

# 知识图谱及其在零售行业的多场景应用

吴友政

京东AI研究院 - 算法科学家

京东智联云 - 总监

2020.10.22

# 京东智联云，京东集团**技术服务**核心输出平台

## 云业务

- 全球最大规模Docker容器集群之一，中国最大的GPU集群之一；
- 整体架构经受住历年618、双十一大促的极端考验；
- 公有云IaaS市场前十



## IoT 业务

- 物联网IOT连接设备数量超过**1.5亿台**，累计服务**2500万**家庭；
- 接入**1000+**品牌，**230+**种品类，**2100+**款商品
- 国内最大的智能家居平台之一

- 国家新一代人工智能开放创新平台；
- 核心载体NeuHub京东人工智能开放平台，累计调用量达**万亿次**

## AI 业务

## 4 大全球研发中心



## 3 大实验室：专注于人工智能的前沿创新，80%的研究由业务需求驱动

- 语音语义实验室
- 计算机视觉与多媒体实验室
- 机器学习实验室

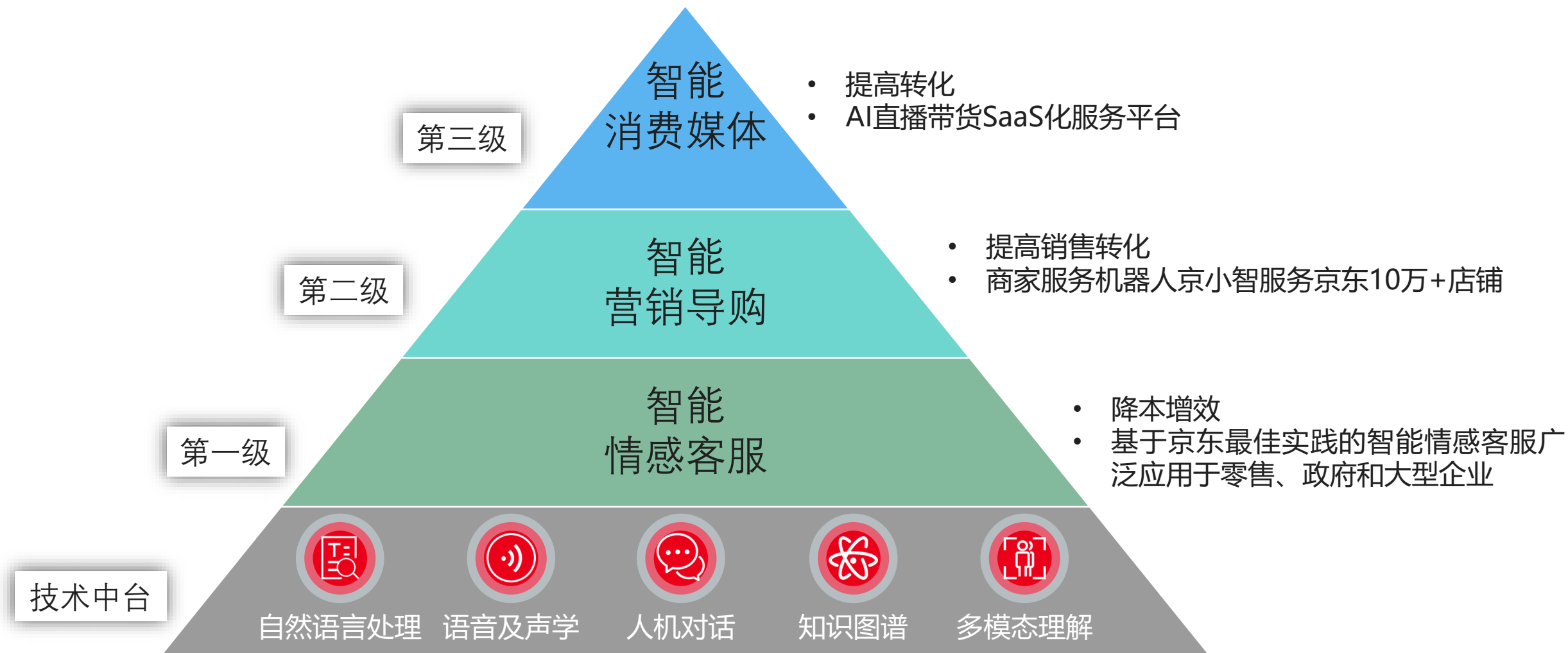
## 8 承担国家级研发项目 8 项

- 科技部 - 国家人工智能开放创新平台
- 科技部 - 科技创新2030—“新一代人工智能”重大项目2项
- 科技部 - 火炬计划和重点研发计划2项
- 发改委 - 人工智能创新发展重点工程
- 工信部 - 新一代人工智能产业创新重点任务2个

