HTI Open Access Collection 1

doi:10.3233/SHTI200015

Construction of Guideline-Based Decision Tree for Medication Recommendation

Wei ZHAO, Xuehan JIANG, Ke WANG, Xingzhi SUN, Gang HU and Guotong XIE[1]
Ping An Health Technology, Beijing, China

Abstract. Clinical decision support system (CDSS) plays an essential role nowadays and CDSS for treatment provides clinicians with the clinical evidence of candidate prescriptions to assist them in making patient-specific decisions. Therefore, it is essential to find a partition of patients such that patients with similar clinical conditions are grouped together and the preferred prescriptions for different groups are diverged. A comprehensive clinical guideline often provides information of patient partition. However, for most diseases, the guideline is not so detailed that only limited circumstances are covered. This makes it challenging to group patients properly. Here we proposed an approach that combines clinical guidelines with medical data to construct a nested decision tree for patient partitioning and treatment recommendation. Compared with pure data-driven decision tree, the recommendations generated by our model have better guideline adherence and interpretability. The approach was successfully applied in a real-world case study of patients with hyperthyroidism.

Keywords. Decision support systems, management, guideline adherence, machine learning

Introduction

Clinical decision support system (CDSS) is one of the major components of personalized health service [1]. Rapid development of CDSS has been driven by explosive increase of electronic health records in recent years. CDSS can provide smart recommendations to both clinicians and patients at almost every stage in a clinical process [2]. In this study, we focus on CDSS for medication recommendation, with the purpose of helping clinicians in primary care to improve the quality and safety of the treatment. This is urgently demanded in China as the national clinical resources are significantly biased to the advanced hospitals in cities, while the majority of the population resides in the rural area only covered by primary care [3].

Usually, CDSS for recommendation is accomplished by partitioning patients so that the patients with similar circumstances are likely to be treated in the same way. In general, there are two types of algorithms for medication recommendation: knowledge-driven and data-driven. The knowledge-driven approach is to develop an expert system based on domain knowledge (e.g., clinical guidelines) [4–6], while the data-driven one is to apply the data mining techniques on medical records to build the mapping between patient's information and medications [7–9].

[1]Address for correspondence. Guotong XIE. Email: xieguotong@pingan.com.cn.

The knowledge-driven approach maintains a set of rules derived from the clinical guideline. By applying the rules, patients satisfied the rule conditions are recommended with corresponding medications, which ensures that the recommendations are always consistent with the guidelines. Following such an approach, Mei et al. developed an OCL-compliant GELLO engine and successfully applied this model in management of chronic diseases [4]. However, the prerequisite of developing adequate knowledge driven CDSS is a comprehensive clinical guideline, which is absent for many types of diseases where the guidelines only cover patient populations with limited clinical conditions. In addition, the knowledge provided in clinical guideline could be too general to provide fine-granular personalized recommendations.

With the large amounts of electronic clinical records (EMRs), data-driven CDSS can be developed to provide personalized medication recommendations. Recently, Liu et al. proposed an algorithm to group the patients based on the similarity metric learnt from the real clinical data [10]. Likewise, Chen and Altman reported a Bayesian conditional probability model for recommendation of clinical orders through the data mining of EMRs [8][9]. The authors demonstrated that automatic recommendations generated by data-driven CDSS could be consistent with clinical guidelines to a large extent. However, although data-driven approach is easy to scale to a large number of diseases and often provides personalized result, the correctness of the result highly depends on the quality and the distribution of the dataset. The recommendations from data-driven CDSS may be not guideline-concordant, which degrades the clinicians' trust to use the system.

In this study, we propose a method to integrate the knowledge-driven and data-driven approach for medication recommendation. Our target is to find a partition of patients such that the patients with the similar clinical conditions are grouped together and there exist the preferred prescription(s) for each group. Such a partition setting can be represented in a form of decision tree, in which the internal nodes on a path define the patient group and the label of the leaf node provides the medication. Specifically, we first build a decision tree based on the clinical rules with rule-obeyed data, then rule-uncovered data are loaded to further split the tree. Intuitively, a nested decision tree is constructed in a process of supplementing and refining the clinical guideline. All clinical rules as well as important hidden information in the EMR data are shown in the nested tree, which makes the recommendation provided by the nested tree not only patient-specific, but also consistent with the knowledge.

