第四部分 社会媒体分析与理解
——跨媒体分析

程 健

jcheng@nlpr.ia.ac.cn

中国科学院自动化研究所

2022.10.24

大 纲

- 跨媒体是什么？
- 跨媒体内容统一表示
- 跨媒体知识图谱构建
- 跨媒体关联分析与推理
- 基于跨媒体分析的应用

大 纲

- ➢ 跨媒体是什么?
- ➢ 跨媒体内容统一表示
- ➢ 跨媒体知识图谱构建
- ➢ 跨媒体关联分析与推理
- ➢ 基于跨媒体分析的应用

跨媒体与人工智能

中华人民共和国中央人民政府
www.gov.cn

国务院关于印发
新一代人工智能发展规划的通知

国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院

2017年7月8日

基础理论：

1. 大数据智能理论
2. **<u>跨媒体感知计算理论</u>**
3. 混合增强智能理论
4. 群体智能理论
5. 自主协同控制与优化决策理论
6. 高级机器学习理论
7. 类脑智能计算理论
8. 量子智能计算理论

跨媒体与人工智能

中华人民共和国中央人民政府
www.gov.cn

国务院关于印发
新一代人工智能发展规划的通知
国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院

2017年7月8日

跨媒体感知计算理论：

研究超越人类视觉能力的感知获取、面向真实世界的主动视觉感知及计算、自然声学场景的听知觉感知及计算、自然交互环境的言语感知及计算、面向异步序列的类人感知及计算、面向媒体智能感知的自主学习、城市全维度智能感知推理引擎。

什么是跨媒体(Cross-media)?

前面介绍的内容往往只针对某种单一形式的媒体数据进行分析，比如图像识别、语音识别、文本识别等。跨媒体分析是指多种形式(文本、音频、视频、图像等)信息协同的多媒体内容分析。

什么是跨媒体(Cross-media)？

特点：跨媒体既表现为包括网络文本、图像、音频、视频等复杂媒体对象混合并存，又表现为各类媒体对象形成复杂的关联关系和组织结构，还表现在具有不同模态的媒体对象跨越媒介或平台高度交互融合。

通过“跨媒体”能从各自的侧面表达相同的语义信息，能比单一的媒体对象及其特定的模态更加全面地反映特定的内容信息。相同的内容信息跨越各类媒体对象交叉传播与整合，只有对这些多模态媒体进行融合分析，才能尽可能全面、正确地理解这种跨媒体综合体所蕴涵的内容信息。

大 纲

- 跨媒体是什么？
- **跨媒体内容统一表示**
- 跨媒体知识图谱构建
- 跨媒体关联分析与推理
- 基于跨媒体分析的应用

跨媒体内容统一表示

多媒体由不同类别媒体组成，因不同类别的媒体数据分别使用不同维数、不同属性的底层特征进行表示，使不同类别的媒体之间无法直接根据特征来计算其相关性，而造成的彼此之间的异构性和不可比性的特性不同，彼此之间存在“鸿沟”。

首先要解决的问题是：

针对跨媒体信息，如何学习一种统一的表达？

典型相关分析（CCA）

典型相关分析（Canonical Correlation Analysis，CCA）利用综合变量对之间的相关关系来反映两组指标之间的整体相关性的多元统计分析方法。

矩阵

设有两组随机变量$\mathbf{X}=(x_1,x_2,\cdots,x_p)'$和$\mathbf{Y}=(y_1,y_2,\cdots,y_q)'$，

分别对两组变量做线性组合：

向量

$$U=a_1x_1+a_2x_2+\cdots+a_px_p=a'\mathbf{X}$$
$$V=b_1x_1+b_2x_2+\cdots+b_qy_q=b'\mathbf{Y}$$

协方差矩阵为$cov(\mathbf{X},\mathbf{Y})=\Sigma_{12}=\Sigma_{21}'$

$cov(X,Y)=\frac{1}{n}(X-\bar{X})(Y-\bar{Y})^T$

Hotelling, H. . Relations Between Two Sets of Variates. 1936

典型相关分析（CCA）

相关系数为： $\rho = corr(U,V) = \dfrac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{12}b}}$

其中U,V称为典型变量，它们之间的相关系数ρ称为典型相关系数。

$$\max \quad \rho = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}}$$

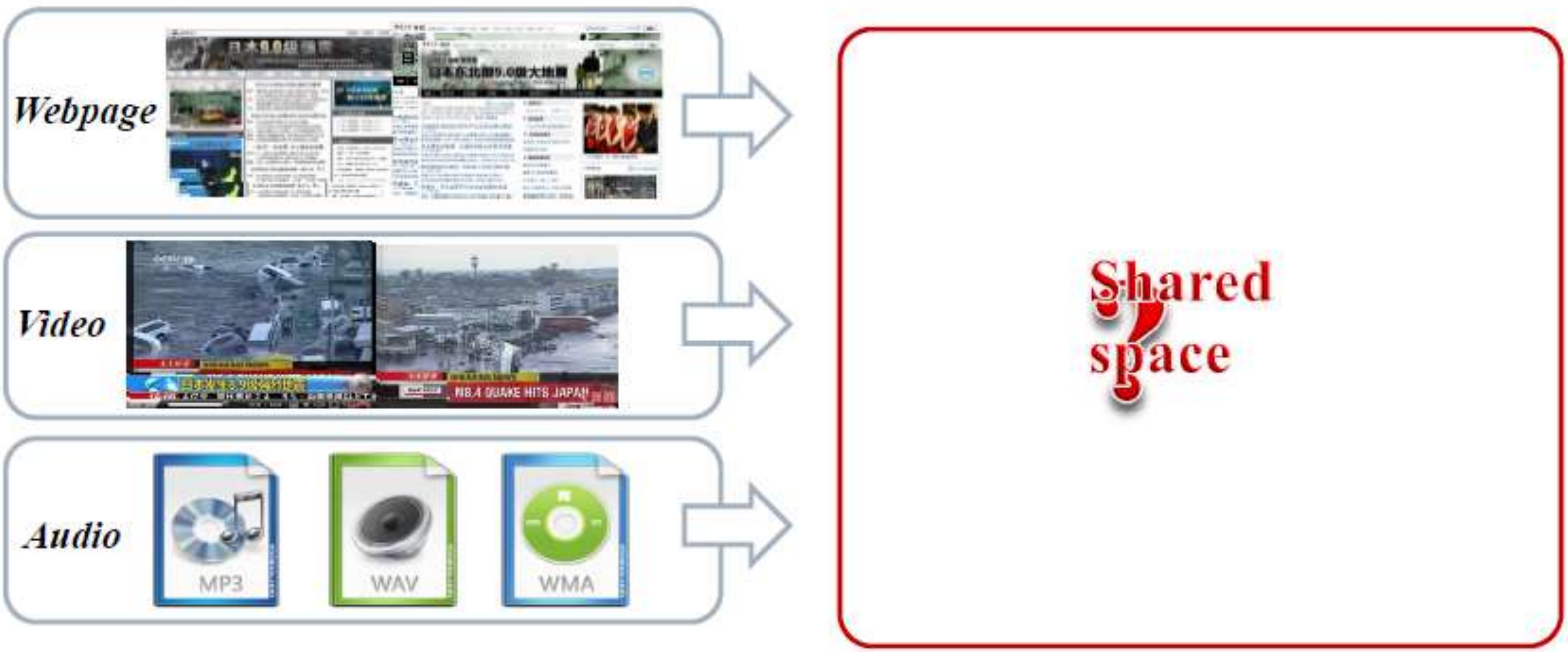
此时，把U，V称为（第一对）典型相关因子。

求解方法：拉格朗日乘数法（省略）

Hotelling, H. . Relations Between Two Sets of Variates. 1936

典型相关分析（CCA）

- CCA (Canonical Correlation Analysis) and its extensions
  - Kernel CCA, Sparse CCA, Sparse Structure CCA
  - 2D CCA, local 2D-CCA, sparse 2D-CCA, 3-D CCA



Kernelized CCA (KCCA)

首先引入一个把数据映射到高维特征空间的非线性映射：

$$\phi : \mathbf{x} = (x_1, \ldots, x_m) \mapsto \phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_n(\mathbf{x})), (m < n)$$

存在一个核K，对所有的x,z有： $K(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$

设两组向量的样本矩阵为：

$$X_{p\times N} = (\mathbf{X}_1, \ldots, \mathbf{X}_N), Y_{q\times N} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$$

设ϕ_X, ϕ_Y分别表示作用于X, Y上的变换，即：

$$\phi_X(\mathbf{X}) = (\phi_X(\mathbf{X}_1), \ldots, \phi_X(\mathbf{X}_N)), \phi_Y(\mathbf{Y}) = (\phi_Y(\mathbf{Y}_1), \ldots, \phi_Y(\mathbf{Y}_N))$$

变换后的$\phi_X(\mathbf{X})$、$\phi_Y(\mathbf{Y})$均为n x N维矩阵。

S. J. Hwang and K. Grauman, “Learning the relative importance of objects from tagged images for retrieval and cross-modal search,” International Journal of Computer Vision, 2012.

