

3. Diday, E.: The symbolic approach in clustering and related methods of Data Analysis. In: Bock, H. (ed.) Proc. IFCS Classification and Related Methods of Data Analysis, Aachen, Germany. North-Holland (1988)
4. Diday, E.: The symbolic approach in clustering and related methods of Data Analysis. In: Bock, H. (ed.) Classification and Related methods of Data Analysis, pp. 673–684. North-Holland, Amsterdam (1988)
5. Goldberger, J., Gordon, S., Greenspan, H.: An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures. In: Proceedings of CVPR, pp. 487–494 (2006)
6. Gowda, K.C., Diday, E.: Symbolic clustering using a new dissimilarity measure. Pattern Recognition 24(6), 567–578 (1991)
7. Katayama, K., Suzukawa, A., Minami, H., Mizuta, M.: Linearly Restricted Principal Components in k Groups. In: Electronic proceedings of Knowledge Extraction and Modeling, Villa Orlandi, Island of Capri, Italy (2006)
8. Katayama, K., Yamaguchi, R., Imoto, S., Matsuura, K., Watanabe, K., Miyano, S.: Clustering for Visual Analogue Scale Data in Symbolic Data Analysis. Procedia Computer Science 6, 370–374 (2011)
9. Kullback, S.: Information theory and statistics. Dover Publications, New York (1968)

Connection between Traditional Medicine and disease

Kotoe Katayama

Institute of Medical Science,
University of Tokyo
4-6-1 Shirokanedai, Minato-ku Tokyo,
Japan 108-8639
k-kata@ims.u-tokyo.ac.jp

Rui Yamaguchi

Institute of Medical Science,
University of Tokyo
4-6-1 Shirokanedai, Minato-ku Tokyo,
Japan 108-8639

Seiya Imoto

Institute of Medical Science,
University of Tokyo
4-6-1 Shirokanedai, Minato-ku Tokyo,
Japan 108-8639

Keiko Matsuura

Center of Kampo Medicine,
Keio University School of Medicine
35 Shinano-machi, Shinjyuku-ku Tokyo,
Japan 160-8639

Kenji Watanabe

Center of Kampo Medicine,
Keio University School of Medicine
35 Shinano-machi, Shinjyuku-ku Tokyo,
Japan 160-8639

Seiya Imoto

Institute of Medical Science,
University of Tokyo
4-6-1 Shirokanedai, Minato-ku Tokyo,
Japan 108-8639

ABSTRACT

In Japanese traditional medicine, “Monshin” plays an important role. “Monshin” is a questionnaire that asked the patient's lifestyle and subjective symptoms. Specialists decide traditional herbal medicine by using of “Monshin”. In this research, we connect “Monshin” to disease through building the Network.

Keywords

traditional Medicine, regression model, elastic net, network.

1. INTRODUCTION

Interest in traditional medicine has increased globally in recent years. The World Health Organization (WHO) recommends the use of traditional medicine for care and treatment of people with health conditions. Further, the WHO suggests the integration of traditional medicine into the next edition of the International Statistical Classification of Diseases and Related Health Problems (ICD-11).

Kampo medicine -the Japanese adaptation of traditional medicine- originated in ancient China and arrived in Japan via the Korean peninsula in the 6th century, after which it developed uniquely. Kampo medicine was approved in 1967 under the National Health Insurance policy. Of the 50 years of National Health Insurance for whole nations, 44 years dealt with Japanese traditional medicine.

A remarkable feature of the Japanese medical licence is that a doctor with that license is allowed to combine modern and Kampo medicine for a medical treatment and prescribe both kinds of drugs; while in most of other countries, e.g., China and Korea, they have different licenses for modern and traditional medicine, thus, it is hard to utilize them in a combinatorial way.

2. “Monshin” NETWORK

In Japanese traditional medicine, “Monshin” plays an important role. “Monshin” is a questionnaire that asked the patient's lifestyle and subjective symptoms. Specialists decide traditional herbal medicine by using of “Monshin”.

However, in order to determine herbal medicine or traditional cure, technical knowledge and experience are required. In Keio University School of Medicine, we analyze “Monshin” data to establish an indicator for non Kampo specialist without technical knowledge to perform suitable traditional medicine. In this

research, we connect “Monshin” to disease through building the “Monshin” Network. To build it, we identified items of “Monshin” relevant to a disease by using of logistic regression model and elastic net.

3. RESULT

We focus atopy as a target disease and build the “Monshin” Network and show it in figure 1 and 2.

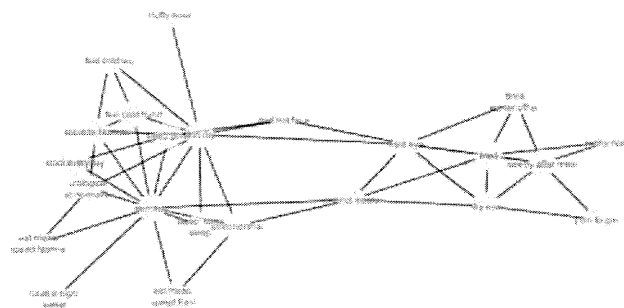


Figure 1. Monshin Network for Atopy (positive)

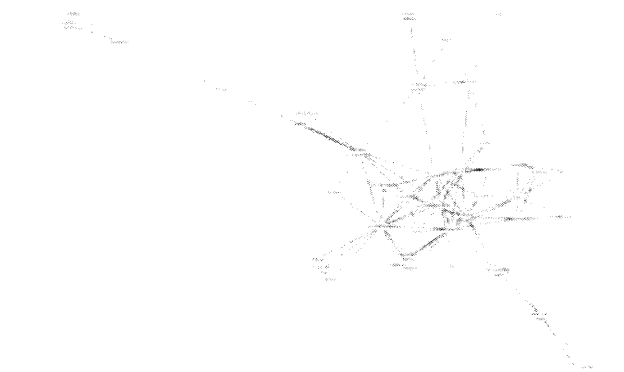


Figure 2. Monshin Network for Atopy (negative)

4. ACKNOWLEDGMENTS

This study was partially supported by Health and Labour Sciences Research Grants for Clinical Research and Research on Statistics and Information from the Ministry of Health, Labour, and Welfare of Japan.

