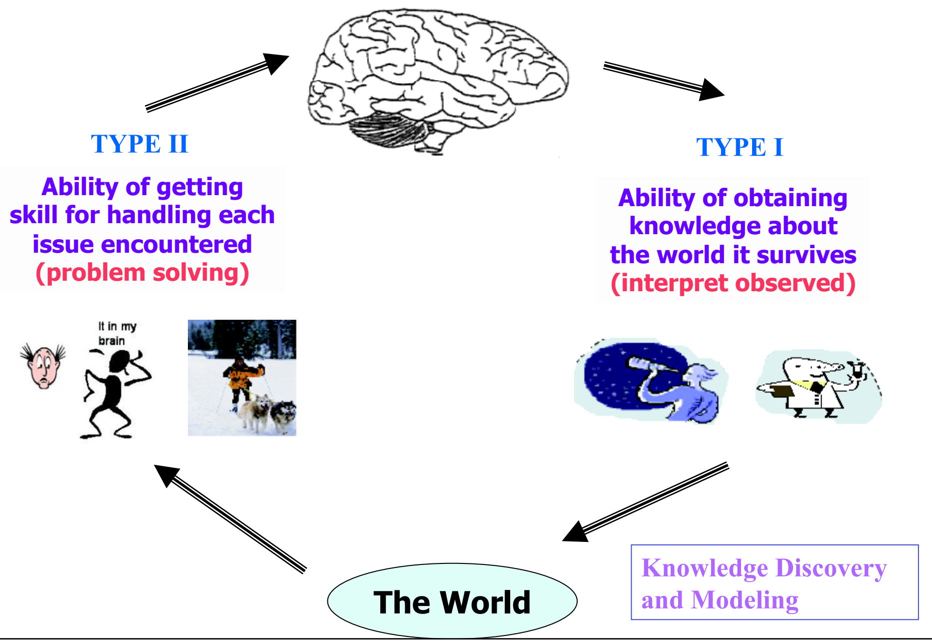
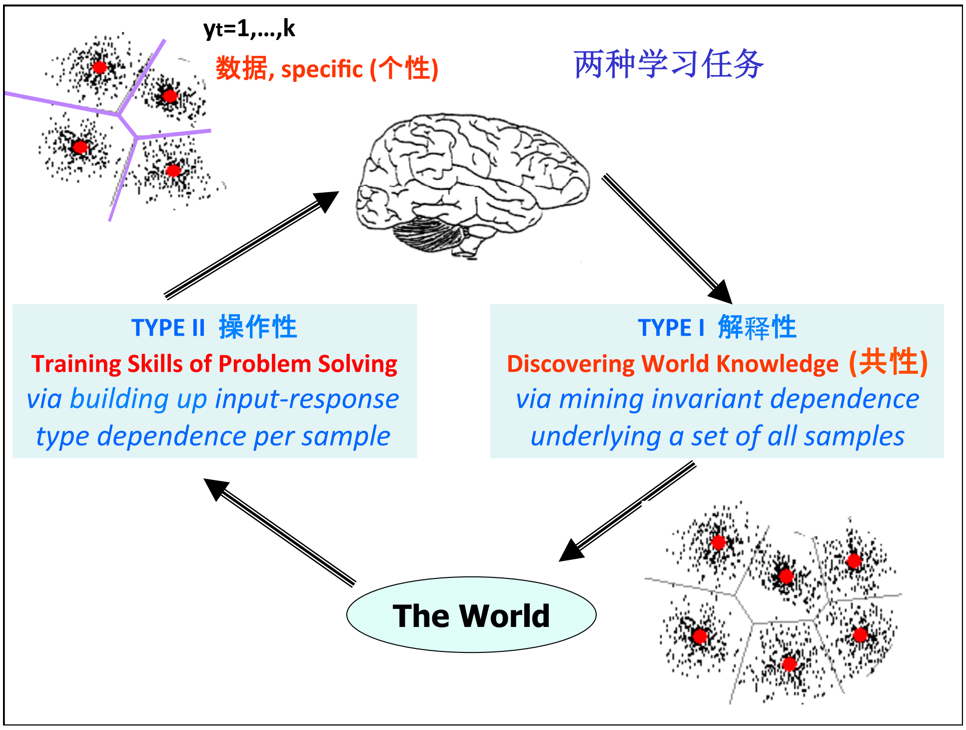
BYY和谐学习理论的新进展

徐雷 Lei Xu
http://www.cse.cuhk.edu.hk/~lxu/
Department of Computer Science and Engineering
The Chinese University of Hong Kong

