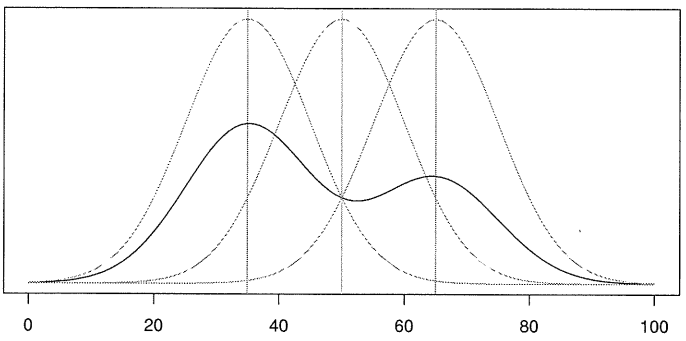
Figure 3 Transform the VAS to PD: first time = 35 second time = 65



4.1 The clustering algorithm

We extend the idea of a hierarchical clustering in the framework of conventional data analysis. Let n be the number of PD and K be the number of cluster (the algorithm is same for PCD).

Step 1 Begin with K clusters, each containing only a single PD, $K = n$. Calculate distance between PD.

Step 2 Search the minimum distance among K clusters. Let combine the pair selected among the clusters. Combine PDs into a new cluster, It is described by mixture distribution of the member, where mixture weight is equal. Let K be $K - 1$. If $K > 1$, go to Step 3, otherwise Step 4.

Step 3 Calculate the distance between new cluster and other cluster, and go back to Step 2.

Step 4 Draw the dendrogram.

Kullback-Leibler divergence is the natural way to define a distance measure between probability distributions (Kullback, 1968), but not symmetry. We would like to use the symmetric Kullback-Leibler (symmetric KL) divergence as a distance between *concepts*. The symmetric KL-divergence between two distributions s_1 and s_2 is

$$\begin{aligned} D(s_1(\boldsymbol{x}), s_2(\boldsymbol{x})) &= D(s_1(\boldsymbol{x})||s_2(\boldsymbol{x})) + D(s_2(\boldsymbol{x})||s_1(\boldsymbol{x})) \\ &= \int_{-\infty}^{\infty} s_1(\boldsymbol{x}) \log \frac{s_1(\boldsymbol{x})}{s_2(\boldsymbol{x})} d\boldsymbol{x} + \int_{-\infty}^{\infty} s_2(\boldsymbol{x}) \log \frac{s_2(\boldsymbol{x})}{s_1(\boldsymbol{x})} d\boldsymbol{x}, \end{aligned} \tag{1}$$

where $D(s_1||s_2)$ is KL divergence from s_1 to s_2 and $D(s_2||s_1)$ is one from s_2 to s_1.

4.2 Distance between normal distribution

In Section 4.1, we use symmetric KL-divergence as distance between PDs.

Let PDs be d dimensional $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ and $N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$. Symmetric KL-divergence in Step 1 is

$$D\left(p\left(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right), p\left(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)\right)$$

$$= tr\left(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j^{-1}\right) + tr\left(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i^{-1}\right) \tag{2}$$

$$+ \; tr\left(\left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T\right) - 2d.$$

Let PDs be $d = 1$,

$$D\left(p\left(x|\mu_i, \sigma_i\right), p\left(x|\mu_j, \sigma_j\right)\right)$$
$$= \frac{1}{2}\left\{\log\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2 + (\mu_i - \mu_j)^2}{\sigma_j^2}\right\} + \frac{1}{2}\left\{\log\frac{\sigma_i^2}{\sigma_j^2} + \frac{\sigma_j^2 + (\mu_j - \mu_i)^2}{\sigma_i^2}\right\} - 1. \tag{3}$$

4.3 Distance between Gaussian mixture distributions

After Step 2 or the case of PCD, we need to calculate symmetric KL-divergence between Gaussian mixture distributions. However, it cannot be analytically computed. We can use, instead, Monte-Carlo simulations to approximate the symmetric KL-divergence. The drawback of the Monte-Carlo techniques is the extensive computational cost and the slow converges properties. Furthermore, due to the stochastic nature of the Monte-Carlo method, the approximations of the distance could vary in different computations.

In this paper, we use unscented transform method proposed by Goldberger et al. (2006).

We show approximation of $D(s_1||s_2)$ in (1). Let cluster c_1 contains d-dimensional distribution $N_d(\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})(m = 1, \ldots M)$. Expression formula of c_1 is $s_1(\boldsymbol{x}) = \sum_{m=1}^{M} \omega_m^{(1)} p(\boldsymbol{x}|\boldsymbol{\theta}_m^{(1)})$, where $\omega_m^{(1)}$ is a mixture weight, $p(\boldsymbol{x}|\boldsymbol{\theta}_m^{(1)})$ is m^{th} probability density function of $N_d(\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})$ and $\boldsymbol{\theta}_m^{(1)} = (\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})$. Simmilary, cluster c_2 contains d-dimensional distribution $N_d(\mu_l^{(2)}, \Sigma_l^{(2)})(l = 1, \ldots L)$. Expression formula of c_2 is $s_2 = \sum_{l=1}^{L} \omega_l^{(2)} p(\boldsymbol{x}|\boldsymbol{\theta}_l^{(2)})$.

Approximation of KL-divergence from s_1 to s_2 by using unscented transform method is

$$D(s_1||s_2) \approx \frac{1}{2d}\sum_{m=1}^{M}\omega_m \sum_{k=1}^{2d} \log\frac{s_1(\boldsymbol{o}_{m,k})}{s_2(\boldsymbol{o}_{m,k})}, \tag{4}$$

where $\boldsymbol{o}_{m,t}$ are sigma points. They are chosen as follows:

$$\boldsymbol{o}_{m,t} = \boldsymbol{\mu}_m^{(1)} + \left(\sqrt{d\boldsymbol{\Sigma}_m^{(1)}}\right)_t, \tag{5}$$
$$\boldsymbol{o}_{m,t+d} = \boldsymbol{\mu}_m^{(1)} - \left(\sqrt{d\boldsymbol{\Sigma}_m^{(1)}}\right)_t,$$

such that $\left(\sqrt{\boldsymbol{\Sigma}_m^{(1)}}\right)_t$ is t^{th} column of the matrix square root of $\Sigma_m^{(1)}$. Then,

$$\boldsymbol{o}_{m,t} = \boldsymbol{\mu}_m^{(1)} + \sqrt{d\lambda_{m,t}^{(1)}}\boldsymbol{u}_{m,t}^{(1)} \tag{6}$$
$$\boldsymbol{o}_{m,t+d} = \boldsymbol{\mu}_m^{(1)} - \sqrt{d\lambda_{m,t}^{(1)}}\boldsymbol{u}_{m,t}^{(1)},$$

where $t = 1, \ldots, d$, $\boldsymbol{\mu}_m^{(1)}$ is mean vector of m^{th} normal distribution in s_1, $\lambda_{m,t}^{(1)}$ is t^{th} eigenvalue of $\boldsymbol{\Sigma}_m^{(1)}$ and $\boldsymbol{u}_{m,t}^{(1)}$ is t^{th} eigenvector. If $p = 1$, the sigma points are simply

$$\mu_m^{(1)} \pm \sigma_m^{(1)}.$$

We can calculate approximation of $D(s_2||s_1)$. Substituting these approximations into (1), we obtain the symmetric KL-divergence. We set the divergence as distance between cluster c_1 and c_2.

5 An application to PD

In this section, we apply our proposal method to real VAS data from Keio University School of Medicine. This is masked data and is not be tied to any information that would identify a patient. To compare the traditional method, we apply centroid method to the same data.

5.1 Medical questionnaire in Keio University School of Medicine

Centre for Kampo Medicine, Keio University School of Medicine, have a questionnaire to patients to help medical decision. The questionnaire includes one set of questions about their subjective symptoms. There are 244 yes-no questions and 118 VAS questions, for example, "How do you feel pain with urination?". Patients answer these questions every time when they come to Keio University. Doctors can understand patients' fluctuate in severity.

5.2 Data description and result

For our analysis, we deal with a question which ask about how patient feel cold: "Do you feel cold in your left leg?". The data contain 435 patients' first and second VAS value. We transform this dataset to PD. Table 1 shows extracts taken from the original data and their translation.

Table 1 Original data and their translation

Patient ID	*First VAS value*	*Second Vas value*	$N(\mu, \sigma^2)$
1	100	78	$N(89, 121)$
2	0	50	$N(25, 625)$
⋮	⋮	⋮	⋮
435	42	5	$N(23.5, 342.25)$

5.3 Result

The result of our simulation show in Figure 4. Vertical axis of this dendrogram means distance between PDs. There seem to be three large Cluster, A, B and C.

The PDs of Cluster A have large variance. The member of Cluster B has small variance. The member of Cluster C has small variance and large mean. The level that patients' expression of sense of pain appears in features of clusters.

The result of centroid method show in Figure 5.

