COLUMN グローバル化時代の漢方③

漢方医学をめぐる国際的諸問題

国際化の潮流のなかでアイデンティティーを失いつつある漢方医学

ICDの改訂に東アジア伝統医学が取り入れられようとしていることについてはコラム2で述べた．一見順調のようにみえる伝統医学のグローバル化であるが，逆に日本漢方のアイデンティティーが失われてしまう可能性も含んでいる．

日本漢方を推進する著者らの立場は，漢方医学は中国由来ではあるが，すでに1,500年間の日本での発達を遂げているので，日本の伝統医学と考えている．事実同じ処方でも日中韓ではその使い方に相当差がある．

しかし世界をみると，1988年に伝統医学推進のための中国政府組織である"国家中医薬管理局"が創設されて以来，政府主導で中医学の国際化の推進を行ってきた結果，欧米の多くの医師・患者が中医学(traditional Chinese Medicine：TCM)を認識しているのに対し漢方医学(Kampo Medicine)を認識する人はほとんどいない．

中国はTCMという言葉をブランドとして広めたい意向があり，世界各国にネットワークを張っている．そのもっとも大きなものが世界中医薬学会連合会(WFCMS)であろう[1]．2003年に中国政府の援助によって創設され，いまや57の国と地域の195のTCM学術団体から構成される，一大学術コンソーシアムである．当然のことながらこの組織は中医学の国際化を推進するための大きな機動力を担っている．

中国のISOへの提案

コラム2に掲載したようなWHO ICD-11への改訂のなかに伝統医学を入れる計画が進行していくなかで，中国は2008年4月に突然ISO(国際標準化機構)のTC(technical committee)215(保健医療情報)[2]に中国国内の医療情報を国際標準にするように要求した．このときは唐突だったので受け入れられなかったが，2009年10月のダーラムの会議でついに伝統医学のワーキンググループ(WG)をつくることが決定した．ただし，取りまとめは韓国代表が行い，中国の主張したTCM(伝統中医学)のWGではなく，TM(伝統医学)のWGとなった．

それとはまったく別の動きが2009年，中国から既存のISO専門委員会ではなく，新しい専門委員会をつくる，という提案がなされ，TC249として承認された[3]．その委員会名はまだ正式ではないが，"伝統中医学"である．中国の意図としては日本・韓国を抜かして世界標準を自分たちで決めていこうというものである．事務局は上海におかれることになった．

ISOの活動

ISOに関する会議は2010年1月に上海で開かれた準備委員会に加え，第1回会議を2010年6月に北京で，第2回会議は2011年5月にオランダ・ハーグで，第3回会議は2012年5月に韓国・大田で開催された．具体的な作業は5つのワーキング・グループに分かれて進めている．議長国は，**表1**のように中国，韓国，ドイツで分け合っている．このISO TC249が漢方にどのような影響を及ぼすのかについてはまだ明らかでない．たとえば，WG1で議論している生薬の品質としては同じ生薬名でも日中韓で使用植物が異なる場合がある．そうした場合，日本の植物が標準からはずれないか？　また，WG2で議論されている伝統医薬製剤の議論では日中韓で同じ製剤名でも生薬の配合比が異なる．その場合に日本の配合比が否定されないか？　また製造方法が日本と異なるものに標準化されないか？　WG3では鍼灸の規格が議論されているが，質のよい日本の鍼灸の規格がはずされないか？　などいろいろな可能性が想定されるため，注意深くかかわっていく必要がある．

情報発信の欠如による漢方医学の存在の希薄化

中国は国策として中医学(TCM)の国際化をはかっている．2006年7月には科学技術部・衛生部・国家中医薬管理局が共同で，中医薬の現代化

表 1　ISO の 5 つのワーキング・グループ

ワーキンググループ	課題	議長国
WG1	Quality and safety of raw materials used in TCM	中国
WG2	Quality and safety of manufactured TCM products	ドイツ
WG3	Quality and safety of acupuncture needles	中国
WG4	Quality and safety of TCM medical devices other than acupuncture needles	韓国
WG5	Informatics of TCM	韓国

と国際化のための“中医薬国際科技合作企画綱要(2006-2020)[3]を公布し，国家戦略として行っている．国家中医薬管理局[4]には 70 人あまりの専従職員がおり国際合作部も存在し，中医学の国際化をはかっている．韓国も政府には伝統医学専門の部局があり，16 名の専従職員がいる．

