Tech Science Press

An Intelligent Tumors Coding Method Based on Drools

Panjie Yang[1,*,#], Gang Liu[2,#], Xiaoyu Li[1,*], Liyuan Qin[1] and Xiaoxia Liu[3]

[1]School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]The First Affiliated Hospital of Chengdu Medical College, Chengdu, China
[3]University of Electronic Science and Technology School of Medical Affiliated Tumor Hospital, Chengdu, China
[#]Contributed equally to this work
[*]Corresponding Authors: Panjie Yang. Email: ypjngup@163.com; Xiaoyu Li. Email: xiaoyu33521@163.com

Abstract: In order to solve the problems of low efficiency and heavy workload of tumor coding in hospitals, we proposed a Drools-based intelligent tumors coding method. At present, most tumor hospitals use manual coding, the trained coders follow the main diagnosis selection rules to select the main diagnosis from the discharge diagnosis of the tumor patients, and then code all the discharge diagnoses according to the coding rules. Owing to different coders have different familiarity with the main diagnosis selection rules and ICD-10 disease coding, it will reduce the efficiency of the artificial coding results and affect the quality of the whole medical record. We first analyze the ICD library information, doctor's diagnostic information, radiotherapy information or chemotherapy information, surgery information, hospitalization information and other related information, and then generated Drools rule files based on the main diagnostic selection principles and coding principles, we also combined the text similarity analysis algorithm to construct an intelligent diagnostic information coding method. Practice shows that the coding method can be used to make the work efficiently and at the same time obtain the coding results which meet the standard and have high accuracy, so that the coders can be free from the repeated work and pay more attention to coding quality control and the coding logic adjustment.

Keywords: Tumor coding; rule engine; drools; ICD-10

1 Introduction

In recent years, the process of industrialization and urbanization has accelerated the problems of poor life style and environmental pollution [1]. Malignant tumors have become major diseases that seriously threaten the lives and social development of Chinese residents. One out of every five dead patients are a tumor patient, the Chinese malignant tumor discipline Development Report (2017) pointed out. With the increase number of tumor patients, the demand for coding in hospitals is gradually increase. In China, most of the tumor hospitals encode the tumor disease according to the International Statistical Classification of the Diseases and Related Health (ICD-10) [2]. However, with the increase of medical knowledge and diseases classified in the ICD-10 disease coding library, the complexity and professionalism required for diagnostic coding have become the main challenge for medical professionals [3,4].

Compared with the application of Drools rule engine in other fields, Drools is less used in medical field, and most of the applications in medical field are disease self-check and medical insurance. For example, Mu et al. [5] and others proposed in 2012 Drools to be used in the disease self-check system that helps people to conduct self-examination and rapid registration of diseases, Wang [6] proposed medical insurance system based on Drools rule engine. However, there is little research on disease coding.

Manual coding is mainly based on the existing patient diagnosis data, which requires the coders to master the knowledge of disease classification. The coders select the main diagnosis of the disease according to the etiology, pathology, purpose of hospitalization and the amount of health hazard caused by the disease, the amount of medical energy spend in the treatment of the disease, and the length of hospitalization [7], then the key segments in the tumor diagnosis are identified manually and code against the ICD-10 disease coding library. However, a tumor patient may have a variety of diseases, in the face of a large number of disease diagnosis, the use of manual coding for tumor disease diagnosis coding is difficult to ensure the accuracy of coding because of human factors and low coding efficiency. The rule engine Drools is used to code the tumor disease, and the selection principle of the main diagnosis of the tumor disease and the comparison rule of the tumor disease diagnosis with the ICD-10 disease knowledge base are compiled into the rule file through the Java language. We use Drools which has the characteristics of managing rules and efficiently executing rules. It can efficiently select the main diagnosis of tumor diseases from the patient's disease diagnosis and encode all disease diagnosis. Therefore, it is very useful and necessary to study Drools-based intelligent tumors coding method (hereinafter referred to as DTIC), so as to reduce the workload of coders and improve the coding efficiency and the quality of medical records.

