



北京交通大学
BEIJING JIAOTONG UNIVERSITY



大数据在医疗行业中的应用





医疗大数据的概述

01 医疗大数据的定义与来源



医疗大数据的定义

医疗大数据是指在医疗活动中产生的数据的集合。既包括个人全生命周期过程中，因体检、门诊、住院等健康活动产生的大数据，又包括医疗服务、疾病防控、健康保障和食品安全、养生保健等多方面的数据。



医疗大数据的来源

- (1) 患者就医流程所产生的数据
- (2) 检验中心数据
- (3) 制药企业、基因测序数据





医疗大数据的概述

02 医疗大数据的特点

多态性

医疗数据包括纯数据，信号图谱，文字描述，声音资料，动画数据等等。

不完整性

在就医过程中会有各种原因导致医疗数据的不完整。

时间性

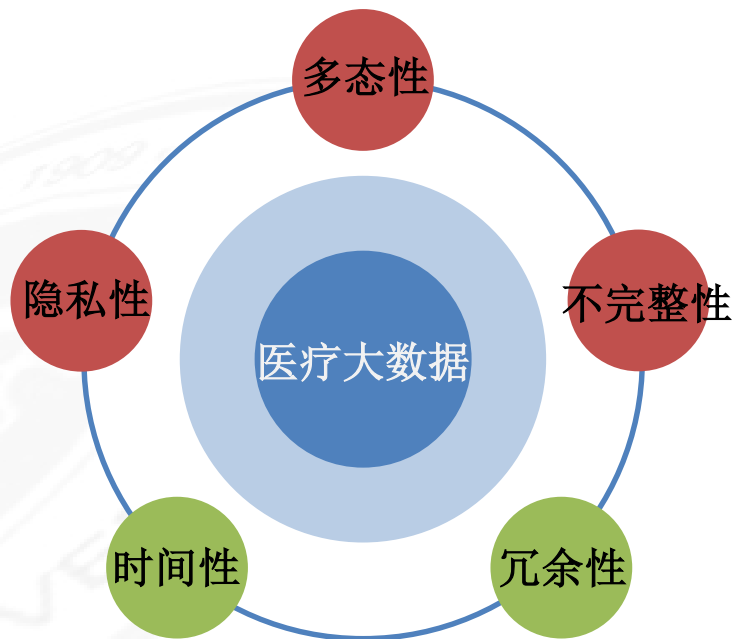
大多医疗数据都是具有时间性、持续性的。

隐私性

隐私性也是医疗数据的一个重要特性，很多医院的临床数据系统都是相对独立的局域网络。

冗余性

同一个患者在不同的医院就医可能会产生相同的数据信息；在同一家医院中也可能存储着大量重复的数据。





医疗大数据的应用

- **服务居民。**个性化健康保健指导，使居民能在医院、社区及线上的服务保持连续性。
- **服务医生。**临床决策支持，如用药分析、药品不良反应、疾病并发症、治疗效果相关性分析、抗生素应用分析；或是制定个性化治疗方案。
- **服务科研。**包括疾病诊断与预测、临床实验数据的分析与处理等方面；提供最佳治疗途径。
- **服务管理机构。**规范性用药评价、管理绩效分析。
- **公众健康服务。**危及健康因素的监控与预警、网络平台、社区服务等方面。





医疗大数据的现状

- 医疗数据量呈爆炸式增长；
- 医疗信息系统孤岛众多，数据的使用和管理越来越复杂；
- 临床业务人员对临床数据的访问和应用提出更高的要求；
- 以患者为中心的信息在医院内外、甚至跨机构、跨领域都有共享的需求。如何更高效的管理医疗大数据？是一个值得深思的问题。
- 各种临床数据以患者为中心的统一管理、统一归档，包括影像、文字、图片、视频等；

IDC：医疗行业数据量每年增长48%

2014-12-04 10:04 作者佚名 来源：腾讯在线

根据IDC撰写的一项最新垂直行业简报，医疗行业在数字宇宙中占比显著，数据量每年以48%的速度增长，是增速最快的行业之一。这份名为《数字宇宙驱动医疗行业数据增长》的报告揭示了未来几年里，带来更多数据增长并且更加依赖数据的医疗行业的关键趋势。





医疗大数据的建设难点

数据存在多源异构问题：

- 医院信息系统不同时期建设，缺乏统一规划。
- 各医院的信息系统发展不均衡。
- 部分医院的信息系统厂商多，系统多，软件及硬件 服务都各自独立，缺乏合理接口。
- 医疗数据收集的渠道不畅。



