Protein-protein interaction network constructing based on text mining and reinforcement learning with application to prostate cancer

Fei Zhu[1,2,3], Quan Liu[1,3], Xiaofang Zhang[1], Bairong Shen[2]
1 School of Computer Science and Technology, Soochow University, Suzhou 215006, China
2 Center for Systems Biology, Soochow University, Suzhou 215006, China
3 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, P. R. China

***Abstract*—As a notoriously lethal human disease, cancer has obtained much concern for a long time. There have accumulated huge amounts of literature and experimental data on cancer-related research. It is impossible for people to deal with these texts manually to discover novel information and knowledge. However, text mining has an advantage of extracting previously unknown and understandable knowledge from large amounts of texts, and forming well-defined knowledge, providing the possibility to fully taking use of the existed texts. With the proceeding of biomedical research, people have gradually realized that complex biological functions and the phenomenon of life are the results of complex interactions among a variety of biological entities, such as protein. Deeply studying protein interaction network is essential to understand life. We, adopting reinforcement learning idea, put forward an algorithm for protein interaction network constructing. With the algorithm, nodes are used to represent proteins and edges denote interactions. During the evolutionary process, a node selects with which nodes in the network it tends to interact. Keep selecting and carrying on iteration, until eventually attaining an optimal network. The network is the result of the dynamic nature of learning behavior. As a malignancy, prostate cancer has been concerned for a long time. We attain biological texts from PubMed and establish a prostate cancer protein interaction networks by the proposed methods. The results show that our proposed method is pretty good. Network topology analysis results also show that the network node degree distribution is scale-free.**

Keywords- systems biology; text mining; prostate cancer; reinforcement learning; protein interaction network

I. INTRODUCTION

As an infamous lethal disease, cancer has caused more than millions of human deaths. Curing cancer is always an important goal of the struggle for researchers and practitioners from diverse domains. Accordingly, various researches on cancer have been one of the most influential study areas for biomedical researchers for more than 100 years. The corresponding huge body of cancer research texts is a valuable resource for people. The enormous scale of the publications on cancer research keeps increasing every year. By searching PubMed with "cancer" as keyword, we retrieved more than three million of publications.

The mass amounts of biomedical texts provide us a rich source of knowledge for research. However, it is impossible to manually process these materials. Text mining, which can help people to mine information and extract knowledge from mountains of texts, has now been extensively applied in biomedical research. Many have benefited from the convenience of text mining technology to discover novel knowledge to improve the development of biomedical research, especially those pertaining to malignant diseases, such as cancer. Realizing the advantages of text mining will facilitate cancer research, by contributing to find new knowledge for cancer diagnostics, treatment, and prevention.

Text mining, with the goal of finding new exciting outcomes hidden in unstructured texts, involves many computational fields and technologies, such as machine learning, natural language processing, statistics, and pattern recognition. There are many cancer-related text mining applications, e.g. identifying malignant tumor related biomedical mentions (genes, proteins, etc.), discovering relationships among biomedical entities (protein-protein, gene-disease, etc.), extracting knowledge from texts and generating hypotheses, and constructing or improving pathways as well.

Reinforcement learning provides a framework to learn directly from the interaction and achieve goals. Reinforcement learning can optimize in unknown environments so as to be suitable for the unknown and uncertainty of the new biological system. The versatility and openness of reinforcement learning ensure that it can make full use of biological knowledge.

Many complex networks in real world, including human protein interaction networks and metabolic network are scale-free. Using reinforcement learning method, agent can repeatedly try selecting interaction, rewards and returns determine which interactions will be reinforced, and network structure is output of dynamics of learning results. The scale-free properties of the network will be preserved.

Many systematic complex diseases, such as cancer, involve some mechanisms that have not yet been understood. Reinforcement learning provides an approach to learn the best decision in an unknown system. Reinforcement learning methods guarantee networks converge to an optimal state.

