文章编号：1003-0077(2016)02-0136-06

基于视觉显著计算的图像语义检索方法

柳 伟[1]，陈 旭[2]，梁永生[1]

(1. 深圳信息职业技术学院 信息技术研究所，广东 深圳 518172；
2. 深圳市海思半导体有限公司，广东 深圳 518000)

摘 要：网络标签已经开始广泛地用于图像内容的标注和分享，由于图像本身的差异和人们对图像的不同理解，对图像语义检索提出了新的挑战。该文首先引入视觉显著模型，突出图像的显著信息；然后提取视觉显著特征，建立图像内容的相似关系；最后基于随机漫步模型平衡图像内容及网络标签间的关系。实验表明该文提出的方法能够有效地实现图像的语义理解并用于图像检索。

关键词：随机漫步；图像分析；标签；网络标注

中图分类号：TP391 **文献标识码**：A

Image Semantic Retrieval Based on Visual Saliency Computation

LIU Wei[1], CHEN Xu[2], LIANG Yongsheng[1]

(1. Shenzhen Key Lab of Visual Media Processing and Transmission, Shenzhen Institute of Information Technology, Shenzhen, Guangdong 518172, China;
2. HiSilicon Technologies Co., Ltd, Shenzhen, Guangdong 518000, China)

Abstract: Internet labeling and tags have been used extensively to describe the image contents on the Web. To understand and utilize these tags for image semantic retrieval, this paper introduces a visual saliency model to emphasize the salient information, and then, extracts the visual feature to describe the similarity between images. At last, a novel random walk is proposed to balance the influences between the image contents and tags. Experiments show the effectiveness and feasibility of the proposed method when applied in image understanding and retrieval.

Key words: random walk; image analysis; tag; internet labeling

1 引言

图像检索广泛应用于搜索引擎，包括 Google，Yahoo，Bing 等。虽然目前基于内容图像检索技术有了一定发展，然而仍然未能跨越语义鸿沟，未能从图像中获取语义信息。因此，也就无法真正实现通过语义信息检索到相关图像。因此，如何能够突破语义鸿沟障碍，建立一种新颖有效的图像检索模型成为当前的研究热点和难点。

Web2.0 技术的发展给可视媒体信息处理带来新的变革，通过社会多媒体计算(Social Multimedia Computing)和协同标注(Collaborative Tagging)，用户既是信息消费者，又是信息提供者。用户可以为网络中的图像设立语义标签进行共享，如 Flickr，ESP 等[1]。这些手动标注的标签为图像提供了很有意义的描述信息。然而，这种社会化信息严重依赖领域知识和先验模型，不可避免地加入了许多噪声。同时，图像具有与常规字符数据完全不同的特性，如感知特性和时间、空间特性等。从模拟生物视觉的角度出发，将认知科学领域的研究成果应用于语义关联可在新计算模型方面有新突破。

近年来，有学者对此做了一些研究，文献[2]通过对纹理空间的相似性估计来构造语义关系图。文献[3-4]以聚类方式对语义距离，图像内容进行估计得到图像的语义信息，文献[5-6]利用反馈信息提高

收稿日期：2014-01-05 **定稿日期**：2014-04-15

基金项目：国家自然科学基金(61172165)；广东省自然科学基金(S2011010006113)；深圳市科技计划项目(JCYJ20120615103240795)

语义检索性能。这些方法缺乏普适性且复杂度较高。文献[7]根据图像显著物体区域对图像内容信息进行估计，然而该方法需要依靠准确的图像分割及显著区域检测。

不同于基于内容的图像检索，本文通过视觉显著计算提取符合人眼特性的视觉特征，基于图论模型描述图像视觉特征与语义概念之间的关系，取代复杂模型和基于训练样本的机器学习方法。采用马尔科夫随机漫步模型使视觉特征和语义标签的关联趋于稳定，最终实现高效、准确的图像分析和语义检索。

2　视觉显著计算

视觉显著计算(Visual Saliency Computation)模拟人类视觉系统面对复杂场景时迅速选择少数几个显著区域进行优先处理的过程，是人类视觉特性研究中引起广泛关注的一个分支。基于视觉显著计算可以实现人类认知的重要视觉信息的自动提取，定位和挖掘，从而降低分析难度，提高计算效率。由于可视媒体具有非结构化、高维度和语义多样性等特点，采用视觉显著计算能为海量信息的表达与组织、高效分析、信息推理和知识提取提供新的思路和方法。