Analysis of questionnaire for Traditional Medical and develop decision support system

Kotoe Katayama

Institute of Medical Science
University of Tokyo

4-6-1 Shirokanedai, Minato-ku Tokyo, Japan 108-8639
k-kata@ims.u-tokyo.ac.jp

Seiya Imoto

Institute of Medical Science
University of Tokyo

4-6-1 Shirokanedai, Minato-ku Tokyo, Japan 108-8639

Kenji Watanabe

Center of Kampo Medicine,
Keio University School of Medicine

35 Shinano-machi, Shinjyuku-ku Tokyo, Japan 160-8582

Rui Ymaguchi

Institute of Medical Science
University of Tokyo

4-6-1 Shirokanedai, Minato-ku Tokyo, Japan 108-8639

Keiko Matsuura

Center of Kampo Medicine,
Keio University School of Medicine

35 Shinano-machi, Shinjyuku-ku Tokyo, Japan 160-8582

Satoru Miyano

Institute of Medical Science
University of Tokyo

4-6-1 Shirokanedai, Minato-ku Tokyo, Japan 108-8639

Abstract—*In Japanese traditional medicine, “Monshin” plays an important role. “Monshin” is a questionnaire that asked the patient's lifestyle and subjective symptoms. Specialists decide traditional herbal medicine by using of “Monshin”. In this research, we analyze “Monshin” and predict “Sho” which is the name of a disease.*

Keywords- *Traditional medicine; random forest; machine learning;*

I. INTRODUCTION

Interest in traditional medicine has increased globally in recent years. The World Health Organization (WHO) suggests the integration of traditional medicine into the next edition of the International Statistical Classification of Diseases and Related Health Problems (ICD-11).

Kampo medicine -the Japanese adaptation of traditional medicine- originated in ancient China and arrived in Japan via the Korean peninsula in the 6th century, after which it developed uniquely. A remarkable feature of the Japanese medical license is that a doctor with that license is allowed to combine modern and Kampo medicine for a medical treatment and prescribe both kinds of drugs; while in most of other countries, e.g., China and Korea, they have different licenses for modern and traditional medicine, thus, it is hard to utilize them in a combinatorial way.

In Japanese traditional medicine, “Monshin” plays an important role. “Monshin” is a questionnaire that asked the patient's lifestyle and subjective symptoms. Specialists decide traditional herbal medicine by using of “Monshin”.

However, in order to determine herbal medicine or traditional cure, technical knowledge and experience are required. In this paper, we analyze “Monshin” data to establish an indicator for non-Kampo specialist without technical knowledge to perform suitable traditional medicine.

II. DIAGNOSIS BY KAMPO

The diagnosis by Kampo is called “Sho” and determined by completely different view from Western medicine. Kampo uses a unique cognitive paradigm, making use of tools such as “Deficiency and Excess”, “Cold and Heat”, Qi, Blood, Fluid. The diagnosis by Kampo is composed one item from each “Deficiency and Excess” category and “Cold and Heat” category, and two items from “Qi, Blood, Fluid” category. Specialists will gather all available information to best determine patients’ “Sho”. There are some types of Kampo examinations: “Monshin”, tongue diagnosis, Palpation, and so on. Kampo specialists consider all the various factors together. However it causes difficulties for non-Kampo specialist. It is hard for them to connect result of “Monshin” and other diagnosis.

III. DATA OF KEIO UNIVERSITY

Since 2006, Center of Kampo medicine, Keio university school of medicine has been collected data about patients’ “Monshin”, “Sho”, western disease name (ICD-10 code), and prescribed herbal medicine. Patients enter “Monshin” information via touch panel operation. “Monshin” has 362 items, ranges in content from physical sign to food preference and is important for Kampo diagnosis. There are

two type questions, yes-no question and VAS question. From April, 2006 till December, 2011, we collected 16805 records, and the number of first visit patients was 2830.

IV. PREDICT OF “SHO” AND RESULTS

We predict “Sho” by using 2830 first visit patients’ “Monshin” data. In this paper, we focus on Deficiency and Excess category as a target and adopt random forests. Random forests was proposed by Breiman and is an algorithm for classification that uses an ensemble of classification trees. Our case is supervised learning. We set training and test data that has labels consistent with that type of classification.

TABLE I. “Sho” of Deficiency and Excess category: 2830 first visit patients

Deficiency and Excess	patients
Deficiency pattern	437
Slightly deficiency pattern	395
Between deficiency and excess	1500
Slightly excess pattern	268
Excess pattern	230
Total	2830

Our target data is Table I. We selected randomly 200 patients as a training data (each 100 patients from deficiency pattern and excess pattern). And others are test data.

TABLE II. Result of test data

		Deficiency pattern	Excess pattern	Discriminant ratio
Predict	Deficiency pattern	231	48	67.0%
	Excess pattern	106	82	
	Total	337	130	

The discriminant ratio of training data was perfect but of test data is 67.0%. It was far from practical use (Table II).

V. PREDICT OF “SHO” WITH BODY MASS INDEX AND ITS RESULT

To cover the shortcomings of our questionnaire, we added Body Mass Index (BMI) data to “Monshin” data. BMI is a simple index of weight-for-height that is commonly used to classify underweight, overweight and obesity in adults. It is defined as the weight in kilograms divided by the square of the height in meters.

Center of Kampo medicine, Keio university school of medicine has patients' BMI data on 2011 year (Table VIII). We selected randomly 40 patients as a training data (each 20 patients from deficiency pattern and excess pattern). And others are test data.

TABLE III. “Sho” of Deficiency and Excess category: year 2011

Deficiency and Excess	patients
Deficiency pattern	75

Slightly deficiency pattern	28
Between deficiency and excess	223
Slightly excess pattern	39
Excess pattern	37
Total	402

The discriminant ratio of training data was perfect and of test data is 91.2% (Table IV).

TABLE IV. Result of test data

		Deficiency pattern	Excess pattern	Discriminant ratio
Predict	Deficiency pattern	51	2	91.2%
	Excess pattern	4	15	
	Total	55	17	

VI. CONCLUDING REMARKS

In Japanese traditional medicine, Kampo, “Monshin” plays an important role. The diagnosis by Kampo is called “Sho” and determined by completely different view from Western medicine. And this is reason why non-Kampo specialists without technical knowledge are feel difficulties to use traditional medicine. Since 2006, Center of Kampo medicine, Keio university school of medicine has been collected 2830 first visit patients’ data. We predict “Sho” by using Random forests which is powerful algorithm for classification. First, we use all the 2830 first visit patients’ data. The discriminant ratio of training data was perfect but of test data is 67.0%. Originally, deficiency and excess category means that patient is strongly built or poor built and our “Monshin” didn’t include such indicator. If we use only BMI for classification, it is not working well. So we use both “Monshin” and BMI, and the discriminant ratio of test data is 91.2%. To get good classification, we have to know what is feature of the target and check the data. In this research, prediction of deficiency and excess category is enough for practical use, but other categories are remained and are our future targets.

ACKNOWLEDGMENT

This study was partially supported by Health and Labour Sciences Research Grants for Clinical Research and Research on Statistics and Information from the Ministry of Health, Labour, and Welfare of Japan.

REFERENCES

- [1] Stafford L. HerbalEgram: Volume 8, Number 1, January 2011 WHO Developing New Traditional Medicine Classification <http://cms.herbalgram.org/heg/volume8/01January/WHOCClassifiesT M.html?t=1294841964> (accessed April 10, 2012).
- [2] Watanabe K, Matsuura K, Gao P, et al. Traditional Japanese Kampo medicine: clinical research between modernity and traditional medicine-the state of research and methodological suggestions for the future, Evid Based Complement Alternat Med 2010; published on line June 16. DOI:10.1093/ecam/neg067.
- [3] L. Breiman. Bagging predictors. Machine Learning, 24 (2):123–140, 1996.