Kernelized CCA (KCCA)

和CCA一样，寻找向量a和b，

使得：$U = a^T \phi_X(\mathbf{X})$ 和 $V = b^T \phi_Y(\mathbf{Y})$ 相关系数最大

$$\rho = corr(U, V) = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}}$$

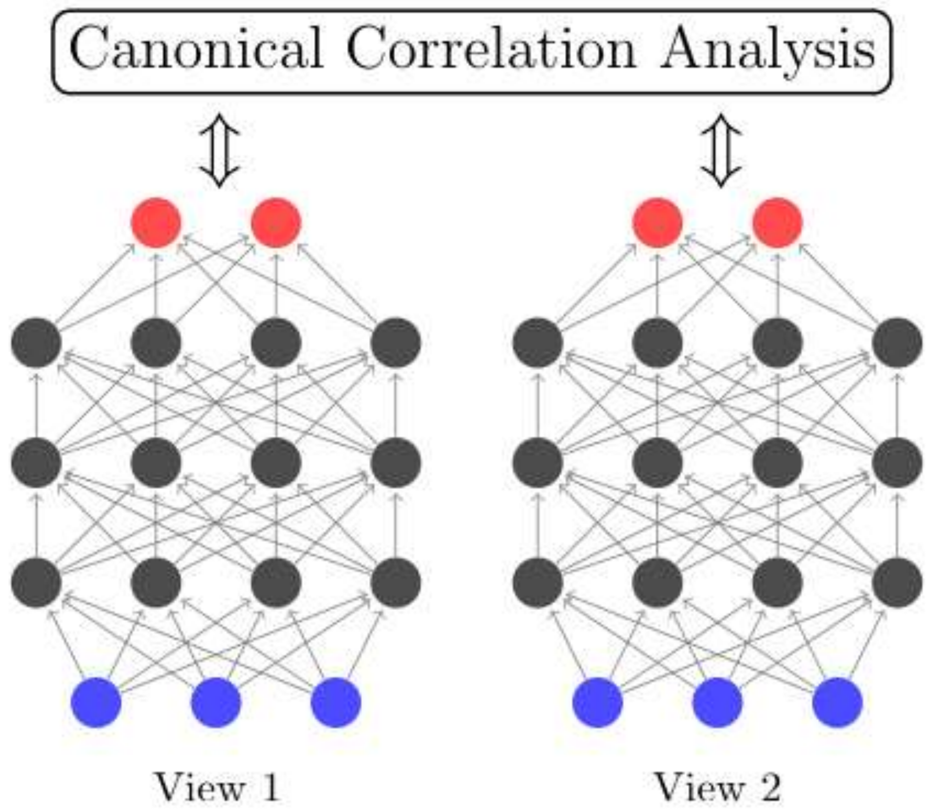
其中：

$$K_X(i,j) = K_X(X_i, X_j) = \phi_X(\mathbf{X}_i)^T \phi_X(\mathbf{X}_j)$$

$$K_Y(i,j) = K_Y(Y_i, Y_j) = \phi_Y(\mathbf{Y}_i)^T \phi_Y(\mathbf{Y}_j)$$

S. J. Hwang and K. Grauman, “Learning the relative importance of objects from tagged images for retrieval and cross-modal search,” International Journal of Computer Vision, 2012.

Deep CCA (DCCA)



View 1网络每层输入输出:

$$h_1 = s(W_1^1 x_1 + b_1^1) \in \mathbb{R}^{c_1}$$

$$h_2 = s(W_2^1 h_1 + b_2^1) \in \mathbb{R}^{c_1}$$

$$f_1(x_1) = s(W_d^1 h_{d-1} + b_d^1) \in \mathbb{R}^{o}$$

$$\frac{\mathrm{cov}(f_1, f_2)}{\sqrt{D(f_1)}\sqrt{D(f_2)}}$$

目标是相关系数最大化:

$$(\theta_1^*, \theta_2^*) = \underset{(\theta_1,\theta_2)}{\operatorname{argmax}} \operatorname{corr}(f_1(X_1;\theta_1), f_2(X_2;\theta_2))$$

G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in Proc. Int. Conf. Mach. Learn. (ICML), 2013

Deep CCA (DCCA)

表1. DCCA与CCA、KCCA在MNIST数据集上的相关性比较，

	CCA	KCCA (RBF)	DCCA (50-2)
Dev	28.1	33.5	**39.4**
Test	28.0	33.0	**39.7**

表2. 网络DCCA-112-d的总相关性比较, d=3,…8

layers (d)	3	4	5	6	7	8
Dev set	66.7	68.1	70.1	72.5	76.0	**79.1**
Test set	80.4	81.9	84.0	86.1	88.5	**88.6**

G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in Proc. Int. Conf. Mach. Learn. (ICML), 2013

