

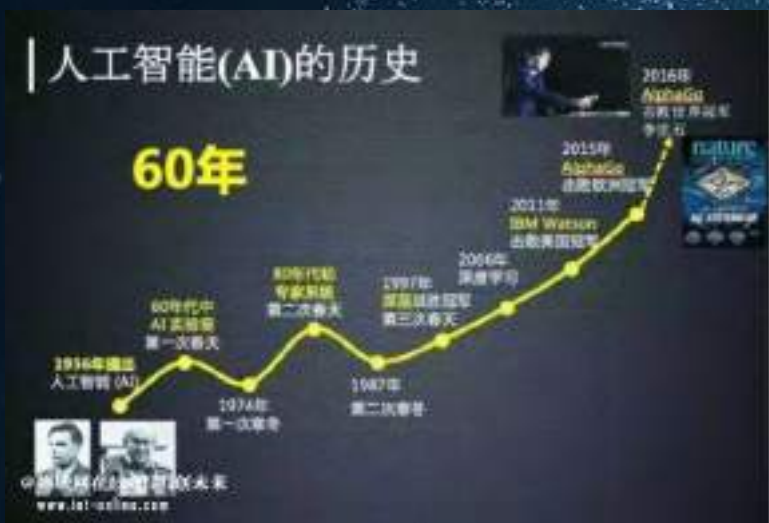


# 人工智能技术与实践

刘大双

北京神州泰岳软件股份有限公司

# 以深度学习为代表的机器学习成为了AI算法的基础



大数据  
带来的样本易获取



GPU  
带来的算力大提升



# 从感知智能的角度 成就十分巨大

看得到



听得到  
说得清

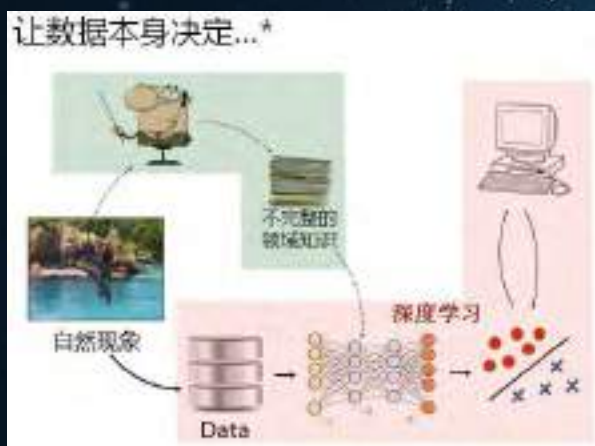


众创时代  
让每一个梦想实现腾飞

机器学习的算法是普惠的  
GPU带来的算力是普惠的  
学习样本已经更加易于获得

# 从认知智能的角度来看深度学习效果不如预期 路还漫长

## DEEP LEARNING



大量的样本数据

数据的“善良性”



互联网上最多最容易获得的样本数据

或者样本数据易于获得且易于积累的领域



没有较高的答案专业性要求

# 尤其是在行业领域深度学习还将面临三个挑战



## 缺乏足够丰富的样本数据

---

算法训练、测试、调优困难



## 中文歧义性的挑战

---

相同的含义不一样的表达  
相同的表达不一样的含义



## 应答结果的专业性要求

---

应答准确性不可让步

# 泰岳**认知+技术** 自成一派的智慧语义认知技术



## Market Guide for Conversational Artificial Intelligence in China Gartner 2017.6

2017年6月全球范围内公认的最具权威的IT研究与顾问咨询公司——Gartner公布了“Market Guide for Conversational Artificial Intelligence in China”，**神州泰岳作为国内AI技术的代表厂商荣誉上榜**，同时入选的包括百度、微软、腾讯等多家知名企业。

报告强调：目前在中国，语音识别技术已趋向成熟，但语义理解技术还处于萌芽阶段，会话的主要挑战在于中国的文化及语言的歧义性。**神州泰岳基于智慧语义认知技术，在自然语言处理领域自成一派**，获得了业界广泛认可，由此得到了Gartner的关注和肯定。



Ultrapower  
神州泰岳

### 具有中国气派的语义分析流派

神州泰岳 NLP: 基于概念计算的智慧语义认知技术, 突破自然语言处理技术瓶颈, 解决中文歧义性。

在技术创新之时亦创理论之精, 形成具有中国气派的信息处理流派, 为国增光。  
——许蔚蔚  
著名语言学家  
第九届全国人大常委会副秘书长  
神州泰岳 & 北京师范大学人工智能联合实验室学术委员会主任

