An Inference Model of Medical Insurance Fraud Detection: Based on Ontology and SWRL†

Xiao-Bo Tang*, Wei Wei**, Guang-Chao Liu*** and Juan Zhu****

*/**/***Center for the Studies of Information Resources, Wuhan University, Wuhan, Hubei, China
*<xbtang@126.com>, **<503175355@qq.com>, ***510675457@qq.com>,
****Jiujiang University, Jiujiang, Jiangxi, China, <zhujuan@whu.edu.cn>

Xiao-Bo Tang is a professor of information science at the Information Management School of Wuhan University in China. He holds a PhD in management science and engineering from Wuhan University and a master’s degree in computer science and technology from Wuhan University of Hydraulic and Electrical Engineering. He is the director of the Information Systems Research Center in Wuhan University, member of the Association for Information Systems, the director of the China Branch of the Association for Information Systems, and the executive director of the Information Society of Hubei Province. His main research interests include knowledge organization and intelligence analysis.

Wei Wei has been working as a lecturer at Wuhan University since 2007 and she has been a PhD student at the Information Management School of Wuhan University since September 2014. She attained her master’s and bachelor’s degrees in management information systems. Her research interests lie in knowledge organization and intelligence service. She is currently working on knowledge representation and inference applied to ontology in the medical domain as part of her PhD thesis.

Guang-Chao Liu is a doctoral student at the Information Management School of Wuhan University since September 2015. He completed his MS in financial management from the Lubin School of Business, Pace University. He is currently working on text mining in the financial domain at the semantic level. His research interests also include sentiment analysis and knowledge discovery.

Juan Zhu has been a lecturer at Jiujiang University and she has been a doctoral student at the Information Management School of Wuhan University. As part of her PhD thesis, she is currently working towards building knowledge bases to represent the user entity and the item entity for recommender systems. Her research interests also include knowledge organization and intelligence service.

Tang, Xiao-Bo, Wei Wei, Guang-Chao Liu and Juan Zhu. 2017. “An Inference Model of Medical Insurance Fraud Detection: Based on Ontology and SWRL.“ *Knowledge Organization* 44(2): 84-96. 23 references.

Abstract: Medical insurance fraud is common in many countries’ medical insurance systems and represents a serious threat to the insurance funds and the benefits of patients. In this paper, we present an inference model of medical insurance fraud detection, based on a medical detection domain ontology that incorporates the knowledge base provided by the Medical Terminology, NKIMed, and Chinese Library Classification systems. Through analyzing the behaviors of irregular and fraudulent medical services, we defined the scope of the medical domain ontology relevant to the task and built the ontology about medical sciences and medical service behaviors. The ontology then utilizes Semantic Web Rule Language (SWRL) and Java Expert System Shell (JESS) to detect medical irregularities and mine implicit knowledge. The system can be used to improve the management of medical insurance risks.

Received: 21 August 2016; Revised: 13 January 2017; Accepted 18 January 2017

Keywords: medical insurance frauds, inference model, ontology,

† This study is part of the State National Sciences Foundation Project “Research on Integrated Retrieval and Semantic Analysis Method of Social Media“ (No. 71273194).

1.0 Introduction

False medical claims, excessive treatment, forgery of medical records and case history, and other methods have been commonly employed to illegally obtain medical insurance funds from many countries' medical insurance systems. Those actions can be deemed as medical insurance frauds. According to Zuo (2014), frauds may be committed by medical employees, beneficiaries, or by the collusion of both. As its medical insurance system continues to develop, China has been well on its way to achieving the goal of universal access to health care. However, a new problem has emerged that participants attempt to take advantage of the system's loopholes. In order to identify medical insurance frauds, we propose an inference model of medical insurance fraud detection based on ontology and SWRL technologies. As Cristiane (2016) explained, ontology is an effective method for representing the knowledge of a given field. It can provide an extremely powerful, meta-level resource that uses inference to analyze explicit knowledge in the ontology. The model presented in this study is aimed at detecting medical irregularities by mining implicit knowledge based on the characteristics of medical insurance. It can be used to improve the quality of medical risk management and prediction.

The remainder of this article is arranged as follows: section 2 is a literature review that summarizes background knowledge related to medical insurance frauds, ontology, OWL, SWRL and JESS. Section 3 describes the general design goals and framework of the system. Section 4 describes the construction of the ontology, and section 5 describes the working of SWRL rules and JESS inference engine. The method is tested in section 6 using information from 200 patient cases to detect medical irregularities from their medical treatment records and validate results. Section 7 consists of the concluding remarks and problems for future research.

2.0 Literature review

2.1 Medical insurance frauds

Medical insurance frauds include any behavior aimed at gaining illegal access to medical insurance funds by using forged information or concealing truth. In China, the primary cause of medical insurance frauds is the "third-party payment" system. Specifically, the payment for the providers of medical services is not directly paid by the beneficiaries of said services, but entirely by third-party medical insurance institutions. The frauds are motivated by the financial gain of medical service providers and beneficiaries.

As presented by Liu (2014), medical insurance frauds can be classified into internal frauds and external frauds. Internal frauds are frauds committed by officials or employees of medical insurance or governmental institutions, and external frauds are those committed by beneficiaries or employees of medical service providers. Internal frauds are essentially embezzlement of the insurance fund, and outside the scope of this study, which focuses on external frauds. The main forms of external frauds include:

– Frauds committed by beneficiaries
– Frauds committed by employees of medical service providers
– Collusion between beneficiaries and employees of providers

In principle, the medical insurance system is created to mitigate the beneficiaries' financial damage caused by health risks. But this does not prevent some beneficiaries from attempting to exploit the system. Common behaviors include imposture by non-participants of medical insurance, providing non-factual information in order to receive excessive or unnecessary services, and forgery of medical records for indemnification. Providers of services consist of designated pharmacies and designated health care institutions. As shown by Zhang (2014), frauds committed by employees of the provider account for the largest proportion of medical frauds. These mainly take the forms of admitting patients who do not need hospitalized treatment, excessive services including inspections, treatments, and drug usage, and illegal transactions of documents and prescriptions related to medical insurance. It is also possible for providers and beneficiaries to conspire in medical frauds. These include issuing fictitious items that require insurance payment, reporting diseases not applicable for insurance (e.g., traffic accidents, fights, etc.) as applicable, and forgery of medical information (Li 2006).

2.2 Ontology

Ontology is a formal description of concepts and their relations in a specific domain (David et al. 2009), consisting of concepts, attributes, and instances (Chakkrit and Michael 2009). Ontologies are built to share and reuse knowledge using artificial intelligence and machine language and to facilitate communications between humans and computer systems. They have been widely utilized for information retrieval and knowledge organization. Due to the complexity and diversity of medical knowledge, building the domain ontology of medical sciences would enable us to establish a knowledge set with unambiguous specifications and semantic features. This set can describe medical concepts, their relationships, and the general principles of medical science.

Much work has been done on ontologies for medical sciences. The Unified Medical Language System (UMLS) (http://www.nlm.nih.gov/research/umls/) was developed by the National Library of Medicine of the United States to solve the problem of different medical sources having different expressions for the same concept, and is capable of providing semantic unification of medical concepts. The Open Biomedical Ontology (OBO) was built by Barry et al. (2007) from the National Institute of Health of the United States, and consists of multiple sets of knowledge related to biology and medicine, including the gene ontology (GO), plant ontology (PO), cell type ontology (CT), sequence ontology (SO) and more. In China, researchers have developed systems including NKIMed (Zhou 2003), the Traditional Chinese Medical Language System (Zeng and Wang 2006), and the Chinese medical information semantic indexing system and semantic retrieval model (Li and Pang 2003). In general, the theories on medical ontologies have become increasingly more mature, and the practical application of medical ontology is also attracting widespread attention.