CCA及其变种比较

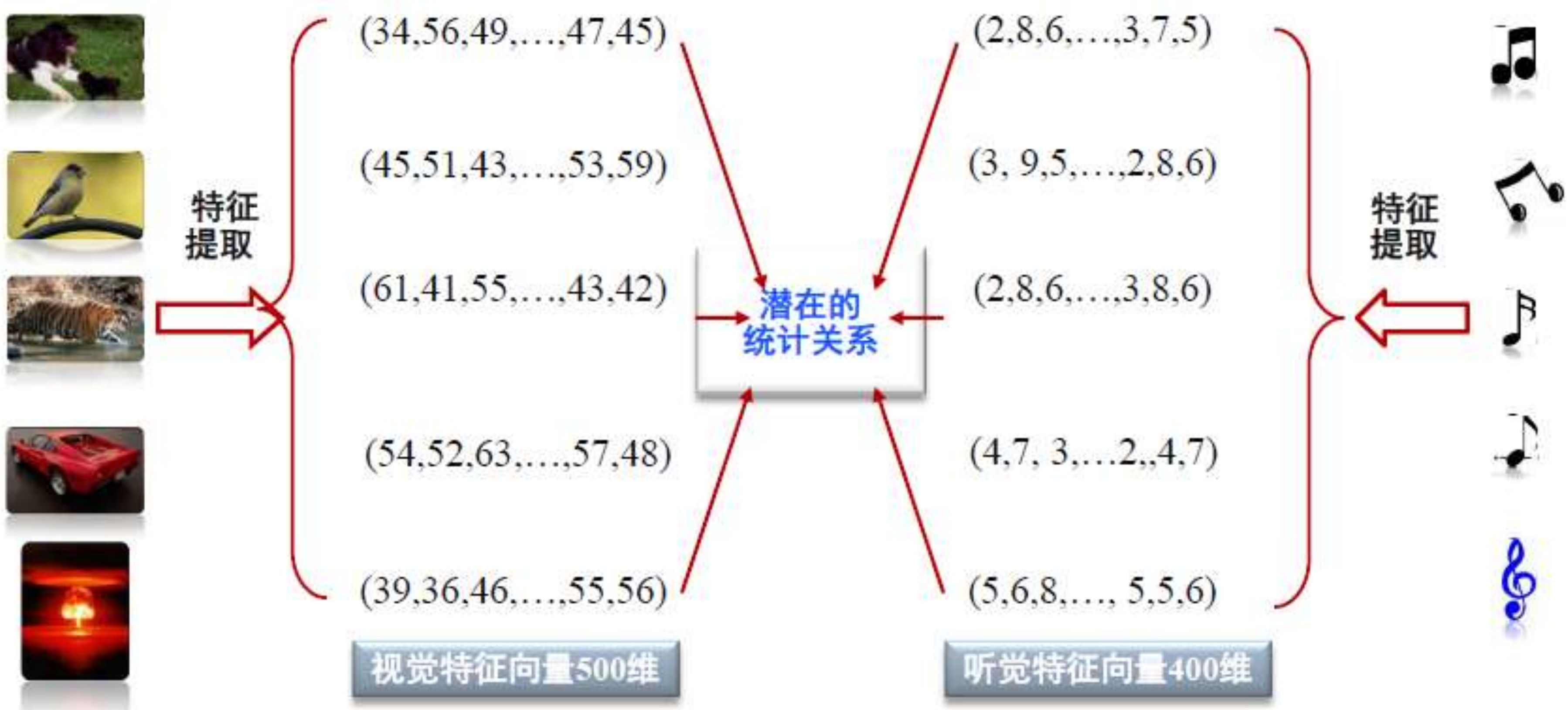
CCA：线性、简单，相关性一般

KCCA：非线性，较复杂，相关性较高

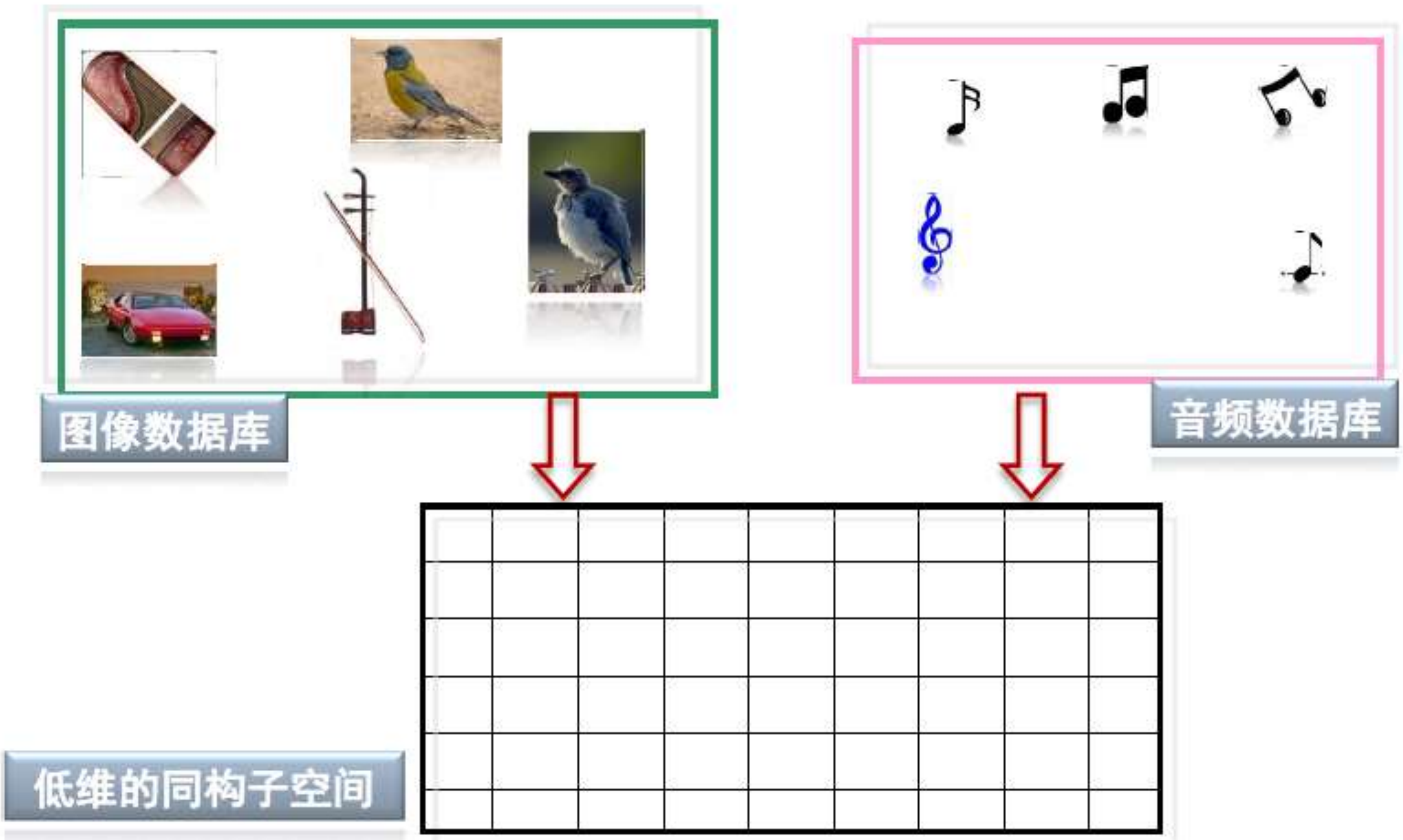
DCCA：非线性，复杂，相关性高

跨媒体内容统一表示

通过典型相关性分析学习不同类型媒体数据在底层特征上的统计相关性，建立跨媒体同构空间，从而实现了不同类型媒体数据度量的有效机制。



跨媒体内容统一表示



跨媒体内容统一表示

用于子空间映射的矩阵计算过程：

$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times p}$ $\quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times q}$

从图像训练集中提取的视觉特征矩阵

从音频训练集中提取的听觉特征矩阵

$X'_{n\times m} = X W_x$

$Y'_{n\times m} = Y W_y$

$\rho(X', Y')$

$$X = \begin{bmatrix} x_{11}, x_{12}, \ldots, x_{1500} \\ x_{21}, x_{22}, \ldots, x_{2500} \\ \cdots \\ x_{n1}, x_{n2}, \ldots, x_{n500} \end{bmatrix}$$

$X' = XWx \quad Y' = YWy$

$\max \rho(X', Y') \approx \rho(X, Y)$

采用典型相关性分析计算两者间的统计关系

$$Y = \begin{bmatrix} y_{11}, y_{12}, \ldots, y_{1400} \\ y_{21}, y_{22}, \ldots, y_{2400} \\ \cdots \\ y_{n1}, y_{n2}, \ldots, y_{n400} \end{bmatrix}$$

通过拉格朗日算法找到两个转换矩阵Wx和Wy

线性降维之后两个矩阵之间的相关性最大程度地与降维之前保持一致

主题模型(Topic Model)

主题模型（Topic Model）是以非监督学习的方式对文档中隐含语义结构（latent semantic structure）进行聚类（clustering）的统计模型。常用于自然语言处理、文本挖掘等。

$$p(w_i|d_j) = \sum_{k=1}^{K} p(w_i|t_k) \times p(t_k|d_j)$$

简单统计数据集可知 $p(w_i|d_j)$

求：$p(w_i|t_k)$ 主题上的词分布

$p(t_k|d_j)$ 文档上的主题分布

主题模型(Topic Model)

- 潜在语义索引（Latent Semantic Indexing, LSI）
- 概率潜在语义索引（Probabilistic LSI, PLSI）
- 隐式狄利克雷分布（Latent Dirichlet Allocation, LDA）

隐式狄利克雷分布(LDA)

隐含狄利克雷分布（LDA）是由David Blei等人在2003年提出的，是无监督启发式的贝叶斯概率模型。 p(主题|文档)

1、对每个文档，从狄利克雷分布中采样生成文档的主题分布：$\theta_d \sim Dir(\alpha)$

2、对文档中的第i个词，

（a）从主题的多项式分布 θ_d 中采样生成文档的主题 $z_i \sim \theta_d$

（b）从多项式分布 β_{z_i} 中采样最终生成词语 $w_i \sim \beta_{z_i}$

p(词|主题)

$$p(w_i, \theta_d) = \sum_k p(w_i|\beta_k)p(z_i = k|\theta_d)$$

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. The Journal of Machine Learning Research, 2003.

Multimodal LDA

多模态数据包括文档d 和非文档信息(如图像) f

1、对每个文档，从狄利克雷分布中采样生成文档的主题分布：$\theta_d \sim Dir(\alpha)$

2、对文档中的第i个（词，f）对，

（a）从主题的多项式分布 θ_d中采样生成文档的主题 $z_i \sim \theta_d$

（b）从多项式分布β_{z_i} 中采样最终生成词语 $w_i \sim \beta_{z_i}$

（c）从多项式分布ψ_{z_i} 中采样最终生成图像 $f_i \sim \psi_{z_i}$

（d）观察到(w_i, f_i)