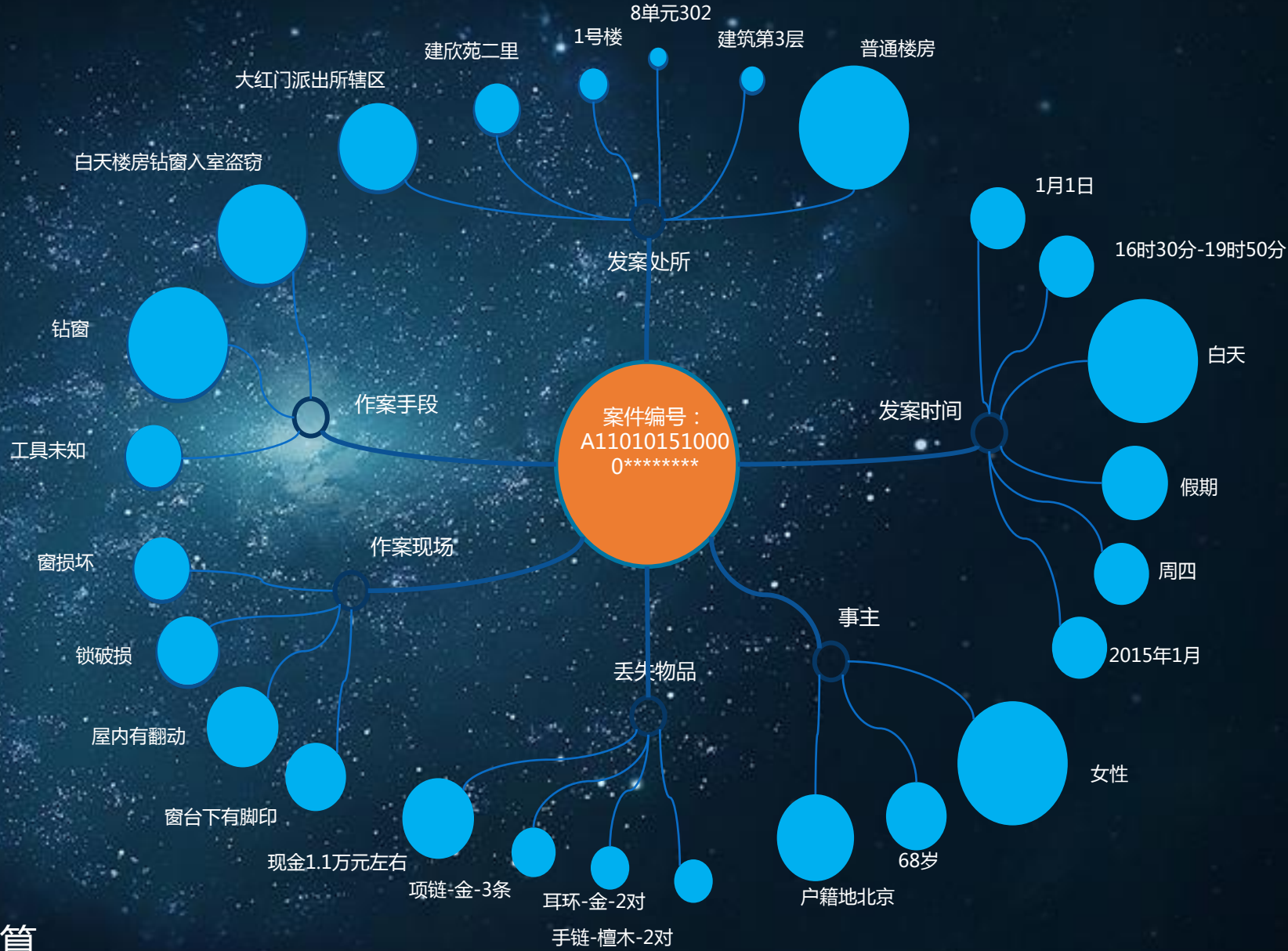
神州泰岳AI研究院荣誉院长  
——李生  
ACL 终身成就奖得主  
哈尔滨工业大学党委书记  
自然语言处理 (NLP) 领域专家



# 智脑公安案情分析系统

序号	案发时间	案发地点	简要案情
1	2016-11-23 08时—2016-11-23 20时	北京	2016年11月23日08时50分,我所... (110报警称:2016年11月23日08时,朝阳区北苑街道... 独居1/7号楼1202的出租房内(100元人民币)被盗,接报后,我所特巡队立即开展工作,经工作了解,2016年11月23日8时50分,出租房主... 报警称:100元人民币,无通商工... 为承租,同住在此租房居住的一男一女... 被盗。
2	2016-09-07 19时	天津静海区33-2-802	2016年9月7日19时12分许,我所接报110报警称:在天津静海区33-2-802室被盗,门锁被撬,我所特巡队立即开展工作,经工作了解:2016年9月7日19时10分,月9日下午3时左右发现出租房出租人... 财产损失:1.苹果手机一台:MacBook Air 11.6寸,内存4G,硬盘128G,Core i5处理器,2016年4月以人民币6388元购买。2. 苹果4s手机:白色,内存8G,自行,联通版,2014年11月以1800元购买,手机型号:15623746162,无发票。3. 华为手机:型号不详,黑色,内存4G,2012年9月以1000多元购买,手机型号不详,里面没有手机卡,无发票。4. 小米NOTE:白色,内存32G,中兴牌,2015年12月以1799元购买。5. 联想牌:品牌型号不详,2015年以400元购买。6. 半镯:千禧,品牌不详,2014年以1800元购买。7. 手表:品牌不详,白色,2015年以300元购买。8. 魅族MX3:白色,内存16G,2014年5月以1999元购买。与同住的郑女士发现被盗:1. 一台笔记本电脑联想牌,黑色,型号为680,硬盘,内存不详,于2013年4月14日以3200元人民币,无发票。2. 一台平板电脑,苹果牌,银色,型号为m12,内存不详,于2014年购买的,价值2000元人民币,无发票;3. 一个P5V(手机),联想牌,黑色,型号为S1(第一代),于2011年在以250元从旧书摊购得,无发票。4. 一手机,小米牌,白色,型号为小米3S,内存16G,手机型号不详,于2016年6月18日以人民币999元购得,无发票。
3	2016-04-25 19时	昌平区天通苑三区18号楼7单元702室	2016年4月25日19时30分我所接报报警称:在昌平区天通苑三区18-7-702室中被盗,接报后我所特巡队立即开展工作,经工作了解,昌平区天通苑三区18号楼7单元702室为出租房,原定于2016年4月25日19时由昌平区天通苑三区18号楼7单元702室对发现租房户的... 财产损失:1. 18寸黑色华硕华硕笔记本电脑一台。(18寸黑色华硕华硕笔记本电脑,10代处理器,英特尔酷睿i5,2014年以3000多元的价格购买,无发票);2. 租户手机发现被盗笔记本电脑一台。(14寸银色华硕华硕笔记本电脑,2011年以3000多元的价格在沃尔玛购买,无发票);3. 租户手机发现被盗笔记本电脑一台,手机一部,黑曜石项链一条。(18寸黑色华硕华硕笔记本电脑,2015年以4800元的价格购买,无发票;手机为小米2,2014年7月,以1400多元的价格购得,无手机卡,黑曜石项链,2015年以2200元的价格购买);4. 租户手机发现被盗笔记本电脑一部APD(华硕笔记本电脑,13.3英寸,白色,内存4G,硬盘是128G的,2015年10月份以2000元购买的);APD是9英寸的,银色,重量16G,2014年10月份以320元购买的。);5. 租户手机发现被盗笔记本电脑一部,联想笔记本电脑一部和一部华为智能手机。(华硕笔记本电脑,15英寸,黑色,内存2G,硬盘是500G的,2012年以4800元购买的);联想笔记本电脑,15英寸,黑色,内存4G,硬盘500G,2012年以4800元购买的。)

