DOI:10.13374/j.issn1001-053x.2008.07.049
第30卷 第7期 北京科技大学学报 Vol.30 No.7
2008年 7月 Journal of University of Science and Technology Beijing Jul. 2008

基于结构模型的知识发现技术

范玉妹[1)] 单 平[2)] 艾冬梅[1)] 张德政[2)]

1) 北京科技大学应用科学学院，北京 100083 2) 北京科技大学信息工程学院，北京 100083

摘 要 将一种结构建模方法用于分析知识结构中概念之间的关系，并将其用于中医医案解析，有效地获取医案中隐含的深层次的辨证论治规律. 医案实例分析结果表明，由结构分析所得到的结构模型，不仅可根据系统中少量的、零散的已知概念关系推导出其他的大多数未知的概念之间的关系，并可使得已有结构关系凸现出来. 结构的图分析可展现症状的关联度和重要程度，分析结果得到了医案验证和专家肯定.

关键词 系统分析；知识获取；结构建模；中医

分类号 N945.1

Knowledge discovery based on structural modeling

FAN Yumei[1)], *SHAN Ping*[2)], *AI Dongmei*[1)], *ZHANG Dezheng*[2)]

1) School of Applied Science, University of Science and Technology Beijing, Beijing 100083, China

2) School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China

ABSTRACT A structure model was used to analyze the relations between concepts in knowledge structure and parse the medical records of traditional Chinese medicine to gain connotative rules of treatment based on syndrome differentiation effectively. The analytic results of medical record examples show that the structure model gained by structural analysis can not only discover many unknown relations by few scattered existing information, but also protrude existing structural relations. The picture analysis of structure displays the correlation degree and importance of symptoms, and the results have been validated by medical records and affirmed by experts.

KEY WORDS system analysis; knowledge acquisition; structural modeling; traditional Chinese medicine

名老中医学术思想与临床经验是通过传承、实践以及创新而形成的独特的知识体系，知识隐含在名老中医辨证施治过程以及所形成的医案之中. 有效地获取医案中的知识[1]，深层次挖掘隐藏在诊疗过程中的隐性知识，最大限度地获取与保留名医数十年积累的诊疗经验，是实现中医传承的关键. 名老中医的知识体系大多由技巧、经验和过程等隐性知识为主，包含辨证论治、理法方药等多个层次，知识结构复杂多变. 已有数据挖掘技术，对于非结构化、知识关系复杂的中医知识挖掘与获取难以奏效. 中医知识体系依附于人体生理与病理过程，是一个复杂的系统，其复杂性不仅是处方用药规律的复杂多变，更为突出的是表现为理法方药内在联系的多层次多维度. 从知识构建的角度来看，复杂的知识结构可以利用系统分析方法来探索知识结构中概念之间的关系[2]. 结构分析结果体现在结构模型中，该模型代表着认知结果或复杂知识结构的显化. 结构模型描述了系统概念之间的关系，记录了人们对系统的定性的认识，同时在认识过程中它又是激发进一步认识的媒介. 结构分析就是一个分析过程、学习过程，更是一个知识获取的过程.

中医医案分析就是一个关于某种疾病知识构建的过程，通过知识构建来再现名老中医临床施治情景，探索辨证思维过程，进而获取疾病诊疗的知识. 结构建模[3]是通过人机交互来完成认知获取知识的过程，通过分解、列举、集结、结构化、扩大、分类与

收稿日期：2007-06-18 **修回日期**：2007-08-18

基金项目：国家"十五"科技攻关项目(No.2004BA721A01H07)；北京市自然科学基金资助项目(No.4062022)

作者简介：范玉妹(1948—)，女，教授，E-mail：lsffym@vip.sina.com

抽象以及分组等过程将隐性的思维过程转化为显性知识构建和知识获取过程. 结构建模的分析过程在原理上与医案解读有着相似之处. 因此，结构建模分析可以作为医案知识获取方法用于医案解读分析. 根据提供的部分医案内部知识节点的关系，运用结构模型本身的分析逻辑，通过推理求出其他未知知识节点之间关系，不断地重复这个过程可以根据认知规律利用人机交互过程，把人脑中隐含的中医知识结构模型逐步地引导出来.

本文将结构建模分析技术用于我国著名肝病中医专家钱英教授诊治的肝病医案分析，利用结构模型分析技术进行隐性知识的获取的实验研究，以验证该方法的有效性[4].