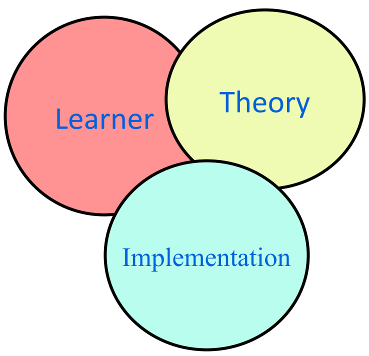
两种基本智能



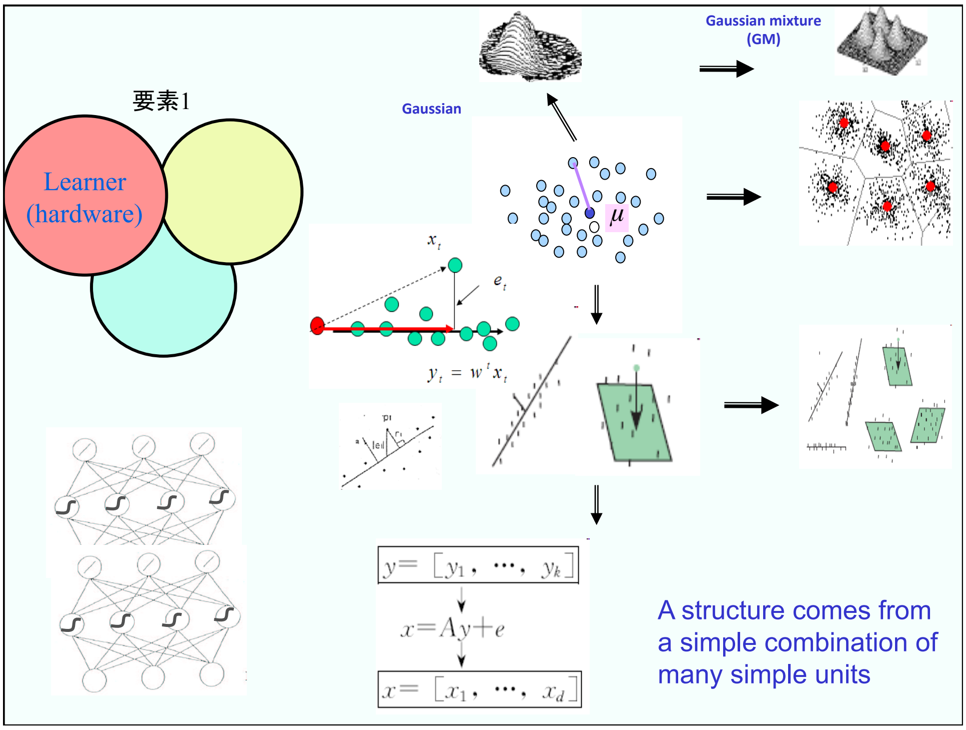


Statistical learning : three Ingredients and one key challenge

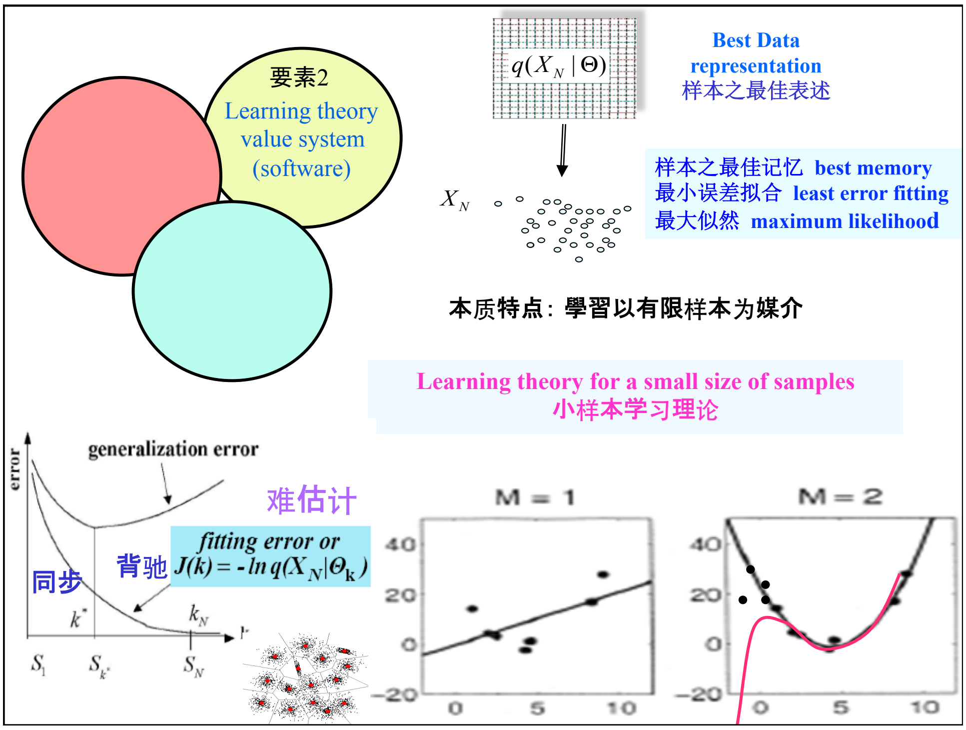
三个基本要素



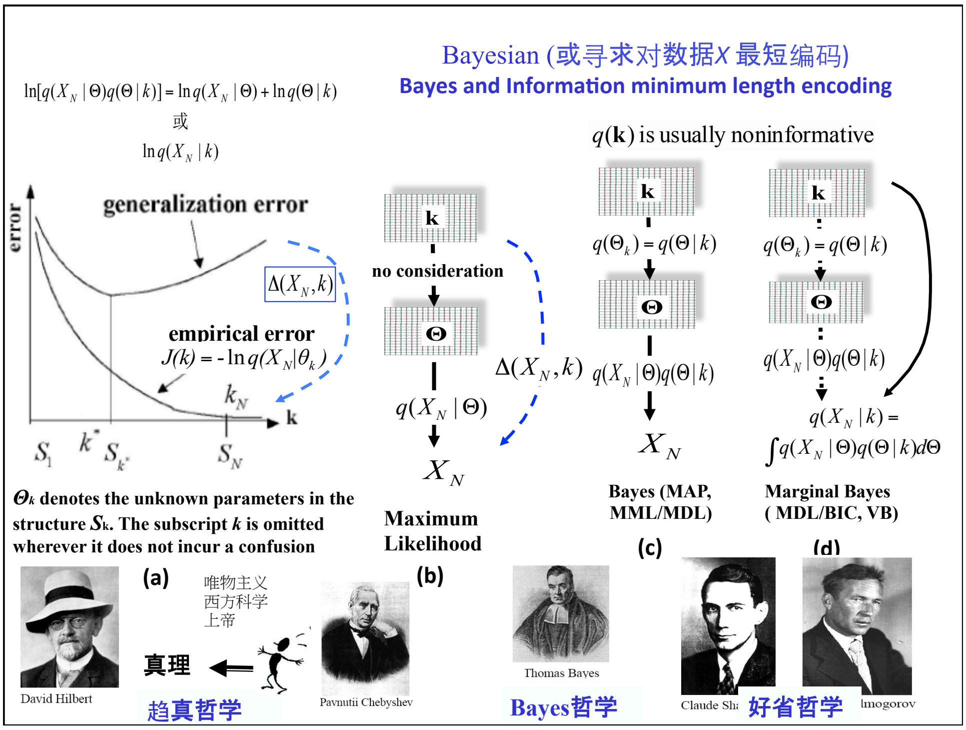
various learning algorithms



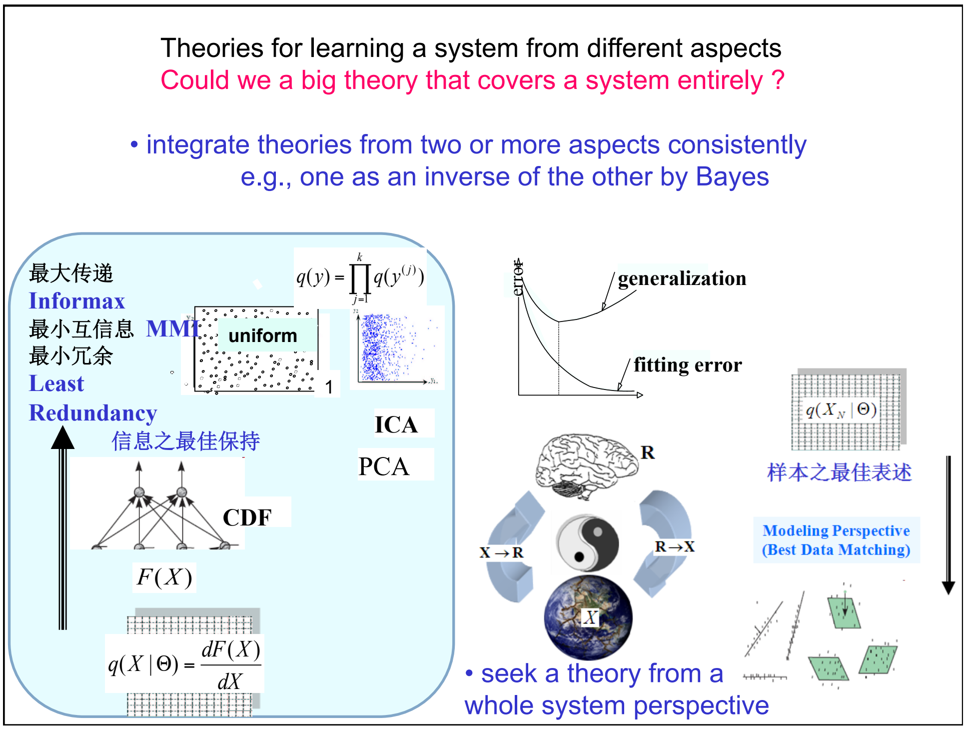
要素1
Learner
(hardware)
Gaussian
Gaussian mixture
(GM)
μ
x_t
e_t
$y_t = w^t x_t$
$y = [y_1, \cdots, y_k]$
$x = Ay + e$
$x = [x_1, \cdots, x_d]$
A structure comes from
a simple combination of
many simple units



要素2
Learning theory
value system
(software)
$q(X_N \mid \Theta)$
Best Data
representation
样本之最佳表述
X_N
样本之最佳记忆 best memory
最小误差拟合 least error fitting
最大似然 maximum likelihood
本质特点：學習以有限样本为媒介
Learning theory for a small size of samples
小样本学习理论
generalization error
error
难估计
fitting error or
$J(k) = -\ln q(X_N|\Theta_k)$
同步
背驰
k^*
k_N
S_1
S_{k^*}
S_N
M = 1
M = 2
40
20
0
−20
0
5
10



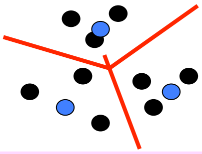
Bayesian (或寻求对数据X 最短编码)
Bayes and Information minimum length encoding
$\ln[q(X_N|\Theta)q(\Theta|k)] = \ln q(X_N|\Theta) + \ln q(\Theta|k)$
或
$\ln q(X_N|k)$
error
generalization error
$\Delta(X_N,k)$
empirical error
$J(k) = -\ln q(X_N|\theta_k)$
k_N
k
S_1 k^* S_{k^*} S_N
Θ_k denotes the unknown parameters in the structure S_k. The subscript k is omitted wherever it does not incur a confusion
(a)
唯物主义
西方科学
上帝
真理
David Hilbert
趋真哲学
k
no consideration
Θ
$q(X_N|\Theta)$
X_N
$\Delta(X_N,k)$
Maximum Likelihood
(b)
Pavnutii Chebyshev
$q(\mathbf{k})$ is usually noninformative
k
$q(\Theta_k) = q(\Theta|k)$
Θ
$q(X_N|\Theta)q(\Theta|k)$
X_N
Bayes (MAP, MML/MDL)
(c)
Thomas Bayes
Bayes哲学
k
$q(\Theta_k) = q(\Theta|k)$
Θ
$q(X_N|\Theta)q(\Theta|k)$
$q(X_N|k) = \int q(X_N|\Theta)q(\Theta|k)d\Theta$
Marginal Bayes (MDL/BIC, VB)
(d)
Claude Sh
好省哲学
lmogorov

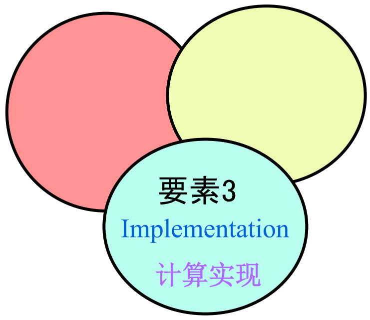


Theories for learning a system from different aspects
Could we a big theory that covers a system entirely ?
• integrate theories from two or more aspects consistently
e.g., one as an inverse of the other by Bayes
最大传递
Informax
最小互信息 MMI
最小冗余
Least
Redundancy
$q(y) = \prod_{j=1}^{k} q(y^{(j)})$
uniform
1
ICA
PCA
信息之最佳保持
CDF
$F(X)$
$q(X|\Theta) = \frac{dF(X)}{dX}$
error
generalization
fitting error
$q(X_N|\Theta)$
样本之最佳表述
R
X→R
R→X
X
Modeling Perspective
(Best Data Matching)
• seek a theory from a whole system perspective

组合与优化
e.g.

C_N^M choices

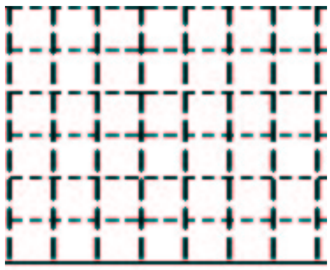


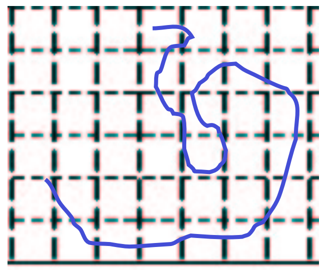


Combinatorial + continuous variable optimization

积分： $(\frac{1}{\Delta})^n$ 上计算之总和

积分与优化



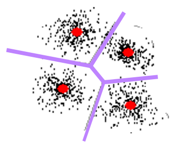


$\max q(X_N \mid k) = \int q(X_N \mid \Theta) q(\Theta \mid k) d\Theta$

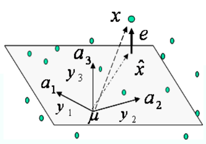
优化： $(\frac{1}{\Delta})^n$ 上之搜索轨迹

理论上的计算复杂性：一般说来是NP问题

Two Steps of Implementation



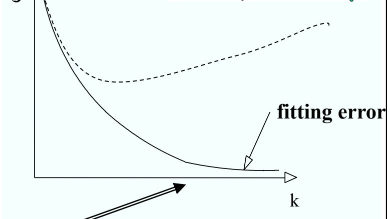
Step 1 Enumerate k for a set of candidate values, fixed at each candidate, make learning



$$\theta^*(k) = \arg\min_\theta \ln p(X_N \mid \theta)$$

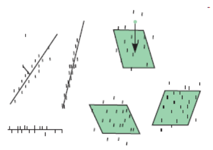
Step 2 Select the best k^* by

$$k^* = \arg\min_k J(\theta, k)$$

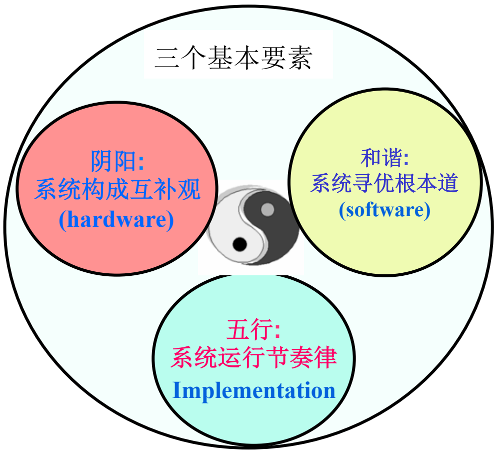


Very computational extensive !

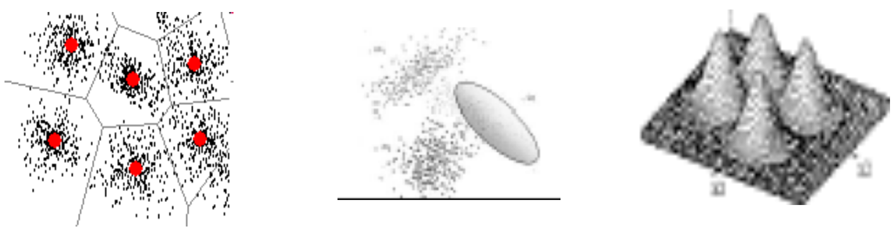
Inaccurate for large k !!!



