

RESEARCH ON THE DIAGNOSIS AND TREATMENT PROCESS OF DERMATOSIS BASED ON DATA-DRIVEN APPROACH

XINGLIANG QI¹, CHIN WEI², TING YUAN¹, ZHENGUO WANG¹ AND YANG ZHOU¹

¹Institute of Chinese Medical History and Literature
Shandong University of Traditional Chinese Medicine
No. 4655, Daxue Road, Jinan 250355, P. R. China
zhenguow@126.com

²Institute of Scientific and Technical Information of China
No. 15, Fuxing Road, Haidian District, Beijing 100038, P. R. China

Received April 2016; accepted July 2016

ABSTRACT. *Taking dermatosis as an example, this study makes use of natural language processing (NLP) techniques and methods, segments and tags the texts automatically, reorganizes the knowledge of dermatosis, and constructs specialized corpus and Ontology knowledge base to diagnose and treat dermatosis. The purpose is to supply the gap of unbalance of doctors' personal level of the dermatosis diagnosis and treatment, improve the level and accuracy of clinical diagnosis and treatment of dermatosis, and to achieve effective drive of medical literature data for clinical diagnosis and treatment.*

Keywords: Dermatoses, Clinical diagnosis and treatment, Natural language processing, Corpus, Ontology

1. Introduction. Traditionally, the diagnosis process and treatment is based on the first symptoms of the patients. After an analysis of the first symptoms, doctors gradually get more diagnosis and treatment value of the symptoms, signs or laboratory tests, test results, and obtain the diagnosis and treatment plan. In the diagnosis process and treatment, the thought of the doctors and the knowledge of the diseases play a vital role. If the doctors' medical knowledge is not enough, they are prone to be of misdiagnosis, of poor efficacy, etc. On the other hand, after nearly a hundred years of development, the study of all the clinical symptoms, signs, and laboratory tests on dermatosis, has been mature. A large amount of literature has been produced. So how to make full use of this knowledge becomes a problem to be worth studying.

In the view of the problems mentioned above, we construct the process of dermatosis diagnosis and treatment system based on the data-driven approach in order to fully use the vast amount of literature, and to mine its mature diagnosis and treatment method and technology. Firstly, we farthest collect the literature. Then, we carry out data processing and knowledge organization by using the methods and techniques of NLP. Finally, using the existing diagnostic criteria, disease diagnosis standard items are ranked according to the level of diagnostic value to distinguish between the diagnostic weight and select the correct node of diagnosis and treatment. It will ensure the diagnosis process and treatment to be highly efficient without danger.

So, the work in this paper focuses on the following aspects. Firstly, according to the existing knowledge, we form the multi-layer concept system of dermatosis. In this system, the corpus of dermatosis is pretreated, automatically tagged and identified in order to construct the dermatosis corpus. Secondly, based on context calculation model, we obtain the semantic recognition of dermatosis, form the concept of dermatosis, and construct the dermatosis Ontology. Then in natural language we describe (the symptoms,

signs and laboratory findings are mainly described.) the feature information of dermatosis and carry out semantic retrieval (analyzing, reasoning on Ontology, etc.). Finally, we construct a corpus based and Ontology based model of diagnosis process and treatment of dermatosis. In this paper, we realize the construction of the diagnosis process and treatment of dermatosis based on data-driven approach by using question-answer style, which significantly improves the efficiency of dermatosis's diagnosis process and treatment and reduce the errors.

2. Construction of Dermatitis Corpus and Ontology. This paper describes the automatic diagnosis process and treatment based on the dermatosis corpus and Ontology. The general program and technical method are shown in Figure 1 which can be referred to in the above paragraph.