2.3 OWL

According to Mercedes (2008), OWL (Web Ontology Language) is a framework proposed by the World Wide Web Consortium (W3C), and a recommended language for semantic web-based ontology. It is an extension of RDF (Resource Description Framework) that builds on XML-based RDF syntax structure and is designed for description logic. OWL is fully compatible with RDF, complements RDF's deficiency in describing relationships, and is better suited for expression and inference. Based on purposes, three different levels of OWL can be used: OWL Full, OWL DL and OWL Lite (http://www.w3.org/TR/owl-features/).

2.4 SWRL

SWRL (Semantic Web Rule Language) was proposed to complement OWL's deficiency in ontology inference. As presented by Jeff (2005), the current W3C standard provides a semantic description of rules allowing users to build a rules-based ontology, and achieve semantic inference for the ontological knowledge base. By describing the relationships between rules, SWRL can improve the ontology's capability for semantic expression. Its canonical format is *antecedent* → *consequent*. "Antecedent" and "consequent" are disjunctions of atoms in ontology, and atoms are classes or attributes from the ontology.

2.5 JESS

Researchers have proposed more specialized inference engines such as Racer, FaCT and Pellet. While efficient and easy to use, they are not very expandable or compatible with other languages. This team chooses JESS for its efficiency, flexibility and excellent in porting and embedding. Developed by Friedman-Hill et al., JESS (Java Expert System Shell) is a rule engine and scripting environment written entirely in Java (Martin et al. 2005). The core JESS language is compatible with CLIPS (C Language Integrated Production System), which uses the Rete algorithm to process JESS rules. Rete is a high-efficiency algorithm used to solve many matching problems, particularly known for its speed in inference operations. JESS has inherited the advantages of CLIPS, and added some new features, such as forward inference, reverse inference, running memory check, and the ability of directly calling and operating a Java class library. JESS also supports the traditional "if … then …" grammar structure.

3.0 Research framework

3.1 Targets of detection

This study is aimed at developing a model for the expression and inference of behaviors related to medical insurance, in order to detect and prevent irregularities. As presented by Li (2015), medications, inspections, and treatments are important components of medical services, and also the main source of frauds**.** The model proposed by this study targets the behaviors detailed in Table 1.

Irregular medications	Prescriptions from service providers include excessive amounts of drugs or drugs unrelated to the disease	e.g., prescription contains multiple drugs with similar or the same effects
Excessive services	Services are unnecessary or beyond the regulations related to basic medical insurance	e.g., unnecessary inspections were performed; hospitalization for trivial diseases
Multiple prescriptions	One large prescription is divided into several smaller ones	e.g., several unnecessary prescriptions are made for the treatment of one disease

Table 1. Targets of detection.

All of the irregularities take place as patients receive medical services from the medical service institutions. For this

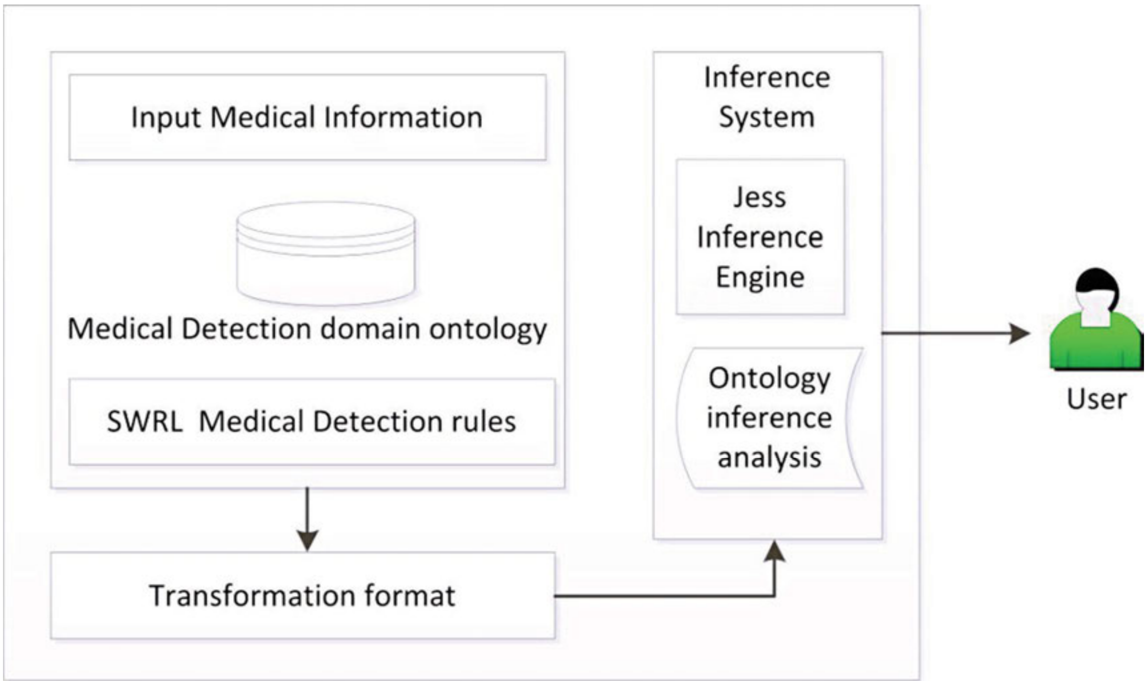

Figure 1. Framework of medical insurance fraud detection.

reason, the detection of the inference model should primarily be performed on the medical records, prescriptions, and inspection documents issued by health service providers. It is relatively difficult to discover medical insurance-related misconducts by simply matching strings or calculating numerical differences. This model instead approaches the problem by analyzing the patients' medical activities using the knowledge representation of medical behaviors and inferring the relationships intrinsic to the knowledge.

3.2 Model framework

Figure 1 shows the inference model for fraud detection.

1. The medical fraud detection ontology normalizes the knowledge of medical service behaviors and expert knowledge from the medical domain. The preprocessed data of instances are added to the ontology through a medical information system to create a knowledge base, as shown in Figure 2.

2. The inference rules of medical fraud detection are described using SWRL. The SWRL rules are built with existing classes, attributes and instances from the ontology. The system will warn users of irregularities found in the medical information, such as repeated prescriptions of the same drug, and excessive inspections.

3. The Pellet inference engine is used to convert the knowledge classes and concepts of the medical detection ontology into an acceptable format for operating on the medical instances. Because SWRL cannot be operated on directly, the SWRL rules need to be converted through XSLT (Extensible Stylesheet Language) into a format that can be executed by JESS. Additionally, Pellet can also detect inconsistencies and conflicting knowledge in the ontology (Rung et al. 2012).

4. JESS is used for the inference process, to identify which classes, axioms or instances are applicable to the current medical service behavior. OWL and SWRL are converted into a format acceptable for JESS during the process; afterwards the results of inference are written in OWL for updating the medical detection ontology.

4.0 Construction of medical detection domain ontology

In this study, we construct a medical fraud detection ontology mainly used for detecting illegal behaviors related to the medical insurance industry by beneficiaries and service providers. The ontology utilizes Medical Terminology (Medical Term Validation Committee 1998), NKIMed, and Chinese Library Classification (Editorial board of the National Library 2010) to define the terminologies and concepts in the field of medical detection. Terminologies are the basic elements of the ontology, used to describe the knowledge representations of the concepts. Table 2 shows a selection of important terms used in this study.