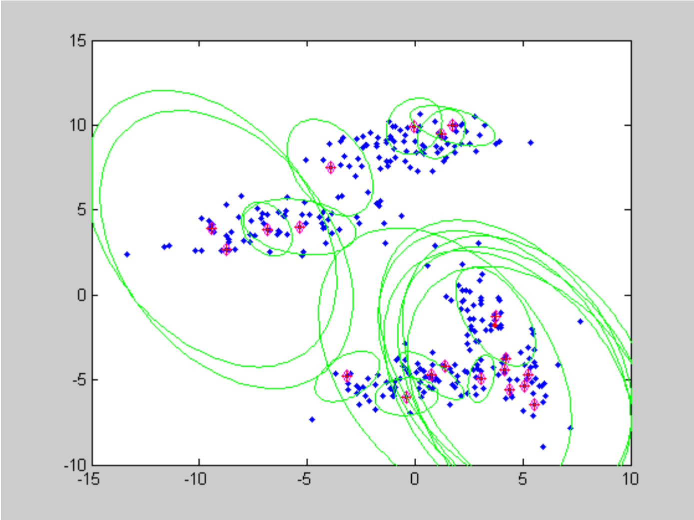
- 阴阳五行和谐是定性描述的古典系统理论
- 学习智能系统理论：基于现代概率论、信息论，可有效计算

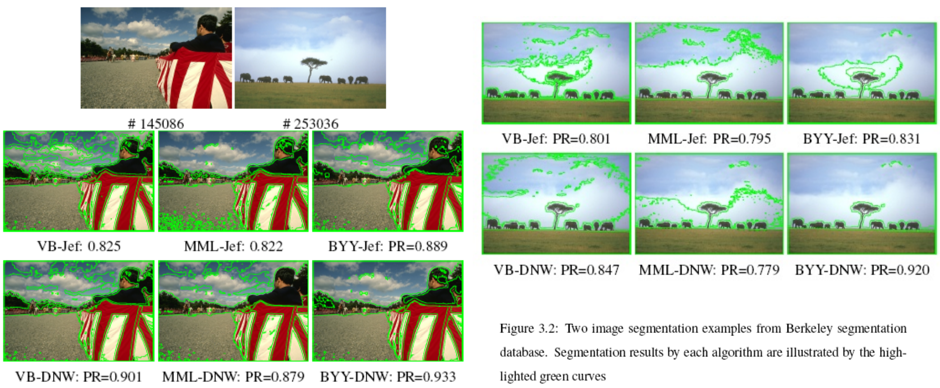


Learning Gaussian mixture $q(x \mid \theta)=\sum_{t=1}^{k}\alpha_t G(x_t|m_t,\Sigma_t)$



自动模型选择





Lei SHI, Shikui TU,Lei Xu (2011), " Learning Gaussian mixture with automatic model selection:A comparative study on three Bayesian related approaches", A special issue on Machine learning and intelligence science: IScIDE2010 (B), Journal of Frontiers of Electrical and Electronic Engineering in China 6(2) (2011) 215–244.

Table 3.2: Average PR scores of 5 runs on the 100 testing images of the Berkeley image segmentation database by GMM algorithms (without post-processings)

VB-Jef	MML-Jef	BYY-Jef	VB-DNW	MML-DNW	BYY-DNW
0.772	0.752	0.816	0.803	0.788	0.851

Unsupervised Natural Image Segmentation via Bayesian Ying-Yang Harmony Learning Theory.
By: Shaojun Zhu, Jieyu Zhao, Lijun Guo, Yuanyuan Zhang
Neurocomputing, 121, 9 December 2013, Pages 532–539

(a) Result from [18] (b) Result of proposed method

Figure 5: Comparison between BYY-DNW [18] and our method.

Table 2: Comparison on the BSD500 using the PRI, GCE, VOI, and BDE indices. For PRI, higher values indicate better segmentation; for GCE, VOI and BDE, lower values indicate better segmentation.

Method	PRI	GCE	VOI	BDE
BYY	0.7785	0.2076	2.4281	8.8760
CTM	0.7628	0.2093	2.0788	9.4038
NCuts(k=5)	0.6930	0.3255	2.4456	15.5592
MeanShift	0.7573	0.0698	4.7244	8.9005
GBMS	0.7362	0.3570	2.1737	15.8600
GraphBased	0.7872	0.2315	2.8308	8.4047
DCM(k=5)	0.6923	0.2950	2.0106	17.7190
SCKM	0.7762	0.2303	2.1141	10.0908
MD2S	0.7735	0.2348	2.3614	10.3685

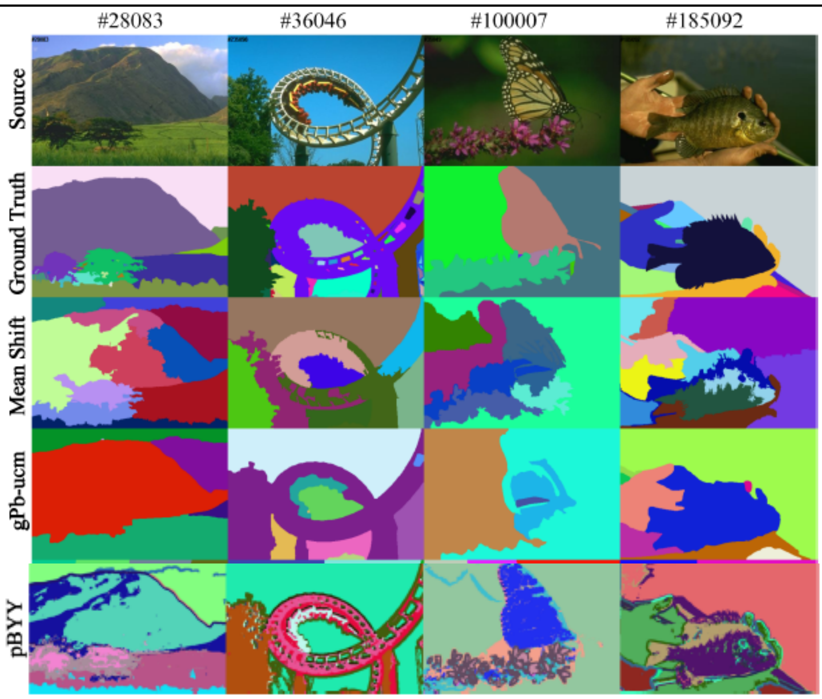
Chen, G., Heng, P.-A., Xu, L.: Projection embeded byy learning algorithm for gassuian mixture based clustering, Applied Informatics (2014)

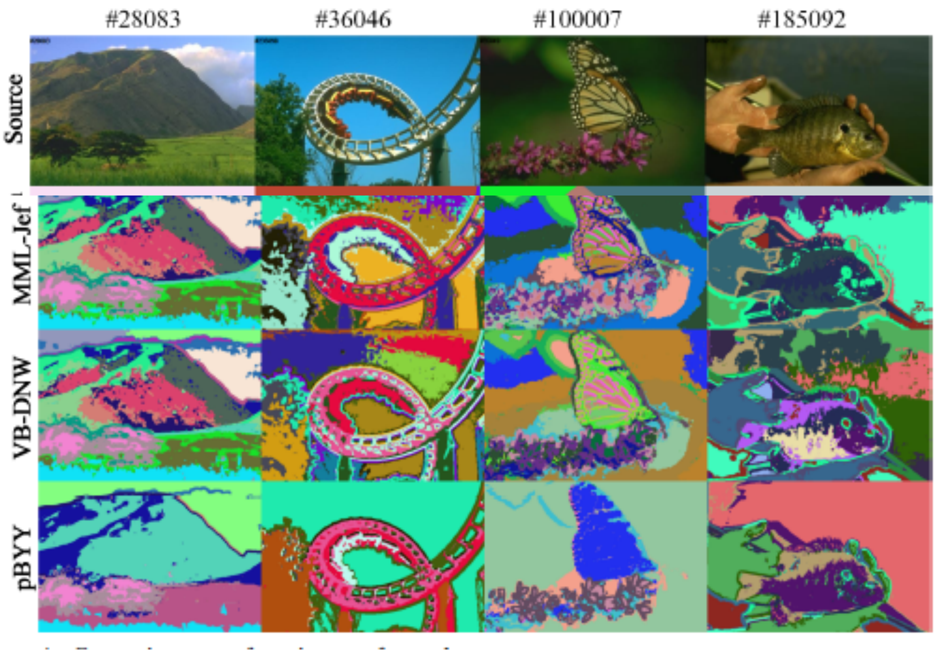
Data Set	GMM-a			GMM-b			GMM-c		
Algorithms	CSR	VI	PRI	CSR	VI	PRI	CSR	VI	PRI
VB-DNW	0.4660	1.0243	0.7730	0.5160	0.6264	0.8599	0.1060	1.3337	0.6469
MML-Jef	0.1700	3.2637	0.7345	0.1600	4.8235	0.7573	0.4140	58.0039	0.6388
BYY-Jef	0.2167	1.1135	0.7006	0.5533	0.6650	0.8257	0.0100	1.6889	0.4732
BYY-DNW	0.1433	1.1947	0.7039	0.0700	**0.5373***	0.8760	0	1.7948	0.4622
pBYY	**0.7260***	**0.5852***	**0.8692***	**0.8840***	0.5482	**0.8779***	**0.6100***	**1.1328***	**0.7451***

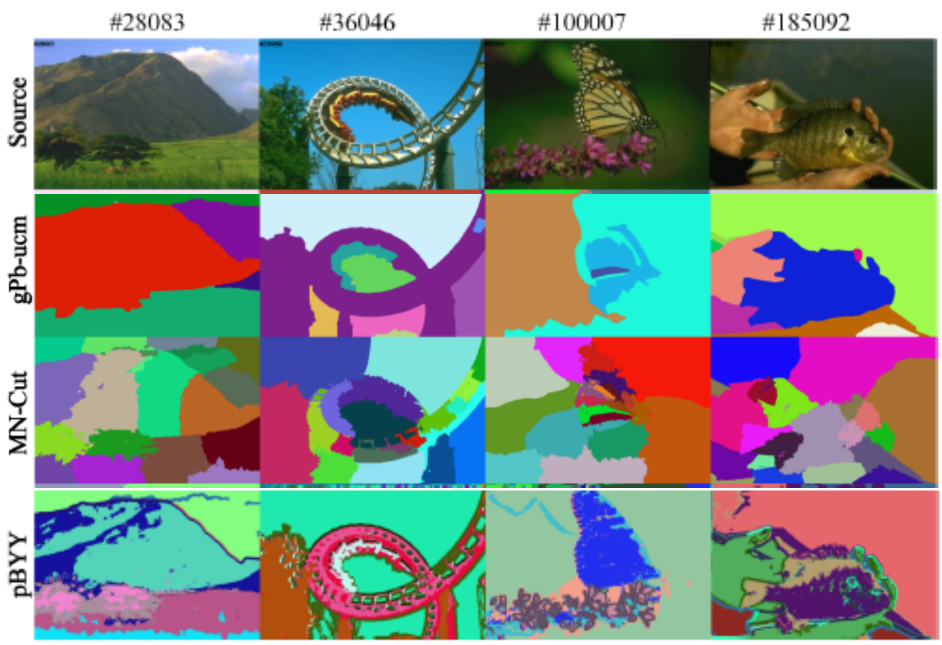
Table 1: The performance of each algorithm on three synthetic data sets after 500 trials, with the initial number of Gaussian components is set as $k = 20$, where "*" indicates the best within its column. For a good performance, we expect that the values of CSR and PRI are big and that the VI value is small.

	BSDS500						
	Human	Mean Shift	MN-Cut	gPb-owt-ucm	MML-Jef	VB-DNW	pBYY
PRI	0.88	0.8157	0.8066	0.7489	0.7851	0.7866	0.8196
VI	1.17	2.2912	2.5163	1.7539	3.4966	3.5589	2.8140
CR	0.72	0.439	0.393	0.439	0.325	0.325	0.487

Table 3: Performance scores on the BSDS. The performance of each algorithm is evaluated separately against each of five human-drawn ground-truth segmentations per image, and then their average is obtained as the score on this image. For the Covering Rate (CR) metrics, a larger value indicates a better performance.