1　结构建模

隐含的、具有复杂结构的中医理论体系显化于知识结构之中就构成了中医知识结构模型. 结构建模的过程是结构分析的过程，这个过程把主观世界中不可见的分析过程以及思维建构过程变成了计算机世界中可见的形象分析过程和形象构建过程，它也是一个创新的过程，利用人的知识和已有的实例，通过交互来启发[5]. 在中医知识模型中，模型的节点是由知识结构中的概念组成，对应于中医药理论中的证、症、因、机以及方药等常用中医诊疗术语. 本文采用基于核心要素的结构建模方法[6]，其基本原理是在已知系统的一个初始关系矩阵的前提下，能根据少量已知关系推导出其他的绝大多数未知关系，从而建立起系统的结构模型. 本文对其算法进行了简化并予以实现，提出了相应的结构模型建模与知识获取方法. 医案来自国家科技攻关课题“基于信息挖掘技术的名老中医临床诊疗经验及传承方法研究”综合数据库[7]. 结构建模所涉及的初始关系可由大量医案和中医专家给出.

如果以领域知识中的一种和一类知识作为概念，以知识之间的关联作为概念之间的关系，则领域知识实质上形成了一个知识系统[8]. 系统可以表示为(S, M_1). 其中 $S=\{s_1, s_2, \cdots, s_n\}$表示具有不同内涵的知识的集合，称之为知识点或概念，$M_1=\{(s_i, s_j)\}$表示不同概念之间关系的集合. 系统 S 的关系矩阵 $\boldsymbol{M}=[m_{ij}]$. $m_{ij}=1$ 表明概念 i 可达概念 j，并且可达具有传递性. 对任一概念 s_i，其他概念 s_j $(i\neq j)$属于 s_i 的下列集合之一[9-10]：(1) s_i 可到达一些概念，即 s_i 要影响它们，这些概念构成 s_i 的“上位集”，此时又细分为有反馈上位集 $F(s_i)$和无反馈上位集 $\mathrm{NF}(s_i)$，反馈上位集 $F(s_i)$的概念亦可达 s_i，而无反馈上位集 $\mathrm{NF}(s_i)$中的概念不可达 s_i. (2) 有一些概念可到达 s_i，即影响 s_i，这些概念构成 s_i 的“下位集”$D(s_i)$. (3) 有一些概念既不被 s_i 影响，也不影响 s_i，这些概念构成 s_i 的“无关集”$V(s_i)$. (4) 有一些概念与 s_i 的关系不清楚，这些概念构成 s_i 的“无知集”$\mathrm{UNK}(s_i)$. 各部分关系如图 1 所示.

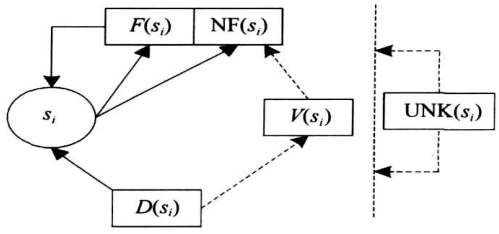

图 1　各部分的关系

Fig·1　Relation of all parts

$D(s_i)$中的概念都可达 s_i，s_i 可达 $\mathrm{NF}(s_i)$中所有的概念，s_i 与 $F(s_i)$中的概念是相互可达的，s_i 到 $D(s_i)$中的概念一定是不可达的，否则这个概念应该属于 $F(s_i)$，$\mathrm{NF}(s_i)$到 s_i 一定是不可达的，否则这个概念也应该属于 $F(s_i)$. 同时，由于可达具有传递性，$D(s_i)$中的概念可达 $F(s_i)$和 $\mathrm{NF}(s_i)$中的概念 $(D(s_i)\rightarrow s_i\rightarrow F(s_i)/\mathrm{NF}(s_i))$，同理，$F(s_i)$中的概念可达 $\mathrm{NF}(s_i)$中的概念$(F(s_i)\rightarrow s_i\rightarrow \mathrm{NF}(s_i))$，$F(s_i)$中的概念自身可达$(F(s_i)\rightarrow s_i\rightarrow F(s_i))$.

取与其他概念关系最多的那个概念为核心概念，记为 s_{ker}. 核心概念初始有反馈上位集$F^0(s_{\mathrm{ker}})$，初始无反馈上位集 $\mathrm{NF}^0(s_{\mathrm{ker}})$，初始下位集$D^0(s_{\mathrm{ker}})$和初始无知集 $\mathrm{UNK}^0(s_{\mathrm{ker}})$的关系如图 2 所示.