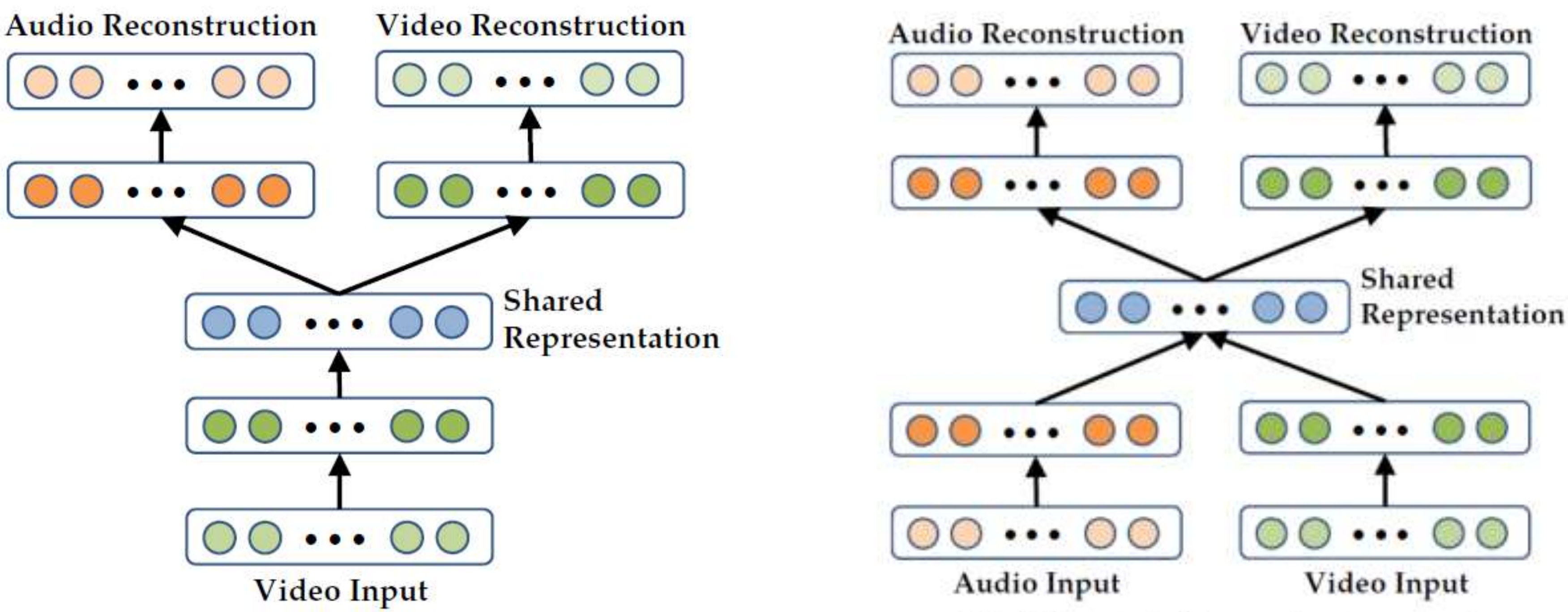
两模态： $$p(w_i, f_i, \theta_d) = \sum_k p(w_i|\beta_k)p(f_i|\psi_k)p(z_i = k|\theta_d)$$

多模态：$$p(w_i, f_i, f_i', \ldots, \theta_d) = \sum_k p(w_i|\beta_i)p(f_i|\psi_k)p(f_i'|\psi_i') \cdots p(z_i = k|\theta_d)$$

Mark Andrews, Gabriella Vigliocco, and David Vinson. Integrating experiential and distributional data to learn semantic representations. Psychological Review, 2009.

深度学习方法

多模态自编码机(AutoEncoder):

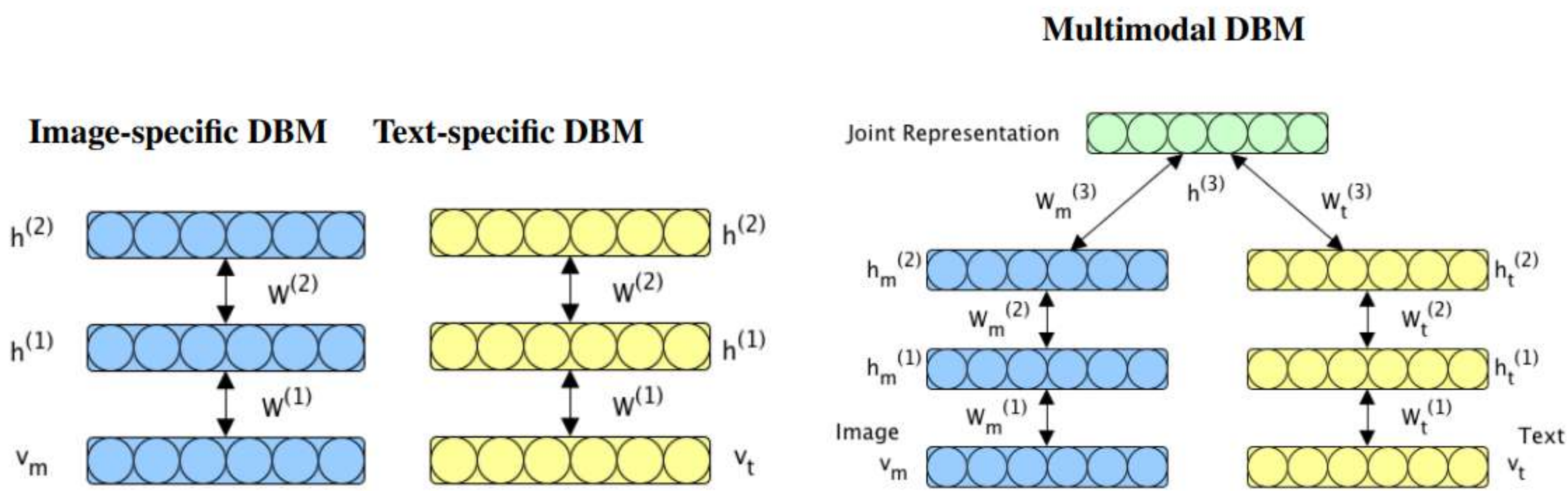


(a) Video-Only Deep Autoencoder (b) Bimodal Deep Autoencoder

Ngiam, J., Khosla, A., Kim, M., et al.. Multimodal deep learning. Int. Conf. on Machine Learning, 2011

深度学习方法

多模态受限玻尔兹曼机(RBM)：



Srivastava, N., Salakhutdinov, R.. Multimodal learning with deep Boltzmann machines. NIPS2012

基于张量的视频镜头表示

图像 声音 文本

每个视频镜头用一个三阶张量 $S \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ 来表示，其中,I_1,I_2和I_3分别是图像特征向量、音频特征向量及文本特征向量的维数：

$s_{i_1,1,1}(1 \le i_1 \le I_1)$ 为图像特征向量对应的值;

$s_{2,i_2,2}(1 \le i_2 \le I_2)$ 为音频特征向量对应的值;

$s_{3,3,i_3}(1 \le i_3 \le I_3)$ 为文本特征向量对应的值;