#28083
#36046
#100007
#185092
Source
MML-Jef
VB-DNW
pBYY


#28083
#36046
#100007
#185092
Source
gPb-ucm
MN-Cut
pBYY

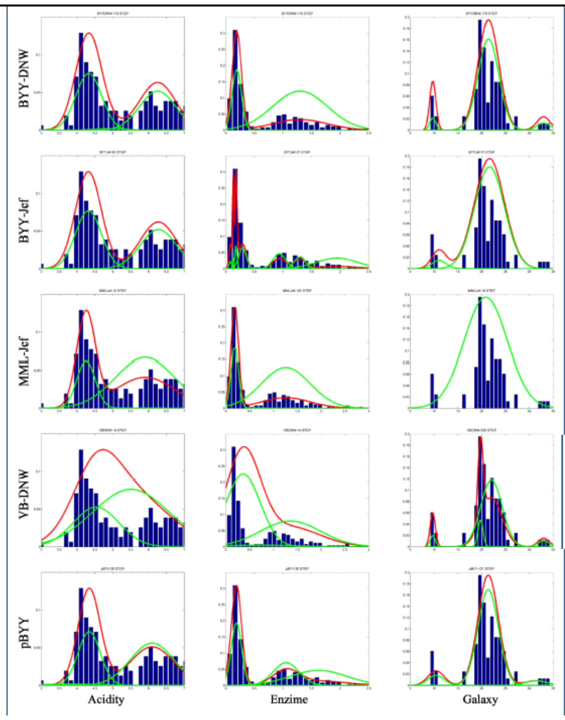


Figure 3: Density fitting on 1-D real-world data sets. From left to right: Acidity data set, Enzime and Galaxy data set; while from top to down: BYY-DNW, BYY-Jef, MML-Jef, VB-DNW and pBYY algorithm. (Red curve indicates the overall density function and green curve the density function per component)

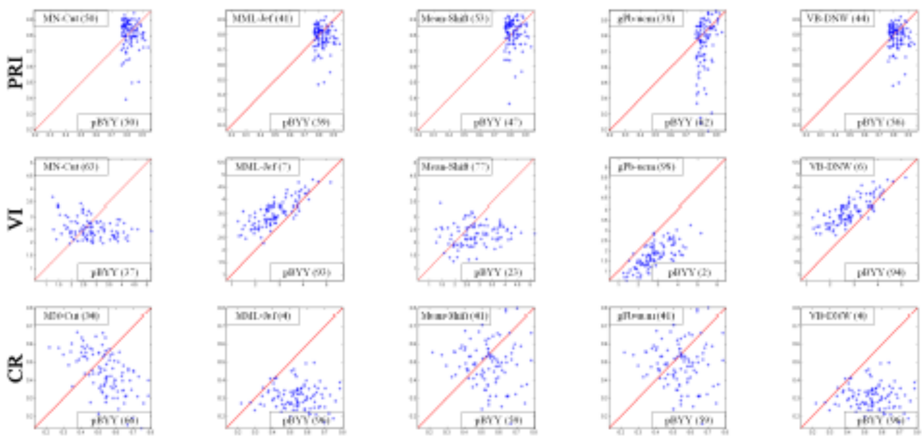
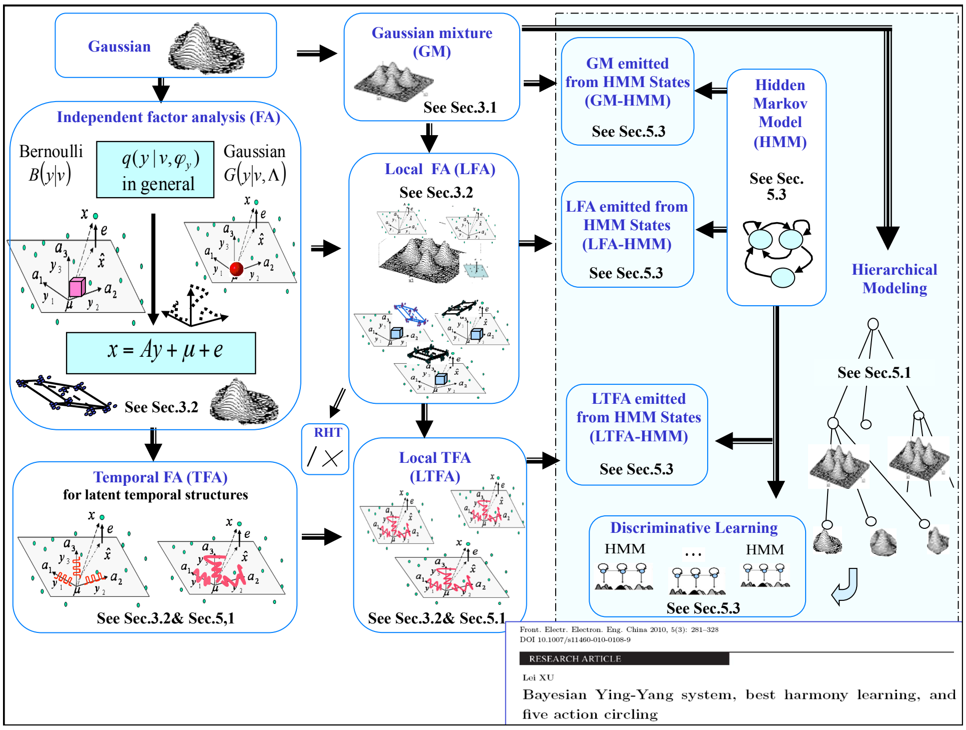
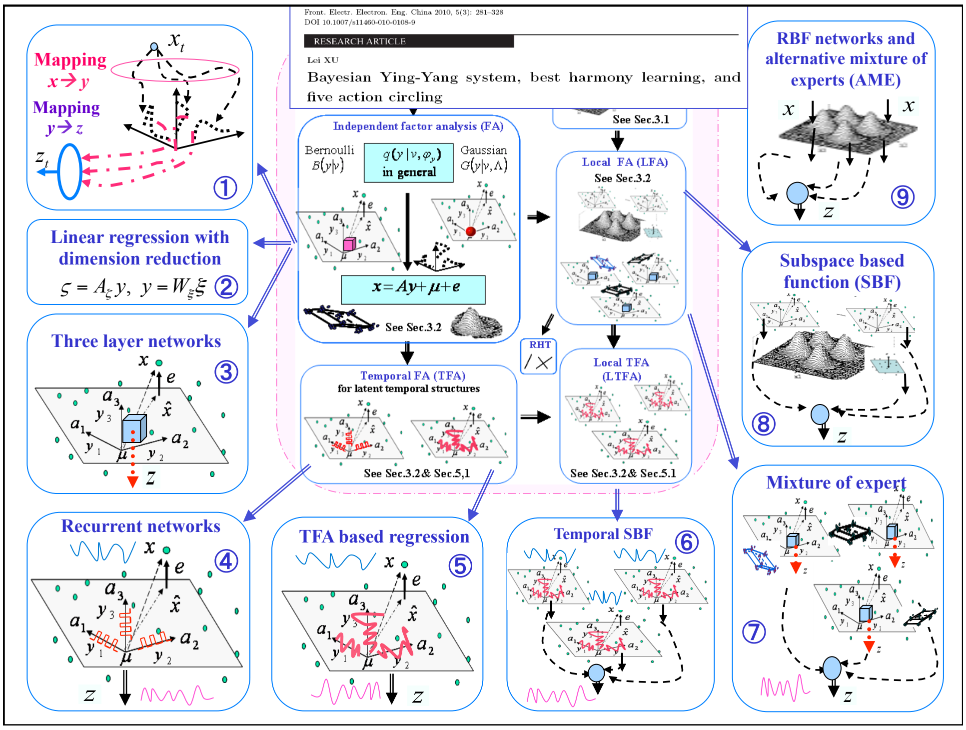
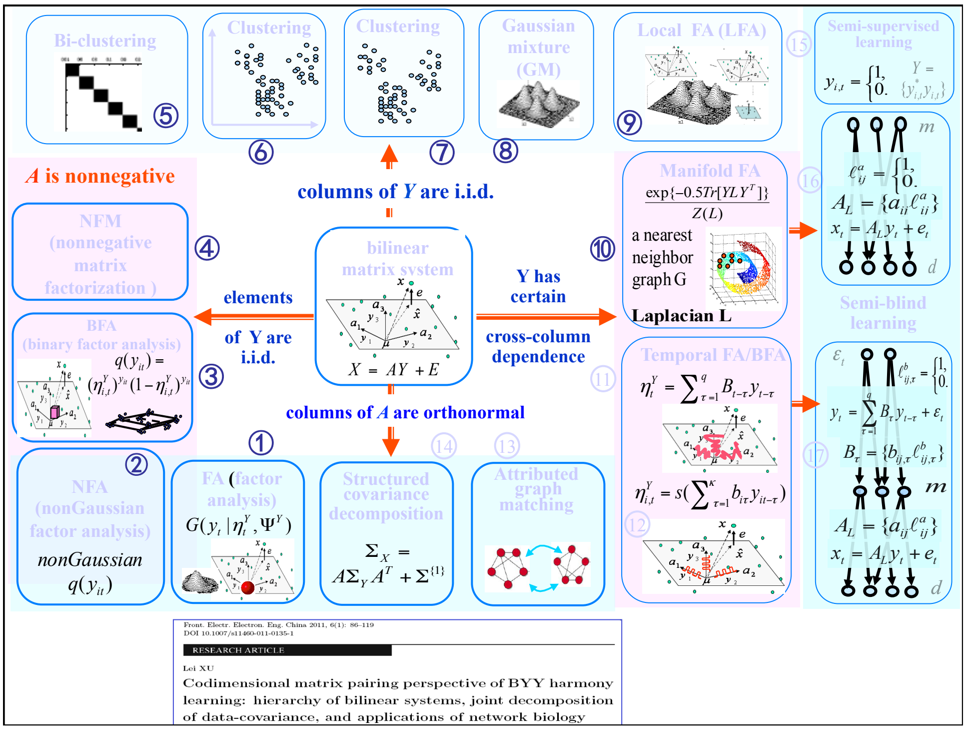


Figure 5: Pairwise comparison of segmentation algorithms on the BSDS500. The coordinates of the blue dots are the PRI, VI and CR scores pBYY and its competing algorithms, obtained per image. Red line represents the boundary of equal performance by the two algorithms, and the boxed digits indicate the number of images where one algorithm is better. For example, the last one shows MN-Cut outscores pBYY merely on 4/100 images.