An Issue-oriented Syllabus Retrieval System based on Terminology-based Syllabus Structuring and Visualization

*Hideki Mima*¹

(1) School of Engineering, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
mima@t-adm.t.u-tokyo.ac.jp

ABSTRACT

The purpose of this research was to develop an issue-oriented syllabus retrieval system that combined terminological processing, information retrieval, similarity calculation-based document clustering, and visualization.

Recently, scientific knowledge has grown explosively because of rapid advancements that have occurred in academia and society. Because of this dramatic expansion of knowledge, learners and educators sometimes struggle to comprehend the overall aspects of syllabi. In addition, learners may find it difficult to discover appropriate courses of study from syllabi because of the increasing growth of interdisciplinary studies programs. We believe that an issue-oriented syllabus structure might be more efficient because it provides clear directions for users. In this paper, we introduce an issue-oriented automatic syllabus retrieval system that integrates automatic term recognition as an issue extraction, and similarity calculation as terminology-based document clustering. We use automatically-recognized terms to represent each lecture in clustering and visualization. Retrieved syllabi are automatically classified based on their included terms or issues. The main goal of syllabus retrieval and classification is the development of an issue-oriented syllabus retrieval website that will present users with distilled knowledge in a concise form. In comparison with conventional systems, simple keyword-based syllabus retrieval is based on the assumption that our methods can provide users, and, in particular, novice users (students), with efficient lecture retrieval from an enormous number of syllabi. The system is currently in practical use for issue-oriented syllabus retrieval and clustering for syllabi for the University of Tokyo's Open Course Ware and for the School/Department of Engineering. Usability evaluations based on questionnaires used to survey over 100 students revealed that our proposed system is sufficiently efficient at syllabus retrieval.

KEYWORDS: Issue oriented, syllabus retrieval, term extraction, knowledge structuring, visualization

1 Introduction

Recently, scientific knowledge has grown explosively because of rapid advancements that have occurred in academia and society.¹ This rapid expansion of knowledge has made it increasingly difficult for learners and educators to comprehend the overall aspects of syllabi. In addition, because of the rapid growth of interdisciplinary studies programs, such as energy studies and earth-environmental studies, learners have found it increasingly difficult to discover appropriate courses of study in their syllabi.

Syllabus retrieval is believed to be one of several solutions to these problems. In fact, several syllabus retrieval systems have been proposed. In general, current syllabus retrieval methods can be classified as query-oriented and/or issue-oriented. Although the query-oriented method is useful and possesses strong retrieval capabilities, it can be difficult to employ, especially for novices, because the generation of queries usually requires users to first clarify their subjects.

The issue-oriented syllabus retrieval method was developed in an attempt to provide clear directions to learners. The issue-oriented syllabus structure is believed to be more efficient for learning and education, because it requires less knowledge about subjects (Mima et al., 2006). However, this system generally requires that users classify all syllabi manually in advance. This can be a time-consuming task. Thus, we can see that it is important to develop a more efficient method for automatic syllabus structuring to accelerate syllabus classification. The advantage of this technique is based on the assumption that automatic methods will enable more efficient processing of enormous amounts of syllabi texts.

In this paper, we introduce an innovative issue-oriented automatic syllabus classification system. We integrate automatic term recognition as issue extraction, terminology-based similarity-calculation for clustering, information retrieval, and visualization. Automatically-recognized terms are used to represent each lecture (or class) in clustering. In the system, provided syllabi are automatically classified and labeled according to the included terms that were automatically extracted. The main goal of syllabus retrieval and clustering is to develop an issue-oriented syllabus retrieval website that will present distilled knowledge to users in a concise form. The advantage of this system, in comparison with conventional syllabus retrieval or classification, is based on the assumption that automatic methods can efficiently process enormous amounts of text. The system has already been put into practical use for syllabus retrieval and clustering for the University of Tokyo's Open Course Ware and for the School/Department of Engineering syllabi. Usability evaluations based on questionnaires used to survey over 100 students revealed that our proposed system is sufficiently efficient at syllabus retrieval.

In the following section of this paper, we briefly explain the process of issue-oriented syllabi retrieval. We also provide an overview of the clustering system. In Section 2, we describe our proposed syllabus retrieval and classification scheme that is based on the use of automatically-extracted terms and on a visualization technique. In Sections 3 and 4, we discuss terminological processing as a feature extraction from each syllabus for similarity calculation and

¹ For example, the Medline database (<http://www.ncbi.nlm.nih.gov/pubmed>) currently contains over 16 million paper abstracts in the domains of molecular biology, biomedicine, and medicine. The database is growing at a rate of more than 40,000 abstracts each month.

visualization. In Section 5, we present our evaluations of data collected from questionnaires used to survey over 100 students. We relied on the collected data to analyze the usability of our proposed scheme and to confirm its feasibility and efficiency. In the final Section, we present a summary of our approach and our conclusions.

2 System Overview

The main purpose of this study was to develop an efficient issue-oriented syllabus retrieval system that would provide clear directions to learners. Our approach to this issue-oriented syllabus classification system is based on the following:

- automatic term recognition (ATR) for automatic issue extraction
- automatic term clustering (ATC) for term variation management
- terminology-based document similarity calculation to develop syllabus classification
- automatic class label inference to clarify general issues of the classes

The system architecture is modular. It integrates the following components (see, Figure 1):

- *Terminology-based issue extraction (TIE)* – A component that conducts automatic term recognition as issue extraction from syllabus texts. It includes term extraction and term variation management.
- *Syllabus retriever (SR)* – It retrieves syllabi based on selected issues that are automatically extracted by TIE. It calculates similarities between each issue and each retrieved syllabus. Currently, we have adopted $tf*idf$ based similarity calculation.
- *Similarity Calculation Engine(s) (SCE)* – It calculates similarities between KSs provided from each KR component by the use of ontology developed by ODE to show semantic similarities between each KSs. We adopted Vector Space Model-based (VSM) similarity calculation and we used terms as features of VSM. Semantic clusters of KSs were also provided.
- *SVM-based learning (SBL)* – A component that learns how to classify syllabi by extraction of classification patterns from features that have also been extracted by TFE. It then produces classification knowledge.
- *Terminology-based syllabus classification (SBC)* – It calculates similarities between syllabi provided by the SR component by the use of terms provided from TIE to develop clusters of syllabi. We adopted Vector Space Model-based (VSM) similarity calculation.
- *Term-based label inference (TLI)* – It infers representing labels for each class developed by TSC. We currently inferred labels based on term frequency (tf) for importance and document frequency (df) for generality.
- *Syllabus class visualizer (SCV)* – It visualizes syllabi structures based on graph expression in which classes of syllabi and representing labels of classes inferred by (TLI) are automatically provided.

As shown in Figure 1 and the flows by numbers, the system extracts issues automatically from syllabi texts in advance and produces classification of lectures based on these terms or issues. Then, representing labels (i.e., class labels) are also inferred by the use of terminological information. Finally, SVC visualizes syllabi structures with respect to selected issues.