The method scores each diagnostic coding result of the patient according to the accuracy of the tumor disease diagnosis compared with the knowledge base of ICD-10 disease. We will select about 30000 patient cases and compare the results of the rule engine reasoning with the manual coding results. Then we choose to take a certain segment as the threshold according to the qualified rate of the segment, those below this segment need to be manually modified. Our purpose is to make the number of cases below the qualified rate as small as possible. Later, we will consider using the K-NN method used in the HEMS [8] system to classify and regress the diagnostic data sets to redefine the threshold of eligibility rate or to reconstruct the tumor rule decision based on the fast decision method of convolutional neural network [9]. We will also use multi-label text classification algorithms to classify diagnostic diseases to improve coding accuracy [10].

The structure of the paper is as follows: The second part describes the Drools rules engine, the Rete algorithm which is core algorithm of Drools and the rule reasoning process. The third part describes the implementation process of Drools-based intelligent tumors coding method. In the fourth part, we experiment the method and analyze the result of the experiment. The fifth part is conclusion.

2 Drools Rule Engine and Rule Inference

Rule engine is the product of inference engine [11], DTIC uses Drools as the rule engine for rule inference, which can effectively separate business decisions from procedure word. It makes no coupling between code and business, which can not only reduce the complexity of the experimental method, but also make the method meet the requirements of high efficiency and strong robustness. The rule-based inference includes the main diagnosis selection inference and the disease coding inference. The next part introduces the rule engine Drools used in DTIC and its core algorithm and rule inference.

2.1 Drools Rule Engine

Drools, also known as JBoss rules, is an open source rule engine customized based on the Java language. It is a hybrid linking engine, which means that it can respond quickly to changes in data and provide advanced query functions [12], this is one of the reasons we use Drools. The traditional logical judgment is to match one by one through if-else, and execute the corresponding rules when the corresponding conditions are matched, but when the requirements change, the source code of the corresponding rules needs to be modified, which is very difficult to manage. If we use traditional logic to select the main diagnosis of disease at this time, it will lead to poor code readability and maintenance because of the large number of rules, the different rules of patients with different hospitalization times and the change of a certain rule in a certain period of time. In addition, we may cause errors due to neglecting the semantic relationship of context when modifying the rules, so that the risk of modification is very high. In contrast, it is more appropriate to choose Rete algorithm. With the advantage of the Rete algorithm, the rules engine Drools is very convenient to manage the rules, it can also meet the demand of higher frequency

of rule change, and Drools is also more efficient than traditional logical judgment for the application scenario with huge rule variables.

2.2 Rete Algorithm

Rete algorithm is one of the core algorithms of Drools, and it is an efficient pattern matching algorithm. It was designed and invented by Dr Charles L. Forgy of Carnegie Mellon University [13]. The efficiency of the rule matching determines the performance of the rule engine [14], but is not affected by the number of rules. Objects can be filtered through the Rete network to find out the matching pattern. At present, many of the top commercial business rules in the world mainly adopt the Rete algorithm [15]. Rete network is usually composed of six nodes and can be divided into two parts, that is, the Root node, the Type node, the Alpha node, the LeftInputAdapterbeta node, the Beta node, the Terminal node, and the rule compilation and runtime execution. The rule compilation is to create a corresponding network according to the rules in the rule base, the steps are as follows:

Step 1: Passing the facts (a set of data describing the relationship between objects and attributes) to the Type node through the created Root node (Entrance of the Rete network). Type nodes hold the various types of facts.

Step 2: Taking out a pattern (the smallest match in the rule) in the rule library and checking the type in the schema. If it is a new fact type, adding a Type node of new type.

Step 3: Checking the Alpha node corresponding to the mode (used to evaluate literal conditions). If it exists, recording the location of the node, and if not, adding the schema to the Rete network as a new Alpha node, and establishing the Alpha memory table according to the pattern.

Step 4: Repeating Step 3 until all modes are processed.

Step 5: The Beta node consists of a left Alpha node (left input node) and a right Alpha node (right input node), so it may have multiple Beta nodes. The LeftInputAdapterbeta node converts a fact into a tuple and provides a function for the Beta node.

Step 6: Repeating Step 5 until all Beta nodes are processed and then encapsulating the Then part as the last Beta (n) node.

Step 7: Repeating from Step 2 to step 6 until entering the Terminal node, it indicates that all the rules in the rule base have been completed.

Taking whether an advanced liver cancer patient is over 60 years old or not as an example, the Rete etwork is shown in Fig. 1.