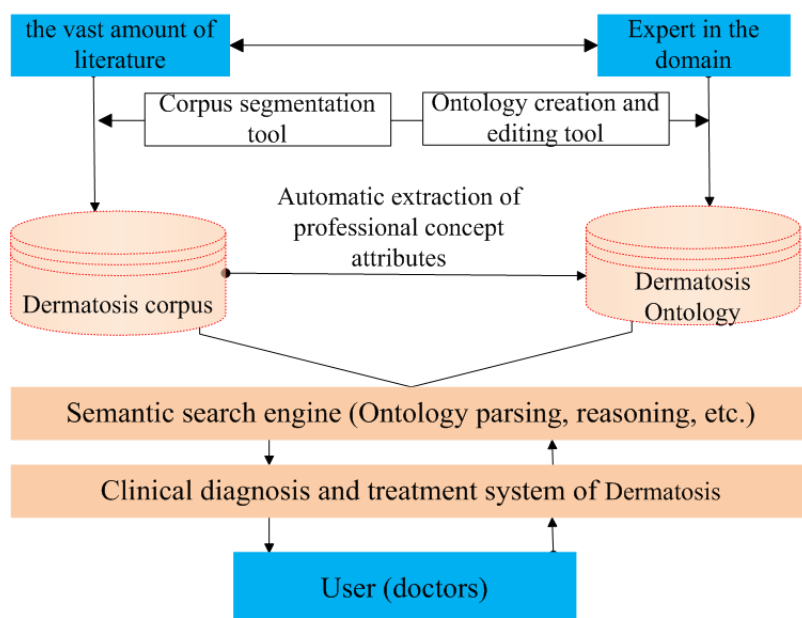


FIGURE 1. The general program and technical method

2.1. Construction of dermatosis corpus. Knowledge in the corpus reflects the actual use of language. However, the knowledge contained in the original data is unstructured and implicit. The tacit knowledge can be revealed by reformatting the form of the word with some tags [1]. In order to construct a corpus of dermatosis, we need to classify the contents of the text of the dermatosis, determine the principles of word segmentation and part of speech tagging, and then construct the semantic dictionary of dermatosis.

2.1.1. Classification of dermatology knowledge. This study carries out the classification of dermatosis by using the existing classification system of dermatosis, such as MESH, and Dermatology Textbook. Our classification includes two categories: the general medical knowledge and the professional knowledge of dermatology.

2.1.2. Construction of the corpus. Language information processing system also needs the support of professional knowledge. We use the tool “Chinese Word Segmentation and Part-of-speech Tagging” developed by Peking University [2], to generate a primary corpus. Then, we construct a professional corpus of dermatology by using the segmentation dictionary of dermatology to mark the label of the words.

2.1.3. *Dermatology semantic dictionary.* The semantic dictionary of dermatology is a machine dictionary in the domain. It can be used in the language information processing directly. In this dictionary, the class system for information processing is set up to complete the classification of words, and researchers can describe the different attributes of each word according to the classification. The dictionary uses two-dimensional form of the relational database to describe the grammatical properties of the word, establishes a file according to the parts of speech and describes a series of attributes of a word using a record. This dictionary plays an important role in the semantic filtering of literature knowledge discovery.

2.2. **Construction of dermatosis Ontology.** The dermatosis Ontology mainly contains the definition of the concept category, properties and instances.

2.2.1. *Construction of the top-level concept category.* The concept relation of the top level refers to the concept of the highest level in the network and multidimensional concept, which plays a leading role in the whole concept of system structure [3]. The classification of the concept of dermatology, in fact, is to use the dermatosis Ontology to restructure the knowledge. In this paper, the concept category of the top layer of the dermatosis Ontology inherits the knowledge classification system of the dermatosis specialized corpus.

2.2.2. *Construction of the attribute description system.* Concept is a reflection of the unique attributes. From general attributes to essential attributes, it reflects the understanding of knowledge from the shallower to the deeper [4]. Attribute is used to describe the nature of the concept. It is a feature that distinguishes it from other concepts. It is a link among the concepts. Attribute explicitly defines the concept. Reconstruction of dermatology knowledge description system is the use of a variety of related attributes of conceptual description.