## 17 世界级人工智能赛事获得冠军17项

## 150 累计发表有影响力的学术论文150 篇，4篇最佳论文

- 探索语音与声学、自然语言处理、多轮对话、知识图谱、多模态在智能情感客服，智能营销和智能消费媒体三大场景下科研创新和商业化的落地实践



01

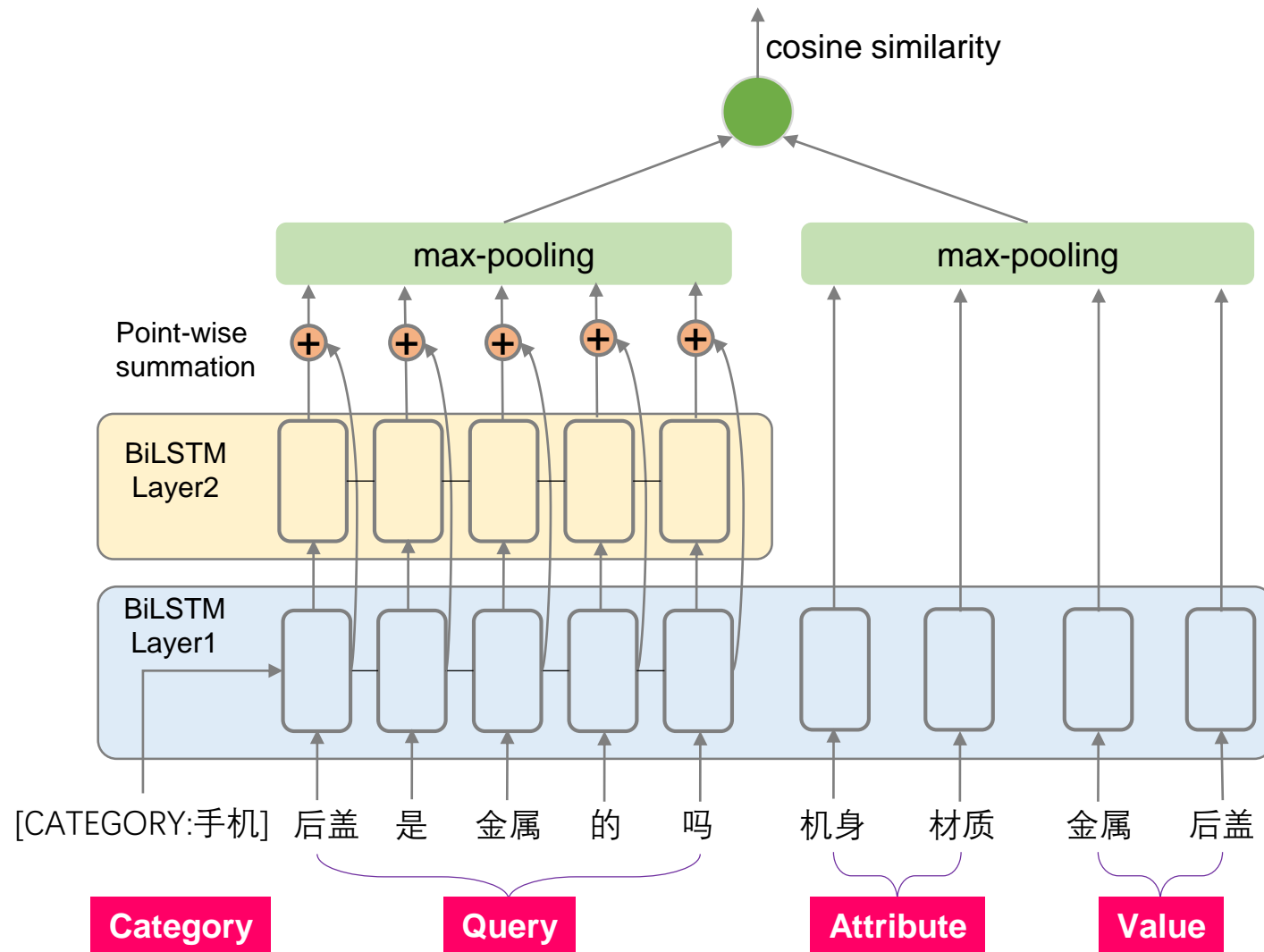
商品知识图谱：以商品为中心

02

药学知识图谱：以药品为中心

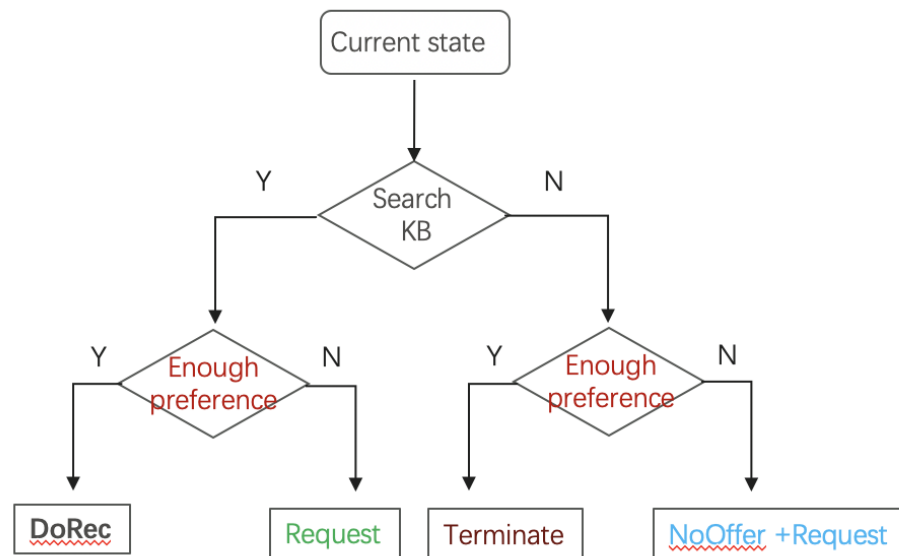
# 应用场景1 - 商品图谱问答

- 智能营销导购机器人京小智: 以“智慧服务, 高效创收”为使命帮助商家降低服务成本、提升店铺转化



# 应用场景2 – 对话式商品推荐

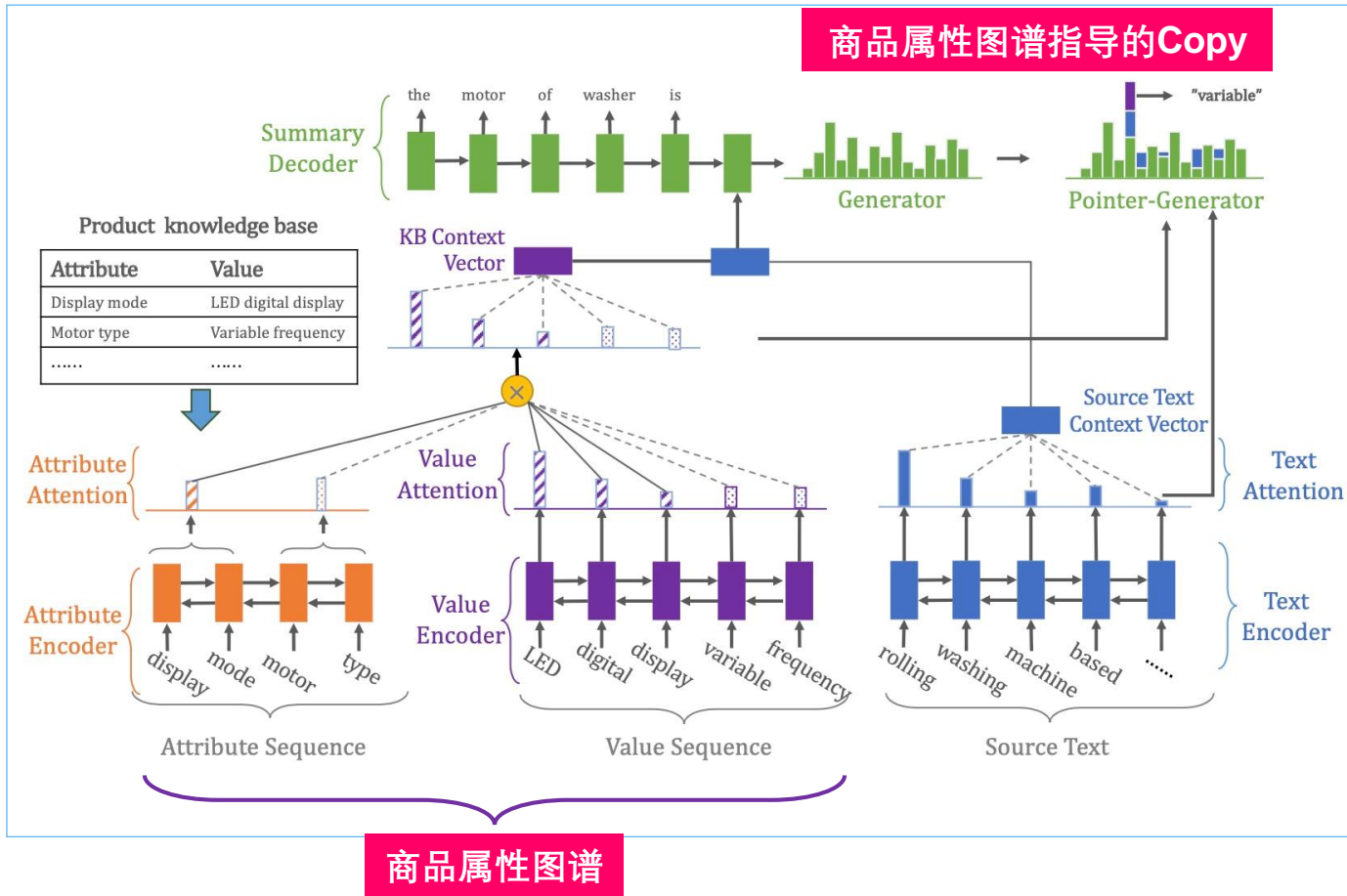
- 基于任务型多轮对话的人机交互形式、融合多种对话技术和推荐技术实现更高质量的自动推荐



意图识别: BERT  
槽值提取: BERT + Bi-LSTM + CRF  
DST: Rule-based  
DP: Rule-based  
NLG: Rule-based

