

第四届中国数据挖掘学术会议

特邀报告

跨媒体数据挖掘和理解

史忠植

智能信息处理重点实验室
中国科学院计算技术研究所

2011.5.7

史忠植 跨媒体数据挖掘和理解



中国科学院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

致谢

- **973** 国家重大基础研究计划

No.2007CB311004

- 国家自然科学基金项目 No. 61035003

内容提要

引言

认知模型CAM

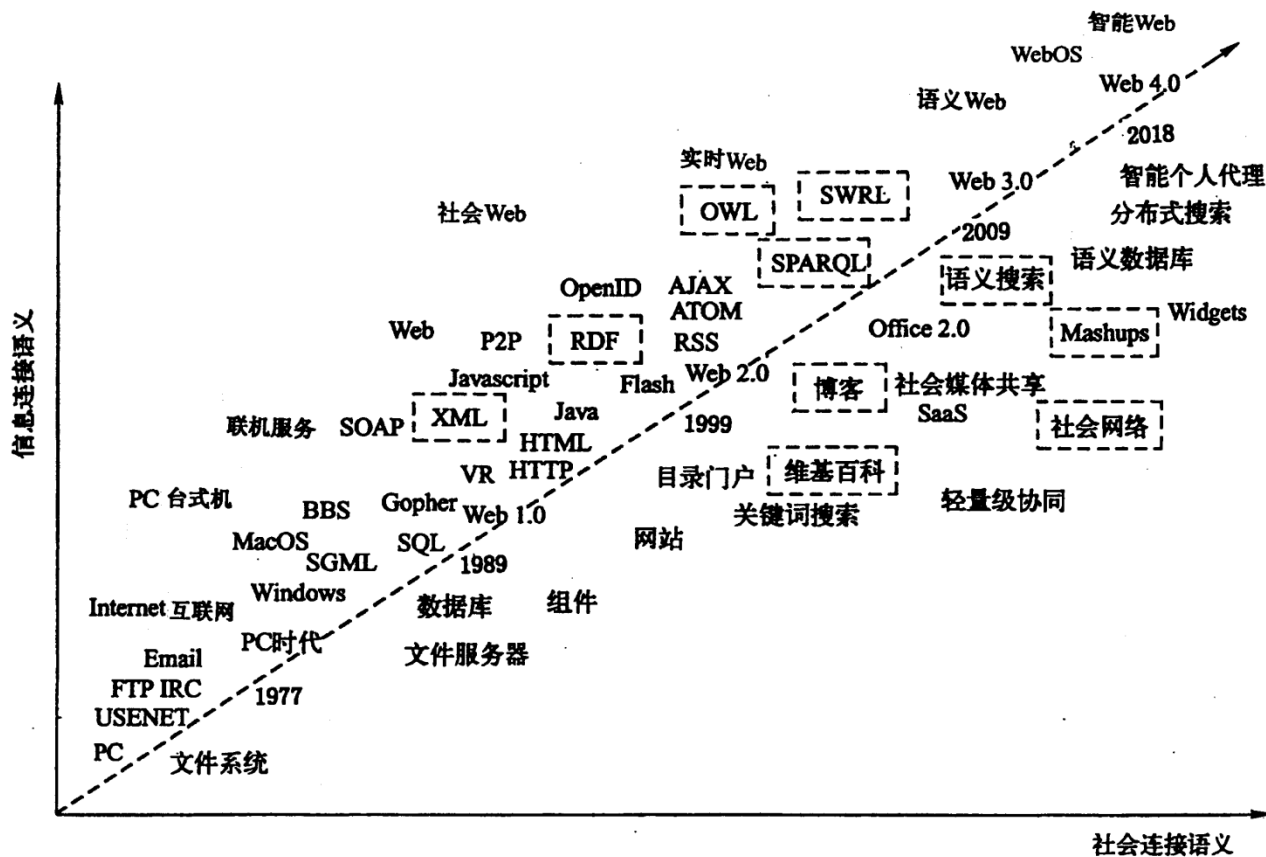
跨媒体语义分析

跨媒体语义关联

跨媒体语义理解

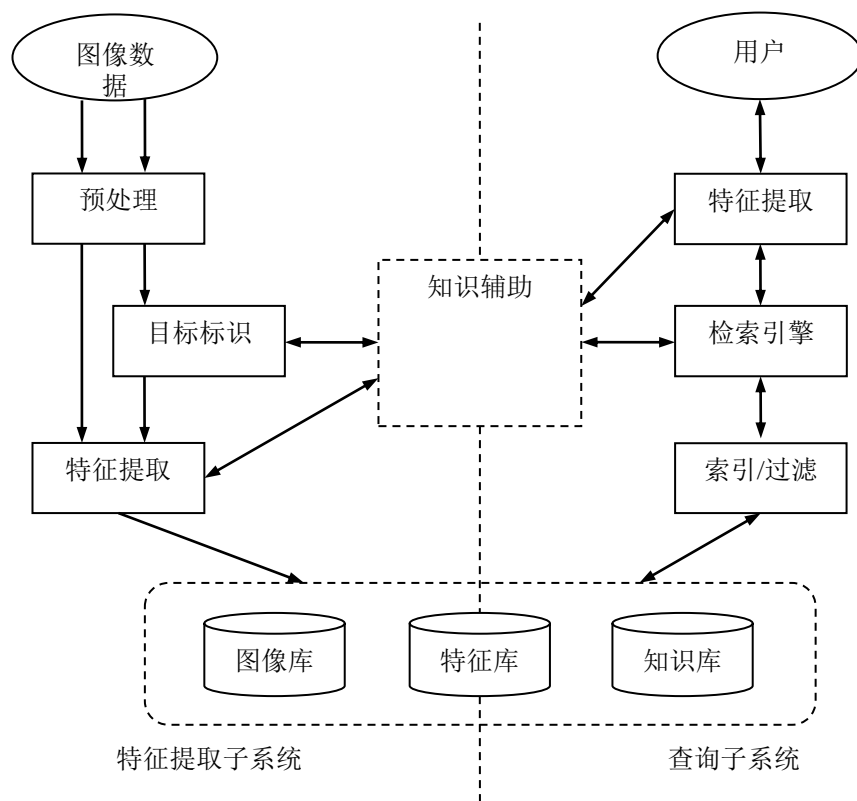
结束语

互联网发展路线图

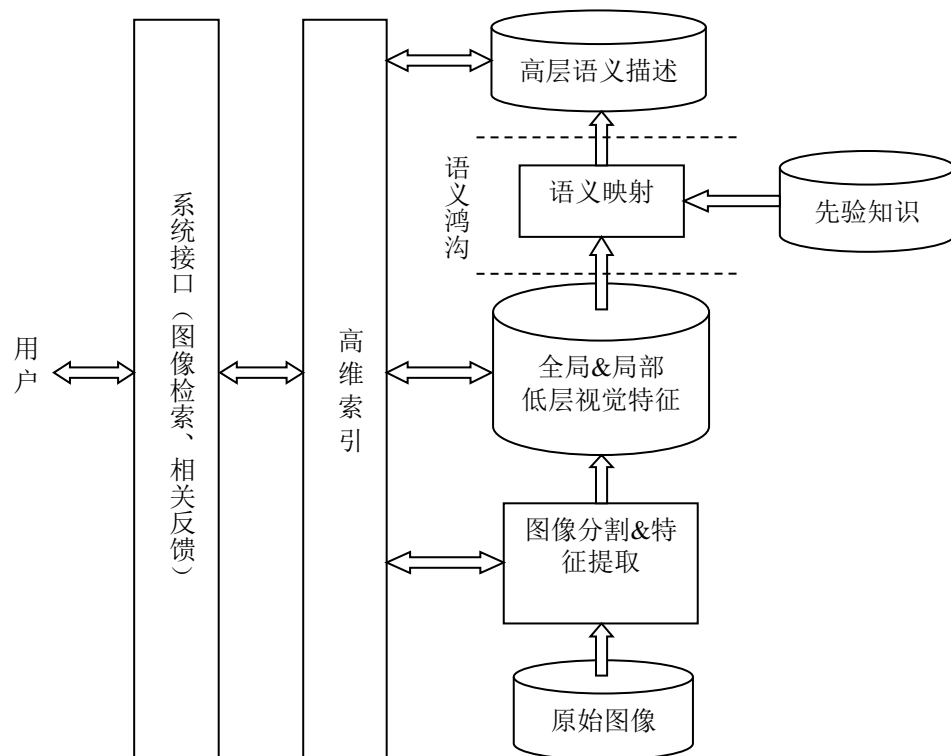


Spivack 2008

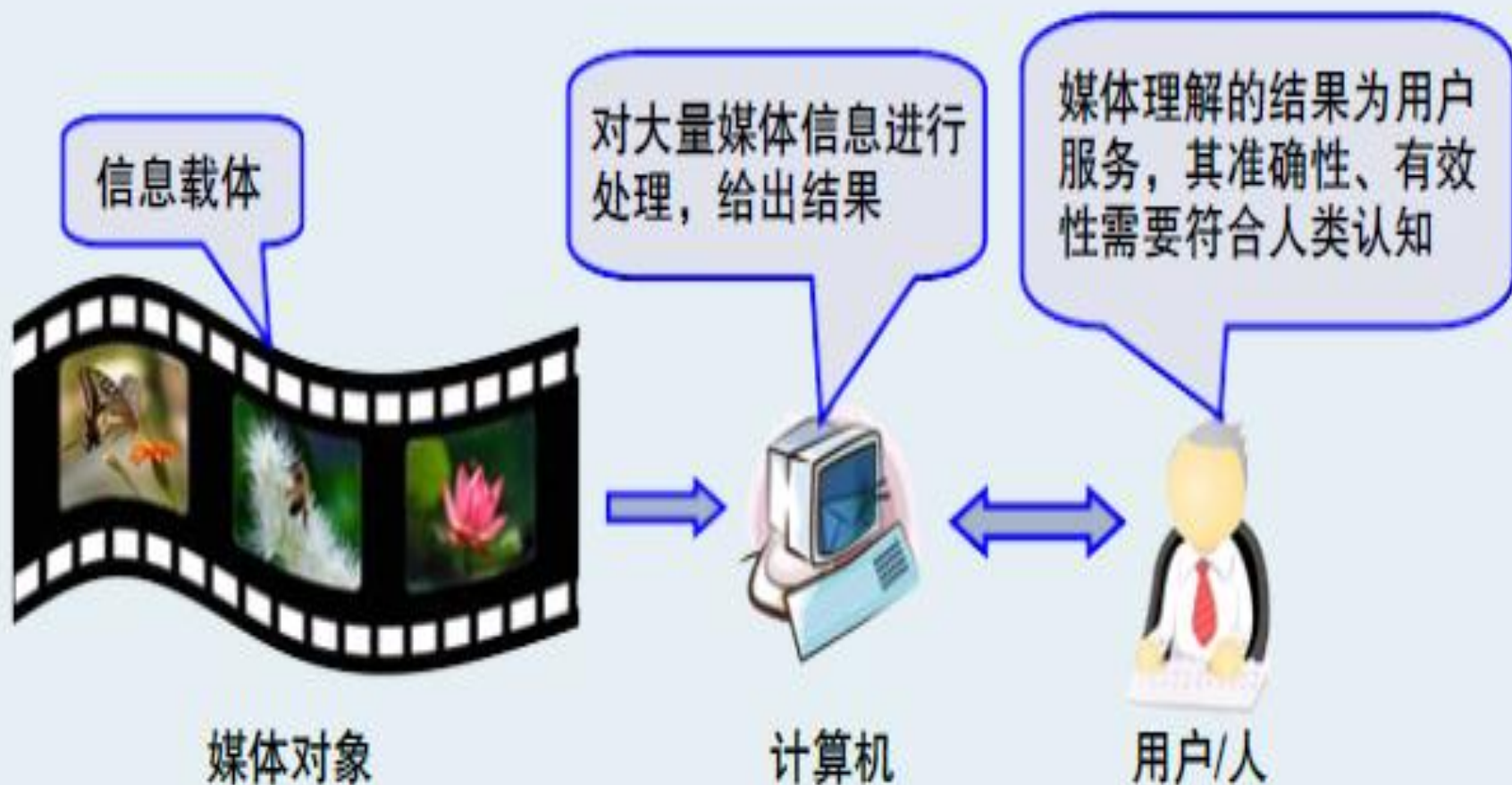
视频图像基于内容检索



视频图像基于语义检索



跨媒体挖掘



内容提要

引言

认知模型CAM

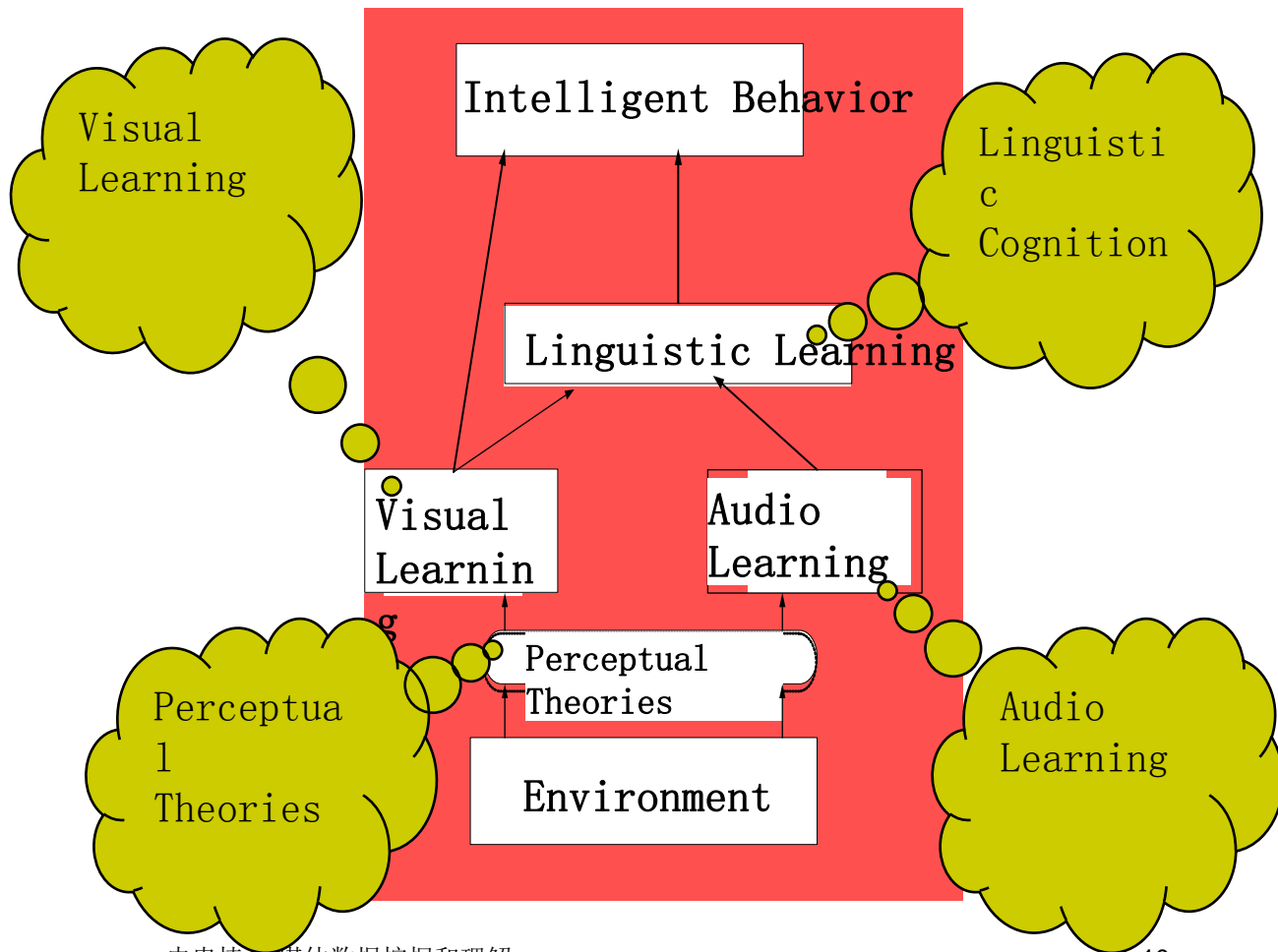
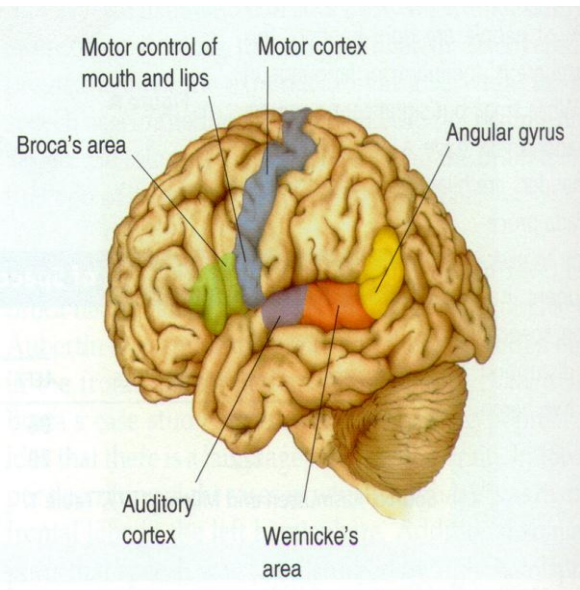
跨媒体语义分析

跨媒体语义关联

跨媒体语义理解

结束语

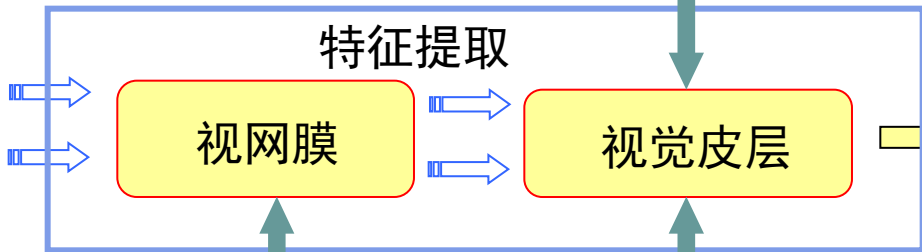
感知学习



视觉感知信息处理

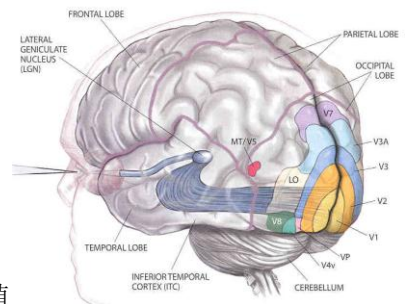
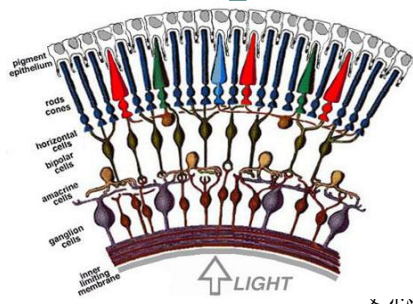