	NF^0	F^0	s_{ker}	UNK^0	D^0
NF^0	$M_{\mathrm{NF,NF}}$	$M_{F,\mathrm{NF}}$	0	$M_{\mathrm{NF,UNK}}$	$M_{\mathrm{NF},D}$
F^0	1	1	1	$M_{F,\mathrm{UNK}}$	$M_{F,D}$
s_{ker}	1	1	1		0
UNK^0	$M_{\mathrm{UNK,NF}}$	$M_{\mathrm{UNK},F}$		$M_{\mathrm{UNK,UNK}}$	$M_{\mathrm{UNK},D}$
D^0	1	1	1	$M_{D,\mathrm{UNK}}$	$M_{D,D}$

图 2　四个部分的划分

Fig·2　Partition of four parts

图中 1 为可达，0 为不可达，其中加重边框的是三个部分的内部关系，重要的是着色的九个子矩阵，其中浅灰色的是无知集中的概念和其他部分的关系，其余三个是深灰色的.

对于浅灰色的部分，无知集中的概念和其他部分的关系依据以下变换规则：如果 $m_{nu}=1$ 或 $m_{fu}=1$，那么 $s_u\in \mathrm{NF}$，$m_{\mathrm{ker}\,u}=1$，$m_{u\mathrm{ker}}=0$ (即 $m_{nu}=1$表明 NF 中有一概念可达 UNK 的一个

概念 s_u，于是有 $s_{\ker}\rightarrow \mathrm{NF}(s_{\ker})\rightarrow s_u$，核心概念 $s_{\ker}$可达概念 s_u，所以无知集中的 $s_u\in \mathrm{NF}$，$m_{\ker u}=1$，$m_{u\ker}=0$).

同理，如果 $m_{uf}=1$ 或 $m_{ud}=1$，那么 $s_u\in D$，$m_{\ker u}=1$，$m_{u\ker}=0$；如果 $m_{un}=1$ 或 $m_{du}=1$，而且 $m_{nu}\leqslant 0$，$m_{fu}\leqslant 0$，$m_{uf}\leqslant 0$，$m_{ud}\leqslant 0$，那么 $s_u\in V$.

对于深灰色的部分：如果 $m_{nd}=1$，那么 $s_n\in F$，$s_d\in F$，$m_{u\ker}=1$，$m_{\ker d}=1$；如果 $m_{nf}=1$，那么 $s_n\in F$，$m_{u\ker}=1$；如果 $m_{fd}=1$，那么 $s_d\in F$，$m_{\ker d}=1$.

通过上述变换，初始矩阵中的 1 已全部利用，结果矩阵中的绝大部分值都已确定，仍未知的可分为两部分. 第 1 部分包括 $\boldsymbol{M}_{\mathrm{NF,NF}}$，$\boldsymbol{M}_{V,V}$ 和 $\boldsymbol{M}_{D,D}$，子矩阵中未知概念的消除采用与系统整体关系矩阵同样的方法进行处理；第 2 部分包括 $\boldsymbol{M}_{V,\mathrm{NF}}$ 和 $\boldsymbol{M}_{D,V}$，它们所在的下面两个矩阵

$$\begin{bmatrix} \boldsymbol{M}_{\mathrm{NF,NF}} & \boldsymbol{M}_{\mathrm{NF},V} \\ \boldsymbol{M}_{V,\mathrm{NF}} & \boldsymbol{M}_{V,V} \end{bmatrix},\begin{bmatrix} \boldsymbol{M}_{V,V} & \boldsymbol{M}_{V,D} \\ \boldsymbol{M}_{D,V} & \boldsymbol{M}_{D,D} \end{bmatrix}$$

具有同样的结构

$$\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{X} & \boldsymbol{B} \end{bmatrix}.$$

由于它是可达矩阵，所以

$$\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{X} & \boldsymbol{B} \end{bmatrix}\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{X} & \boldsymbol{B} \end{bmatrix}=\begin{bmatrix} \boldsymbol{A}^2 & 0 \\ \boldsymbol{XA}+\boldsymbol{BX} & \boldsymbol{B} \end{bmatrix}=\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{X} & \boldsymbol{B} \end{bmatrix} \tag{1}$$

则得 $\boldsymbol{XA}+\boldsymbol{BX}=\boldsymbol{X}$，这就是自蕴含方程，需要人机交互求解 $\boldsymbol{X}$.