2.3 Rule Inference

We choose to use Drools for rule reasoning to reason the diagnosis of patients efficiently according to the main diagnosis selection rules and disease coding rules of the tumor when inputting the diagnosis of many patients. Main diagnostic selection inference. One of the characteristics of medical reasoning is the focusing mechanism, which is used to select the final diagnosis from many candidate diagnoses [16]. In DTIC, if we infer the patient who is hospitalized for the first time according to the number of hospitalization, its main diagnostic selection of tumor diseases is as follows:

Firstly, looking for the keywords of cancer and tumor in the diagnosis, if these keywords exist, the diagnostic string before the keywords is taken as the primary diagnosis, which is based on the principle that the first time inpatient is the primary tumor for primary diagnosis.

Secondly, looking for keywords such as radiotherapy, chemotherapy in the diagnosis and choosing radiotherapy diagnosis or chemotherapy diagnosis. If the cancer and tumor keywords are not found in the previous step, then the radiotherapy diagnosis or chemotherapy diagnosis is used as the main diagnosis; otherwise, the other diagnosis is given first place.
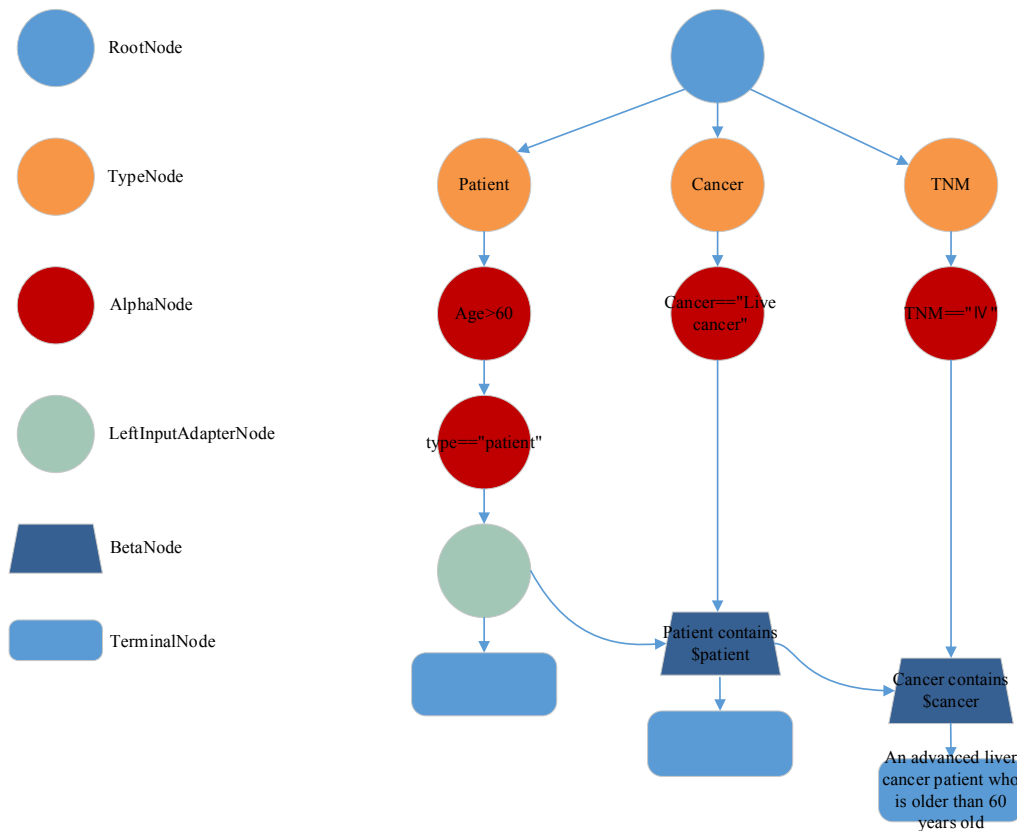
Figure 1: Rete network example

Thirdly, determining whether there are keywords, such as secondary, transferred in the diagnosis. If so, it is inferred that the tumor patient has tumor metastasis, whether the secondary tumor is treated as the primary diagnosis depends on the first and second steps. If the first two steps fail to match, the secondary diagnosis is selected as the primary segment, otherwise, put it first or postpone position in other diagnosis.