作案事件中  
“简要案情”、“回访记录”、  
“现勘记录”、“讯问笔录”等  
电子文本信息进行特征提取和关联计算



# 以入室盗窃案情记录为例

1. 文本篇幅短小，属于短文本类型。由于单词频率低，因此统计方法和深度学习是不容易奏效的
2. 包含丰富案件属性信息(案发时间、案发地点、相关人物、作案手段、丢失物品等)，且信息表达具有极大灵活多样性。

A1102216600002015020006

2015年2月18日8时许，刘某从昌平区南邵镇\*\*\*村临117号暂住地外出，于下午13时许，回到暂住地时发现放在电视机上苹果迷你平板电脑被盗。

A1101015300002015010003

2015年1月2日19时左右，事主吴某某（男，38岁，身份证号码：61232119751218\*\*\*\*，户籍所在地：陕西省南郑县喜神坝乡新建村\*\*\*街09号）从居住地（北京市东城区清华街3号）家中离开外出，走时用钥匙将屋门暗锁锁好。2015年1月2日20时30分左右，事主回来后发现屋门暗锁被撬坏，屋内物品被翻动，后到派出所报案。经清点，被盗铂金pt950结婚钻戒一对（于2010年2月14日购买，价值分别为人民币1796元、人民币3199元，有发票）；黄金项链一条、铂金项链两条（特征不详，总价值约为人民币10000余元）；玉质手镯两个（特征价值不详）；瑞士摩凡陀手表一块（是20世纪70年代我岳父以人民币700余元购买，无发票）；黄金耳环一对、铂金耳环两对（特征不详，总价值约为人民币4000余元）等物品。



# 偷盗的物品是“电视机”还是“电脑”呢？

2015年2月18日8时许，刘XX从昌平区南邵镇北XXX村临117号暂住地外出，于下午13时许，回到暂住地时发现放在电视机上苹果迷你平板电脑被盗。



“苹果”是水果还是平板电脑呢？

# 不同种类时间标签提取同样十分困难

2015年1月2日19时左右，事主吴某（男，38岁，身份证号码：61232119751218\*\*\*\*，户籍所在地：陕西省南郑县喜神坝乡新建村\*\*\*街09号）从居住地（北京市东城区\*\*\*\*街3号）家中离开外出，走时用钥匙将屋门暗锁锁好。2015年1月2日20时30分左右，事主回来后发现屋门暗锁被撬坏，屋内物品被翻动，后到派出所报案。经清点，被盗铂金pt950结婚钻戒一对（于2010年2月14日购买，价值分别为人民币1796元、人民币3199元，有发票）；黄金项链一条、铂金项链两条（特征不详，总价值约为人民币10000余元）；玉质手镯两个（特征价值不详）；瑞士摩凡陀手表一块（是20世纪70年代我岳父以人民币700余元购买，无发票）；黄金耳环一对、铂金耳环两对（特征不详，总价值约为人民币4000余元）等物品。