视觉信息
(非结构化)

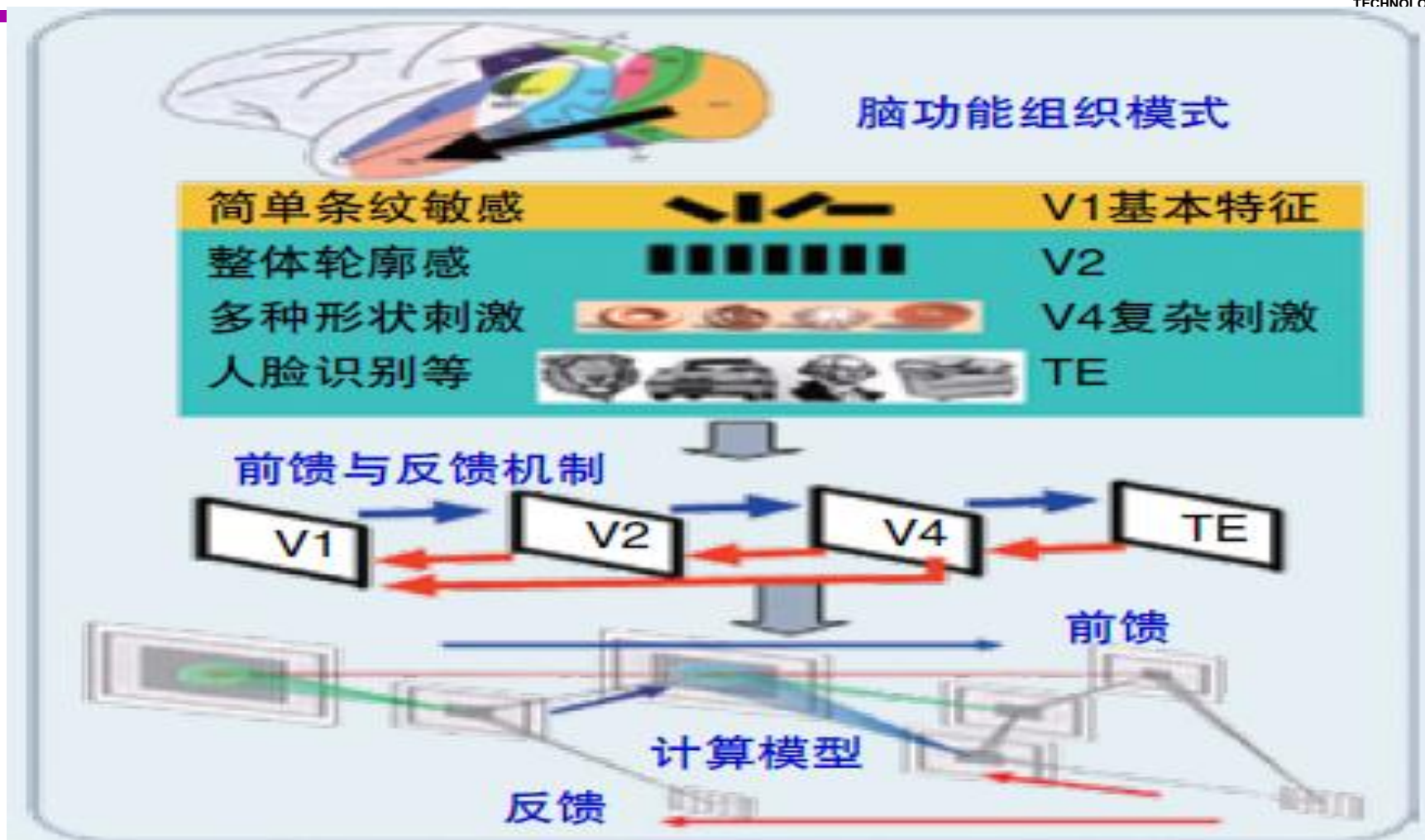


语义描述 (结构化)

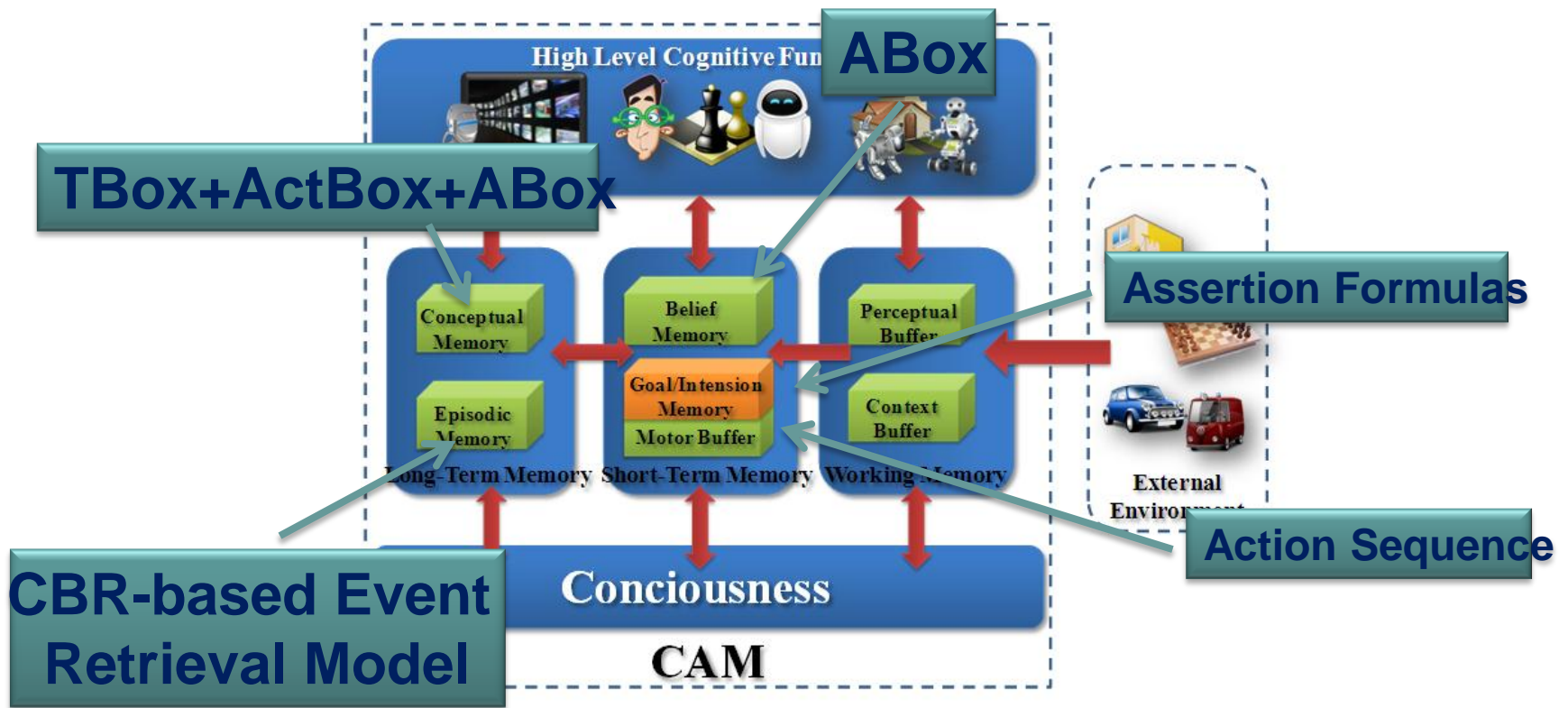
一只虎



视觉感知认知过程

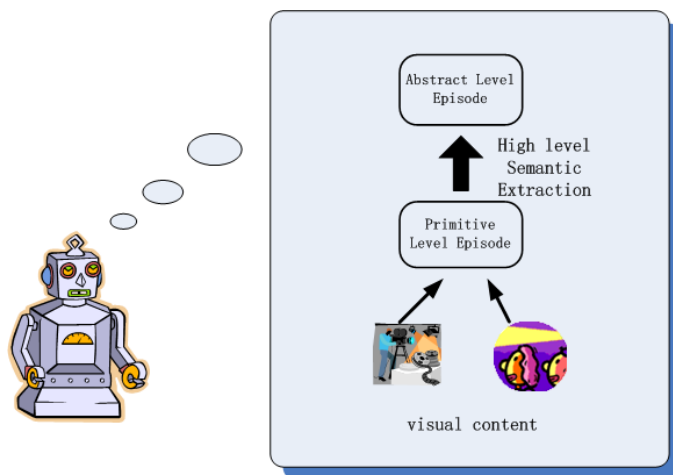


认知模型CAM



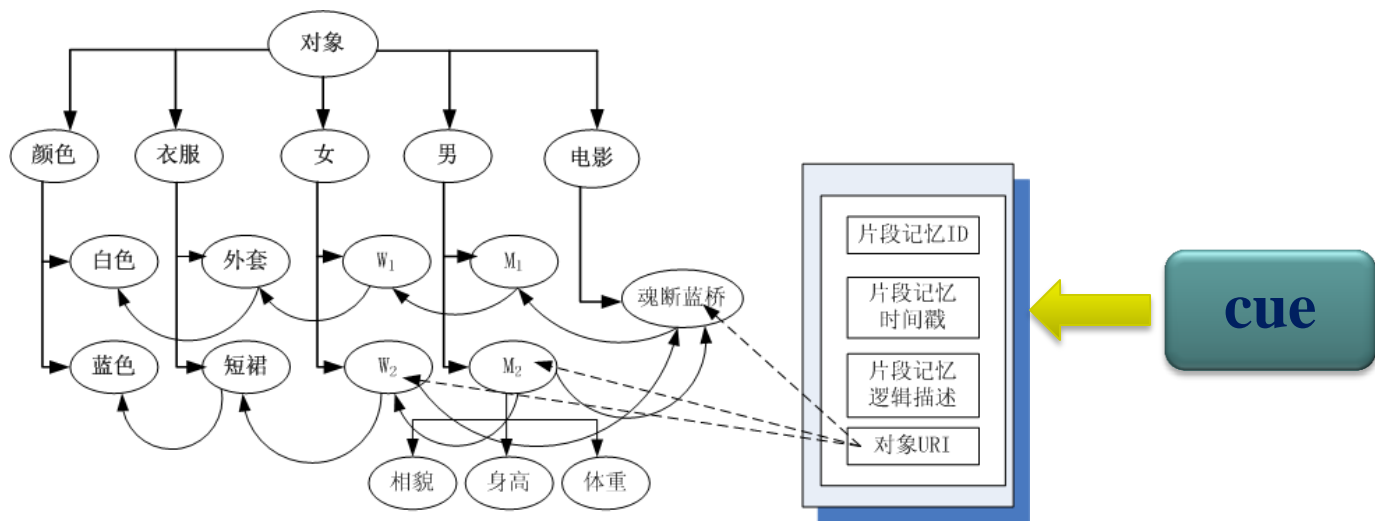
■ Episodic Memory(情景记忆)

存储了具有上下文的长期信息。典型的情景记忆包括个体过去所经历的事件，如昨天看过的某场电影，或者上个暑假参加的某项旅游活动等。



■ 信息存储

以记忆片段(episode)为存储单元,记忆片段包括了关于记忆的符号描述以及与符号关联的感知信息。我们以RDF graph形式对与记忆片段相关的对象,感知信息等进行了描述,表示以及存储。



案例推理



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY

检索RETRIEVE

find similar
problems

存储RETAIN

integrate in
case-base

CBR

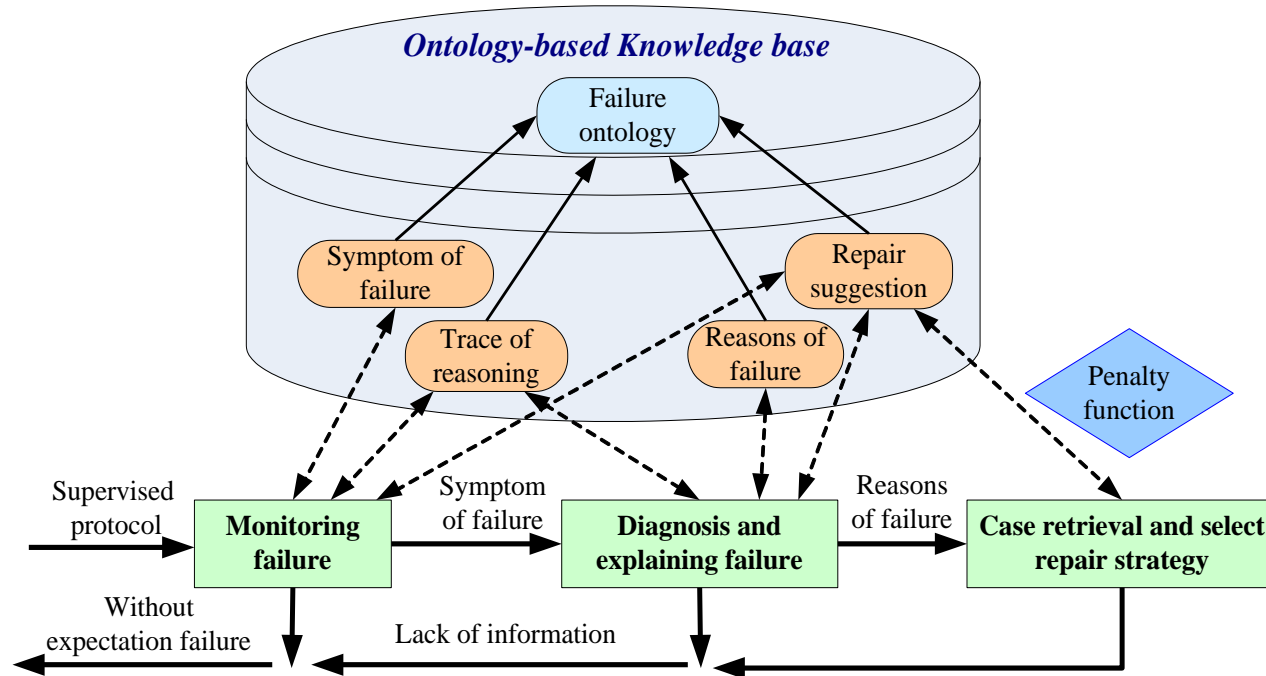
重用REUSE

propose solutions
from retrieved cases

修改REVISE

adapt and repair
proposed solution

内省学习



内容提要

引言

心智模型CAM

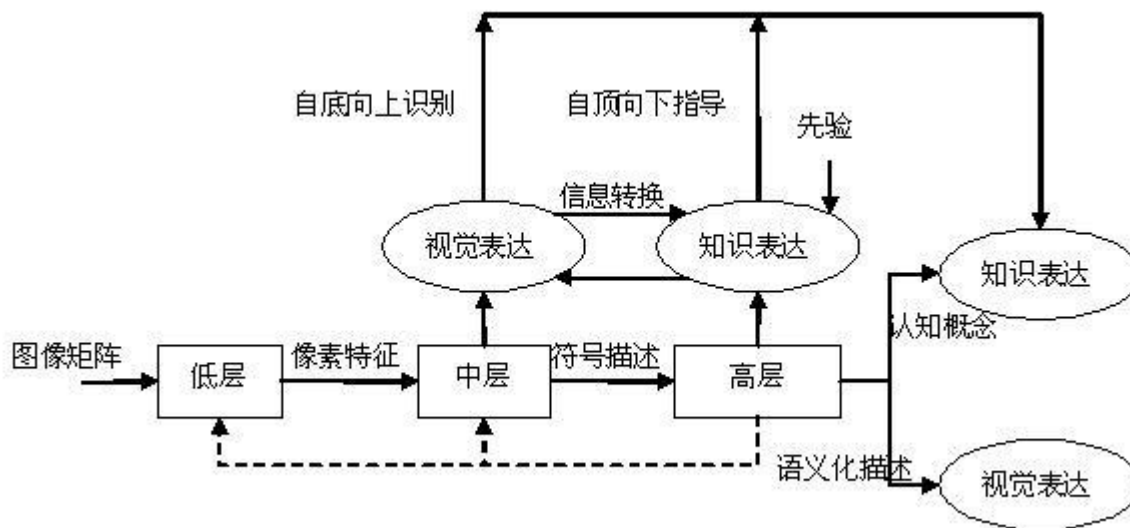
跨媒体语义分析

跨媒体语义关联

跨媒体语义理解

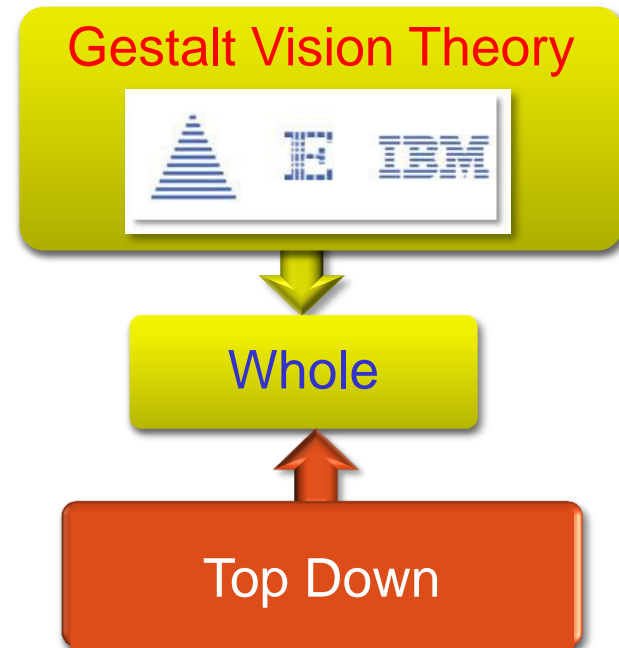
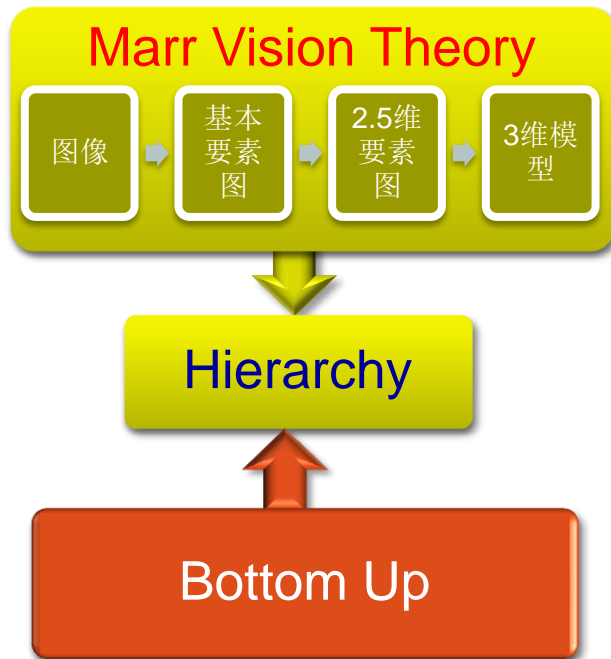
结束语