# 应用场景3 – 面向忠实度的文案写作

- 商品知识图谱提供了重要的商品属性信息，加强模型对商品属性信息的认知，增强了模型生成的商品文案的忠实度





# 应用场景4 – 企业采购

- 面向大型企业和政府客户，提供一系列智能化（生产材料、办公用品、福利等）采购产品/工具



处方编号: 000477950800|1551 处方时间: 2019-02-01 15:33:36 就诊科室: 南区心血管内科 医生姓名: 徐顺霖

## 患者信息

患者: 苗\*\* 年龄: 59 性别: 女 身高: 163cm 体重: 63kg  
是否怀孕: 否 孕期: 是否哺乳: 否 过敏信息: 对辣椒、海鲜过敏; 对青霉素过敏;

 查看检查单

 查看过往病史

 查看手术

 查看家族病史

## 诊断信息




心衰 (150.911w11) ; 房缺 (Q21.102w11)

## 用药信息

HIS药品名称	药品规格	天数	总量	单次量	频次	给药途径
螺内酯片	20mgx20片/袋	1	80mg	40mg	1天2次	口服
伊迈格(托拉塞米片)	5mgx14片/盒	1	10mg	10mg	1天1次	口服
氯化钾缓释片	0.5g/片	1	1.5g	0.5g	1天3次	口服<饭后>