在建模过程中涉及的人机交互信息，既可采用专家的意见来补充概念间的联系，也可利用中医常识知识库来补充所需概念间的关系. 这种建模方法可以发挥已有知识和计算机的优势，将专家头脑中隐含的知识通过交互来逐步获取，把人和计算机单独都不能完成的工作交给人机结合系统，发挥系统功能的优势.

2 结构模型算法

利用初始矩阵中已有的数字 1，按照变换规则尽可能地确定矩阵中的数字 -1，辅之以人机交互，从而获得系统的可达矩阵.

步骤 1　生成系统概念. 通过专家医案分析，已知的相关症状、病因机分析和诊断概念生成系统的概念 $S=\{s_1,s_2,\cdots,s_n\}$.

步骤 2　建立初始矩阵. 初始关系矩阵的建立是基于中医常识知识库，或通过人机交互由中医专家输入中医概念或术语之间的关系，“知道什么，输入什么”. 记初始矩阵

$$\boldsymbol{M}=[\mathrm{m}_{ij}],\ \mathrm{m}_{ij}=\begin{cases} 1 & \text{概念 } i \text{ 可达概念 } j \\ -1 & \text{否则} \end{cases}$$

步骤 3　求核心概念. 核心概念取其他概念到达该概念和由该概念出发到达其他概念最多的那个概念，即矩阵中概念所在行和列中数值 1 最多的那个概念.

步骤 4　把除核心概念之外的全部其他系统概念分为四个部分 NF^0、F^0、UNK^0 和 D^0，如图 2 所示. 未定的有 12 个子矩阵，其他部分由定义和可达性均已确定为 0 或 1.

步骤 5　用矩阵中已有的 1，按照变换规则尽可能地确定未知的部分. 比如 $\boldsymbol{M}_{\mathrm{NF},U}$ 中有 $m_{ij}=1$，则根据变换规则，概念 j 应属于 MF. 如 $\boldsymbol{M}_{\mathrm{NF},D}$ 中有 $m_{ij}=1$，则概念 i 和概念 j 都应属于 F. 每个子阵的变化直到其中所有的概念小于等于 0 时为止.

步骤 6　经过上述变换，结果矩阵中判断子阵 $\boldsymbol{M}_{\mathrm{NF,NF}}$、$\boldsymbol{M}_{V,V}$ 和 $\boldsymbol{M}_{D,D}$ 是否还有 -1. 若有则采用与系统整体关系矩阵同样的方法处理，即把它们各自看成一个小系统，返回步骤 3.

步骤 7　判断子阵 $\boldsymbol{M}_{V,\mathrm{NF}}$ 和 $\boldsymbol{M}_{D,V}$ 中是否还有 -1. 若有则需要通过求解自蕴含方程来确定其值，此过程需要人机交互获得一些初始值.

步骤 8　至此能获得值的概念均已获得. 如若还有 -1，则用人机交互的方式消除，转至步骤 5，反复迭代直至最终得到系统的可达矩阵.

3 基于结构模型的知识获取

根据上述算法就我国中医肝病专家钱英教授诊断肝病医案进行了分析. 在所分析的病例中病人主要症状为脉沉细、舌质淡、苔白厚、舌下静脉粗、手末梢暗和眠差. 钱英教授认为肝藏血，主疏泄，达阳气于四末，慢性肝病患者，常有痰、瘀阻于肝络，出现手背末梢发暗. 舌下静脉曲张、增粗亦往往为肝络不通之表现. 人体为统一的整体，有诸内必形诸外，体内血液循环受阻亦必形之于外. 验之临床，从西医角度凡有以上指征者往往伴有明显肝纤维化或早期肝硬化. 利用中医知识库对医案的分析，得到了医案中主要的已知关系[11]，并给出了与症状相关的主要病因机概念，如图 3 所示.