With reinforcement learning framework, the biological knowledge can be seamlessly embedded into the network construction approach. Moreover, multiple sources of biological data are capable of being used to build the network. Thereby, reinforcement learning framework ensures that the constructed network is with high confidence.

In this work, we take advantage of the reinforcement learning framework to build up a prostate cancer protein interaction network from text data. In this work, we establish the protein interaction network with a reinforcement learning approach, where a node represents a protein and an edge

978-1-4799-5669-2/14/$31.00 ©2014 IEEE

denotes an interaction. Node *i* select node to interact with under the decision of reinforcement learning agent, thus getting a decision. Node will get a reward after each attempt. The value of reward determines which interactions will be reinforced. The structure of the interaction network is yielded by the continuing iteration of the agent.

II. TEXT MINING IN CANCER SYSTEMS BIOLOGY

Text mining is the process of extracting previously unknown and understandable knowledge from large amounts of texts, and forming well-defined knowledge. As the source data of text mining is unstructured and the goal is to discover unknown knowledge, the problem that text mining tries to solve is very complex. Technically, text mining combines multiple disciplines, covering information technology, text analysis, pattern recognition, statistical, data visualization, databases, machine learning, natural language processing, information retrieval, data mining, and other fields.

With the development of biomedical research, people research work is published in the form of texts. Much biomedical knowledge exists in a non-structured form in a variety of texts. Biomedical text provides a wealth of resources for biomedical research. However, due to the magnificent scale, it is hard to fully process throughout by human being. Effective use of biomedical knowledge embedded in the text to analyze vast amounts of biomedical data has been a task and challenge. More and more interests have been put in text mining. By retrieval results from PubMed with "text mining" as keyword, we can see that there is a very large increasing in publication number since 2000, as showed in Figure 1. Many researchers use data mining techniques to discover new knowledge in order to promote biomedical research, particularly in some areas of malignant diseases, such as cancer research.

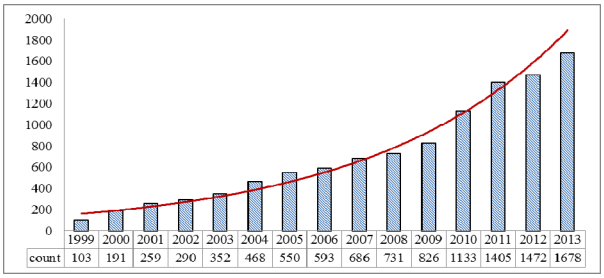

Figure 1. Retrieval results and its trend by using "text mining" keyword from 1999 to 2013

Text mining is composed of four stages: information retrieval, information extraction, knowledge discovery, and hypothesis generation. Information retrieval systems aim to get desired texts on a certain topic; information extraction systems are used to extract predefined types of information such as relation extraction; knowledge discovery systems help us to extract novel knowledge from texts; hypothesis generation systems infer unknown biomedical facts based on texts. Consequently, the general tasks of biomedical text mining include information retrieval, named entity recognition, relation extraction, and hypothesis generation.

Extracting biomedical relations, such as protein-protein interaction and gene-disease relation, is a hot spot. Lots of work has been presented by researchers. In the current genomic era, researchers are interested in mining interactions in genome-wide associations which provide beneficial scaffolds for further biomedical research, and other comprehensive relationships as well.

As a sophisticated disease, cancer has relation to amounts of biological entities, e.g., genes and proteins. How to fully take advantage of existed biomedical texts has not only become a challenge but also a hot field. Biomedical researchers concentrate on finding out cancer-related biological entities from the literature for further cancer diagnostics, treatment, and prevention study. It is generally believed that early detection, evidence-based strategies for prevention and patient management can be used to reduce and control the causes of cancer. Hence, cancer risk assessment, which evaluates the probability of suffering from cancer by evaluating the available evidence, is an important part of cancer research, where the existed texts are available materials and therefore biomedical text mining is a practical approach.