医疗大数据的建设难点

数据互联互通问题：

- 观念保守：医院管理者对信息开放持保守观念；对数据安全、如何保护患者隐私等问题心里没底；整个行业缺少数据开放案例。
- 政策滞后：数据开放法律法规制定滞后。
- 标准不足：现在制定发布的一些标准，偏重于数据集标准，而对数据字典关注不足，对诊断编码（ICD-10）、药品编码、耗材编码、收费项目编码关注不足，没有统一标准。



医疗大数据的建设难点

医疗大数据的数据安全



个人 (Personal)

- 患者隐私保护

医疗机构 (Medical Institution)

- 自身与患者双重保护

第三方 (Third party)

- 如何合作

政府 (Government)

- 政策、国家安全



医疗大数据-文字

1. 语言处理技术

建立标准
术语系统



医学知识
模型



信息提取



文本模块
化

2. 数据挖掘技术

机器学习

统计学习

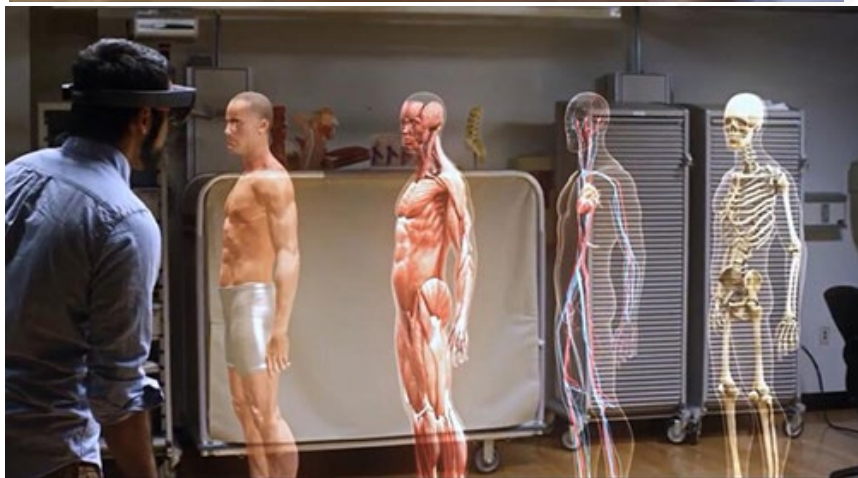
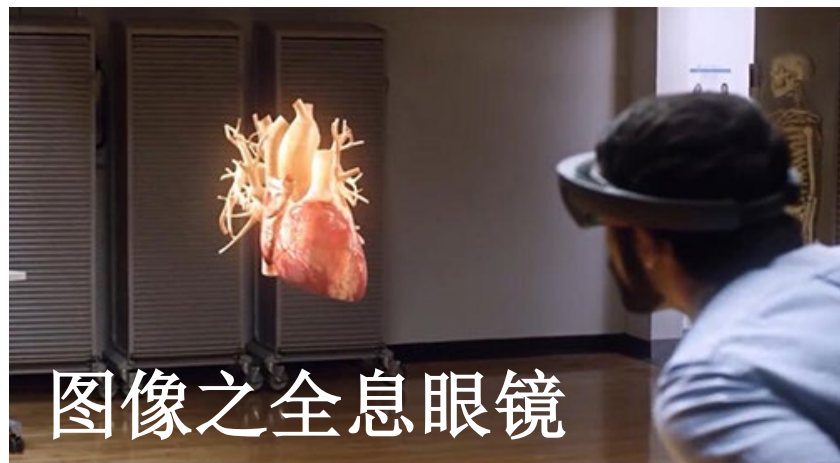
规则归纳

