SYMPTOMS REDUCTION USING ROUGH SETS BASED TRADITIONAL CHINESE MEDICINE STOMACHACHE DIAGNOSIS

RUIXIANG WANG, YANG ZHOU, HAO YUAN AND XINGLIANG QI

Institute of Chinese Medical History and Literature
Shandong University of Traditional Chinese Medicine
No. 4655, Daxue Road, Jinan 250355, P. R. China
xinxiguanliyu@163.com

ABSTRACT. *As an important concept in rough sets theory, attributes reduction can provide "the most valuable symptoms" for different stomachache syndromes in Traditional Chinese Medicine (TCM). To reduce the symptoms related to stomachache to meet the requirements of TCM expert systems, including the different diagnostic experiences of TCM specialists and the fewest core symptoms, this study used Rosetta, which combines a heuristic search algorithm that allows core attribute selection and determination, to reduce symptoms based on 18 decision tables according to the TCM syndrome differentiation theory and the rough sets theory. Finally, 18 symptoms reduction sets to differentiate each TCM syndrome and 30 core symptoms of stomachache were achieved based on the rough sets theory and 279 TCM diagnostic cases. This study shows that it is useful to attempt an approach utilizing sound mathematical techniques to reduce redundant symptoms in the differentiation of TCM syndromes.*
Keywords: Data mining, Expert system, Rough sets, Attribute reduction

1. **Introduction.** In ancient times, eminent Chinese physicians explored the association between the symptoms of a certain disease and the application of herbs in the practice of medicine, thereby creating the logical theoretical basis for TCM. As a part of TCM theory, syndrome differentiation is the theoretical analysis of clinical information gained using the four main TCM diagnostic methods. The goal of syndrome differentiation in TCM is to determine the syndrome of an illness according to the symptoms. The syndrome is the etiology and pathogenesis of the illness on the ground of TCM theory.

In recent years, people have attempted to discover the mechanism of TCM theory using experimental approaches on the one hand; while on the other hand, researchers hope to find a logical algorithm to replace experienced TCM specialists, thereby ignoring the reasoning of TCM theory and depending only on medical records.

Researchers in China have made significant contributions in many fields of artificial intelligence, and artificial intelligence in TCM is a very important field. In the work of [1,2], some self-learning expert systems for diagnosis in TCM have been built by incorporating several data mining techniques such as the naive Bayes classifier, fuzzy and certainty reasoning.

Although some of these TCM expert systems have a satisfactory ability to identify the patterns of syndromes, these systems suffer from certain limitations, such as users in real-life being hampered by redundant symptoms. TCM medical records reveal that different TCM doctors pay attention to different symptoms for a certain disease due to varied experiences and theories. Those learning TCM and researchers in TCM expert systems are puzzled by a mass of symptoms. Therefore, most TCM expert systems were designed for a certain TCM specialist with a limited number of symptoms. In the work of [3,4], the You-bo Guan TCM expert system and Jian-hua Dong TCM expert system were built. However, for a certain disease, the medical records from a single TCM doctor

are limited, so researchers cannot gain enough rules to construct a sufficiently effective TCM expert system. Symptoms redundancy thus leads to an exploration of symptoms reduction.

Our research in TCM expert systems began in 2010 when we collected 279 TCM stomachache cases obtained from the literature from books in the Han dynasty to papers in the current academic journals. A decision table included 279 records; 128 symptoms condition attributes and 18 diagnosis decision attributes were created. Clearly, there are too many symptoms to meet the requirements of expert systems. One approach to this problem is to reduce redundant attributes.

Rough set theory is a new mathematical tool to address vagueness and uncertainty. The reduction of knowledge based on a rough set model is an effective mathematical tool to reduce redundant attributes in a relational database system.

The application of rough sets to the field of TCM has been reported in several papers. In the work of [5], the authors investigated the application of singular rough sets to increasing or decreasing the prescribed dose of drug in TCM physician diagnosis-recognition. In the work of [6], the authors mined rules by applying rough sets theory to the clinical records of asthma in TCM. In the work of [7], the authors gained rules by applying rough sets theory to the clinical records of rheumatoid arthritis.