## 预警信息

### 相互作用

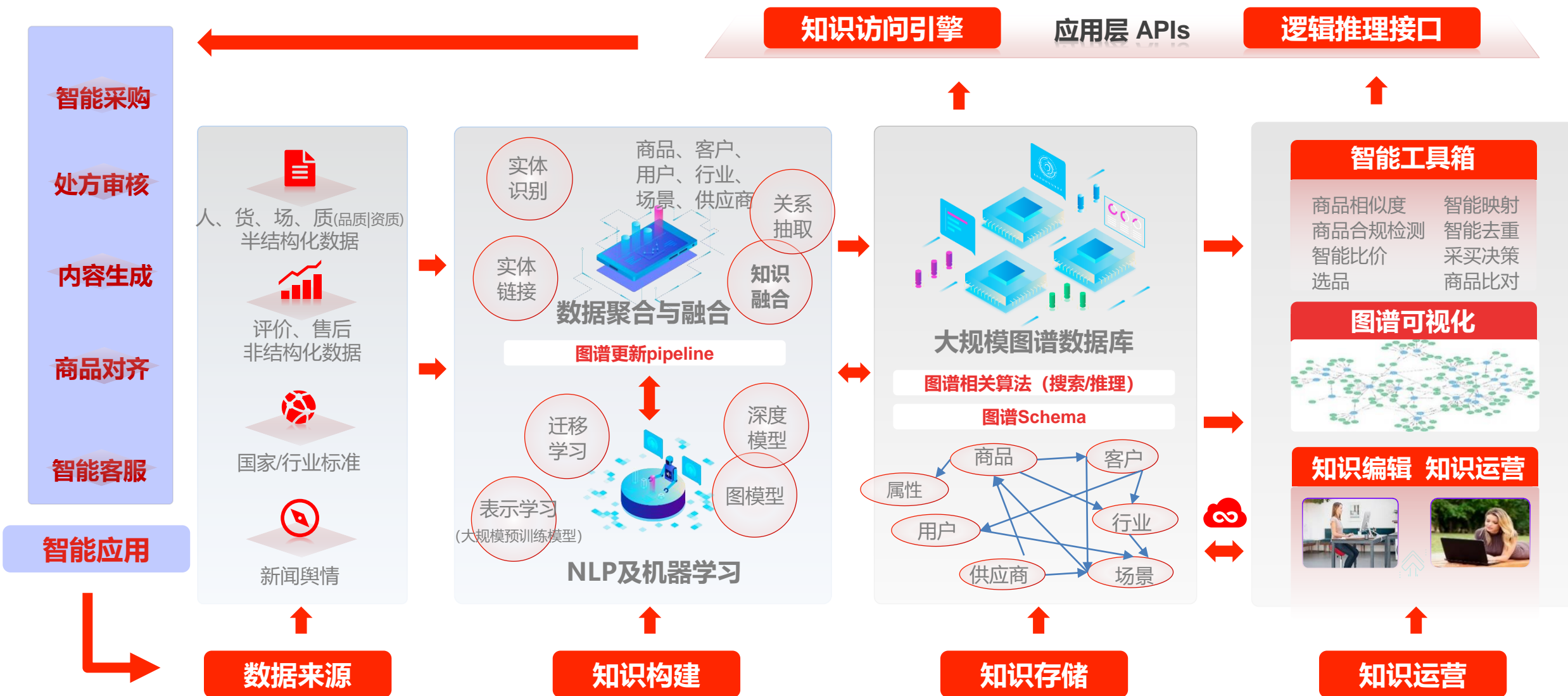
-  【3级】与XXXX合用, 可导致XXX。
-  【2级】本药宜静脉给药
-  【1级】给药频率不合适。成人宜1天N次