3. 应用

电子病历数据挖掘支持：临床决策、建立临床路径、临床指南，诊疗过程追踪



医疗大数据-图像



生物特征识别技术通常按照，扫描、数字化处理、分析、特征提取、存储、匹配分类几个步骤处理。目前扫描数字化处理已经相对成熟，主要的研究集中在分析和特征提取方面。

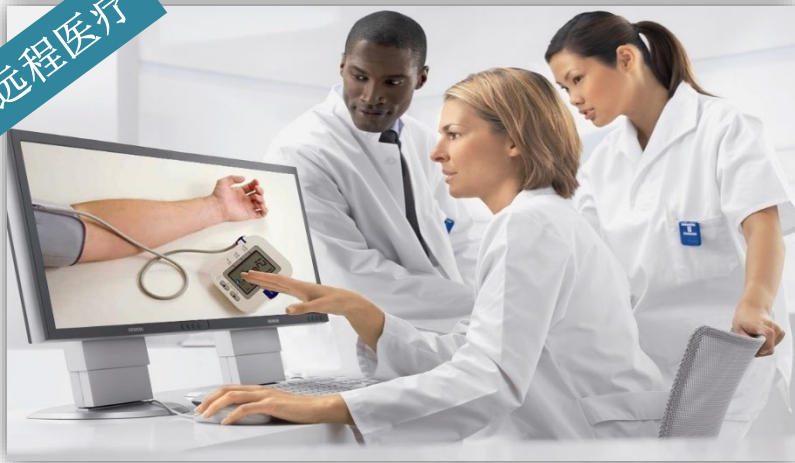
计算机辅助诊断技术：医学CAD

- 1、图像预处理：灰度直方图、去噪、图像增强
- 2、图像特征提取：边缘分割、灰度共生矩阵
- 3、图像分类：支持向量机算法（神经网络的一种衍生算法）



医疗大数据处理-视频

远程医疗



在线教育





医疗大数据-设备



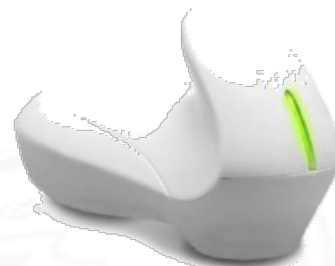
智能血糖仪



智能血压计



智能心电仪



智能胎心仪



智能温度计



智能健康秤



智能腕带-心率、运动



智能孕表



智能手表-实时监测心率、
体温、运动和睡眠追踪



肌肤检测仪



医疗大数据的应用

01

大数据与智慧医疗

智慧医疗通过整合各类医疗信息资源，构建药品目录数据库、居民健康档案数据库、影响数据库、检验数据库、医疗人员数据库、医疗设备数据库等卫生领域的**六大基础数据库**。

医生可以随时查阅病人的病历、病史、治疗措施和保险细则，随时随地快速制定诊疗方案；也可以让患者自主选择更换医生或医院，患者的转诊信息及病历可以在任意一家医院通过医疗联网的方式调阅。

智慧医疗的核心就是“以患者为中心”，给予患者以全面、专业、个性化的医疗体验。

智慧医疗的优点：

- (1) 促进优质医疗资源的共享
- (2) 避免患者重复检查
- (3) 促进医疗智能化

智慧
医疗





医疗大数据的应用

02

流行疾病监控预警

以搜索数据和地理位置信息数据为基础，分析不同时空尺度的人口流动性、移动模式和参数，进一步结合病原学、人口统计学、地理、气象和人群移动迁徙、地域之间的因素和信息，可以建立流行病时空传播模型，确定流感等流行病在各个区域传播的时空路线和规律，得到更准确的态势评估和预测。





医疗大数据的应用

03

健康管理

健康管理是通过客户的基因数据、历史疾病信息、电子健康档案、用药数据等开展系统的管理，为客户提供形式多样、内容丰富的个性化医疗及健康管理服务。

借助物联网、智能医疗器械、智能穿戴设备，实时收集居民的健康大数据，通过对体征数据的监控，实现健康管理。

利用大数据技术，对个人健康进行全生命周期管理，实现任何时间、任何地点都可以访问相关信息，从而保证了健康信息的一致性、连续性。

智能健康管理是将人工智能技术应用到健康管理的具体场景中，目前主要借助于可穿戴设备来实现。





医疗大数据的应用

04

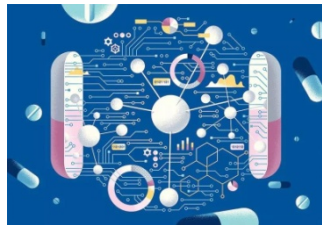
药物研发

从药物研发前到药物投入市场销售的每一个阶段，大数据都能起到积极的作用。

在药物研发的早期阶段，通过医疗大数据的整理分析，可以了解各区域药品的需求走势。医药公司如果能够对医疗大数据进行充分的运用，不仅能大幅度地降低运营成本，还有可能比预计时间提前将新药投入市场销售，这样可以争取更早获得利润回报。