图 3 中凡是有箭头连接的两个症状视为有关系[12]，设为 1，且关系是有方向的. 比如虚证有箭头指向气血两虚，则二者有关系，值为 1，关系是由虚证到气血两虚的. 其他没有箭头相连的均设为 -1，表示关系未定. 在进行结构建模分析以前，已知的

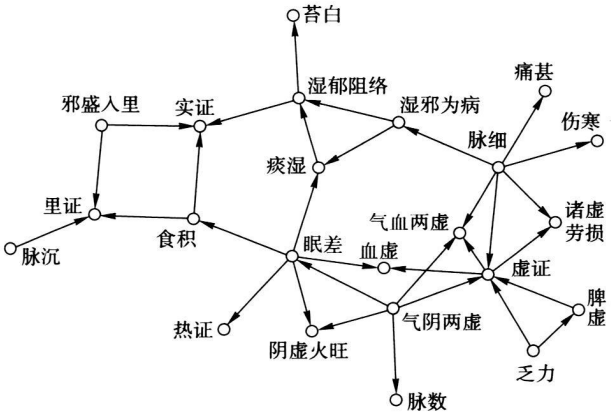

图 3　初始关系图

Fig.3　Initial relations

相关症状、病因机分析和诊断概念包括乏力、里证、脉沉、脉数、脉细、眠差、脾虚、气血两虚、气阴两虚、热证、伤寒、食积、湿邪为病、湿郁阻络、实证、苔白、痰湿、痛甚、邪盛入里、血虚、虚证、阴虚火旺和诸虚劳损，设其索引为 0～22. 医案按语给出病机分析与治疗是痰湿致肝络不通、湿郁阻络、气阴两虚，其中以气阴两虚为病机的主要方面，凡有以上依据者，疏通肝络为重要治法，用益气养阴、化湿通络之法治疗.

经过程序运算[13]，可确定核心概念为虚证. $F^0(s_{ker})$为空；$NF^0(s_{ker})$：7, 19, 22；$D^0(s_{ker})$：0, 4, 6, 8；$UNK^0(s_{ker})$：1, 2, 3, 5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21. 经过变换之后的结果关系图，如图 4.

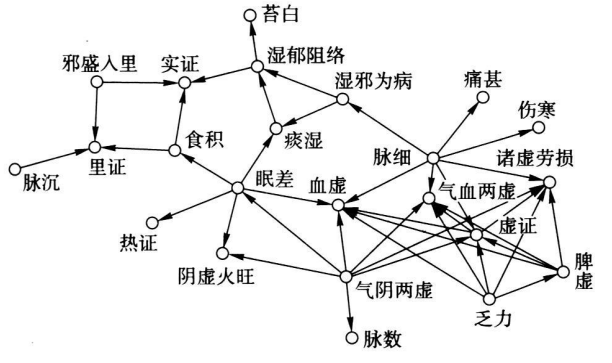

图 4　结果关系图

Fig.4　Result relations

比较图 3 和图 4 可以看出：血虚多了四个入度，即增加了关系：气阴两虚→血虚，乏力→血虚，脾虚→血虚，脉细→血虚. 诸虚劳损多了三个入度，增加的关系为：气阴两虚→诸虚劳损，乏力→诸虚劳损，脾虚→诸虚劳损. 气血两虚也增加了两个入度，增加的关系为：脾虚→气血两虚，乏力→气血两虚. 脾虚增加了三个出度，即脾虚→血虚，脾虚→诸虚劳损，脾虚→气血两虚. 乏力也多了两个出度：乏力→血虚，乏力→气血两虚. 此外还增加了关系：脉细→血虚，气阴两虚→诸虚劳损.

可见，结果图提高了血虚、诸虚劳损、脾虚、气血两虚和乏力的入度. 依据结构建模分析结果可以看出，本医案以气阴两虚、脉细、虚证、血虚四个概念出入度最大，与医案所定气阴两虚为病机的主要方面的结论一致. 由此可见，基于结构建模的知识获取不仅能得到已有系统中未知的、不确切的知识，并可以很直观地看出新增加的关系，所得到的新的概念关系经专家论证符合中医理论. 与此同时，结构建模分析所得到的新的概念联系为理解医案和分析名老中医诊断思路提供了指导.

4　结语

中医知识许多来自经验，不同知识之间又形成了纵横交错的关联关系，利用结构模型能根据少量的已知关系推导出其他的绝大多数未知关系，并可以很直观地看出新增加的关系. 本文的理论与方法不仅限于中医领域，在其他领域也有好的应用前景. 以领域中的一种或一类知识作为一个概念，以知识之间的关联作为概念之间的关系，则可将其作为一个知识系统进行分析推理.