Figure 4 Dendrogram for PDs (see online version for colours)

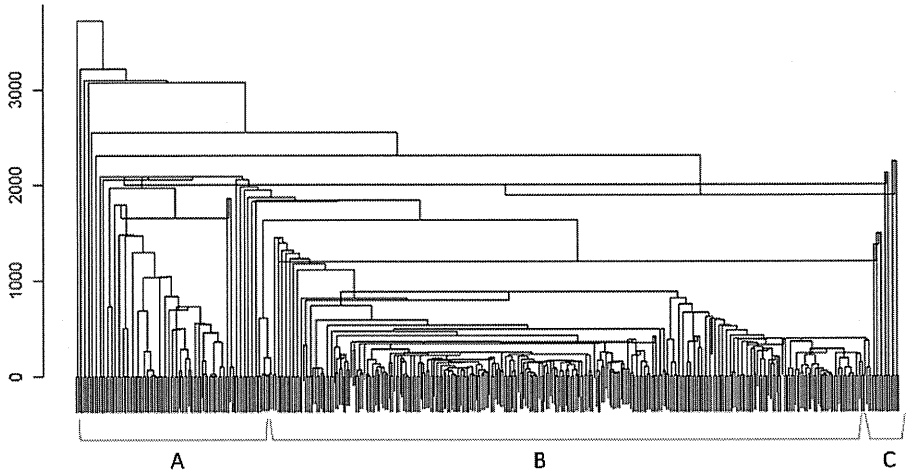
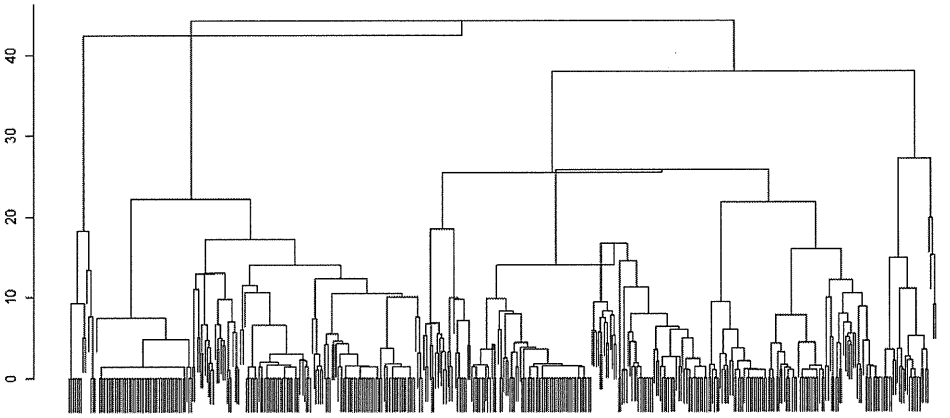


Figure 5 Dendrogram of traditional method



6 An application to PCD

In this section, we show the case of PCD.

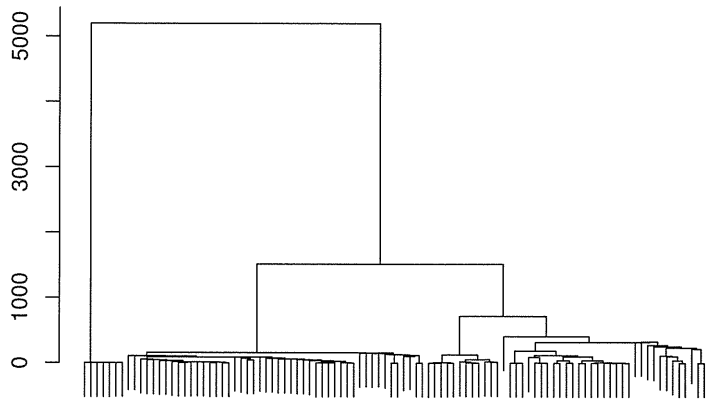
6.1 Data description and result

We also use the medical questionnaire in Keio University School of Medicine. We deal with four question: "Do you feel cold in your leg?", "Do you feel pain in your leg?", "Do you feel cold in your hand?", "Do you feel pain in your hand?". The data contain 113 patients' first and second VAS value. We transform this dataset to PCD.

6.2 Result

The result of our simulation is shown in Figure 6. Vertical axis of this dendrogram means distance between PCDs.

Figure 6 Dendrogram for PCDs



7 Concluding remarks

In this paper, we defined PD and PCD that are from transformation of the VAS to distribution-valued data. We also proposed hierarchical clustering method for them. Comparing across a group of patients by using the VAS is difficult, but our method can do it. Through the simulation, we verified our model.

References

Billard, L. and Diday, E. (2006) *Symbolic Data Analysis*, Wiley, NewYork.

Dexter F. and Chestnut, D.H. (1995) 'Analysis of statistical tests to compare visual analog scale measurements among groups', *Anesthesiology*, Vol. 82, No. 4, pp.896–902.

Goldberger, J., Gordon, S. and Greenspan, H. (2006) 'An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures', *Proceedings of CVPR*, pp.487–494.

Kullback, S. (1968) *Information Theory and Statistics*, Dover Publications, New York.

Price, D.D., Bush, F.M., Long, S. and Harkins, S.W. (1994) 'A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales', *Pain*, Vol. 56, No. 2, pp.217–226.

Revealing Modern History of Japanese Philosophy Using Natural Language Processing and Visualization

Hideki Mima, Katsuya Masuda, Susumu Ota, and Shunya Yoshimi
Center for Knowledge Structuring, University of Tokyo

Type of presentation: Poster
Keywords: Natural language processing, visualization, Japanese philosophy, thoughts, knowledge structuring
Contact email address: mima@t-adm.t.u-tokyo.ac.jp
Postal address: 7-3-1 Hongou Bunkyou-ku Tokyo 113-8656, Japan

Abstract

The purpose of this study was to reveal the modern history of Japanese philosophy using natural language processing (NLP) and visualization. Knowledge[1] has been increasing at an exponential rate with advances in science and technology in recent years resulting in massive amounts of knowledge that have been extremely difficult to process manually. Thus, it is important to utilize information technologies (IT) to support new discoveries of knowledge from large numbers of resources, such as literature. To implement the study, we have developed:

1) A corpus representing a modern history of Japanese philosophy,

2) A computational model for extracting ontology[2] from the corpus, and

3) An interactive user interface (UI) to support new discoveries of knowledge.

We chose "Shisou" (thoughts) by the Japanese publisher Iwanami Shoten for the target corpus, which is one of the most representative journals of philosophy in Japan that has an almost 90 year history from 1921 to the present-day. It is comprised of about 8,600 papers and more than 160,000 pages of textual data. The first step in this study was to develop a technology to digitize such large amounts of textual data from physical books (semi-) automatically. Because the target was too huge to digitize manually (i.e.,

---

[1] Although the definition of knowledge is domain-specific, our definition of knowledge here is the particles represented by ontology, which is the (hierarchical) collection and classification of (technical) terms used to recognize their semantic relevance.

[2] Although the definition of ontology is also domain-specific, our definition of ontology h ere is, as previously mentioned, the (hierarchical) collection and classification of (technical) terms used to recognize their semantic relevance.

by typing), a rapid, accurate and low-cost approach was required. Thus, we developed an Optical Character Reader (OCR) based (semi-) automatic book-digitizing system, in which we integrated three processes:

i) Book scanning
ii) OCR
iii) Automatic document style recognition

The input for the system were physical books and the output was a full-text corpus with meta-data, i.e. titles, authors, page numbers, and dates.

We propose a knowledge structuring (KS) system[1] to integrate NLP and the visualization-based interactive UI for the model of ontology extraction and UI. The system architecture is modular, and it integrates five components (Fig. 1): a) information (ontology) extraction, b) a corpus database, c) information retrieval, d) similarity calculations, and e) visualization.