人眼视觉具备排他性(即同一时刻只能存在一个注视内容)，同时人眼视觉系统具备“Cognitive Miser”特性[8]，具体表现在视觉注意转移的眼动行为。因此，视觉显著计算包括视觉注意计算模型和注视转移模型两部分。视觉显著计算的过程如图 1 所示，输入图像经过初始显著计算获得显著图，结合注意转移分析得到注视区域，最后获得总显著图。

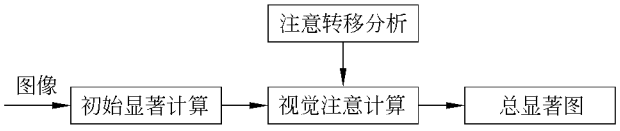

图 1　视觉显著计算过程

由于 Harel 方法[9]采用图论和随机行走来检测突变信号，适合描述视觉转移机制，因此，本研究在初始显著计算部分采用该方法。最终显著区域计算在初始显著区域计算结果的基础上，结合考虑了视觉注视转移及延迟方面的视觉特性进一步优化。

在获取初始显著计算结果后，由于人眼注视画面过程中会发生注视、眼跳和追随运动，而人眼视觉及心理学相关研究表明[7,9]，人眼对区域平均眼跳延迟，即注视的时长约为 350 毫秒，平均眼跳时长约为 70 毫秒。因此为方便计算，设置时间参数 Δt 如式(1)所示。

$$\Delta t^n \propto R^2 \ (mean(\Delta t^n)) \text{ s.t. } \quad \text{mean}(\Delta t^n) = 420 \tag{1}$$

其中，n 表示第 n 个显著区域；$R2$ 表示显著区域的面积；$\propto$ 表示注视时间和区域面积成正比。构造注视转移矩阵 P_t 如式(2)所示。

$$\begin{cases} P_t = \begin{bmatrix} m_{11}^n & m_{12}^n & \cdots & m_{1m}^n \\ m_{21}^n & m_{22}^n & \cdots & m_{2m}^n \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1}^n & m_{n2}^n & \cdots & m_{nm}^n \end{bmatrix} \\ m_{ij,s}^{n+1} = m_{ij,s}^n + \Delta k \quad s.t. \quad m_{ij,k}^n \leqslant 1 \\ \Delta k \propto \dfrac{1}{R^2} \Delta t \\ m_{ij,s}^n = [m_{ij}^n \in R^n \ (p_x, p_y)] \end{cases} \tag{2}$$

对于同一幅图像，该矩阵表达了不同时间段范围内人眼所注视区域的可能性，模拟 Δt 时间段内人眼的扫视范围。初始注视转移矩阵 P_t 为全 0 矩阵；n 为显著区域序；$m_{ij,s}$ 为所在显著区域中的元素；Δk 为注视增长单元，它与显著区域大小成反比关系。随时间增长增加显著区域注视值，当显著区域内注视值均为 1 时，根据人眼排他性，视点转移到下一个显著区域。

根据人眼返回抑制性及邻近优先性，结合初始显著图 $F(x,y)$、返回抑制图 $I^n(x,y)$ 及邻近优先图 $M^n(x,y)$ 信息，得到综合显著图 $D^{n+1}(x,y)$，下一个显著区域 R^{n+1} 位置计算公式如式(3)所示。

$$\begin{cases} D^{n+1}(x,y) = F(x,y) I^n(x,y) M^n(x,y) \\ (p_x^{n+1}, p_y^{n+1}) = \arg\max_{(x,y)} D^{n+1}(x,y) \\ (p_x^{n+1}, p_y^{n+1}) \in R^{n+1} \end{cases} \tag{3}$$

则最终的显著区域结果可表示为式(4)。

$$S' = P_t \cdot S \tag{4}$$

其中 S 为初始显著计算结果；P_t 为注视转移矩阵。

3　基于视觉显著计算的语义检索方法

3.1　视觉显著区域特征提取

根据图像质量分析的理论，图像中包含的信息并不等于人眼视觉系统获得的信息[10]。因此，将总显著图中的像素值，即显著度作为线性变换的调节因子，对原图像进行线性变换滤波，使之与人眼视觉感知系统一致，如式(5)所示。其中，P_s 为原图像，S' 为式(4)得到的总显著图，P 为经过显著图线性变换