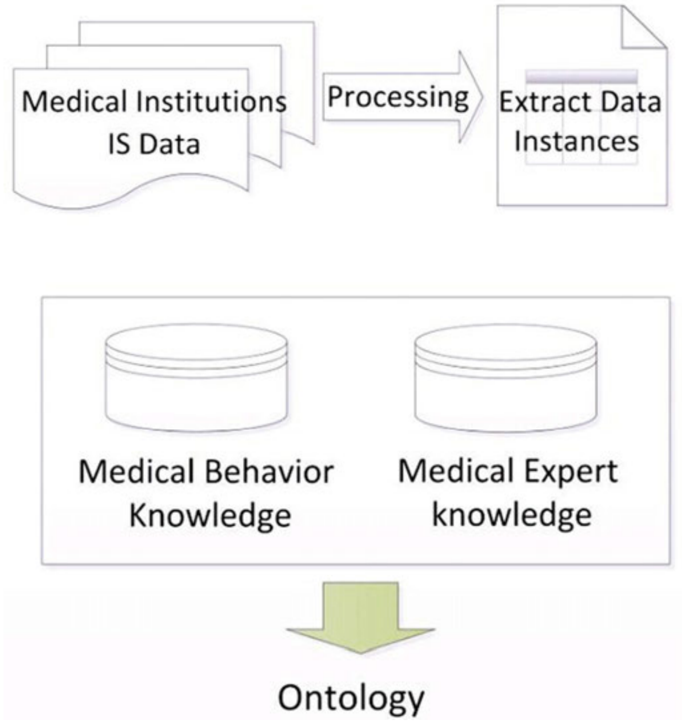

Figure 2. The development of domain ontology.

Medical information	doctors, patients, diseases, drugs, symptoms, inspections, treatments
Diseases	endocrine disease, respiratory disease, cardiovascular diseases, digestive diseases, viral diseases, dermatology diseases, ophthalmology disease, nervous diseases, bacterial infections and fungal diseases, otorhinolaryngology diseases, cancer
Drugs	chemical medicine, Chinese patent medicine, Chinese herbal medicine, endocrine system medicine, oral medicine, respiratory medicine, cardiovascular medicine, digestive medicine, vitamin drugs, codeine phosphate, aspirin, Proglumide
Inspections	general clinical test, medical imaging, ultrasonic inspection, blood pressure check, blood inspection, bone marrow inspection, urine inspection, CT scan, B-scan ultrasonography
Treatments	surgical treatment, medication, chemotherapy, physic therapeutics, cinesiatrics, acupressure treatment, acupuncture, dietary therapy, blood transfusion, biotherapy
Symptoms	headache, astriction, haemorrhage, gasp, hematochezia, sharp pain, back pain, lumbago, vertigo, macula, arthronalgia, frequent micturition, fever, claudication, odontalgia
Irregularities	split prescription, excessive services, irregular medication, similar drug efficacy, excessive prescriptions

Table 2. Some important terms in the medical insurance fraud detection domain ontology.

The ontology is constructed using Protégé, a Java-based, open source graphical application, developed by Stanford University Information Center for accessing, creating and maintaining ontologies. Protégé uses RDF to describe the relationships between web pages and other resources, and to support description of ontologies' main components, classes and attributes. The main components of Protégé are OWL classes, properties, and individuals and so on. Classes are the semantic representations of concepts in knowledge. Properties describe the relationships of class linked to the basic data types, and the relationships can be connected to other classes or instances. Individuals represent objects in a domain that can be considered as instances of classes. Property here is a bilateral relationship between two individuals, or can be considered as a bridge between the two individuals, making it different from the attributes in an object-oriented programming language. Protégé has good support for ontology languages including RDF, OWL, and Schema XML. Being Java-based, it ensures the expandability of the ontology's application environment, and provides fast compiler for ontology construction.

Figure 3 shows the medical detection ontology built in Protégé, with the classes and their hierarchical structure to the left, and the specified constraints to the right.

We adopted the knowledge engineering method (Natalya and Deborah 2006) to construct the medical detection domain ontology. Knowledge engineering is an approach

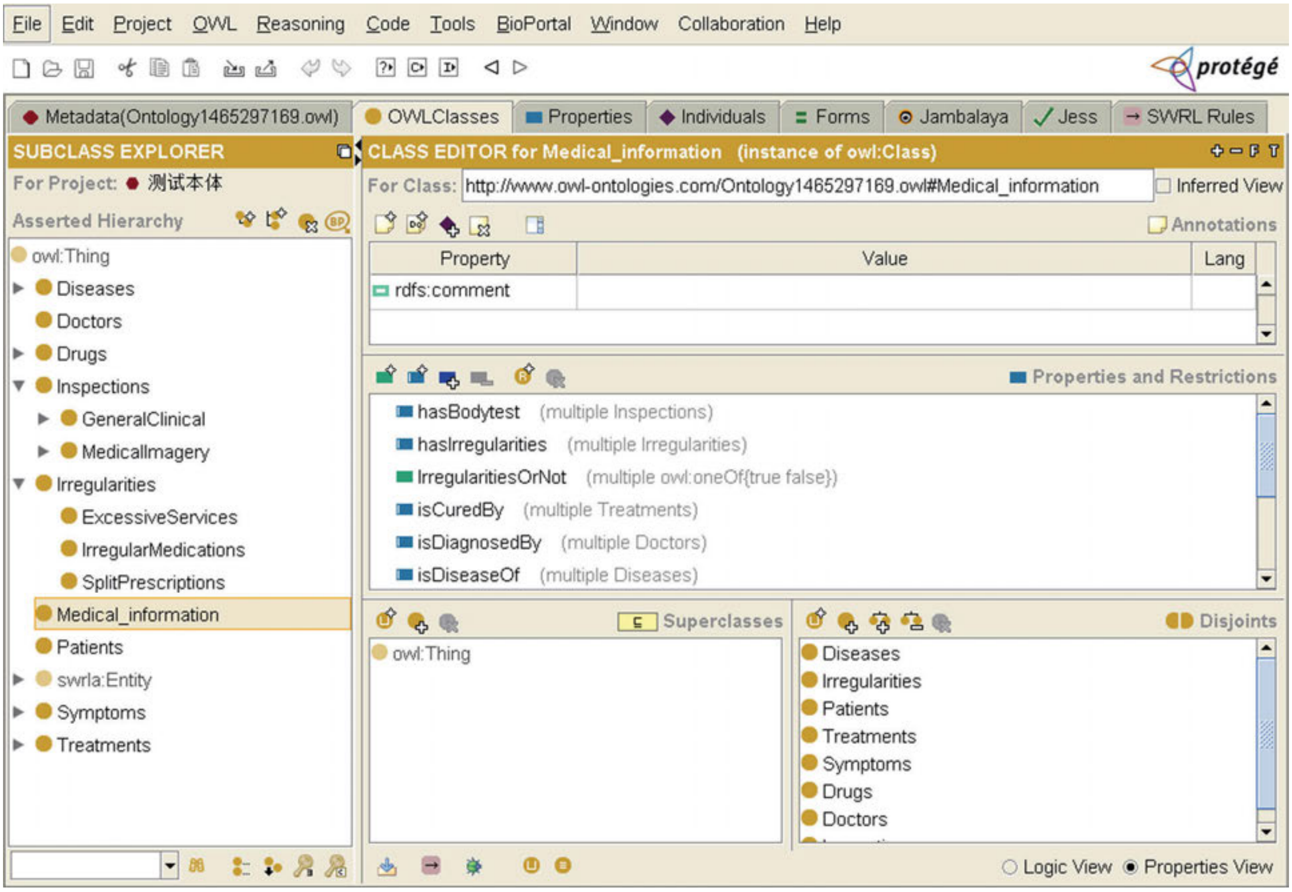

Figure 3. Medical detection domain ontology created in Protégé.

to ontology development for Protégé proposed by Natalya et al. at Stanford University. They believed that there is no absolutely correct way to model a domain, all solutions must be adapted to practical application, and the process of ontology development is one of continued iteration. The method consists of the process of determining the ontology's domain and scope, enumerating essential concepts, defining classes and their hierarchical structure, defining attributes, creating instances in the ontology, and evaluating the ontology. The following discussion describes the construction of the medical detection ontology in this study.