このように国家戦略として伝統医学の国際化を推進している中韓に比べ，わが国には専従部門が存在しない．2012 年 4 月，中国の商務部，外交部，国家中医薬管理局など 14 部門が中医薬のサービス貿易の発展に関する意見を公表した．これによると中医学の輸出のために，貿易の発展を重視し，その後押しを国家をあげて行うことを鮮明にしたものである．日本ではとうていありえないが，中国においても 14 もの政府部門が揃って意見を提出することはきわめて異例であり，中医学の国際振興がいかに産業として大きいかを物語っている．そのなかには中医学を担う人材の育成や海外進出を明記してある．中国が中医学を広める戦略として国際中医師の資格がある．これはもともと中医師を自称する多くの無資格者を取り締まるための資格基準をつくるための資格認定試験であったが，国際中医師試験として発展し，日本においても 1996 年より毎年試験が実施されるようになっている．2004 年からは資格認定は国家中医薬管理局から世界中医薬学会連合会に移管されている．すでにこうした資格を認めて診療を許可している国も出始めている．わが国にも中医学大学の日本校があり，中医学を広めるために活動している．

日本では正規医療，しかし海外では補完・代替医療

漢方の世界でよくいわれるのが“漢方は補完・代替医療ではない．日本ではれっきとした正規医療だ．だから海外の補完・代替医療の学会にはいかない”と．しかし，そうであるのであれば，なおさら海外でこのことを宣伝すべきではなかろうか．海外に漢方の学会がない以上，欧米における補完・代替医療の学会に積極的に出向き，日本の漢方を堂々と主張すればいい．しかし，海外でのこうした学会で日本の研究者をみることはほとんどない．

一方，中国・韓国は大挙してそうした国際学会の場で発表する．海外に積極的に出向くことは情報を与えるのみならず，情報収集にも重要な機会なのであるが，日本はそうした努力を怠ってきたため，大きく世界の伝統医学の潮流から遅れを取っている．

日本からの情報発信を積極的に

中国のやり方は戦略性に富んでいて脅威に思えるが，彼らの認識は違う．日本でも中医学をやっているから中国の中医学国内標準を国際化すれば日本にもメリットがあるであろう，という考えである．そこで“日本の伝統医学はたしかに中国から伝来したが，日本に来て 1,500 年の間に独自の発展を遂げ，現代の中医学とは似て非なものである”という説明をすると驚かれる．それは我々からの情報発信が足りないせいである．

欧米でも然り．多くの人が TCM は知っているが，Kampo は知らないという．また，日本式の鍼管のついた鍼だけが FDA で認可されているため，多くの施術者たちが日本鍼を使っているが，“TCM acupunture”と称しているのである．これも明らかに日本からの情報発信が足りないためである．

このように中国やその他の諸国において日本の

漢方の存在を幅広く情報発信する必要がある．情報発信をすることで，情報収集も可能となるからである．

日中韓での協力体制の確立

WHOで日中韓の取りまとめを過去8年にわたり行ってきた経験から最後に述べたい[5-9]．

まずはたがいの理解を深めることである．中国は国内では中医学の権威は失墜してきており，海外に活路を見出そうとしている．また，韓国は西洋医学と韓医学との対立のなかで新しい道を模索している．

こうしたたがいの国の事情がわかってくると，助言をしあいながら，そうした問題をどのように克服したらいいかという知恵が湧き上がってくる．

まずは相互理解を深めることであろう．そのためには民間のみならず国家レベルでの交流も必要である．日中韓には保健大臣会議の枠組があり，伝統医学がひとつのトピックなのであるが，日本政府が対応しきれず進まない．

政府，民間を問わずいろいろな交流を推進するために，学会のみならず政府に専門組織が必要であり，わが国の国家戦略をしっかり定める必要がある．

おわりに

世界的な補完・代替医療の潮流により，伝統医学は否が応でも国際舞台に立たされることになった．そのようななかで日本漢方のアイデンティティーをどのように保っていくのか，また，何を売りにしていくのかについて真剣に考える時期に来ている．

なぜならばICDはじめ，国際的展開によって国内状況が影響を受けることは必至だからである．国際的に漢方をアピールすることは国内的に漢方を守っていくことにほかならない．わが国の文化として育んできた漢方医学が今後も継続して世代を超えて継承されていくためにも，重要な時期に来ているものと考える．

読者の方々のお力により漢方医学が永続的に発展していくことを請い願うものである．

文献

1) 中医薬学会世界連合会HP．http://www.wfcms.org/
2) ISO TC215 HP. http://www.iso.org/iso/iso_technical_committee?commid=54960
3) ISO TC249 HP. http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=598435
4) 中華人民共和国国家中医薬管理局HP．http://www.satcm.gov.cn/.
5) Cameron, S. et al.: Asian medicine：Japan's paradigm. Nature, 482：35, 2012.
6) 渡辺賢治：日本人が知らない漢方の力．祥伝社，2012.
7) 渡辺賢治：21世紀の日本の東洋医学の進路を探る漢方の国際化に向けての戦略．日本東洋医学雑誌，58：594-599，2007.
8) 渡辺賢治：世界で高まる伝統医療のニーズに後れをとる日本漢方．クリニックマガジン，486：28-33：2010.
9) 渡辺賢治：今こそ日本型医療の創生を．日本医事新報，4468：1，2010.