Researchers get to analyze and understand complex biological systems, such as cancer, from a systems biology viewpoint. Some work showed that text mining can be used to facilitate the development of systems biology. For instance, systems biology-based networks can be constructed by aggregating previously reported associations from texts or various databases. The network of genes, genetic diseases, and brain areas introduced by Hayasaka et al. used extracted associations from the texts as construction basis. The interaction network through literature mining provided by Sharma et al. demonstrated 19 genes were confirmed to be related to prostate cancer. As it can be seen, the full utilization of text mining to enhance cancer systems biology research is a new hot topic.

III. REINFORCEMENT LEARNING

Reinforcement learning is based on the idea that the system learns directly from the interaction during the process of approaching the goals. The reinforcement learning framework has five fundamental elements: a controller, environment, state, reward, and action, showed as Figure 2. The controller, which learns knowledge by interacting with outside environment and then chooses an action in accordance with the decision made by established controlling model, is the agent of the system; accordingly, the state will then be changed; the environment will return a reward to evaluate the action taken.

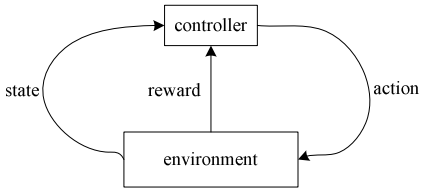

Figure 2. Framework of reinformcement learning. Controller selects an action; the environment responds to the action, generates new scenes to the agent, and then returns a reward.

Reinforcement learning introduces an approach to learning in an unknown environment. During each episode of trying to reinforcement learning model, agent chooses an action from the available action set and receives reward. In recent years, reinforcement learning has attained the attention from more and more researchers, and has been used in many biomedical fields. Farhang Sahba et al utilized reinforcement learning to segment CT images so as to identify suspected cancer part. The diagnosis and health decision system based on reinforcement learning has also been utilized. However, very few people use reinforcement learning method and analysis of protein-protein interaction networks.

A Markov decision process (Markov decision process, MDP) is often used to model the reinforcement learning problem. Usually a MDP model can be represented by a tuple M=<X, U, f, ρ, γ>, where X is the state space, u is the action space, f is the state transfer function, ρ is the reword function and γ is the discount factor.

In the reinforcement learning framework, the controller interacts with the environment, gets the representation of environment denoted by state $x_k \in X$, and then chooses an action $u_k \in U$ according to its policy h:$X \rightarrow U$ such that $u_k=h(x_k)$, where u is all available actions. By taking the action. The controller will be returned a reward $r_{k+1}=\rho(x_k, u_k)$ and gets to a new state x_{k+1}. In the long term of trying, learning and optimizing, the controller will try to get the maximal sum of the rewards suggesting the optimal action sequence.

Reinforcement learning is intended to get a solution solving a stable policy that will not change over time, and make decisions by the attained policy. In reinforcement learning, strategy be given in h: $X \times U \rightarrow [0,1]$. The value of $h(x,u)$ refers to the probability of taking the action of u under state x. If h is a deterministic strategy, in any state x, agent can get a deterministic action $u=h(x)$. We can use value function V^h and action value function Q^h to evaluate the merits of the policy of a Markovian decision process, where $V^h(x)$ is the expected accumulative reword under state x by policy h, and Q^h is the expected accumulative reword under current state-action (x,u) by policy h.

$$V^h(x) = \sum_{u \in U} h(x,u)[\rho(x,u) + \gamma \sum_{x' \in X} f(x,u,x')V^h(x')] \quad (1)$$

$$Q^h(x,u) = \rho(x,u) + \gamma \sum_{x' \in X} f(x,u,x') \sum_{u' \in U} h(x',u')Q^h(x',u') \quad (2)$$