后的图像，C 为常量。通过视觉显著性滤波，增强了图像中的显著信息，抑制了非显著信息，使提取的视觉显著特征更符合人眼视觉感知的要求。

$$P=\frac{\left(1+\frac{S'}{C}\right)\times P_s}{\max\left(\left(1+\frac{S'}{C}\right)\times P_s\right)}\times 255 \qquad (5)$$

图像全局特征在图像处理中有着大量的研究基础，从计算简单且易于存储的角度出发，本文提取四种全局特征。

(1) 颜色特征

将图像分割为 3×3 单元，然后计算每个单元 R，G，B 分量对应的颜色均值，颜色方差及颜色偏度，共 81 维特征。

(2) 局部二元模式特征(LBP)

LBP 能有效纹理描述算子，度量和提取图像局部纹理信息，其均匀模式共有 58 个，加上非均匀模式编码，共 59 维特征。

(3) Gabor 小波纹理特征

为了能够获取 Gabor 纹理特征，每个图像变换为 64×64 像素，利用 Gabor 小波变换获取五个等级八个方向纹理子图像，对于每个子图像分别计算对应的均值，方差和偏度，共 120 维 Gabor 纹理特征向量。

(4) 边缘特征

提取每个图像的边缘方向直方图，首先将图像转换为灰度图，然后利用 Canny 边缘检测算子获取 36 个(每 10 度为一单位)边缘方向直方图，加上额外非边缘像素个数信息，共组成 37 维边缘特征。

由四个全局特征构造 297 维向量，进一步归一化(零均值、单位方差)来表示每一幅图像特征信息。

3.2 相似性度量

在视觉显著计算的基础上，结合提取的图像特征信息建立图像相似图。D 表示图像集，使得 $d_p\in D, p\in[1,|D|]$，d_p 为图像特征信息，采用加权余弦函数来估计图像 d_p 和 d_q 间的相似性，如式(6)所示。

$$Sim(d_p,d_q)=\frac{d_p\cdot d_q}{\|d_p\|\ \|d_q\|} \qquad (6)$$

根据计算图像之间的相似性度量值，可以构建图像间的相似图。通常构建相似图包括 K 最邻近节点算法(KNN)，eNN，exp-weighted 图等方法，其中 KNN 方法在测试中能获得较好性能，因此采用 KNN 算法。当 d_q 为 d_p 的 k 最近邻节点之一时，建立从 d_p 到 d_q 的连接，权重为 $\mathrm{sim}(d_p,d_q)$。图 2 为 k 为 1 时的图像相似关系示意图。

3.3 语义检索模型

除了图与图之间的相似关系，还需要建立图像和标签之间的相似关系。由于图像与标签间存在着双向关系，不能直接简单的将两者间的相似关系结合到图与图间的相似图中。为了解决这个问题，本文将双向关系转换为单向关系，具体如图 3 所示。

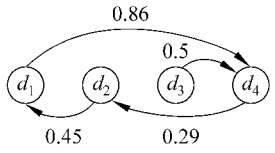

图 2 图像相似性关系

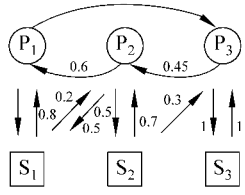

图 3 随机漫步模型示意图

图 3 中，P_i 和 S_i 分别为图像和语义标签。图像与语义标签箭头表示为图像与该语义标签的相关度，在马尔科夫模型中用单步跳转概率来描述，具体如式(7)所示。

$$P''_{t+1|t}(j\mid i)=\begin{cases}\dfrac{C_{ij}}{\sum_{p\in S}C_{ip}} & i\in P, j\in S\\ \dfrac{C_{ij}}{\sum_{p\in P}C_{ip}} & i\in S, j\in P\end{cases} \qquad (7)$$

若 $i\in P, j\in S$，则 C_{ij} 为语义 S_j 被标注为图像 P_i 的次数，$\sum_{p\in S}C_{ip}$ 为图像 P_i 被标注的次数。若 $i\in S, j\in P$，则 C_{ij} 为语义 S_i 被标注为图像 P_j 的次数，$\sum_{p\in P}C_{ip}$ 为语义 S_i 标注到所有图像的次数。