视频图像处理的三层模型

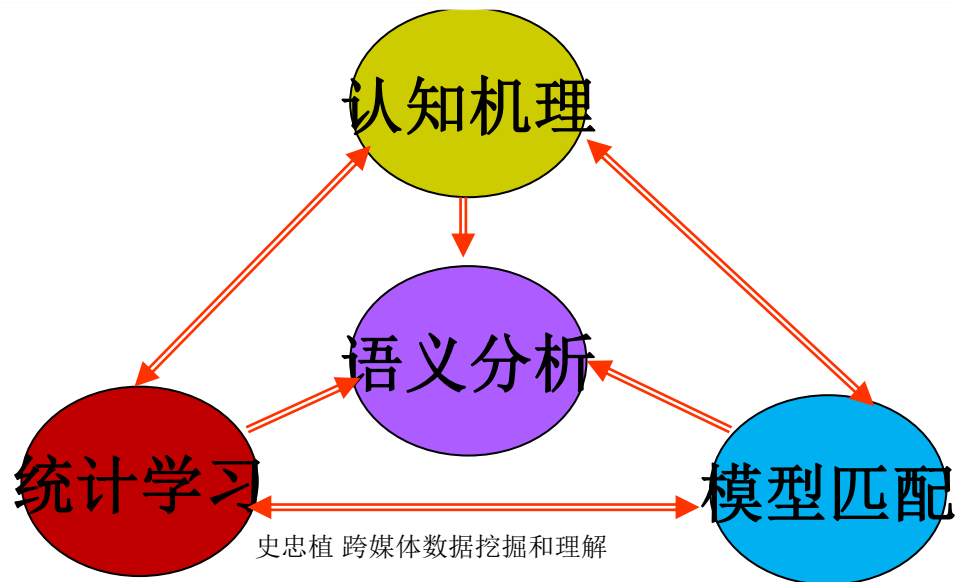
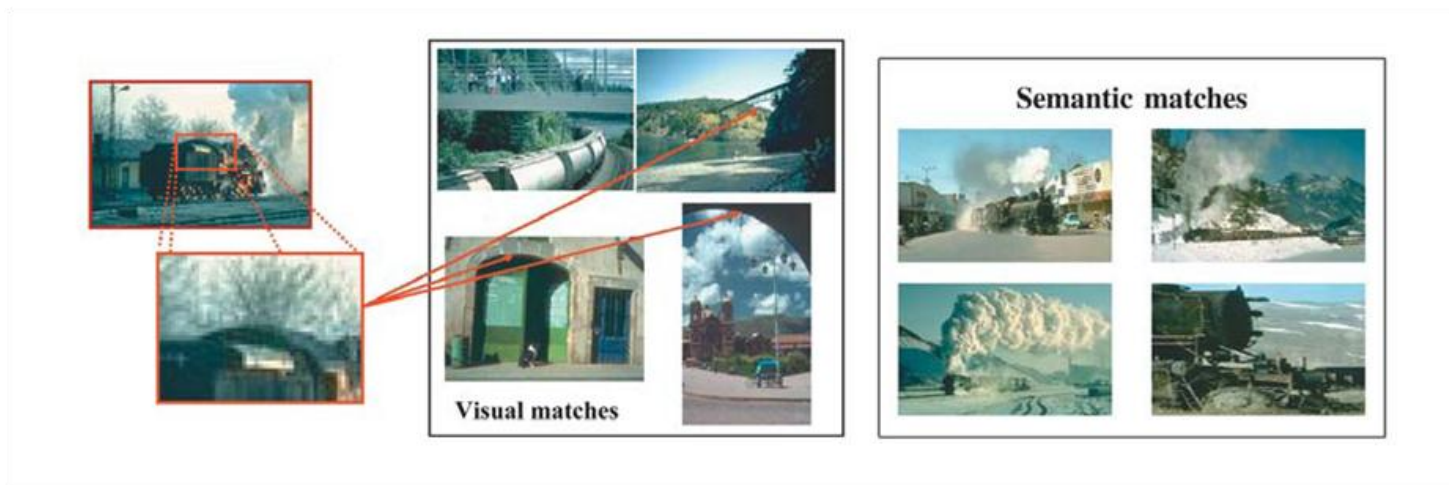


- 视频图像处理中视觉信息的表示、低层特征分析、中层目标识别以及高层语义分析等问题都还需要进一步研究。

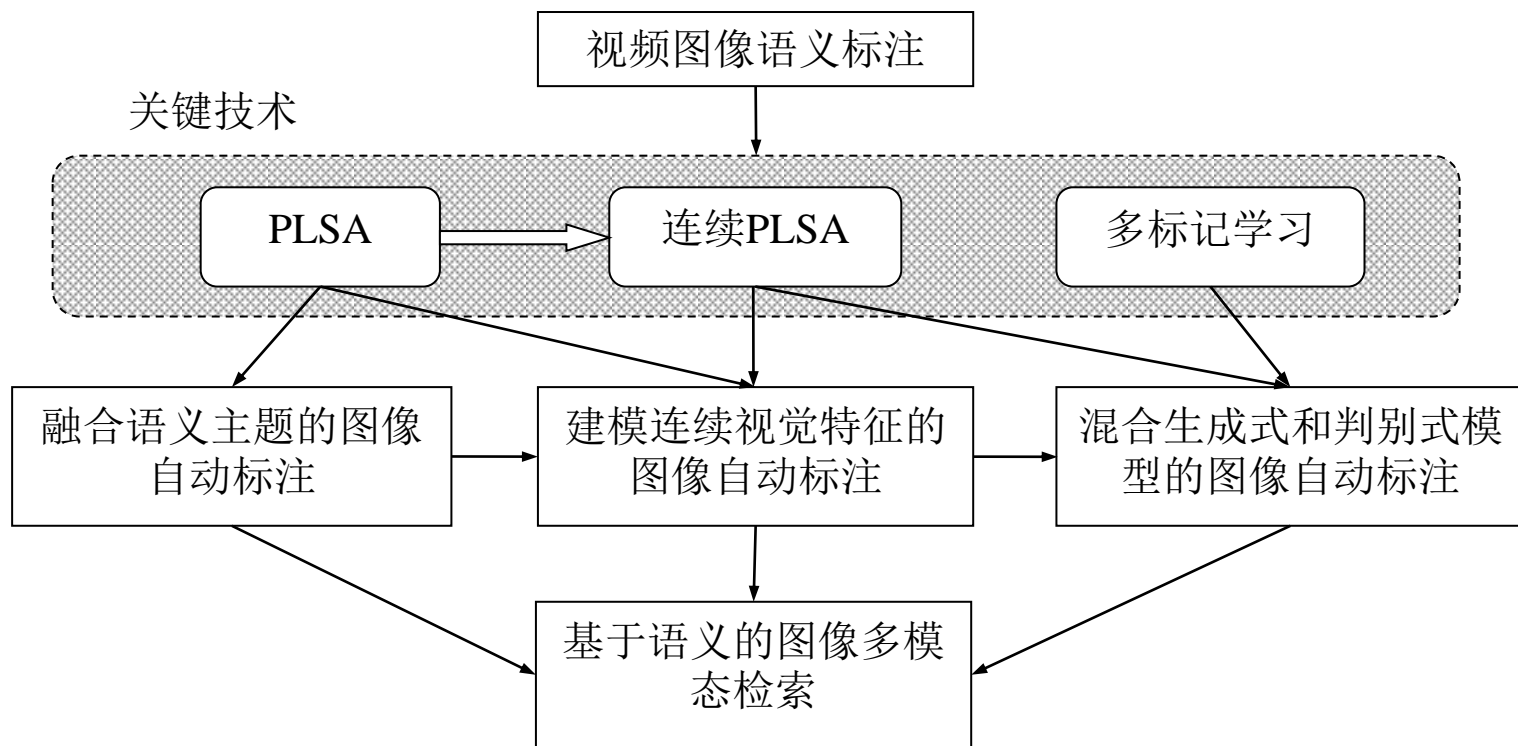
视觉计算



图像语义分析



图像语义分析

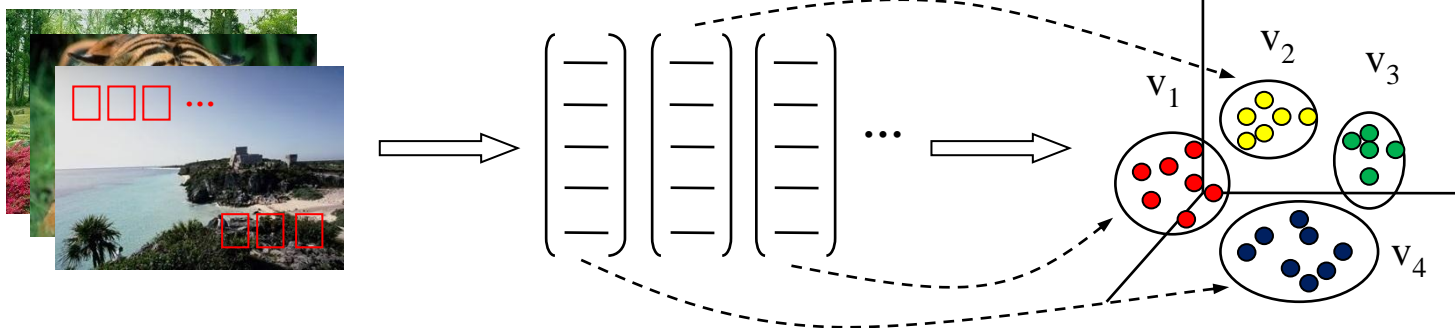


图像词袋表示

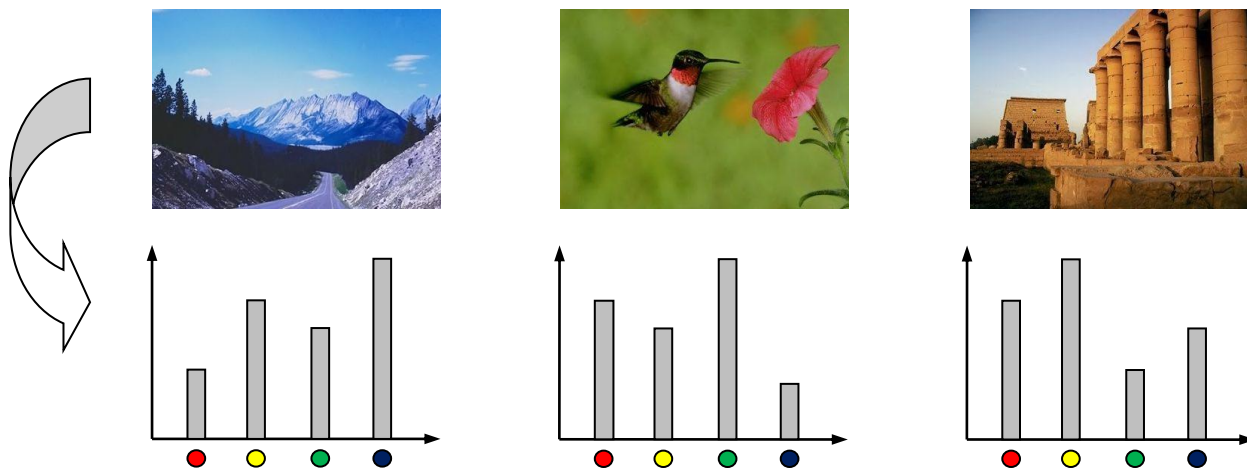
区域检测

特征提取

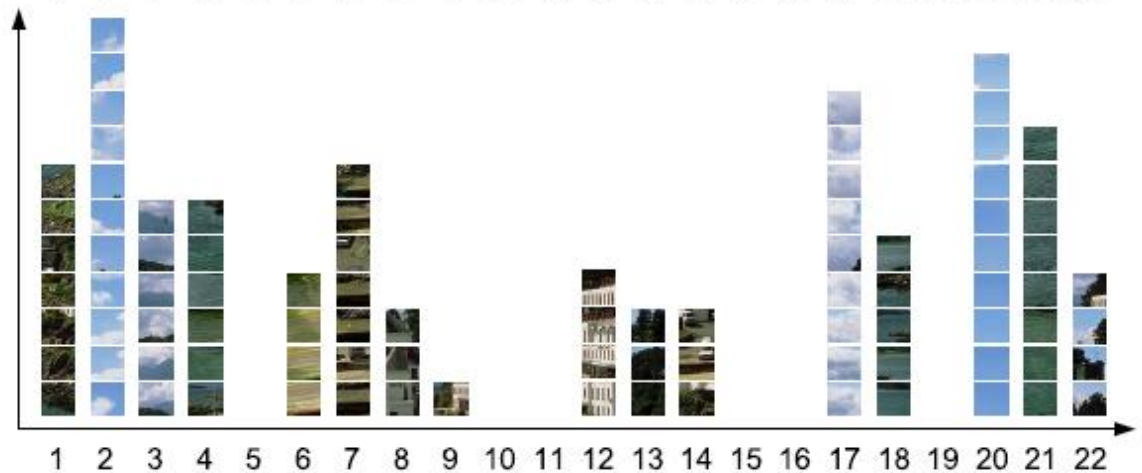
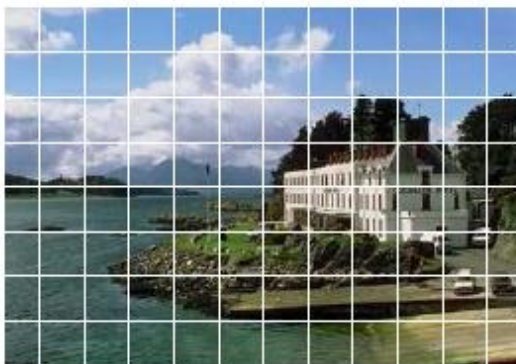
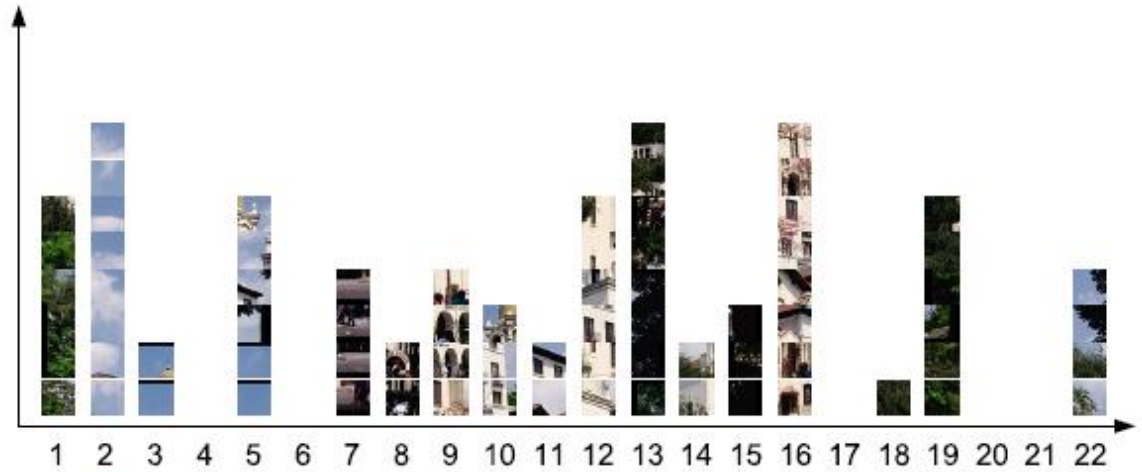
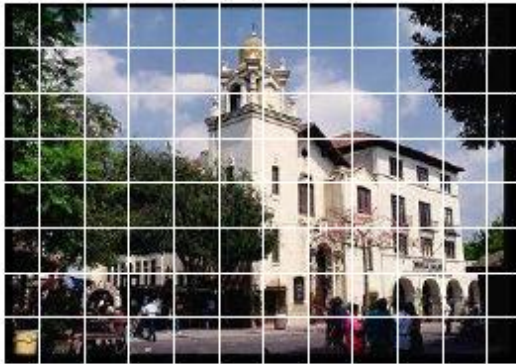
向量量化



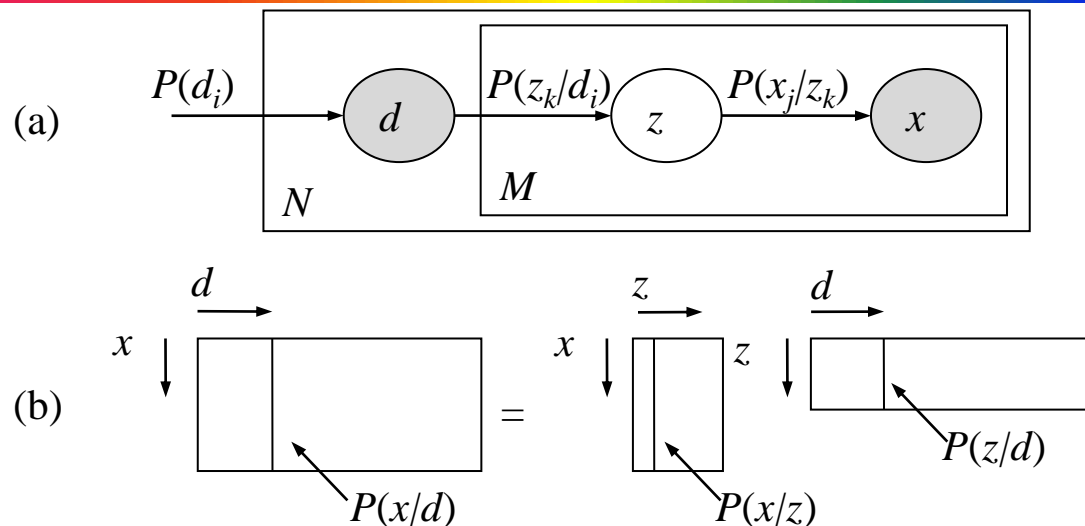
图像的词袋表示



图像词袋表示



概率潜在语义分析PLSA



- PLSA是Hofmann提出的最早的主题模型，它的初衷是为改进基于线性代数的方法LSA(latent semantic analysis)[DDF+90]。PLSA提出了对文本集进行主题建模的基本思想，可以将一个文本文档建模为若干潜在主题的混合，而每个主题都表示为一个词的多项分布。
- PLSA的图模型表示如图(a)所示。使用PLSA模型可以将一个文档表示为一个对应于主题分布的 K 维向量，这等价于图(b)中所示的矩阵分解。
- PLSA的模型参数是两个条件概率分布 $P(x/z)$ 和 $P(z/d)$ ，这两个参数都满足多项分布。可以通过EM算法或EM算法的不完全版本——folding-in算法来估计模型参数。