We implemented such an approach and applied it in a real-world case study of patients with hyperthyroidism. A nested decision tree was built based on clinical guideline and we compared it with the baseline decision tree using CART algorithm [11], which is the pure data-driven approach. We found that, with little compromise in prediction accuracy for rule-uncovered data, the nested decision tree had better guideline adherence and increased the interpretability of the model. With demonstrated effectiveness, we believe that medication recommendation for other diseases could generally benefit from the proposed approach, especially when the clinical guidelines of diseases only address the limited clinical conditions of the patients.

Methods

For a type of disease *d*, we are given the clinical guideline of *d*, and a EMR dataset for patients diagnosed with disease *d*, in which each sample consists of the clinical condi-

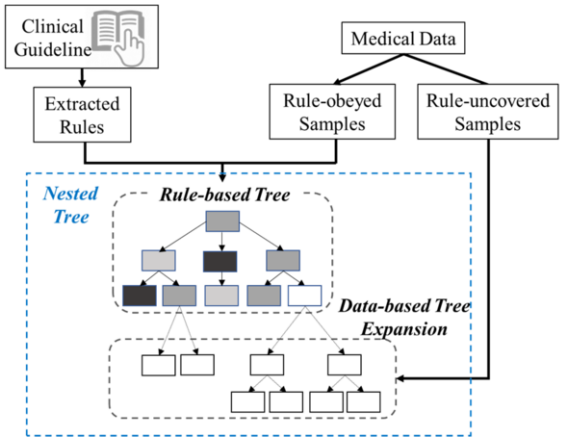

Figure 1. Construction of the nested decision tree by combining knowledge-driven and data-driven approaches.

tions of a patient and the prescription made by the clinician. Clinical rules for treatment of *d* are extracted from guidelines. And medical data is splited into two parts according to whether the clinical condition are mentioned by the rules, namely rule-coverd and rule-uncovered samples. Rule-obeyed samples are further selected from rule-covered samples by the agreement of prescriptions with guideline rules. As illustrated in Fig. 1, using rule-obeyed samples whose prescriptions are consistent with rules, the basic part of our tree (called rule-based tree) is built under the guide of extracted rules. By loading the rule-uncovered samples, the rule-based tree gets further expansion (called data-based tree expansion). The basic tree part and its data-based tree expansion part form a nested tree.

The method section mainly focused on methodology of key steps in the construction of the nested decision tree: guideline rule extraction, construction of rule-based decision tree and decision tree expansion.

Guideline Rule Extraction

To integrate the guideline knowledge into the construction of decision tree, the first step is to extract structured rules from clinical guidelines. An extracted rule contains information of related feature conditions and corresponding medication recommendation.

For example, information like "patients of hyperthyroidism coexistent cardiovascular disease should take β blocking agents [12][13]" in guideline would be converted to a rule like: *if prediag_cvd* is true, *then* the prescription should contain β blocking agents. In this rule, "*prediag_cvd*" is the feature name to represent the status whether patients suffering from cardiovascular disease, and the rule is satisfied when the value of *prediag_cvd* equals to 1. With the rule satisfied, our candidate recommendations will be β blocking agents, or the drug combinations that occur frequently in the data and contain β blocking agents.

Construction of Rule-Based Decision Tree

In general, decision tree was constructed by selection of features that had largest information entropy gain or least Gini impurity/Gini index [11]. Gini index for a binary feature A is calculated as:

$$\text{Gini}(D_A) = \frac{|D_{A=0}|}{|D|} Gini(D_{A=0}) + \frac{|D_{A=1}|}{|D|} Gini(D_{A=1}) \quad (1)$$

where D represents the total data of parent node. The symbol $|D|$ gives the sample number of the node. The subscript (0 and 1) denotes possible values for a binary feature. Gini index for feature equals to i can be calculated as:

$$Gini(D_{A=i}) = 1 - \sum_{j=1}^{n} p_{ij}^2 \quad (2)$$

where i belongs to {0,1} for binary data. And n is the number of possible classes with p_{ij} denoting the fraction of class j in the set of $D_{A=i}$.