Gaussian
Gaussian mixture (GM)
See Sec.3.1
Independent factor analysis (FA)
Bernoulli $B(y|v)$
$q(y|v,\varphi_y)$ in general
Gaussian $G(y|v,\Lambda)$
$x = Ay + \mu + e$
See Sec.3.2
GM emitted from HMM States (GM-HMM)
See Sec.5.3
Hidden Markov Model (HMM)
See Sec. 5.3
Local FA (LFA)
See Sec.3.2
LFA emitted from HMM States (LFA-HMM)
See Sec.5.3
Hierarchical Modeling
See Sec.5.1
RHT
Local TFA (LTFA)
See Sec.3.2& Sec.5.1
LTFA emitted from HMM States (LTFA-HMM)
See Sec.5.3
Temporal FA (TFA)
for latent temporal structures
See Sec.3.2& Sec.5,1
Discriminative Learning
HMM … HMM
See Sec.5.3
Front. Electr. Electron. Eng. China 2010, 5(3): 281–328
DOI 10.1007/s11460-010-0108-9
RESEARCH ARTICLE
Lei XU
Bayesian Ying-Yang system, best harmony learning, and five action circling


Mapping $x \to y$
Mapping $y \to z$
①
Linear regression with dimension reduction
$\varsigma = A_\varsigma y,\ y = W_\xi \xi$ ②
Three layer networks ③
Recurrent networks ④
TFA based regression ⑤
Temporal SBF ⑥
Mixture of expert ⑦
Subspace based function (SBF) ⑧
RBF networks and alternative mixture of experts (AME) ⑨
Front. Electr. Electron. Eng. China 2010, 5(3): 281–328
DOI 10.1007/s11460-010-0108-9
RESEARCH ARTICLE
Lei XU
Bayesian Ying-Yang system, best harmony learning, and five action circling
Independent factor analysis (FA)
$x = Ay + \mu + e$
See Sec.3.2
Local FA (LFA)
See Sec.3.1
Temporal FA (TFA)
for latent temporal structures
See Sec.3.2& Sec.5,1
Local TFA (LTFA)
RHT



阴阳五行和谐系统理论 versu Learning System

《道德经》 《周易》

《黄帝内经》

(A) 阴阳:

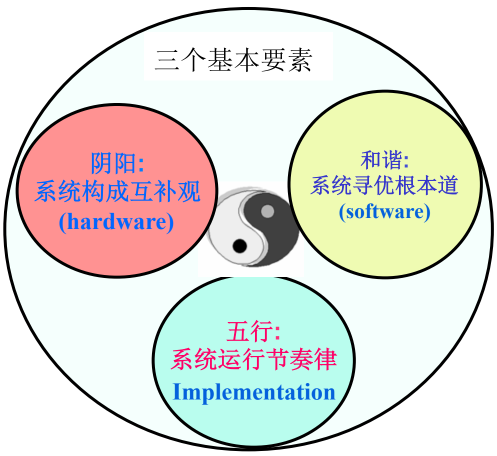
系统构成互补观

周易《系辞上》“乾以易知，坤以简能”

《周易集解》虞翻曰“陽見稱易，陰藏為簡，簡，閲也”

宋邵雍 《观物外篇》“阳以阴为基”“阴以阳为唱”

阴阳交错，流转，消長，轉化



(C) 和谐:

系统寻优根本道

《易・彖传》“保合太和, 乃利贞”

《正蒙》“太和所谓道”

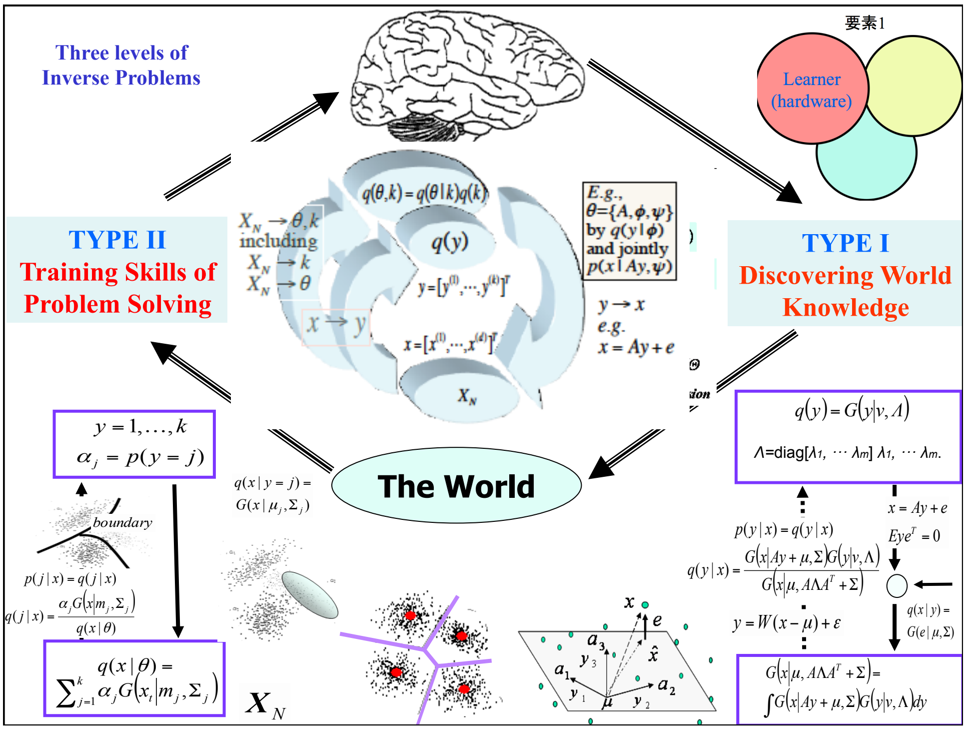
和谐指阴阳以最默契或交换信息最小的方式达到最大共识。

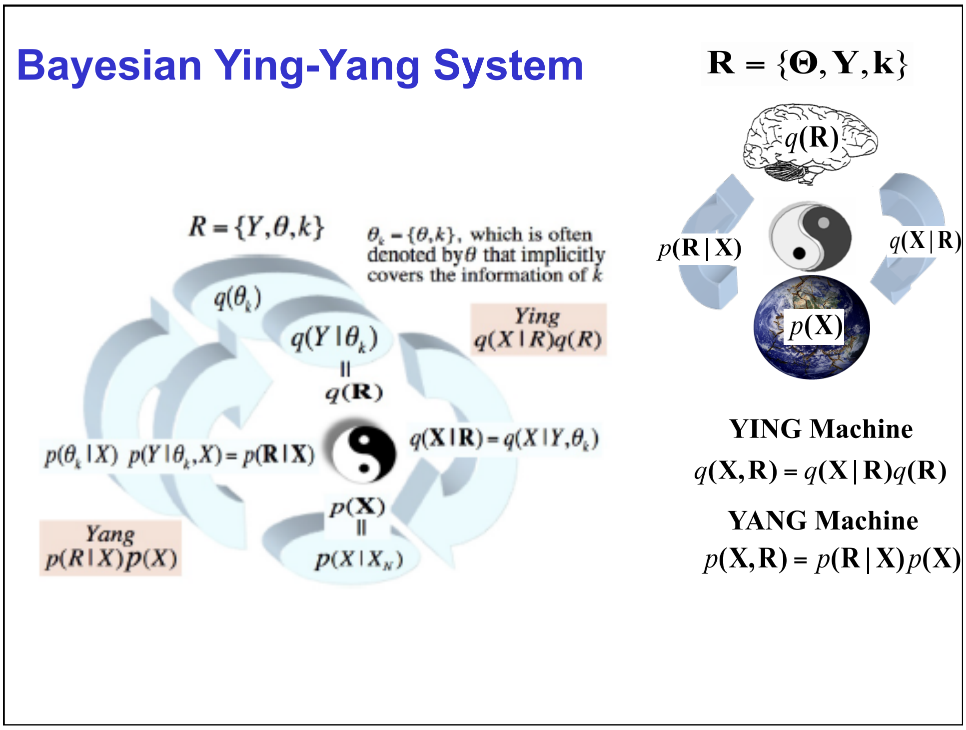
(B) 五行:系统运行节奏律

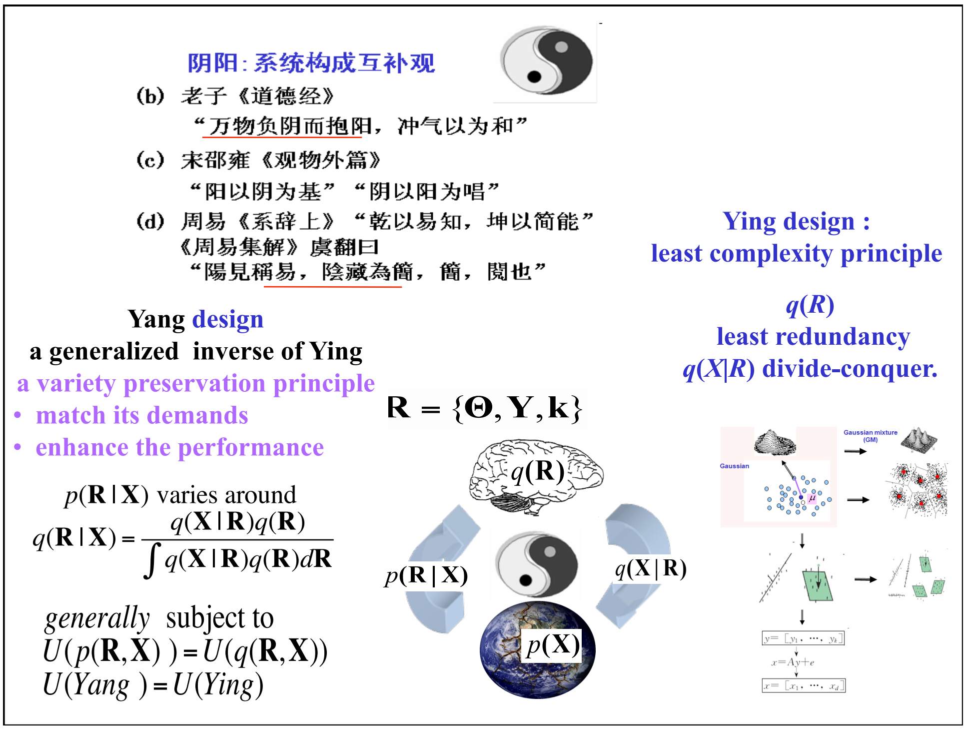
《尚书・洪范》《黄帝内经》 陰陽二氣之流動而存在的五種“存在形式”