概率潜在语义分析PLSA

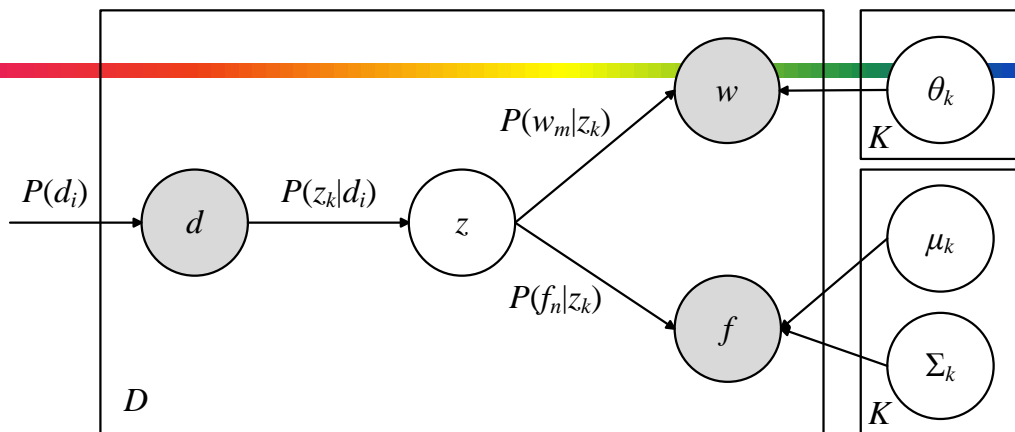
- 使用PLSA进行图像自动标注的基本原理是：
 - 首先学习训练集中各个图像的主题分布 $P(z|d)$;
 - 然后依据这个参数进一步学习视觉词和文本词在给定各个语义主题下的分布 $P(v|z)$ 和 $P(w|z)$ 。由PLSA的条件独立假设，这两个分布**独立于具体的训练图像**，对于训练集之外的图像也是有效的。
 - 于是，给定一幅未知图像 d_{new} ，可以依据自动获取的视觉词表示 $v(d_{new})$ 和训练得到的参数 $P(v|z)$ ，使用folding-in算法计算该图像的主题分布 $P(z|d_{new})$ ，从而**计算概率 $P(w|d_{new})$** 并经过排序得到对应的标注关键词集合。

连续视觉特征的GM-PLSA

- GM-PLSA (Gaussian-multinomial PLSA) 根据不同模态数据各自的特点而对其分别进行处理，采用连续PLSA建模图像的视觉特征，采用标准的PLSA建模文本关键词，并通过共享相同的潜在主题分布进行关联。
- 每幅图像既可以视为视觉模态下的连续的高斯分量的混合，同时也可以视为在文本模态下的离散的关键词的混合。
- 由于GM-PLSA结合连续PLSA和标准PLSA建模图像训练集的视觉和文本信息，并采用不对称学习方法估计模型参数，故而能更准确地获取未知图像的语义信息。

李志欣, 施智平, 刘曦, 史忠植. 建模连续视觉特征的图像语义标注方法[J]. 计算机辅助设计与图形学学报.

GM-PLSA的图模型表示



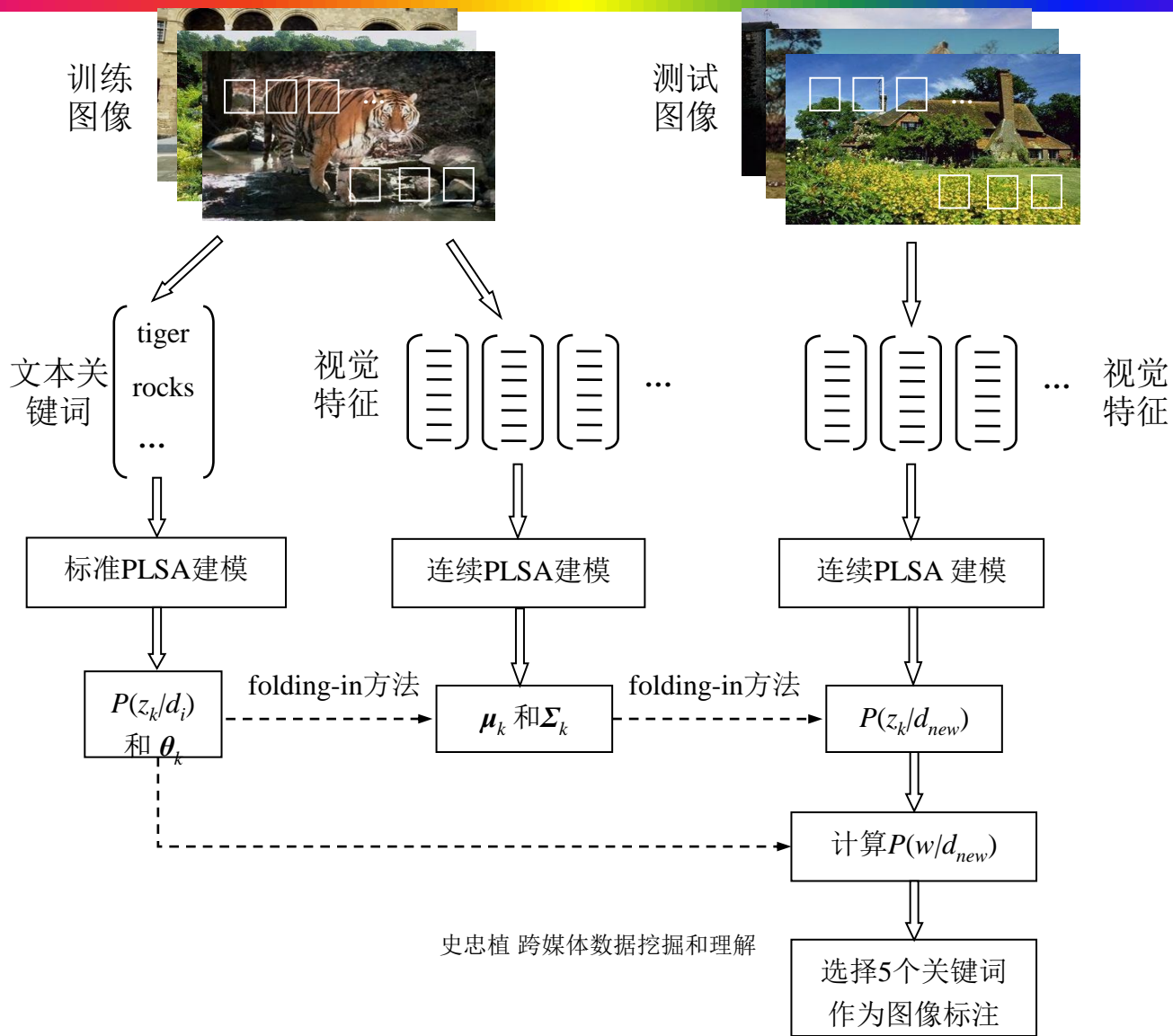
■ GM-PLSA可用下列的生成式过程描述：

- 以概率 $P(d_i)$ 选择一个文档 d_i ;
- 在给定文档 d_i 的条件下，以概率 $P(z_k|d_i)$ 采样满足多项分布的潜在主题 z_k ;
- 在给定潜在主题 z_k 的条件下，以多项分布 $\text{Mult}(\mathbf{x}|\boldsymbol{\theta}_k)$ 采样各个文本关键词 w_m ;
- 在给定潜在主题 z_k 的条件下，以高斯分布 $N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ 采样各个视觉特征向量 f_n 。

GM-PLSA的标注算法

- 为了估计未知的模型参数，采用了**不对称的学习算法**，因为不对称学习算法能在潜在空间的定义中更好的控制各个模态的数据产生的影响。
- 该算法在给定的图像文档中首先选择文本模态的数据来估计各个主题的混合分布，然后再根据主题分布估计视觉数据的混合高斯分布。这使得潜在空间被限制在文本模态数据上并**保持连续性**，同时也保持关联视觉模态数据的能力。

GM-PLSA的学习和标注过程



多标签分类

- 传统的单标记分类任务是将单一的标记 l (来自于标记的有限集合 L , $|L| > 1$) 赋给一个示例 \mathbf{x} 。一个单标记数据集 D 由 n 个训练样本 $(\mathbf{x}_1, l_1), (\mathbf{x}_2, l_2), \dots, (\mathbf{x}_n, l_n)$ 组成。
- 多标记分类任务将一个标记子集 $Y \subseteq L$ 赋给每一个示例。一个多标记数据集 D 则由下面 n 个训练样本组成: $(\mathbf{x}_1, Y_1), (\mathbf{x}_2, Y_2), \dots, (\mathbf{x}_n, Y_n)$ 。
- 多标记学习的方法可分为两类: 一类是“问题转化”, 将学习任务转化为一个或多个单标记分类任务; 另一类是“算法改进”, 直接对特定算法进行改进以处理多标记数据。
- BR(binary relevance) 是一种流行的问题转化方法, 它学习 $|L|$ 个分类器, 每一个对应于 L 中的一个标记。

多标签分类

性能评价指标

1. 汉明距离 (HammingLoss)
2. 准确度、精确度、查全率 (Accuracy, Precision, Recall)
3. $microF1$ 、 $macroF1$ 测度

$$F1_{micro} = \frac{2 * \frac{\sum_{i=1}^Q N_{i-true-pos} * \sum_{i=1}^Q N_{i-true-pos}}{\sum_{i=1}^Q N_{i-pos} * \sum_{i=1}^Q N_{i-true}}}{\frac{\sum_{i=1}^Q N_{i-true-pos}}{\sum_{i=1}^Q N_{i-pos}} + \frac{\sum_{i=1}^Q N_{i-true-pos}}{\sum_{i=1}^Q N_{i-true}}} \quad (10) \quad F1_{macro} = \frac{1}{Q} \sum_{j=1}^Q \frac{2p_j r_j}{p_j + r_j} \quad (11)$$

其中

$$p_i = \frac{N_{i-true-pos}}{N_{i-pos}} \quad (12) \quad r_i = \frac{N_{i-true-pos}}{N_{i-true}} \quad (13)$$

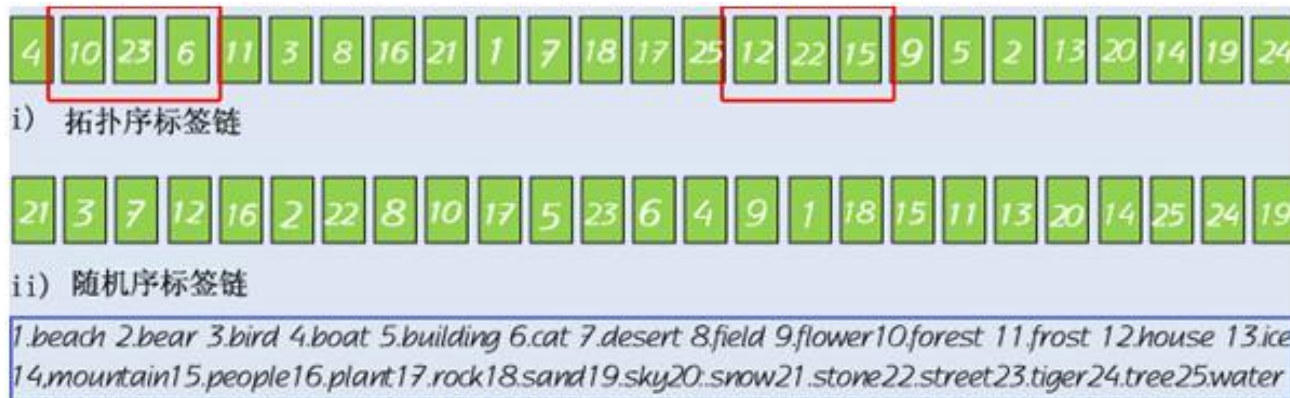
N_{i-true} 是测试集中标签 i 的正例个数, N_{i-pos} 是标签 i 预测的正例个数,
 $N_{i-true-pos}$ 是对标签 i 预测中正确的正例个数

基于全标注集的拓扑序标签链分类器方法



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY

所生的拓扑序标签链



Core1集上生成的标签链

10(森林,forest)-23(老虎,tiger)-6(猫,cat),
12(房子,house)-22(街道,street)-15(人,people)



MediaMi11集上生成的标签链

7(人脸,face)-1(主持人,anchor)-20(分屏,split screen),
8(政府大楼,government building)-4(法庭,court)-25(暴力,violence)

基于全标注集的拓扑序标签链分类器方法







中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY

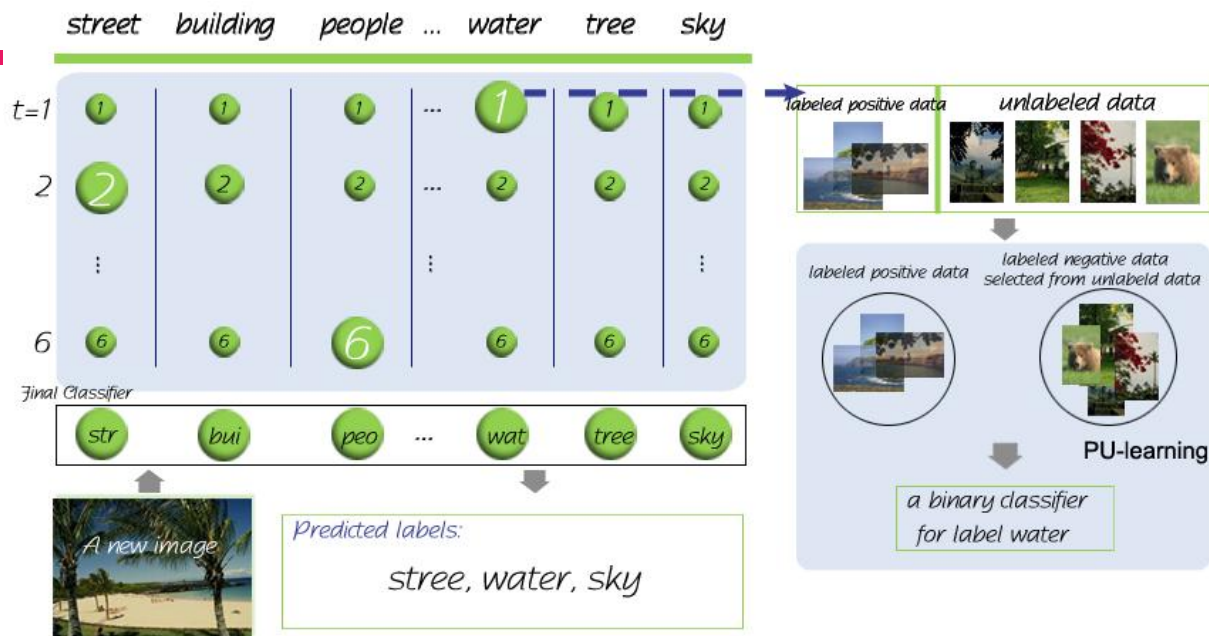
比较结果 在Core1集和MediaMill集上与其他方法比较

数据集 \ 方法		本文方法	ECC	BR	MLSSM
Core15k (25类)	$microF_1$	46.72	43.29	35.69	47.08
	$macroF_1$	41.82	39.54	26.16	40.14
MediaMill (25类)	$microF_1$	39.56	37.28	33.49	40.72
	$macroF_1$	34.99	30.36	25.42	32.33

标签预测示意图

输入图像				
真实标签	bear,ice,snow	bear,beach,water	tiger,cat,forest	field,mountain
(本文方法)	bear,ice,snow	bear,beach,water	cat,tiger,forest	field,mountain
ECC	bear,ice,snow, beach	bear,beach,water	cat,tiger,forest	field,mountain, rock
BR	beach ,bear,ice,snow, sand	beach,bear,water, people , bird	cat,tiger, field ,forest	desert ,mountain,rock, field
MLSSM	bear,ice,snow	bear, sky ,water, people	cat,tiger,forest, tree	field,mountain

基于弱标注集的模式共享分类方法



- ▶ 该框架采用Boosting框架训练若干轮，在每一轮中训练一个关于某标签的二类分类器，每一轮训练的标签分类器将根据标签之间的关联程度以对其他标签进行不同信任度的分类。
- ▶ 采用一种基于正例和无标注样例学习（Positive and Unlabeled data Learning, PU-learning）算法来训练二类分类器。

基于弱标注集的模型共享分类方法

弱标注学习算法Roc-SVM

▶ 将未标定样本集 U 中所有样本视为负例，正样本集 P 中所有样本视为正例，初始化一个负例集 $RN = \Phi$ 。

▶ 由公式(22)和(23)分别计算正例和负例模板

$$\bar{c}^+ = \alpha \frac{1}{|P|} \sum_{x \in P} \frac{x}{\|x\|} - \beta \frac{1}{|U|} \sum_{x \in U} \frac{x}{\|x\|} \quad (22) \quad \bar{c}^- = \alpha \frac{1}{|U|} \sum_{x \in U} \frac{x}{\|x\|} - \beta \frac{1}{|P|} \sum_{x \in P} \frac{x}{\|x\|} \quad (23)$$

▶ 按 $sim(x, c) = \frac{x \cdot c}{\|x\| \cdot \|c\|}$ 计算每个样本与正例和负例模板的相似度，如果与负例模板相似度更大，那么该样本认为是负样本，加入到 RN 中