随机漫步模型只能跳转到相连接的节点中，实际上，非连接节点之间也存在着随机连接关系。比如图像间的相似性可以通过余弦函数计算，但是某些图像间也存在隐含关系。因此，设定 m 为概率因子(实验测试中 m 设为 0.95)，g 为归一化随机分布矩阵，则修改的转移概率矩阵如式(8)所示。

$$P'_{t+1|t}(j\mid i)=mP''_{t+1|t}(j\mid i)+(1-m)g \qquad (8)$$

如图 3 所示，跳转图包括了图像间相似图和图像与语义标签之间的双向图。这两种图的权重针对不同应用场合可以互不相同，使得模型更具灵活性。模型中对转移概率矩阵引入权重因子 λ，如式(9)所示。

$$P_{t+1|t}(j\mid i)=\begin{cases}\lambda P'_{t+1|t}(j\mid i) & i\ or\ j\in S\\ (1-\lambda)P'_{t+1|t}(j\mid i) & i,j\in P\end{cases} \qquad (9)$$

当 $\lambda=1$，随机漫步只作用于图像与语义标签之间的双向图；当 $\lambda=0$，随机漫步只作用于图像与图像之间的双向图。

确定转移概率矩阵 P 后，采用马尔科夫模型进行随机漫步，计算从节点 i 到节点 j 的转移概率如式(10)所示。

$$P_{t|0}(j\mid i)=[(((v_i)P)P)\cdots P]_j \quad (10)$$

其中，v_i 为初始跳转行矢量，起步矢量元素 i 为 1，其他元素为 0。t 控制随机漫步步数。通过随机漫步模型，实现了图像与图像、图像与语义标签之间的准确关联。

4　实验结果与分析

实验数据集为 flickr 上有标注的 2 000 幅自然图像，其中包括高速公路、街道、房屋、城市、森林、海滩、风景和人物等八种语义，每帧图像的大小为 480×360 像素。

4.1　视觉显著计算

首先由十名专业人员使用 SMI RED250 型眼动仪观看图像，根据眼动测试结果获取人眼所可能注视的区域范围，构造客观注视图（Ground Truth）。获取客观注视图后，将本文算法与其他三种视觉显著模型[11-13]计算结果进行对比，如图 4 所示。在性能比较上，接受者特性曲线（ROC）是很多算法中评价视觉效果最常用指标，形式主要包括 ROC 曲线及 ROC 曲线下面积 AUC，各种视觉显著计算的 ROC 曲线如图 5 所示。从图 4 和图 5 可以看出，本文提出的视觉显著计算模型获得的视觉显著区域与人眼视觉感知的结果更加一致。