In our approach, we also used Gini index and all features were converted into a binary dataset. The novelty is that we applied the rules to guide the tree construction. Specifically, we added weight factors to rule-related features, making them favored to be selected as splitting features during tree construction. We defined weighted Gini gain to select splitting features, and the one with largest weighted Gini gain was selected to group the data. The weighted Gini gain for feature A was calculated as follows:

If feature A is not mentioned in all rules,

$$\Delta Gini_A^{weight} = \Delta Gini_A \quad (3)$$

If feature A is mentioned in any rule,

$$\Delta Gini_A^{weight} = \xi \Delta Gini_A (1 + Weight_A) \quad (4)$$

where

$$\Delta Gini_A = Gini(D) - Gini(D_A) \quad (5)$$

and $Gini(D_A)$ as well as $Gini(D)$ can be calculated according to Formula (1) and (2), respectively. In Formula (4), ξ is a coefficient of constant larger than 1 to ensure the preference for features extracted from the rules. $Weight_A$ is the weight of feature extracted from the rules to ensure the preference for features with high percentage of satisfied samples, and its value depends on the corresponding conditions mentioned in the rules. The way to determine the value of $Weight_A$ is listed in Table 1, where $\text{Percent}_{\text{featureA=i}}$ is defined as the percentage of the samples that satisfy the condition featureA=i.

Table 1. Determination of $Weight_A$.

Value	Condition
$\text{Percent}_{\text{featureA=0}}$	Feature A is 0 in all the rules
$\text{Percent}_{\text{featureA=1}}$	Feature A is 1 in all the rules
$\text{Max}(\text{Percent}_{\text{featureA=0}}, \text{Percent}_{\text{featureA=1}})$	Feature A=1 in some rules, and FeatureA =0 in other one(s)

To summarize, $\Delta Gini_A$ accounts for the baseline feature importance in splitting data. ξ makes the features extracted from the rules have higher priority than the features not mentioned in any rule. $Weight_A$ differentiates importance among rule extracted features.

Decision Tree Expansion

To further expand the decision tree, clinical data not covered by the rules is loaded into the rule-based tree. By going through the rule-based tree, samples are grouped into several leaf nodes and those leaf nodes can be further split as long as the stop criteria is not satisfied. Tree expansion in this phase follows exactly the general CART algorithm. Therefore, after the expansion, a nested tree is constructed to provide additional information, including 1) new branches for the patients whose conditions are not covered by the guideline and/or 2) finer partition for the existing patient groups.

Another consideration in our approach is to avoid overfitting of the decision tree. Therefore, two parameters are used to constrain the tree. They are maximum depth of the tree (max_depth) and the minimum number of samples at a leaf node (min_samples_leaf). Decision tree constructed under such constraints has limited layers and is easy to be interpreted.

Results

To illustrate our approach, a real-world case study of patients with hyperthyroidism is introduced here. Like most diseases, the guideline of hyperthyroidism does not provide a hierarchical way to group patients, i.e., only addresses the patient groups with some clinical conditions.

Data Set

We used medical data from 90 clinical institutes in China with time spanning from 2012 to 2016. In total, there were 83360 records with the main diagnosis as hyperthyroidism, corresponding to 17166 unique patients. Females accounted for 80% of the whole data. The distribution of age centered around 20 to 40 years old. Those population characteristics described from data agree well with the fact that young women are more susceptible to hyperthyroidism.

We further investigated the history of disease and medication. It was found that 99% of patients had disease history, and about 65% and 5% of patients had merely hyperthyroidism and Graves disease history, respectively. In terms of medication history, 78% of patients had taken drugs for hyperthyroidism medication, among which 91% of patients followed previous prescriptions.