! 案发时间 !	报警时间：
	案发时间：2015-01-02 19时——2015-01-02 20时
	案发时段：白天

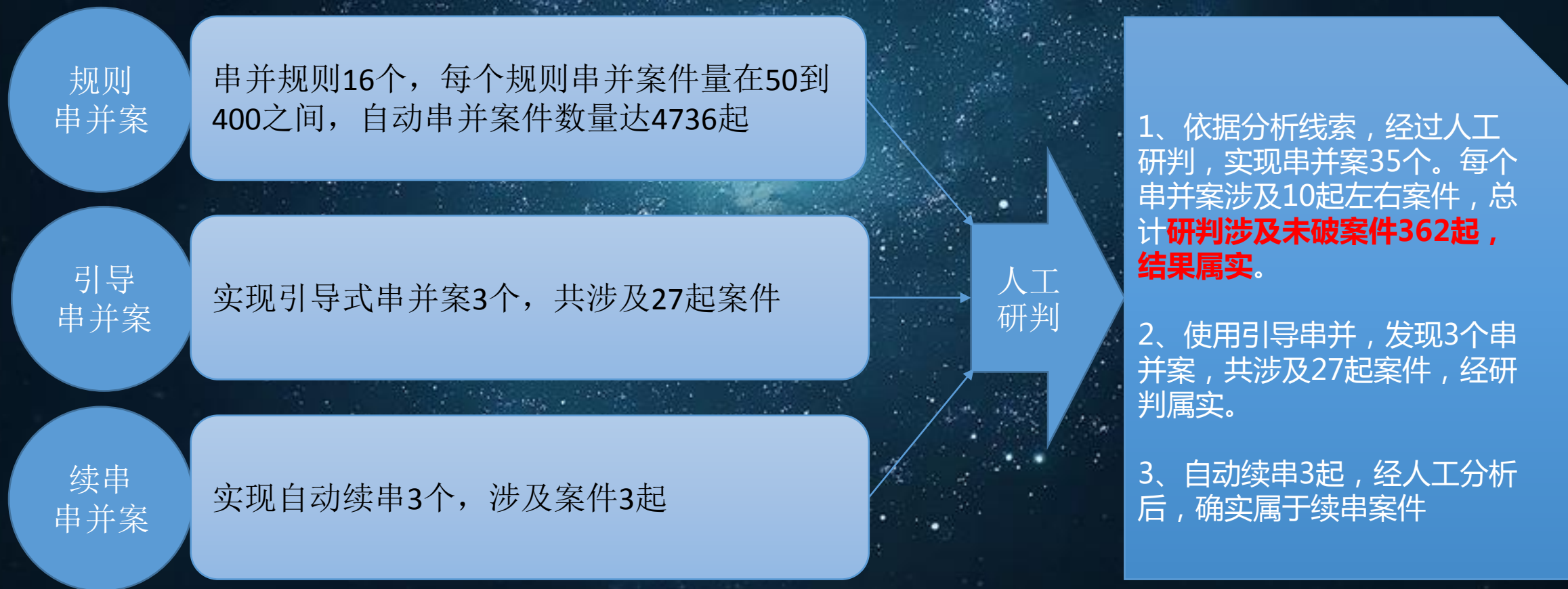
1. “2015-1-2 19:00—2015-1-2-20:30”
2. 去噪：09号的“号”非时间
3. 消歧：“2010年2月14日”非案发时间
4. 格式标准化：“20世纪70年代”标准格式应为“1970年”

# 入室盗案件的侦破实战成效

## 当前针对入室盗案情分析

涉及分析案件：2016年1-12月期间的案件，总数量：22710起案件。

涉及续串案件：经分析2017年1月1日-16日期间的案件，总数量：900起案件。



注：系统自动续串有利的**补充了人工分析时续串的难题，有效提高了案件分析侦办及破案效率。**

# 神州泰岳认知+：基于概念计算的智慧语义认知技术

节点表示式	中文命名	英文命名	高层
jw60a	原生物	protoplast	
jw60t			
jw61	植物	Plant	t=b, \k=6,
jw61\3	草与花	grass and flowers	
jw61\4	禾本与爬蔓	gramineous and trailing	
jw61\5	灌木	frutex	
jw61\6	树	tree	
jw61\k		type of plant	
jw617	植物生存环境	environment of plant	
jw619	食用植物	edible plant	t=b
jw6199	主食	staple food	
jw619a	蔬菜	vegetable	
jw619b	果实	fruit	
jw619b*1	苹果		
jw619b*2	香蕉		

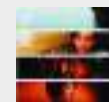
“苹果”的符号化语义表示

- ✓ jw619b\*1 (水果)
- ✓ pea26 (苹果公司)
- ✓ gwa35\*5 (苹果电影)



Pea26 表示苹果公司, pe现代社会机构->pea2经济组织-> pea26 公司

Gwa35\*5 表示苹果电影, gw信息物->gwa3文化->gwa35信息产物->gwa35\*5 电影

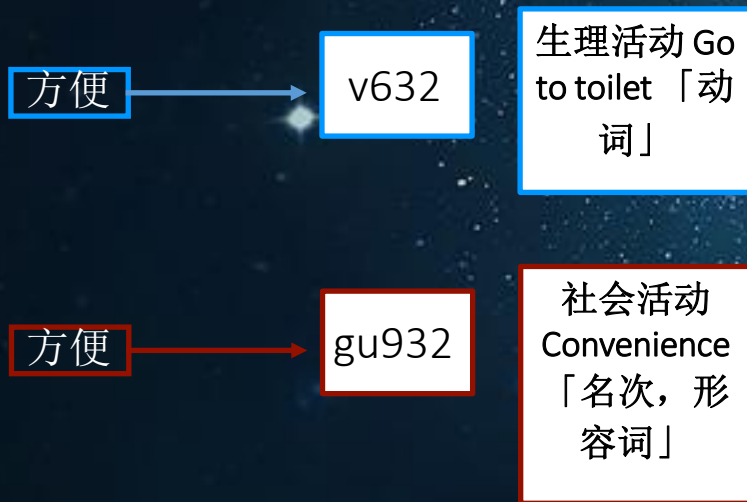


符号化语义表示的DNA语义词库

语义=语言本体+领域范畴+意义表示

# 神州泰岳认知+：基于概念计算的智慧语义认知技术

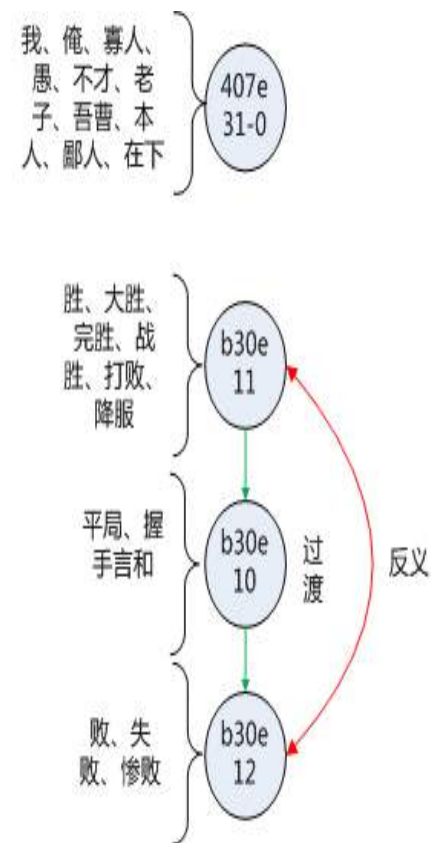
一老外学了些汉语，被公司派到中国任职欢迎宴上有人说要去方便一下，他问方便是什么意思，那人说方便就是上厕所，一会儿有人来敬酒，说下次去美国请多提供方便，老外很纳闷。当地一电台女记者说等你方便时做个专访，老外连忙说不行，女记者说等你方便时我请你吃饭，老外晕倒