4.1 Defining classes

Classes are the core of ontology, providing an abstract description of the physical objects in the domain. The establishment of the medical insurance fraud detection domain ontology is based on current medical domain knowledge, allowing us to define the general concepts first, and then narrow down the definitions to suit our need. The classes and hierarchical structure of the ontology should provide a comprehensive coverage of topics related to behaviors in medical services, before its issues and constraints can be identified. The design of the structure was done by consulting medical experts and standards including *Chinese Library Classification* and NKIMed. A total of nine main classes are used: medical information, doctors, patients, symptoms, inspections, diseases, treatments, drugs and irregularities. Each concept is then classified and arranged into a hierarchical structure with super-classes and sub-classes (Subhashis and Sayon 2016). Figure 4 shows the medical detection domain ontology model.

4.2 Defining attributes

The definition of classes and their hierarchy alone cannot represent all knowledge information in a field, as their internal structures must also be described by defining attributes. Object attributes and datatype attributes are two important attribute types in an OWL ontology. Object attributes represent the mutual relationships between two classes, and a class's attributes are linked to instances through its domain and range. Datatype attributes are used to designate a class's unique attributes, and represent the relationship between instances and data. In this study, the important attributes of the medical detection domain ontology include the purpose of detection, treatment

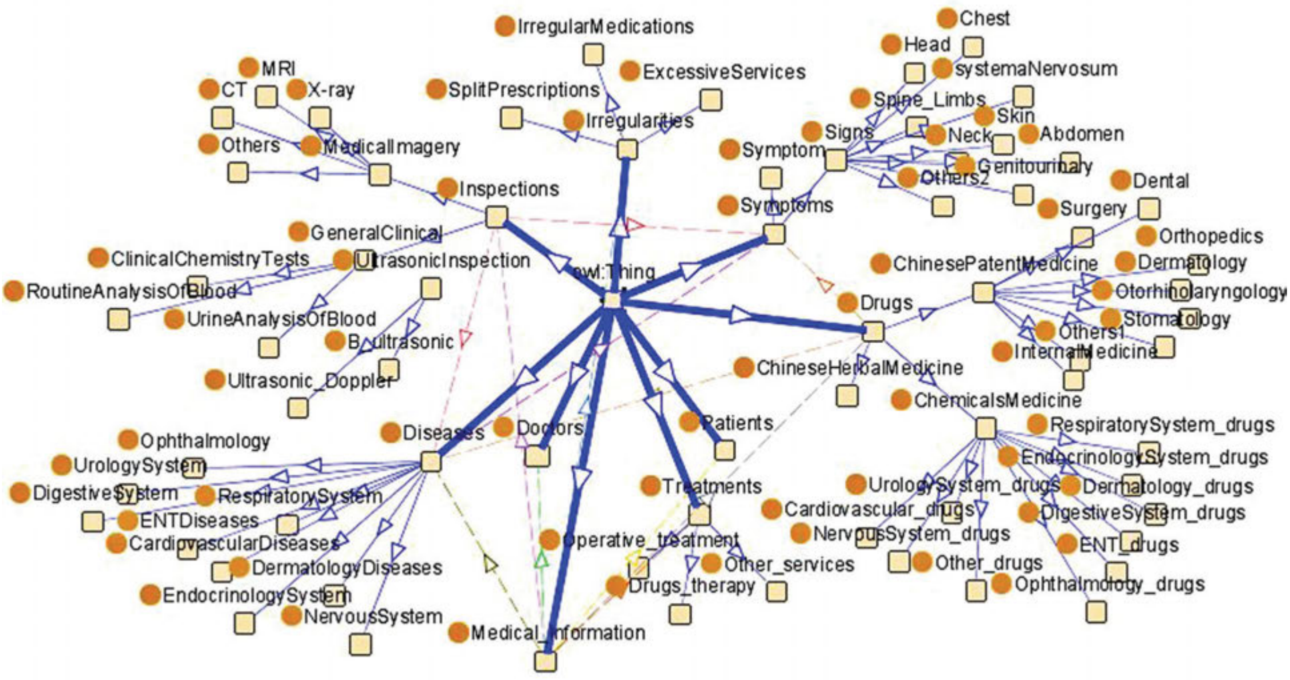

Figure 4. The model of medical detection domain ontology.

measures, and related diseases, which are object attributes, and admission IDs, times of admission, and patient names, which are datatype attributes.

4.3 Adding instances

A class can contain multiple instances, while one instance can belong to one or more classes. In this study, instances were added according to the need of the experiment. Medical information instances were created for testing the system's inference rules.

4.4 Integrity and consistency test

A robust ontology is fundamental for the proper execution of the inference model. The Pellet inference engine can detect conflicts in the ontology, and was used to test the integrity and consistency of the ontology, as shown in Figure 5.

The construction of ontology is by necessity a complex process that requires cooperation between experts and knowledge engineers. Iterative testing and improvement are needed to gradually add details and optimize the ontology. The ontology constructed in this study is only used as an experimental model to verify the validity of the inference system.

5.0 The inference of medical detection domain ontology

Inference means obtaining the implicit knowledge from a given set of knowledge, and ontology inference involves extracting the implicit knowledge from explicit knowledge defined in ontology. Ontology inference can serve functions including semantic-based query, knowledge inference, and ontology checking. It is based on the idea of description logic. Description logic is a logic-based knowledge representation, and a decidable subset of first-order logic, that can describe the provability of a given statement within limited time. It consists of concepts, attributes or roles, and individuals. The basic problems of description logic inference include the inclusion relation of concepts, the satisfiability problem, and entity detection (Shi and Sun 2006).

Adela et al. (2009) show that much work has been done on inference engines for description logic-based inference. FaCT, Racer and Pellet are several inference engines based on OWL. Figure 6 depicts the workflow of inference in this study. The main components of the inference engine are the parser, the tuples, and the rules. First, the engine reads the ontology in OWL format. The OWL parser then converts the ontology from OWL into the tuples format of < subject, verb, object >. Rules are restrictions that focus the ontology on a particular domain. Finally, the ontology of classified OWL is output by the description logic engine.