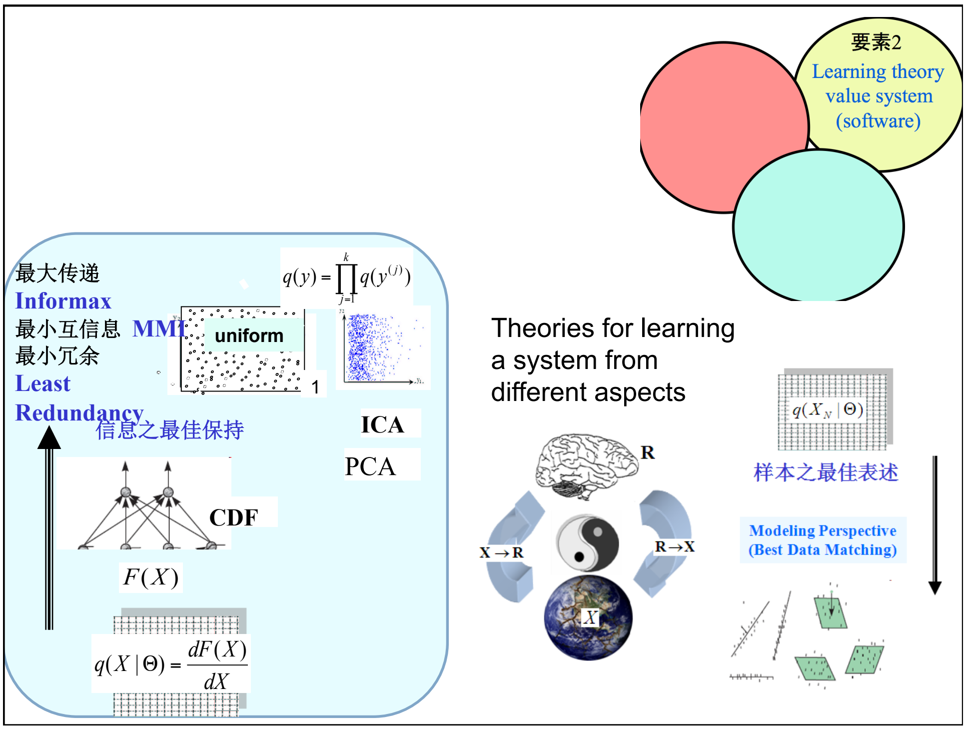
宋张载《正蒙・乾称篇》 “凡象皆气”

生克乘诲 Sheng-Ke-Cheng-Hui: 相生, 相克, 相乘, 相诲


Three levels of
Inverse Problems
要素1
Learner
(hardware)
TYPE II
Training Skills of
Problem Solving
TYPE I
Discovering World
Knowledge
$q(\theta,k)=q(\theta|k)q(k)$
$q(y)$
$X_N \to \theta, k$
including
$X_N \to k$
$X_N \to \theta$
$y=[y^{(1)},\cdots,y^{(k)}]^T$
$x \to y$
$x=[x^{(1)},\cdots,x^{(d)}]^T$
X_N
E.g.,
$\theta=\{A,\phi,\psi\}$
by $q(y|\phi)$
and jointly
$p(x|Ay,\psi)$
$y \to x$
e.g.
$x=Ay+e$
The World
$y=1,\ldots,k$
$\alpha_j=p(y=j)$
boundary
$p(j|x)=q(j|x)$
$q(j|x)=\frac{\alpha_j G(x|m_j,\Sigma_j)}{q(x|\theta)}$
$q(x|y=j)=G(x|\mu_j,\Sigma_j)$
$q(x|\theta)=\sum_{j=1}^{k}\alpha_j G(x_t|m_j,\Sigma_j)$
X_N
$q(y)=G(y|v,\Lambda)$
Λ=diag[$\lambda_1,\cdots\lambda_m$] $\lambda_1,\cdots\lambda_m$.
$p(y|x)=q(y|x)$
$q(y|x)=\frac{G(x|Ay+\mu,\Sigma)G(y|v,\Lambda)}{G(x|\mu,A\Lambda A^T+\Sigma)}$
$x=Ay+e$
$Eye^T=0$
$y=W(x-\mu)+\varepsilon$
$q(x|y)=G(e|\mu,\Sigma)$
$G(x|\mu,A\Lambda A^T+\Sigma)=\int G(x|Ay+\mu,\Sigma)G(y|v,\Lambda)dy$


Bayesian Ying-Yang System
$\mathbf{R}=\{\Theta,\mathbf{Y},\mathbf{k}\}$
$q(\mathbf{R})$
$p(\mathbf{R}|\mathbf{X})$
$q(\mathbf{X}|\mathbf{R})$
$p(\mathbf{X})$
$R=\{Y,\theta,k\}$
$\theta_k=\{\theta,k\}$, which is often denoted by θ that implicitly covers the information of k
$q(\theta_k)$
$q(Y|\theta_k)$
Ying
$q(X|R)q(R)$
$q(\mathbf{R})$
$p(\theta_k|X)\ p(Y|\theta_k,X)=p(\mathbf{R}|\mathbf{X})$
$q(\mathbf{X}|\mathbf{R})=q(X|Y,\theta_k)$
$p(\mathbf{X})$
$p(X|X_N)$
Yang
$p(R|X)p(X)$
YING Machine
$q(\mathbf{X},\mathbf{R})=q(\mathbf{X}|\mathbf{R})q(\mathbf{R})$
YANG Machine
$p(\mathbf{X},\mathbf{R})=p(\mathbf{R}|\mathbf{X})p(\mathbf{X})$


阴阳: 系统构成互补观
(b) 老子《道德经》
“万物负阴而抱阳，冲气以为和”
(c) 宋邵雍《观物外篇》
“阳以阴为基” “阴以阳为唱”
(d) 周易《系辞上》“乾以易知，坤以简能”
《周易集解》虞翻曰
“陽見稱易，陰藏為簡，簡，閱也”
Ying design :
least complexity principle
$q(R)$
least redundancy
$q(X|R)$ divide-conquer.
Yang design
a generalized inverse of Ying
a variety preservation principle
• match its demands
• enhance the performance
$\mathbf{R} = \{\mathbf{\Theta}, \mathbf{Y}, \mathbf{k}\}$
$p(\mathbf{R}|\mathbf{X})$ varies around
$q(\mathbf{R}|\mathbf{X}) = \frac{q(\mathbf{X}|\mathbf{R})q(\mathbf{R})}{\int q(\mathbf{X}|\mathbf{R})q(\mathbf{R})d\mathbf{R}}$
generally subject to
$U(p(\mathbf{R},\mathbf{X})) = U(q(\mathbf{R},\mathbf{X}))$
$U(Yang) = U(Ying)$
$q(\mathbf{R})$
$p(\mathbf{R}|\mathbf{X})$
$q(\mathbf{X}|\mathbf{R})$
$p(\mathbf{X})$
Gaussian mixture (GM)
Gaussian

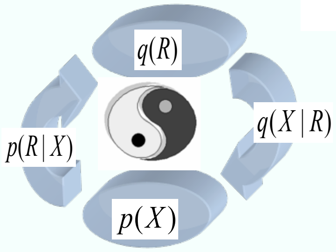

要素2
Learning theory
value system
(software)
最大传递
Informax
最小互信息 MMI
最小冗余
Least
Redundancy
信息之最佳保持
$q(y) = \prod_{j=1}^{k} q(y^{(j)})$
uniform
1
ICA
PCA
CDF
$F(X)$
$q(X|\Theta) = \frac{dF(X)}{dX}$
Theories for learning
a system from
different aspects
$q(X_N|\Theta)$
样本之最佳表述
Modeling Perspective
(Best Data Matching)
R
X→R
R→X
X

最大和谐原理

和谐: 系统寻优根本道

(e) 《正蒙》“太和所谓道”

(f) 阴阳以最默契或交换信息最小的方式达到最大一致。



Taoism philosophy perspective :

$$p(X,R)=p(R\mid X)p(X) \xrightarrow{\textbf{Best Harmony}} \xleftarrow{} q(X,R)=q(X\mid R)q(R)$$

Best Harmony Learning Principle

best matching + ***(least complexity or most firm)***

$$Max\, H(p\,\|\,q,\mathbf{k},\boldsymbol{\Theta}),\ H(p\,\|\,q)=\int p(R\mid X)p(X)\ln[q(X\mid R)q(R)]dXdR$$

$$H(p||q)=\int p(R|X)p(X)\ln[q(X|R)q(R)]dX\,dR$$
$$\text{subject to } U(p(X,R))=U(q(X,R)).$$

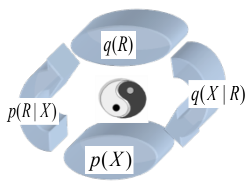
$$\max H(p||q),\quad H(p||q)=\int p(R\mid X)p(X)\ln[q(X\mid R)q(R)]dXdR$$
$$=\int p(R\mid X)p(X)\ln p(R\mid X)p(X)dXdR-KL(p||q)$$

$$\min KL(p||q),\ KL(p||q)=\int p(R\mid X)p(X)\ln\frac{p(R\mid X)p(X)}{q(X\mid R)q(R)}dXdR$$

和谐: 系统寻优根本道

(e) 《正蒙》“太和所谓道”

(f) 阴阳以最默契或交换信息最小的方式达到最大一致。



以最默契的方式

Ying and Yang seeks a best agreement via minimizing *KL*(*p*||*q*) in a most tacit manner via minimizing the information -*H*(*p*||*p*) that is transferred by Yang.

Front. Electr. Electron. Eng. China 2010, 5(3): 281–328
DOI 10.1007/s11460-010-0108-9

RESEARCH ARTICLE — See Sec.4.1 and esp. Eq.(24) in

Lei XU

Bayesian Ying-Yang system, best harmony learning, and five action circling

最大和谐泛函：阴阳和谐的信息理论解释

$$p(X,R)=p(R\mid X)p(X) \quad \textbf{Best Harmony} \longleftarrow \quad q(X,R)=q(X\mid R)q(R)$$

$$Max\, H(p\,\|\,q),\ H(p\,\|\,q)=\int p(R\mid X)p(X)\ln[q(X\mid R)q(R)]dXdR$$

Matching nature

$$\max_q H(p\|q) \quad \rightarrow \quad q(X,R) \text{ approaches } p(X,R)$$

The trend is $q(X,R)=p(X,R)$

Least complexity nature

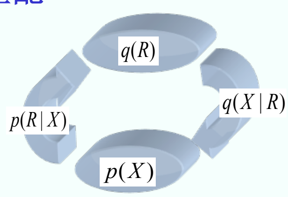
At the limit, $\max H(p\|p)$ actually minimizes the system entropy

$$H(p\|p)=-\int p(X,R)\ln p(X,R)dXdR$$

最佳匹配+最小复杂度

best matching + (least complexity or most firm)

BYY Best Matching 最佳匹配



See Sec.4.1 in

Front. Electr. Electron. Eng. China 2010, 5(3): 281–328
DOI 10.1007/s11460-010-0108-9

RESEARCH ARTICLE

Lei XU

Bayesian Ying-Yang system, best harmony learning, and five action circling

$$\frac{dP(R,X)}{dRdX}=p(R\mid X)p(X)>0 \quad \int p(R\mid X)p(X)dXdR=1$$

$$\frac{dQ(R,X)}{dRdX}=q(X\mid R)q(R)>0 \quad \int q(X\mid R)q(R)dXdR=1$$

$$Max\,\mathrm{KL}(P\,\|\,Q),\ KL(P\,\|\,Q)=-\int\ln\frac{dQ(R,X)}{dP(R,X)}dP(R,X)$$