From the natural language analysis and knowledge mining angle to obtain the construction method and technology, the attribute description of each concept is divided into three ways: an overview description, NLP semantic category, and professional description. Professional attributes include two ways: natural language description and subject word description [5].

2.2.3. *Definition of dermatology instances.* Instances contain static data instances, and patient data instances. Instances of static data contain dermatosis classification coding (classification codes) and drug classification (drug classes). Patient data instance is the actual patient data in electronic prescription records, including prescription and laboratory data.

3. Construction of Clinical Diagnosis and Treatment of Dermatoses Based on Data-Driven Approach.

3.1. **Diagnosis process of dermatosis.** Diagnosis process is to start with one or more symptoms or signs, and then to be led by doctors. Taking complaining scalp erythema as an example, doctors firstly confirm scalp erythema to be the first symptom, and then eliminate inducement contact, and finally check if there is Auspitz's sign. Auspitz's sign can occur in psoriasis and seborrhea. Based on the massive medical real data and the semantic analysis of all the symptoms of the disease, dermatosis diagnosis process and treatment system will automatically identify the following three causes step by step: family history, pruritus, and bumpy hair. If the patient has family history or bumpy hair, he or she can be diagnosed as psoriasis. If the patient has symptoms of pruritus, he or she will be more likely to be diagnosed as seborrhea instead of psoriasis. If the diagnosis remains uncertain after the process mentioned above, the system will further tell doctors to do pathological examination for the final diagnosis.



FIGURE 2. Dermatitis knowledge Q&A



FIGURE 3. Search results for the disease related to the “erythema”

We construct an instance of a dermatosis diagnosis process as Figure 2.

In the department of dermatology, erythema is a very common symptom, associating with a lot of diseases. In the question and answer system, we enter the word – “erythema”. Then, we will get the result of diseases, such as the Steven-Johnson syndrome, multiple myeloma, and erythema-rash, and inflammatory dermatosis, as shown in Figure 3.

In the actual diagnosis process, the symptoms are more specific, such as “scalp erythema”. When the patient visits with such symptoms, the key node of further diagnosis is that if there are any inducement contacts, such as hair dye, and insect bites. If there is no contact reason, we can add the other key nodes to promote the diagnosis process, as shown in Figures 4 and 5. Key nodes 1 refer to “recently dyed hair or mosquito bites”. Key nodes 2 refer to the concomitant symptoms in Figure 5.

Obviously, the diagnosis process handled by the process mentioned above will be more rigorous, accurate and efficient, and will minimize the misdiagnosis.

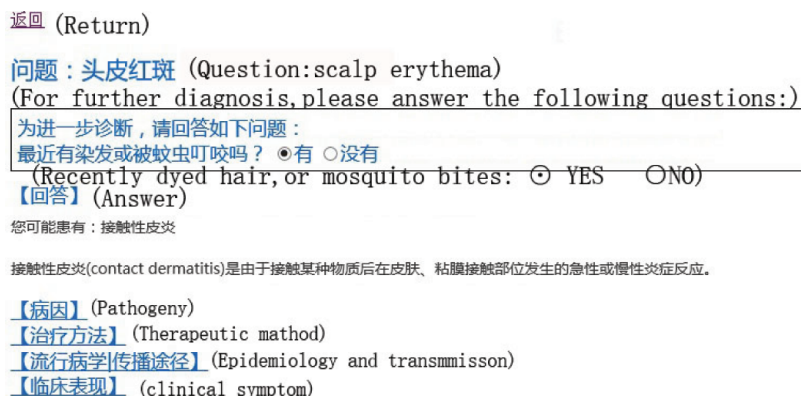


FIGURE 4. Key nodes 1

[返回](#) (Return)

问题：头皮红斑 (Question:scalp erythema)

(For further diagnosis, please answer the following questions:)

为进一步诊断, 请回答如下问题:

最近有染发或被蚊虫叮咬吗? 有 没有

(Recently dyed hair, or mosquito bites: YES NO)