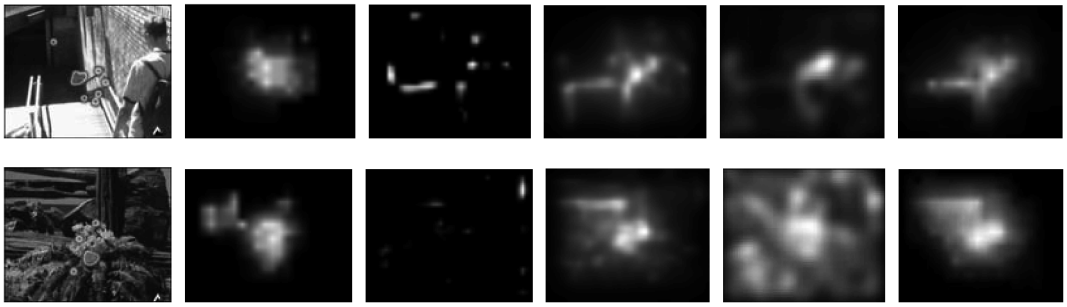

图 4　视觉显著计算结果对比

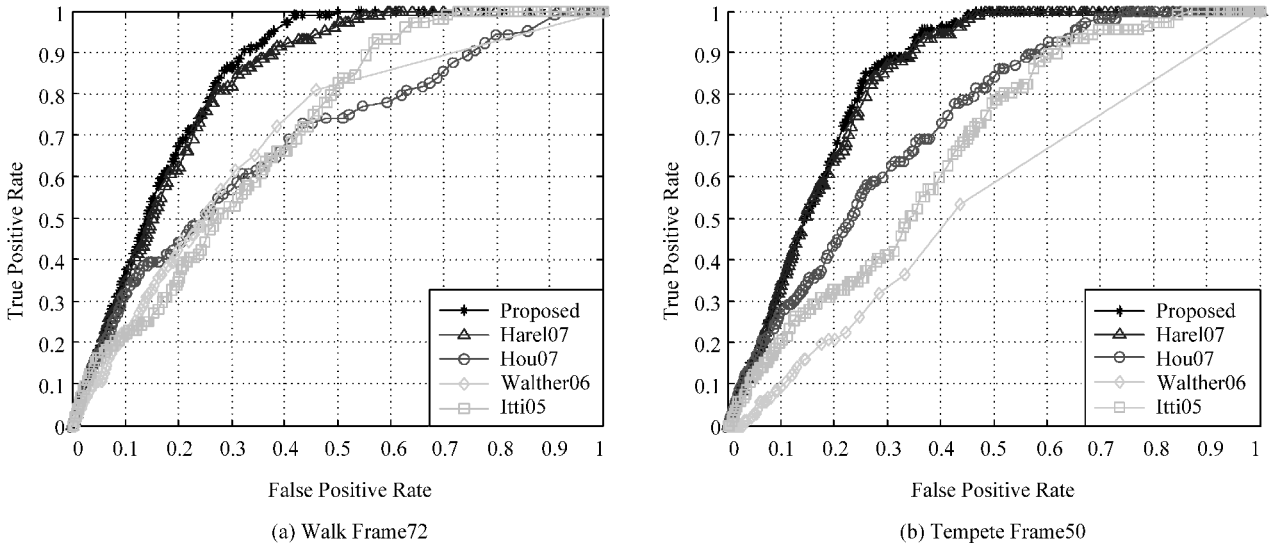

图 5　各种视觉显著计算算法的 ROC 曲线

4.2　图像语义检索

如 3.3 所述，图像检索模型分为图像视觉相似性关联和语义关联两个部分。因此，图像在检索模型中经历了特征提取，相似性计算和语义标签关联等操作。选取图 6 所示的示例图像，由于网络语义

高山 森林 风景 人物 …

图6 示例图像及其网络标注的语义标签

标签可能包含有着大量的语义标签信息，为了降低计算复杂度及降低筛选干扰，需要通过语义筛选提取出准确的语义信息。经过随机漫步模型迭代以后，语义筛选的结果如图7所示。可以看出，最优的语义标注为“人物”和“风景”。

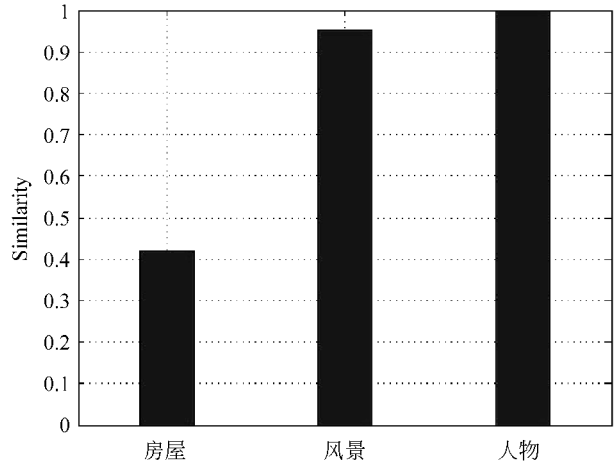

图7 语义标签筛选

仅根据图像视觉特征信息的检索结果如图8所示，检索结果按照相关性依次排序。由于经过语义筛选提取了准确的语义信息，即“人物”和“风景”，可以进一步结合图像视觉特征信息，检索对应语义及内容相似的图像，检索结果如图9所示。由图9可以看出，通过加入语义信息，图像中有人物的相关性排序更加靠前，从而提高了检索的准确率。

示例图像 检索结果

图8 基于视觉相似性的图像检索

示例图像+语义标签 检索结果

图9 结合视觉相似性和语义的检索

实验结果表明，本文所提方法避免了图像分割、识别及学习训练等复杂操作，能够利用网络语义标签有效地实现图像检索。

5 结论

本文提出了一种基于视觉显著计算、网络语义标签及马尔科夫随机漫步模型的检索方法。该方法建立了图像视觉内容特征以及网络语义标签间的关系，由于避免了学习训练等复杂运算，适用于内容丰富的自然图像数据集。数据集测试表明了视觉显著计算模型的有效性以及结合语义检索的准确性。

下一步的工作将研究更好的视觉显著性滤波方法，同时在更大规模数据集上统计图像检索的查到率和查准率。

参考文献

[1] http://www.flickr.com/[OL], 2013

[2] C Wang, L Zhang, H J Zhang. Learning to reduce the semantic gap in web image retrieval and annotation [C]//Proceedings of SIGIR'08, Singapore. 2008: 355-362.