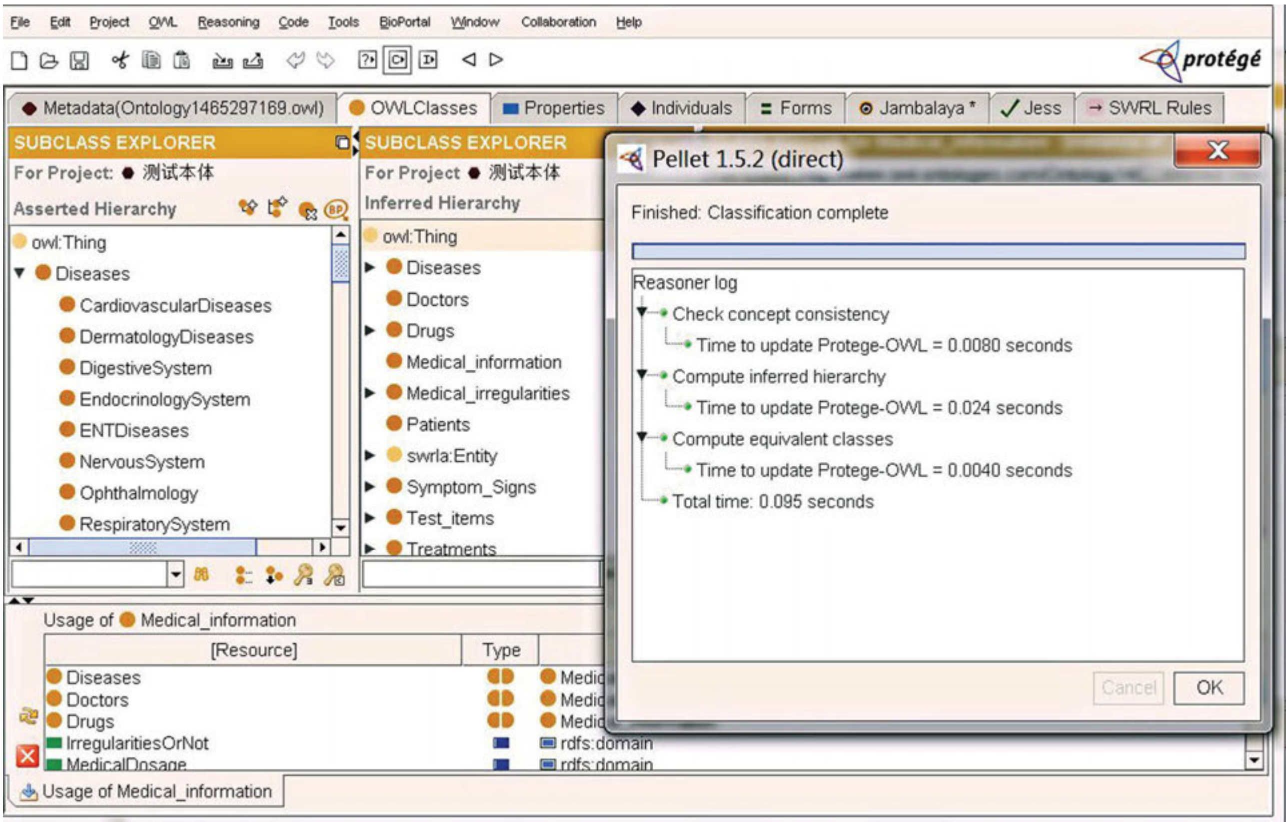

Figure 5. Consistency test in Pellet.

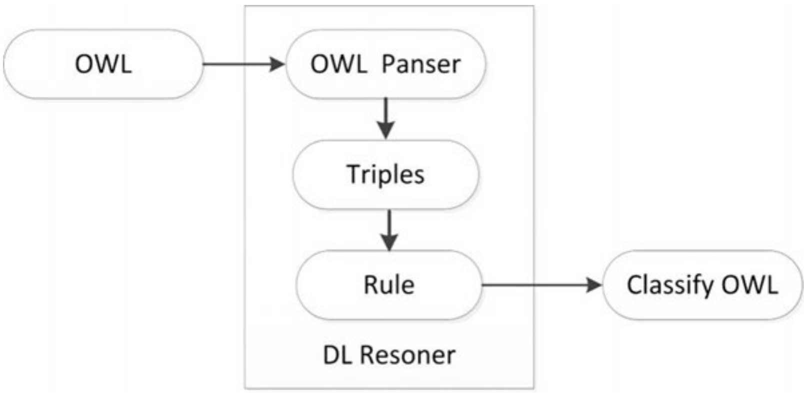

Figure 6. The workflow of inference engine.

Our study uses SWRL format for the inference rules of medical fraud detection, which have XML-based syntax. SWRL rules can supplement the OWL ontology with robust semantic expressions, creating inference rules based on the instances and attributes in the ontology. SWRL rules can directly reference the ontology's classes, attributes and relationships. For example, the following relationships have been defined in the ontology:

$$hasEfficacyOf(x_1, y_1)$$
$$hasEfficacyOf(x_2, y_1)$$

Suppose an efficacy relationship has drugs as its domain, and diseases as its range, with the knowledge that two different drugs x_1 and x_2 can both cure disease y_1, therefore x_1 and x_2 have the same efficacy, described as:

$$hasEfficacyOf(?x_1,?y_1) \wedge hasEfficacyOf(?x_2,?y_1) \wedge differentForm(x_1,x_2) \rightarrow hasSimilarEfficacy(x_1, x_2)$$

SWRL allows us easy editing of inference rules using instances and attributes in the ontology through the SWRL Tab plug-in in Protégé. However, as the inference process

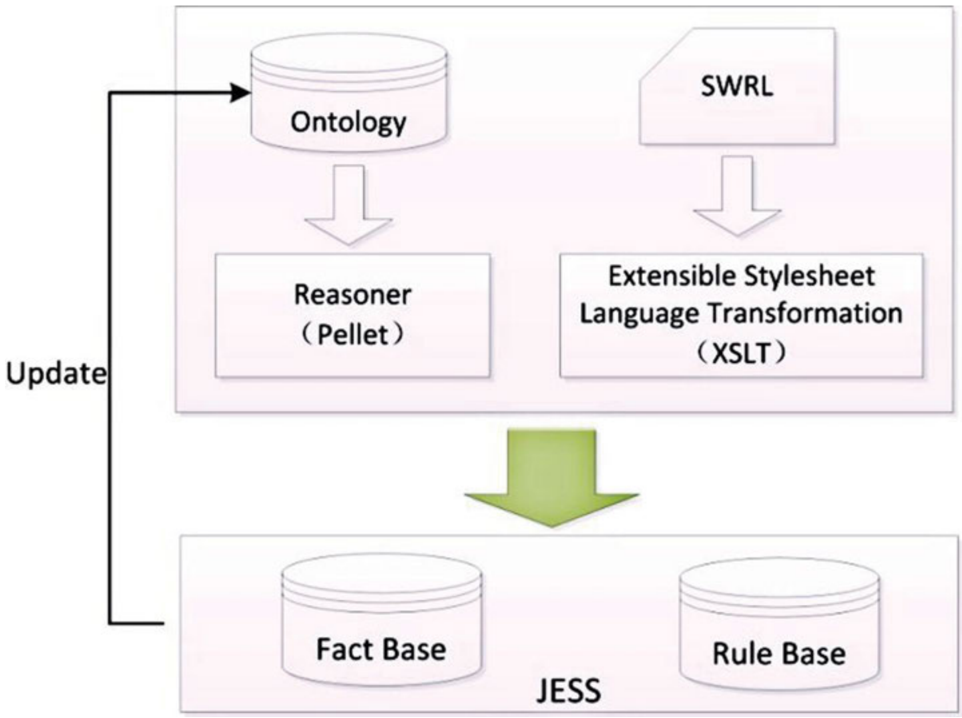

Figure 7. Workflow of JESS inference format conversion.

cannot be performed directly from Protégé, it is necessary to use the ontology knowledge facts as the basis, and infer the rules through the rules engine.