### 伊迈格 (托拉塞米片)

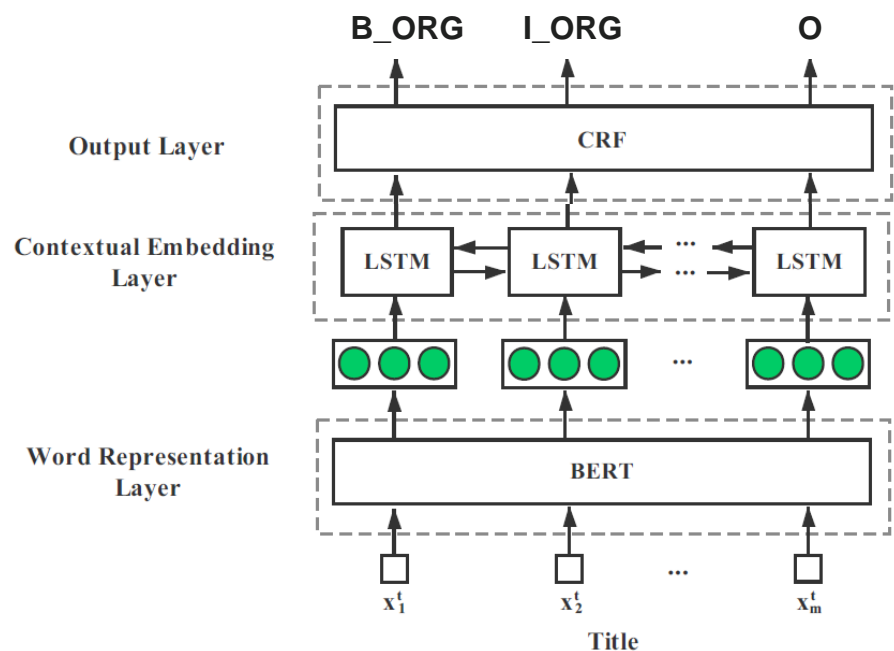
-  【1级】用药用量, 给药频率不合适。成人宜1天N次

药师是否审核: 是

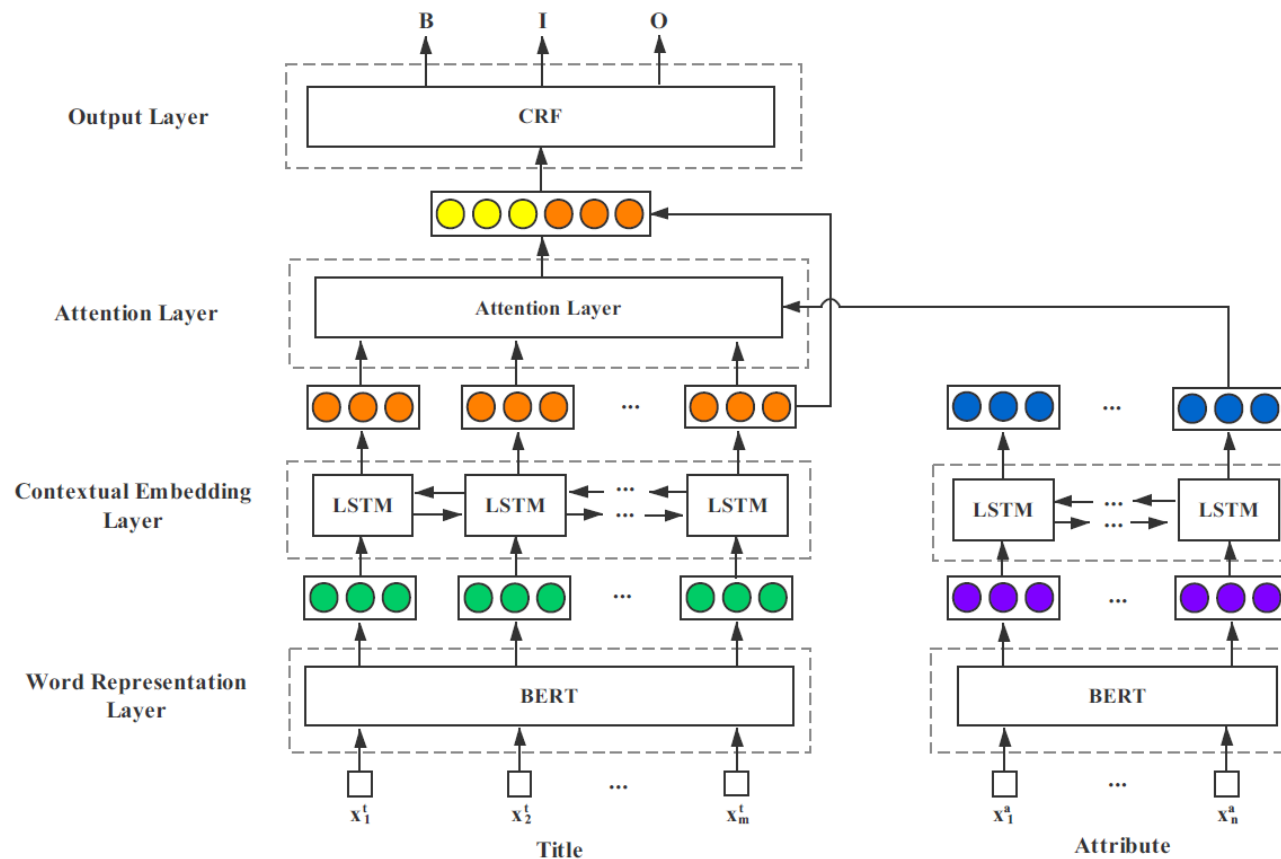
药师审核结果: 通过



- 零售电商平台上海量商品的属性-属性值对普遍不完备，影响用户体验
- 传统的序列标注NER模型（对每一个属性设置一套BIO标签）对于数万商品属性-属性值的抽取效果不佳



BERT+Bi-LSTM+CRF



Attribute-comprehension open tagging model  
[Xu, et al. 2019]

- 借助商品图片信息，基于Multi-task的框架实现对属性和属性值联合预测

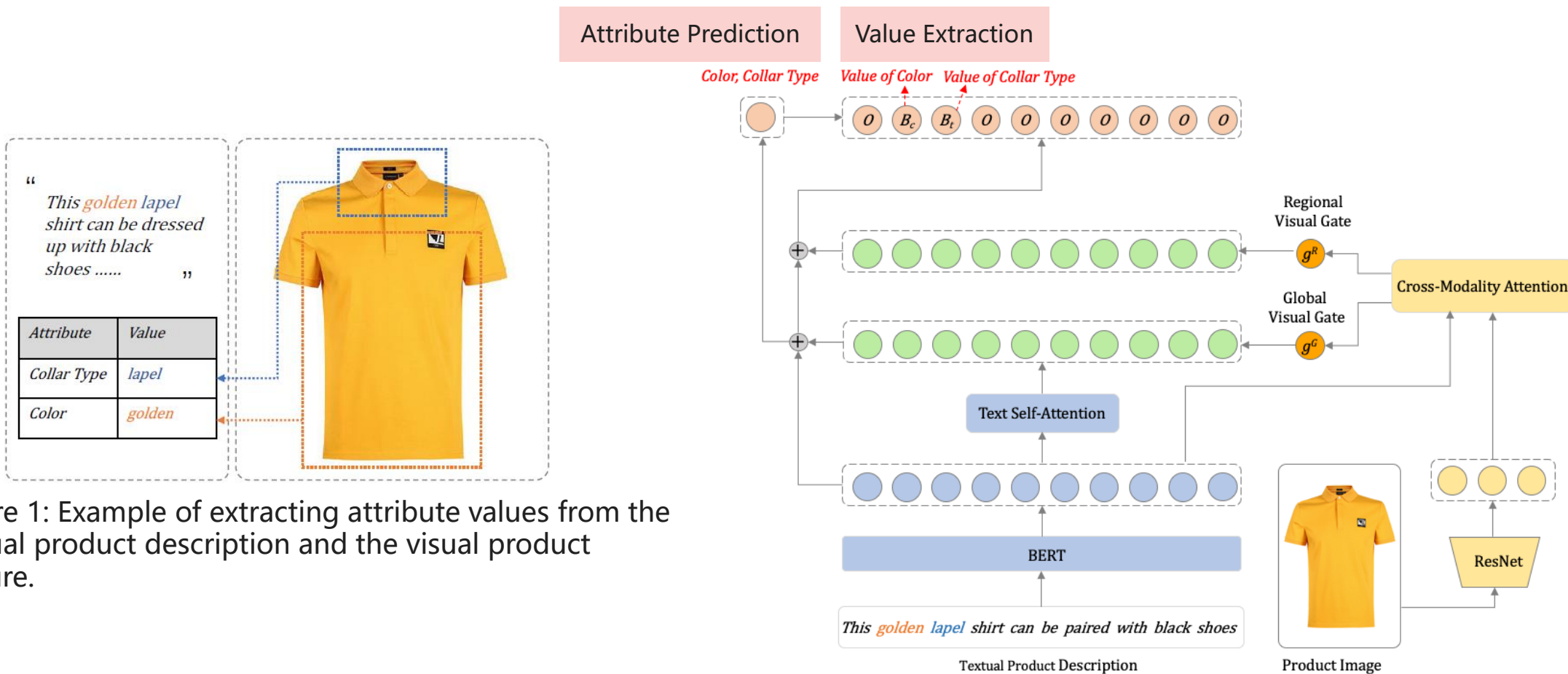


Figure 1: Example of extracting attribute values from the textual product description and the visual product picture.