药品成功研发以后，整合大数据分析所研发药品的市场需求，寻找投入产出比的最佳方案，确定最优资源组合，来达到节约成本的目的。

药品投入市场前，分析药物的副作用以及药品可能会产生的不良反应，通过大数据扩大采样分布范围和样本数，使数据分析结果更科学，有效减少企业运营成本，加快研发成果上市。





医疗大数据的概述

Valence Health

Valence Health 使用MapR公司的数据融合平台 (Converged Data Platform) 来建立一个数据湖并作为公司主要的数据仓库。

Valence每天从3000个数据输入源接收45种不同类型的数据。这些关键数据包括实验室测试结果、患者健康记录、处方、疫苗记录、药店优惠、账单和付款, 以及医生和医院的账单, 用来提升决策来改善医保结果和财务状况。该公司快速增长的客户和日益增加的相关数据量正在压垮现有的技术基础设施。

在采用MapR的解决方案之前, 如果收到一个数据源发来的2千万个实验室测试结果, 他们需要22个小时来处理这些数据。MapR把这个处理时间从22小时降到20分钟, 并且使用更少的硬件。提升了医保结果和财务状况。



医疗大数据的概述

Novartis Genomics

下一代基因测序（NGS）是一个经典的大数据应用，它面临双重的挑战，即巨量原始异构的数据，以及NGS最佳实践的快速变化。另外，许多前沿研究需要与外部组织的不同数据进行大量的交互。这就需要强大的工作流程工具来处理大量的原始的NGS数据，而且足够灵活以跟上快速变化的研究技术。它还需要一个方法来将这些大量外部组织的数据有意义地整合到Novartis的数据，如1000 Genomes, NIH 的 GTEx (Genotype-Tissue Expression, 基因型组织表达)和 TCGA (The Cancer Genome Atlas, 癌症基因组图谱)-特别是临床数据，表型性数据，实验数据和其他相关数据。



医疗大数据的概述

沃森医疗

IBM最新沃森医疗保健内容分析预测技术，该技术允许企业找到大量病人相关的临床医疗信息，通过大数据处理，更好地分析病人的信息。在加拿大多伦多的一家医院，针对早产婴儿，每秒钟有超过3000次的数据读取。通过这些数据分析，医院能够提前知道哪些早产儿出现问题并且有针对性地采取措施，避免早产婴儿夭折。为了让更多的创业者更方便地开发产品，比如可以通过社交网络来收集数据的健康类App。也许未来数年后，这些搜集的数据能让医生给你的诊断变得更为精确，比方说不是通用的成人每日三次一次一片，而是检测到你的血液中药剂已经代谢完成会自动提醒你再次服药。



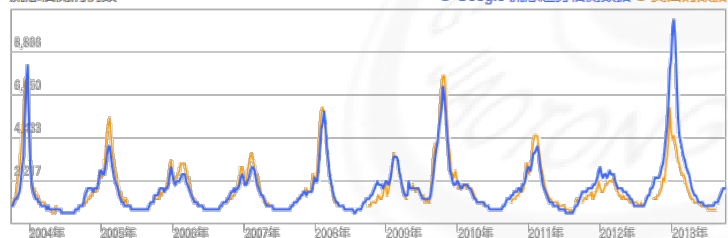
流感趋势预测



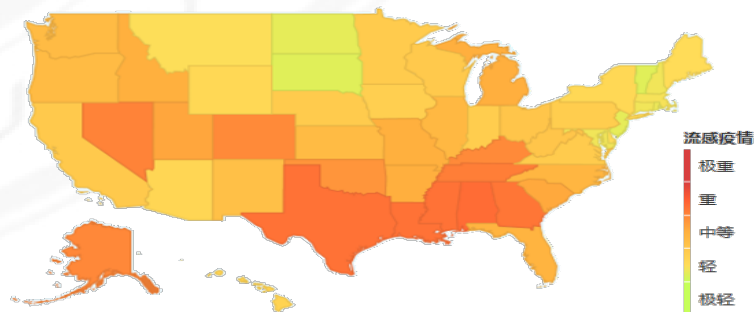
- 2008年，谷歌推出了其著名的**流感趋势**网站（<http://www.google.org/flutrends>）。该网站假定的前提是：如果用户患上了流感，则他们会搜索更多同流感相关的信息。如此一来，如果对任何一个国家或地区有关流感的搜索量进行统计，就能较好推断出某个国家或地区是否正爆发流感。
- 事实上，谷歌的这项统计数据被证实很有效。谷歌的相应数据，同美国疾病控制与预防中心(CDC)等政府机构所统计的数据非常接近（**97%**）。在某些情况下，谷歌甚至能够比CDC提前一周预测出哪些地区将爆发流感。