吴飞等，基于张量表示的直推式多模态视频语义概念检测，软件学报，2008

大 纲

- 跨媒体是什么？
- 跨媒体内容统一表示
- **跨媒体知识图谱构建**
- 跨媒体关联分析与推理
- 基于跨媒体分析的应用

跨媒体知识图谱构建

跨媒体知识图谱构建的目的是为了提供基本的可计算的知识表达结构，从而在跨媒体环境中语义关系分析以及认知层级的推理。

关键问题：

- 跨媒体知识图谱创建：实体提取以及关系构建
- 基于跨媒体知识图谱的信息查询与检索
- 跨媒体知识图谱中对的挖掘与推理
- 知识驱动的跨媒体学习模型

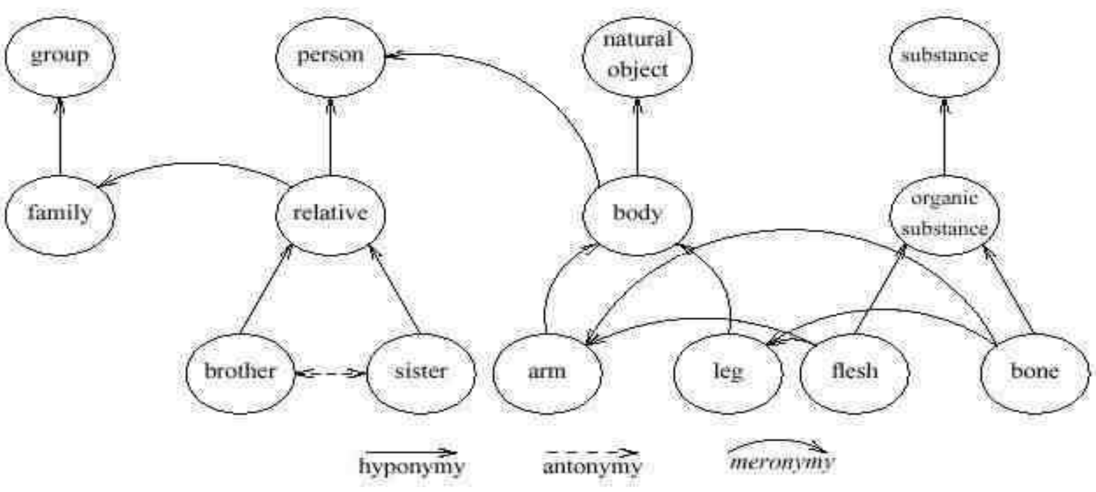
典型知识图谱—WordNet

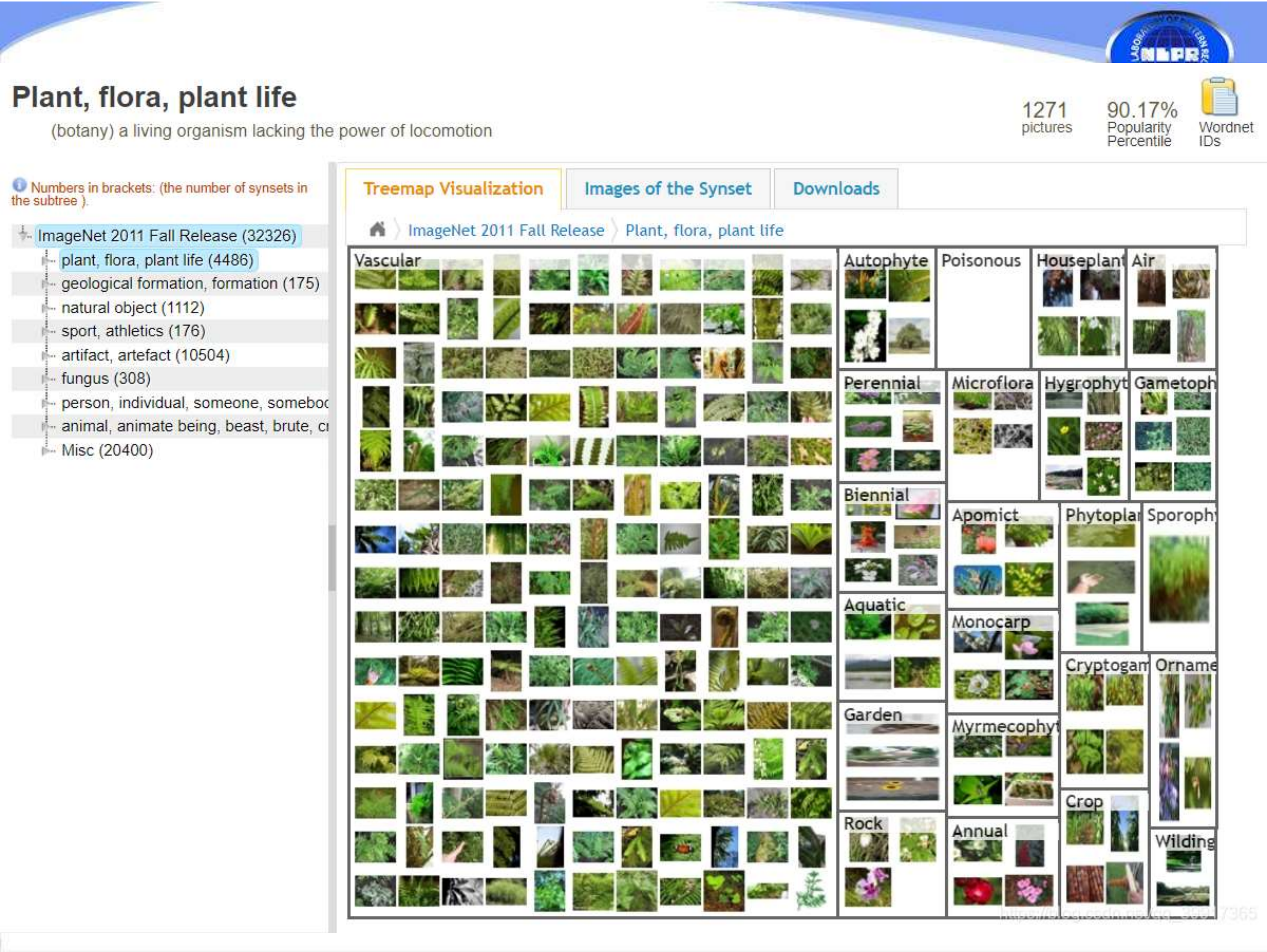
Wordnet是一个由普林斯顿大学认识科学实验室在心理学教授乔治·A·米勒带领建立和维护的大型的英语词典。将词汇分成五个大类：名词、动词、形容词、副词和虚词。

知更鸟：

（1）属性（attributes）：恒温脊椎动物，
（2）部件（Parts）：beak, feathers, wings
（3）功能（functions）：sings, flies, lays eggs

Figure 2. Network representation of three semantic relations among an illustrative variety of lexical concepts




Plant, flora, plant life
(botany) a living organism lacking the power of locomotion
1271
pictures
90.17%
Popularity Percentile
Wordnet IDs
Numbers in brackets: (the number of synsets in the subtree).
ImageNet 2011 Fall Release (32326)
plant, flora, plant life (4486)
geological formation, formation (175)
natural object (1112)
sport, athletics (176)
artifact, artefact (10504)
fungus (308)
person, individual, someone, somebod
animal, animate being, beast, brute, cr
Misc (20400)
Treemap Visualization
Images of the Synset
Downloads
ImageNet 2011 Fall Release
Plant, flora, plant life
Vascular
Autophyte
Poisonous
Houseplant
Air
Perennial
Microflora
Hygrophyt
Gametoph
Biennial
Apomict
Phytoplar
Sporophy
Aquatic
Monocarp
Garden
Myrmecophyt
Cryptogam
Orname
Rock
Annual
Crop
Wilding

典型知识图谱—Wikipedia

Wikipedia是一个基于维基技术的多语言百科全书式的协作计划，由网民自发形成共同参与创建、维护、编辑、修改的一个网络空间，是全球网络上最大的参考工具书

典型知识图谱—其它

Freebase:4千多万个实体(entity)，20亿多个实体与实体之间关系描述的facts

NELL(CMU): Never-Ending Language Learning：5千多万个实体与实体之间关系描述

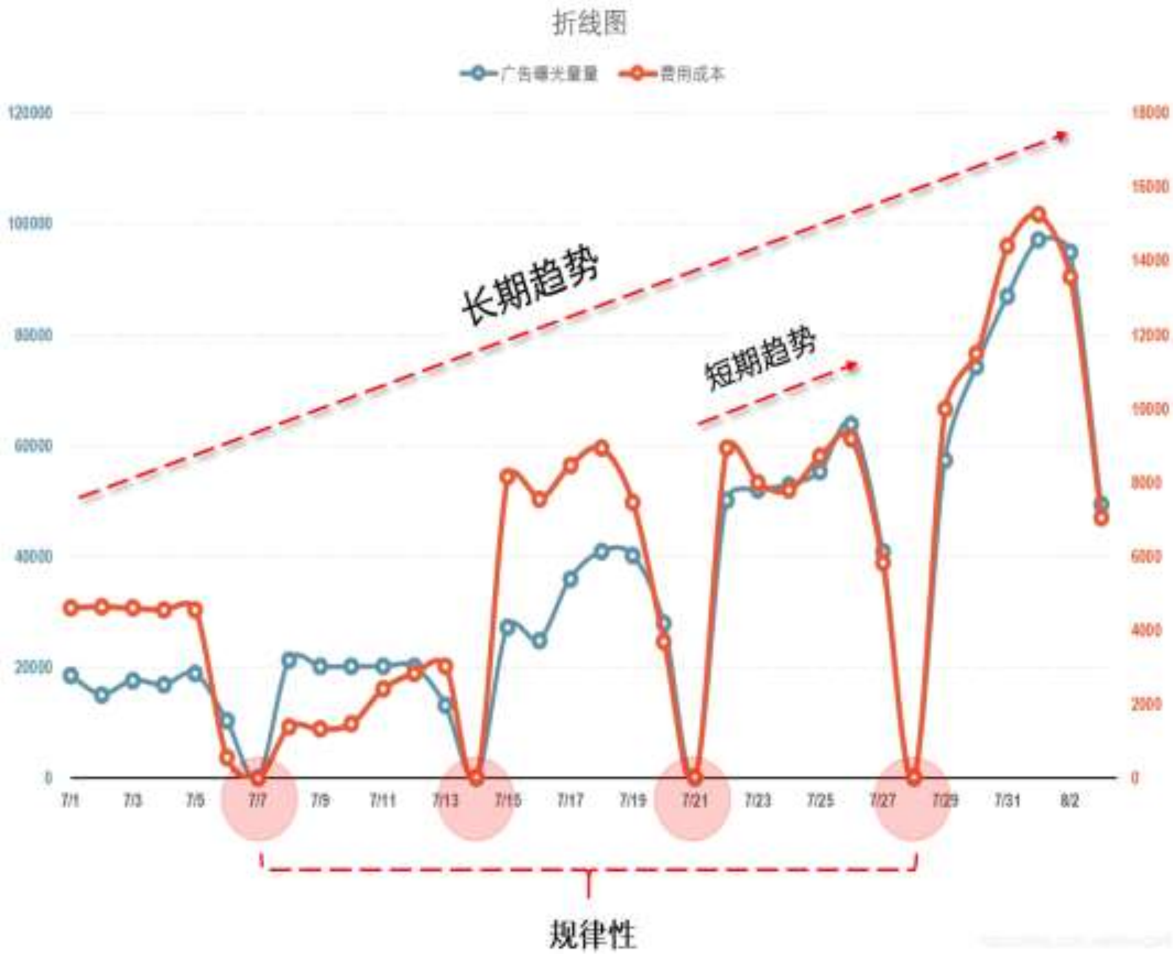
ReVerb: 1千5百万条实体与实体之间的关系描述

大 纲

- 跨媒体是什么？
- 跨媒体内容统一表示
- 跨媒体知识图谱构建
- **跨媒体关联分析与推理**
- 基于跨媒体分析的应用

关联性分析

图表相关分析法：



投放时间	广告曝光量(y)	费用成本(x)
2016/7/1	18,481	4,616
2016/7/2	15,094	4,649
2016/7/3	17,619	4,600
2016/7/4	16,825	4,557
2016/7/5	18,811	4,541
2016/7/6	10,430	568
2016/7/7	18	-
2016/7/8	…	…
2016/7/9	…	…