For this study we chose JESS (Java Expert System Shell) as the inference engine due to its good portability, embeddedness and high efficiency. While JESS cannot directly read ontologies from the OWL format or rules from the SWRL format, the inference can be done after performing format conversions. Figure 7 shows the format conversion and inference process for JESS. First, the ontology is converted to JESS by Pellet inference engine. Second, the SWRL rules are converted to JESS by XSLT (Stylesheet Language Extensible), and the inference results are written to the OWL to update the ontology.

6.0 Experiment and discussion

6.1 Constructing the medical insurance fraud detection system

We used Protégé to build the medical insurance fraud detection ontology, with the hierarchical structure shown in the previous sections. Figure 3 depicts the Protégé OWL ontology editor. The root node of the ontology is "owl: thing," and its child nodes consist of the nine subclasses: medical information, doctors, patients, symptoms, inspections, diseases, treatments, drugs and irregularities. Two hundred medical treatment records were established to verify the effectiveness of the rules inference. Each medical record contains the patient's conditions (the patient's basic information as well as some physiological parameters) and the prescriptions given by the doctor, and the final inference results will be compared with the experts from Medical Insurance Regulatory.

By selecting the SWRL Tab plug-in, the SWRL rules can be edited in Protégé on the OWL platform. This study uses JESS as a rules engine to embed the Protégé platform to perform rules inference. JESS software consists of the rules base, the fact base and the execution engine (Jing and Elena 2004). Figure 8 shows SWRL's operation interface. The two rectangular boxes with dotted lines are the SWRL rules editor and the JESS operation interface. The rules are written in the editor, and the JESS operation interface provides the triggers for calling the rules engine.

We developed four SWRL rules for medical fraud detection. The rules can be expressed as the following.

RULE 1:
If drug x1 can cure disease y1, and drug x2 can also cure disease y1, then they have similar efficacy, and a case of irregular medication is detected.

Medical_information(?I) $\wedge$ *Drugs(?*x_1*)* $\wedge$ *UseDrugs(?I, ?*x_1*)* $\wedge$ *Drugs(?*x_2*)* $\wedge$ *UseDrugs(?I, ?*x_2*)* $\wedge$ *Symptoms(?*y_1*)* $\wedge$ *hasEfficacyOf(?*x_1*, ?*y_1*)* $\wedge$ *hasEfficacyOf(?*x_2*, ?*y_1*)* $\rightarrow$ *hasIrregularities(?I,SimilarEfficacy)* $\wedge$ *IrregularitiesOrNot(?I, true)*

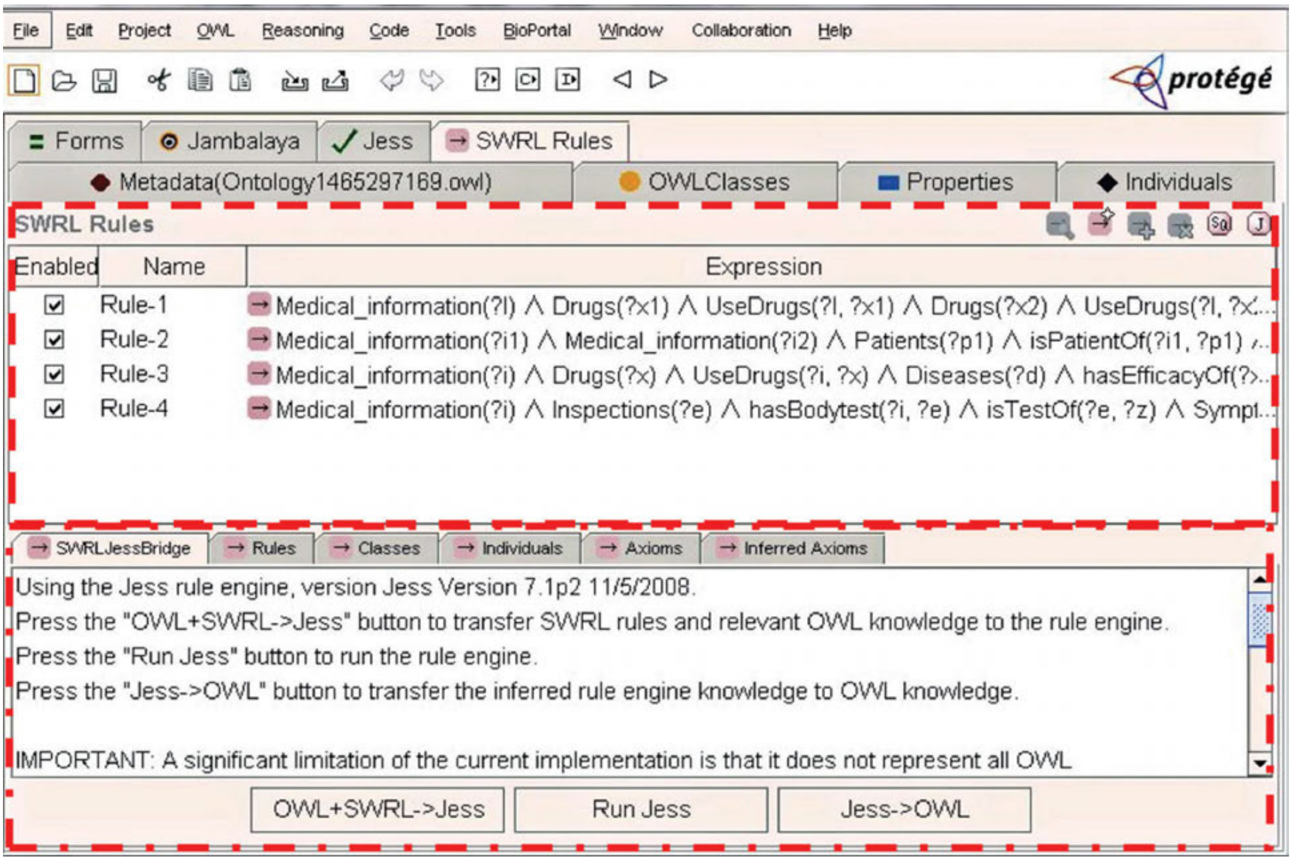

Figure 8. The joint interface of SWRL rule editing and JESS operation.

RULE 2:
If the actual dosage of a drug exceeds its proper dosage, then a case of excessive prescription is detected.

Medical_information(?i) ∧ Drugs(?x) ∧ UseDrugs(?i, ?x) ∧ Diseases(?d) ∧ hasEfficacyOf (?x, ?d) ∧ isDiseaseOf(?i, ?d) ∧ MedicalDosage(?i, ?y) ∧ Dosage(?x, ?z) ∧ swrlb:lessThan (?z, ?y) → hasIrregularities(?i, ExcessivePescriptions) ∧ IrregularitiesOrNot(?i, true)

RULE 3:
If the medical information entry i contains separate inspections for clinical responses and disease determination, then a case of excessive inspection is detected.

Medical_information(?i) ∧ Test_items(?e) ∧ hasBodytest(?i, ?e) ∧ isTestOf(?e, ?z) ∧ Symptoms(?y) ∧ hasClinicalResponse(?i, ?y) ∧ Diseases(?d) ∧ isDiseaseOf(?i, ?d) ∧ differentFrom(?z, ?y) ∧ differentFrom(?z, ?d) → hasIrregularities(?i,ExcessiveInspections) ∧ IrregularitiesOrNot(?i, true)

RULE 4:
If the medical information entries i1 and i2 for patients p1 and p2 have the same citizen's ID number, then a case of multiple split prescriptions for one beneficiary is detected.