- 多模态商品属性-属性值抽取数据集 <https://github.com/jd-aig/JAVE>

Category	#Product	#Instance	#Attr	#Value
Clothes	12,240	34,154	14	1,210
Shoes	9,022	20,525	10	1,036
Bags	3,376	8,307	8	631
Luggage	1,291	2,227	7	275
Dresses	4,567	12,283	13	714
Boots	713	2,090	11	322
Pants	2,832	7,608	13	595
Total	34,041	87,194	26	2,129

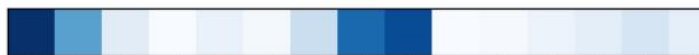
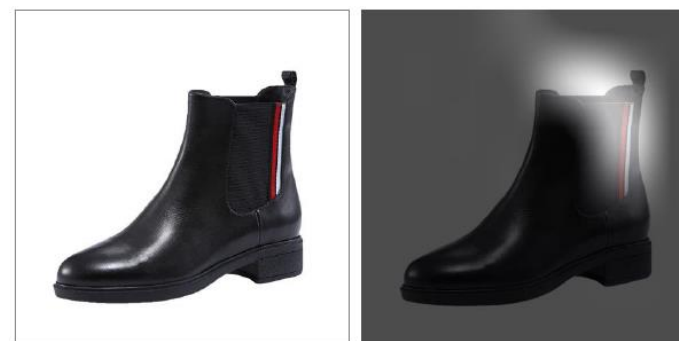
Table 1: Statistics of the our dataset

Model	Attribute	Value
WSM	77.20	72.52
Sep-BERT	86.34	83.12
RNN-LSTM (Hakkani-Tür et al., 2016)	85.76	82.92
Attn-BiRNN (Liu and Lane, 2016)	86.10	83.28
Slot-Gated (Goo et al., 2018)	86.70	83.35
Joint-BERT (Chen et al., 2019)	86.93	83.73
ScalingUp (Xu et al., 2019)	-	77.12
JAVE (LSTM based)	87.88	84.09
JAVE (BERT based)	87.98	84.78
M-JAVE (LSTM based)	90.19	86.41
M-JAVE (BERT based)	<b>90.69</b>	<b>87.17</b>

Table 2: F1 score (%) of comparative methods and variants of our model



利落的 **小翻领** 搭配甜美撞色绑带  
neat **little lapel** with a sweet contrast band



**黑色** 稳重大气与 **高帮** 设计更好加持  
the stable **black** matches the **high-top** design very well

Figure 2: Heat maps for global (blocks above the text) and regional (images on the right) visual gates

# Entity Alignment

Finding entities in a KG or from different KGs that refer to the same real-world identity

- 同一知识体系中的实体对齐
- 不同知识体系中的实体对齐



零售商



TMALL天猫

供应商

序号	物料名称	规格型号
1	水晶板	1m宽*1mm厚
2	聚氨酯封堵-利通	ZY-SX031-C2-2196
3	机床滤芯	0063RN010BN4HC HYDAC
4	空气过滤器-FESTO	159594 LFP-D-MIDI-5M
5	螺钉	M6*15
6	纽扣头螺钉-Mapal	10029150
7	西霸王连接器-西霸王	H24B-TS-PG21
8	电烙铁芯(外热式)	
9	聚氨酯封堵-大连智云	ZY-S266-20112
10	半圆头螺钉	M4*20

一级分类	二级分类	品牌	编码	产品名称
个人护理	女性护理	护舒宝	2067636	护舒宝(Whisper)超薄日用 云感超净棉卫生巾 240mm 16片(绵柔 瞬吸 透气 贴身)
个人护理	女性护理	护舒宝	1443315	护舒宝瞬洁洁丝薄日用卫生巾5+1
个人护理	女性护理	护舒宝	1375798	护舒宝(Always)考拉呼吸极薄夜用卫生巾, 5片 360mm 夜用
生活用品	凑单商品	美丽雅	1439944	棉签 HC047457 100支双头棉签
个人护理	手部护理	隆力奇	1365032	隆力奇蛇油护手霜 120g
个人护理	身体护理	洁丽雅	1124283	洁丽雅(Grace) 纯棉强吸水舒适面巾 6982-1A 74*34cm, 蓝色
个人护理	身体护理	洁丽雅	1347829	洁丽雅 紫色缎档面巾 6638 70*34cm 红色
个人护理	身体护理	金号	1287125	金号浴巾 G3955H 灰色
个人护理	身体护理	金号	1307648	金号(KING SHORE)纯棉面巾 2185H 74*33cm 110g, 红