▶ 如果 $|RN| > |P|$ ，随机地从 RN 中删除 $|RN| - |P|$ 个样本
否则，随机地从 P 中删除 $|P| - |RN|$ 个样本。

▶ 利用 P 和 RN 训练一个SVM分类器 h

基于弱标注集的模型共享分类方法

目标函数

- ▶ $D = \{x_1, x_2, \dots, x_n\}$ 是图像数据集，每幅图像是由 d 维向量表示且由一个二值向量 $y_i = (y_{i1}, y_{i2}, \dots, y_{in})^T$ 标注，其中 $y_{ik} = 1$ 表明第 i 个样本有第 k 个标签，但是 $y_{ik} = 0$ 不能说明第 i 个样本不含有第 k 个标签
- ▶ 借助于 LogitBoost 和 harmonic 函数的思想，我们可以得出如下目标函数：

$$F_{\text{cost}} = F_{LL} + \beta(F_{LU} + F_{UU}) \quad (18)$$

$$F_{LL} = \sum_{l=1}^L \frac{1}{N_{Ll}} \sum_{i=1}^{N_{Ll}} \log(1 + \exp(-y_{il}(H_l(x_i) + \alpha_l^t h^t(x_i)))) \quad (19)$$

$$F_{LU} = \sum_{l=1}^L \frac{1}{N_{Ll} \cdot N_{Ul}} \sum_{i=1}^{N_{Ll}} \sum_{j=1}^{N_{Ul}} S_{ij} \log(1 + \exp(-y_{il}(H_l(x_j) + \alpha_l^t h^t(x_j)))) \quad (20)$$

$$F_{UU} = \sum_{l=1}^L \frac{1}{N_{Ul} \cdot N_{Ul}} \sum_{i=1}^{N_{Ul}} \sum_{j=1}^{N_{Ul}} S_{ij} \log(1 + \exp(\frac{1}{2}(H_l(x_i) + \alpha_l^t h^t(x_i) - H_l(x_j) - \alpha_l^t h^t(x_j)))) \quad (21)$$

基于弱标注集的模型共享分类方法

模型共享弱标注学习MSWL算法

Alg5: 模型共享的弱标注学习算法

Input: 样本集 $D=\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 标签集 L , $|D|=n$, $|L|=m$,

$y_i=(y_{i1}, y_{i2}, \dots, y_{iL})$, $y_{il} \in \{0, +1\}$, $i=1, \dots, n$, 训练迭代轮数 T 。

Output: 最终的分类器 $y_i = \text{sign}[H_i(x)]$, $l=1, 2, \dots, m$ 。

Process:

1. 对所有样本的初始输出设为 0, $H_l(x)=0$, $l=1, 2, \dots, m$

2. 初始化一个候选模型池 $H_p=\{h_1(x), \dots, h_m(x)\}$, 其中模型 $h_l(x)$ 是利用 Roc-SVM 在标注正样本以及未标定样本上训练得到的关于标签 l 的 SVM 分类器

3. **for** $t=1, 2, \dots, T$

4. 找到最优的 $h' \in H_p$ 和对应的 $\alpha' = \{\alpha'_l\}, l=1, \dots, L$, 以最小化公式(18)

5. $H_l(x) = H_l(x) + \alpha'_l h'(x)$, $l=1, \dots, m$

6. 将被选定的 h' 从模型池 H_p 中删除, 同时在 h' 所对应的标签上再利用 Roc-SVM

训练一个新的 SVM 分类器 \bar{h} , 并加入到 H_p

7. **end**

$$\alpha'_l = -\left[\frac{\partial F_{\text{cost}}^2(\alpha'_l)}{\partial \alpha_l'^2}\right]^{-1} \frac{\partial F_{\text{cost}}(\alpha'_l)}{\partial \alpha_l'} \Big|_{\alpha_l'=0}$$

$$\frac{\partial F_{\text{cost}}(\alpha'_l)}{\partial \alpha_l'} \Big|_{\alpha_l'=0} = \frac{1}{N_L} \sum_{i=1}^{N_L} \frac{y_{il} h(x_i)}{1 + e^{y_{il} h(x_i)}} - \beta \frac{1}{N_L} \frac{1}{N_u} \sum_{i=1}^{N_L} \sum_{j=1}^{N_u} S_{ij} \frac{y_{il} h(x_j)}{1 + e^{y_{il} H_l(x_j)}} - \beta \frac{1}{N_{ul}} \frac{1}{N_{ul}} \sum_{i=1}^{N_u} \sum_{j=1}^{N_u} S_{ij} \frac{(h(x_j) - h(x_i)) / 2}{1 + e^{(H_l(x_j) - H_l(x_i)) / 2}}$$

$$\frac{\partial F_{\text{cost}}^2(\alpha'_l)}{\partial \alpha_l'^2} \Big|_{\alpha_l'=0} = \frac{1}{N_L} \sum_{i=1}^{N_L} \frac{1}{2 + e^{y_{il} H_l(x_i)} + e^{-y_{il} H_l(x_i)}} + \beta \frac{1}{N_L} \frac{1}{N_u} \sum_{i=1}^{N_L} \sum_{j=1}^{N_u} S_{ij} \frac{1}{2 + e^{y_{il} H_l(x_j)} + e^{-y_{il} H_l(x_j)}} + \beta \frac{1}{N_u} \frac{1}{N_u} \sum_{i=1}^{N_u} \sum_{j=1}^{N_u} S_{ij} \frac{[(h(x_j) - h(x_i)) / 2]^2}{2 + e^{(H_l(x_j) - H_l(x_i)) / 2} + e^{-(H_l(x_j) - H_l(x_i)) / 2}}$$

基于弱标注集的模型共享分类方法

实验结果分析

数据集

COREL5K- 20个label 1500幅图像 MEDIAMILL – 20个label 1500幅图像

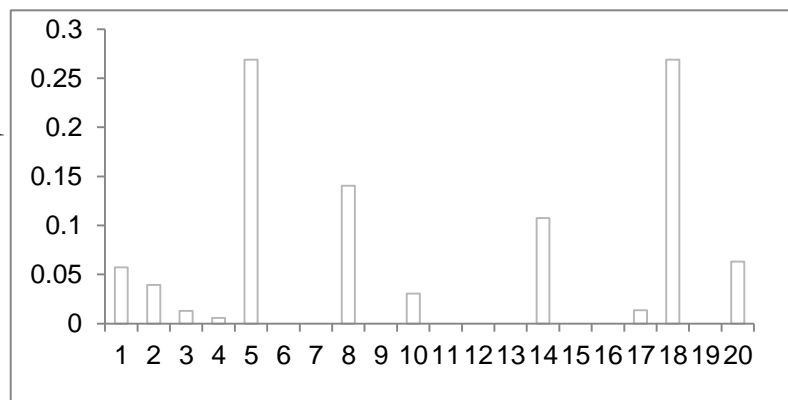
与其他方法比较

1) RS-BR方法 2) S-EM-BR方法 3) ML-KNN方法 4) MLSSM方法

采用评价指标: *microF1*、*macroF1*

方法自动发掘的标签关联:

5-cat, 8-forest, 14-rock, 18-tiger



基于弱标注集的模型共享分类方法



在Core1集上与其他方法比较

评价指标	标定百分比 p	方法				
		MSWL	RS-BR	S-EM-BR	ML-KNN	MLSSM
$microF_1$	10%	0.208	0.233	0.188	0.110	0.083
	20%	0.289	0.275	0.256	0.185	0.192
	30%	0.334	0.310	0.309	0.246	0.208
	40%	0.435	0.367	0.341	0.303	0.257
	50%	0.485	0.410	0.407	0.353	0.327
$macroF_1$	10%	0.192	0.194	0.167	0.071	0.064
	20%	0.266	0.256	0.209	0.162	0.162
	30%	0.301	0.244	0.233	0.199	0.186
	40%	0.387	0.302	0.296	0.245	0.225
	50%	0.422	0.341	0.332	0.324	0.305

基于弱标注集的模型共享分类方法



在MediaMill集上与其他方法比较

评价指标	标定百分比 p	方法				
		MSWL	RS-BR	S-EM-BR	ML-KNN	MLSSM
$microF_1$	10%	0.142	0.153	0.138	0.098	0.077
	20%	0.224	0.215	0.214	0.142	0.168
	30%	0.315	0.293	0.285	0.216	0.242
	40%	0.365	0.311	0.317	0.268	0.287
	50%	0.445	0.377	0.344	0.321	0.311
$macroF_1$	10%	0.101	0.114	0.098	0.086	0.055
	20%	0.179	0.176	0.173	0.128	0.133
	30%	0.287	0.252	0.249	0.205	0.216
	40%	0.348	0.292	0.288	0.266	0.256
	50%	0.382	0.336	0.329	0.302	0.281

SSIM更适合评价图像质量

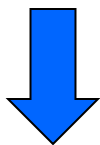


原图像和各类与原图像之间MSE均为210的失真图像。(a) 原图像；(b) 对比拉伸后的失真图像， $SSIM = 0.9168$ ；(c) 均值平移后的失真图像， $SSIM = 0.9900$ ；(d) JPEG 压缩后的失真图像， $SSIM = 0.6949$ ；(e) 图像模糊后的失真图像， $SSIM = 0.7052$ ；(f) 添加椒盐噪声后的失真图像， $SSIM = 0.7748$ 。根据主观感知的结果，显然SSIM更加适合评价图像质量。

基于结构相似度的稀疏编码模型

- 引入结构相似度到稀疏编码模型中，使得重构图像块尽量保持原图像块的结构信息，提出了基于结构相似度的稀疏编码模型（SS_SC: Sparse Coding Model Based on Structural Similarity）。

$$E(A, S) = \lambda_1 * \sum_{i=1}^N (I_i - Y_i)^2 + \lambda_2 * (1 - SSIM(I, Y)) + \lambda_3 * \sum_{i=1}^M S\left(\frac{a_i}{\sigma_i}\right)$$



$$E(A, S) = \lambda_1 * \sum_{i=1}^N (I_i - Y_i)^2 + \lambda_2 * \left(1 - \frac{(2\mu_I\mu_Y + C_1)(2\sigma_{IY} + C_2)}{(\mu_I^2 + \mu_Y^2 + C_1)(\sigma_I^2 + \sigma_Y^2 + C_2)} \right) + \lambda_3 * \sum_{i=1}^M S\left(\frac{a_i}{\sigma_i}\right)$$

李志清, 施智平, 李志欣, 史忠植. 基于结构相似度的稀疏编码模型. 软件学报.(已录用)

Zhiqing Li, Zhiping Shi, Xi Liu and Zhongzhi Shi. A Novel Sparse Coding Model Based On Structural Similarity. The 35th IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, 2010 (Accepted).

模型推导

$$\vec{\nabla}_{a_i} E(A, S) = \lambda_1 * \vec{\nabla}_{a_i} B_1 - \lambda_2 * \frac{B_{21} * B_{22}}{B_{23} * B_{24}} * \left(\frac{\vec{\nabla}_{a_i} B_{21}}{B_{21}} + \frac{\vec{\nabla}_{a_i} B_{22}}{B_{22}} - \frac{\vec{\nabla}_{a_i} B_{23}}{B_{23}} - \frac{\vec{\nabla}_{a_i} B_{24}}{B_{24}} \right) + \lambda_3 * \vec{\nabla}_{a_i} B_3$$

$$\vec{\nabla}_{\phi_{i,j}} E(A, S) = \lambda_1 * \vec{\nabla}_{\phi_{i,j}} B_1 - \lambda_2 * \frac{B_{21} * B_{22}}{B_{23} * B_{24}} * \left(\frac{\vec{\nabla}_{\phi_{i,j}} B_{21}}{B_{21}} + \frac{\vec{\nabla}_{\phi_{i,j}} B_{22}}{B_{22}} - \frac{\vec{\nabla}_{\phi_{i,j}} B_{23}}{B_{23}} - \frac{\vec{\nabla}_{\phi_{i,j}} B_{24}}{B_{24}} \right)$$

其中,

$$\vec{\nabla}_{a_i} B_1 = (-2) * \sum_{k=1}^N (I_k - Y_k) \phi_{k,i}$$

$$\vec{\nabla}_{\phi_{i,j}} B_1 = (-2) * (I_i - Y_i) * a_j$$

$$\vec{\nabla}_{a_i} B_{21} = (2 / N) * \mu_I * \sum_{k=1}^N \phi_{k,i}$$

$$\vec{\nabla}_{\phi_{i,j}} B_{21} = (2 / N) * \mu_I * a_j$$

$$\vec{\nabla}_{a_i} B_{22} = (2 / (N - 1)) * \sum_{k=1}^N ((I_k - \mu_I) * \phi_{k,i})$$

$$\vec{\nabla}_{\phi_{i,j}} B_{22} = (2 / (N - 1)) * (I_i - \mu_I) * a_j$$

$$\vec{\nabla}_{a_i} B_{23} = (2 / N) * \mu_Y * \sum_{k=1}^N \phi_{k,i}$$

$$\vec{\nabla}_{\phi_{i,j}} B_{23} = (2 / N) * \mu_Y * a_j$$

$$\vec{\nabla}_{a_i} B_{24} = (2 / (N - 1)) * \sum_{k=1}^N ((Y_k - \mu_Y) * \phi_{k,i})$$

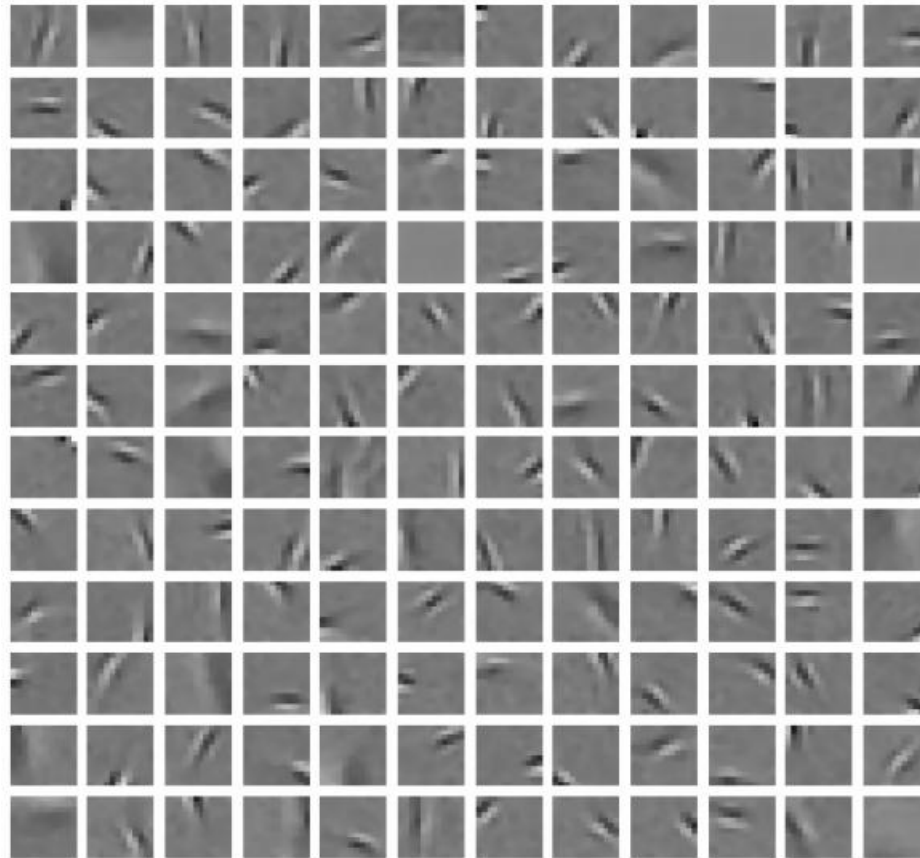
$$\vec{\nabla}_{\phi_{i,j}} B_{24} = (2 / (N - 1)) * (Y_i - \mu_Y) * a_j$$

优化学习算法

■ 获取基函数集合的算法.

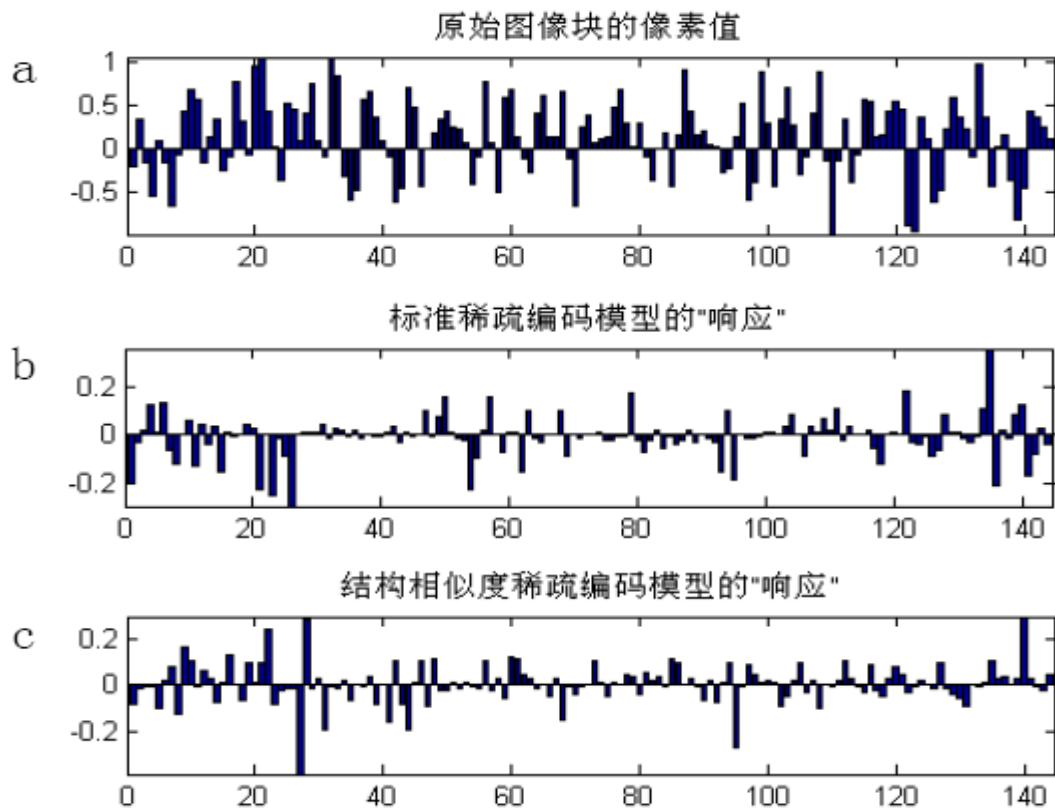
- Step 1. 对训练数据集进行白化低通滤波预处理;随机初始化基函数矩阵A;
- Step 2. 固定A,使用共轭梯度下降法和式(13)优化 $E(A, S)$,得到"响应" S ;
- Step 3. 固定 S ,使用简单梯度下降法和式(14)优化 $E(A, S)$,得到A;
- Step 4. 若训练次数达到设定值,得到训练好的基函数矩阵A; 否则,转入Step 2继续进行优化.

