Gamma(s_j)$ denote the set of neighbors of symptom s_i and symptom s_j in the symptom-symptoms co-occurrence network with frequency respectively. Note that we consider the frequency of the edge between neighbors when counting $\Gamma(s)$, i.e., we take a symptom s_j as a neighbor of s_i when the co-occurrence frequency of s_i and s_j is beyond a threshold τ . The constraint incentivizes the vector representations of frequently co-occurred symptoms to be close, the larger w_{ij} , the greater penalty. The following constraint incentivizes the vector representations of diseases which are the different stages of the same patient to be close, the larger v_{mn} , the greater penalty.

$$R_2(G^{DD}) = \sum_{(d_n, d_m) \in E^{ev}} v_{mn} \|\mathbf{d}_n - \mathbf{d}_m\|_2 \quad (10)$$

where $v_{mn} = \frac{|\Gamma(d_n) \cap \Gamma(d_m)|}{|\Gamma(d_n) \cup \Gamma(d_m)|}$ is again the the Jaccard similarity, $\Gamma(d_n)$ and $\Gamma(d_m)$ denote the set of neighbors of disease d_n and disease d_m in the disease evolution network respectively. Note that we make use of ℓ_2 on evolutionary diseases rather than ℓ_1 like the penalty on co-occurred symptoms. The reason is 1) we want treat these two kinds of relations differently and 2) put more emphasis on disease evolution because of more significant effect on compacting diseases.

4. EMPIRICAL RESULTS

4.1 Mining Medical Domain Causality

We compare our method with different kinds of baselines for causal mining on two datasets SemEvalExtend¹ and MedArtic², which is shown in Table 2.

Table 2 shows that our method outperforms the previous methods on all the three metrics on two different datasets. Our CAUSALSIM encode three kinds of different resources (i.e., frequency co-occurrence, contexts based semantic and

¹detailed in our [12]

²We labeled domain-specific medical articles in Chinese

Table 2: Comparison with baselines on causal mining from text (%).

Method	SemEvalExtend			MedArtic		
	P	R	F1	P	R	F1
Ittoo [2]	65.7	68.3	66.9	53.6	23.4	32.6
Sorgente [8]	76.7	71.4	73.9	57.2	26.1	35.8
Do [1]	69.1	74.2	71.6	54.3	48.3	51.1
Zhao [12]	87.3	84.1	85.6	62.8	26.3	37.1
CAUSALSIM	89.6	86.3	87.9	63.4	57.2	60.1
-Distributional	76.2	72.4	74.3	57.8	42.3	48.8
-Distributed	73.8	84.5	78.8	52.3	52.6	52.4
-Angle	88.9	85.7	87.3	60.8	55.6	58.1
-Triangle	89.1	84.1	86.5	61.2	52.1	56.3

inference rules) and model them in a unified factor graph for causal mining. [8] just used lexico-syntactic features and some self-constructed rules. [2] just encode lexico-syntactic patterns in their bootstrapping framework. [1] just exploit frequency co-occurrence and causal triggers to find causal candidate. lexico-syntactic patterns and causal triggers are used in [12]. Therefore, CAUSALSIM leverages more kinds of resources leading to a better performance. We also compare CAUSALSIM with removing four factors on CAUSALSIM separately to show the power of different factors. The comparison shows that “Distributed”, i.e., mutual information between cause and effect, plays the most important role in causal mining, and “Distributed”, i.e., contexts based semantic has the most power on improving recall, and “Angle” and “Triangle”, i.e., causal transitivity rules are both important for causal mining.

Experiment results on two different data sets show that our model can effectively mine causality. The results also show that our model improves over CAUSALSIM-Distributional, CAUSALSIM-Distributed, CAUSALSIM-Angle and CAUSALSIM-Triangle, suggesting the necessity of combing them.

4.2 Disease Prediction

We conduct experiments on a real Chinese clinical narrative data, which are recorded by medical professionals. In the clinical case of each patient, we have a set of symptoms in text, diagnosed current disease given by the doctor-in-charge and historical diseases proposed by the patient. In our dataset, there are 17,803 clinical cases with 18,899 symptoms and 1,066 diseases in total. We split all clinical cases into a training/validation/test set randomly, with the ratio of 8:1:1. The first part is used for model training, the second for hyper-parameter tuning, and the third for evaluation. Results are given in Table 3.

Table 3 shows the accuracy of the baseline methods as well as our CONTEXTCARE on disease prediction. Each row of Table 3 represents a model for disease prediction. Each column stands for a type of resource combination used to predict disease. For example, the column “*Diagnosis Network*” denotes the use of bipartite symptom-disease diagnosis network including a large number of symptom-disease paths and the column “*+Co*” and “*+Ev*” denotes that the co-occurrence of symptoms and evolution of diseases are taken into account separately besides symptom-disease paths.