Best Harmony Learning Principle 最大和谐

best matching + (least complexity or most firm)

$$Max\, H(p\,\|\,q),\ H(P\,\|\,Q)=\int\ln\frac{dQ(R,X)}{d\mu(R,X)}dP(R,X)$$

Harmony measure : A Randon-Nikodym Derivative perspective

$$H_\mu(P\|Q)=\int \frac{dP}{d\mu}\ln\left(\frac{dQ}{d\mu}\right)d\mu$$

$Q = P$

$$H_\mu(P\|P)=\int \frac{dP}{d\mu}\ln\left(\frac{dP}{d\mu}\right)d\mu$$

$\mu = Lebesque$

$$H(p\|p)=\int p(X)\ln p(X)dX$$

$-H(p\|p)$ is an entropy

$\mu = P$

$$KL(P\|Q)=-H_P(P\|Q)=\int \ln\left(\frac{dP}{dQ}\right)dP$$

$$\mu = P,\ p(X)=\frac{dP}{dX},\ q(X)=\frac{dQ}{dX}$$

$$H_P(P\|Q)=-KL(p\|q)$$

$$KL(p\|q)=\int p(X)\ln\frac{p(X)}{q(X)}dX$$

$KL-divergence$

$\mu = Lebesque$ ①a

$$H_\mu(P\|Q)=\int p(X)\ln q(X)dX$$

$$p(X)=\frac{dP}{dX},\quad q(X)=\frac{dQ}{dX}$$

Front. Electr. Electron. Eng. 2012, 7(1): 147–196
DOI 10.1007/s11460-012-0190-2

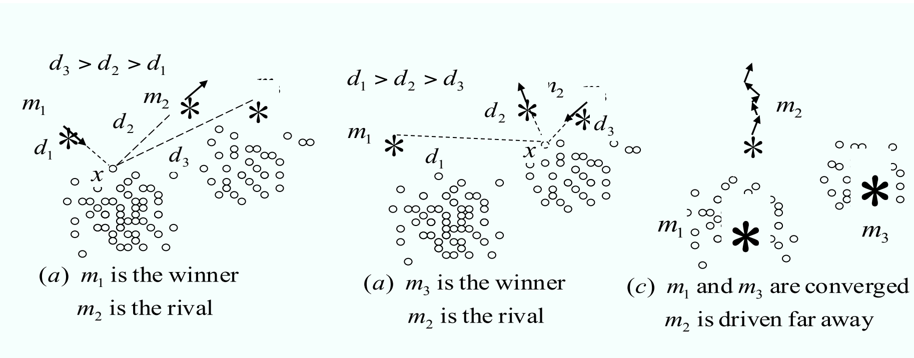
See Fig.5 in

RESEARCH ARTICLE

Lei XU

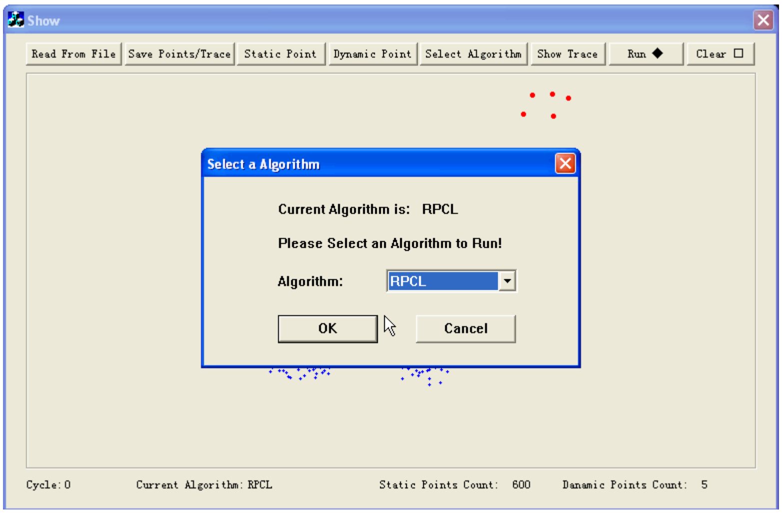
On essential topics of BYY harmony learning: Current status, challenging issues, and gene analysis applications

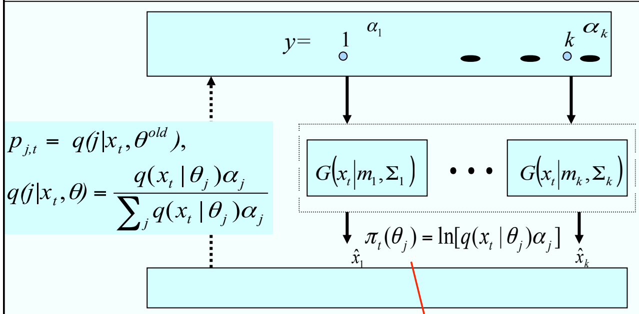
Rival Penalized Competitive Learning (RPCL)
(penalizing particulars)



See Xu, Oja & Krzyzak, ICPR92, then
Xu, L., A. Krzyzak, and E.Oja, IEEE Tr. on Neural Networks, Vol.4, No.4, 1993, pp636-649.

Google Scholar > 600 SCI-Expanded > 300





$$N_\ell = \sum_t p_{\ell,t},\quad \alpha_\ell^{new} = \frac{N_\ell}{N},$$

$$\mu_\ell^{new} = \frac{1}{N_\ell}\sum_t p_{\ell,t} x_t,$$

$$\Sigma_\ell^{new} = \frac{1}{N_\ell}\sum_t p_{\ell,t}(x_t-\mu_\ell^{old})(x_t-\mu_\ell^{old})^T$$

EM algorithm

BYY阴阳和谐学习的微观特征

$$p_{\ell,t} = q(\ell|x_t,\theta^*)(1+\Delta\pi_{\ell,t}),$$

$$\Delta\pi_{\ell,t} = \begin{cases} \boxed{\pi_t(\theta_\ell^*) - \sum_j q(j|x_t,\theta^*)\pi_t(\theta_j^*), \quad (a)} \\ -E_t(\theta_\ell^*) + \sum_j q(j|x_t,\theta^*)E_t(\theta_\ell^*), \ (b) \end{cases}$$

$$\pi_t(\theta_j) = \ln[\alpha_j G(x_t|\mu_j,\Sigma_j)],$$

$$E_t(\theta_j) = -\ln q(j|x_t,\theta) = -\ln\frac{\alpha_j G(x_t|\mu_j,\Sigma_j}{q(x_t|\theta)}, \quad (10)$$

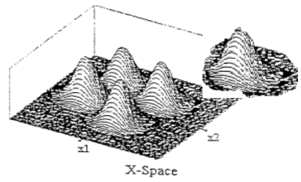
鼓励先进
惩罚落后
宽容多数

Automatic Model Selection : A new road

• For $\theta_k = \bigcup_{j=1}^{k} \theta^{(j)}$, there is a subset $\theta_{SR}^{(j)} \subset \theta^{(j)}$.
an indicator $\Psi(\theta_{SR}^{(j)}) \to 0$ or even some parameters in $\theta_{SR}^{(j)}$ towards 0 means effectively a model with its scale reduced by one (e.g., $k \to k$-1).

$$\alpha_\ell = 0 \text{ and } Tr[\Sigma_\ell] = 0.$$

• There is either a learning algorithm (e.g., RPCL learning, 1992) or a learning principle (e.g., Ying-Yang harmony).

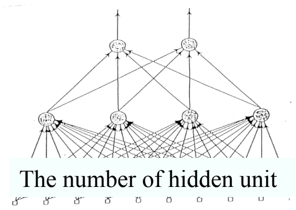


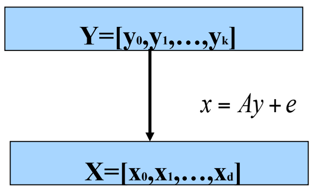
• Learning (either an implementation of the algorithm or an optimization of the principle) drives that $\Psi(\theta_{SR}^{(j)}) \to 0$ or/and some parameter in $\theta_{SR}^{(j)}$ towards 0, if corresponding components are redundant and thus discarded.

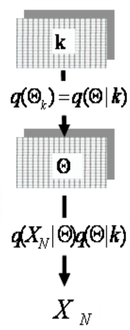
featured with either $J(\theta_{k+1}) \to \infty$ or $J(\theta_{k+1}) = J(\theta_k)$

Model Selection

Search a $\hat{\theta}_k$ in a dimension reduced space Θ for a model with a less number of free parameters.







Regularization

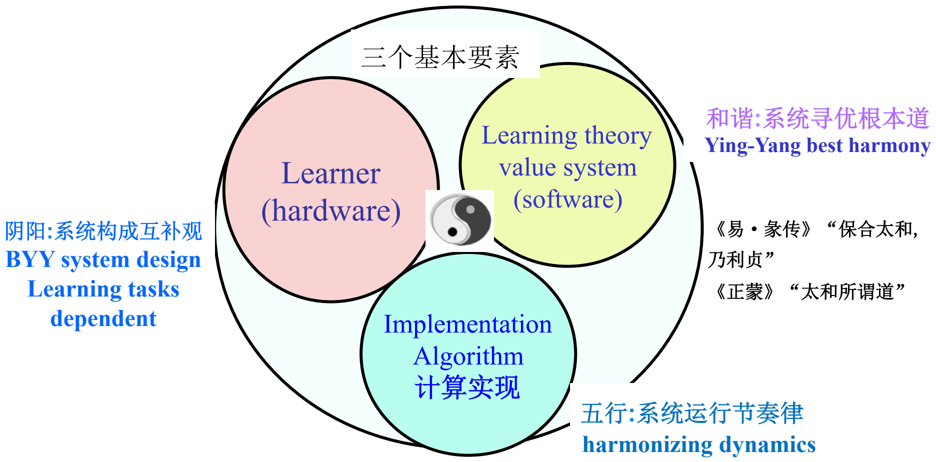
Large enough to accommodate the true structure
But impose certain constraint on regularity

Sparse learning since 1995

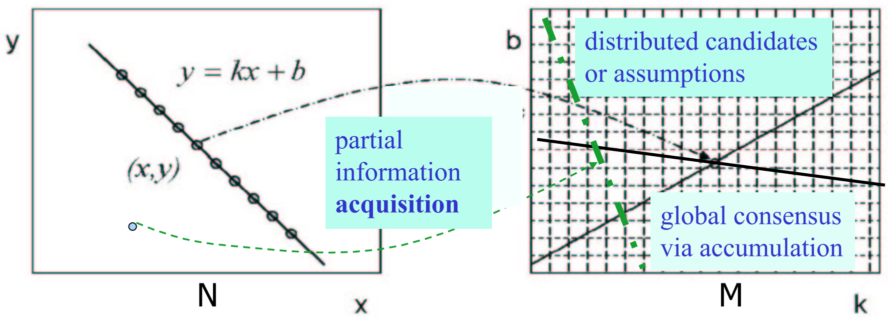
•model selection prunes away extra individual columns of *A*,
•focusing on pruning away individual parameters in *A* per element.