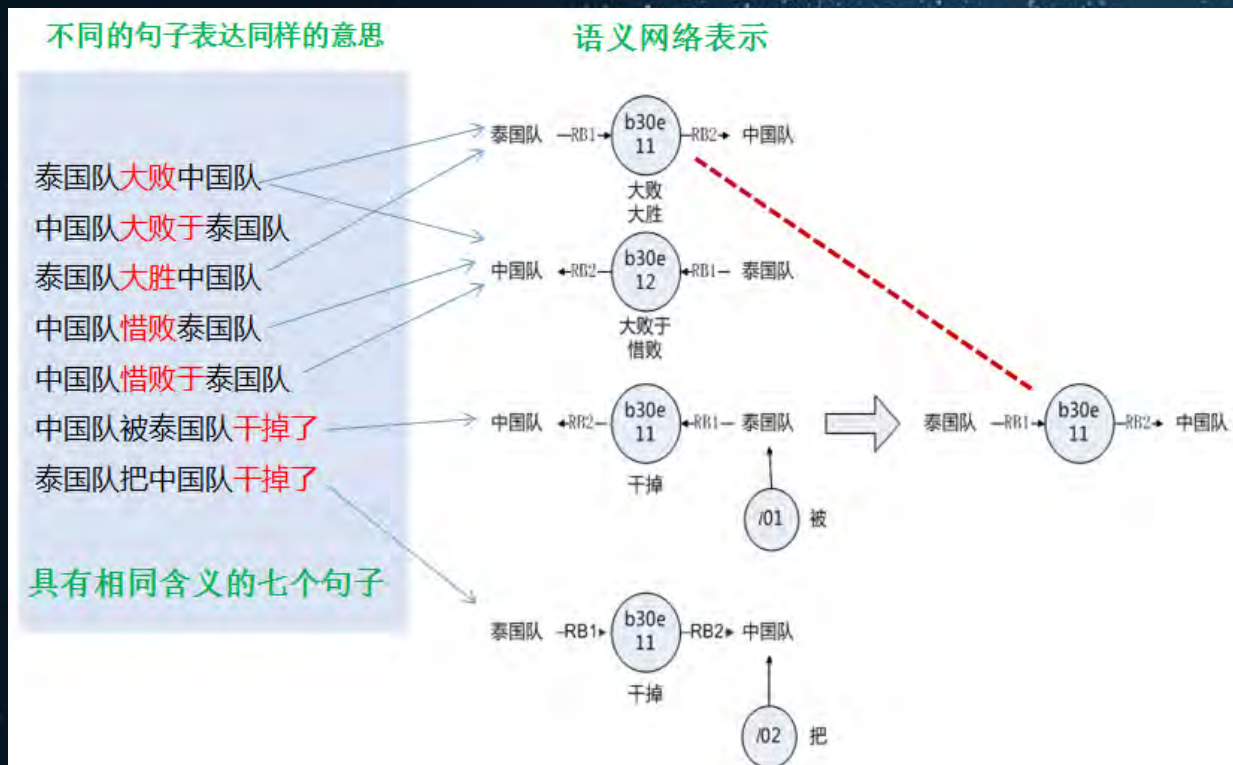
**多词一义：**多个词语表达同一概念

概念的同义、反义、上下位体现在符号上，**可进行计算**

价值：词语的语义感知  
**通过有限的概念，处理无限的词汇**

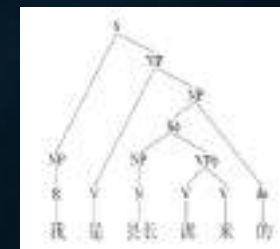
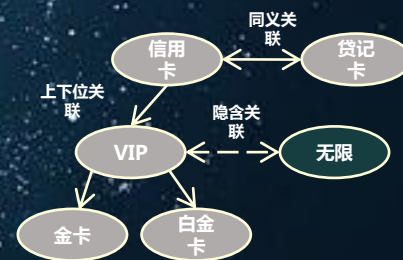


# 神州泰岳认知+：基于概念计算的智慧语义认知技术



智慧语义认知技术的  
句子语义感知

通过有限的语义网络，处理无限的语句表示



传统句法分析



# 顺丰速递客服工单智能分拣



客户声音智能分析挖掘系统CVMS  
(客户服务部门)

## CVMS系统带来的变化和价值

**50秒 进化到 毫秒**  
实现自动归类分析, 提高效率并节省人力成本

**60+分类 进化到 500+ 分类**  
提高归类标准, 利于改善流程, 提升客户满意度

**单语种 进化到 多语种**  
降低客服能力要求, 减少培训复杂度

**进化到 绩效考评和大数据分析挖掘**



# 顺丰速递客服工单智能分拣

## 工单多样性表达

例1、第三方自称王小姐来的投诉，此件实际重量1公斤，但是单子上却写着2公斤，收取其17元钱，客希望给予合理解释，请跟进，客服8515

例2、第三方自称收方邵S\*\*\*\*\*来电查件，称此件是其的妹妹签收的，业务员收去17元运费，可是面单上填写是12元，称我司业务员多收费，烦请跟进，客服8049

The screenshot displays a software interface for analyzing customer complaints. It features a sidebar with navigation options like '数据测试', '发起工单测试', and '结束工单测试'. The main area shows a list of complaints and a detailed view of one. The detailed view includes a '分析结果' (Analysis Results) section with a table summarizing the complaint and its analysis.