Each sample in the medical data we used recorded one visit of hyperthyroidism patient, which consisted of basic information of the patient, diagnosis made by the clinician, items of image tests and laboratory tests the patient carried out, and the medication prescription. All features were converted to binary data for further analysis.

Table 2. Extracted rules from guideline for hyperthyroidism.

ID	Feature	Value	Medication
1	prediag_cvd	True	include β blocking agents
2	prediag_thyroid_crisis	True	exclude MMI
3	prediag_hypothyroidism	True	include thyroid hormones
4	age0_18	True	include MMI exclude PTU
	prediag_graves	True	

Note: Features are named in a way of "*category_item*". *Prediag* means previous diagnosis, namely disease history. Abbreviations: cvd for cardiovascular disease, thyroid_crisis for thyrotoxic crisis, graves for Graves' disease, age_0_18 stands for patients with age in range of 0 to 18. MMI for methimazole which is the representative of sulfur-containing imidazole derivatives, PTU for propylthiouracil which belongs to thiouracils.

Rule Extraction Result from Guidelines

According to the management guidelines of hyperthyroidism and the related reference [12][13], we summarized 23 rules for medication recommendations. For simplicity of presentation, we only selected top 4 rules (see Table 2) ranked by covered sample number in our data to illustrate our approach. Each rule is represented in the *if-then* fashion, including the feature and its value in rule's condition, and recommended medication. For example, the rule of ID 4 represents that patients with age<=18 and with disease of graves should be treated with MMI therapy, but avoiding PTU therapy.

Data Preprocessing

Medication data was cleaned by fuzzy matching the generic name with the ATC (Anatomical Therapeutic Chemical) coding system and 5-digit ATC code was used to represent the category of drugs. Our recommendation is also on the level of drug categories instead of drug generic name as the medication principles mentioned in guidelines are mainly at this level. Another reason is that statistic errors and bias would be much larger with drug generic names than with drug categories. In total, seven drug categories are identified for hyperthyroidism medication: sulfur-containing imidazole derivatives (ATC code: H03BB), thiouracils (ATC code: H03BA), non-selective β blocking agents, (ATC code: C07AA), selective β blocking agents (ATC code: C07AB), thyroid hormones (ATC code: H03AA), glucocorticoids (ATC code: H02AB) and various thyroid diagnostic radiopharmaceuticals (ATC code: V09FX).

Diagnosis and history diseases data were prepared and standardized using ICD10. Records about image tests and laboratory tests were standardized by mapping to our standard clinical term sets.

Feature Selection and Medication Pattern Mining

After data-preprocessing, feature selection was carried out to make decision tree model with better performance [14]. Features were selected based on following criteria. First, features with less than 50 samples of true value were excluded. Second, we dropped the features with no obvious relationship with any drug category, which was quantified by

Table 3. Description of medication pattern for hyperthyroidism.

Index	Medication Patterns	Ratio
0	Glucocorticoids	0.9%
1	Various thyroid diagnostic radiopharmaceuticals	0.5%
2	Selective β blocking agents	1.7%
3	**Sulfur-containing imidazole derivatives**	**36.8%**
4	Glucocorticoids + Sulfur-containing imidazole derivatives	0.2%
5	Selective β blocking agents + Sulfur-containing imidazole derivatives	3.0%
6	**Thyroid hormones**	**3.9%**
7	Selective β blocking agents + Thyroid hormones	0.1%
8	**Sulfur-containing imidazole derivatives + Thyroid hormones**	**4.6%**
9	Selective β blocking agents + Sulfur-containing imidazole derivatives + Thyroid hormones	0.4%
10	Non-selective β blocking agents	3.3%
11	Glucocorticoids + Non-selective β blocking agents	0.1%
12	**Sulfur-containing imidazole derivatives + Non-selective β blocking agents**	**14.5%**
13	Thyroid hormones + Non-selective β blocking agents	0.1%
14	Sulfur-containing imidazole derivatives + Thyroid hormones + Non-selective β blocking agents	0.8%
15	**Thiouracils**	**18.1%**
16	Selective β blocking agents + Thiouracils	1.6%
17	Thyroid hormones + Thiouracils	1.4%
18	**Non-selective β blocking agents + Thiouracils**	**7.7%**
19	Thyroid hormones + Non-selective β blocking agents + Thiouracils	0.3%

Note: In combination prescriptions, each drug category is connected using the "+" sign. Medication patterns with ratio larger than 3.5% are marked in bold.

relative risk (RR) analysis between each feature and each drug category. In this way, feature number was reduced from larger than 200 to about 20. Combining with features in the extracted rules, there were 28 features in total, including 4 image test features, 6 laboratory test features, 17 features about previous diagnose/disease and 1 feature about basic information. All these features were selected as data input in decision tree construction.