The ultimate goal of reinforcement learning is to get an optimum strategy h^*. The corresponding value function $V^h(x)$ and action-value function $V^h(x,u)$can be represented as

$$V^*(x) = \max_{u \in U} \{\rho(x,u) + \gamma \sum_{x' \in X} f(x,u,x')V^*(x')\} \quad (3)$$

$$Q^*(x,u) = \rho(x,u) + \gamma \sum_{x' \in X} f(x,u,x')\{\max_{u' \in U} Q^*(x',u')\} \quad (4)$$

IV. Protein Interaction Network

Along with the deepening in biomedical research, it is gradually realized that complex biological functions and biological phenomena are determined by complex interactions between various basic bio-unit rather than the structure and function of a single biomolecular. Hartwell proposed that modern biology need not only study cell structure and function of biological molecules, but also to study these interactions between biological molecules, and how life functions is realized via the various interactions of biomolecules, which requires us to study a variety of bio-molecular interaction through the formation of networks.

Individual proteins interact with other proteins and then constitute a protein interaction network to participate in biological signal transduction, regulation of gene expression, metabolism and cell cycle regulation. Systematic analysis of protein interactions with a systems biology view has a significant role in understanding the working principle of the biological system, biological signals and reaction mechanisms of energy metabolism in a disorder or disease physiological condition, and the functional relations between proteins.

Interactions between proteins are the basis of many biological functions. In protein interaction networks, a node represents protein, and an edge denotes a physical connection between proteins which, in most cases, are undirected. Generally, we can use three different methods to construct protein interaction networks. The first one is to utilize text mining technology; the second one is to perform prediction based on known homologous; the third one is take advantage of high throughput experiment.

There are plenty of important and complex networks. Barabási found that in most cases, the node degree of a network of real world rarely obey power rate law rather than Poisson distribution. For a randomly selected node, the probability of its degree being k is $p(k) \propto \frac{1}{k^r}$, where r and k are constants. The network thereby is called scale-free network. Researchers conclude that the characteristic of scale-free is a natural feature of many complex networks. Many biological networks, such as protein interaction networks of yeast, Caenorhabditis elegans and Drosophila, are of scale-free. Analyzing the topology of protein interaction network will usually bring out some significant biological knowledge.

V. Modelling With Reinforcement Learning

We utilize reinforcement learning framework to build up a prostate cancer protein interaction network from text data. In this work, we establish the protein interaction network with a reinforcement learning approach, where a node represents a protein and an edge denotes an interaction. The node i select node to interact with under the decision of reinforcement learning agent, thus getting a decision. Node will get a reward after each attempt. The value of reward determines which interactions will be reinforced. The structure of the interaction network is the result of continuing iteration of the agent. In this way, both the evolution of the interaction network and the evolution of the individual protein are taken into consideration.

Node i randomly selects other nodes to establish an interaction with the probability that is assigned to other nodes. Each node has the policy to choose nodes and get reinforced in each iteration. Each node maintains a weight

vector $\langle w_{i1},\ldots,w_{in}\rangle$. As each time newly added nodes connect existing node i with a given probability, any node i is selected by weight vector $w_i(t)$ at time t. The network establishing process can be seen as a Markov chain, and thus can utilize reinforcement learning method.

In this work, we will introduce a reinforcement learning model for protein interaction network establishing. The fundamental elements of reinforcement learning, reward, actions, and states, will be discussed.