（執筆：渡辺賢治）

Symbolic Hierarchical Clustering for Visual Analogue Scale Data

Kotoe Katayama, Rui Yamaguchi, Seiya Imoto, Hideaki Tokunaga,
Yoshihiro Imazu, Keiko Matsuura, Kenji Watanabe, and Satoru Miyano

Abstract. We propose a hierarchical clustering in the framework of Symbolic Data Analysis(SDA). SDA was proposed by Diday at the end of the 1980s and is a new approach for analysing huge and complex data. In SDA, an observation is described by not only numerical values but also "higher-level units"; sets, intervals, distributions, etc. Most SDA works have dealt with only intervals as the descriptions. In this paper, we define "*pain distribution*" as new type data in SDA and propose a hierarchical clustering for this new type data.

Keywords: Visual Analogue Scale, Distribution-Valued Data.

1 Introduction

Conventional data analysis usually can handle scalars, vectors and matrices. However, lately, some datasets have grown beyond the framework of conventional data analysis. Most statistical methods do not have sufficient power to analyze these datasets. In this study, we attempted to extract useful information from such datasets.

Symbolic data analysis (SDA) proposed by Diday [3] is an approach for analyzing new types of datasets. "Symbolic data" consist of a *concept* that is described by intervals, distributions, etc. as well as by numerical values. The use of SDA enriches data description, and it can handle highly complex datasets. This implies that complex data can be formally handled in the framework of SDA. However, most SDA works have dealt with only intervals as the descriptions and are very few studies

Kotoe Katayama · Rui Yamaguchi · Seiya Imoto · Satoru Miyano
Human Genome Center, Institute of Medical Science, The University of Tokyo,
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
e-mail: k-kata@ims.u-tokyo.ac.jp

Hideaki Tokunaga · Yoshihiro Imazu · Keiko Matsuura · Kenji Watanabe
Center for Kampo Medicine, Keio University School of Medicine,
35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, SIST 10, pp. 799–805.
springerlink.com

based on this simple idea. The case that *concept* is described by intervals is simple, but ignores detailed information in the intervals. We propose distribution-valued data to describe the *concept*.

In this study, we focus on the case in which a *concept* is described by distribution and develop a new method to analyze this dataset directly using SDA.

2 The Visual Analogue Scale

The visual analogue scale (VAS) has developed to allow the measurement of individual's responses to physical stimuli, such as heat. The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. It was originally used in the field of psychometrics, and nowadays widely used to assess changes in patient health status with treatment.

A VAS consists of a line on a page with clearly defined end points, and normally a clearly identified scale between the two end points. For guidance, the phrase "no pain" and "worst imaginable pain" are placed at the both side of the line, respectively. Minimum value 0 of the VAS means "no pain" and maximum value 100 means "worst imaginable pain".

These scales are of most value when looking at change within patients, and are of less value for comparing across a group of patients because patient have a different sense of pain. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling such data. Many researchers prefer to use a method of analysis that is based on the rank ordering of scores rather than their exact values, to avoid reading too much into the precise VAS score.

3 Transform the Visual Analogue Scale to Distribution-Valued Data

We transform the VAS to distribution-valued data to compare across a group of patients. VAS varies according to patients, because sense of pain varies a great deal depending on people. Changing VAS score within patients means their sense of pain. If they have big change of VAS score, their expression of sense of pain is rough. On the contrary, if they have small change, their expression is sensitive. We suggest that these sense of pain is described by normal distribution and call it "*pain distribution*(PD)".

Let VAS score of patient's first time be x_1 and second time be x_2. We define the middle point of x_1 and x_2 as mean of PD μ, and $(\mu - x_1)^2 = (\mu - x_2)^2$ as variance. We describe PD as $N(\mu, \sigma^2)$. In case that the number of VAS score is d, PD is d-dimensional normal distribution. In this case, a diagonal matrix is used as a variance-covariance matrix of d-dimensional normal distribution.