单据号	物料编码	旧物料编码	物料类型	大类	中类	小类	短描述	长描述	审批状	特征项	特征值	特征项	特征值	特征项	特征值
3811943	19082132	HL_JBFSWM019	Z100-国能物08-燃油油	0807-润滑	080712-复	复合锂基润滑脂\SHC	复合锂基	AS1-物资	名称	复合锂基	牌号	SHC GREAS	品牌	美孚	
3812563	19082144	HL_JBFSWM015	Z100-国能物08-燃油油	0807-润滑	080732-特	开齿开齿液\CEPLATY	开齿开齿液	AS1-物资	名称	开齿开齿液	牌号	CEPLATY	品牌	福斯	
3810949	12336167		Z100-国能物08-燃油油	0803-润滑	080306-液	抗磨液液压油\DTE 25 V	抗磨液液压油	AS1-物资	名称	抗磨液液压油	牌号	DTE 25 V	品牌	美孚	
3810528	12336168		Z100-国能物08-燃油油	0803-润滑	080320-热	导热油\Therminol 55	导热油	AS1-物资	名称	导热油	牌号	Therminol	品牌	首诺	
3809680	12336166		Z100-国能物08-燃油油	0803-润滑	080304-齿	重负荷工业齿轮油\CKI	重负荷工业	AS1-物资	名称	重负荷工业	牌号	CKD 320	品牌	加士孚	
3809678	12336165		Z100-国能物08-燃油油	0803-润滑	080304-齿	重负荷工业齿轮油\CKI	重负荷工业	AS1-物资	名称	重负荷工业	牌号	CKD 220	品牌	加士孚	
3809721	19081958	3590000000449	Z100-国能物08-燃油油	0803-润滑	080306-液	液压油\PULSALUBE 1#	液压油	AS1-物资	名称	液压油	牌号	PULSALUBE	品牌	帕斯菲达	
3802534	12335198		Z100-国能物08-燃油油	0807-润滑	080732-特	电重负荷开式齿轮油	电重负荷	AS1-物资	名称	电重负荷	牌号	0#;1×15#	品牌	昆仑	
3802504	12335191		Z100-国能物08-燃油油	0807-润滑	080732-特	电重通用航空润滑油	电重通用	AS1-物资	名称	电重通用	牌号	7008;1×1	品牌	昆仑	
3802501	12335155		Z100-国能物08-燃油油	0807-润滑	080732-特	电重全能润滑油\船用	电重全能	AS1-物资	名称	电重全能	牌号	船用 SUP6	品牌	昆仑	
3802520	12335154		Z100-国能物08-燃油油	0807-润滑	080732-特	电重轮轨润滑油\KL-B	电重轮轨	AS1-物资	名称	电重轮轨	牌号	KL-B;12×	品牌	昆仑	
3802517	12335152		Z100-国能物08-燃油油	0807-润滑	080732-特	电重硅脂\7057;12×0	电重硅脂	AS1-物资	名称	电重硅脂	牌号	7057;12×	品牌	昆仑	
3802513	12335188		Z100-国能物08-燃油油	0807-润滑	080732-特	电重硅脂\7057;1×15	电重硅脂	AS1-物资	名称	电重硅脂	牌号	7057;1×1	品牌	昆仑	
3802492	12335186		Z100-国能物08-燃油油	0807-润滑	080726-极	电重轴承润滑油\HTHS	电重轴承	AS1-物资	名称	电重轴承	牌号	HTHS;1×1	品牌	昆仑	
3802484	12336149		Z100-国能物08-燃油油	0803-润滑	080312-压	真空泵油\100#;1×16	真空泵油	AS1-物资	名称	真空泵油	牌号	100#;1×1	品牌	长城	
3802491	12335184		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	MEP-1;1×	品牌	昆仑	
3802490	12335151		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	MEP-1;1×	品牌	昆仑	
3802488	12335182		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	MEP-II;1	品牌	昆仑	
3802463	12335148		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	MEP-II;1	品牌	昆仑	
3802403	12335181		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	3#;1×17	品牌	昆仑	
3802401	12335146		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	2#;1×17	品牌	昆仑	
3802453	12335177		Z100-国能物08-燃油油	0807-润滑	080726-极	电重极压锂基润滑油	电重极压	AS1-物资	名称	电重极压	牌号	000#;1×1	品牌	昆仑	

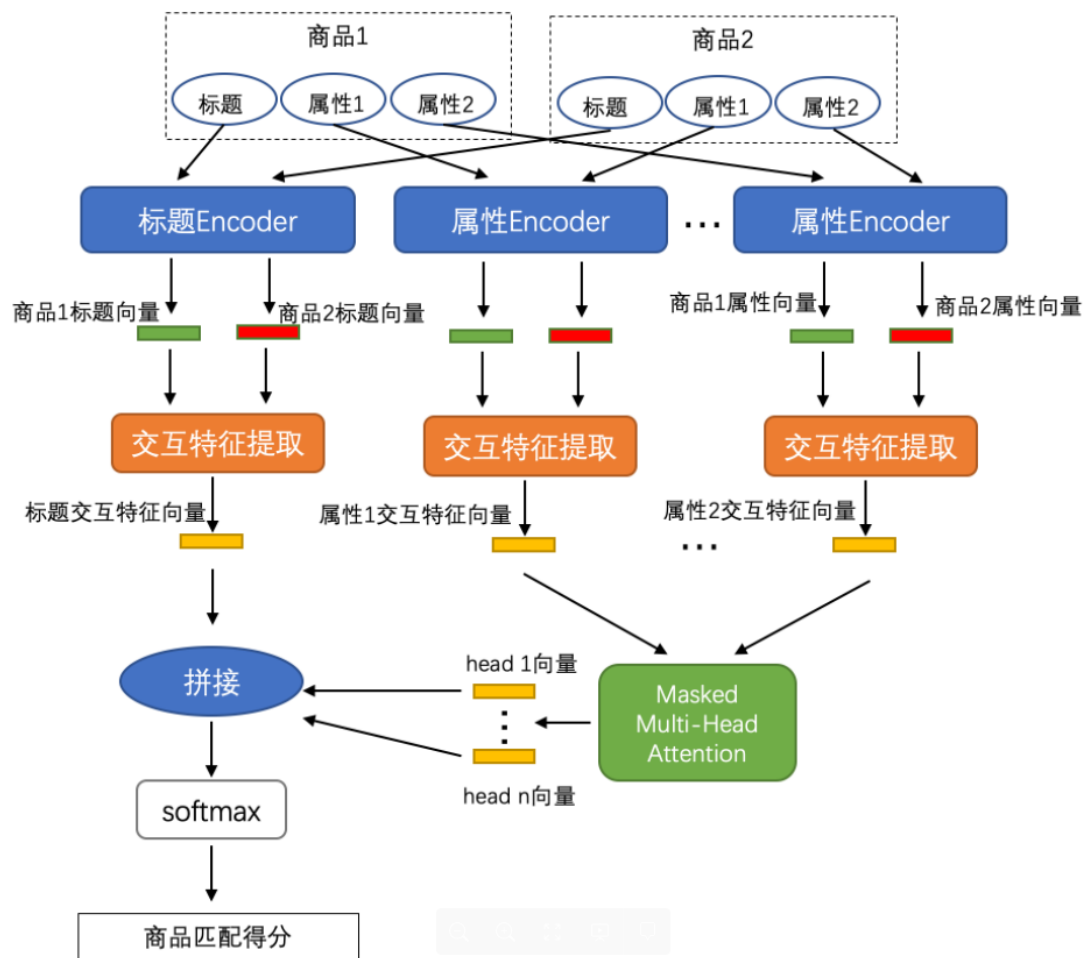
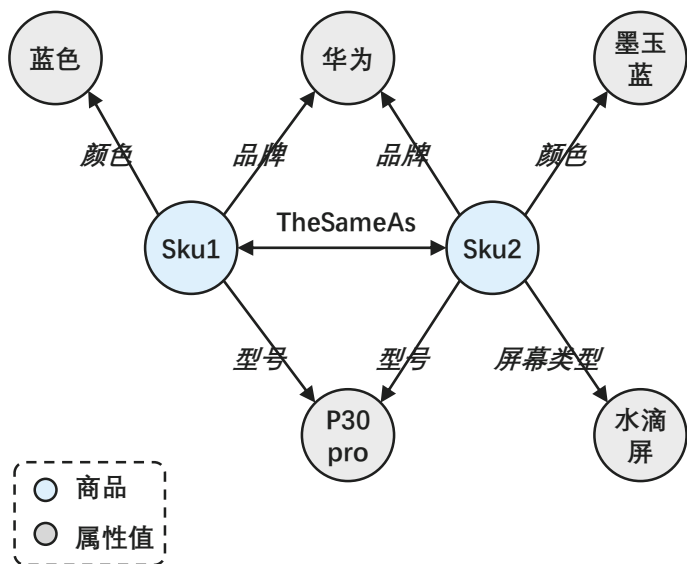
采购商