[3] G J Qi, X S Hua, H J Zhang. Learning semantic distance from community-tagged media collection [C]// Proceedings of MM'09, Beijing, China. 2009:243-252.

[4] L Wu, X S Hua, N Yu, W Y Ma, et al. Flickr distance[C]//Proceedings of MM'08, Vancouver, BC, Canada. 2008:31-40.

[5] Y Rui, T S Huang, M Ortega, et al. Relevance feedback: A power tool in interactive content-based image retrieva[J]. IEEE Trans. Circuits Syst. Video Technol. 1998, 8(5): 644-655.

[6] S Tong, E Chang. Support vector machine active learning for image retrieval[C]//Proceedings of MM'01, Ottawa, ON, Canada. 2001:107-118.

[7] J Fan, Y Gao, H Luo, et al. Automatic image annotation by using concept-sensitive salient objects for image content representation[C]//Proceedings of SIGIR'04, Sheffield, U. K. 2004:361-368.

[8] T D Keech, L Resca, Eye movements in active visual search: a computable phenomenological model [J], Attention, Perception, & Psychophysics, 2010, 72 (2): 285-307.

[9] Jonathan Harel, Christof Koch, Pietro Perona. Graph-based visual saliency [C]//Proceedings of Neural Information Processing Systems(NIPS),2006:545-552.

[10] H R Sheikh, A C Bovik. Image information and visual quality[J]. IEEE Transaction on Image Process, 2006, 15(2): 430-444.

[11] Walther D, Koch C. Modeling attention to salient proto-objects[J]. Neural Networks, 2006, 19(9): 1395-1407.

[12] Jonathan H, Christof K, Pietro P. Graph-based visual saliency[C]//Proceedings of the International Conference on Advances in Neural Information, 2007: 545-552.

[13] Hou X, Zhang L. Saliency detection: A spectral residual approach[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

柳伟(1973—),博士,教授,主要研究领域为图像处理、视频编解码、基于内容的多媒体信息处理、多媒体数据库系统。
E-mail: liuwei@sziit.edu.cn

梁永生(1971—),博士,教授,主要研究领域为计算机网络与数据通信、信号处理与模式识别。
E-mail: liangys@sziit.edu.cn

陈旭(1984—),博士,工程师,主要研究领域为视频编码。
E-mail: anderson.chen@huawei.com

知识图谱与问答系统前沿技术研讨会暨清华大学“计算未来”博士生论坛顺利召开

2016 年 4 月 17 日,知识图谱与问答系统前沿技术研讨会暨清华大学“计算未来” 博士生论坛在 FIT 大楼多功能报告厅召开。本次研讨会由中国中文信息学会语言与知识计算专业委员会、中国中文信息学会青年工作委员会青工委和清华大学计算机科学与技术系联合举办。研讨会由清华大学李涓子教授、中科院自动化所刘康博士和清华大学刘知远博士担任主席,博士生论坛由林衍凯同学担任主席。

研讨会邀请多位在知识图谱领域享有盛名的学者专家进行专题报告,他们是:自然语言处理著名学者、前 Google 高级科学家林德康博士,中科院软件所副研究员韩先培博士,文因互联网创始人鲍捷博士,中科院自动化所副研究员刘康博士,百度自然语言处理部高级研究员马艳军博士。研讨会还邀请清华大学四名博士生同学和中科院自动化所两位博士生同学做口头报告,以及 11 名清华大学的博士生同学做海报展示。

本次研讨会共吸引了学术界和产业界 200 余名老师和同学参加,会场座无虚席,学术交流气氛浓厚。研讨会首先邀请林德康博士做了题为“From Search Engine to Answer Engine”的特邀报告,介绍他在谷歌研制自动问答系统的实践经验与思考。接下来,研讨会分 “知识图谱的构建与表示”和“智能问答系统”两个专题,分别开展了深入的报告与热烈的研讨。与会老师和同学均表示收获颇丰。