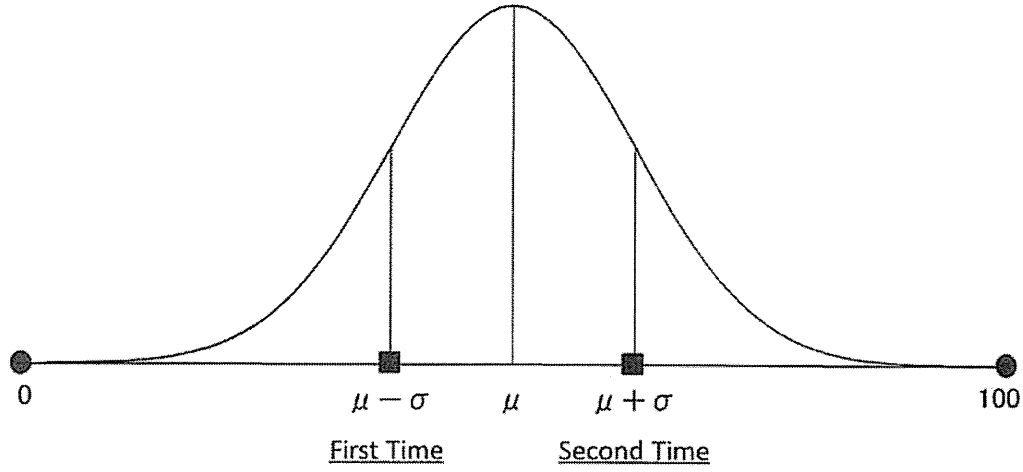


Fig. 1 Transform the Visual Analogue Scale to Distribution-Valued data

4 Hierarchical Clustering for PD

Cluster analysis groups data objects only on the bases of information found in the data that describes the objects and their relationships. The goal is that the objects within a group should be similar (or related) to one another and different from the objects in other groups.

In this section, we propose a hierarchical clustering for distribution-valued data, especially for PD.

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.

<Step1> Begin with K clusters, each containing only a single PD, $K = n$. Calculate distance between PD.

<Step2> Search the minimum distance in K clusters. Let the pair the selected 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 Step3, otherwise Step4.

<Step3> Calculate the distance between new cluster and other cluster, and go back to Step2.

<Step4> Draw the dendrogram.

Kullback-Leibler divergence is the natural way to define a distance measure between probability distributions [8], but not symmetry. We would like to use the symmetric Kullback-Leibler (symmetric KL) divergence as distance between *concepts*. The symmetric KL-divergence between two distributions s_1 and s_2 is

$$D(s_1(x), s_2(x)) = D(s_1(x)||s_2(x)) + D(s_2(x)||s_1(x))$$
$$= \int_{-\infty}^{\infty} s_1(x) \log \frac{s_1(x)}{s_2(x)} dx + \int_{-\infty}^{\infty} s_2(x) \log \frac{s_2(x)}{s_1(x)} dx, \quad (1)$$

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 PDs

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

$$D(p(x|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), p(x|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j))$$
$$= tr(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{-1}) + tr(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i^{-1}) + tr((\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T) - 2d. \quad (2)$$

Let PDs be $d = 1$,

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

After Step2, we need 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*[5].

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 $s_1(x) = \sum_{m=1}^{M} \omega_m^{(1)} p(x|\boldsymbol{\theta}_m^{(1)})$, where $\omega_m^{(1)}$ is a mixture weight, $p(x|\boldsymbol{\theta}_m^{(1)})$ 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 $s_2 = \sum_{l=1}^{L} \omega_n^{(2)} p(x|\boldsymbol{\theta}_l^{(2)})$.

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

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

where $o_{m,t}$ are sigma points. They are chose as follows:

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

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

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

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 the VAS Data

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 same data.

5.1 Medical Questionnaire in Keio University School of Medicine

Center 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 visual analogue scale 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 data set to PD. Next table show extracts taken from the original data and their translation.

Table 1 VAS value and PD

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	$N(23.5, 342.25)$

The result of our simulation show in figure2. 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.