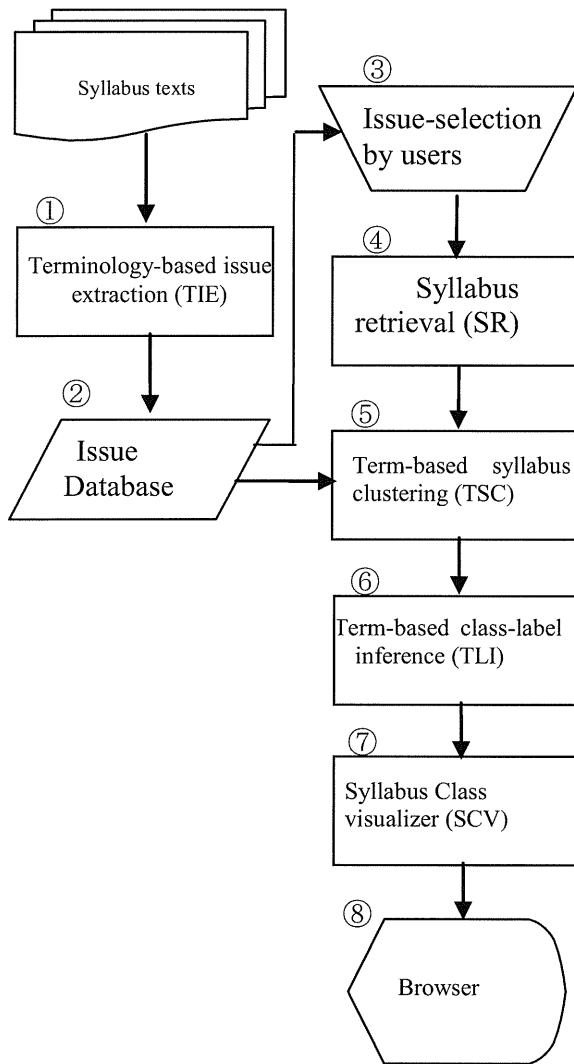


FIGURE 1 – The system diagram

3 Terminological processing as an ontology development

The lack of clear naming standards within a domain (e.g., biomedicine) makes ATR a non-trivial problem (Fukuda et al., 1998). Also, this lack of standards may typically cause many-to-many relationships between terms and concepts. In practice, two problems stem from this issue: (1) some terms may have multiple meanings (i.e., *term ambiguity*), and, conversely, (2) some terms may refer to the same concept (i.e., *term variation*). Generally, term ambiguity exerts negative effects on IE precision; term variation decreases IE recall. These problems reveal the difficulty involved in the use of simple keyword-based IE techniques. Therefore, the development of more sophisticated techniques, such as the identification of groups of different terms that refer to the same (or similar) concept(s) that could benefit from reliance on efficient and consistent ATR/ATC and term variation management methods, is needed. These methods are also important tools that can be used to organize domain-specific knowledge because terms should not be treated

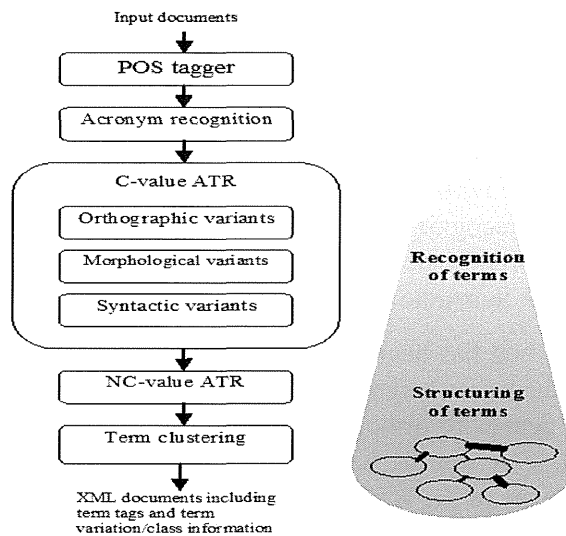


FIGURE 2 – Term recognition as issue extraction

in isolation from other terms. Rather, they should be related to one another so that relationships that exist between corresponding concepts are, at least partially, reflected in the terminology.

3.1 Term recognition

For our system, we used an ATR method based on *C/NC-value* methods (Mima et al., 2001; Mima and Ananiadou, 2001). The *C-value* method recognizes terms by combining linguistic knowledge and statistical analysis. The method extracts multi-word terms,² and it is not limited to a specific class of concepts. It is implemented as a two-step procedure. In the first step, term candidates are extracted by the use of a set of linguistic filters that describe general term formation patterns. In the second step, the term candidates are assigned termhood scores (referred to as *C-values*) based on a statistical measure. The measure amalgamates four numerical corpus-based characteristics of a candidate term: (1) frequency of occurrence, (2) frequency of occurrence as a substring of other candidate terms, (3) the number of candidate terms that contain the given candidate term as a substring, and (4) the number of words contained in the candidate term.

The *NC-value method* further improves the *C-value* results by considering the context of the candidate terms. The relevant context words are extracted and assigned weights based on the frequency with which they appear with top-ranked term candidates extracted by the *C-value* method. Subsequently, context factors are assigned to candidate terms according to their co-occurrence with top-ranked context words. Finally, new termhood estimations, referred to as *NC-values*, are calculated as a linear combination of the *C-values* and context factors for the respective terms. Evaluation of the *C/NC-methods* (Mima and Ananiadou, 2001) has revealed that contextual information improves term distribution in the extracted list because it places real terms closer to the top of the list.

² More than 85% of domain-specific terms are multi-word terms (Mima and Ananiadou, 2001).

3.2 Term variation management

Term variation and ambiguity have caused and continue to cause problems for ATR, as well as for human experts. Several methods for term variation management have been developed. For example, the BLAST system (Krauthammer et al., 2000) used approximate text string matching techniques and dictionaries to recognize spelling variations in gene and protein names. FASTR (Jacquemin, 2001) handles morphological and syntactic variations by means of meta-rules used to describe term normalization. Semantic variants are handled via WordNet.

The basic *C-value* method has been enhanced by term variation management (Mima and Ananiadou, 2001). We consider a variety of sources from which term variation problems originate. In particular, we deal with orthographical, morphological, syntactic, lexico-semantic, and pragmatic phenomena. Our approach to term variation management is based on term normalization as an integral part of the ATR process. Term variants (i.e., synonymous terms) are addressed in the initial phase of ATR when term candidates are singled out. This differs from the process that is used in other approaches (e.g., FASTR handles variants subsequently by application of transformation rules to extracted terms). Each term variant is normalized (see, Table 1, as an example) and term variants that have the same normalized form are then grouped into classes to link each term candidate to all of its variants. In this way, a list of normalized term candidate classes, rather than a list of single terms, is statistically processed. The termhood is then calculated for a whole class of term variants, rather than for each term variant separately.