美国流感疫情

流感估计病例数

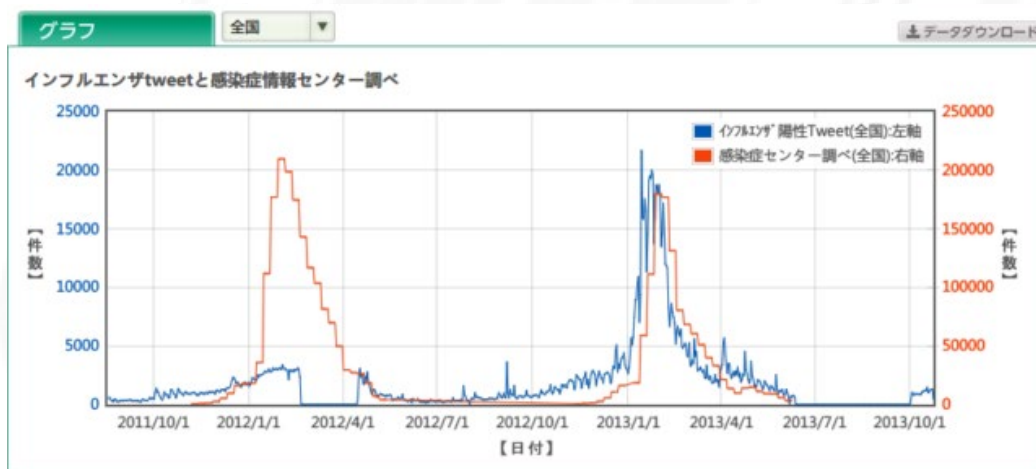
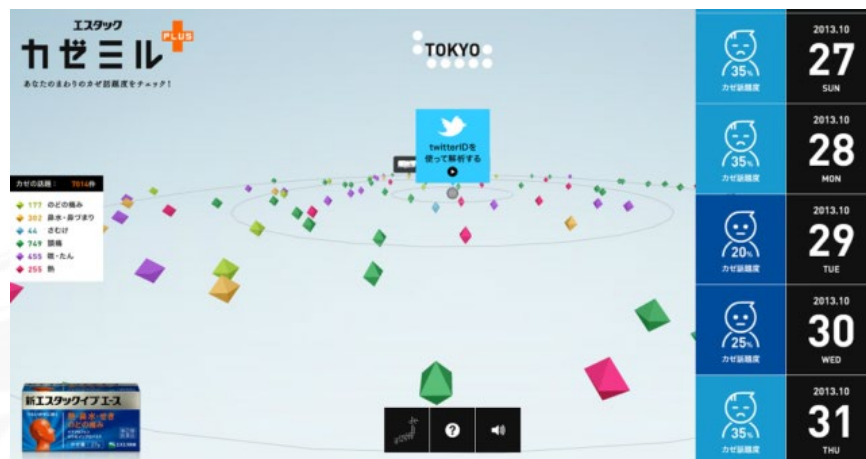
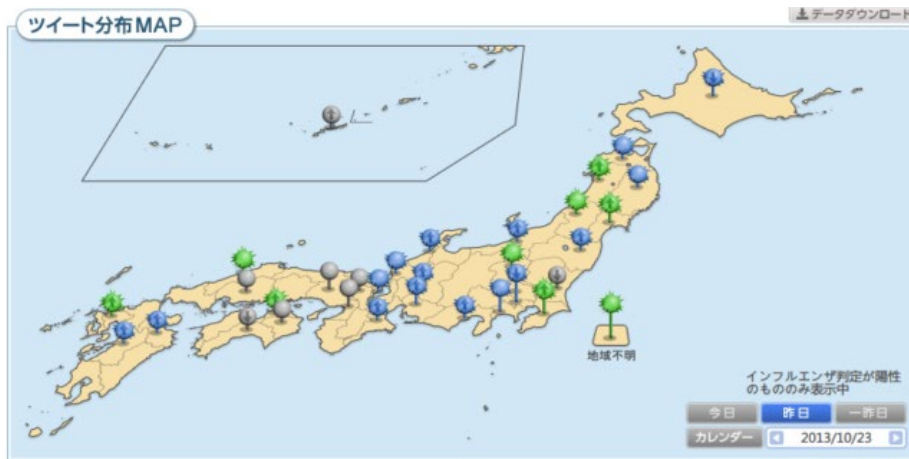