SS_SC的基函数



SS_SC模型学习得到的144个基函数，它们具有与视皮层简单细胞类似的性质，表现为良好的位置选择性、方向选择性以及空间频率选择特性。

稀疏响应



一个图像块的原始像素值和两种“响应”。无论是应用标准稀疏编码模型得到的“响应”，还是应用SS_SC模型得到的“响应”，都比图像块原始像素值的分布要稀疏得多。

模型性能比较

(方法, 图像块数目)	<i>Avg_SSim</i>	<i>Avg_Error</i>	<i>Avg_Sparseness</i>
(SC, 100)	0.8667	0.0793	8.0061
(SS_SC, 100)	0.9317	0.0675	8.1922
(SC, 500)	0.8563	0.0849	7.9467
(SS_SC, 500)	0.9236	0.0733	8.1138
(SC, 2000)	0.8574	0.0836	7.9988
(SS_SC, 2000)	0.9249	0.0723	8.1526
(SC, 10000)	0.8542	0.0842	8.0153
(SS_SC, 10000)	0.9231	0.0753	8.1600

注：SC指标准稀疏编码模型，SS_SC指基于结构相似度的稀疏编码模型；*Avg_SSim*表示平均结构相似度，*Avg_Error*表示平均误差平方和，*Avg_Sparseness*表示平均稀疏代价函数值。

内容提要

引言

心智模型CAM

跨媒体语义分析

跨媒体语义关联

跨媒体语义理解

结束语

相关性分析

- 典型相关性分析(Canonical Correlation Analysis. CCA 是一种标准的统计分析方法，最早应用于经济学、医学和气象等领域，
- CCA的核心思想是：寻找两个变量场 X ， Y 对应的两组基向量 W_x ， W_y ，使得 X ， Y 在 W_x ， W_y 上投影后的结果最大程度地保持与 X ， Y 之间的相关性一致。CCA方法已经被应用到Web图像标注领域。
- 由于典型相关性分析是建立在两个不同变量场所对应矩阵的基础上，因此，同样也适用于对图像与音频、音频与文本等跨媒体特征的相关性分析。

典型相关分析

- 设 $\mathbf{x}=(x_1, x_2, \dots, x_p)'$ 和 $\mathbf{y}=(y_1, y_2, \dots, y_q)'$ 是两组随机变量，且 $V(\mathbf{x})=\Sigma_{11}(>0)$ ， $V(\mathbf{y})=\Sigma_{22}(>0)$ ， $\text{Cov}(\mathbf{x}, \mathbf{y})=\Sigma_{12}$ ，即有

$$V\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

其中 $\Sigma_{21}=\Sigma_{12}'$ 。

- 研究 $u=\mathbf{a}'\mathbf{x}$ 与 $v=\mathbf{b}'\mathbf{y}$ 之间的相关关系，其中

$$\mathbf{a}=(a_1, a_2, \dots, a_p)', \quad \mathbf{b}=(b_1, b_2, \dots, b_q)'$$

现来计算一下 u 与 v 的相关系数。

$$\text{Cov}(u, v) = \text{Cov}(\mathbf{a}'\mathbf{x}, \mathbf{b}'\mathbf{y}) = \mathbf{a}'\text{Cov}(\mathbf{x}, \mathbf{y})\mathbf{b} = \mathbf{a}'\Sigma_{12}\mathbf{b}$$

$$V(u) = V(\mathbf{a}'\mathbf{x}) = \mathbf{a}'V(\mathbf{x})\mathbf{a} = \mathbf{a}'\Sigma_{11}\mathbf{a}$$

$$V(v) = V(\mathbf{b}'\mathbf{y}) = \mathbf{b}'V(\mathbf{y})\mathbf{b} = \mathbf{b}'\Sigma_{22}\mathbf{b}$$

典型相关分析

u 与 v 的相关系数

$$\rho(u, v) = \frac{\mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b}}{\sqrt{\mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a}} \sqrt{\mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b}}}$$

由于对任意非零常数 k_1 和 k_2 , 有

$$\rho(k_1 u, k_2 v) = \rho(u, v)$$

因此, 为避免不必要的结果重复, 我们常常限定 u 与 v 均为标准化的变量, 即附加约束条件

$$V(u) = 1, \quad V(v) = 1$$

即

$$\mathbf{a}' \boldsymbol{\Sigma}_{11} \mathbf{a} = 1, \quad \mathbf{b}' \boldsymbol{\Sigma}_{22} \mathbf{b} = 1$$

在此约束条件下, 求 $\mathbf{a} \in R^p$ 和 $\mathbf{b} \in R^q$, 使得

$$\rho(u, v) = \mathbf{a}' \boldsymbol{\Sigma}_{12} \mathbf{b}$$

达到最大。

跨媒体相关分析

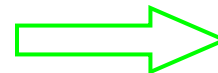
输入： 图像特征矩阵

$$X = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ x_{21}, x_{22}, \dots, x_{2p} \\ \dots \\ x_{n1}, x_{n2}, \dots, x_{np} \end{bmatrix}$$

声音特征矩阵

$$Y = \begin{bmatrix} y_{11}, y_{12}, \dots, y_{1q} \\ y_{21}, y_{22}, \dots, y_{2q} \\ \dots \\ y_{n1}, y_{n2}, \dots, y_{nq} \end{bmatrix}$$

X 与 Y 不同维数!



典型相关分析

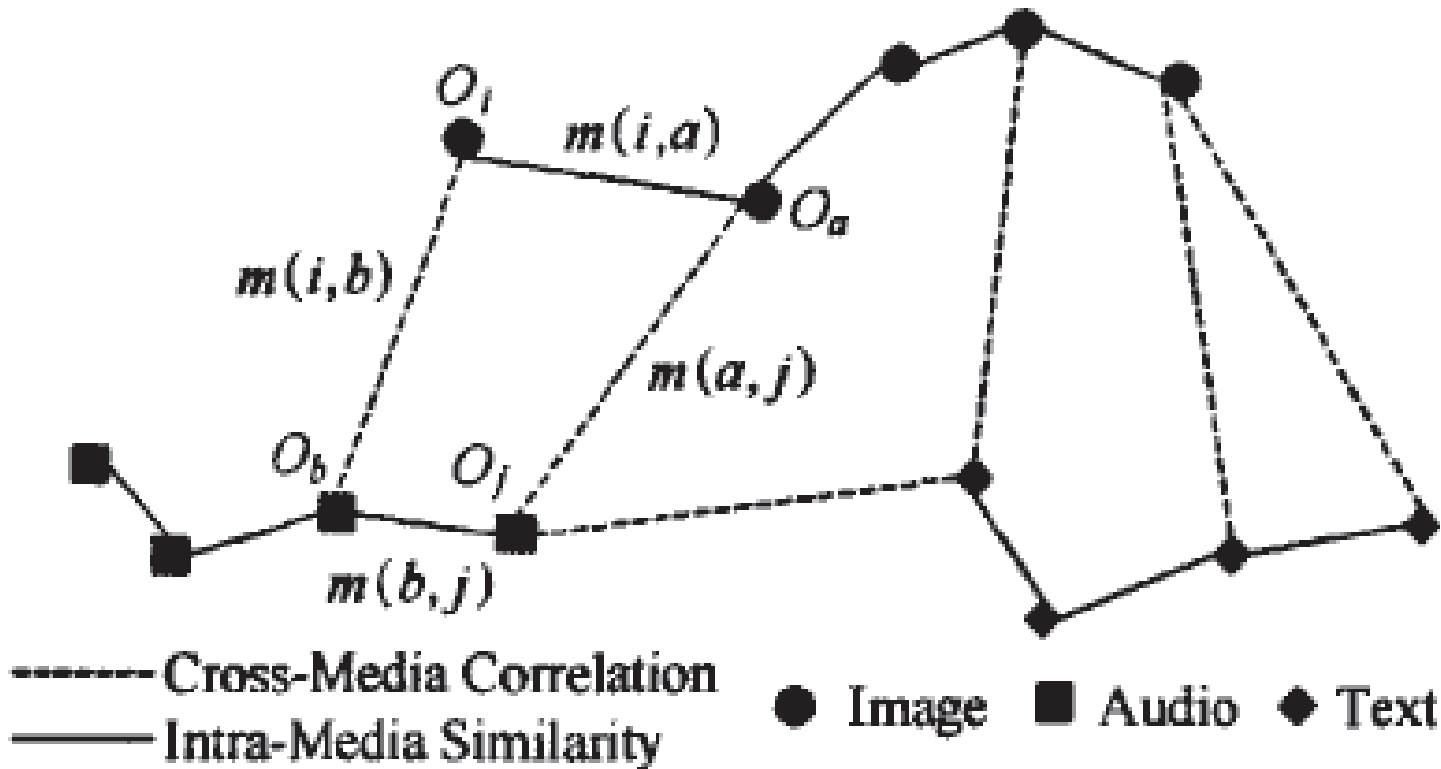
输出：

$$X' = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1m} \\ x_{21}, x_{22}, \dots, x_{2m} \\ \dots \\ x_{n1}, x_{n2}, \dots, x_{nm} \end{bmatrix}$$

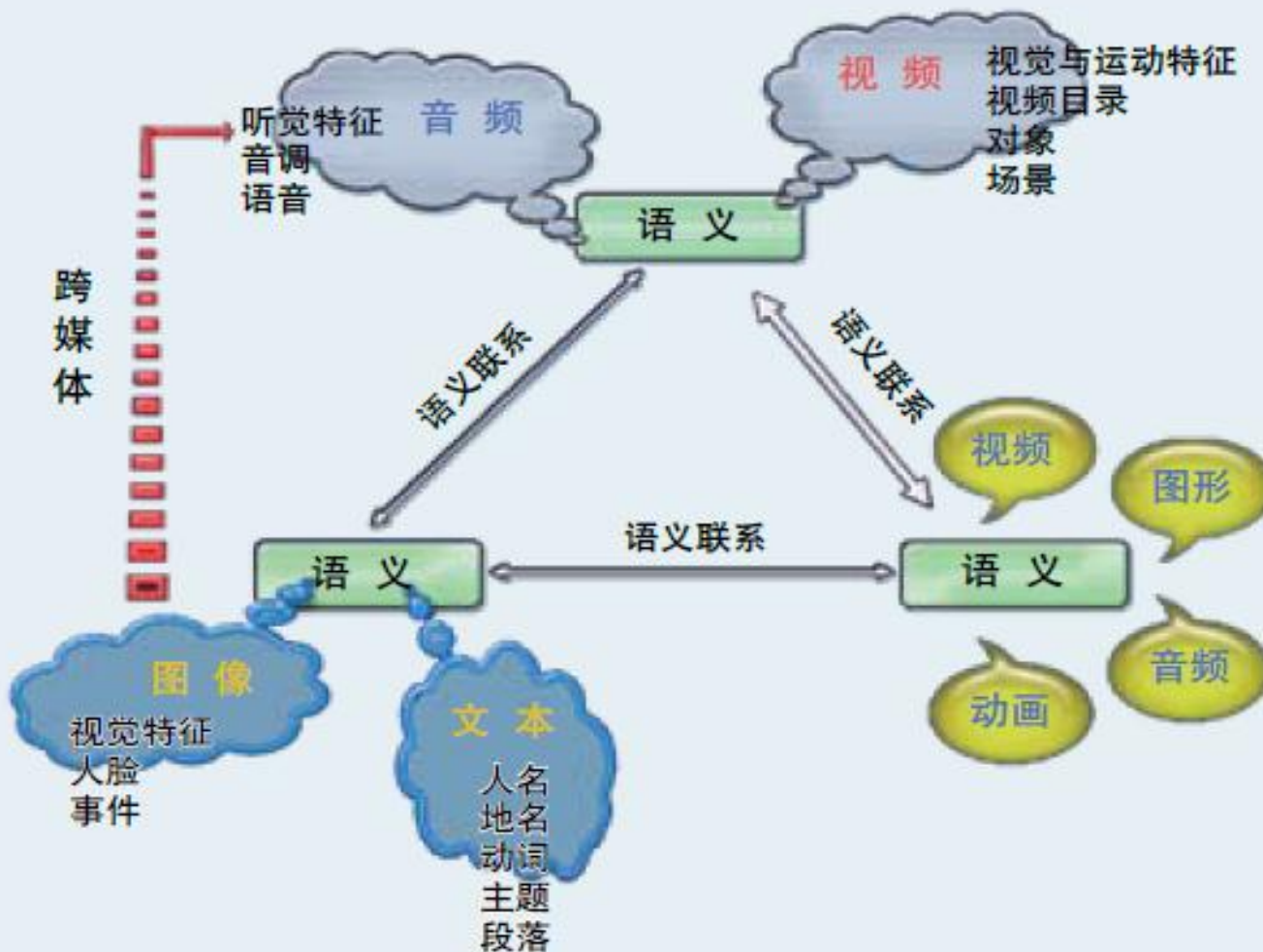
$$Y' = \begin{bmatrix} y_{11}, y_{12}, \dots, y_{1m} \\ y_{21}, y_{22}, \dots, y_{2m} \\ \dots \\ y_{n1}, y_{n2}, \dots, y_{nm} \end{bmatrix}$$

X 与 Y 相同维数!