A. Description of Reward

The construction of a network can be modeled as a Markov decision process model, and can get optimal strategies by reinforcement learning algorithms. In interaction networks, each node can be regarded as an agent that has intelligence. A node chooses other nodes as interaction node according to a certain probability. There are many methods to get the probability. In this work, we get the access probability by computing the rate of occurrence of node i choosing node j to all occurrence of node i.

$$p_{ij} = \frac{\text{occurence of node } i \text{ and node } j}{\sum_{\text{k}} \text{occurence of node } i \text{ and node } k} \tag{5}$$

However, texts from different sources have different levels of authority and credibility. Therefore, we need to consider the weight of text when using text data. We choose impact factor (IF) value as a weight factor. Generally, if the journal has a larger impact factor, the internal research texts are tending to have higher weight. The weight can be calculated as

$$w_{ij} = \frac{\sum_{\text{t}} \text{occurence of node } i \text{ and node } j * \text{impact factor of text } t}{\sum_{\text{k,t}} \text{occurence of node } i \text{ and node } k * \text{impact factor of text } t} \tag{6}$$

We can view the probability of selecting a node is the weight assigned by other nodes, which is used to measure the probability of interaction between two nodes. Thus, in a n-node network, any node i, with a weight vector, the access probability of node j is $\frac{w_{ij}}{\sum_k w_{ik}}$. Finally we obtain a probability matrix.

Through the matrix, node i can get the probability of interacting with node j. At each stage, there is a probability matrix which represents the probability of interaction. The network topology changes with the procedure of interaction. During each iteration, the probability matrix of a node selecting other nodes will be updated by equation 6. When the probability matrix keeps stable, the evolution process can be considered completed, and the topology of the network is formed.

B. Description of Action

In reinforcement learning, policy which defines an agent at a given time is a mapping from the state to the available action in the state. Reinforcement learning algorithm obtains a reward by mapping the scene to action. Actions don't only affect the direct rewards, but can also affect the next scene, which will affect all subsequent rewards.

In the process of construction of a protein interaction network, the actions of an agent can be described as determining whether there is an interaction between the current node and another node, as showed in Table I.

TABLE I. AVAIABLE ACTIONS

#	Action
0	unable to determine
1	no interaction between nodes
2	interaction between nodes

C. Description of State

Agent in the current state chooses an action, and enters the next state. Hereby the state can be regarded as the static image of the external environment in a given state after taking some action.

In protein interaction networks, the state at time t is the description of all proteins and interactions. With a graph view, it is a description of all nodes and edges between nodes, as well as the corresponding transition probability matrix w.

D. Algorithm for Network Construction

In this work, we introduce an algorithm, in which each node is to choose with which node to interact, and the change of the transfer probability matrix interference the topology of the network. The matrix W can be used for action selection. The final attained matrix w can be seen as the topology of the network, and update process can be regarded as the construction process of the network.

Algorithm 1: Protein Interaction Network Construction Using Reinforcement Learning Approach

Input: nodes N
Output: transfer probability matrix X

```
initialize graph G
initialize weight matrix W
Repeat
  foreach node i∈N
    select node j, j∈N
    connect node i and node j
    if interaction ij is incorrect then
      w_ij ← −∞
    else if interaction ij ∈ S then
      compute and get new value function V'
      if V' < V then
        V ← V'
        update W
      end if
    end if
  end for
until terminal
return W
```

VI. EXPERIMENT AND ANALYSIS

Prostate cancer is one of the most common cancers worldwide. In 2008, it is the top one cancer for men according to United States statistics. In recent years, with aging population and changes in lifestyle, a significant growth trend in the incidence of prostate cancer has become one of the main men of malignant neoplasms. We can get more than 110,000 texts from PubMed, and the trend is increasing. As we can see, prostate cancer has always been an important biomedical research.

A. Workflow and Data Acquisition

Protein interaction text mining refers automatically dig out protein interactions from biomedical texts. The process can be divided into 5 steps: text data acquisition, protein named entity recognition, relationship extraction, networks establishing and network visualization.

PubMed contains over 22 million from MEDLINE and life science journals and biomedical literature online books, including life sciences, behavioral sciences, chemistry, biology and other fields. These free resources provide researchers a wealth of resources to carry out biological text mining research.

We use E-utilities provided by PubMed as an application programming interface to get 80,841 texts from PubMed with keyword "prostate cancer". The downloaded texts are used for subsequent processing.