Stomachache is a common type of disease for TCM. The present study attempts to reduce symptoms using rough sets theory based on the diagnostic cases of various TCM specialists. According to TCM syndrome differentiation theory, stomachache is grouped into 18 syndromes, including Qi stagnation, YANG deficiency, food stagnation, etc. The reason for the classification is the difference in symptoms. The classification leads to different treatment prescriptions.

The goal of our study was to reduce symptoms based on rough sets models without a loss of classification information.

2. **Philosophy and Notions of Rough Sets.** In the work of [8], proposed by Pawlak in 1982, rough sets are a novel mathematical approach to knowledge discovery. The rough sets philosophy is founded on the assumption that we associate some information with every object in the universe of discourse. For example, if objects are patients suffering from a certain disease, the symptoms of the disease form information about the patients. If the symptoms information is adequate, we can differentiate all of the patients according to symptoms. If the patients are grouped into a few classes, we can determine which class a patient belongs to based only on that patient's symptoms information. According to the principle of Occam's razor, if less symptoms information is adequate to determine which class any patient belongs to, then redundant attributes should be cut. The essence of TCM syndrome differentiation is classification according to symptoms for a certain disease. Without the loss of classification information, a reduction maintains the fewest symptoms that can differentiate a patient's etiology and pathogenesis on the grounds of a mathematical analysis.

As a mathematical tool, we recall some basic notions related to rough sets.

An information system is a pair $A = (U, A)$, where U is a non-empty, finite set called the universe and A is a non-empty, finite set of attributes; $a: U \to Va$ for $a \in A$, where Va is called the value set of a.

We consider a special case of information systems known as decision tables. A decision table is an information system in the form of $A = (U, A \cup \{d\})$, where $d \notin A$ is a distinguishable attribute called a decision. The elements of A are called conditions.

For the information system $S = (U, A)$, every subset $IND \subset A$ generates an equivalence relation. An order pair $AS = (U, IND)$ is called an approximation space. Equivalence classes of IND are called elementary sets within AS because these sets represent the smallest discernible groups of objects.

The lower approximation of X in AS is defined as $LOW_{IND}(X) = \{x \in U | [x]_{IND} \subseteq X\}$.

The upper approximation of X in AS is defined as $UPPER_{IND}(X) = \{x \in U | [x]_{IND} \cap X \neq \phi\}$.

The set $UPPER_{IND}(X) - LOW_{IND}(X)$ is called the IND-doubtful region of IND in (U, IND).

A lower approximation forms certain rules, while the IND-doubtful region forms uncertain rules.

The reduction of rough sets is the essential part of an information system that can discern all objects that are discernible in the original information system. In the work of [9,10], mathematicians have presented certain attributes reduction algorithms.

Example 2.1. *Let us consider the following example.*

TABLE 1. TCM diagnosis cases

U	cough	night-sweat	low-fever	syndrome
$X1$	1	1	1	*blood stagnancy*
$X2$	1	1	1	*blood stagnancy*
$X3$	1	1	0	*blood stagnancy*
$X4$	1	1	0	*deficiency-of-blood*

$U = \{X1, X2, X3, X4\}$ is the collection of patients;

$A = \{$cough, night-sweat, low-fever$\}$ is the collection of valuable symptoms of a certain disease;

$D = \{$blood stagnancy, deficiency-of-blood$\}$ is the decision attribute;

The lower approximation of blood stagnancy is $\{X1, X2\}$;

The upper approximation of blood stagnancy is $\{X1, X2, X3, X4\}$;

The IND-doubtful region of blood stagnancy is $\{X3, X4\}$.

The reduction of the decision table is {low-fever} because only symptom including low-fever also discerns all discernible objects; that is, the lower approximation of the decision table, which only includes low-fever, is the same as the original information system.

3. **Symptoms Reduction Based TCM Stomachache Diagnosis.** In this study, 279 TCM stomachache diagnostic cases were collected from the literature from the book Shanghan Zabing Lun to current academic journals.

Let us consider TCM stomachache diagnostic cases in Table 2.

TABLE 2. TCM stomachache diagnostic cases

no	stomach-distending-pain	...	nausea	...	syndrome
1	1	...	0	...	*deficiency-of-blood; deficiency-of-Yin*
⋮			... ...		
279	0	...	0	...	*deficiency-of-spleen-YANG*

The decision attributes from Table 2 include 18 syndromes such as deficiency-of-blood, deficiency of Yin, and blood stagnancy. Due to the atomicity of the decision attribute, we divided Table 2 into 18 tables such as Table 3.