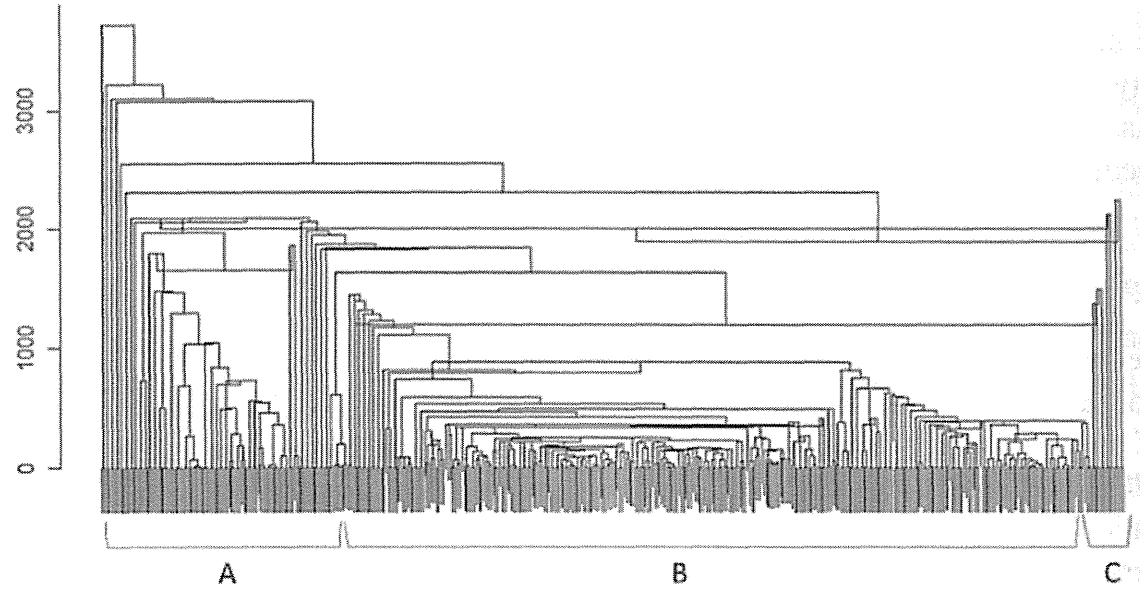


Fig. 2 Dendrogram for PDs

The result of centroid method show in figure3.

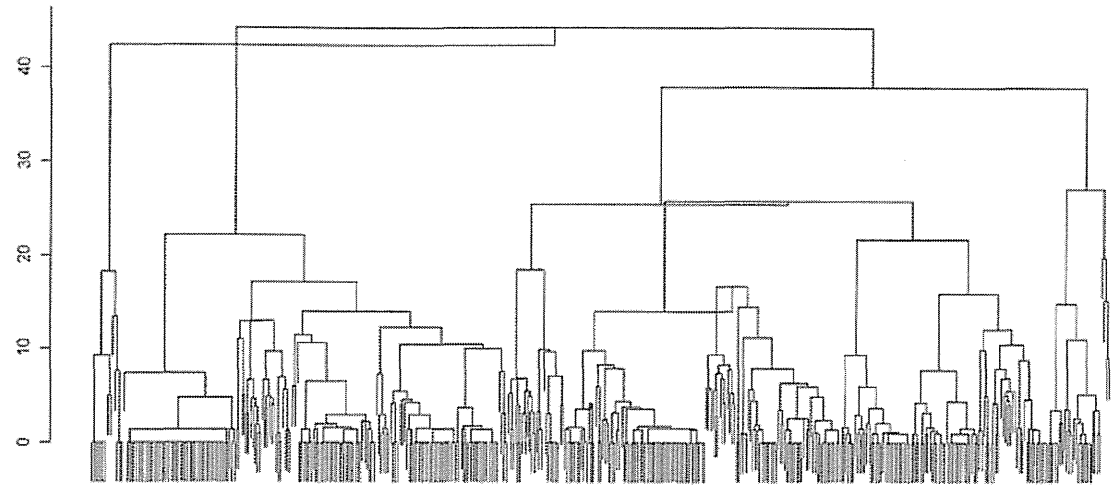


Fig. 3 Dendrogram of Traditional Method

6 Concluding Remarks

In this paper, we defined PD that is from transformation of the VAS to Distribution-Valued data. We also proposed hierarchical clustering method for it. 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.

In the future, we will define multidimensional PD and apply our clustering method.

References

1. Billard, L., Diday, E.: Symbolic Data Analysis. Wiley, NewYork (2006)
2. Bock, H.-H., Diday, E.: Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data. Springer, Berlin (2000)
3. Diday, E.: The symbolic approach in clustering and related methods of Data Analysis, Classification and Related Methods of Data Analysis. In: Bock, H. (ed.) Proc. IFCS, Aachen, Germany. North-Holland, Amsterdam (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. Kullback, S.: Information theory and statistics. Dover Publications, New York (1968)

Transform of visual analogue scale data and their clustering

Kotoe Katayama*, Rui Yamaguchi and Seiya Imoto

Human Genome Centre,
Institute of Medical Science,
The University of Tokyo,
4-6-1 Shirokanedai, Minato-ku,
Tokyo 108-8639, Japan
E-mail: k-kata@ims.u-tokyo.ac.jp
E-mail: ruiy@ims.u-tokyo.ac.jp
E-mail: imoto@ims.u-tokyo.ac.jp
*Corresponding author

Keiko Matsuura and Kenji Watanabe

Centre for Kampo Medicine,
Keio University School of Medicine,
35 Shinano-machi, Shinjuku-ku,
Tokyo 160-8582, Japan
E-mail: cherry@hop.ocn.ne.jp
E-mail: toyokeio@sc.itc.keio.ac.jp

