一个商业化AI项目的实施过程

——深圳奇点信息技术有限公司/极智金融云 创始人 李翛然 FRM

所有成功的商业化AI项目都是在进行解放和发展 生产力

公司简介

总公司成立于2014年,业务始于金融科技 (极智金融云)

2017年,"也许"量产全球第一款专业金融机器人"股神SAI"

2017年 获得KPMG 全国金融科技双创大赛 TOP 30

2018年 全国最大的肺癌数据中心结构化数据 服务

2019年 微众银行区块链应用大赛一等奖





我们来严肃和认真的谈一谈"钱"的问题

吴恩达离职 2017-03

"AI创业阵亡" 2017-10

"寒冬将至" 2018-05

"AI存在问题" 2018-07

"资本极寒" 2018-08

李菲菲离职 2018-09

【重磅】吴恩达终于离开,没有人感到意外

2017年03月22日 14:10 xteacher

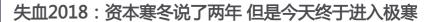






2018年07月05日 13:39 新智元





2018年08月08日 15:22 《商界》杂志









李飞飞确认离职,谁来接管谷歌云AI?

07:30 收藏23 评论4

"在危险环境中表示绝望的人,在黑暗中看不见光明的 人,只是懦夫与机会主义者"

—— 毛泽东文集,第二卷

让我们仔细的分析一次商业化AI的全过程

- 需求挖掘
- 确定工程目标
- 概念验证
- 工程实施
- 交付验收

"我们积累了大概30年的历史数据,您看看能做些什么,要求就 是要有新的技术及具体的产出,不能是概念化的"