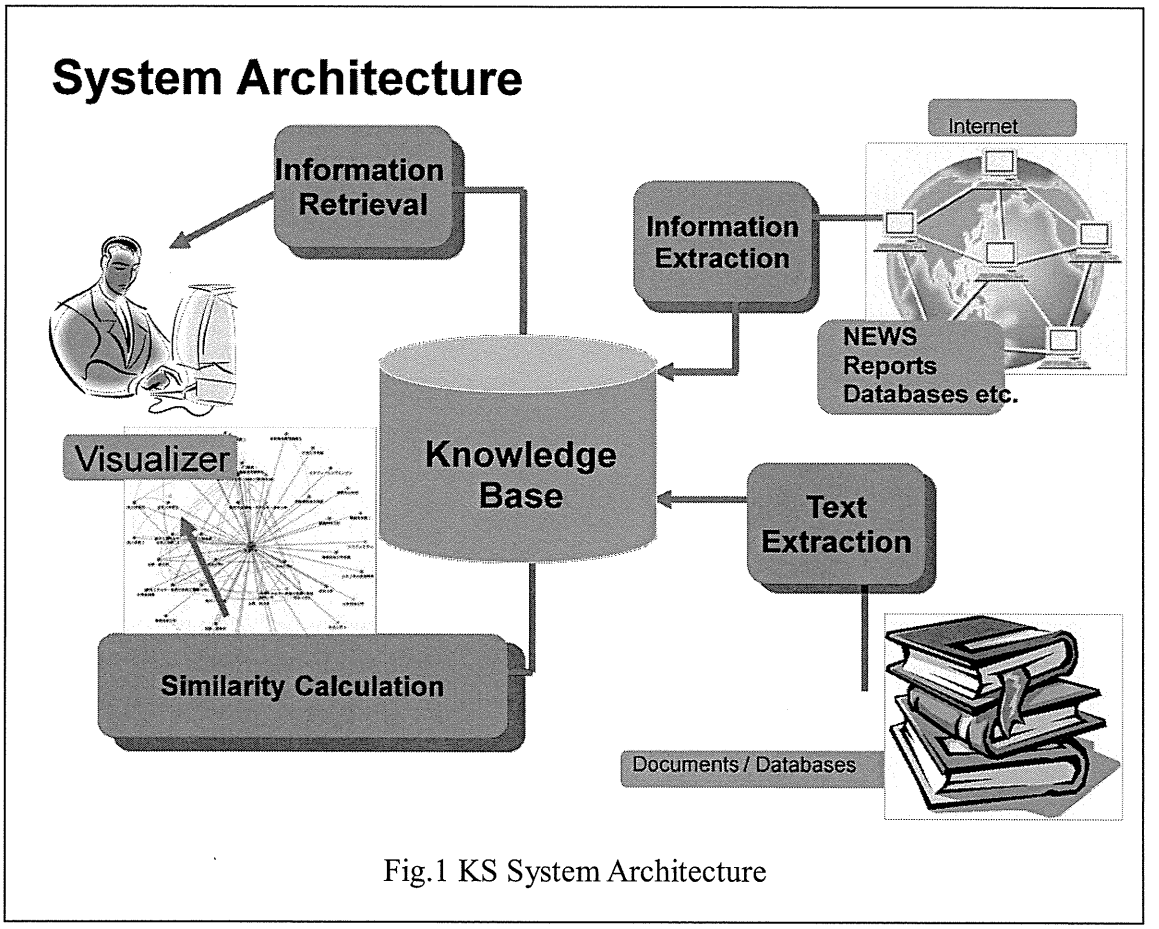


Fig.1 KS System Architecture

The main objective of the system was to facilitate knowledge acquisition from documents and generate ideas through terminology-based real-time calculations of document similarities and their visualization with an interactive UI. Fig. 2 outlines the visualization of knowledge structures for *shisou* papers relevant to the keyword "shisou (thoughts)" in the 1930s. The system constructs a graph to structure knowledge in which

the nodes (dots) reflect relevant papers with the keyword, and the links between the nodes reflect semantic similarities that are calculated based on terminological information in the papers. Additionally, the locations of all nodes are calculated and optimized when the graph is drawn. The distance between each node depends on how close they are in meaning. Cluster recognition is also carried out based on the detection of groups of papers in which every combination of papers that are included is strongly linked (i.e., their similarity exceeds a threshold). As seen in Fig. 2, several clusters are automatically recognized and category names such as "Marxism", "socialism" and "right-wing thoughts" are also automatically assigned to clusters to facilitate an overview of thoughts discussed in these papers.

We have currently finished digitizing and creating a "Shisou" textual database of the 20 years from 1940 to 1959 and installed it in the KS system. Several experiments on text digitization were conducted to evaluate the OCR and style recognition process to improve accuracy. We obtained more than 98% accuracy in OCR, about 90% accuracy in style recognition according to the latest evaluation.

We expect to discover new knowledge on the historical flow of Japanese thinking during one of its most important eras from before World War II to the present-day by digitizing and analyzing huge amounts of historical textual data with the system.

References

[1] Mima, H. and Ananiadou, S. "An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-word units in Japanese." *International Journal on Terminology*, 6 (2), pp. 175–194, 2000.

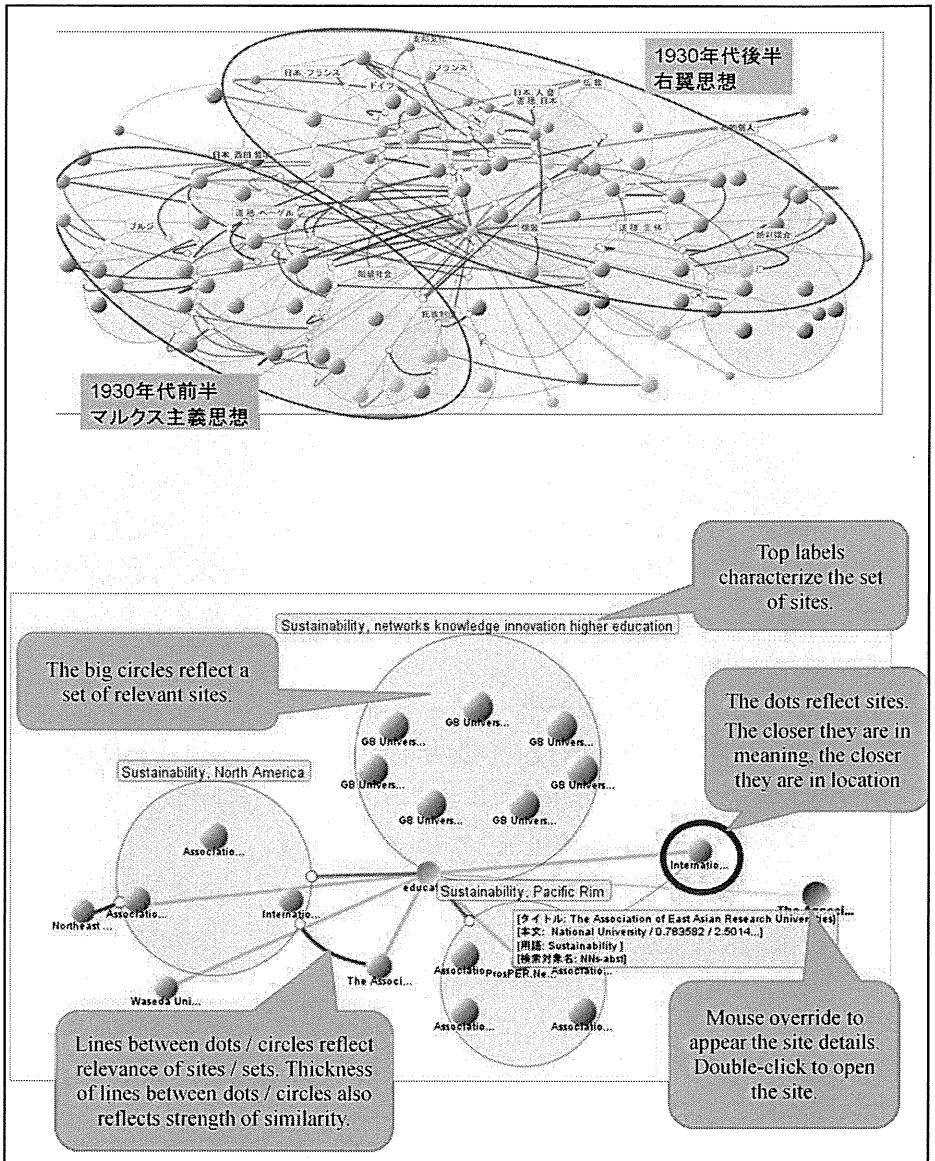


Fig.2 Visualization of knowledge structures

生命科学における知の構造化

東京大学大学院 工学系研究科（兼任）東京大学 知の構造化センター　美馬 秀樹

Key words　知の構造化 ／ 自然言語処理 ／ 可視化 ／ オントロジー ／ 漢方医学

増え続ける“知識”

1800万件／3万件……、これらはそれぞれ、MEDLINE(医学・生命科学分野の文献データベース)に登録されている文献数、および毎月のおおよその増加数(2010年時)である。ICT(情報通信技術)の発展、科学の拡大、専門分野の深化を背景に、生命科学分野のみならず、あらゆる分野において知識の量が爆発的に増加しており、非専門家はもとより専門家にとっても知識の全体像の把握が非常に困難な状況となっている。さらには、環境やエネルギーのような地球規模での複雑で多様な問題が顕現化し、学際的、分野横断的に知識の活用を促す仕組みの構築がより重要性を増している。

東京大学 知の構造化センターでは、このように自律分散的に創造される膨大な知識を構造化し、新しい知的価値、経済的価値、社会的価値、文化的価値に結びつける「知の構造化」の研究開発を進めている。「知の構造化」により、時勢や学問分野間、また、人や組織間を越えた知識の「インターフェーシング」を行い、知の要素と要素の関係からその全体像を明らかにすることで、多様な知を関連づけ、新しい価値を創出することを目指している。

例えば、医療において、近年、医学と工学の連携により発明された技術として、「3次元血管造影診断技術」がある。血管造影は心筋梗塞の重症度診断等、さまざまな診断で利用されているが、従来は腕や大腿部の動脈から細い管(カテーテル)を入れて造影剤を流し込み、映画撮影するものであった。これは、治療法の選択等にも欠くことのできない有用な検査であるが、患者の時間的、体力的な負担が大きく、簡単にくり返して行えるものではなかった。