Predicting the category of disease is valuable because of the following two scenarios. First, it is necessary to early determine the department that a patient should make an appointment with. Second, doctors and experts can provide quick consulting if they have the correct category of the

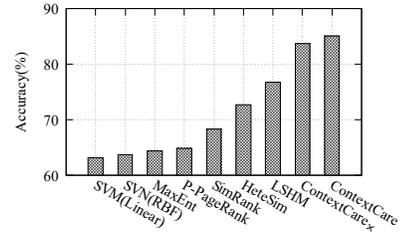


Figure 4: Comparison with baseline methods on disease category prediction in accuracy (%).

potential disease in advance. Figure 4 shows the results that CONTEXTCARE stands out from all the baseline methods in disease category prediction.

5. FUTURE ISSUES FOR DISCUSSION

The potential research issues for discussion are described in this part. For **medical domain causality mining**, the issue is how to qualify the certainty of obtained causality candidates to get large amount of valid causal knowledge. For **treatment effectiveness measurement**, the issue is that is there a causal analysis model could model the effectiveness of drugs combination and find most effective combinations. For **disease prediction via causal representation**, the issue is upon representation learning model and how to encode expert knowledge (logic and rules) into representations. For **explicable medical treatment recommendation**, the issue might be how to leverage treatment effectiveness measurement model and provide effective treatment program and present reasons in rational way.

6. ACKNOWLEDGEMENT

This work was supported by the National Key Basic Research Program of China (973 Program) via grant 2014CB340503 and the National Natural Science Foundation of China (NSFC) via grants 61133012 and 61472107.

7. REFERENCES

- [1] Q. X. Do, Y. S. Chan, and D. Roth. Minimally supervised event causality identification. In *EMNLP'11*, pages 294–303. ACL, 2011.
- [2] A. Ittoo and G. Bouma. Minimally-supervised extraction of domain-specific part-whole relations using wikipedia as knowledge-base. *Data Knowl. Eng.*, 85:57–79, May 2013.
- [3] Y. Jacob, L. Denoyer, and P. Gallinari. Learning latent representations of nodes for classifying in heterogeneous social networks. In *WSDM*, pages 373–382, 2014.
- [4] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *KDD*, pages 538–543, 2002.
- [5] G. Jeh and J. Wiom. Scaling personalized web search. In *WWW 2003*, pages 271–279, 2003.
- [6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *ICLR*, 2013.
- [7] C. Shi, X. Kong, Y. Huang, P. S. Yu, and B. Wu. Hetesim: A general framework for relevance measure

Table 3: Comparison with baseline methods in accuracy and precision@N (%).

Method	Accuracy				Precision@5	Precision@10
	<i>Diagnosis Network</i>	<i>+Co</i>	<i>+Ev</i>	<i>+Co,Ev</i>		
SVM (linear)	16.02	-	-	-	-	-
SVM (RBF)	16.79	-	-	-	-	-
Decision Tree (C4.5)	17.31	-	-	-	-	-
MaxEnt	18.98	-	-	-	-	-
P-PageRank [5]	17.22	19.71	17.25	19.74	43.17	62.16
SimRank [4]	19.36	21.97	19.38	21.98	46.52	64.33
HeteSim [7]	20.62	23.03	20.69	23.32	55.31	70.48
LSHM [3]	21.38	25.87	22.55	25.77	65.74	82.45
ContextCare _x	22.35	28.09	24.74	30.66	69.38	85.76
CONTEXT CARE	23.57	30.32	27.26	31.73	73.21	87.36

in heterogeneous networks. *TKDE*, 26(10):2479–2492, 2014.

- [8] A. Sorgente, G. Vettigli, and F. Mele. Automatic extraction of cause-effect relations in natural language text. In *AI*IA*, pages 37–48, 2013.
- [9] J. Sun, J. Hu, D. Luo, M. Markatou, F. Wang, S. Ebadollahi, Z. Daar, and W. F. Stewart. Combining knowledge and data driven insights for identifying risk factors using electronic health records. In *AMIA*, volume 2012, pages 901–10, 2012.
- [10] F. Wang, P. Zhang, B. Qian, X. Wang, and I. Davidson. Clinical risk prediction with multilinear sparse logistic regression. In *KDD*, pages 145–154, 2014.
- [11] S. Zhao and T. Liu. Causality and its applications in social media: A survey. *Journal of Software*, 25(12):2733–2752, 2014.
- [12] S. Zhao, T. Liu, S. Zhao, Y. Chen, and J.-Y. Nie. Event causality extraction based on connectives analysis. *Neurocomput.*, 173(P3):1943–1950, Jan. 2016.
- [13] S. Zhao, Q. Wang, S. Massung, B. Qin, T. Liu, B. Wang, and C. Zhai. Constructing and embedding abstract event causality networks from text snippets. In *WSDM '17*, 2017.