锁定用户需求

检查类型	影像所见 影像诊断	B:告日期	检查科室4检查日期
СТ	双侧胸廓对称,纵隔气管居中,各叶段支气管开口通畅,双肺野纹理增强,右戶、右肺上叶周围型肺癌,并远端少许阻塞性肺炎,纵隔、右肺门淋20	011-03-18T10:55:29+08:00	放射线科 2011-03-18T10:41:02+08:00
DR	右胫腓骨骨质不光整,见骨痂影,对位对线可,见金属固定物影,右踝关节诸样符合右胫腓骨陈旧性骨折术后改变。	012-03-14T09:41:54+08:00	放射线科 2012-03-14T09:26:09+08:00
DR	右膝关节髁间隆起变尖,关节面不光滑,髌骨缘欠光滑,关节间隙稍窄。右侧上右膝关节骨质增生,右侧腓骨陈旧性骨折术后改变。 20	012-03-14T09:41:14+08:00	放射线科 2012-03-14T09:25:59+08:00
DR	右侧髋关节间隙正常,关节面骨质光滑、连续,关节骨质骨小梁略稀疏,未见F右髋关节骨质略疏松 20	012-09-14T09:47:44+08:00	放射线科 2012-09-14T09:03:54+08:00
DR	右膝关节髁间隆起变尖,关节面不光滑,髌骨缘欠光滑,内侧关节间隙稍窄。有右膝关节骨质增生。右侧腓骨骨折术后。	013-09-24T09:00:53+08:00	放射线科 2013-09-24T08:49:40+08:00
DR	右踝关节诸骨骨小梁稀疏,腓骨下段骨质不连续、见金属固定物影。胫骨下段!右腓骨术后改变。右胫骨异常密度影,请结合病史。右踝关节骨质1/20	013-09-24T09:02:19+08:00	放射线科 2013-09-24T08:49:51+08:00
	TEN MATERIAL CONTROL DATE DISTRICT STATE S)13-12-17T09:44:59+08:00	放射线科 2013-12-17T08:44:59+08:00
СТ	TALLING WALLING AND)13-12-17T09:58:23+08:00	放射线科 2013-12-17T09:01:32+08:00
СТ	THE MAINTAIN THE PROPERTY OF T		放射线科 2011-03-21T21:35:11+08:00
СТ	双侧胸廓对称、气管纵隔居中,各叶段支气管开口通畅,左肺门影稍增大,局挡、左肺门影增大,以淋巴结肿大可能性大,建议临床进一步检查,20)11-03-16T15:41:46+08:00	放射线科 2011-03-16T15:26:01+08:00
	101/101/101/101/101/101/101/101/101/101	"	放射线科 2014-05-22T09:36:10+08:00
	患者屏气不佳,部分层面有呼吸伪影,影响观察。肝影不大,表面光滑,各叶和、考虑肝内多发小囊肿,必要时进一步检查2、肝表面钙化灶3、右20		放射线科 2011-12-16T07:53:54+08:00
СТ	腰椎生理曲度变直,胸11-12、腰1-5椎体边缘变尖,骨质密度减低。腰1-2、腰2-胸11-12、腰1-5椎骨质增生、骨质疏松腰1-2、腰2-3、腰3-4、腰4-5、20		放射线科 2012-04-02T15:10:21+08:00
СТ	7/6/7/1/2 XIII/1/1/4 X	*	放射线科 2012-10-06T19:40:52+08:00
	胸廓对称,气管纵隔居中。各叶段支气管开口通畅。双肺上叶见条索状高密度11、考虑左肺上叶周围型肺癌2、左肺下叶、右肺上叶结节,建议隔120		放射线科 2010-04-28T13:12:10+08:00
-447 72/F7	左膝关节间隙变窄,关节面欠光滑,各组成骨关节面边缘增生变尖,髌骨髌股;1、左膝关节退行性骨关节病伴髌骨髌股关节、股骨远端及胫骨近端20		放射线科 2012-05-18T07:21:29+08:00
	双侧胸廓对称,纵膈、气管居中,各叶、段支气管开口通畅。左肺舌叶可见条片、考虑左肺舌叶周围型肺癌,对比前片(2013-01-24)病灶范围缩了20		放射线科 2013-04-28T08:36:38+08:00
	腰椎曲度加大,腰3椎体略向后移位,腰3-4椎间隙变窄。腰1-5椎体骨质边缘变5腰3椎体轻度后滑脱腰1-5椎体骨质增生、骨质疏松胸11-12、腰1-2、20		放射线科 2013-08-23T09:32:03+08:00
-447 72/F7	桥脑、双额叶、双侧放射冠区、双侧半卵圆中心见多发斑点状等及稍长T1稍长T1、考虑双侧多发腔隙性脑梗死及点状缺血灶,结合病史建议MRI增了20		放射线科 2014-06-20T11:58:16+08:00
	双侧胸廓对称,纵膈、气管居中。左肺舌叶见团片状高密度影,最大层面大小汽、考虑左肺舌叶周围型肺癌伴远端阻塞性肺炎肺不张可能性大,对20		放射线科 2014-05-14T09:45:27+08:00
	胃呈术后改变,吻合口未见明显增厚,胃腔明显扩张,腔内容物多。肝影不大,1、胃呈术后改变;残胃略扩张、内容物多,请结合临床2、肝脏多岁20		放射线科 2015-08-13T14:50:50+08:00
	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廓大致对称,气管红、符合右肺上叶切除术后改变;右肺下叶支气管改变,请结合临床20		放射线科 2016-09-17T15:35:47+08:00
_	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廓大致对称,气管、1、符合右肺术后改变;右肺中叶纵隔旁缝合线周围病变,与2018-0520		放射线科 2018-07-11T15:28:39+08:00
	右侧胸壁可见瘢痕影,右侧第5肋骨骨质不连续,双侧胸廓大致对称,气管纵隔1、符合右肺上叶切除术后改变,请结合临床病史2、考虑残余右肺、20		放射线科 2015-05-14T12:02:44+08:00
_	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廓大致对称,气管、1、符合右肺上叶切除及右肺中叶术后改变;右肺中叶纵隔旁缝合线20		放射线科 2018-05-12T15:23:40+08:00
	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廓大致对称,气管、1、符合右肺上叶切除及右肺中叶术后改变;右肺中叶纵隔旁缝合线20		放射线科 2018-03-05T08:31:15+08:00
	胃呈毕试术后改变,吻合口未见明显增厚,未见明确异常强化影。肝影不大,1、胃呈术后改变,注意复查2、肝脏多发囊肿3、胆囊底部改变,考20		放射线科 2018-03-06T14:30:47+08:00
	前列腺体积增大,左右径5.8cm,前后径4.3cm,密度欠均,内见多发结节样钙/1、前列腺增生伴多发钙化灶;前列腺密度欠均匀,必要时请进一步20	-	放射线科 2018-03-01708:41:07+08:00
	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廓大致对称,气管、1、符合右肺上叶切除及右肺中叶术后改变;右肺中叶纵隔旁缝合线20		放射线科 2018-01-14T14:45:46+08:00
CT	右侧胸壁可见瘢痕影,右侧第4、5肋骨骨质不连续,双侧胸廊大致对称,气管、1、符合右肺上叶切除及右肺中叶术后改变;右肺中叶纵隔旁缝合线20	017-11-10T16:34:18+08:00	放射线科 2017-11-10T14:50:58+08:00