Once feature part was done, we need to process the label part by identifying the medication patterns. Here we used association rule mining method (Aprior algorithms [15]) to identify the frequent sets of medications with a support threshold of 0.1%. As shown in Table 3, there were 20 medication patterns identified for hyperthyroidism treatment. Among all the prescriptions, the most popular one was sulfur-containing imidazole derivatives only, which accounted for about 37% of samples. This is in consistent with the fact that sulfur-containing imidazole derivatives is the first priority in antithyroid drug therapy except for special cases [12][13].

Nested Decision Tree Construction

The construction of nested guideline-based decision tree contains two steps. First, we selected rule-obeyed samples, as defined in the method section, to construct the rule-based tree. Among the total 83360 medical records, only 4246 of them could satisfy the

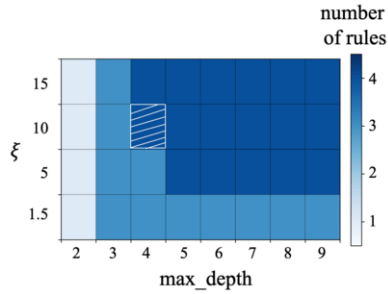

Figure 2. Parameter exploration for rule-based tree: the grids in the plot are colored according to the number of rules emerged on the decision trees.

condition in one or more rules and were selected as rule-covered samples, since the guideline only covers patient populations with limited clinical conditions. In rule-covered samples, the samples which had prescription consistent with rule recommendations were selected as rule-obeyed samples for train set and the rest were chosen as test set. Second, rule-uncovered samples were used to further expand the rule-based tree, where 80% of these samples were chosen as the train set and the rest 20% as the test set. In that way, a nested decision tree was constructed and it could be validated that recommendations given by this tree not only obey guidelines, but also are more patient-specific.

Rule-Based Decision Tree Construction

The key idea to construct a rule-based decision tree is adding a bias factor ξ to rule-related features when selecting features to split the tree. Therefore, we first explored the parameter space to maximize the number of rules emerged on the rule-based decision tree. For explanation, if a rule was emerged on the tree, we mean all features in a rule are used as the splitting features in one or more paths of the decision tree.

In the decision tree construction, three parameters are involved. Figure 2 shows the change of number of rules emerged on the decision tree with max_depth and ξ in a fixed min_samples_leaf. With min_samples_leaf setting to 0.3% of the sample size of train data, we tested ξ with 1.5, 5, 10, 15 and max_depth from 2 to 9, we found that the best case where all rules are emerged in the tree occurs when ξ and max_depth are large. Finally, we chose ξ to be 10, and max_depth to be 4 as a balance between rule compliance and complexity of the tree. The chosen parameters enabled all 4 rules to be satisfied in the rule-based tree as denoted in a slash-shaded box in Fig. 2.

Further Expansion to Nested Tree

With the rule-based tree ready, a nested tree is further built with rule-uncovered samples. Figure 3 shows the final nested tree with the rule-based tree on the top and two giant branches derived from rule-uncovered data at the bottom. Abbreviations of feature names are listed on the upper-right in the figure. The leaf nodes are represented in curved boxes where top three medication patterns are listed with corresponding ratios. Note that, for the rule-based tree, the satisfied (emerged) rules are marked in the corresponding paths.