处理内容	分析结果
第三方自称收方邵S*****来电查件，称此件是其的妹妹签收的，业务员收去17元运费，可是面单上填写是12元，称我司业务员多收费，烦请跟进，客服8049	103000000:派件环节->103070000:费用收取->103070200:收错费用 103000000:派件环节->103070000:费用收取->103070300:恶意额外收费

投诉，此件实际重量1公斤，但是单子上却写着2公斤，收取其17元，客希望给予合理解释，请跟进，客服8515

投诉，收取17元运费，可是面单填写是12元 => 派件环节/费用收取/收错费用

投诉，.....，称我司业务人员多收费 => 派件环节/费用收取/恶意额外收费

# 某国有银行客服大数据分析挖掘系统



# 更多应用场景

针对重点行业，提供公有云/私有云/混合云的SAAS服务

## 企业风险监测应用



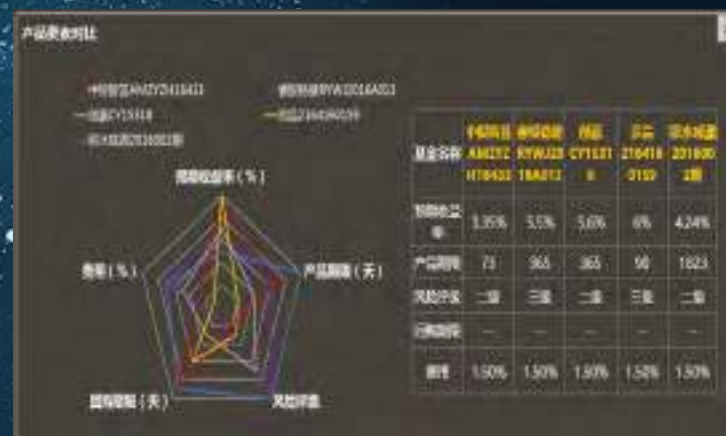
- 提供分类分级风险预警体系，风险及时预警；
- 企业上下游、分子公司、高管、相关企业关系；
- 挖掘企业风险，掌握企业信誉动态；
- 掌握市场真实声音，优化风控体系；
- 掌握市场风险动态，优化投资和信贷策略；
- 共享权威成果，提升自身风控能力；

## 内参产品应用



- 用户自定义数据源分析，灵活、实时掌握舆情动态；
- 根据关联结果，发现潜在的、有价值的商业情报；
- 及时提供相关预警，并对其发展态势进行持续跟踪；
- 提供新型市场调研的解决方案，助力商业决策；
- 快速实现符合自身需求的定制化分析报告；
- 帮助客户快速了解事件发展态势、展开应对；
- 系统的报告和数据列表自由推送给指定人员；

## 竞品分析应用



- 网点客户经理进行同类竞品比对；
- 可依据产品分析师推荐理由营销；
- 指定一至多个竞品同屏对比展示；
- 分析师可指定该推荐产品的对比竞品；
- 竞品对比显示推荐产品优势；
- 竞品对比显示同类竞品劣势；
- 竞品对比结合维度雷达图示；