<component>

<section>

大小为3.2*2.0,伴部分肺不张,故为T2;因未见淋巴结转移,故为N0;因胸水中找到腺癌细胞,左侧胸膜转移,故为M1a,故临床诊断为左肺腺癌IV期(cT2M0M1a)。3、治疗:EGFR检测提示21外显子L858R突变,给予口 服吉非替尼分子靶向治疗。 明确临床诊断: 左肺腺癌IV期伴EGFR 21-L858R突变 左侧恶性胸腔积液 左侧胸膜转移 诊断依据:患者气短、胸闷1月余 辅助检查:肺CT:1、左肺各叶、右肺中叶、下叶炎症;2、左肺下叶钙化灶;3、左侧胸腔积液及叶间积液,部分包裹,继发邻近左肺舌叶及下叶部分被动不张;PET-CT:1、左肺下叶 降主动脉旁高代谢灶,考虑周围型肺癌;左侧胸膜转移癌;左侧胸腔积液,相邻左肺各叶部分肺不张。2、左肺炎症及炎性索条。3、脾脏钙化灶;右肾结石;直肠末端代谢增高,考虑炎性。余全身PET-CT显像未见明显异常 治疗原则: 分子靶向药物治疗。]]></content> <html><![CDATA[杨雷副教授代主治医师查房: 1、吉大三院及我院脱落细胞均找到腺癌细胞,计2次。2、分期: 因左肺下叶主动脉旁病灶大小为3.2*2.0,伴部分肺 不张,故为T2;因未见淋巴结转移,故为N0;因胸水中找到腺癌细胞,左侧胸膜转移,故为M1a,故临床诊断为左肺腺癌IV期(cT2M0M1a)。3、治疗:EGFR检测提示21外显子L858R突变,给予口服吉非替尼分子靶向治疗 ,
 明确临床诊断:
 在肺腺癌IV期伴EGFR 21-L858R突变

 左侧胸膜转移
 诊断依据:患者气短、胸闷1月余 辅助检查:肺CT:1、左肺各叶、右肺中叶、下叶炎症;2、左肺下叶钙化灶;3、左侧胸腔积液及叶间积液,部分包裹,继发邻近左肺舌叶及下叶部分被动不张;PET-CT:1、左肺下叶降主动 脉旁高代谢灶,考虑周围型肺癌;左侧胸膜转移癌;左侧胸腔积液,相邻左肺各叶部分肺不张。2、左肺炎症及炎性索条。3、脾脏钙化灶;右肾结石;直肠末端代谢增高,考虑炎性。余全身PET-CT显像未见明显异常。
 治疗原则:
 分子靶向药物治疗。]]></html> <entry> </entry> </section> </component> </entry> </section> </component> <component> <section> <title></title> <code code="Control 13" codeSystemName="" tab="13" pagebreak=""/> <value><![CDATA[2016年01月26日15时42分{COL} 教授查房{COL}记录医生:{COL}0{COL}JSCF{COL}11029087{COL}{COL}{COL}JCOL}D0213{COL}0{COL}3{COL}1{COL}1{COL}16</p> \D0213.bmp|\Sign\Q0001.bmp]]></value> <text><![CDATA[2016年01月26日15时42分{COL} 教授查房{COL}记录医生:{COL}0{COL}JSCF{COL}{COL}11029087{COL}{COL}COL}COL}COL}COL}COL}COL}OCL}15COL}0{COL}3{ \D0213.bmp|\Sign\Q0001.bmp]]></text> <entry>