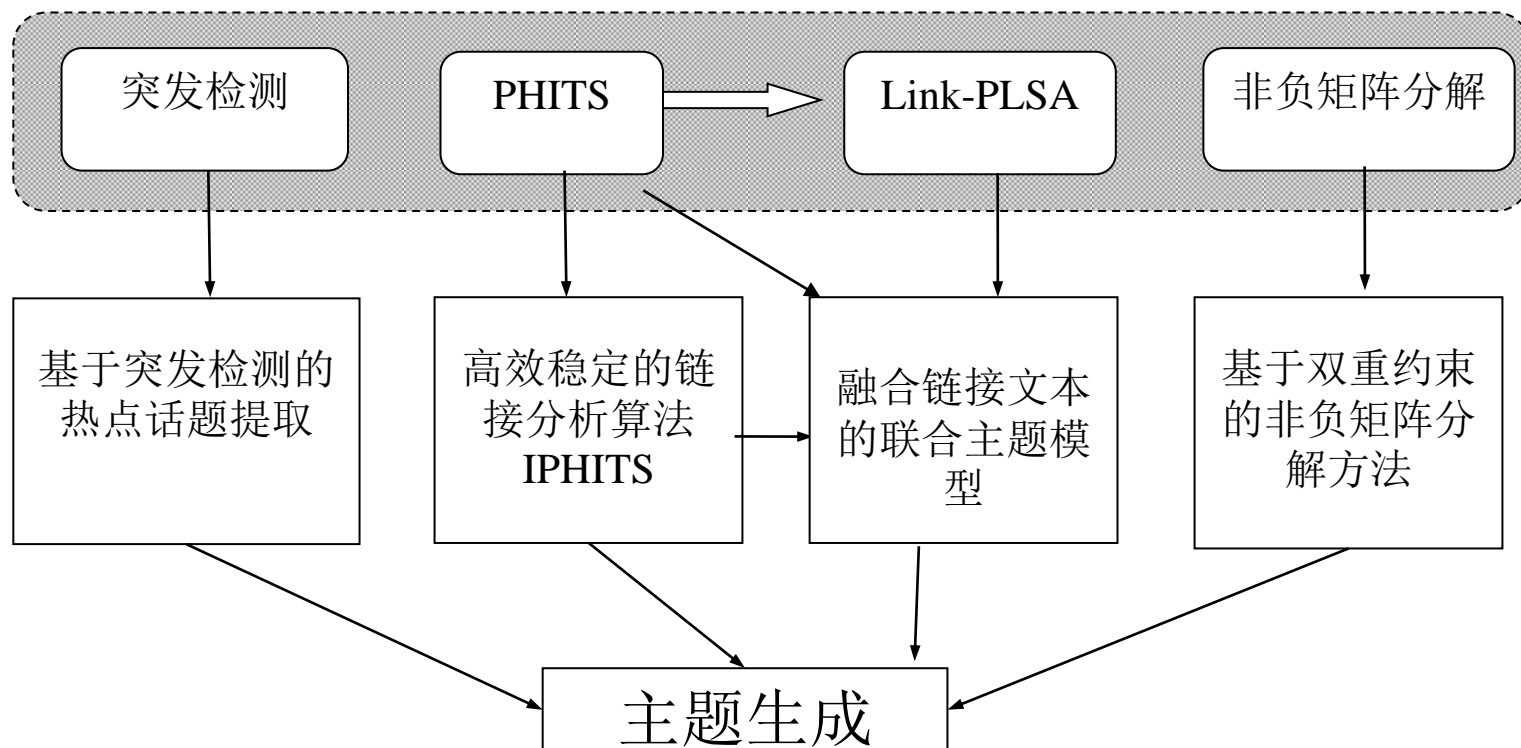
典型相关分析



跨媒体语义关联

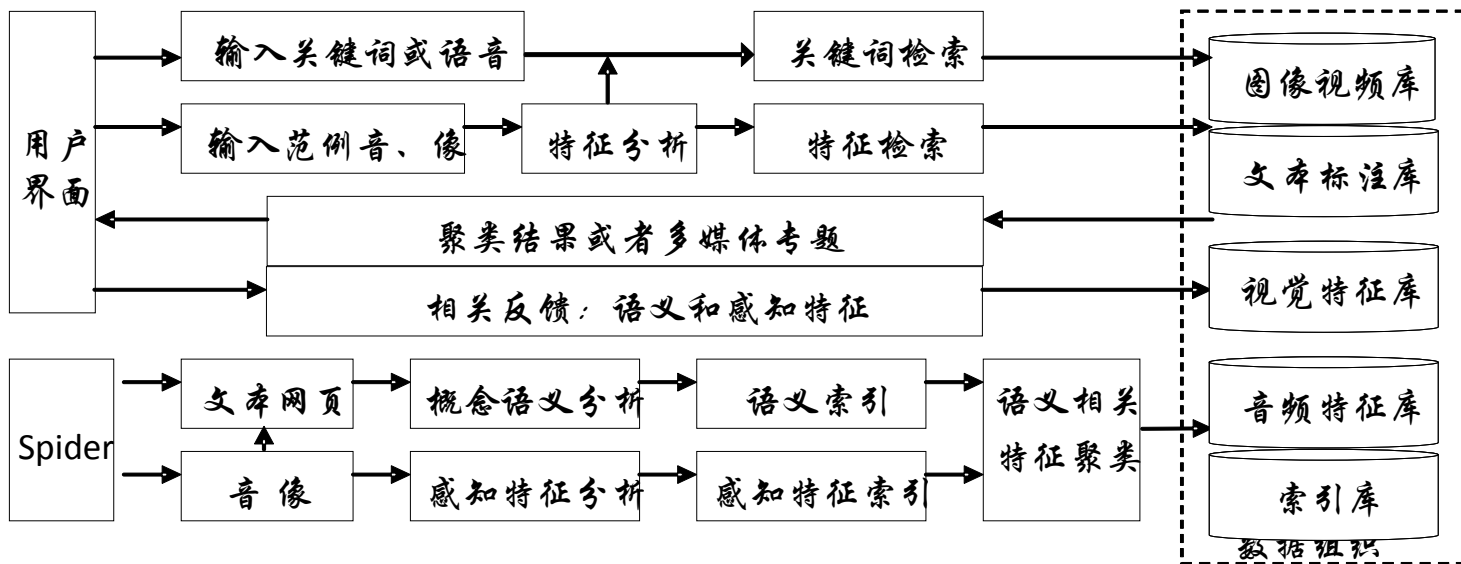


主题生成



跨媒体搜索引擎演示系统

- 基于信息粒度的专题组织与展示
- 基于概念语义空间的多媒体信息融合
- 开发了多媒体搜索引擎演示系统，有效地支持多媒体信息的理解与检索。



内容提要

引言

心智模型CAM

跨媒体语义分析

跨媒体语义关联

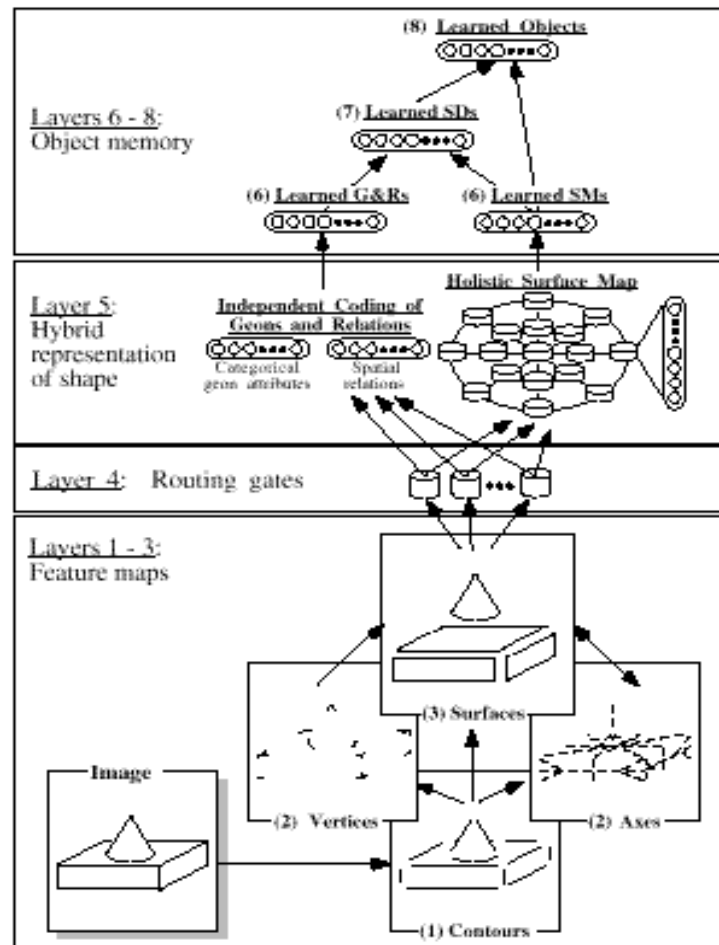
跨媒体语义理解

结束语

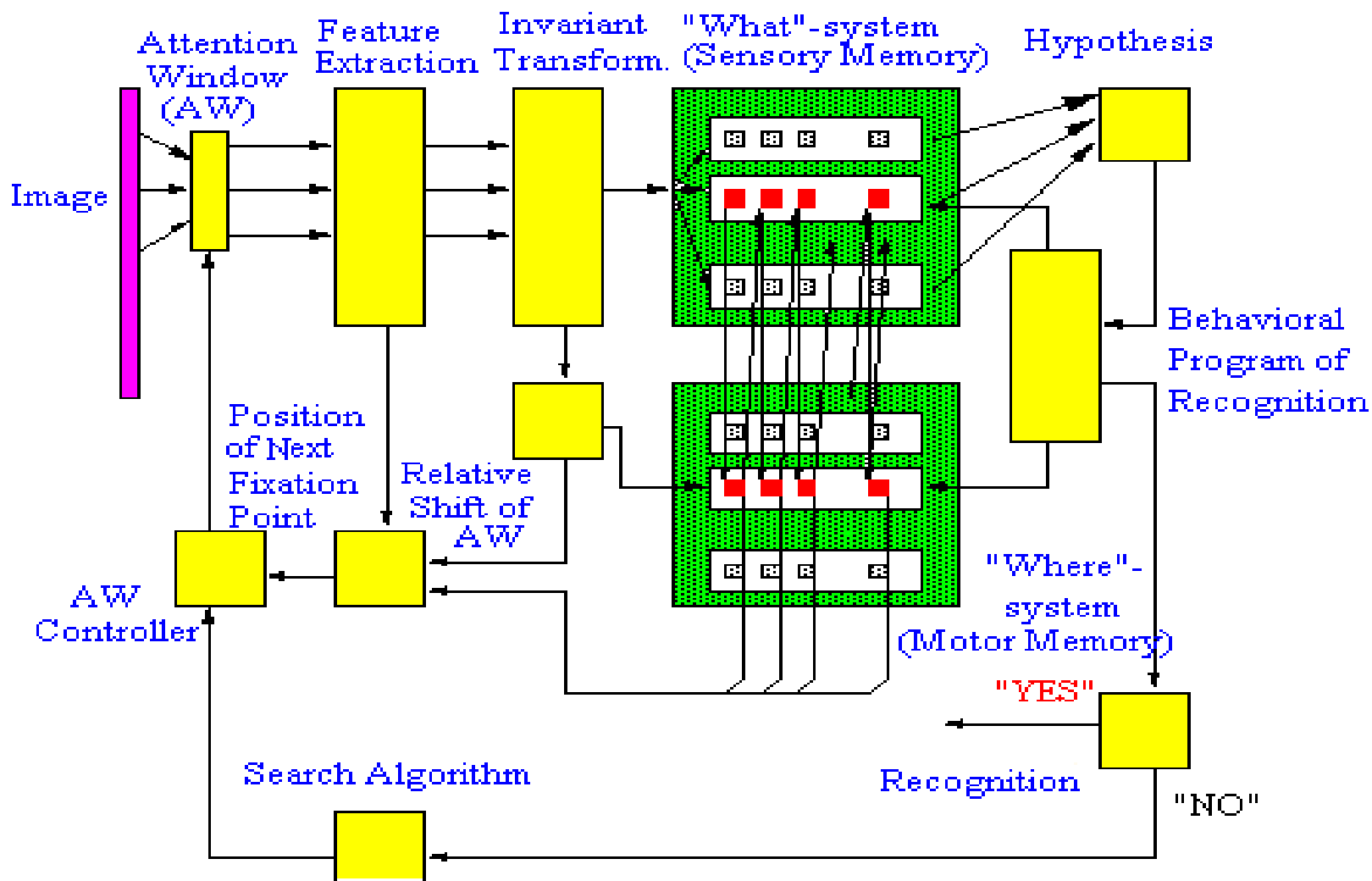
特征捆绑



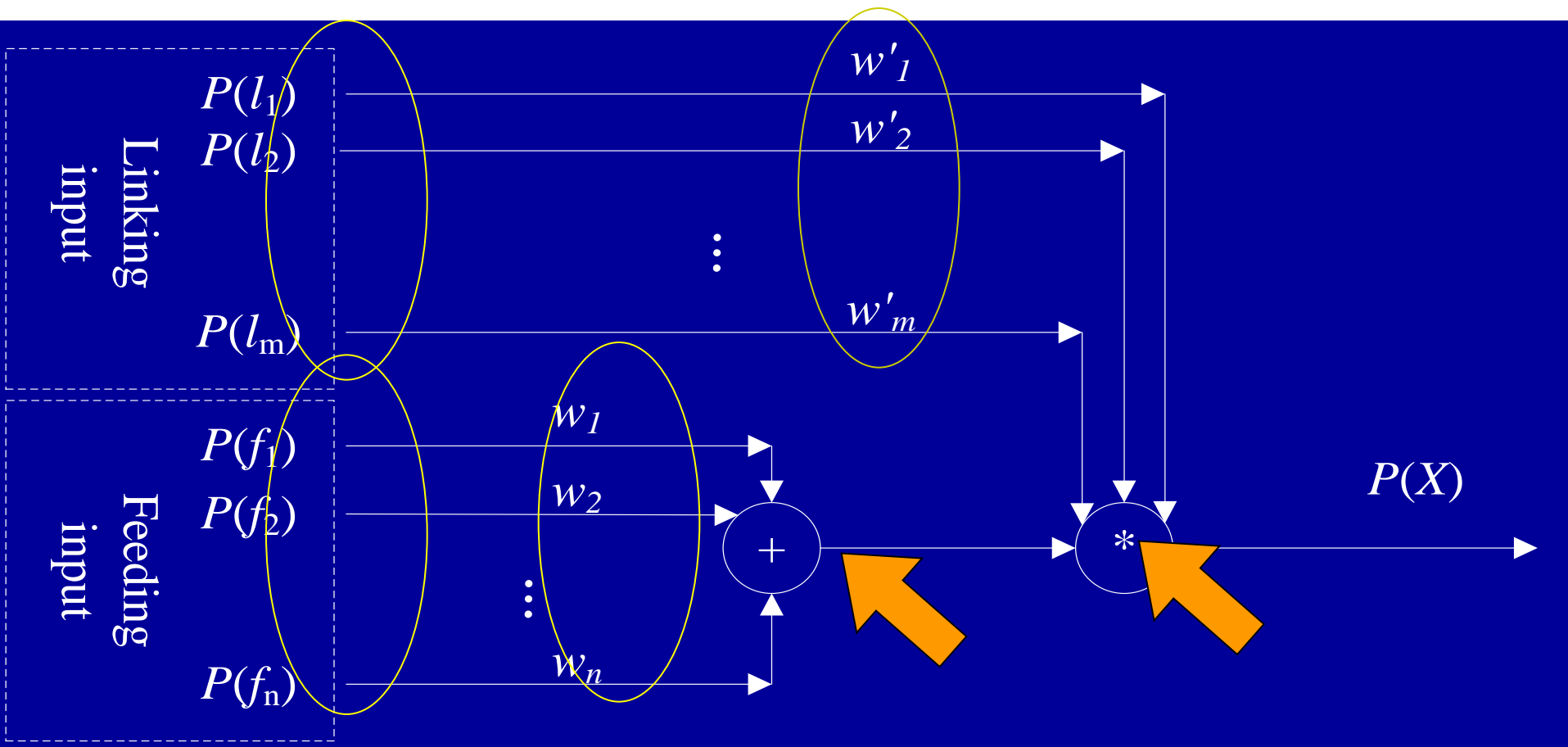
- Scenes are broken down and analyzed via many pathways, each with different feature preference.
- How are these disparate features bound into a single percept?
- How are different objects distinguished?



特征抽取



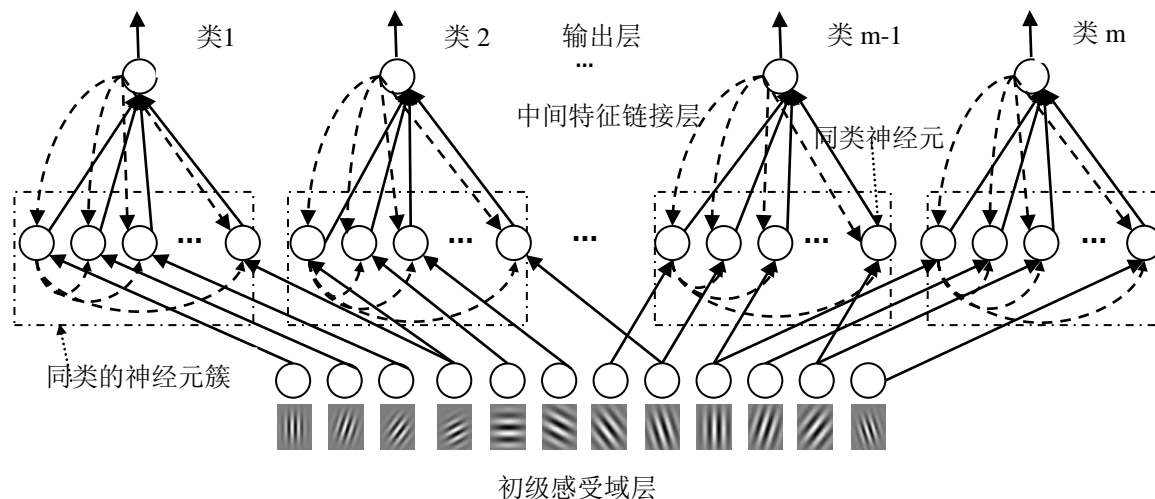
脉冲耦合神经网络模型



特征捆绑计算模型



- ▶自然界中的物体是由各种不同基本属性组成的，如颜色、大小、方位等，这些属性在大脑不同部位被加工。为了能够把物体知觉成一个整体而不仅是零散的特征，需要把散布于不同皮层区的属性信息合理地组合在一起，这就是所谓的“捆绑问题”。
- ▶Bayesian Linking Field模型是一种三层结构特征捆绑计算模型，自底向上分别为初级感受域层、中间特征连接层、输出层。每个神经元包括两部分输入：feeding输入和linking输入，输出由两部分输入耦合获得。



Bayesian Linking Field模型

史忠植 跨媒体数据挖掘和理解

基于特征捆绑计算模型的物体识别

初级特征选择

- 每维Gabor特征的贡献度： $TotalConA(i) = 0.2 * KurA(i) + 0.8 * LinkConA(i)$ ， $LinkConA(i) = \frac{TotalNum * NearMaxValASampleNum(i)}{TotalNumA * NearMaxValAAllSampleNum(i)}$ (1)
 $KurA$ 是类A在各维Gabor特征下的峰度值，
- 计算出各维Gabor原始特征的贡献值并由大到小排序，选择前指定个数（如20）为类A对应的特定维数的Gabor特征。

特征捆绑的实现

- 模型中所有的同类神经元组合构成同类神经元群，同类神经元之间相互连接，非同类神经元之间以及输出层神经元之间相互竞争
- 底层神经元到中间层神经元的连接权值均设为1；中间层神经元有两种连接，分别是来自于同类其他中间层神经元和相应输出神经元之间连接。

基于特征捆绑计算模型的物体识别

连接权值计算

- 同类中间层神经元之间的连接权值由(2)计算，中间层神经元与输出神经元之间连接权值由(3)计算。

$$W_{j, j'(m)} = \frac{TotalNum * NearMaxValSampleNum(j, j', m)}{NearMaxValSampleNum(j, m) * NearMaxValSampleNum(j', m)} \quad (2)$$

$$W_{j, k(m)} = \frac{TotalNum * NearMaxValMSampleNum(j, m)}{TotalNumM * NearMaxValSampleNum(j, m)} \quad (3)$$

c 为 O 与“ $X = A$ ”在训练集共同出现的次数， N 为训练集样本总数， a 为 O 在训练集中出现的次数， b 为“ $X = A$ ”在训练集中出现的次数

$$MI(O, "X = A") = \log\left(\frac{P(O | "X = A")}{P(O)}\right)$$



$$w_{ij} = MI(O, "X = A") = \log\left(\frac{cN}{ab}\right)$$

基于特征捆绑计算模型的物体识别

物体识别模型建立

▶选择初级特征，确定模型所有神经元；同时由特征捆绑实现部分计算出神经元之间的连接权值。

▶初级感受域层神经元的feeding输入 $P_{fi1}(t)$ 为相应的Gabor特征值，该层神经元的输出由公式 $P_{ilayer1}(t) = P_{fi1}(t)$ (4) 计算，

中间层神经元的feeding输入由 $P_{fi2}(t) = \exp(-1 * (P_{i2layer1}(t) - MaxVal)^2)$ (5) 计算，linking输入 $P_{lj2}(t)$ 是与其连接的同类其他神经元以及对应输出神经元的输出，这一层神经元输出 $P_{ilayer1}(t)$ 最后用 $P_{ilayer2}(t) = (\sum_{i2} w_{i2} P_{fi2}(t)) (\prod_{j2} w'_{j2} P_{lj2}(t))$ (6) 计算求得；

输出层神经元的feeding输入 $P_{fi3}(t)$ 是与其相连的中间层神经元的输出，它的输出由 $P_{ilayer3}(t) = \sum_{i3} w_{i3} P_{fi3}(t)$ (7) 计算

▶模型输出层神经元的初始 $t=0$ 时输出为 $1/m$ ，模型中各神经元的输出将进行不断更新调整 $t=1,2,\dots$ ，当整个过程收敛时，输出层中具有最大输出的神经元所表征的类别即为所识别的结果

实验结果分析



Caltech256物体识别



► Caltech256数据库，6类，100幅图像训练，50幅图像测试

分类器 对应每类的 中间层神经元个数	平均识别精度		
	Feature Binding Model	SVM	AdaBoost
20	78.7%	71.0%	73.0%
40	82.3%	77.7%	78.0%

刘曦，史忠植，石志伟，施智平. 一种基于特征捆绑计算模型的物体识别方法，
2004, 软件学报.