As a rough sets toolkit for data analysis, Rosetta can automatically reduce symptoms with a heuristic algorithm. In accordance with the algorithm, if the lower approximation from a decision table in which a symptom was deleted remains the same as that from a decision table that includes all symptoms, the symptom is a core symptom. We obtained symptoms reduction by deleting all non-core symptoms. The data in Table 3 was analysed using Rosetta software.

TABLE 3. Divided TCM stomachache diagnostic cases

no	stomach-distending-pain	...	nausea	...	deficiency-of-blood
1	1	...	0	...	1
⋮			... ...		
279	0	...	0	...	0

no	stomach-distending-pain	...	nausea	...	deficiency-of-Yin
1	1	...	0	...	1
⋮			... ...		
279	0	...	0	...	0

⋮

no	stomach-distending-pain	...	nausea	...	deficiency-of-spleen-YANG
1	1	...	0	...	0
⋮			... ...		
279	0	...	0	...	1

4. **Primary Results.** The result of this study is a determination of what the fewest symptoms are for each syndrome and what the core symptoms are for a stomachache in TCM.

The following is the symptoms reduction for 6 TCM syndromes for a stomachache diagnosis based on 279 TCM stomachache diagnoses since ancient times:

deficiency-of-YANG: stomach distending pain, acid regurgitation, liking for warmth and fearing to cold, emaciation, belching, cold limbs;

blood stagnancy: anorexia, acid regurgitation, liking for warmth and fearing to cold, stomachache if hungry and remain even if eat, vomit, stomach stabbing pain;

fire stagnation: stomach distending pain, anorexia, sticky stool, dry stool, bitter taste, vomit;

obstruction of pulmonary Qi: cough;

deficiency-of-blood: squamous and dry skin;

coldheat complex: stomach distending pain, anorexia, acid regurgitation, liking for warmth and fearing to cold, emaciation, dry stool, dry mouth, stomachache aggravation with pressure, astriction, vomit, thin lingual fur, red tongue.

There are approximately 30 symptoms that are the combination of the 18 sets of symptoms reduction. They are as follows: stomach distending pain, anorexia, acid regurgitation, emaciation, belching, lacking in strength, palpitation, liking for warmth and fearing to cold, cold limbs, dry mouth, frequent micturition, squamous and dry skin, bitter taste, stomach stuffy pain, astriction, sticky stool, dry stool, nausea, vomit, melena, aversion to cold, watery stool, stomachache if hungry and remain even if eat, stomach stabbing pain, diarrhoea, stomachache relief with pressure, pallor of face, stomach severe pain, sleeplessness, cough.

We regard these 30 symptoms as core symptoms based on the mathematical logic of the rough sets theory. These 30 core symptoms are enough to distinguish each stomachache syndrome of TCM. The symptoms reduction approach based on rough sets can solve the problem of a bottleneck in TCM expert systems in which the knowledge base is derived from the medical records of different TCM specialists.

5. **Discussion.** This symptoms reduction meets the requirement for a TCM diagnostic expert system. Based on symptoms reduction, a TCM stomachache diagnostic expert
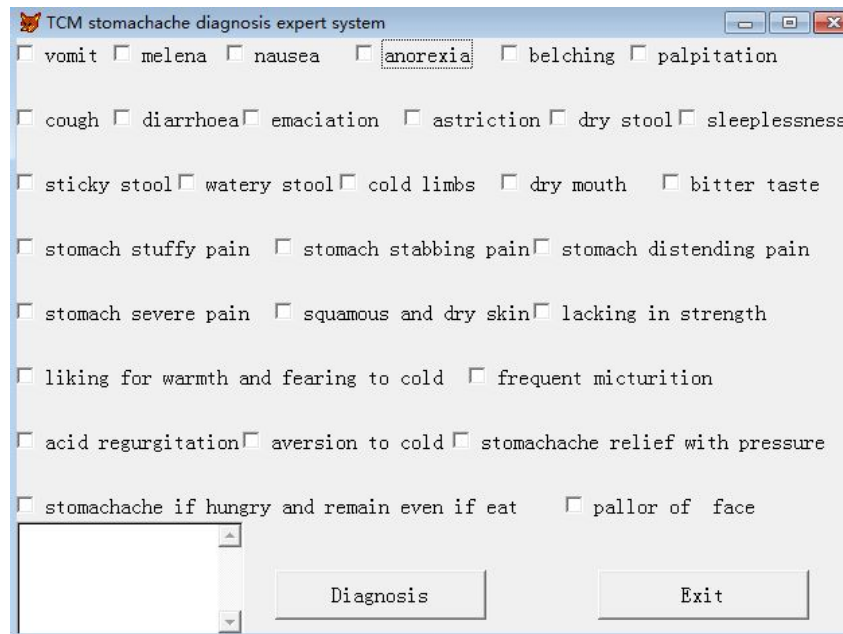
FIGURE 1. Interface of TCM stomachache diagnosis expert system