Term variants	Normalized term
human cancers	} → human cancer
cancer in humans	
human's cancer	
human carcinoma	

TABLE 1 – Automatic term normalization

3.3 Term clustering

In addition to term recognition, term clustering is an indispensable component of the literature mining process. Because terminological opacity and polysemy are very common in molecular biology and biomedicine, term clustering is essential for the semantic integration of terms, the construction of domain ontologies, and for semantic tagging.

In our system, ATC is performed by the use of a hierarchical clustering method in which clusters are merged based on average mutual information that measures the strength of the relationships between terms (Ushioda, 1996). The system uses terms automatically recognized by the *NC-value* method and their co-occurrences as input. A dendrogram of terms is produced as output. Parallel symmetric processing is used for high-speed clustering. The calculated term cluster information is encoded and used for calculation of semantic similarities in the SCE component. More precisely, the similarity between two individual terms is determined based on their position in a dendrogram. In addition, a commonality measure is defined as the number of shared ancestors between two terms in the dendrogram. A positional measure is defined as the

sum of their distances from the root. Similarity between two terms corresponds to a ratio between commonality and positional measure.

Table 3 shows a sample of automatically-recognized terms (issues) that occur in an Engineering domain syllabus text that consists of 850 lectures (Faculty of Engineering, University of Tokyo, 2006). As we can see from the Table, reasonable and representative issues were successfully extracted by our method.

Automatically-Recognized Terms	Termhood
基礎知識 (basic knowledge)	144.55
線形代数 (linear algebra)	77.35
統計力学 (statistical mechanics)	74.00
固体物理 (solid-state physics)	67.20
ベクトル解析 (vector calculus)	65.01
偏微分方程式 (partial differential equation)	62.40
材料力学 (mechanics of materials)	62.13
環境問題 (environmental issues)	60.17

TABLE 2 – Sample of recognized issues

Further details of the methods and their evaluations can be found in Mima et al. (2001) and Mima and Ananiadou (2001).

4 The Use of Visualization to Generate Issue-oriented Syllabus Structures

In our system, the TSC, TLI, and SCV are implemented by the integration of terminology-based issue extraction from syllabi and by clustering of syllabi based on semantic similarities that are also calculated based on terms in syllabi. Graph-based visualization for the automatic generation of issue-oriented syllabus structures is also provided to help in retrieval of lectures. Figure 3 shows an example of the visualization of issue-oriented syllabus structures relevant to the issue, “environment and energy,” that occurs in the engineering syllabus. To structure knowledge, the system constructs a graph in which the nodes are used to indicate relevant syllabi for the key issues selected by the user. Links among the syllabi indicate semantic similarities that are calculated by the use of terminological information developed by our TIE components. Semantic similarity is based on comparisons of terminological information extracted from each syllabus, whereas conventional similarity calculation is generally based on extracted nouns. In addition, the locations of each node are calculated and optimized when the graph is drawn. The distance between nodes depends on the closeness of their meanings. The complete algorithm of this issue-structuring method is presented below:

begin

$Q \leftarrow$ issues specified to IR

$R \leftarrow \text{IR}(Q)$ // retrieving relevant syllabi to Q and putting them into R

for every x in R do

$w(Q, x) \leftarrow \text{IRscore}(Q, x)$ // calculate IR score between Q and x

for every y in R do

if $x \neq y$ then

$p \leftarrow \text{Ont}(x)$ // retrieving terminological information of x

$q \leftarrow \text{Ont}(y)$ // " " y

$w(x, y) \leftarrow \text{Sim}(p, q)$ // calculate similarity using p and q

fi

end

end

Visualize graph based on every $\{w(i, j) | i=Q \text{ or } i \in R, j \in R, i \neq j\}$

end.

We generate an issue-oriented syllabus structure based on (1) cluster recognition and (2) terminology-based cluster label inference. Cluster recognition is performed by detection of groups of nodes in which every combination of included nodes is strongly linked (i.e., their similarity exceeds a threshold). Automatic cluster label inference is performed by the use of terminological information included in each cluster with respect to *tf* (term frequency) and *df* (document frequency (i.e., term generality)).

5 Evaluation

We performed a practical application of the system for syllabus retrieval for the University of Tokyo's Online Course Catalogue (UTOCC),³ for the Open Course Ware (UT-OCW)⁴ site, and for the syllabus-structuring (SS) site⁵ for the School/Department of Engineering. All of these syllabi are available to the public over the Internet. The UT-OCW's course search system is designed to search the syllabi of courses posted on the UT-OCW site and on the Massachusetts Institute of Technology's OCW site (MIT-OCW). In addition, OCC and SS site's search is designed to search the syllabi of more than 9,000 lectures from all schools/departments at the University of Tokyo, and 1,600 lectures from the School/Department of Engineering at the University of Tokyo. Both systems display search results based on relationships that exist among the syllabi as a structural graphic (see, Figure 3). Based on terms that were automatically-extracted terms (issues) from the syllabi and on similarities calculated by the use of those terms, the system displays the search results in a network format that uses dots and lines. In other words,

³ <http://catalog.he.u-tokyo.ac.jp/>

⁴ <http://ocw.u-tokyo.ac.jp/>.

⁵ <http://ciee.t.u-tokyo.ac.jp/>.

Positive statements	#
Advantage of visualization	45
Improvement in retrieval efficiency	41
Clarity of results	22
User-friendly interfaces	20
Misc.	23
Total	151

TABLE 3 – Breakdown of positive statements

Statements that recommended further improvement	#
Complexity of visualization	67
Additional linkage to other syllabi	23
Lack of clarity about relationships that exist among lectures	11
Linkage to other systems (e.g., lecture management, etc.)	13
Quality of issue extraction	10
Difficulty of operation	5
Speed of calculation	1
Misc.	38
Total	168

TABLE 4 – Statements that recommend further improvement

Conclusion

We developed an issue-oriented syllabus retrieval system that combined terminological processing, information retrieval, similarity calculation-based document clustering, and

visualization. The system provides visualizations of issue-oriented syllabus structuring during retrieval. This differs from conventional syllabus retrieval that solely provides a list of retrieved results relevant to a specific query.

We evaluated the system based on data collected from questionnaires used to survey over 100 students. Based on our results, we can reasonably state that the system provides relatively efficient syllabus retrieval.

References

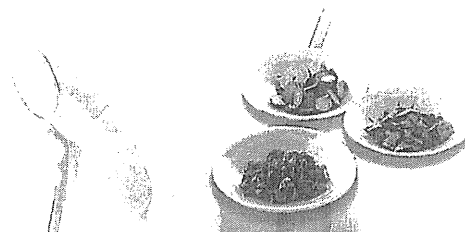
- Fukuda, K., Tsunoda, T., Tamura, A. and Takagi, T. (1998). *Toward information extraction: Identifying protein names from biological papers*, Proc. of PSB-98, Hawaii, pp. 3:705–716.
- Mima, H., Ananiadou, S. and Matsushima, K. (2006). *Terminology-based Knowledge Mining for New Knowledge Discovery*, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 5(1), pp. 74–88.
- Mima, H., Ananiadou, S. and Nenadic, G. (2001). *TRACT workbench: An automatic term recognition and clustering of terms*. In V. Matoušek, P. Mautner, R. Mouček, K. Taušer (eds.) *Text, Speech and Dialogue*, LNAI 2166, Springer Verlag, pp. 126–133.
- Mima, H. and Ananiadou, S. (2001). *An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-word units in Japanese*, *International Journal of Terminology*, Vol. 6(2), pp. 175–194.
- Krauthammer, M., Rzhetsky, A., Morozov, P. and Friedman, C. (2000). *Using BLAST for identifying gene and protein names in journal articles*. *Gene* 259, pp. 245–252.
- Jacquemin, C. (2001). *Spotting and discovering terms through NLP*. MIT Press, Cambridge MA, p. 378.
- Ushioda, A. (1996). *Hierarchical clustering of words*. In Proc. of COLING '96, Copenhagen, Denmark, pp. 1159–1162.