目标确认

首要目标

整理全中国(也许是全球)最大的肺癌结构化知识数据库

目标效果

- 构建可能的知识图谱
- 可视化展示
- 建模及实验教学辅助

概念验证

胸廓对称,气管纵隔居中。各叶段支气管开口通畅。双肺上叶见条索状高密度影,左肺上叶可见结节状高密度影,最大层面约为2.1cm×2.1cm,CT值为22HU,周围可见浅分叶及毛刺,邻近胸膜凹陷。左肺下叶、右肺上叶可见结节样高密度影,直径约0.3-0.4cm,边界尚清。右肺中叶及左肺下叶可见小片状高密度影;左肺舌叶见钙化影。纵隔内气管前腔静脉后间隙及主肺动脉窗间隙可见略肿大淋巴结影,直径0.8-1.1cm。心影不大,冠状动脉及胸主动脉可见钙化影。

概念验证

胸廓 各叶段 双肺 叶见条 索状 高密度 左肺 叶可见状高 左肺下右肺 样高 尚清 舌叶 前腔 主肺心影

工程标准

- 可服务化部署,本地网络离线使用
- 总体ROC97%以上, AOC95%以上
- 故障升级方便,最好是windows 环境

具体案例,公众版删除

项目难点

- 目标不确定
- 没有数据集,也不可能有数据集
- 由于HIS系统经过多年使用,缺少维护数据迁移过程中有很多的数据丢失及格式错乱
- 目标结果为探索发现型,因此必须做到快速迭代,任何工程人员可快速上手

TDD交付过程

确定功能/效果点 → 快速开发 → 测试反馈 → 迭代升级 → 交付

长期沟通机制

渐进式部署



例:气管镜分析

胸廓对称,气管纵隔居中。各叶段支气管开口通畅。双肺上叶见条索状高密度影,左肺上叶可见结节状高密度影,最大层面约为2.1cm×2.1cm,CT值为22HU,周围可见浅分叶及毛刺,邻近胸膜凹陷。左肺下叶、右肺上叶可见结节样高密度影,直径约0.3-0.4cm,边界尚清。右肺中叶及左肺下叶可见小片状高密度影;左肺舌叶见钙化影。纵隔内气管前腔静脉后间隙及主肺动脉窗间隙可见略肿大淋巴结影,直径0.8-1.1cm。心影不大,冠状动脉及胸主动脉可见钙化影。

结构化信息提取

关键字提取?

TF-IDF, TextRank, CRF, BERT, Transformer? ?

分词

jieba

'胸廓/对称/,/气管/纵隔/居中/。/各叶段/支气管/开口/通畅/。/双肺/上/叶见/条/索状/高密度/影/,/左肺/上叶/可见/结节/状/高密度/影/,/最大/层面/约/为/2.1/cm/×/2.1/cm/,/CT/值为/22HU/,/周围/可见/浅分/叶及/毛刺/,/邻近/胸膜/凹陷/。/左肺/下叶/、/右肺/上/叶/可见/结节/样/高密度/影/,/直径约/0.3/-/0.4/cm/,/边界/尚清/。/右肺/中叶/及/左/肺/下叶/可见/小/片状/高密度/影/;/左肺/舌叶/见/钙化/影/。/纵隔/内/气管/前腔/静脉/后/间隙/及主/肺动脉/窗/间隙/可见/略/肿大/淋巴结/影/,/直径/0.8/-/1.1/cm/。/心影/不大/,/冠状动脉/及/胸/主动脉/可见/钙化/影/。'

thulac

'胸廓/对称/,/气管/纵隔/居中/。/各/叶段/支气管/开口/通畅/。/双/肺/上/叶见条/索状/高密/度影/,/左/肺/上叶可见/结/节状/高/密度/影/,/最/大/层面/约/为/2/./1/cm×/2/./1/cm/,/CT值/为/22HU/,/周围/可见/浅/分叶/及/毛刺/,/邻近/胸膜/凹陷/。/左/肺/下叶/、/右肺/上叶可见/结/节样/高/密度影/,/直径/约0/./3/-0/./4/cm/,/边界/尚/清/。/右肺/中叶/及/左/肺/下叶可见/小/片状/高密/度影/;/左肺/舌叶/见/钙/化影/。/纵隔/内/气管/前腔/静脉/后/间隙/及/主肺/动脉/窗间/隙可见/略/肿大/淋巴结影/,/直径/0/./8/-1/./1/cm/。/心影/不/大/,/冠状动脉/及/胸/主动脉/可见/钙/化影/。'