Medical_information($?i_1$) ∧ *Medical_information*($?i_2$) ∧ *Patients*($?p_1$) ∧ *isPatientOf*($?i_1, ?p_1$) ∧ *Patients*($?p_2$) ∧ *isPatientOf*($?i_2, ?p_2$) ∧ *PatientID*($?p_1, ?id$) ∧ *PatientID*($?p_2, ?id$) ∧ *Diseases(?d)* ∧ *isDiseaseOf*($?i_1, ?d$) ∧ *isDiseaseOf*($?i_2, ?d$) → *hasIrregularities*($?i_1$, *MultiplePrescriptions*) ∧ *hasIrregularities* ($?i_2$,*MultiplePrescriptions*) ∧ *IrregularitiesOrNot*($?i_1$, *true*) ∧ *IrregularitiesOrNot* ($?i_2$, *true*)

Through the implementation of the "J" button, the JESS engine is triggered. Figure 9 shows the integration of OWL ontology and SWRL, and the inference results from JESS. The results show that the process had selected 4 rules, 64 classes and 59 individuals.

It can be seen that nine axioms have been added, which are the results of JESS inference. As shown in Figure 10, the instance "Medical_information_01" has its attribute "hasIrregularities" assigned the value "true," with the attribute "SimilarEfficacy" attached. This shows the validity of the inference.

The mechanism of medical data detection is based on medical fraud detection domain ontology. The medical

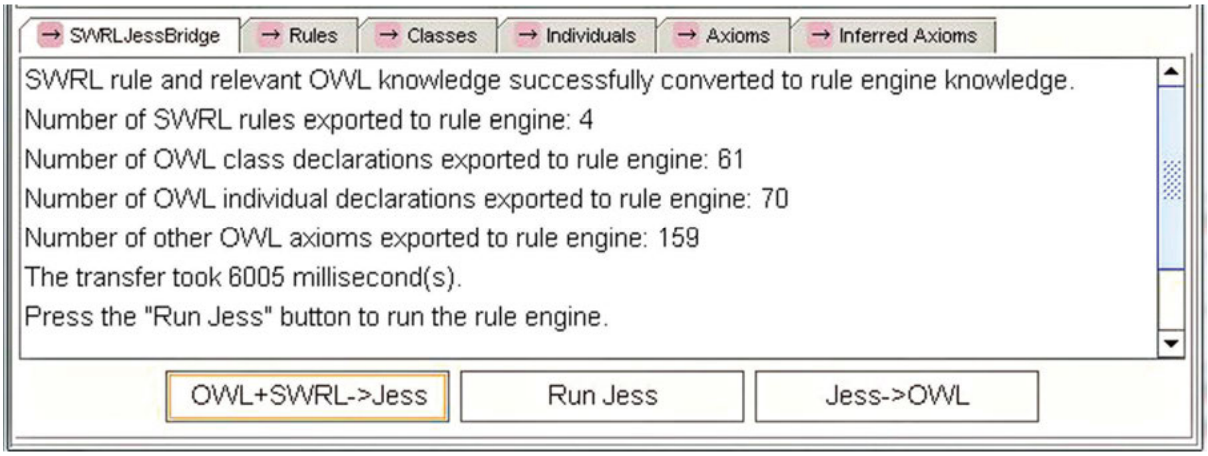

Figure 9. The inference result of OWL+SWRL→Jess.

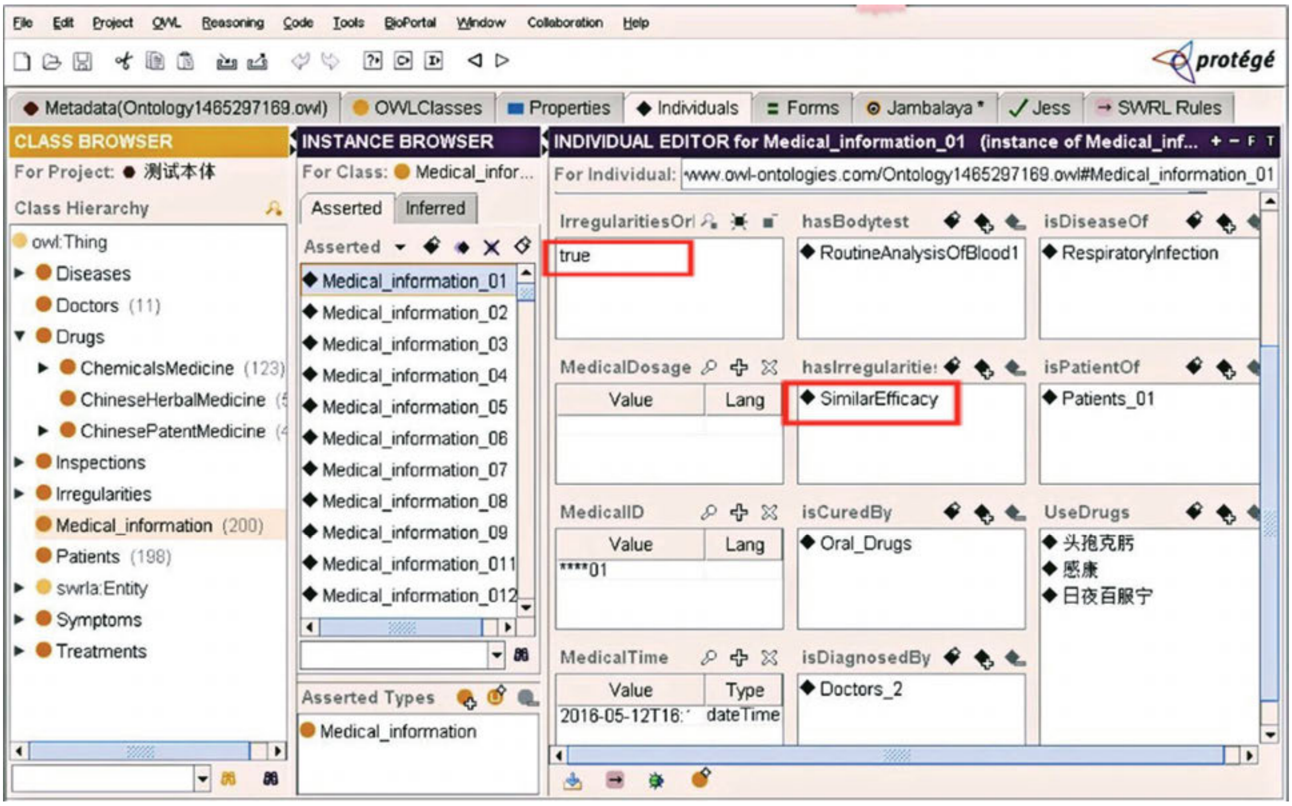

Figure 10. The inference result of medical fraud detection ontology.

treatment data can be processed based on the rules of detection. Finally, the obtained explicit knowledge is preserved in the knowledge base, which can be parsed and extracted for inquiries.

6.2 Evaluating the medical insurance fraud detection system

We applied the system to the sample data, and had experts on medical insurance regulation evaluate its performance. The test was performed on two hundred medical records to check the system's precision and recall. Precision refers to the percentage of correctly detected (i.e. true) irregular records among the detected records, and recall refers to the percentage of correctly detected records among all irregular records.