Automatic Model Selection : **since1992, e.g., RPCL (Xu, Kryzak, and Oja)**

BYY harmony learning



Hough transform (Hough, 1962)



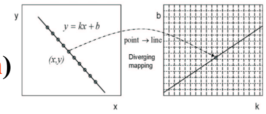
- Accumulation window
- Quantization resolution

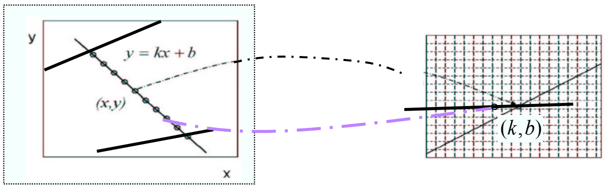
critical drawbacks

- Missing objects due to window
- Blurring peaks due to low quantization
- Cost increasing exponentially with d
- Disturbing accumulations: useful pixel and noise

Randomized Hough Transform (RHT)

(tackling the problems via a new fundamental mechanism)

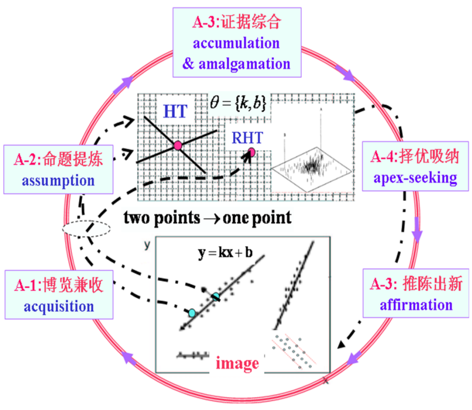


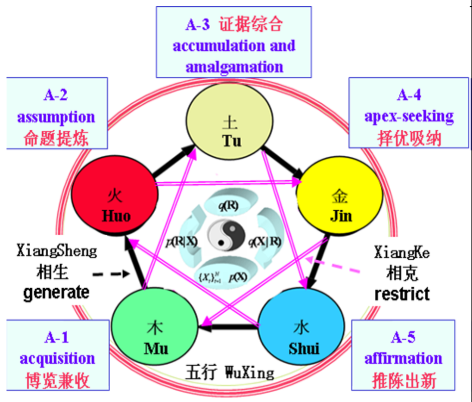


- Flexible window
- Flexible quantization
- Dynamic storage
- Reduced disturbances

A new curve detection method: randomized Hough transform (RHT)
L Xu, E Oja… - Pattern Recognition Letters, 1990 – Elsevier
SCI-citations >350 Google Scholar >820

Randomized Hough transform (RHT): basic mechanisms, algorithms, and computational complexities
L Xu, E Oja - CVGIP Image Understanding, 1993
SCI-citations > 200 Google Scholar >530

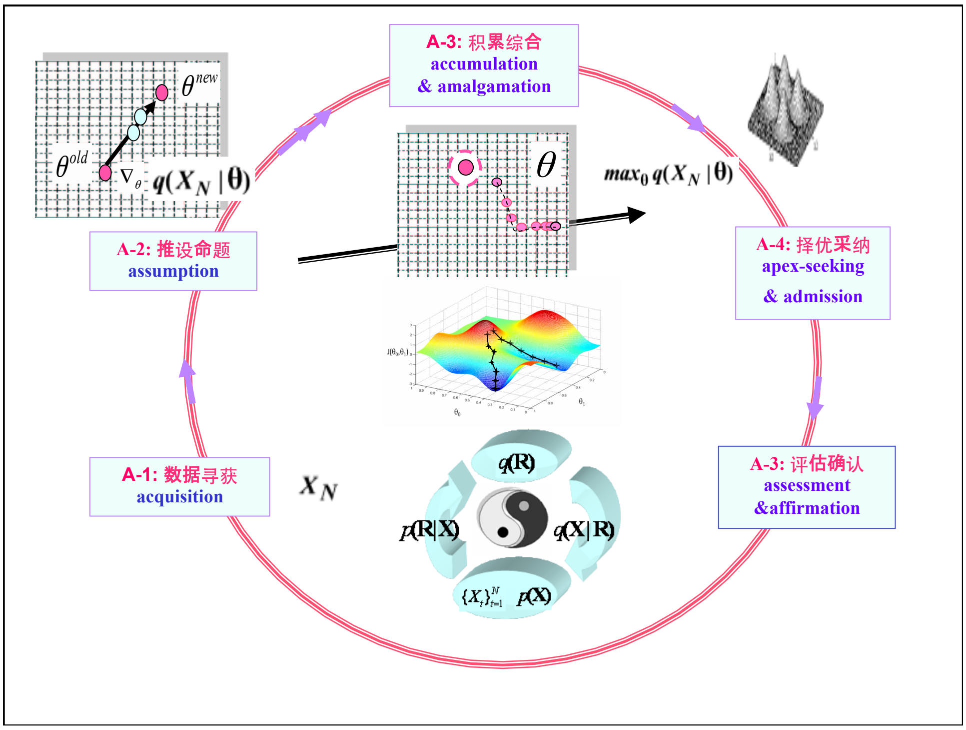


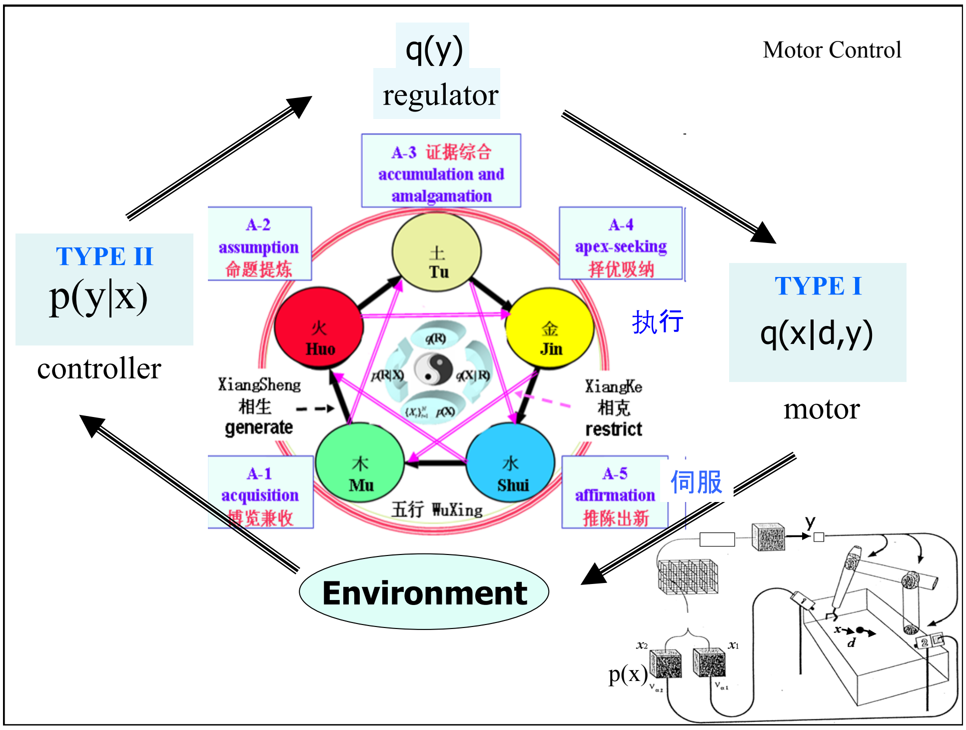


问题求解之A5框架 (since 2003)

A unified perspective and new results on RHT computing, mixture based learning, and multi-learner based problem solving☆

Pattern Recognition 40 (2007) 2129–2153 Lei Xu*


A-3: 积累综合
accumulation
& amalgamation
θ^{new}
θ^{old}
$\nabla_\theta\, q(X_N\,|\,\theta)$
θ
$max_\theta\, q(X_N\,|\,\theta)$
A-2: 推设命题
assumption
A-4: 择优采纳
apex-seeking
& admission
A-1: 数据寻获
acquisition
X_N
q(R)
p(R|X)
q(X|R)
p(X)
A-3: 评估确认
assessment
&affirmation


q(y)
regulator
Motor Control
A-3 证据综合
accumulation and
amalgamation
A-2
assumption
命题提炼
A-4
apex-seeking
择优吸纳
TYPE II
p(y|x)
controller
土
Tu
火
Huo
金
Jin
执行
TYPE I
q(x|d,y)
motor
XiangSheng
相生
generate
XiangKe
相克
restrict
木
Mu
水
Shui
A-1
acquisition
博览兼收
五行 WuXing
A-5
affirmation
推陈出新
伺服
Environment
y
p(x)

参见以下综述

Lei Xu (2012), " On essential topics of BYY harmony learning: Current status, challenging issues, and gene analysis applications", A special issue on Machine learning and intelligence science: IScIDE (C), Journal of Frontiers of Electrical and Electronic Engineering 7(1) (2012) 147–196.

Lei Xu (2011), "Another perspective of BYY harmony learning: representation in multiple layers, co-decomposition of data covariance matrices, and applications to network biology. A special issue on Machine learning and intelligence science: IScIDE2010 (A), Journal of Frontiers of Electrical and Electronic Engineering in China 6(1) (2011) 86–119.

Lei Xu (2010), "Bayesian Ying-Yang system, best harmony learning, and five action circling", A special issue on Emerging Themes on Information Theory and Bayesian Approach, Journal of Frontiers of Electrical and Electronic Engineering in China, 5(3):281–328, 2010.

论坛上报告结束后，没有时间让我来回答一个质疑：几千年前的老子需要一个现代的解释吗？
字面上，除了老子本人，这是一个没有别人能给出答案的问题。
以下三点或许能令人反思，这样问题是否有意义？

(a) 我已在论坛上回应，西方科学以bottom-up方式寻求具有普适性和恒常性的依赖关系，而阴阳五行原理以top-down方式寻求具有普适性和恒常性的依赖关系，不只是个哲学（哲学的定义是对自然与社会的基本看法，不需接受普适性和恒常性的检验）。比西方科学缺少的是，没有定量描述，只停留在定性描述。但是，这不意味不能够进行定量研究，或者后人不应该去研究。东西方学问不是对立的，而是互补的。
(b) 论坛上我在报告中已介绍，我的努力是基于现代概率论、信息论，进行定量研究，并有解决多个典型学习问题的应用结果为支撑。
(c) 目前学习领域里称之为Bayes学习，Boltzmann学习，Helmotz学习的方法和理论都不同于Bayes，Boltzmann，Helmotz的原来研究，只是继承了他们的基本精神。他们才作古百来二三百年，要不要去问他们，需不需要当今的这些解释呢？

的确，过去现在中国都不乏以迷信的方式推崇阴阳五行说。但是，并不等于阴阳五行全是糟粕，搞科学的人就不能碰。若要碰的话，客气地说也是‘雷人’。

巧了，本人姓徐名雷。不过，不是所谓‘雷人’的雷，而是一个受西方科学方式教育而后从事研究工作三十多年的学者。发表期刊論文百余篇，論文之被引用總量，據SCI-citations逾4500（其中前十篇的被引用量之和逾2500 、最大单篇为937）。據Google Scholar逾10000(其中前十篇的被引用量之和近6000 、最大单篇为2086)。

或许也可以反问，具有这样研究经历的学者，有无资格以科学研究方式去探究我国古人的经典？而对我进行阴阳五行研究发出疑问者，又有多少理据支撑？对经典的阴阳五行说知多少？对我的研究内容知多少？