# 小富机器人 —中国富二代专业智能机器人



# 传统机器人面临的四大痛点

## 一问一答式对话

无法支持多轮会话  
问答死板  
用户体验较差

## 交互沟通友好性需要提高

问答准确率存在较大提升空间



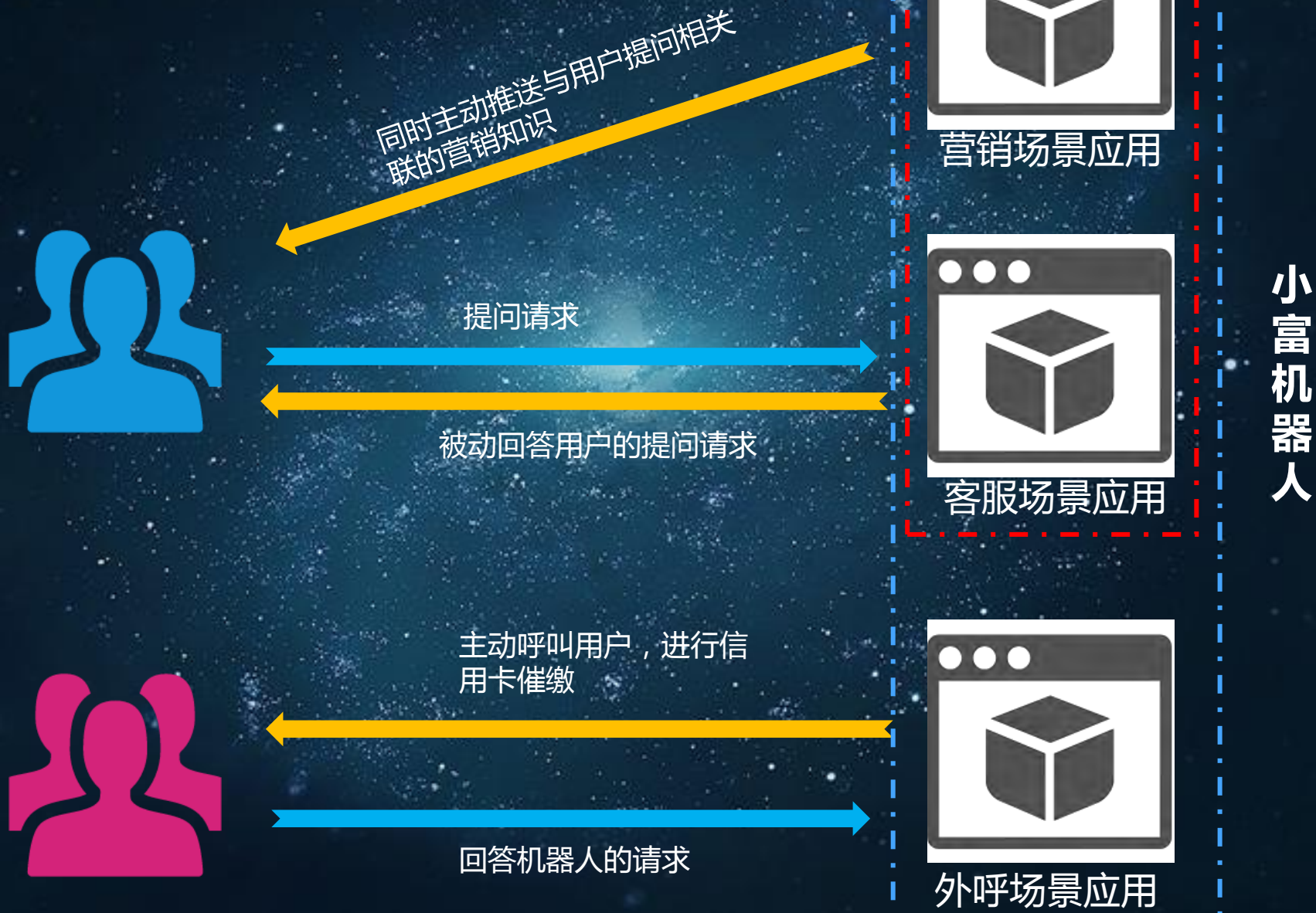
## 智能服务机器人应用场景的多样性

客服、外呼、营销、个人助理等  
场景差异性较大  
无法快捷便利支持机器人研发

## 知识加工自动化程度极低

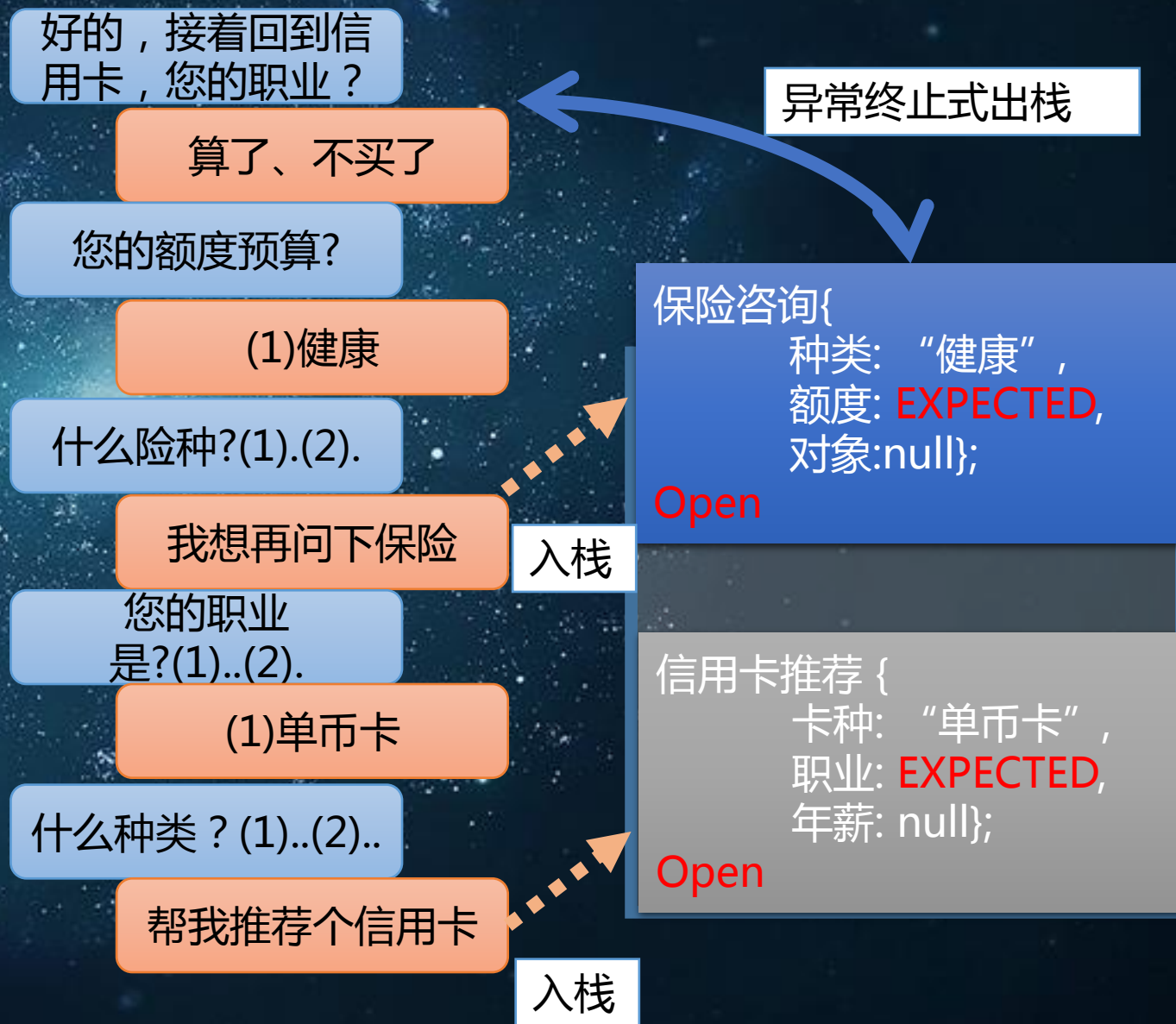
FQA问答编写加工工作量巨大  
问题扩展性支持极弱  
无法实现碎片化知识的自动加工