Satoru Miyano

Human Genome Centre,
Institute of Medical Science,
The University of Tokyo,
4-6-1 Shirokanedai, Minato-ku,
Tokyo 108-8639, Japan
E-mail: miyano@ims.u-tokyo.ac.jp

Abstract: We propose a hierarchical clustering for the visual analogue scale (VAS) in the framework of symbolic data analysis (SDA). The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. VAS is of most value when looking at change within the same people, and is of less value for comparing across a group of people because they have different sense. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling VAS. We describe VAS as distribution and handle it as new type data in SDA. SDA was proposed by Diday at the end of the 1980s and is a new approach for analysing huge and complex data. In SDA, an observation is described by not only numerical values but also 'higher-level units'; sets, intervals, distributions, etc. In this paper, we define 'VAS distribution' and 'VAS changes distribution' as new type data in SDA and propose a hierarchical clustering for these new type data.

Keywords: visual analogue scale; VAS.

Reference to this paper should be made as follows: Katayama, K., Yamaguchi, R., Imoto, S., Matsuura, K., Watanabe, K. and Miyano, S. (2011) 'Transform of visual analogue scale data and their clustering', *Int. J. Knowledge Engineering and Soft Data Paradigms*, Vol. 3, No. 2, pp.143–151.

Biographical notes: Kotoe Katayama received her PhD in Information Science in 2010 from the Graduate School of Information Science and Technology, Hokkaido University. She is a Project Researcher in Human Genome Centre, Institute of Medical Science, The University of Tokyo. Her research interests include multivariate statistics, symbolic data analysis, statistical inference on genomescience and bioinformatics.

Rui Yamaguchi received his PhD in 2003 from the Graduate School of Science, Kyushu University. He is a Lecturer at Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo.

Seiya Imoto received his PhD in 2001 from the Graduate School of Mathematics, Kyushu University. He is an Associate Professor at Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo. Keiko Matsuura has MD and PhD. She is a Project Assistant Professor in Center for Kampo Medicien, Keio University School of Medicine.

Kenji Watanabe is a Kampo Medicine Specialist with an MD and PhD in Internal Medicine from Keio University School of Medicine. He is a Fellow of the Japanese Society of Internal Medicine, as well as a Fellow of American College of Phisicians.

Satoru Miyano is a Professor at Human Genome Center, Institute of Medical Science, The University of Tokyo. He received his BS, MS and PhD all in Mathematics from Kyushu University, Japan, in 1977, 1979 and 1984, respectively. He joined Human Genome Center in 1996.

This paper is a revised and expanded version of a paper entitled 'Symbolic hierarchical clustering for visual analogue scale data' presented at the 3rd International Conference on Intelligent Decision Technologies, (IDT 2011), University of Piraeus, 20–22 July 2011.

1 Introduction

The visual analogue scale (VAS) has been developed to allow physical stimuli, such as heat to the measurement of individual's responses to. The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. It was originally used in the field of psychometrics, and nowadays widely used to assess changes in patient health status with treatment. VAS is very useful to measure the changes in sensation within a patient, but it is difficult to compare more than one patients. Some researches tried to compare VAS among groups. They are based

on the premise of knowing the group (Dexter and Chestnut, 1995; Price et al., 1994). However it has not argued how we divide the groups.

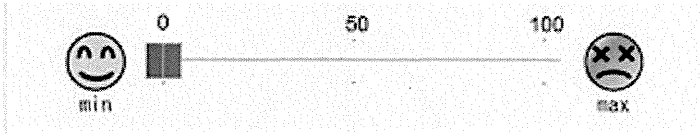
To compare VAS among groups, we transform VAS to 'symbolic data'. Symbolic data analysis (SDA) was proposed by Diday (Billard and Diday, 2006) is an approach for analysing new types of datasets. 'Symbolic data' consist of a *concept* that is described by intervals, distributions, etc., as well as by numerical values. The use of SDA enriches data description, and it can handle highly complex datasets. We propose distribution-valued data to describe the *concept*.

In this study, we define two new type of *concepts* and develop a new clustering method to them.

2 The VAS

A VAS consists of a line on a page with clearly defined end points, and normally a clearly identified scale between the two end points. For guidance, the phrase 'no pain' and 'worst imaginable pain' are placed at the both side of the line, respectively. Minimum value 0 of the VAS means 'no pain' and maximum value 100 means 'worst imaginable pain' (Figure 1).

Figure 1 Vas scale: minimum value 0 means 'no pain', maximum value 100 mean 'worst imaginable pain' (see online version for colours)



These scales are of most value when looking at change within patients, and are of less value for comparing across a group of patients because patient have a different sense of pain. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling such data. Many researchers prefer to use a method is based on the rank ordering of scores rather than their exact values, to avoid reading too much into the precise VAS score.