物体识别



测试样本	分类器	人脸正确识别率		
		Feature Binding Model	SVM	AdaBoost
所有测试样本(1000幅)		83.0%	77.9%	81.1%
所有正面测试样本(200幅)		96.5%	87.5%	93%
所有非正面测试样本(800幅)		79.6%	75.5%	78.1%
平均每幅图像所用时间(ms)		2.64	4.48	0.16

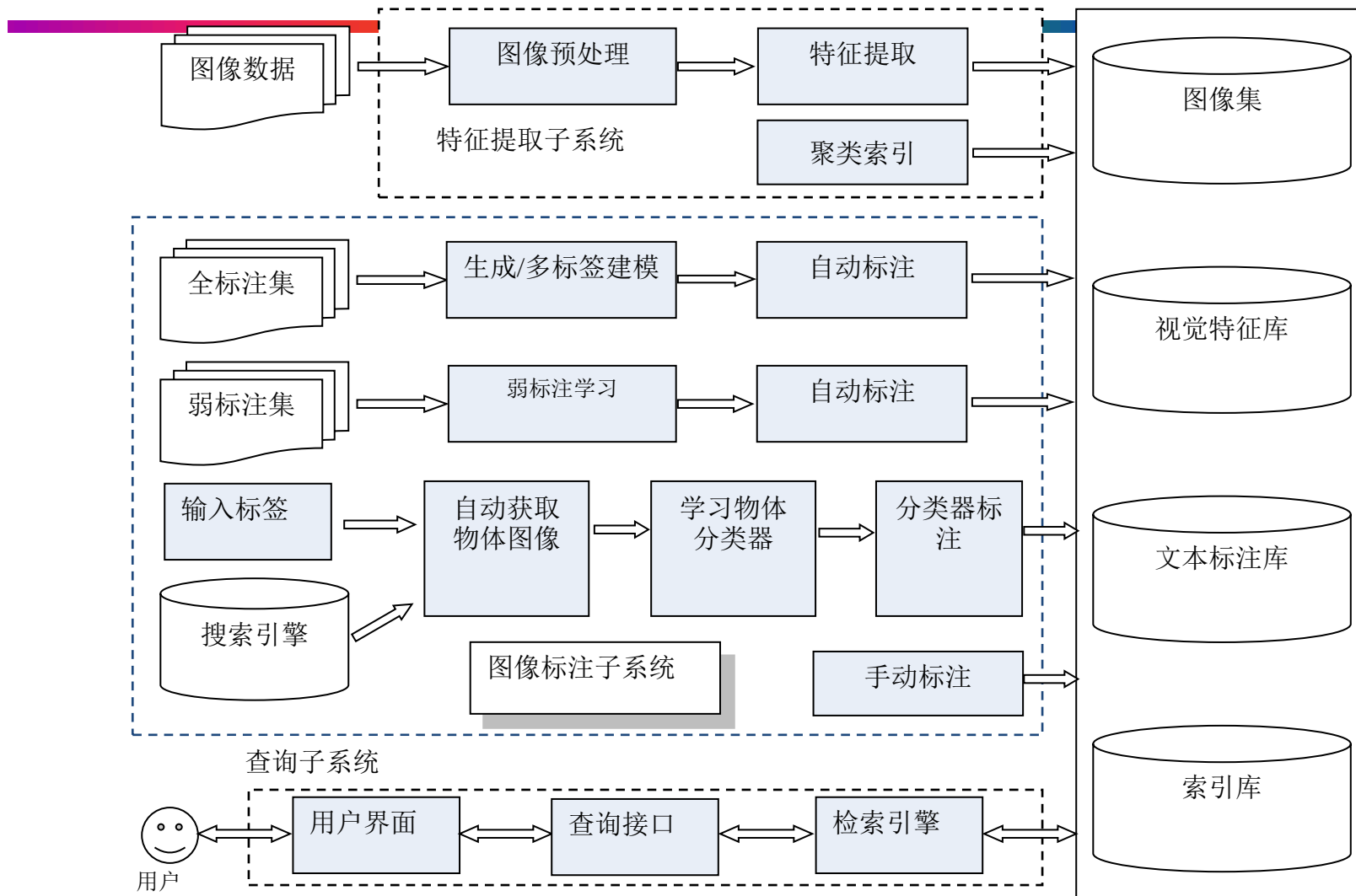


(每类物体中间层神经元个数)	分类器	平均正确识别率		
		Feature Binding Model	SVM	AdaBoost
20		78.7%	71.0%	73.0%
40		82.3%	77.7%	78.0%

跨媒体检索系统框架



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY



跨媒体检索系统框架

■ 特征提取子系统

图像格式转换，图像的增强与去噪等图像预处理功能；图像自动分割算法；多种图像视觉特征的提取算法及图像标注的预处理算法。

■ 图像标注子系统

- (1) 基于全标注集自动标注，拓扑序标签链分类器或者PLSA
- (2) 基于弱标注集自动标注，基于弱标注集模型共享分类方法
- (3) 基于图像搜索引擎返回结果的自动标注，基于外观及空域特征的半监督物体识别方法

■ 查询子系统

2011/5/7

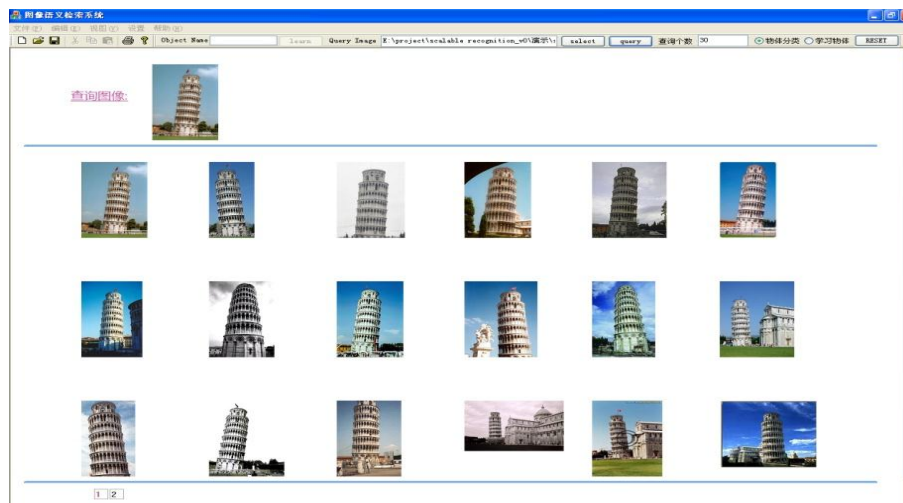
史忠植 跨媒体数据挖掘和理解

- (1) 基于图像范例查询
- (2) 基于标签查询

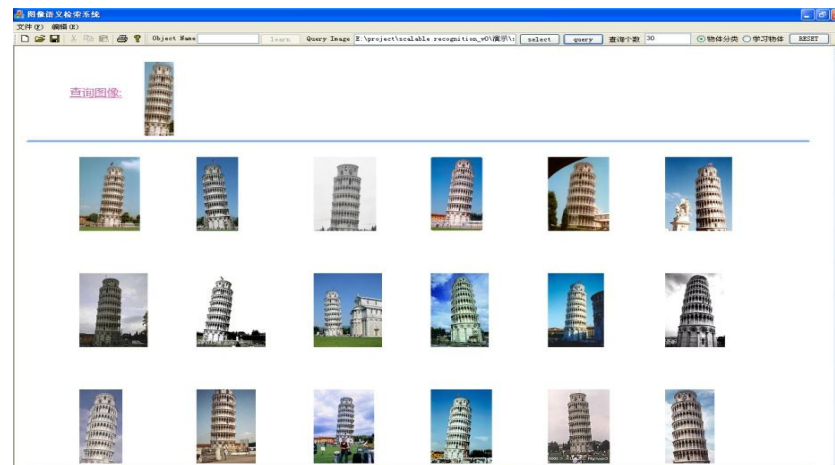
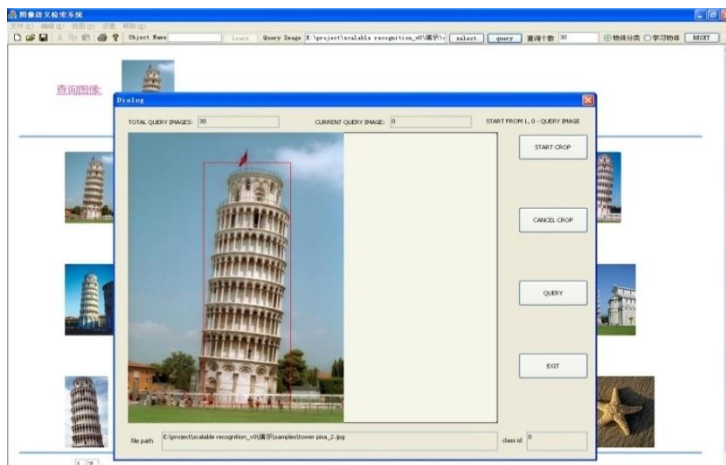
系统演示



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY



基于图范例查询



2011/5/7

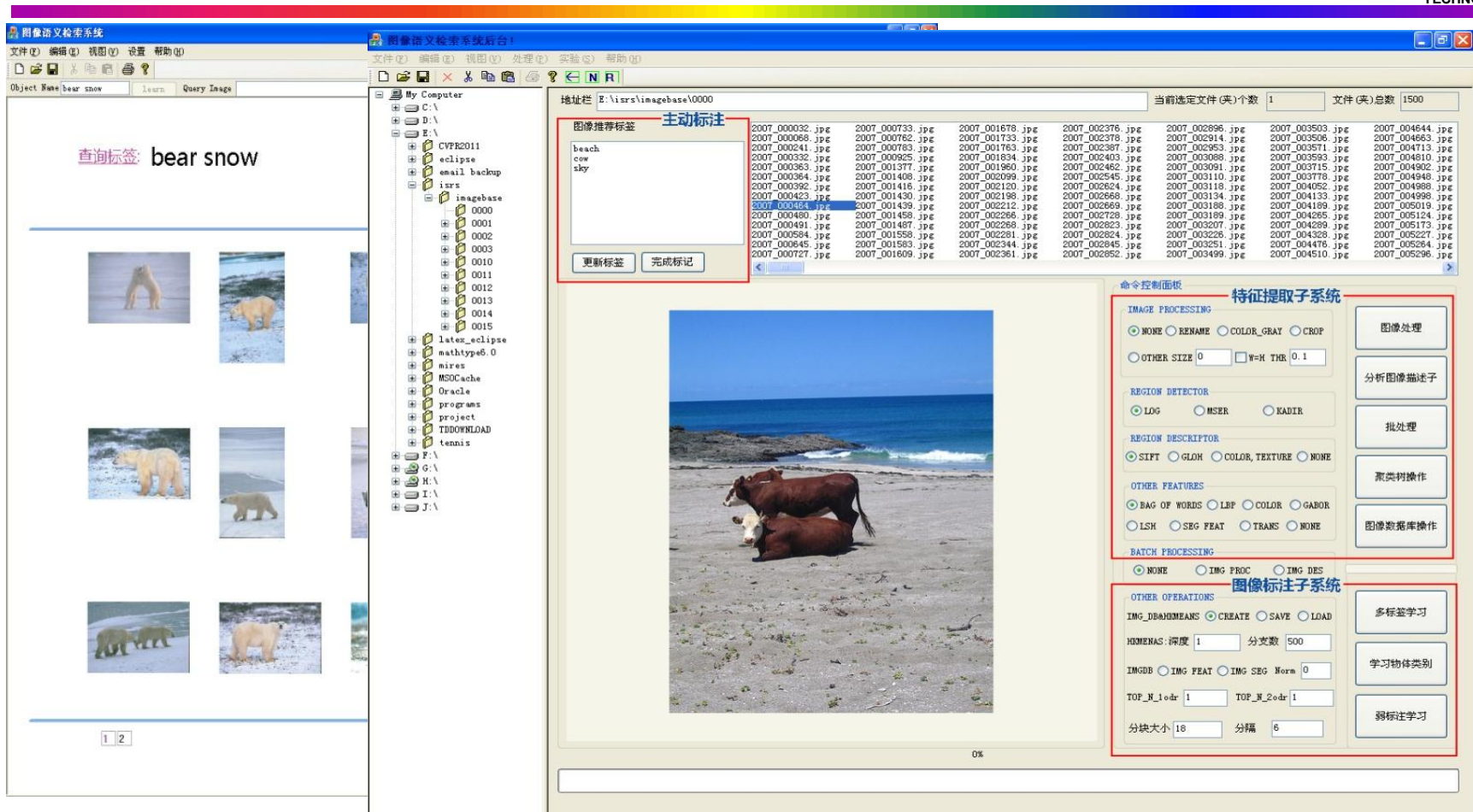
抓取图区域

史忠植 跨媒体数据挖掘和理解

基于图区域范例查询

70

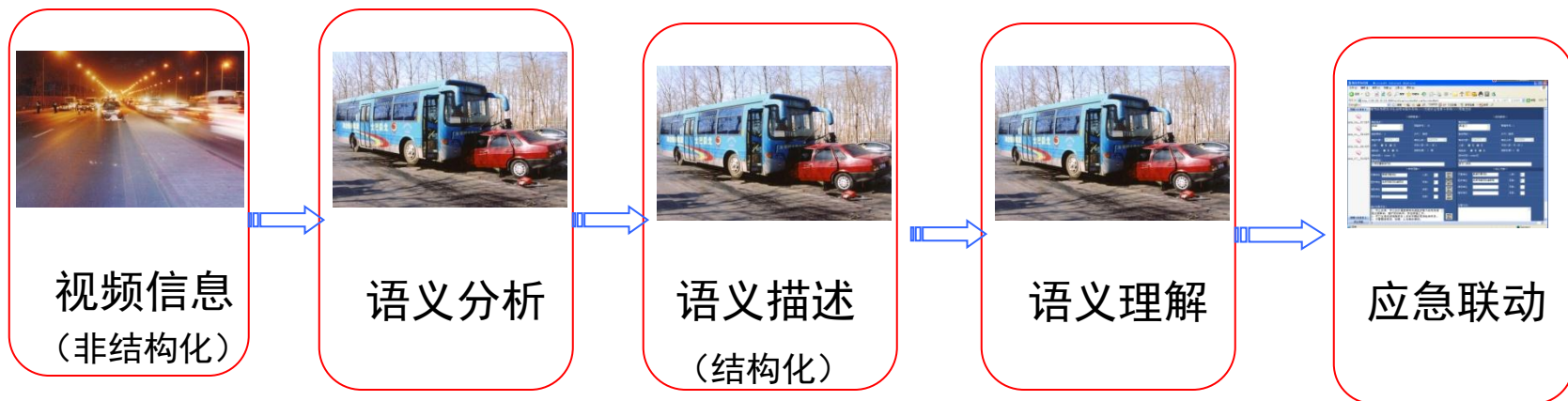
系统演示



基于标签查询

系统后台界面

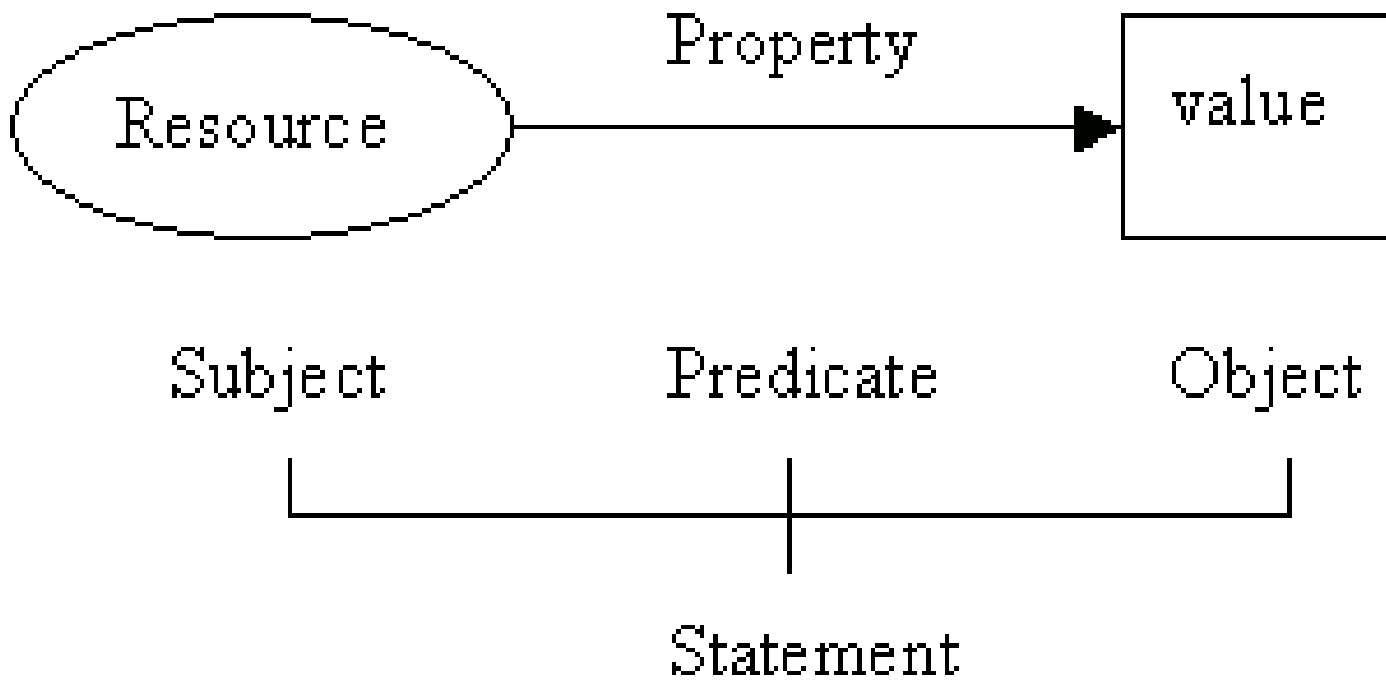
交通事故图像的理解



在CAM的基础上，提出一种图像理解模式。首先对图像（视频）进行特征抽取和语义分析，在此基础上采用本体描述图像（视频），然后通过基于案例推理或逻辑推理理解图像的语义。

图像语义表征

基本RDF



图像语义表征

我们将基本的RDF模型扩展为五元组

$\langle S, P, O, NG, T \rangle$,

- S表示主体 (Subject),
- P表示谓词 (Predicate),
- O表示客体 (Object),
- NG表示命名图 (named graph) : 标识RDF图中来自不同数据源的节点。
- T表示时间: 根据Allen's Interval Algebra, 定义两个客体在不同时间段的关系。

展望

- 认知模型
- 跨媒体语义分析
- 跨媒体关联挖掘
- 跨媒体理解

Thank You



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY

Question!

Intelligence Science

<http://www.intsci.ac.cn/>