美国：流感样疾病 (ILI) 数据由美国疾病控制中心提供。



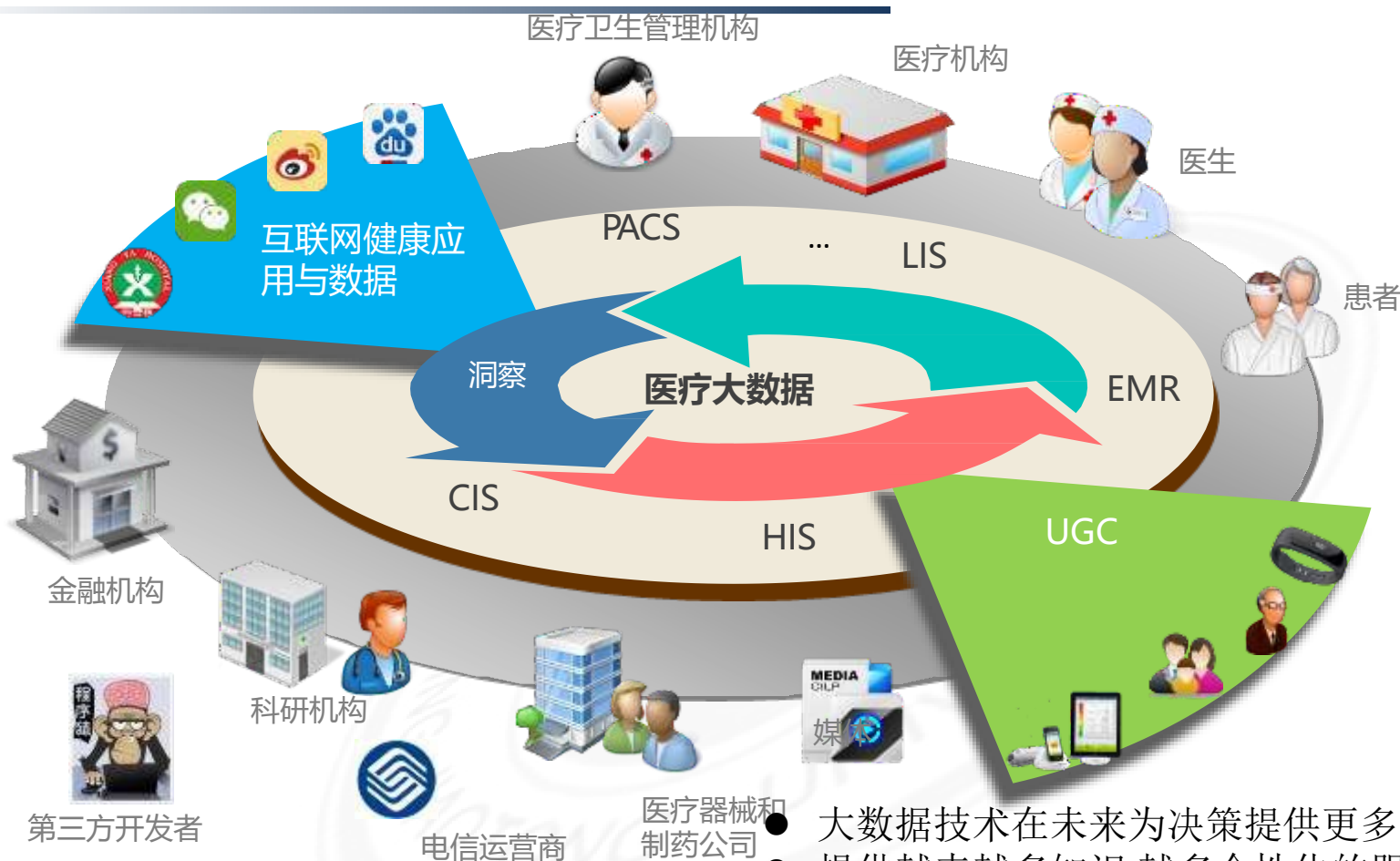


流感趋势预测





医疗大数据的发展趋势



- 大数据技术在未来为决策提供更多的支持
- 提供越来越多知识 越多个性化的服务
- 数据催生新的业务模式和服务模式
- 处理过程及用户传输产生的内容实时化、及时化
- 大数据技术在医疗领域的不断创新