关联性分析

Pearson相关系数：

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

相关系数r的取值范围为$-1 \le r \le 1$。

$$\begin{cases} r>0 \text{ 为正相关}, r<0 \text{ 为负相关} \\ |r|=0 \text{ 表示不存在线性相关} \\ |r|=1 \text{ 表示完全线性相关} \end{cases}$$

$0<|r|<1$表示存在不同程度线性相关。

* Pearson相关系数要求连续变量的取值服从正态分布

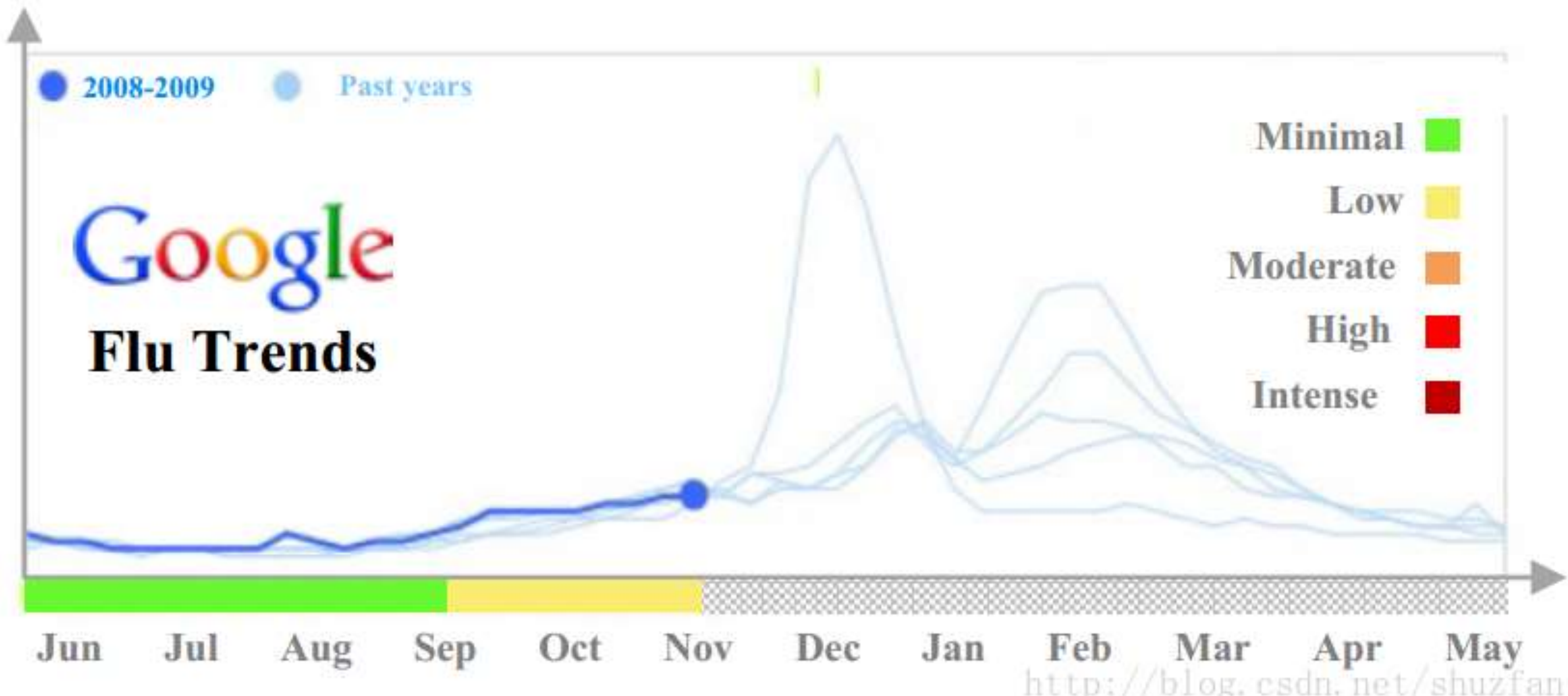
关联性分析

Spearman相关系数，也称为等级相关系数：

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(R_i - Q_i)^2}{n(n^2 - 1)}$$

对两个变量成对的取值分别按照从小到大（或者从大到小）顺序编秩，Ri代表xi的秩次，Qi代表yi的秩次，Ri-Qi为xi、yi的秩次之差。

谷歌流感



2008-2009
Past years
Google
Flu Trends
Minimal
Low
Moderate
High
Intense
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
http://blog.csdn.net/shuzfan

谷歌流感

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[5,6,3]

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can predict *x* has become commonplace (*5*–*7*) and is often put in sharp contrast with traditional methods and hypotheses. Although these studies have shown the value of these data, we are far from a place where they can supplant more traditional

ability and dependencies among data (*12*). The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scien-

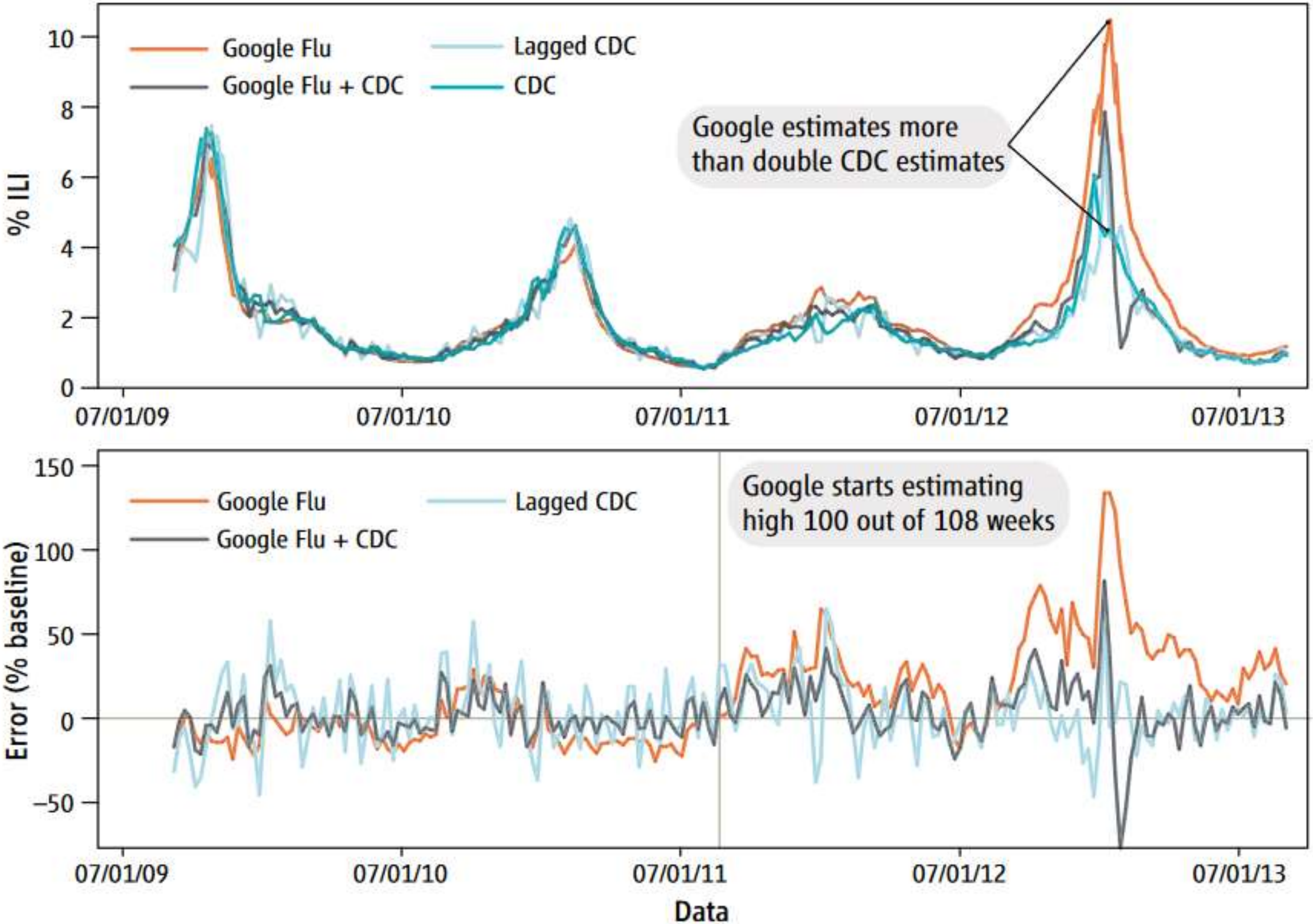
run ever since, with a few changes announced in October 2013 (*10*, *15*).