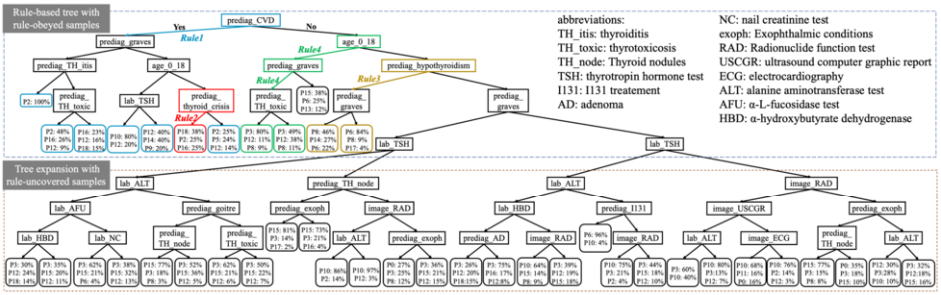

Figure 3. The final nested tree: the upper dashed box denotes rule-based tree and the lower dash box shows the tree expansion using rule-uncovered data.

From Fig. 3, some branches refine the rule-based decision tree. For example, a feature named *prediag_TH_toxic* (meaning thyrotoxicosis) could further split Rule4-satisfied samples into two groups where the preference to treatment pattern 3 and 12 are diverged. The only difference between treatment pattern 3 and 12 is the existence of non-selective β blocking agents as listed in Table 3. Therefore, according to this partition, patients who were not suffered from thyrotoxicosis have higher chance to take non-selective β blocking agents than those who were. Such information revealed by data refines the rule-based tree and provide preciser partition to the whole cohort.

In addition to refining the rule-based tree, the rule-uncovered data contributes in generating new branches that supplement the tree. Those new branches contain features of diseases history (prefix: *prediag_*) as well as laboratory (prefix: *lab_*) and image (prefix: *image_*) test which enrich our knowledge of hyperthyroidism treatment.

Assessment of Nested Decision Tree Model

As our nested decision tree was constructed by a combination of rule-based part and data-based expansion part with respective training sets, we also assessed the tree separately base on two kinds of test sets as mentioned before. For the rule-based part, the objective was to achieve best compliance with guideline rule, so the adherence of medicine recommendations with extracted rules was evaluated. Note that the true labels in this assessment were the output of applied rules. For data-based expansion part, we compared the agreement of the top three recommendations with true prescription in the data.

To evaluate our nested decision tree in the aforementioned two perspectives, we constructed a baseline decision tree using general CART algorithm and selected the same tree parameters (e.g. max_depth) with the nested decision tree. We used rule-covered test samples to validate the guideline adherence of the nested tree, while the rule-uncovered test samples were used to test the accuracy of nested tree. The result is summarized in Table 4. We found that for rule-covered test data, the guideline compliance of medication recommendation was significantly increased using our model compared with the baseline. For rule-uncovered test data, the accuracy of top 3 recommendations using our model was slightly lower than the baseline due to the early partition by rule-related features.

Table 4. Assessment of the nested and baseline decision trees.

Indicator	Test set	Nested tree	Baseline
Adherence of top 1 recommendation with rules	Rule-covered test data	95.2%	7.0%
Adherence of top 3 recommendations with rules*	Rule-covered test data	98.1%	86.4%
Accuracy of top 3 recommendations	Rule-uncovered test data	69.5%	70.5%

Note: * means any result in top 3 recommendations agrees with rules.

Discussion

In this paper, we proposed a method to build a nested decision tree that can provide guideline-concordant and patient-specific medication recommendation. As mentioned in [16], integrating information from guideline and data becomes a trend in electronic health record data mining. By combining knowledge-driven and data-driven models, our approach enables the personalized recommendation and ensures guideline adherence.

Our approach was tested with hyperthyroidism data. The method can be generalized to other diseases. By adjusting related parameters, our approach can be applied in different scenarios. For example, in case of strong demand to make rules occurring on the decision tree, ξ should be large. Proper parameter setting requires to adjust back and forth according to the research purpose. In case that there is no rule-covered sample, the rule-based tree can still be constructed by enumerating all combinations of the extracted rules.

Patterns revealed by the data can provide us hints of potential important considerations in treatment, which can be validated through randomized controlled trials before being adopted by clinical guideline.