If it is not a first inpatient, the difference in the reasoning process is to give priority to judging whether the patient is a surgical patient, if so, the primary diagnosis selects the reasoning process and the first inpatient agreement. The reasoning process of main diagnosis selection in non-surgical patients is to exchange the first step and the second step in the reasoning process of the main diagnosis selection of the first inpatient.

Disease coding and reasoning. Disease coding and reasoning. The disease code is coded on the primary and non-primary diagnoses after the main diagnostic inference has been completed. The disease coding and reasoning are as follows:

Step 1, analysing the text similarity between the diagnosis and the disease in the ICD-10 disease coding base, diseases with the highest similarity were selected as the matching results, then the corresponding coding of the matching result is the disease code of the diagnosis, and the corresponding similarity is taken as the weight of the reasoning link and the score of the coding result. The text similarity analysis adopts Gensim method based on TF-IDF and Jaro method in Levenshtein, comparing the two methods and selecting the best results as the results of the analysis.

Step 2, the coding results are further modified by exclusive reasoning. The exclusive reasoning is based on the patient's attributes. For example, one tumor patient has a reproductive organ carcinoma, and if the patient is a male, the female genital organ carcinoma is excluded, whereas the male genital organ carcinoma is excluded.

Taking the first inpatient as an example, its main diagnostic reasoning process code is shown in Appendix A, and the whole rule reasoning process is shown in Fig. 2.
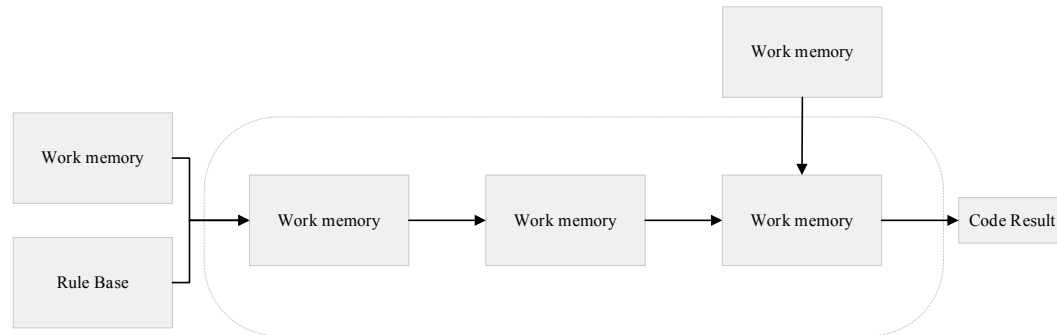

Figure 2: Reasoning process

3 Drools-based Intelligent Tumors Coding Method

As a Drools-based intelligent tumors coding method, its research focuses on the structure of the rule base. The effectiveness of the rule base directly affects the accuracy of the final coding results. The core rules of this method are the selection rule of tumor main diagnosis and exclusivity rule in the process of disease coding and reasoning, it generated completely by experts. Drools-based intelligent tumors coding method can enable experts who are not familiar with software coding technology to focus on the description of rule task logic without paying attention to the implementation of code.

3.1 Establishment of a Tumor Coding Rule Base

In the notification issued by the Executive Office of the National Health and Family Planning Commission [17], the main diagnosis is selected for tumor-like diseases in accordance with the following principles:

- If this hospitalization is for tumor surgical treatment, selecting the tumor as the main diagnosis.
- If the secondary tumor is treated or diagnosed in the hospital, even if the primary tumor still exists, the secondary tumor is selected as the main diagnosis.
- Only radiotherapy or chemotherapy is performed for malignant tumors in the hospital, choosing radiotherapy or chemotherapy of malignant tumors as the main diagnosis.
- If this hospitalization is aimed at treatment of tumor complications or diseases other tumors, choosing complications or that disease as the main diagnosis.

For patients with different times of hospitalization, the difference is that the order of the above rules, which is described in detail in the process of main diagnostic reasoning in the second part. According to the above rules and exclusive rules described in the second part, the tumor coding rule base is established as shown in Fig. 3.