system was completed. The interface of the TCM stomachache diagnosis expert system is illustrated in Figure 1.

TCM has been practiced for at least 2000 years in China. In practice, for a certain disease, a TCM doctor first determines the etiology and pathogenesis according to the patient's symptoms and TCM theory. We have presented a rough sets method and the role of this approach in the selection of core symptoms for pattern recognition. In this way, we can precisely judge which aetiology and pathogenesis a patient belongs to according to only the fewest symptoms without reasoning based on the TCM theory. This self-learning TCM expert system based on rough sets is a helpful research tool for TCM artificial intelligence diagnoses, from both a philosophical and technological perspective.

Different TCM doctors have different clinical experiences and different perspectives on TCM theory; therefore, different TCM doctors often make different or contradictory diagnoses for a certain patient. The rough sets theory provides a mathematical method for processing uncertain TCM experience. A rough sets reduction can confirm core symptoms for distinguishing TCM syndromes based on the medical records of different doctors. In future studies, incorporating the insights of physiopathology and molecular biology, the core symptoms could provide inspiration to reveal the scientific mechanism of syndrome differentiation in TCM diagnoses.

REFERENCES

[1] X. W. Wang, H. B. Qu, P. Liu and Y. Y. Cheng, A self-learning expert system for diagnosis in traditional Chinese medicine, *Expert Systems with Applications*, pp.557-566, 2003.
[2] J. S. Shieh and Q. Yang, Traditional Chinese Medical diagnosis based on fuzzy and certainty reasoning, *IEEE International Symposium on IT in Medicine and Education*, pp.12-14, 2008.
[3] B. Li and Z. Zhong, The model of fuzzy mathematics in hepatitis diagnosis of Chinese traditional medicine, *Journal of Guangxi Medical University*, vol.15, no.1, pp.52-54, 1998.

[4] L. Chen and H.-Q. Li, Study on the expert system of animal model with traditional Chinese medicine symptoms evaluation, *Chinese Archives of Traditional Chinese Medicine*, vol.28, no.1, pp.83-84, 2010.

[5] Y. M. Hu and L. Jia, Knowledge support system of doctor diagnosis-recognition based on singular rough sets, *Computer Engineering and Applications*, vol.16, no.6, pp.31-32, 2004.

[6] H.-C. Yan, L.-H. Li, H.-X. Ma et al., Analysis on clinical medicine asthma medical records association rules based on rough set theory, *Liaoning Journal of Traditional Chinese Medicine*, vol.39, no.7, pp.1218-1220, 2012.

[7] G.-M. Xie, The establishment of TCM diagnosis model based rough sets, *Journal of Mathematical Medicine*, vol.18, no.4, pp.302-304, 2005.

[8] Z. Pawlak, J. Grzymala-Busse, R. Slowinski et al., Rough sets, *Communications of the ACM*, vol.38, no.11, pp.88-95, 1982.

[9] J. G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak and J. Wróblewski, Rough set algorithms in classification problem, *Rough Set Methods and Applications*, pp.1-40, 2000.

[10] F. Wang, J. Xu and L. Li, A novel rough set reduct algorithm to feature selection based on artificial fish swarm algorithm, *Advances in Swarm Intelligence*, pp.24-33, 2014.