Although not widely reported until 2013, the new GFT has been persistently overestimating flu prevalence for a much longer time. GFT also missed by a very large margin in the 2011–2012 flu season and has missed high for 100 out of 108 weeks starting with August 2011 (see the graph). These errors are not randomly distributed. For example, last week's errors predict this week's errors (temporal autocorrelation), and the direction and magnitude of error varies with the time of year (seasonality). These patterns mean that GFT overlooks considerable information that could be extracted by traditional statistical methods.

Even after GFT was updated in 2009, the comparative value of the algorithm as a stand-alone flu monitor is questionable. A study in 2010 demonstrated that GFT accuracy was not much better than a fairly simple projection forward using

"**Big data hubris**" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis...The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis.

Lazer, D., Kennedy, R., King, G., Vespignani, A., The Parable of Google Flu: Traps in Big Data Analysis, *Science*, 343:1203-1205, 2014



GFT overestimation. GFT overestimated the prevalence of flu in the 2012–2013 season and overshot the

大 纲

- 跨媒体是什么？
- 跨媒体内容统一表示
- 跨媒体知识图谱构建
- 跨媒体关联分析与推理
- 基于跨媒体分析的应用

应用：跨媒体描述生成

实现跨媒体数据间的交叉翻译，并使用自然语言描述符联系理解跨媒体数据。

关键问题：

- 针对文本、图像、视频等的跨媒体描述符
- 认知、情感、推理间的联系。

(a) A dog is wearing a red sombrero

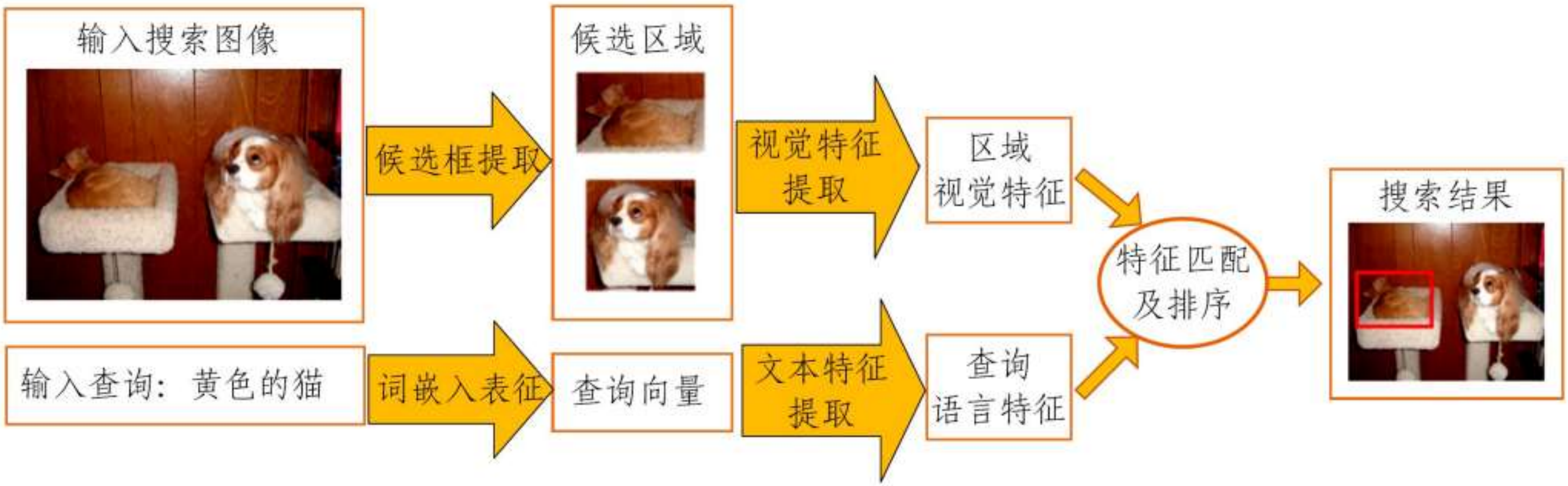
(b) Several cars and a motorcycle are on a snow covered street