頭痛の漢方治療：最新のエビデンス

Novel evidence for Japanese traditional medicines for headache

上野眞二 村松慎一 Shinji UYENO¹ and Shin-ichi MURAMATSU²

鷺谷病院脳神経外科¹, 自治医科大学東洋医学部門², 同神経内科学部門²



◎頭痛の治療薬として多くの漢方薬が経験的・伝統的に有用とされている。片頭痛に対しては、呉茱萸湯、五苓散、桂枝人参湯などが、緊張型頭痛に対しては川芎茶調散、葛根湯、釣藤散などが頻用され、薬物乱用性頭痛をはじめ二次性頭痛にも使用されている。多数の成分を含む漢方薬は作用機序の詳細が未解明で、これまでエビデンスレベルの高い臨床報告は少ない。しかし最近、基礎研究とともに従来の経験則を考慮した臨床治験も進められている。

Keywords 漢方, 頭痛, 五苓散, 呉茱萸湯, 川芎茶調散

頭痛は日常診療のなかでよく遭遇する疼痛疾患群である。日本における15歳以上の人口の頭痛有病率は39.6%で、緊張型頭痛の有病率がもっとも高く22.4%、ついで片頭痛が8.4%とされる¹⁾。現在、頭痛診療に際しては、国際頭痛分類第2版(ICHD-II)に従った診断と治療が推奨されている²⁾。一次性頭痛(機能性頭痛)は片頭痛、緊張型頭痛など症候により診断され、さしあたり生命に影響のない頭痛である。

片頭痛は10歳代後半から40歳代の女性に多く、緊張型頭痛の約1/5の頻度である。発作性に出現して痛みは強く、体動や力みにより頭痛が増悪するためquality of life(QOL)は著しく低下する。悪心・嘔吐、光・音・臭過敏などの症状を伴う拍動性の頭痛が4~72時間持続する。典型例では、頭痛発作の20~30分前に閃輝暗点などの前兆を認める(表1)。むくみ、嘔気、嘔吐、利尿などは漢方医学の水毒(体内における水分代謝調節異常)の症候である(図1)。片頭痛には肩こり(75%)やストレス(72%)を伴うことも多く、頭痛は両側性(40%)や非拍動性(50%)の場合も少なくないので、緊張型頭痛と誤診しないように注意する。

緊張型頭痛はもっとも頻度が多い。痛みの程度は中等度で日常生活に著しい支障をきたすことは

表1 前兆のない片頭痛の診断基準

頭痛発作>5回
持続時間4~72時間
以下の特徴のうち2項目以上
1. 片側性
2. 拍動性
3. 中等度~重度の頭痛
4. 日常的な動作で増悪
発作中に以下の1項目を満たす
1. 悪心または嘔吐(あるいはその両方)
2. 光過敏および音過敏

ないが、ほぼ毎日起こる。頭部の圧迫感・絞扼感、肩こりとして自覚されることが多い。片頭痛のような発作性や前兆は認めないが、同一患者に緊張型頭痛と片頭痛が併存することはまれではない。

二次性頭痛(症候性頭痛)は器質的疾患を含む他の原因による頭痛である。頭痛診療に際しては二次性頭痛のうち西洋医学的治療が優先される、くも膜下出血、髄膜炎、脳腫瘍などを鑑別することがもっとも大切である。初発した頭痛、いつもと違う頭痛、最近増悪している頭痛、発熱・筋力低下・意識の変容などを伴う頭痛などは二次性頭痛を疑う。

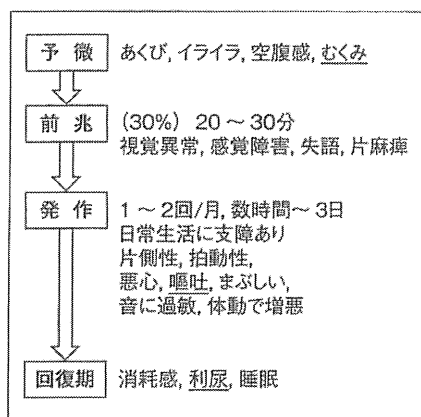


図1 片頭痛の経過

下線を引いた症候を、漢方医学では水毒(水分代謝調節異常)としてとらえる。



頭痛の漢方治療

漢方薬の適応としては、①西洋薬で十分な鎮痛効果が得られない、②西洋薬で副作用がある、③西洋薬の副作用に不安があり、西洋薬とは異なる治療を求める、④西洋薬による薬物乱用頭痛の予防・離脱に使用する、⑤頭痛の背景に心因性の要素を含む、などが考えられる。実際の日常臨床では片頭痛、緊張型頭痛などの一次性頭痛に頻用されているが、器質性疾患に伴う頭痛にも西洋医学的な治療とともに使用される。



頭痛の西洋薬治療

片頭痛の急性期治療は、セロトニン受容体(5HT_{1B/1D})作動薬であるトリプタン製剤の登場により急速に進歩した。トリプタン製剤は頭蓋内の血管平滑筋の5HT_{1B}受容体と血管周囲に分布する5HT_{1D}受容体に結合して神経ペプチドの放出を抑制し、血管を収縮することにより頭痛を頓挫させる。日本では現在、5種類のトリプタン製剤が経口錠、口腔内崩壊・速溶錠、点鼻薬、注射薬の剤型で保険適応となっている。各製剤の特徴を考慮して、頭痛発作の進行が早い症例には最高血中濃度に到達する時間(T_{max})の早いものを、発作

持続時間の長い症例には血中濃度半減期の(T_{1/2})の長いものを使用すれば、片頭痛の急性期治療においてある程度は個別の症例に適した薬の選択が可能となる。しかし、トリプタン製剤は、①血管収縮作用を有するため、虚血性心疾患の既往を有する症例には使用禁忌であり、あらたに虚血性心疾患を惹起する危険がある、②発作時の頭痛を軽減するが予防効果はない、③薬物乱用頭痛の多くは市販薬によって生じるが、トリプタン製剤は他の頭痛薬に比べて薬物乱用頭痛となる期間が短い、④無効例が存在する、⑤高価であり経済的な負担が大きい、など問題点は少なくない。