参　考　文　献

[1] Helene A F, Xavier G F. Working memory and acquisition of implicit knowledge by imagery training, without actual task performance. *Neuroscience*, 2006, 139(1): 401

[2] James D, Il-Yeol S, Ioanna L. An analysis of structural validity in entity-relationship modeling. *Data Knowl Eng*, 2003, 47(2): 167

[3] Zhang H J, Wu Y J, Song L L. System method for structural modeling. *J Heilongjiang Inst Technol*, 2006, 20(2): 68
(张海君，武玉婧，宋丽丽. 系统结构模型的生成. 黑龙江工程学院学报，2006，20(2)：68)

[4] Yee L, Wu W Z, Zhang W X. Knowledge acquisition in incomplete information systems: a rough set approach European. *J Oper Res*, 2006, 168(1): 164

[5] Katharina M, Michael I, Peter B, et al. Knowledge discovery and knowledge validation in intensive care. *Artif Intell Med*, 2000, 19(3): 225

[6] Su C T, Chen L S, Yuehwern Y. Knowledge acquisition through information granulation for imbalanced data. *Expert Syst Appl*, 2006, 31(3): 531

[7] Dang Y Z, Wang Z T. A kernel element transposition method for structural modeling in systems analysis. *J Syst Eng*, 1997, 12(4): 1
(党延忠，王众托. 系统分析中结构建模的核心变换法. 系统工程学报，1997，12(4)：1)

[8] Yang B R, Zhou Y. Inner mechanisms' research of knowledge discovery system. *J Univ Sci Technol Beijing*, 2002, 24(2): 345
(杨炳儒,周颖. 知识发现系统内在机理.北京科技大学学报, 2002,24(2):345)

[9] Azuma O, Ikuo K. Correction procedures for flexible interpretive structural modeling. *IEEE Trans Syst Man Cybern*, 1989, 19 (1): 85

[10] Wang Y L. *System Engineering*. Beijing: China Machine Press, 2003: 76
(汪应洛.系统工程.北京:机械工业出版社,2003:76)

[11] Zhang B, Gong J H, He C Z. OSA-based interpretative structural modeling. *Syst Eng Electron*, 2005, 27(3): 453
(张宾,龚俊华,贺昌政.基于客观系统分析的解释结构模型. 系统工程与电子技术,2005 27(3):453)

[12] Emmanuel J, Christian D. A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy. *Comput Stat Data Anal*, 2007, 51(8): 3666

[13] Wang Y H, Zhang S K, Liu Y, et al. Ripple-effect analysis of software architecture evolution based on reachability matrix. *J Software*, 2004, 15(8): 1107
(王映辉,张世琨,刘瑜, 等.基于可达矩阵的软件体系结构演化波及效应分析.软件学报,2004, 15(8):1107)

(上接第 825 页)

[12] Ma Q G, Li A. E commerce and firm informationalization: empirical study on the effects of organizational learning. *J Ind Eng Eng Manage*, 2004,18(2):11
(马庆国,李艾. 电子商务与企业信息化组织学习效应实证研究. 管理工程学报,2004,18(2):11)

[13] Day, George S. The capabilities of market-driven organizations. *J Marketing*, 1994, (10): 37

[14] Bell M, Pavitt K. The development of technological capabilities // *Trade, Technology, and International Competitiveness*. Working Paper, 1995

[15] Dess G G. Measuring organizational performance in the absence of objective measures: the case of the privately held firm and conglomerate business unit. *Strategic Manage J*, 1984, 5: 265

[16] Huber G P. Organizational learning: the contributing processes and the literatures. *Organ Sci*, 1991, 9: 88

[17] Crossan M, Lane H, White R. An organizational learning framework: from intuition to institution. *Acad Manage Rev*, 1999, 24(3): 522

[18] Edmondson A. The view through a different lens: investigating organizational learning at the group level of analysis // *Proceedings of 3rd International Coherence on Organizational Learning*. Lancaster, 1999, 299

[19] Santons. *On the Management of Knowledge: from the Transparency of Co-location and Cosetting to the Quandary of Dispersion and Differentiation*. Working Paper, 1997

[20] Tsai C T. *Organizational Factors, Creativity of Organizational Members and Organizational Innovation*. [Dissertation]. Taipei: Taiwan University, 1997
(蔡启通. 组织因素、组织成员整体创造性与组织创新之关系 [学位论文]. 台北:台湾大学,1997)

[21] Goh S, Richards G. Benchmarking the learning capability of organizations. *Eur Manage J*, 1997, 15(5):575