これに対し、「3次元血管造影診断」は、ITによる高速センシングと3次元CG(Computer Graphics)を利用した可視化技術により、短時間に検査を行うことができ、患者の負担軽減の観点からもその価値は計り知れない。この発明は、図1に示すように、「医療」と「情報工学」に係る知の構造化、さらには「造影」と「可視化」という知の合成なしには、なし得なかったものである。

本稿では、このような、知識を効

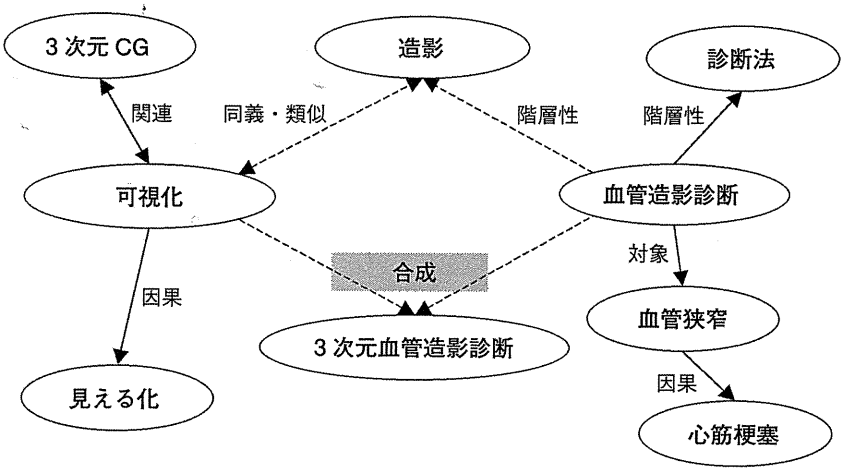


図1　生命科学における知の構造化の例

「情報」「知識」、および「知」という言葉は、ドメインや文脈、状況等によりさまざまな意味を持つ。例えば、Webを対象とした情報抽出は、HTML文書から特定の部分のテキストを抽出することを指す場合が多いが、自然言語処理では、さらにテキストから固有名詞等の特定の情報を抽出することを指す。よって、本稿においても、それらを厳密に定義しないが、知の構造化の対象をテキストとした際の「知」および「知識」の対象としては、(専門)用語等の属性により特徴付けられたパーティクル(文、段落、節、文書等の単位、またはそれらと関連付けられたコンテンツ)を示すものとする。

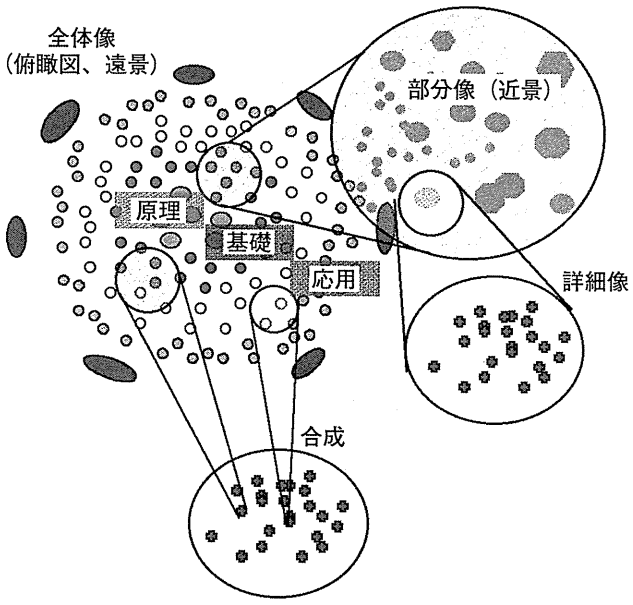


図２　知の構造化のモデル

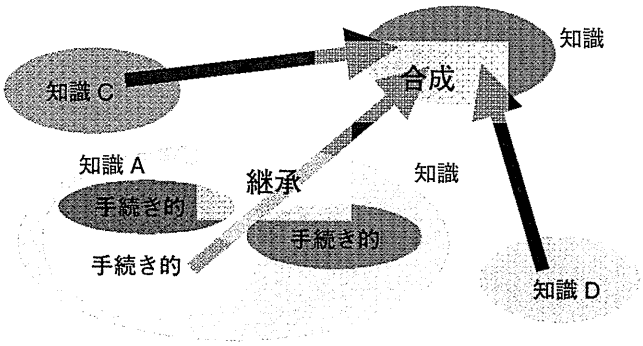


図３　知識の継承と合成

率的に活用し、価値を生み出す技術、方法論としての「知の構造化」を概説するとともに、生命科学分野への適用例をもとに、いかに可視化、構造化の技術が知の創出、活用に貢献するかについて述べる。

知の構造化の必要性

科学の拡大、深化、それにともなう分野の分化を背景に、自律分散的に創造、管理される知識 の活用に際し、問題点として以下を考える。

・ 情報過多、知識過多
・ 過度の細分化
・ 縦割り型(階層型)知識管理

結果として、

・ 知識の相互の繋がり
・ 知識の重複、差分
・ 知識の抜け

が不明瞭となっているのが現状である。

例えば、コンピュータ2000年問題、大銀行の統合のように、誰一人システムの全体像を把握していないという状況が生じる原因となる。分野を超えて知識を理解し、活用するためには、知識の全体像を明らかにすることが先決であり、総じて、「他を知る、他をわかる」ことが非常に重要となる。

知の構造化のモデル

ここでの「知の構造化」の目的は、膨大な知識を対象とし、

a) 知識間の隙間を埋める知識の発見
b) 知識間をまたぐ知識の発見

を行うことにある。一般に、すべての概念の関連は絶対的に定義できるものではないため、すべての場合において上記を厳密に区別し処理を進めることは困難であるが、すでにある例として、a)に対してはバイオインフォマティックス等、b)に対しては環境科学等がそれに当たるといえる。これを実現するための構造化モデルとして以下を考える(図2)。

1) 全体像の把握

知識の既存の関連や属性に基づく関連を抽出し、知識間の個々の関連から全体の関連を明らかにする。細分化や縦割りの弊害等により、失われがちな関連をも見つけ出すことが先決であり、オントロジー、可視化、見える化等の技術が重要な要素となる。

2) 抽象化と詳細化

膨大な量の知識の全体像を把握するためには、抽象化は必須である。抽象化された領域より、必要とする知識を選択した後、その領域の詳細化へと進めることで、必要な知識の絞り込みが容易になる。いわば、「森を見て、木を見る」操作である。

3) 合成

さまざまな知識から新たな知識を創造するためには、既存の知識を如何に再利用するかが重要である。異なる分野の知識を上記、抽象化等の操作により選択し、合成することで、より新しい知識の創出が期待される。また、創出された新たな知識を次の合成の種へとリサイクル、リファインメントをくり返すことで、知識はより成熟する(図3)。

これらの構造化、可視化、および操作が、個々人、および任意の視点によりリアルタイムに行えることが重要である。つまり、任意の視点で詳細化、抽象化の階層を上下しつつ、関連のある、もしくは関連が必要な知識を選択し、合成の要素を探すのである(図2,3)。さらには、次の瞬間

にこれら新たに創出された知識が次代の合成や抽象化の対象となる。このように、知識の連続的創出と活用を促し、さらに高度な知識の再活用へと昇華させるためには、知識創出、活用の「螺旋」を形成できることが重要である。以下では、これらを支援する技術に関し、詳述する。

知の構造化システム —「MIMAサーチ」—

知の構造化センターでは、上記の方法論を実践し、テキスト情報を対象とした知の構造化を支援する機構として「MIMAサーチ[1)]」を開発し、その実用化を行ってきた。「MIMAサーチ」においては、自然言語書処理を活用することで、膨大なテキスト情報より瞬時に必要とする知識を抽出し、さらに抽出した知識間の関連性を自動で計算する。一般に、自然言語処理とは、形態素解析、構文解析、意味解析等により計算機を用いて言語の理解を行うことを指す。従来、これらを用いた仮名漢字変換、機械翻訳システム、用語(概念)抽出システム、全文検索システム等のアプリケーションが開発されており、現在では、計算機の発展により大量の言語情報を高速に処理することが可能となっている。

「MIMAサーチ」の特長は、図4に示すように、論文や、報告、アンケート等に記述されている自由文(テキスト)を自然言語処理により解析し、その統計情報に基づき、オントロジーとして重要な用語(概念)を自動的に認識・抽出することにある。

さらに、抽出したオントロジーを比較し、関連をとらえることで、文書間の意味的関連とその関連の強さを定量的に計算する。そして、それらを視覚的にとらえることができるよう、関連およびその強さをグラフモデルにより可視化する。つまり、単なる個々の文書の内容をとらえるだけではなく、文書間の意味的な関連にもとづいて全体を俯瞰し、知識を抽象化してとらえることができることを意味する。