片頭痛の予防薬として保険適応があるものは、日本で開発されたカルシウム拮抗薬の塩酸ロメリジンのほか、抗セロトニン薬のジメトチアジン塩酸、抑制系神経伝達物質であるγ-aminobutyric acid(GABA)類似作用を有するバルプロ酸と数少ない。外国ではA型ボツリヌス毒素の顔面・頸部への局所投与が片頭痛予防に有用と報告されている³⁾。筋緊張緩和作用のみならず、末梢痛覚神経の神経伝達物質を抑制することにより効果を発揮すると考えられるが、日本では保険適応はない。

緊張型頭痛の治療では、非ステロイド性抗炎症薬(NSAIDs)は胃腸障害、造血器障害などの副作用があり、慢性的な使用により薬剤乱用性頭痛が惹起する可能性がある。また、筋弛緩薬、抗不安薬には眠気、ふらつきなどのQOLを低下させる問題点がある。



頭痛に対する漢方薬の頻用処方

片頭痛に対しては、呉茱萸湯、五苓散、桂枝人参湯などが、緊張型頭痛に対しては川芎茶調散、葛根湯、釣藤散などが頻用され、薬物乱用性頭痛をはじめ二次性頭痛にも使用されている(表2)⁴⁾。



基礎研究のエビデンス：五苓散の利尿作用とアクアポリン

五苓散は後漢の『傷寒論』と『金匱要略』を原

表 2 頭痛に対する漢方薬の頻用処方

処 方	構 成 生 薬
五苓散	沢瀉, 猪苓, 蒼朮(または白朮), 茯苓, 桂皮
呉茱萸湯	呉茱萸, 人參, 大棗, 生姜
川芎茶調散	白芷, 甘草, 羌活, 荆芥, 川芎, 防風, 薄荷, 香附子, 細茶
桂枝人參湯	桂皮, 甘草, 人參, 乾姜, 白朮(または蒼朮)
釣藤散	釣藤鈎, 橘皮(または陳皮), 半夏, 麥門冬, 茯苓, 人參, 防風, 菊花, 甘草, 生姜, 石膏
半夏白朮天麻湯	半夏, 白朮, 陳皮, 茯苓, 天麻, 生姜, 黃耆, 人參, 沢瀉, 黃柏, 乾姜, 麦芽, (神麩), (蒼朮)
苓桂朮甘湯	茯苓, 白朮(または蒼朮), 桂皮, 甘草
当帰四逆加呉茱萸湯	当帰, 桂皮, 芍薬, 木通, 細辛, 甘草, 大棗, 呉茱萸, 生姜
葛根湯	葛根, 麻黄, 大棗, 桂枝, 芍薬, 甘草, 生姜
加味逍遙散	当帰, 芍薬, 蒼朮(または白朮), 茯苓, 柴胡, 牡丹皮, 山梔子, 甘草, 生姜, 薄荷
桃核承気湯	桃仁, 桂皮, 大黄, 芒硝, 甘草
三物黄芩湯	黄芩, 苦参, 地黄
五積散	茯苓, 蒼朮(または白朮), 陳皮, 半夏, 当帰, 芍薬, 川芎, 厚朴, 白芷, 枳殼(実), 桔梗, 乾姜, 桂皮, 麻黄, 大棗, 甘草, (生姜), (香附子)

典とし、沢瀉、猪苓、蒼朮、茯苓、桂皮の5種類の生薬により構成される。元来、急性熱疾患による発汗の後、口渇、嘔吐、下痢、尿量減少のある場合に使用された、消化管内の過剰な水の吸収を促進することで血管内脱水を補正し、さらには排尿により過剰な水分を排出するとされる代表的な利尿剤である。

五苓散は水負荷状態では尿量を増加させ、脱水状態では尿量を減少させる水分代謝調整作用をもつ⁵⁾が、その薬理作用の詳細は明らかでなかった。一般的な利尿薬の尿量増加の機序としては、腎血圧上昇による腎糸球体濾過量の増加と電解質の再吸収阻害による原尿濃縮作用の2種類が知られているが、五苓散にはこのいずれの作用もみられない。最近、水チャネルのアクアポリン(AQP;「サイドメモ」参照)に対する五苓散の作用が報告されている。生体での水の移動は浸透圧や静水圧に依存し、AQPは細胞膜の水透過性、すなわち水の移動効率の調節を行っている。病的な浸透圧の異常が生じるとそれに伴い過剰に水の移動が起こり、

サイドメモ

アクアポリン(AQP)

1992年、細胞膜の水透過性を調節するアクアポリン(AQP)という水チャネルが発見された。AQPは28 kDaの膜蛋白で細胞膜を貫通する形で存在し、内部に水分子が通過する漏斗状の孔がある。現在までに、ヒトでは13種のアイソフォームが各臓器で確認され、脳では血液脳関門を形成している毛細血管周囲のアストロサイトの足突起などにAQP4が分布している。AQP4欠損マウスでは、急性水中毒状態でも野生型に比べアストロサイト足突起の膨化が著明に軽減され、脳浮腫の発生にAQP4が関与することが示されている。また、視神経脊髄炎の患者血液中にAQP4に対する自己抗体が高頻度に認められ、病態形成に関与していることが推察されてきている。細菌、植物、動物と普遍的に存在し、生命活動に広くかわる重要な分子AQPを発見したPeter Agreには2003年のノーベル化学賞が贈られた。

その結果として浮腫(水毒)の病態が生じる。マウス肺上皮細胞株とプロテオリポソームを使用した *in vitro* 実験系で五苓散がAQPの働きを阻害して細胞膜の水透過性を抑制することが報告され、この抑制作用は構成生薬の蒼朮に含まれるマンガンを担っている可能性が推察されている⁶⁾。五苓散が浮腫を抑制(水毒を改善)する利水作用の、すくなくとも一部はAQPの阻害作用による可能性がある。

脳浮腫の西洋薬による治療では浸透圧利尿薬が使用されるが、脱水、電解質異常、溶血、心不全などの副作用がある。五苓散にはこれらの副作用がないため使用しやすい。現在、慢性硬膜下血腫の保存的治療・再発予防、脳腫瘍や急性期脳梗塞の脳浮腫軽減に臨床応用されている。

臨床研究のエビデンス

1. 呉茱萸湯・五苓散は慢性頭痛に有用である

呉茱萸湯は後漢の『傷寒論』と『金匱要略』を原典とする。呉茱萸、人參、大棗、生姜の4種の生薬から構成され、血流増加、体温上昇、鎮痛の薬理作用がある。疲労しやすく、手足の冷え、水毒(体内における水分の偏在)による嘔気・嘔吐を伴う頭痛に使用される。これまでに多くの症例報告と症例集積研究があり、頭痛に対して高い有用性が報告されてきた。

漢方医学では生体防御反応を、寒性・非活動性・沈降性の場合を陰証、熱性・活動性・発揚性の場合を陽証と大きく二分する概念などがある。これらの概念と個々の症例の体質、症状、兆候などを総合して得られる証(陰陽、虚実、寒熱などの漢方医学的診断)に基づいて処方決定される。この証を考慮した頭痛の臨床研究が報告されている。