3.2 Tumor Rule Enforcement

The process of Drools rule engine is start with a fact, after a step-by-step matching operation, it performs the corresponding operation and finally draws the corresponding conclusion [18]. Based on this process, our rule enforcement process is as follows: First of all, when we code each patient with tumor disease, we store the required basic patient information data in Map, and transfer the session object created through the KIE container to the working memory (Working Memory). Secondly, the Drools rule engine dynamically loads the rules of the tumor encoding rules base in the working memory, and then matches the rules currently stored in the workspace when the session object is inserted into the workspace. Before pattern matching, all facts are asserted, and their own WME (Working Memory Element) is established for each fact and then matched down from the root node. Finally, when the rule matches, the inference is carried out according to the rules. The reasoning process is described in the second part of rule reasoning. The result is composed of three parts, namely, the main diagnosis result, the disease coding result and the score

corresponding to each disease coding result. We take all the diagnosis of a tumor patient as a sample, and the final coding accuracy of a sample is the average value of the coding results of each disease in the sample. See the calculation formula is shown in equation.
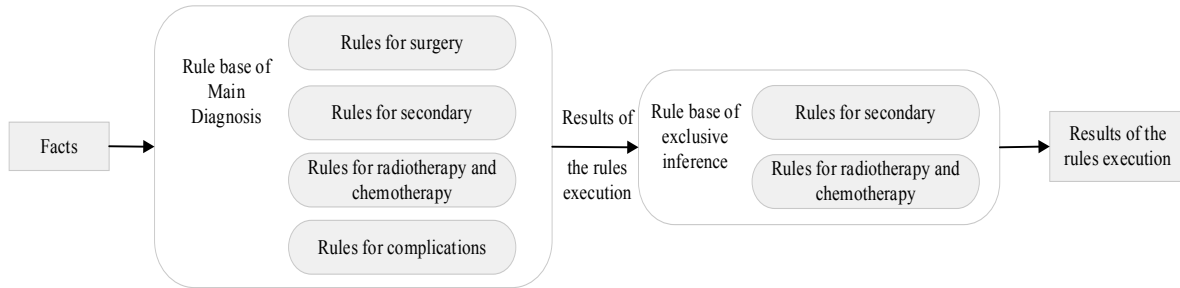

Figure 3: Tumor coding rule base

$$S = \frac{\sum_{1}^{n} Max\{\delta_1......\delta_m\}}{n} \tag{1}$$

where: S is the final score of a sample, n is the number of disease diagnoses in a sample, δ_m is the similarity between a disease diagnosis in the sample and m disease names in the ICD-10 disease code base, $Max\{\delta_1......\delta_m\}$ is the maximum similarity selected from $\{\delta_1......\delta_m\}$.

4 Experiment and Analysis

4.1 Experiment

We designed two groups of experiments to test the method, taking the disease diagnosis of 33293 patients as the test data, and the artificial coding results of these data (neglecting the errors caused by human factors) as the reference standard. The experimental methods are as follows:

Experiment 1. We respectively used Drools-based intelligent tumors coding method and the traditional rule logic judgment method to realize the main diagnosis selection and disease coding of tumor patients. The main diagnosis and selection rules used in the two methods are consistent with the matching rules of the ICD-10 disease code base and the methods of measuring similarity are the same. Experiment 2. We will use Drools-based intelligent tumors coding method to carry out the major diagnostic selection and disease coding, and compare the results with the artificially coded results.

In Experiment 1, it needs average every 8 seconds to complete the main diagnosis selection and disease coding of one tumor patient by using the traditional way. While it only needs 3 seconds to complete the main diagnosis selection and disease coding of one tumor patient by using intelligent tumors coding method. The actual execution time of the two methods depends on the number of disease diagnosis of a tumor patient, the more the number is, the more time it takes. In Experiment 2, for the existing test data, the results are shown in Tab. 1. There are 21360 cases with the score higher than 0.9, of which 20527 cases are consistent with manual coding, and the coding accuracy is 96.1%. There are 8484 cases with the score of 0.8-0.9, and 7720 case in accordance with the artificial codes, the coding accuracy is 91.0%. There are 1545 cases with the score of 0.7–0.8, and 1236 cases in accordance with the artificial codes, the coding accuracy is 80%. There are 929 cases with the score of 0.6–0.7, and 416 cases in accordance with the artificial codes, the coding accuracy is 44.8%. There are 975 cases lower than the score of 0.6 and 17 cases in accordance with manual coding, the coding accuracy is 1.7%.