より具体的には、「MIMAサーチ」は以下のような特徴を持つ。

- 必要とする分野全体の知識、日々創出されるリアルタイムな情報、共創的に創出される知識を含むさまざまな形態の知識群を統合し、データベースとして蓄積する。
- 上記データベースより、ある分野や領域、または分野横断的に任意の知識を抽出し、抽出された知識全体の関連を可視化する。
- 知識間の関連として、あらかじめ定義された情報、もしくは手続きにより導出される類似、包含(差分)、部分全体、因果、を含むオントロジー的関連が参照できる。
- 上記はキーワード等により指定される任意の視点を反映できる。
- 上記により指定、もしくは計算された関連をもとに、関連の強い知識同士をまとめあげる(クラスタリング/クラシフィケーション機能)。
- 上記のまとめあげを任意の抽象度で可視化する(階層的クラスタリング機能)。
- 任意の知識を選択し、また必要な知識を加えることで新たな知識を創出し、データベースに追加できる。

例えば、これらにより、複数の分野から横断的に知識を検索し、関連度指定、抽象度指定により計算されたクラスターから任意の知識を選択、さらにこれらを合成するという流れが実現可能である。

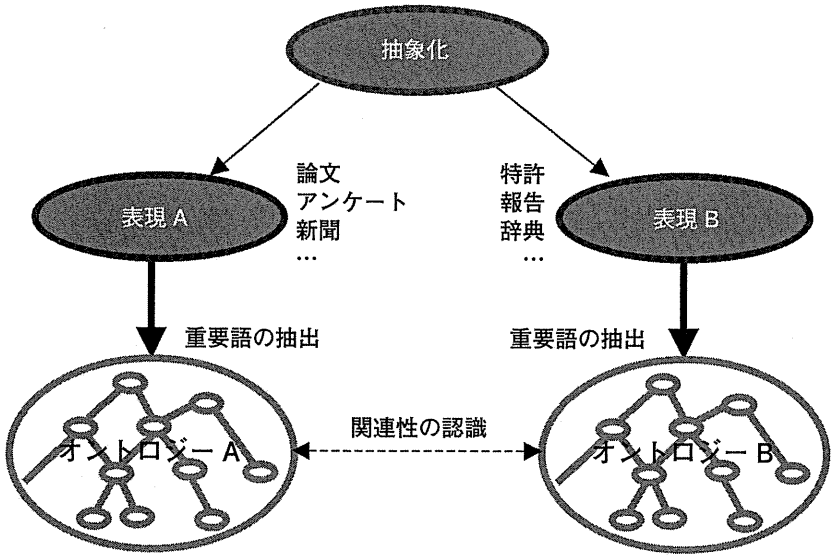


図4　特徴の抽出と関連性の認識

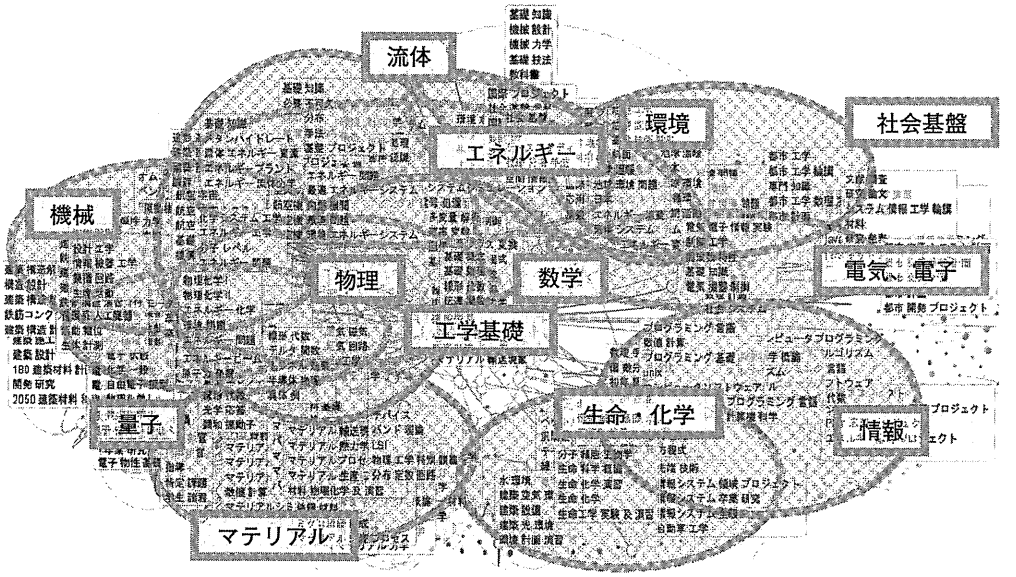


図5　工学シラバスの構造化例

例えば、図5に東京大学工学部講義シラバス(2006年度版の約900講義が対象)に対し、MIMAサーチにより処理をした例を示す。

図では、「数学」、「物理」等の原理、基礎から「情報」、「社会基盤」のような応用に至る工学知の全体像が見てとれる。また、「流体」、「エネルギー」、「環境」のような分野をまたぐ知の存在が確認できるのも特徴である。ここで注目すべきは、「環境」と「生命・化学」の隙間にある知識の"抜け"である。シラバスの充実とともに、「環境生命」のような分野の知を補うことが期待される。

なお、MIMAサーチでは、知識ソースとしてオフィスドキュメント等のファイルサーバやデータベース上にある静的知識、インターネット上の準動的知識に加え、Wikipedia等でも利用されているWikiシステムを統合することで、一般のブラウザで編集が可能な動的知識の管理にも対応している。MIMAサーチとWikiシステムとをシームレスに統合することで、検索した知識の継承や、関連する知識との合成を容易に行うことができる。また、これら継承や合成により新たに創出された知識は、即座に検索対象とすることが可能である。これにより、マルチユーザ環境では、他のユーザとも即時に知識共有が行える。

生命科学論文の構造化

生命科学分野における論文の加速的な増加は冒頭でも述べたが、爆発的に増加する論文のすべてに1人の人間が目を通すのは、すでに不可能な状況になっていることが容易に想像できる。にもかかわらず、論文の査読や発明特許の申請等においては、既知の事項との重複がないか等の、関連する分野の知識を網羅的に把握する必要がある。このような目的においても、知の構造化技術を利用することで、まずは分野全体の知識を俯瞰し、全体のなかでの位置づけをつかんだうえで、さらにその位置の詳細を確認するといった、「全体像」から「詳細像」、さらにはまた「全体像」へといったズームインとズームアウトをくり返すことで、関連する知識をより効率的に探すことが可能である。

より具体的には、まずMIMAサーチで全体を俯瞰し、意味的なまとまりのある部分に絞り込んで検索を進め、主として関連している可能性の高い論文を把握したうえで個々の関連を取り出し、検証するという詳細化のアプローチにより検索や比較の対象を絞ることが考えられる。

例えば、図6(a)(b)はそれぞれ2006年、および2007年に開催された生命科学研究ネットワーク・シンポジウムで発表された論文(それぞれ304件、324件)をMIMAサーチにより可視化したものである。

先にも述べたように、図では、内容が関連する論文がより近くになるように配置されており、よりまとまりのある論文群(クラスター)にはその内容に応じて「分子メカニズム」のような重要な用語を基に計算したトピックラベルが自動で振られている。また、さらに大きなまとまりを円で囲み、「臨床医学」のような分野名のラベルを割り当てている。年度をまたいだ恒常的なテーマが存在するのと同時に、「分子機構」や「メタボリックシンドローム」のような、それぞれの時勢に応じてテーマとなる研究が変遷していく様子が見てとれる。

また、イノベーション支援や知識創造支援の観点では、境界領域の設定のような分野を横断した関連が増加することが望ましいといえるが、図において、2006年度から2007年度への分野の変遷をみると、2006年度に比較して、2007年度には「新学術領域」と「臨床医学」や「工学」との関連が増加し、全体の繋がりがより明確になり、全体像が凝縮されていることが見てとれる。これは主に、シンポジウム等により人的交流が増加したためと推察できるが、このように、ある種の仮説検証のプロセスの一部として、知の構造化と可視化を活用することも可能である。

社会のイノベーションに対する期待が大きいなか、情報過多、知識過多により、十分活用されていない情報や知識が多く存在しているのも事実である。IT、シミュレーションなどにより、社会の効率化、自動化を目指す一方で、それを利用し、価値を生み出すのは、将来においても、やはり「人」自身である。その意味でも、膨大な情報や知識から、有用な知識やその関連を抽出し、人による知の創出、活用、価値化をいかに支援するかが「知の構造化」の本質であろう。