3 New types of *concepts*; 'PD' and 'PCD'

We define two types of *concepts*. First, we focus on sense of pain which patient naturally has. We describe the sense which patient oneself has as distribution. We name it *patient distribution* (PD). We also focus on the changes of pain within a patient and name it *patient changes distribution* (PCD).

3.1 Transform the VAS into 'PD'

We transform the VAS to 'PD'. VAS varies according to patients, because sense of pain varies a great deal depending on people. A change of VAS score within patients means their sense of pain. If they have big change of VAS score, their expression of

sense of pain is rough. On the contrary, if they have small change, their expression is sensitive. We suggest that these sense of pain is described by normal distribution and call it 'patient distribution (PD)'.

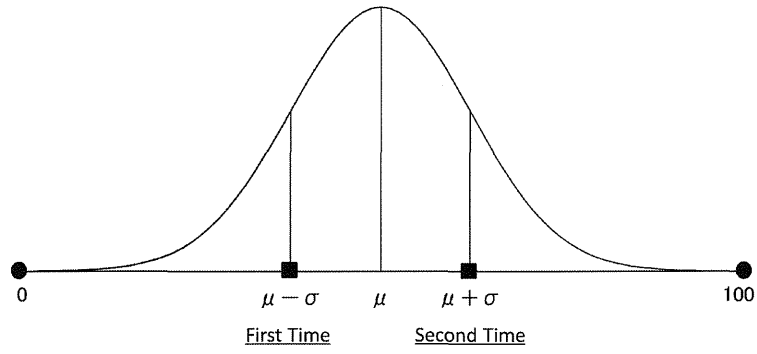
Let VAS score of patient's first time be x_1 and second time be x_2. We define the middle point of x_1 and x_2 as mean of PD μ, and $(\mu - x_1)^2 = (\mu - x_2)^2$ as variance. We describe PD as $N(\mu, \sigma^2)$. In case that the number of VAS score is d, PD is d-dimensional normal distribution. In this case, a diagonal matrix is used as a variance-covariance matrix of d-dimensional normal distribution.

3.2 Transform the VAS into 'PCD'

We transform the VAS to 'PCD' by using PD. We suggest that PCD is described by mixture normal distribution.

We set score of patient's first time x_1 be mean of PD as new distribution of first time (PD_1). New distribution of second time is defined in a similar way(PD_2). By combining PD_1 and PD_2, we get mixture distribution, where set mixture weight for PD_1 as 0.6 and for PD_2 as 0.4. This weight is based on doctors' opinions that first time VAS is more important than others. It is, finally, PCD. Figure 2 shows the case that x_1 is 35 and x_2 is 65.

Figure 2 Transform the VAS to distribution-valued data

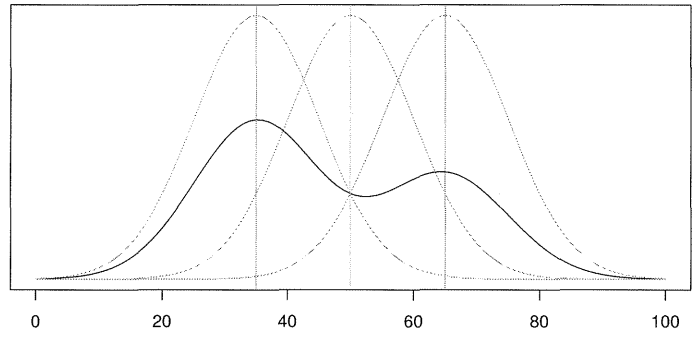


4 Hierarchical clustering for PD and PCD

Cluster analysis groups data objects only on the bases of information found in the data that describes the objects and their relationships. The goal is that the objects within a group should be similar (or related) to one another and different from the objects in other groups.

In this section, we propose a hierarchical clustering for distribution-valued data (PD and PCD).

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)