为进一步诊断, 请选择您还伴有的症状: (For further diagnosis, please choose your symptoms:)

头发呈束状 (Hair bundles) (In the hair there is a white sheath around
头发里有像石棉结晶的纯白色鞘状物环绕 the crystallization of asbestos)

白色糠状鳞屑堆积, 黏着在头发根部和头皮而成大片厚痂 (The accumulation of white bran
毛囊口棘状隆起 (Hair follicle bulge) like scales)

鳞屑呈油腻性, 带黄色 (Yellow greasy scales)

头发易折断 (Hair is easy to break.)

发病者是儿童 (The patient is a child.)

红斑中具有黑点 (There are black spots in the erythema.)

红斑上有丘疹 (There is a rash on erythema.)

起水疱 (Blister)

(Enter)

FIGURE 5. Key nodes 2

3.2. Treatment process of dermatosis. During the diagnosis process, we get symptoms, signs and other information in the form of questions and answers based on dermatosis Ontology. Then, the perfect Chinese Herbs matched with the treatment elements can be selected automatically, and the automatic push of Traditional Chinese Herbs is realized.

For example, the main points of diagnosis of psoriasis may be scalp erythema, positive Auspitz's sign, pruritus, and pathological examination results. The Auspitz's sign and pathological examination results are abandoned because they do not have the significance in the Traditional Chinese Medicine prescription drug database. Therefore, the key points of treatment can only include the scalp erythema, and pruritus. And the selection of treatment elements will be simple and accurate based on the premise of professional Ontology and corpus.

In order to achieve the perfect clinical curative effect, we need the perfect essential treatment elements, so that we can achieve a perfect therapeutic scheme from literature. The value treatment elements need to be extracted from the Traditional Chinese Medicine data. It has been realized in the research and construction of the professional data.

4. Conclusions. This paper makes a study on how to reorganize the knowledge of the vast amount of dermatology literature using the technology and method of natural language processing, and to construct the dermatosis Ontology which will provide the knowledge guarantee for each link in the process of diagnosis and treatment. It will help doctors obtain information according to requirement of the dermatosis, mutual relationships and potential links of a variety of information in order to reduce the subjective bias. It can really improve the level and accuracy of doctors' diagnosis and treatment of dermatosis, and can achieve effective drive of medical literature data for clinical diagnosis and treatment.

The application of this research is universal. It can be popularized and applied to the process of diagnosis and treatment of clinical diseases. It provides examples for comprehensive improvement of the level of diagnosis and treatment of clinical diseases. And it will provide theoretical basis and technical support for the development of the related research.

The process of diagnosis and treatment of dermatosis is based on an amount of medical data. The research of this paper is a gradual process, which needs to be improved and perfected in the future to provide a more abundant and accurate basis for the dermatosis diagnosis and treatment.

Acknowledgment. This work was supported by the Open Foundation of Shandong Key Lab of Language Resource Development and Application.

REFERENCES

- [1] E. Atwell, A cross-language methodology for corpus part-of-speech tag-set development, *Corpus Linguistics Conference*, University of Birmingham, 2007.
- [2] Y. Liu, Z. Zhang and Y. Huang, Research and development of semantic annotation platform for scientific literature, *ICIC Express Letters*, vol.10, no.7, pp.1787-1794, 2016.
- [3] Z. Wang, Y. Zhou, J. Zhang and X. Fu, Research on the representations of Chinese medicine property based on ontology, *ICIC Express Letters*, vol.5, no.10, pp.3895-3900, 2011.
- [4] J. Wang, M. Liang and M. Guan, A web-based ontology concept attribute learning method, *Journal of Information & Computational Science*, vol.11, no.14, pp.4953-4963, 2014.
- [5] Y. Liu, Y. Zhao and Z. Sui, Research on automatic construction of medical ontology based on multidimensional model, *Journal of Computational Information Systems*, vol.5, no.10, pp.1725-1733, 2010.