B. Results and Analysis

We download 80,841 abstracts texts to establish a protein interaction network. There are 4,544 effective protein-protein interactions. The edges under different thresholds are as showed in Table II.

TABLE II. EDGES BY DIFFERENT THRESHOLDS

Threshold	Edges
4.5	406
4	493
3.5	608
3	791
2.5	908
2	1327

By analyzing the distribution of nodes, we can see that the distribution of network nodes meet the degree distributions of scale-free networks, as showed in Figure 3.

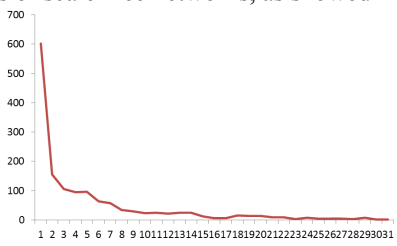

Figure 3. Node degree distribution of established protein interaction network

We check the interaction in HPRD, IntAct and STRING. Table III shows the different edges, number of unmatched edges and overall matching rates under different edge weight thresholds.

TABLE III. THE DIFFERENT EDGES, NUMBER OF UNMATCHED EDGES AND OVERALL MATCHING RATES UNDER DIFFERENT EDGE WEIGHT THRESHOLDS.

Threshold	Edges	Unmatched edges	Matching rate（%）
4.5	416	71	82.93
4	494	97	80.36
3.5	608	136	77.63
3	791	176	77.74
2.5	908	181	80.07
2	1327	207	84.40

We can get more than 77% of the matching rate by the proposed method with different thresholds. It can be seen that higher thresholds bright out significant improvement in matching rate. However, this does not imply that the performance is getting better. It is mainly due to the reduction of the total number under relatively high threshold value. We can be also noted that when the threshold values are 2.5 and 2, the matching rates go up sharply. This is because that the overall performance of the proposed method is good, and decreasing threshold will add a small amount of additional edges that can be found in HPRD, IntAct and STRING. As a result, although unmatched edges increase, the more newly added edges are matching edges.

VII. CONCLUSION

With in-depth research on biology, people gradually realized, complex biological functions and biological phenomena, are the basic unit of the complex interactions between various elements. Building interaction networks and understanding life functions with systems biology view is getting more and more recognized.

In this work, we, within reinforcement learning framework, introduce an algorithm for interaction networks establishing. The nodes are considered as proteins, and the edges are viewed as interactions. During the process of network evolution, a node selects which other node to be interacted with. Iterative evolution forms an optimal interaction network.

As prostate cancer is one of the most highly malignant neoplasms, we use the proposed approach to build up a prostate cancer protein interaction network, based on the texts attained from PubMed. Network topology analysis of the results shows that node degree distribution of established network is consistent with scale-free properties.

Meanwhile, the matching of edge in protein databases only indicates that the two proteins are interacted. There are still much work on analyzing the network, pathways to reveal its role in prostate cancer.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (61303108, 61472262, 31470821), Jiangsu(13KJB520020), Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University(93K172014K04).

REFERENCES

[1] Ahmedin Jemal, DVM, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, David Forman. Global Cancer Statistics[J]. CA: A Cancer Journal for Clinicians, 2011, 61(2): 69-90.

[2] http://www.ncbi.nlm.nih.gov/pubmed/.

[3] Jorrit J Hornberga, Frank J Bruggemana, Hans V Westerhoffa. Cancer: a Systems Biology disease[J]. Biosystems, 2006, 83(2-3): 81-90.

[4] Yen-Ching CR, Tzong-Han Tsai, Wen-Lian Hsu. New Challenges for Biological Text-Mining in the Next Decade[J]. Journal of Computer Science and Technology, 2010, 25: 169-179.

[5] Karin M Verspoor, Judith D Cohn, Komandur E Ravikumar, Michael E Wall. Text Mining Improves Prediction of Protein Functional Sites[J]. PLoS One, 2012, 7(2): 1-16.