The limitation of current work is that the effect of our approach largely depends on the quality of original medical data, especially in the tree expansion phase. Besides, decision tree is a weak classifier and it remains open to investigate combining clinical knowledge with more advanced data-driven models. At current stage, we used decision tree as it was straightforward to visualize the integrated result. More efforts would be invested to explore combination of clinical knowledge with other machine learning methods in the future.

Conclusions

In this paper, we proposed a method to build a guideline-based decision tree for medication recommemdation, by leveraging both the clinical knowledge and electronic medical records. In this way, fine-granular personalized medication recommendations with good guideline adherence can be achieved. The effectiveness of our approach was validated in a real-world case study of patients with hyperthyroidism. Compared with general decision tree method, our approach showed improved guideline adherence with slight compromise on accuracy in the test data.

References

[1] M. A. Musen, B. Middleton, and R. A. Greenes, Clinical decision-support systems, In Biomedical informatics 2014: 643-674.

[2] E. S. Berner, Clinical decision support systems: state of the art. (Agency for Healthcare Research and Quality: Publication No. 09-0069-EF 2009).

[3] S. Anand, V. Y. Fan, J. Zhang, L. Zhang, Y. Ke, Z. Dong, and L. C. Chen, China's human resources for health: quantity, quality, and distribution, *Lancet* **372** (2008), 1774-1781.

[4] J. Mei, H. Liu, G. Xie, S. Liu, and B. Zhou, An OCL-compliant GELLO Engine, *Studies in Health Technology and Informatics* **169** (2011), 130-134.

[5] J. Mei, H. Liu, G. Xie, et al, An engine for compliance checking of clinical guidelines, *Studies in Health Technology and Informatics* **180** (2012), 416-420.

[6] R. C. Chen, Y. H. Huang, C. T. Bau, et al. A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection, *Expert Systems with Applications* **39** (2012), 3995-4006.

[7] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, CADRE: Cloud-assisted drug recommendation service for online pharmacies, *Mobile Networks and Applications* **20** (2015), 348-355.

[8] J. H. Chen, and R. B. Altman, Automated Physician Order Recommendations and Outcome Predictions by Data-Mining Electronic Medical Records, *AMIA Summits on Translational Science Proceedings* (2014), 206-210.

[9] J. H. Chen, R. B. Altman, Data-Mining Electronic Medical Records for Clinical Order Recommendations: Wisdom of the Crowd or Tyranny of the Mob? *AMIA Summits on Translational Science Proceedings* (2015), 435-439.

[10] H. Liu, X. Li, G. Xie, X. Du, P. Zhang, C. Gu, and J. Hu, Precision Cohort Finding with Outcome-Driven Similarity Analytics: A Case Study of Patients with Atrial Fibrillation, *MedInfo* (2017), 491-495.

[11] S. Shan, Decision Tree Learning. *Machine Learning Models and Algorithms for Big Data Classification.* Springer US (2016), 1-28.

[12] Z. Meng, and J. Tan, Reading and analysis on management guidelines for hyperthyroidism published in 2011 by American Thyroid Association and American Association of Clinical Endocrinologists, *International Journal of Radiation Medicine and Nuclear Medicine* **35** (2011), 193-201.

[13] R. S. Bahn, H. B. Burch, D. S. Cooper, et al, American Thyroid Association and American Association of Clinical Endocrinologists Taskforce on Hyperthyroidism and Other Causes of Thyrotoxicosis, *Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists, Thyroid,* **21** (2011), 593-646.

[14] P. Langley, Selection of relevant features in machine learning, *In Proceedings of the AAAI Fall symposium on relevance* **184** (1994): 245-271.

[15] R. Agarwal, and R. Srikant, Fast algorithms for mining association rules, *In Proceedings of the 20th VLDB Conference* (1994), 487-499.

[16] J. Sun, J. Hu, D. Luo, et al. Combining knowledge and data driven insights for identifying risk factors using electronic health records, *In AMIA Annual Symposium Proceedings* **2012** (2012).