4.2 Results Analysis

Comparing the two methods in experiment 1, we find that the traditional method takes about three times as much time as the intelligent tumor coding method, which fully shows that Drools is more efficient than the traditional logical judgment. In addition, by further analyze that degree of association between the execution time and the disease number of the tumor patient, we find that the time consumption of the two method is obviously increase, mainly because the number of the tumor disease diagnosis and the ICD-10 disease code base is increased when the number of the disease of the tumor patient is large, which also explains the problem that the execution efficiency of the Drools is independent of the number of matched rules, but is time-consuming.

Table 1: Results of using intelligent tumors coding method

Score	Number of cases that meet the score	Number of cases that meet manual coding	Coding accuracy
>0.9	21360	20527	96.1%
0.8–0.9	8484	7720	91.0%
0.7–0.8	1545	1236	80%
0.6–0.7	929	416	44.8%
<0.6	975	17	1.7%

In Tab. 1, 28247 cases with a score higher than 0.8 and the coding accuracy is more than 91.0 percent, which accounting for 85% of the test data. In the rest 15% of the test data, the coding accuracy is more than 80%, accounting for 36.0%. Even if it needs to be modified manually, it requires only a very small amount of work. If the error of the reference sample itself and the allowable error of the coding result are taken into account, score 0.8 can be used as the dividing point. It is only necessary to modify the code with a score below 0.8 to ensure that the coding is in accordance with the standard and the high accuracy, liberating the coders from the repetitive labor force and save the labor force at the same time.

5 Conclusion

In order to solve the problem of low coding efficiency of tumor, we proposed a Drools-based intelligent tumors coding method. The purpose of this paper is to describe the research and implementation of this method. Compared with the traditional logic processing, this method can ensure that the rule can be managed flexibly under the condition of extremely low risk of modification, and the influence of the number of rules on the execution efficiency of the rule can be ignored, especially for the application scenario of tumor coding, when processing disease diagnostic strings, there are many branching rules, so it is particularly important to ignore the influence of the number of rules. Compared with manual coding, this method can obtain unified standard and high accuracy coding results while completing tumor coding efficiently. It can make coders focus on modifying the coding results below a certain fraction threshold, reduce the workload of coders, save manpower and strengthen the quality control of coding at the same time.

In the future, we will focus on how to improve the accuracy of coding results. The first is to consider how to refine the rules, for example, the use of exclusive rule reasoning in the second section, it can effectively correct some of the error disease codes. If the disease coding rules can be further refined, the accuracy of the coding of the tumor disease diagnosis can be effectively improved. Secondly, how to improve the accuracy of coding results from the source, that is, how to standardize the writing problem diagnosed by doctors. If clinicians can be further strictly required to make the diagnosis of tumor diseases written by doctors clean and tidy, so that the similarity analysis of the text can ensure that the diagnosis of tumor diseases can best match the disease names in the ICD-10 disease coding base. In addition, we

consider adding natural language processing technology to preprocess the tumor diseases diagnosis is not standard, which can also improve the accuracy from the source.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] F. Wu, G. Z. Lin and J. X. Zhan, "The present situation and trend of malignant tumors in China," *Chinese Oncology*, vol. 21, no. 2, pp. 81–85, 2012.

[2] A. L. Sandra, G. Iouri and D. Vivian, "Development and evaluation of a serious game for teaching ICD-10 diagnosis coding to medical students," *Games for Health Journal*, vol. 8, no. 5, pp. 349–356, 2019.

[3] E. Erin, G, Paul and D, Christina, "Identifying medicare beneficiaries accessing transgender-related care in the Era of ICD-10," *Games for Health Journal*, vol. 6, no. 4, pp. 166–173, 2019.

[4] D. S. Zheng, X. Y. Li and G. W. Yang, "An assertion graph based abstraction algorithm in GSTE and its application," *Integration–The VLSI Journal*, vol. 63, no. 9, pp. 1–8, 2018.

[5] B. Mu, X. Feng and J. Y. Shi, "A rule-based disease self-inspection and hospital registration recommendation system," in *2012 IEEE Intl. Conf. on Computer Science and Automation Engineering*, Beijing, China, pp. 212–215, 2012.

[6] J. C. Leng, K. Peng and W. X. Sun, "Research on medical insurance audit system based on rule engine," *China Digital Medicine*, vol. 9, no. 12, pp. 4–5.