[6] Hayasaka S, Hugenschmidt CE, Laurienti PJ. A network of genes, genetic disorders, and brain areas[J]. PLoS One, 2011, 6: e20907.

[7] Sharma P, Senthilkumar RD, Brahmachari V, Sundaramoorthy E, Mahajan A, et al. Mining literature for a comprehensive pathway analysis: a case study for retrieval of homocysteine related genes for genetic and epigenetic studies[J]. Lipids Health Dis, 2006 5: 1.

[8] Richard S Sutton, Andrew G. Barto. Reinforcement Learning[M]. MIT Press, 1998.

[9] Van Hasselt H. Reinforcement Learning: State of the Art [M]. Berlin: Springer, 2007, 207-251.

[10] Farhang Sahba, Hamid R Tizhoosh, Magdy M A Salama. A Reinforcement Learning Framework for Medical Image Segmentation[C]. In: Proceedings of IEEE International Joint Conference on Neural Networks, 2006, pp.511-517 .

[11] Alison Watts. A Dynamic Model of Network Formation[J]. Games and Economic Behavior, 2001, 34: 331-341.

[12] Arunachalam Vinayagam, Ulrich Stelzl, Raphaele Foulle, et al. A Directed Protein Interaction Network for Investigating Intracellular Signal Transduction[J]. Science Signaling, 2011, 4(189):rs8.

[13] E D Coelho, J P Arrais, J L Oliveira. From protein-protein interactions to rational drug design: are computational methods up to the challenge?[J]. Curr Top Med Chem, 2013, 13(5):602-618.

[14] Xuebing Wu, Rui Jiang, Michael Q Zhang, et al. Network Based Global Inference of Human Disease Genes[J]. Molecular Systems Biology, 2008, 4(189): 1-11.

[15] Albert László Barabási, Natali Gulbahce, et al. Network Medicine: A Network-Based Approach to Human Disease[J]. Nature, 2011, 12(1): 56-68.

[16] Albert László Barabási. Scale-Free Networks: A Decade and Beyond[J]. Science, 2009, 325: 412-413.

[17] Zhu Fei, Shen Bairong. Combined SVM-CRFs for Biological Named Entity Recognition with Maximal Bidirectional Squeezing[J]. PLoS One, 2012, 7(8): 1-8.

[18] Xiao Li, Haoyang Cai, Jiabao Xu, et al. A mouse protein interactome through combined literature mining with multiple sources of interaction evidence[J]. Amino Acids, 2010, 38(4): 1237-1252.

[19] Ahmed J, Meinel T, Dunkel M, et al. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge[J]. Nucleic Acids Research, 2011, 39:960-967.

[20] W Hayes, K Sun, N Przulj. Graphlet-based measures are suitable for biological network comparison[J]. Bioinformatics, 2013, 29(4): 483-491.

[21] Marc Vidal, Michael E. Cusick, Albert-László Barabási. Interactome Networks and Human Disease[J]. Cell, 2011, 144(6): 986-998.

[22] Richard Durrett. Random graph dynamics[M]. Cambridge University Press, 2007.

[23] HPRD, http://www.hprd.org/.

[24] Suraj Peri, J Daniel Navarro, Ramars Amanchy, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans[J]. Genome Research, 2003, 13(10): 2363-2371.

[25] IntAct, http://www.ebi.ac.uk/intact/main.xhtml.

[26] S Kerrien, Y Alam Faruque, B Aranda, et al. IntAct-open source resource for molecular interaction data[J]. Nucleic Acids Research, 2007, 35(Suppl 1): 561-565.

[27] STRING, http://string.embl.de/.

[28] Lars J Jensen, Michael Kuhn, Manuel Stark, et al. STRING 8-a global view on proteins and their functional interactions in 630 organisms[J]. Nucleic Acids Research, 2009, 37(Suppl 1): 412-416.