(c) Some people in chairs and a child watch someone playing a trumpet

(d) A girl is putting her finger into a plastic cup containing an egg

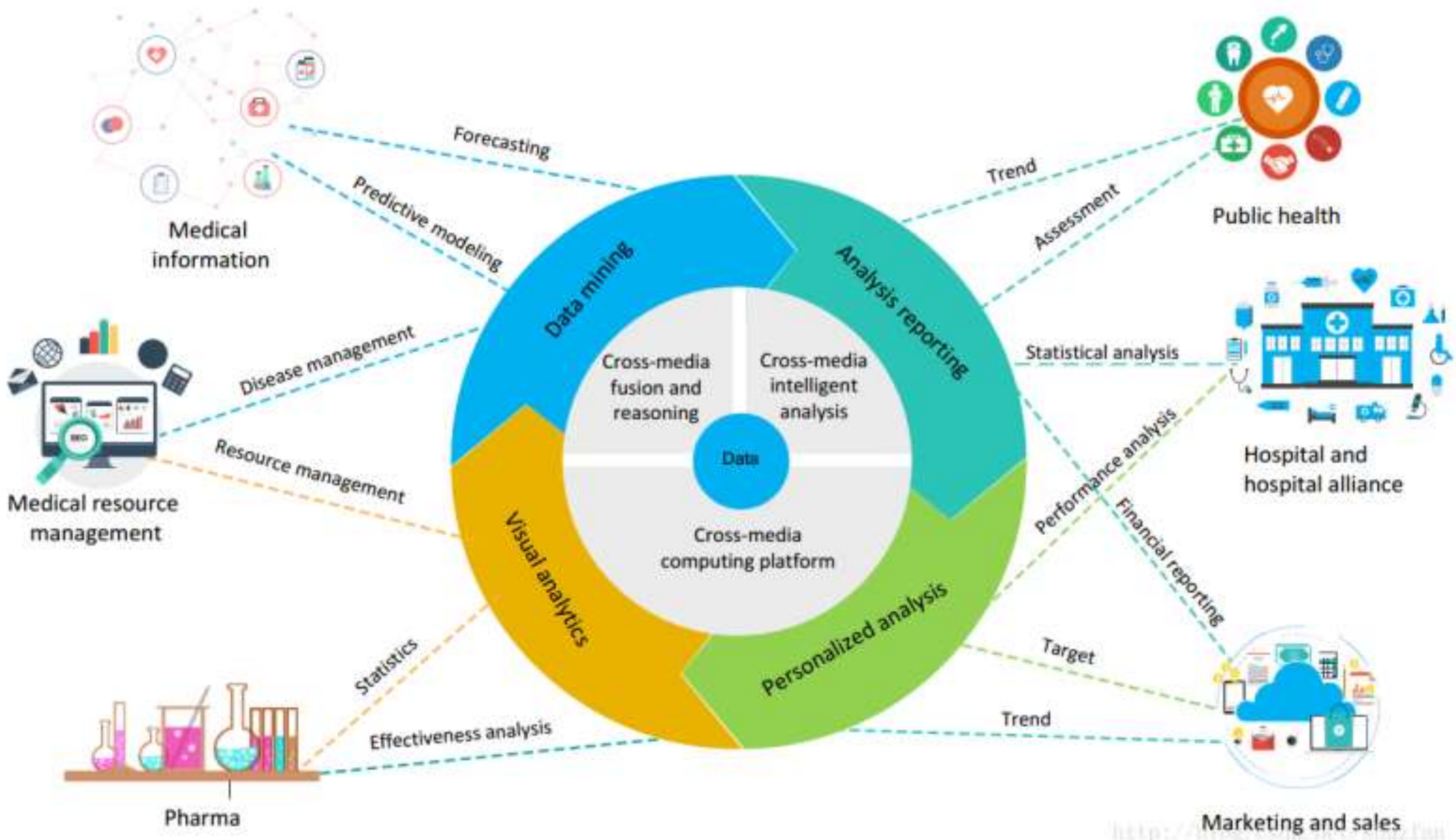
应用：跨媒体检索

用户向计算机提交一种类型的多媒体对象作为查询例子，系统可以自动找到其他不同类型、在语义上相似的多媒体对象，实现跨媒体检索。



应用：跨媒体精准医疗

如医疗数据的融合与推理，从而实现个性化精准医疗。
挑战： 跨媒体数据融合与推理能力不足；缺乏领域
专[illegible]力。
下



应用：反腐

26日凌晨：8.26陕西特大车祸交通事故

16时35分：网友将官员现场微笑照片上传微博

17时53分：微博报料官员姓名与单位

22时29分：网友微博怀疑“表哥”戴3.8万欧元名表

当天点击超16万人

23时57分：渤海论坛的微博发布“表哥”不同场合的名表

8月27日：奢侈品专家比对后证实网友的推测

8月31号：网上信息重组后的“表哥事件”出现各大主流媒体

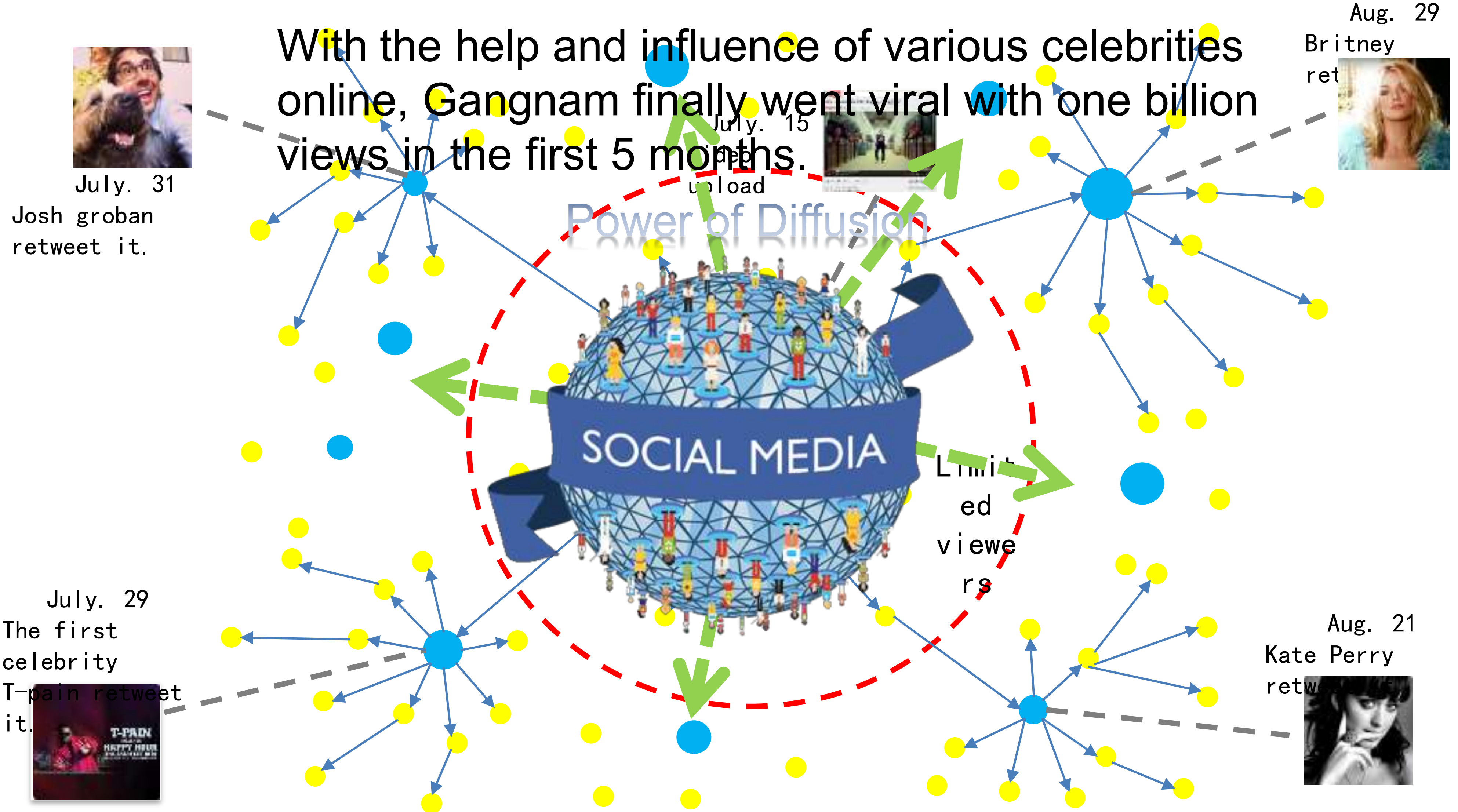
陕西“表哥”杨达才事件

应用：热点话题检测

Why Gangnam Style became a global hit?

应用：热点话题检测

With the help and influence of various celebrities online, Gangnam finally went viral with one billion views in the first 5 months.

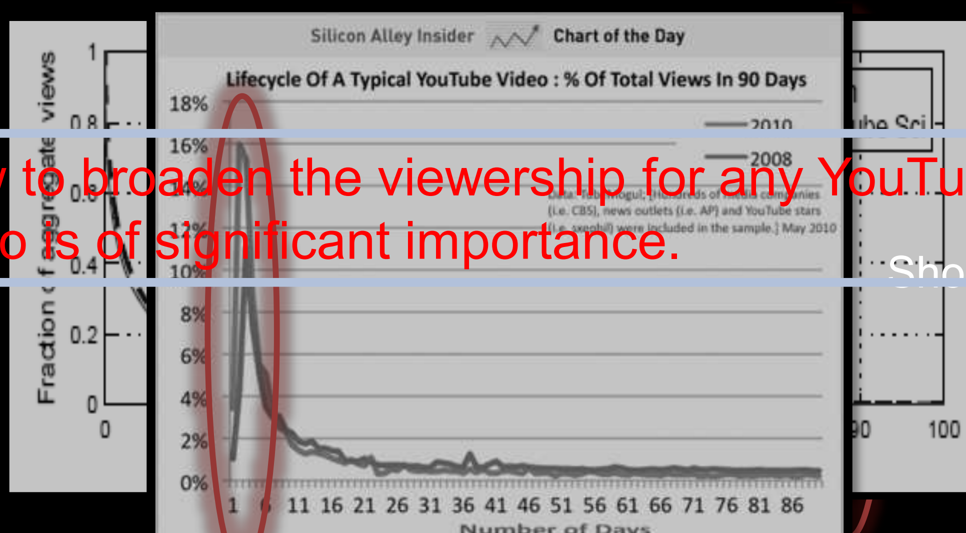


应用：热点话题检测

What Is Happening

➢ In YouTube

- YouTube exhibits limited propagation efficiency and many videos remain unknown to the wide public.
- Large quantities of video are uploaded to YouTube and the trend is growing year by year.



- How to broaden the viewership for any YouTube video is of significant importance.

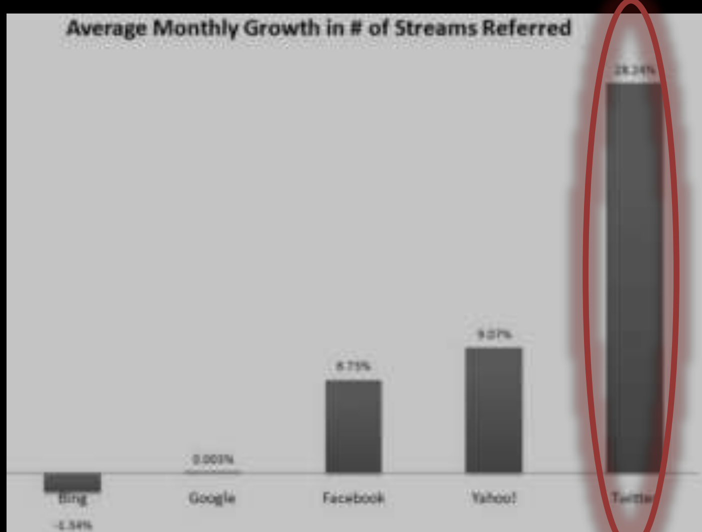
2 billion
Long tail effect
Short active life span

应用：热点话题检测

What Is Happening

- In Social Media

- External referrers such as social media websites arise to be important sources to lead users to YouTube videos.



- Twitter has been quickly growing as the top referrer source for web video discovery.

Thanks!

Q&A