# 小富机器人多场景应用关联技术



# 多论会话

- 话题栈
  - 用来管理对话中**正在解决中 (Open)**的话题集合的机制
- 如何识别一个话题
  - 复杂型知识：要素型、流程型等
  - 主题迁移度：新主题和最近主题语义差异性建模
- 话题栈基本操作
  - 中断：话题之间可以发生中断，按照优先级判断机制
  - 入栈：当前话题未彻底解决完毕，需要进一步引导客户以获取更多必要信息
  - 出栈：当前话题被**完美解决**或**异常终止**



# 概念问答

基于概念的智能问答，解决不同问法，搜索给出同一答案

为什么要买交强险？

买交强险的好处？

一定要买交强险么？

交强险的意义

不买交强险会有什么后果？

开车从不出事，还用买交强险吗？

交强险是不是白交钱？

概念标签自动提取技术

形成内部相同的概念搜索表达式

e\_交强险+c\_必要性

admin2 2014-05-17 16:11:23

买交强险的好处

小富 2014-05-17 16:11:23

## 1. 机动车交通事故责任强制保险必要性

答案：必要性

以前实行的商业机动车第三者责任保险(以下简称“商业三责险”)是按照自愿原则由投保人选择购买。在现实中商业三责险投保比率比较低,致使发生道路交通事故后,有的因没有保险保障或致害人支付能力有限,受害人往往得不到及时地赔偿,也造成大量经济赔偿纠纷。因此,实行交强险制度就是通过国家法律强制机动车所有人或管理人购买相应的责任保险,以提高三责险的投保率,在最大程度上为交通事故受害人提供及时和基本的保障。实行交强险制度是通过国家法规强制机动车所有人或管理人购买相应的责任保险,以提高第三者责任险(简称“三责险”)的投保率,在最大程度上为交通事故受害人提供及时和基本的保障。



# 知识自动化加工

海尔洗衣机介绍 word文档

MS7518BZ51、MS8518BZ51 产品介绍..

一、产品型号定义..

二、随机附件..


三、产品外观图..

四、产品各部件名称..

五、产品功能介绍

洗涤过程

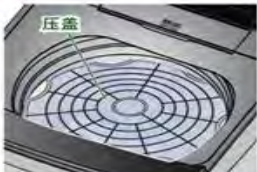
1、打开上盖，放入洗涤物..



(1) 为保证最佳的洗涤平衡效果，避免衣物缠绕，请按照如下注意事项放置衣物。

① 请将衣物直接放入洗衣桶内，并均匀放置在内桶底部，避免叠加。

② 本机机会自动传感衣物的重量，以求最小的耗水量，水位略有不同属于正常现象。



自动加工出 42 条知识  
文档中的图片也自动拆分为独立知识条目

序号	知识标题	知识内容	知识项-洗衣机部件	知识项-产品功能	知识项-洗涤剂	知识项-洗涤对象	知识项-动作
6	产品功能介绍-洗涤过程-打开上盖，放入洗涤物	1、打开上盖，放入洗涤物 (1) 为保证最佳的洗涤平衡效果，避免衣物缠绕，请按照如下注意事项放置衣物： ① 请将衣物直接放入洗衣桶内，并均匀放置在内桶底部，避免叠加。 ② 本机机会自动传感衣物的重量，以求最小的耗水量，水位略有不同属于正常现象。	上盖、洗衣桶	洗涤平衡、自动传感、耗水量			打开、放入、放置
7	产品功能介绍-洗涤过程-打开上盖，放入洗涤物-注意事项	注意：如果洗涤蓬松衣物（比如可以机洗的羽绒服等），请使用压盖将衣物压到平衡环以下，注意压盖的方向要按限压盖中部“此面向上”的标识放置。	压盖	平衡环		蓬松衣物、羽绒服	压到
8	产品功能介绍-洗涤过程-放入洗涤剂	2、放入洗涤剂 用户可在洗涤前将洗衣液、洗衣粉、柔顺剂添加到洗衣液、洗衣粉、柔顺剂投入口。洗衣过程中，洗衣机会将洗衣液、洗衣粉、柔顺剂自动添加到洗涤轴中，无需手动添加。投放方法如下： (1) 如图所示方向打开投入口盖，拿下塞子。 (2) 投放洗涤剂。 (3) 正确添加结束后，请将投入口盖附近擦拭干净，盖好投入口盖。	投入口、洗涤剂、投入口盖、塞子	自动添加、不能储存	洗衣液、洗衣粉、柔顺剂		放入、添加、投放、打开、拿下、擦拭、盖好
9	产品功能介绍-洗涤过程-放入洗涤剂-注意事项	注：本机不能储存洗衣液、洗衣粉、柔顺剂，长时间看加入一次洗衣所需的使用量，在进行下一次洗衣时重新进行添加。			洗衣液、洗衣粉、柔顺剂		投放、加入、添加
10	产品功能介绍-洗涤过程-关闭上盖	3、关闭上盖	上盖				关闭

# 人工智能认知领域客户的认可

业界首款  
面向业务建模的  
DINFO-OEC  
非结构化文本分析挖掘平台  
(基于智慧语义认知技术研发)



## 中国银行业 文本语义分析挖掘平台的首选平台

在与国内外著名公司包括IBM watson、SAS、HP等竞争中，成功服务于四大国有银行中的三家



2014年度  
中国非结构化大数据领域  
最佳解决方案奖



2016年度  
中国大数据非结构化大数据领域  
最佳产品





神州泰岳**认知**+ **用心**做好创新