慢性頭痛23例(片頭痛13例、緊張型頭痛2例、混合型頭痛8例)を、個々の症例の陰証と陽証とを考慮して2群に分け処方を決定した。陰証18例には呉茱萸湯エキス製剤7.5 g/dayを、陽証5例には五苓散エキス製剤7.5 g/dayをそれぞれ4週間

投与の後に、自覚的な頭痛の改善の程度を、著明改善、改善、やや改善、不変、悪化の5段階に分け検討した。その結果、やや改善以上を有効とする有効率は呉茱萸湯群で72.2%、五苓散群では80.0%、全体では73.9%であった⁷⁾。

さらに、上記の結果を踏まえて最近、慢性頭痛(片頭痛・緊張型頭痛)を対象に、第1段階として呉茱萸湯の治療が有効だった症例(レスポnder)を抽出し、続いて対照としてプラセボ群を設定した無作為化二重盲検比較試験が行われた。第1段階では慢性頭痛91例に呉茱萸湯エキス製剤を7.5 g/dayを投与し、レスポnderを選別した。4週間休薬の後、このレスポnder 53例を呉茱萸湯群28例とプラセボ群25例に分け、それぞれの被検薬を12週間投与の後に、頭痛の種類、頻度、随伴症状、その他の頭痛薬の頓服回数などを指標として検討した。その結果、呉茱萸湯群では頭痛発症頻度(日)はプラセボ群と比較して有意に減少し(2.6±3.7 vs. 0.3±1.4, $p=0.034$)、頭痛薬の頓服回数はやや減少(2.2±4.0 vs. 1.4±8.2, $p=0.672$)していた。また、呉茱萸湯群では冷え、月経痛、肩こりなどの随伴症状が50%以上改善していた⁸⁾。

2. 呉茱萸湯は片頭痛予防に有用である

片頭痛予防効果に対する呉茱萸湯と西洋薬・塩酸ロメジリンとを比較した臨床研究において、呉茱萸湯の有効性が報告されている⁹⁾。片頭痛患者14例を、無作為にA群(前半28日間は塩酸ロメジリン10 mg/dayを服用し、2週間休薬後;後半28日間は呉茱萸湯エキス製剤7.5 g/dayを服用する)7例とB群(前半28日間は呉茱萸湯エキス製剤7.5 g/dayを服用し、2週間休薬後、後半28日間は塩酸ロメジリン10 mg/dayを服用する)7例に分けたオープン・クロスオーバー研究である。頭痛発作回数、visual analogue scale(VAS)のピーク値で自己評価した頭痛の程度、トリプタン製剤内服錠数、頭痛発作消失までの時間を指標として開始前、第一相、休薬相、第二相、終了時に評価した。この結果、いずれの指標においても呉茱萸湯のほうが塩酸ロメジリンと比較して有意に治療効果が認められた。

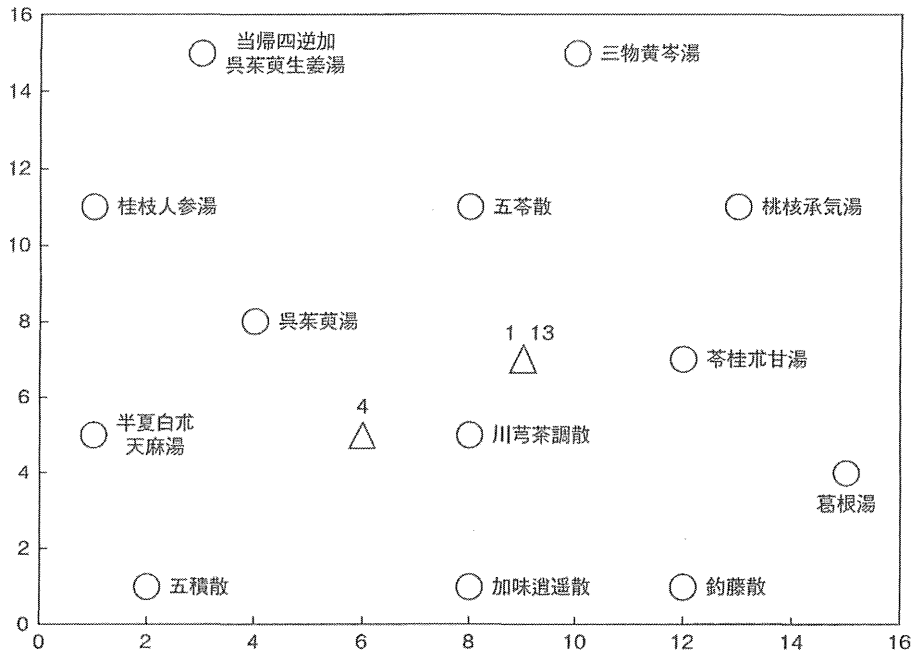


図2 頭痛頻用処方SOM解析¹⁰⁾

13処方の類縁関係を適応症候から、自己組織化マップ(SOM)により解析した。図の右側が実証、左側が虚証対応の処方に相当し、川芎茶調散はほぼ中間に位置する。1、4、13は川芎茶調散の著効症例の位置。

3. 自己組織化マップによる川芎茶調散の適応解析¹⁰⁾

川芎茶調散は太平惠民和劑局方の改訂版(宋代)を原典とし、構成生薬は、白芷、甘草、羌活、荊芥、川芎、防風、薄荷、香附子、細茶、(細辛)である。鼻炎を伴う軽度の感冒から女性の生理に伴う頭痛まで頭痛全般に使用される。川芎茶調散は頭痛に頻用される他の漢方薬に比べて特徴的な適応指標(漢方の証)は知られていない。川芎茶調散の中樞神経における作用としてラット線条体のドパミン濃度増加作用が報告されているが、頭痛に対する鎮痛機序は明らかではない^{11,12)}。

竹田らは階層型ニューラルネットワークを応用し、藤平の特徴判別表¹³⁾に基づき教師あり学習を行った漢方処方診断支援システムを開発してきた。これを使用して、川芎茶調散を処方した17例(著効3例、有効9例、無効5例)を対象として他の頻用12処方との類似性を検討した結果では、呉茱萸湯、釣藤散、葛根湯が適応となる頭痛のなか

に川芎茶調散が有効である可能性が示唆された。さらに、各処方との類似性を散布図的に表現できる自己組織化マップ(self-organizing map: SOM)解析を行った結果を示す(図2)。この図で、漢方医学の虚実分類における実証対応の処方は横軸右、虚証対応は左、中間証対応はその間に位置している。川芎茶調散は中間証群に位置づけられた。古典に“いっさいの頭痛に用いる”と記載された川芎茶調散の幅広い適応を反映した結果となっている¹⁴⁾。

おわりに

本稿では、頭痛に使用される漢方薬の基礎と臨床研究を紹介した。日本頭痛学会の慢性頭痛の診療ガイドラインには、「漢方薬は予防薬あるいは急性期治療薬として長期にわたり使用されており、経験的あるいは伝統的には効果・両面から有用である。これらを裏付ける科学的エビデンスも