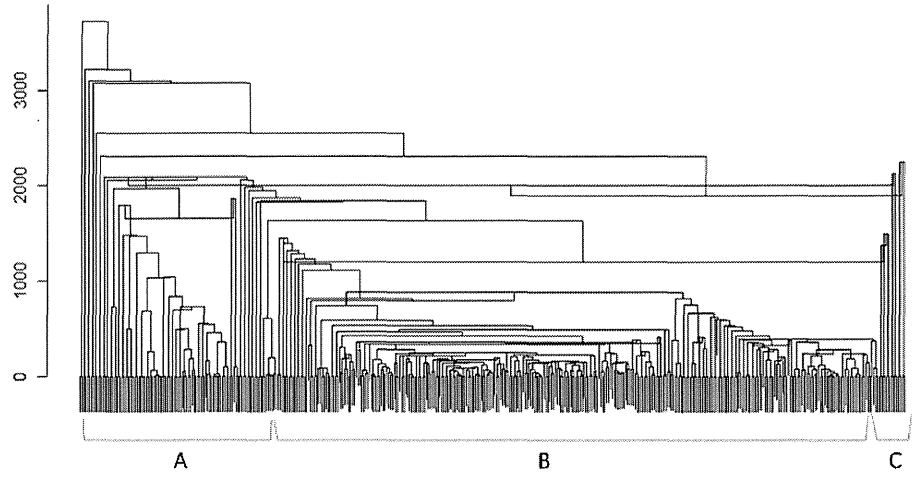
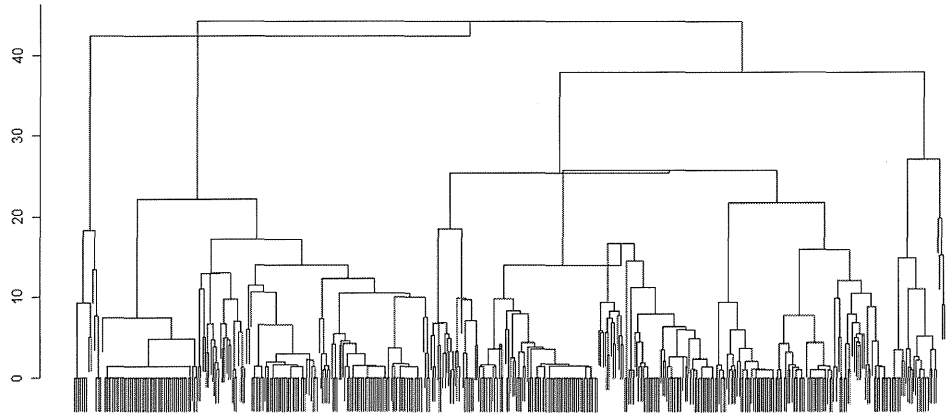


Figure 5 Dendrogram of traditional method



6 An application to PCD

In this section, we show the case of PCD.

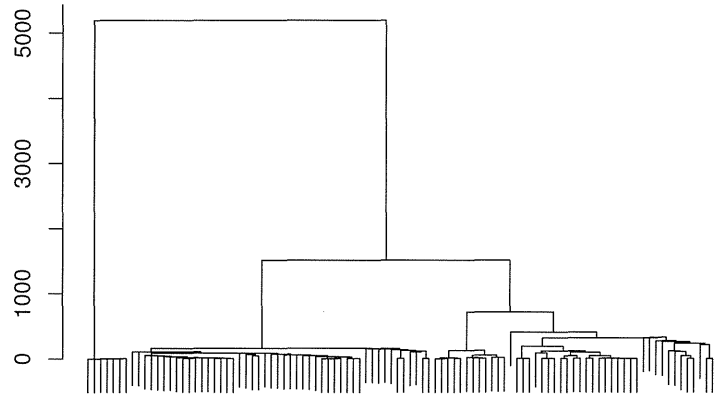
6.1 Data description and result

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.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 6 (2011) 370–374

Procedia Computer Science

Complex Adaptive Systems, Volume 1
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2011- Chicago, IL

Clustering for Visual Analogue Scale Data in Symbolic Data Analysis

Kotoe Katayama[a*], Rui Yamaguchi[a], Seiya Imoto[a], Keiko Matsuura[b], Kenji Watanabe[b], Satoru Miyano[a]

[a]*Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[b]*Center for Kampo Medicine, Keio University School of Medicine, 35 Shinano-machi, Shinjuku-ku, Tokyo 160-8582, Japan*

Abstract

We propose a hierarchical clustering for the visual analogue scale (VAS) in the framework of Symbolic Data Analysis(SDA). The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. VAS is of most value when looking at change within people, and is of less value for comparing across a group of people because they have different sense. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling VAS. We describe VAS as distribution and handle it as new type data in SDA.
In this paper, we define "VAS distribution" as new type data in SDA and propose a hierarchical clustering for this new type data.

Keywords hierarchical clustering; Distribution Valued Data

1. **Introduction**

The visual analogue scale (VAS) has developed to allow the measurement of individual's responses to physical stimuli, such as heat. The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. It was originally used in the field of psychometrics, and nowadays widely used to assess changes in patient health status with treatment. VAS is very useful to measure the changes in sensation within a patient, but it is difficult to compare more than one patients. Some researches tried to compare VAS among groups. They based on the premise of knowing the group[3],[9]. It has not argued how we

* Corresponding Author
Email address: k-kata@ims.u-tokyo.jp

1877–0509 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.procs.2011.08.068