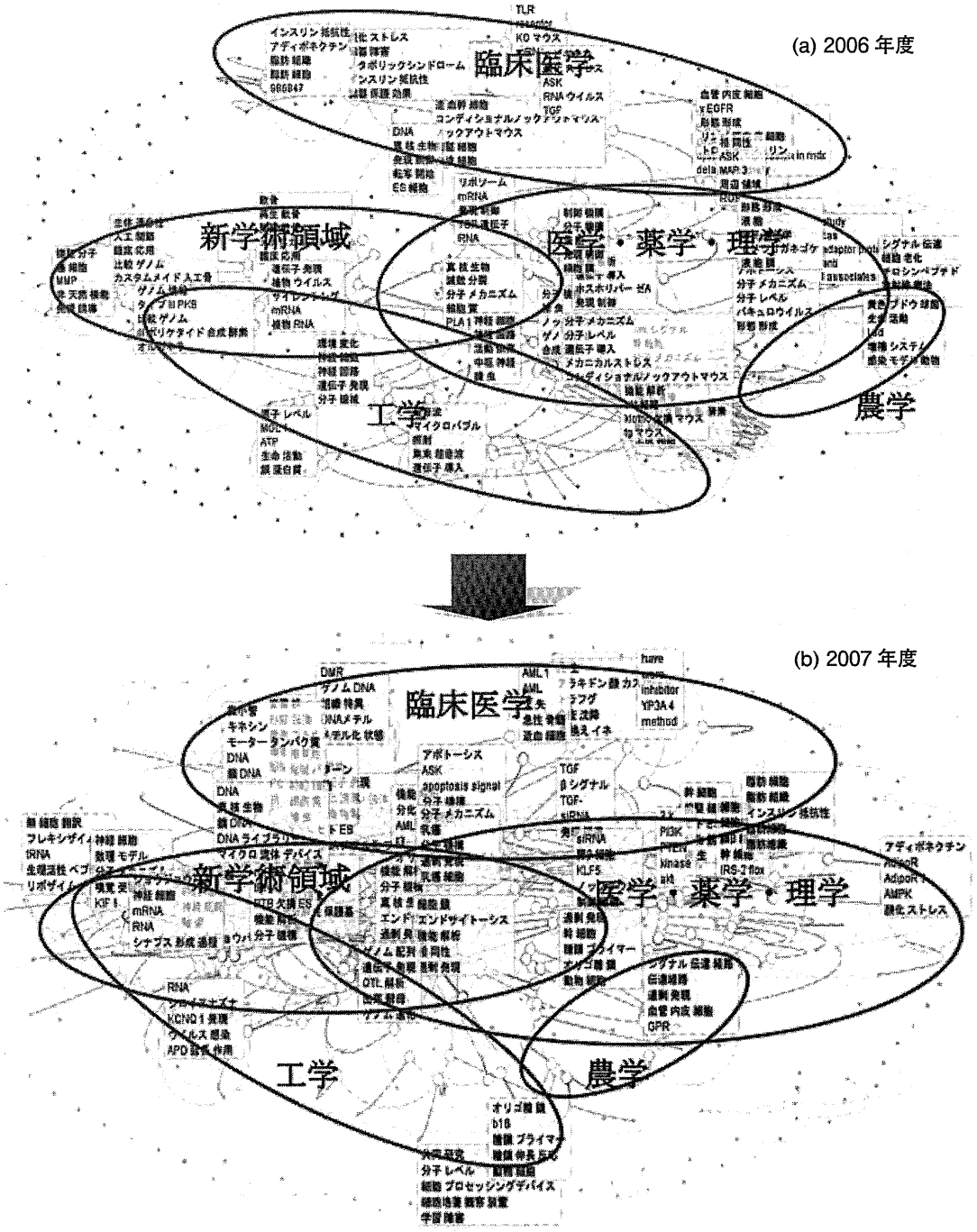


図 6　論文関連性の時間的推移

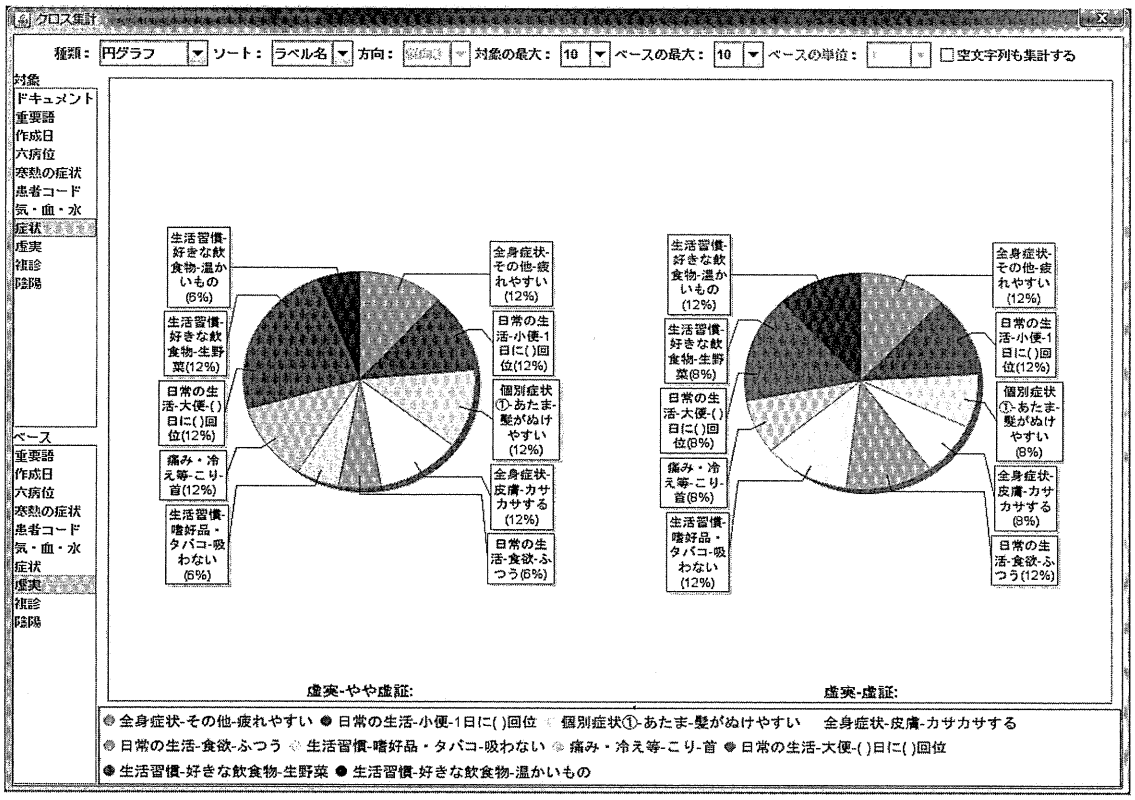


図7 「証」と問診との関連

臨床医学の構造化 ―漢方診療のエビデンス創出と診断支援

本研究では、臨床医学の構造化として、漢方医学を対象とした診療のエビデンス創出、および創出されたエビデンスをもとにした診断支援システムの構築を進めている。今日では、医師の7割以上が日常診療で漢方を用いているにもかかわらず、漢方専門医以外は漢方薬の使用処方は限定されているのが現状である。これは、漢方診療が、同病異治・異病同治に示されるような、個別化診療であることが主な原因である。

また、漢方医学的診断の特徴である「証」の特定はKnow-Howのような経験知から導かれるものであり、いまだ十分に科学的・統計学的な解明が為されているとは言い難い。つまり、臨床的エビデンス(診断・治療)が得られるようになれば、一般臨床医もある一定のレベルで効果的な漢方治療ができると期待される。これを目的とし、本研究では、まず、漢方薬および鍼灸治療にともなう患者の自覚症状の推移を、外来に設置した入力端末を活用して系統的に収集し、電子カルテの医療情報とともにデータベース化を行っている。さらに、知の構造化の技術を利用することによって、治療効果の判定や漢方・鍼灸の診断「証」と症状との関連性を解析する。このようにITを活用した伝統医学の新たな臨床研究の手法を開発することで、漢方の診断と治療の科学的検証を行うことを目指している。例えば、図7には「証」(虚実)における「やや虚証」と「虚証」の診断に対する微妙な差異を統計と可視化により明確化した例であるが、本研究により、このようなKnow-Howにかかる経験的、暗黙的知識の「見える化」が行えると期待される。

実際のデータ収集においては、問診システムにて評価された患者の状態に対し、投与された薬剤や処置を問診終了後に医師が入力することとし、再診時以降は、患者の視点から評価された症状の連続的な変化が、治療経過とともに時系列で記録されている。現在、平成17年度、18年度初診患者1700名余りの診療がデータベース化されており、このデータを基にMIMAサーチによる解析を進めている。従来のこのような研究は、収集したデータの統計情報をもとに、定量的分析を行うのみであったが、さらに、MIMAサーチを利用することで可視化技術を活用した定性的分析手法との統合的解析を行っていることも本研究の大きな特徴である。

図8に、頭痛に関する患者のMIMAサーチによる分析と可視化の結果を示す。図では、グラフモデルによる患者間の関連の可視化が行われ、それぞれの点が患者を表し、患者間に引かれている線の長さと太さがそれ