[7] N. C. Li and Y. Lu, "Case analysis of main diagnostic selections," *Medical Information*, vol. 26, no. 3, pp. 409–410, 2013.

[8] Q. Liu, Z. Yang, X. Liu and S. Mbonihankuye, "Analysis of the efficiency-energy with regression and classification in household using K-NN," *Journal of New Media*, vol. 1, no. 2, pp. 101–113, 2019.

[9] L. Chen, B. Wang, W. Yu and X. Fan, "CNN-based fast HEVC quantization parameter mode decision," *Journal of New Media*, vol. 1, no. 3, pp. 115–126, 2019.

[10] J. He, C. Wang, H. Wu, L. Yan and C. Lu, "Multi-label Chinese comments categorization: comparison of multi-label learning algorithms," *Journal of New Media*, vol. 1, no. 2, pp. 51–61, 2019.

[11] M. Li, W. Lu and D. Xiang, "Design and realization of transformer fault diagnostic expert system based on Drools," in *2015 Intl. Conf. on Computational Intelligence and Communication Networks*, Jabalpur, India, pp. 1583–1588, 2015.

[12] M. Proctor, *Drools: A Rule Engine for Complex Event Processing*. Berlin, Heidelberg, 2011. [Online]. Avaliable: https://doi.org/10.1007/978-3-642-34176-2_2.

[13] D. Liu, G. Tao and P. X. Jiang, "Rule engine based on improvement rete algorithm," in *Intl. Conf. on Apperceiving Computing and Intelligence Analysis Proc.*, Chengdu, Sichuan, China, pp. 346–349, 2010.

[14] J. L. Bao, "Research on rete algorithm based on rule engine technology," *Science & Technology Information*, vol. 32, no. 6, pp. 92–105, 2008.

[15] F. Guo and J. J. Bai, "Rule engine JBoss rules based on rete algorithm," *Computer Era*, vol. 17, no. 1, pp. 8–10, 2008.

[16] S. Tsumoto, *Modelling Medical Diagnostic Rules Based on Rough Sets*. Berlin, Heidelberg. 1999. [Online]. Avaliable: https://doi.org/10.1007/3-540-69115-4_65

[17] General Office of the National Health and Family Planning Commission of China, *Notice on Printing and Distributing the Quality Specifications for the First Page Data of Inpatient Cases (interim) and the Quality Management and Control Indicators for the First Page Data of Inpatient Cases.* Beijing, China, 2016. [Online]. Avaliable: http://www.nhc.gov.cn/yzygj/s2909/201606/fa8a993ec972456097a2a47379276f03.shtml

[18] K. Niu, "Research on application of rules engine drools in manufacturing J2EE system," *Group Technology & Production Modernization*, vol. 31, no. 2, pp. 7–13, 2014.

Appendix A

```
rule "FirstPatient"
    when
        $diagnosis:Diagnosis(Admission==1 || (Admission>1 &&Shoushuleixing==true) ||
                        (Admission>1 && Shoushuleixing==false&& NewTumor == true));
        $readDiseaseService:ReadDiseaseService();
        $readShoushuService:ReadShoushuService();
        $readXingtaiService:ReadXingtaiService();
    then
        drools.setFocus("DealFistPatient");
end

rule "TumorCancer" extends "FirstPatient"
agenda-group "DealFistPatient"
    when
        $firstDiagnsis:Diagnosis($diagnosis.getFistDiagnosis() matches; ".*[Cancer|Tumor]*.");
    then
        String[] arraystr = $diagnosis.getFistDiagnosis().split("Cancer|Tumor");
        tempStr = arraystr[0];
        String res = $diagnosis.getFistDiagnosis().replace(tempStr, "");
        $firstDiagnosis.setFistDiagnosis(res);
        Map TumorCancerRes = $readDiseaseService.getCancer(tempStr);
        $firstDiagnosis.setAllDiagnosis(String.valueOf(TumorCancerRes.get("DiseaseName")));
        $firstDiagnosis.setAllDiagnosisCode(String.valueOf(TumorCancerRes.get("DiseaseCode")));
        $firstDiagnosis.setAccuracy(Double.parseDouble(String.valueOf(TumorCancerRes.get("Accuracy"))));
end
```