In Table 3, the True Positive rate (TP) represents the percentage of cases where the experts agreed with the detected results. The False Negative rate represents cases where the experts disagreed that the detected results are

Parameter	Definition
True Positive rate (TP)	The system detects and the expert agree
False Negative rate (FN)	The system detects but the expert does not agree
False Positive rate (FP)	The system does not detect but the expert detects

Table 3. Evaluation of the medical insurance fraud detection system.

irregularities. The False Positive rate (FN) represents cases where the experts believed the system had failed to detect irregularities. The precision rate is determined by dividing TP with the sum of TP and FN, shown by Equation (1). Equation (2) shows the recall rate, determined by dividing TP with the sum of TP and FP.

$$\text{Precision}=TP/ (TP + FN) \quad (1)$$
$$\text{Recall} = TP / (TP+ FP) \quad (2)$$

Table 4 represents the detected results of the system and the experts. The system achieved a precision rate of 100% and a recall rate of 92.86%. From two hundred medical information instances, it detected thirteen entries with irregularities, which are all consistent with the expert findings. The experts also discovered one additional record with irregular behavior; this is due to that the rules did not account for the misconduct involved in this record, which can be remedied by further refining the rules. The consistency proves the validity of the system.

	System	Experts
Similar Efficacy	7	7
Excessive Prescriptions	3	3
Excessive Inspections	1	1
Multiple Prescriptions	2	2
Other Irregularities	0	1

Table 4. The evaluation results of the system by the experts.

7.0 Conclusions and prospective

The medical insurance system is a foundational institution of national welfare that ensures citizens' access to medical services. The risks of frauds in the field have grown with the development of the industry, and their prevention and control are crucial to saving people's lives. Thus the detection of illicit conduct, the discovery of explicit relationships and the prevention of irregularities in the medical service process have become important topics of research by medical insurance administration and related researchers. Our study presents an analysis of frauds in medical insurance, from the perspective of medical service behaviors. We propose a medical insurance fraud detection inference model, and use the ontology technology to detect fraudulent behaviors. This model is an experimental model based on knowledge of medical domain that can discover irregularities in the rendering of medical services. It provides a new attempt at the detection of medical insurance fraud behaviors.

Irregularities in medical services involve a wide range of knowledge, including the efficacy, pharmacology, indications of drugs, and clinical representations of diseases, and much more. Obtaining such knowledge requires tight cooperation between experts and knowledge engineers. This study mainly focuses on the effective representation of knowledge in the medical domain and the technical implementation of ontology-based inference. Our future work can focus on further improvements on the ontological knowledge base of medical insurance fraud detection, designing more rules to cover additional illegal behaviors and comparing the results with other inference engines and methods, and extending the model to adopt it to big data medical platforms.

References

Baorto, David, Li Li and James J. Cimino. 2009. "Practical Experience with the Maintenance and Auditing of a Large Medical Ontology." *Journal of Biomedical Informatics* 42: 494-503.

Casteleiro, Mercedes Argüello and Jose Julio Des Diz. 2008. "Clinical Practice Guidelines: A Case Study of Combining OWL-S, OWL, and SWRL." *Knowledge-Based Systems* 21: 247–255.

Chen, Rung-Ching, Yun-Hou Huang, Cho-Tsan Bau and Shyi-Ming Chen. 2012. "A Recommendation System Based on Domain Ontology and SWRL for Anti-diabetic Drugs Selection." *Expert Systems with Applications* 39: 3995-4006.

Das, Subhashis and Sayon Roy. 2016. "Faceted Ontological Model for Brain Tumour Study." *Knowledge Organization* 43: 3-12.

Editorial board of the National Library. 2010. *China Library Classification*. Beijing: National Library Press. OCLC has *Chinese Library Classification* = Zhong guo tu shu guan fen lei fa shi yong shou ce / $c Zhong guo tu shu guan fen lei fa bian ji wei yuan hui bian. Di 1 ban. Bei jing: Shu mu wen xian chu ban she, 1999.

Laua, Adela, Eric Tsuib, and W. B. Lee. 2009. "An Ontology-Based Similarity Measurement for Problem-Based Case Reasoning." *Expert Systems with Applications* 36: 6574-79.

Lian, Shi and Sun Jigui. 2006. "Description Logic Survey." *Computer Science* 33: 194-97.

Medical Term Validation Committee. 1998. *Medical Terminology.* Beijing: Science Press.

Mei, Jing and Elena Paslaru Bontas. 2004. "Reasoning Paradigms for OWL Ontologies." Technical Report B-04-12. Berlin: AG Netzbasierte Informationssysteme, 1-23.

Netto, Cristiane Mendes, Gercina Ângela Borém de Oliveira Lima and Ivo Pierozzi Júnior. 2016. "An Application of Facet Analysis Theory and Concept Maps for Faceted Search in a Domain Ontology: Preliminary Studies." *Knowledge Organization* 43: 254-64.

Ni, Zuo and Chu Lifeng. 2014. "Analysis of Monitoring Health Care Fraud based on the Data Analysis Technology." *Insurance Practice and Exploration* 11: 32-40.

Noy, Natalya F. and Deborah L. McGuinness. 2006. "Ontology Development 101: A Guide to Creating Your First Ontology." http://protégé.stanford.edu/publications/ontology_development/ontology101-noymcguinness.html

O'Connor, Martin Joseph, Holger Knublauch, Samson Tu and Mark Alan Musen. 2005. "Writing Rules for the Semantic Web using SWRL and JESS." *Proceedings of the 8th International Protégé Conference, July 18-21, 2005, Madrid, Spain. Protege with Rules Workshop,* 1-7.

Pan, Jeff Z., Giorgos Stoilos, Giorgos Stamou, Vassilis Tzouvaras and Ian Horrocks. 2005. "f-SWRL: A Fuzzy Extension of SWRL." In *Artificial Neural Networks: Formal Models and Their Applications- ICANN 2005: 15th International Conference, Warsaw, Poland, September 11-15, 2005, Proceedings, Part II*, ed. Wlodzislaw Duch, Erkki Oja and Slawomir Zadrozny. Berlin: Springer, 829-34.

Ruyi, Liu. 2014. "Research and Application of Medical Violation Reasoning Detection Model Based on Ontology." Masters thesis, Jiangsu University.

Smith, Barry, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, The OBI Consortium, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H Scheuermann, Nigam Shah, Patricia L Whetzel and Suzanna Lewis. 2007. "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration." *Nature Biotechnology* 25: 1251-55.

Snae, Chakkrit and Michael Brueckner. 2009. "Personal Health Assistance Service Expert System (PHASES)." *International Journal of Biological and Medical Sciences* 26: 157-60.

Xiaobin, Zhou and Cao Cungen. 2003. "Medical Knowledge Acquisition: An Ontology-Based Approach." *Computer Science* 10: 40-45.

Xin, Li. 2006. "Analysis of Potentially Fraudulent Behavior of Basic Medical Insurance." *Journal of Hubei University of economics* 3 no. 9: 40-41.

Xinmin, Zhang. 2014. "Study on the Legal Liability System of Social Medical Insurance Fraud." *Journal of Southwest University for Nationalities* 35: 90-95.

Yazi, Li and You Bin. 2015. "Analysis of Characteristics of Medical Insurance Fraud." *China Social Security* 2015 no. 2: 76-79.

Yi, Li and Pang Jingan. 2003. "Research on Semantic Indexing System and Semantic Retrieval Model for Chinese Medical Information based on Multilayer Conceptual Semantic Network Structure." *Journal of the China Society for Scientific & Technical Information* 22: 403-11.

Zhao, Zeng and Wang Xiaoping. 2006. "UMLS and the Establishment of the Traditional Chinese Medical Language System." *Chinese Journal of Medical Library and Information Science* 15 no. 3: 1-3.