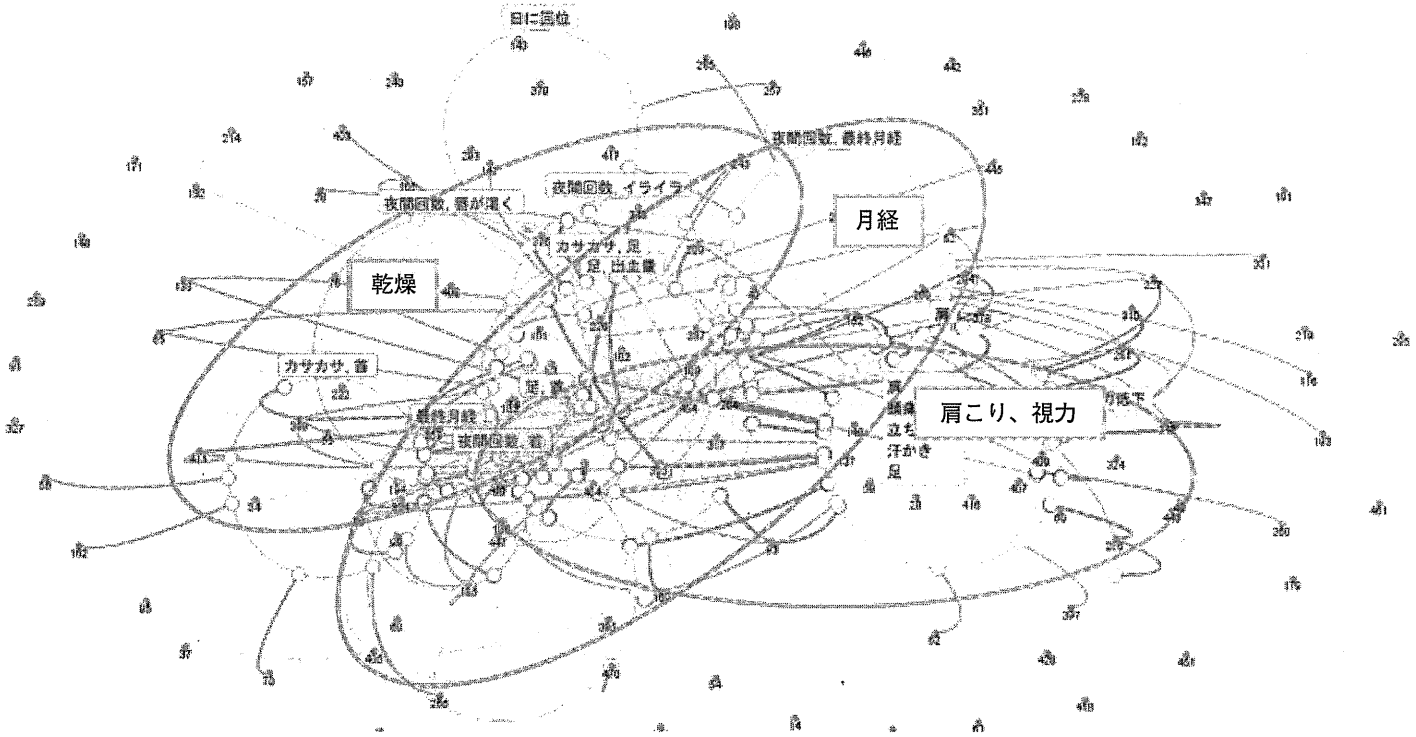


図8　頭痛に関する患者の分析結果

ら患者間の関連の強さを示す。つまり、体質や症状が似ている患者がより近くに配置されるようになる。また、図中の大きな円により関連の強い患者群(クラスター)が示され、これら円の上に示される内容が問診に表れる特徴を示している。

図からも容易に読み取れるように、頭痛に関する患者では、肩こりや視力低下、乾燥(アレルギー)、月経異常等の患者群との関連が示されている。新たな患者がどのパターンに最も近いかということや、どのような位置付けにあるか等の観察により、証の理解や、新たな治療法の開発につながるものと期待される。

溢れる情報をムダなく活かす

本稿では、分野を超えた知識の活用を目的に、可視化をはじめとする知の構造化のモデル化、また、モデル実現のためのシステム設計とその一実装である「MIMAサーチ」に関して述べた。また、実際に「MIMAサーチ」を活用した生命科学分野における知の構造化の研究を紹介した。

社会では種々雑多な情報や知識のやり取りが日々洪水の流れのように行われている。現状では、それらに1つ1つに目を通し、理解するのはほぼ不可能である。それでも、我々はそれらから意味ある情報を読み取っていかなければならない。本稿で紹介した構造化とその技術によって、日々流入する情報を構造化し、可視化すれば、それらを俯瞰することが可能になり、知識の取捨選択は大いに容易になるはずである。

リアルタイムに流れてくる情報や知識を取捨選択しつつ、価値のある知識のみを抜き取り、そこから仮説を構築し、すでに蓄積された知識を活用しながら、仮説を検証するサイクルにおいては、必要な知識を必要なタイミングで獲得できることが不可欠である。 その意味でも、知の構造化と可視化への期待は非常に大きい。

[参考文献]

1) Mima H, et al.: ACM TALIP, 5: 74-88, 2006.

Author 著者

美馬 秀樹(みま ひでき)

東京大学大学院 工学系研究科 国際工学教育推進機構 特任准教授 博士(工学)

1996年、徳島大学工学研究科システム工学専攻博士課程修了。株式会社ジャストシステム研究員、ATR音声翻訳通信研究所 研究員、英国マンチェスターメトロポリタン大学Lecturer、東京大学大学院理学系研究科 研究員、東京大学大学院工学系研究科助手を経て現職。自然言語処理、知的ユーザインタフェース、機械翻訳、知識処理、知識管理、可視化の研究と実用システムの開発に興味を持つ。平成15年度情報処理振興事業機構(IPA)未踏ソフトウェアプロジェクト 天才プログラマ/スーパクリエータ認定。平成16年 The international Daiwa Adrian prize, a triennial award for excellence in scientific collaboration between the UK and Japan, 共同受賞。2006年より英国マンチェスター大学Informatics専攻(School of Informatics)Senior honorary research fellow兼任。

漢方医学をめぐる最近の動向

Current Situation of Kampo Medicine

渡辺賢治 Kenji WATANABE
慶應義塾大学医学部漢方医学センター

日本の９割の医師が使う漢方

2008年の日本漢方生薬製剤協会の調査では医師の83.5%が漢方を使うという数字に驚かされたが[1]，2011年の調査では89%に上昇しており，まさに医療現場にはなくてはならない存在となってきた．この傾向は日本特有のものではない．世界中で伝統医学の見直しが行われ，しかも経済発展している国ほど伝統医学を使用するという実態がWHOの調査でも示されている[2]．

近年の遺伝子治療薬の発達に代表されるように，標的を明らかにした治療と，複合物で作用機序もすべてが明らかになっていない漢方薬という一見矛盾するような医学が並立して，しかも両者が同時に発展している，というこの事実を，どう解釈すべきであろうか？

疾病の発症機序そのものの解明が進み，それに対してピンポイントの治療を行うという医療の方向性は今後ますます発展するであろう．しかしここ数十年の経験から，ピンポイントの標的を有する医薬品が，かならずしも標的だけを攻撃するものではない，ということが明らかになってきた．たとえば，酵素阻害剤である抗高脂血症薬のスタチンが筋肉に影響を与えるなど，まったく標的とは関係のない体の部分に副作用が出現する．当たり前のことであるが，池に一つ石を投げ入れてもその波紋が広がるように，生体はシステムであるので，標的は一つであってもその影響は全身に及ぶ．

学生に漢方と化合物の西洋薬との違いを尋ねると，漢方は複合物であるから複数の標的があり，化合物は標的がひとつで影響が限定されている，という単純化した答えをするが，根本的に誤りであることは容易におわかりいただけると思う．分子標的薬のようにたとえ標的は一つであったとしても，その影響は全身に及ぶのである．

このように薬の開発において，要素還元論的発想のみでは予想できないことが起こりうることを，研究の最前線でも感じはじめている．ましてや医療現場においては，臨床医の多くが，「部分の集合が全体ではない」ことを実感しながら日常の診療を行っているのである．

その意味において漢方医学のもつ複合的全体主義の考え方が，医師に徐々に受け入れられつつあるのかもしれない．

共通保健統計プラットフォーム

このように医療現場で伝統医学が見直されて同時に用いられるようになると，西洋医学との連携が必要となる．そのひとつの動きが，WHOのICD改訂作業であろう．ICDは，正式にはInternational Statistical Classification of Diseases and Related Health Problems(疾病および関連保健問題の国際統計分

類)とよばれ，異なる国や地域から，異なる時点で集計された死亡や疾病のデータの体系的な記録，分析，解釈および比較を行うため，世界保健機関憲章に基づき，世界保健機関(WHO)が作成した分類である．最新の分類はICDの第10回目の修正版として，1990年の第43回世界保健総会において採択されたものであり，ICD-10とよばれている．

1900年にはじまったICDは当初死因統計のためのものであったが，近年では疾病分類にまで広がりつつあり，わが国でも診断群分類包括制度(DPC)がICD-10に準拠している．

現在ICDの改訂作業が進行しているが，そのなかに伝統医学を入れる計画が明らかにされた[3-5]．世界に広がる伝統医学は，いままで保健統計上ほとんど正確なデータは取られてこなかったが，ICD-11に入り西洋医学と共通の統計プラットフォームができることで，どのような疾病に対して用いられているか，西洋医学の病名とどのような対応関係があるかなどのデータが表れてくることが期待される．