pkuseg

胸廓/对称/,/气管/纵隔/居中/。/各/叶段/支气管/开口/通畅/。/双/肺/上叶/见条/索状/高/密度/影/,/左/肺/上叶/可见/结节/状高/密度/影/,/最/大/层面/约/为/2.1cm/×/2.1cm,CT值/为/22/HU/,/周围/可见/浅/分叶/及/毛刺/,/邻近/胸膜/凹陷/。/左/肺/下叶/、/右/肺/上叶/可见/结/节样/高/密度/影/,W/直径/约/0.3-0.4cm/,/边界/尚清/。/右/肺/中叶/及/左/肺/下叶/可见/小/片状/高/密度/影/;/左/肺/舌叶/见/钙/化影/。/纵隔/内/气管/前腔/静脉/后/间隙/及/主肺/动脉/窗间/隙/可见/略/肿大/淋巴/结影/,/直径/0.8-1.1cm/。/心影/不大/,/冠状动脉/及/胸主/动脉/可见/钙/化影/。

于是,我们搜索和整理整个互联网,整理了15W+的专业词库然后……

具体效果 公众版删除

数字和数量

胸廓对称,气管纵隔居中。各叶段支气管开口通畅。双肺上叶见条索状高密度影,左肺上叶可见结节状高密度影,最大层面约为2.1cm×2.1cm,CT值为22HU,周围可见浅分叶及毛刺,邻近胸膜凹陷。左肺下叶、右肺上叶可见结节样高密度影,直径约0.3-0.4cm,边界尚清。右肺中叶及左肺下叶可见小片状高密度影;左肺舌叶见钙化影。纵隔内气管前腔静脉后间隙及主肺动脉窗间隙可见略肿大淋巴结影,直径0.8-1.1cm。心影不大,冠状动脉及胸主动脉可见钙化影。

于是,我将金融当中使用的数字提取迁移了过来,并进行了 升级……

于是~

https://github.com/Wall-ee/chinese2digits

可能是最好用的中文数字提取转换工具 万二,负百分之点三五 缺失信息处理

缺少dtd说明文件……几乎无法解析xml

单位转换,三元语义分析与提取等......

具体效果 公众版删除

架构图公众版删除

交付验收

首要目标

整理全中国(也许是全球)最大的肺癌结构化知识数据库—完成

目标效果

- AOC,ROC均在99%以上
- 肺癌知识图谱基本构建完成,可根据病程记录和诊断进行大致的病症推断
- 可视化展示,可进行病理图可视化显示
- 建模和辅助教学可自动化提取数据
- 新人可在2-3天的培训内对系统进行维护和升级

为什么传统公司/高科技AI公司商业化都遇到了重大的问题?

TDD交付过程

传统 扯皮开发模式

 AI工程让客户看到可预期,循序渐进,带来实在的 生产力提升是给传统公司和团队带来最大的挑战 "持续集成是生产可用AI产品的最重要的能力"

持续集成—TDD的核心

- 架构师(项目经理全权负责制)
- 团队内所有成员必须按模块负责制
- 运维开发一体的责任制度
- 成员相互Code Review 要求必须相互熟悉且向客户负责
- 相应的与TDD相关的绩效和奖惩制度







重试设计

运维计划

范式编程

数据库优化

幂等性设计

熔断机制

通讯优化

重试设置

Al

熔断设计

消息机制

边缘计算

安全加密

路由设计

部署升级

缓存机制

网关模式

AI最难得部分在于项目的工程实现, 团队对于各种工具的性能,功能边界理解

AI春天的到来?

AI创造/优化AI

Q&A

极智金融云 李翛然

WB:机器人老师傅 xrli_office@foxmail.com