作用機序の解明

もうひとつ西洋医学と伝統医学を結ぶ共通プラットフォームが作用機序の解明であろう．漢方が臨床的に有用であることは認められつつあるが，多くの医師が作用機序が明らかでないので，使いにくいという[1]．漢方が医療用として大々的に収載されたのは1976年であり，その間に数多くの質の高い基礎研究がなされている．残念ながらほとんどが日本語であるために世界に知られていないが，研究のレベルはけっして低くない．近年，漢方薬のような複合物の研究が世界の一流紙に掲載されるようになったことは喜ばしいことである．たとえば，CPT-11に対する遅発性の重篤な下痢に対して半夏瀉心湯が有効であることは診療でもよく知られている．これはCPT-11の活性物質であるSN-38が肝でグルクロン酸抱合して胆汁中に排泄され，腸管に達した後，そのまま便中に排泄されれば問題ないのであるが，腸内細菌によりグルクロン酸がはずれるために，ふたたび吸収され腸管循環することによって起こる．半夏瀉心湯は黄芩という生薬が含まれるが，黄芩に含まれるバイカリンが，このグルクロン酸抱合がはずれるのを競合阻害するために再吸収を妨げ，腸管循環しないために下痢を抑制する，という作用機序は1997年にすでにわが国で報告している[6]．しかし，2010年にはエール大学のグループが，黄芩湯という黄芩を含む漢方薬で同様の結果を示しており，このときは『*Science*』誌に掲載されたのである[7]．データの質はほとんど変わらないが，時代の流れであろうか．最近，インパクトファクターの高い英文誌に，漢方関連の論文が掲載されることが多い．しかし，世界的にみると中国，韓国，香港などが盛んに一流の英文誌に投稿しているのに対し，わが国の掲載数はそれほど伸びていない．中国などの友人からは，日本の生薬学の存在感が最近とみに薄いという指摘を受ける．薬学部6年生移行に伴い，日本での生薬研究者が減少しているせいであろうか．伝統医学が見直されている現代において，懸念される点である．

臨床研究

1990年代にevidence based medicine(EBM)の必要性が叫ばれはじめてから，臨床研究で効果の根拠を示すことが求められるようになった．漢方に関しては，和文・英文合わせて345のRCTが日本東洋医学会によって集積されており，構造化抄録も和文・英文で利用可能である[8]．

しかしその多くが和文であり，世界の臨床医に読まれているかというと，残念ながらかならずしもそうではない．前述の基礎研究同様，最近では伝統医学の臨床研究が一流の英文誌に掲載される時代となりつつあるが，やはり中国からは数多くの臨床研究が投稿されるのに比べ，わが国ではまだまだ数が少

ない．

最近では，2009年に流行した新型インフルエンザに対する麻杏甘石湯と銀翹散を合わせた蓮花清瘟カプセルのオセルタミビルとの比較試験が記憶に新しい．『*Annals of Internal Medicine*』誌に掲載されたが[9]，国が主導して新型インフルエンザに対する漢方薬の効果を示したものである．研究費や支援体制など，わが国が学ぶべきものも多い．

また，一方で漢方の臨床研究に関しては，西洋医学と同じ研究デザインで行うことに対して多くの議論がある．すなわち，漢方の診断である“証”を基盤として，①個別化医療であり，②患者主観を重視している漢方に対して，果たして西洋医学的ゴールデンスタンダードである無作為比較試験がふさわしいかどうかという点である．

ICT(情報通信技術)の発達により，システムズバイオロジーで臨床的エビデンスを示せる時代に入りつつあり，すでにいくつかのマルチディメンジョナルな解析法が示されつつある．今後の解析技術の開発により，漢方によりふさわしい研究デザインがなされることを期待したい[10]．

本シリーズの特徴

本シリーズでは最新の漢方の知見を，各領域における第一人者の先生方に紹介してもらうことを目的としている．おもに臨床的エビデンスを示してもらいながら，その作用機序がどこまでわかっているかという解説をお願いしている．漢方がはじめて医療用として収載されてから45年になるが，漢方がここまで解明されてきている，ということを読者の皆様にお示しできることと期待している．

文献/URL

1) Moschik, E. C. et al.: Usage and attitudes of physicians in Japan concerning traditional Japanese medicine (kampo medicine) : a descriptive evaluation of a representative questionnaire-based survey. *Evid. Based Complement. Alternat. Med.*, **2012** : 139818, 2012.
2) Ong, C. K. et al.: WHO Global Atlas of Traditional, Complementary and Alternative Medicine. World Health Organization, Kobe, 2005.
3) Normile, D.: WHO Shines a Light on Traditional Medicine. Science Insider Dec. **6** : 2010. http://news.sciencemag.org/scienceinsider/2010/12/who-shines-a-light-on-traditional.html
4) Watanabe, K. et al.: Asian medicine : A way to compare data. *Nature*, **482**(7384) : 162, 2012.
5) Cameron, S. et al.: Asian medicine : Japan's paradigm. *Nature*, **482**(7383) : 35, 2012.
6) Kase, Y. et al.: Preventive effects of Hange-shashin-to on irinotecan hydrochloride-caused diarrhea and its relevance to the colonic prostaglandin E2 and water absorption in the rat. *Jpn. J. Pharmacol.*, **75**(4) : 407-413, 1997.
7) Lam, W. et al.: The four-herb Chinese medicine PHY906 reduces chemotherapy-induced gastrointestinal toxicity. *Sci. Transl. Med.*, **2**(45) : 45ra59, 2010.
8) 漢方治療エビデンスレポート2010，日本東洋医学会．http://www.jsom.or.jp/medical/ebm/er/index.html
9) Wang, C. et al.: Oseltamivir compared with the Chinese traditional therapy maxingshigan-yinqiaosan in the treatment of H1N1 influenza : a randomized trial. *Ann. Intern. Med.*, **155**(4) : 217-225, 2011.
10) Watanabe, K. 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.*, **2011** : 513842, 2011.

*　　*　　*

特集　Alzheimer病の新しい治療薬—実際の使用経験を含めて—

Alzheimer病と漢方薬*

上野眞二**／村松慎一***

Key Words：Traditional Japanese medicine, Kampo, Alzheimer disease, Chotosan, Kamikihito, Yokukansan

はじめに

後漢時代の医学書『傷寒論』に記載された葛根湯，小青竜湯などをはじめ，漢方薬の多くは近代科学の成立よりはるか昔に創薬された．主に植物由来の天然物からなる生薬を配合したもので，たとえば，葛根湯は葛根，麻黄，桂枝，芍薬，大棗，生姜，甘草の7種類の生薬により構成されている．葛根湯の薬理作用では麻黄中に含有されるエフェドリンが重要なことは間違いないが，麻黄を含む他の漢方薬との適応の違いは単一成分では説明し難い．構成生薬中の多数の成分が複雑な相互作用を呈し，代謝酵素活性や腸内細菌叢など患者の個人差も加わるので，漢方薬の薬理学的な解析は容易ではない．詳細な作用機序が不明なため，現在でも各処方の適応は古典に記載された臨床医の経験則によることが多い．

東洋医学の古典には，『黄帝内経』，『傷寒論』に「善忘」，『素問・霊枢』に「喜忘」，『諸病源候論』に「多忙」との用語がみられるのをはじめ，「健忘」，「好忘」，「易忘」など認知機能障害と推察される病態の記載がある[1]．また，『傷寒論』とほぼ同時代の薬物書『神農本草経』には，生薬の遠志について「知恵を益し，耳目を聰明とし，物事を忘れず，志を強くし，力を倍にする」と解説されている[2]．

漢方薬はAlzheimer病(AD)の治療薬としても期待されている[3]．現在のところ，中核症状の改善に確実に有効な漢方薬はないが，behavioral and psychological symptoms of dementia(BPSD)などの周辺症状には効果が報告されている．本稿ではADに使用される代表的な漢方薬ついて，基礎研究の結果も含め紹介する．漢方薬の原典，構成生薬，効能・効用を**表1**に示した[4][5]．また，臨床試験の概要を**表2**にまとめた．

釣藤散（ちょうとうさん）

ラットでは，釣藤散は一酸化窒素(NO)を介する血管拡張，脂質代謝改善，赤血球変形能増加などの作用により脳循環を改善した[6]．一過性脳虚血モデルマウスに釣藤散を前投与すると，学習記憶障害に対して予防効果を示したが，構成生薬の釣藤鉤(Uncaria Hook)を除くとその効果は著しく減弱したため，効果の主な部分は釣藤鉤の作用によると考えられる[7]．釣藤鉤は，アカネ科(*Rubiaceae*)のカギカズラ*Uncaria rhynchophylla* Miquel, *Uncaria sinensis* Haviland, または*Uncaria macrophylla* Wallichの蔓にある釣状の

* Kampo formulae for Alzheimer disease.

** Shinji UYENO, M.D.: 鷲谷病院〔〒321-0346　栃木県宇都宮市下荒針町3618〕; Washiya Hospital, Utsunomiya, Tochigi 321-0346, Japan.

*** Shin-ichi MURAMATSU, M.D.: 自治医科大学東洋医学部門/神経内科学部門; Divisions of Oriental Medicine and Neurology, Jichi Medical University, Shimotsuke, Tochigi, Japan.