物料编码	物料名称	二级	三级
A260299000014	角磨片125MM 27型	耗材	抛光打磨耗材
B239900000003	多功能擦车纸300×3501910	行政	福利物资
60201212	扎带BHT375M 带宽7.6	辅料	包装材料
60117310	焊钉M4×Φ6×8GB10432.1 拉弧式	设备、装备	设备配件
60273586	透明钢丝软管Φ25	设备、装备	设备配件
A550199000958	带扣透明塑料袋300×200	行政	宣传文化
A260299000081	珩磨条200×20 W150K	耗材	抛光打磨耗材
60286440	镀锌打包铁皮32×0.9	辅料	包装材料
A290000000258	导电嘴1.2	设备、装备	设备配件
170702990005A	汽车遮阳膜1.4M×30M(日高)	耗材	其他耗材

# Entity Alignment

同一知识体系中的实体对齐

- Schema已对齐、品类Schema完整程度不同 (4000+个细分品类)
  - 规则引擎, 人工指定关键属性, 计算2个商品在对应关键属性上的相似度
  - **深度匹配模型**

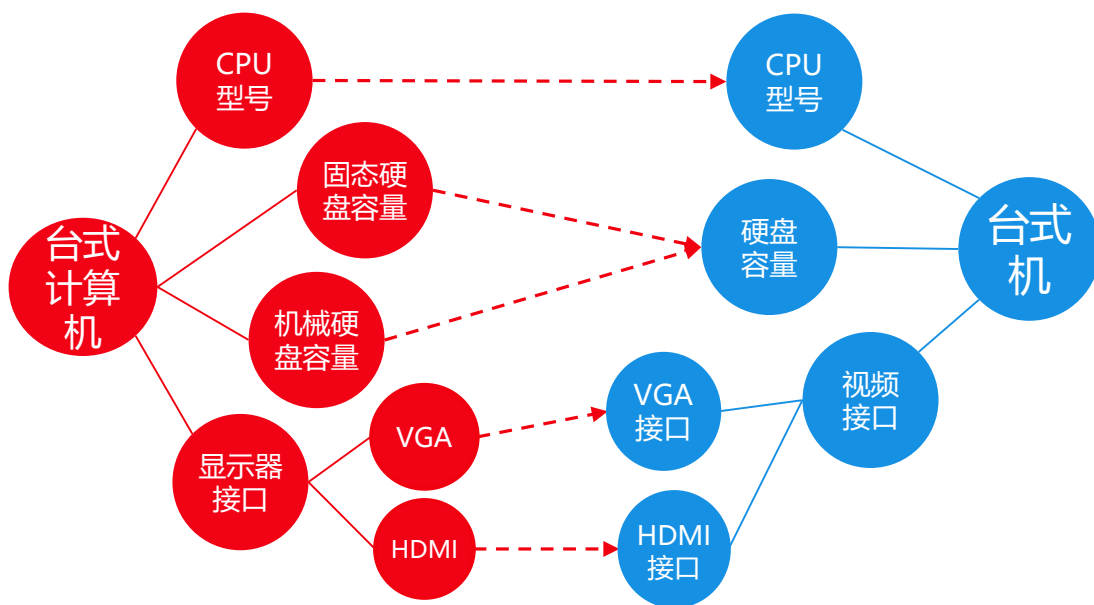




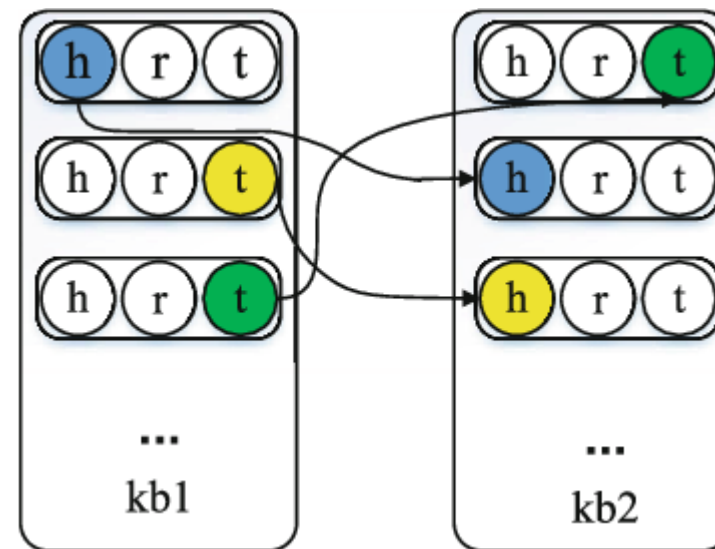
# Entity Alignment

不同知识体系中的实体对齐

- Schema不同、完整程度也不同
  - 属性与属性值的映射
  - Trans家族模型
    - TransE[Bordes, et al. 2013], TransR[Lin, et al. 2015], JE[Hao, et al. 2016], MTransE[Chen et al. 2017]



图谱属性映射 [王雪鹏等. 2017]

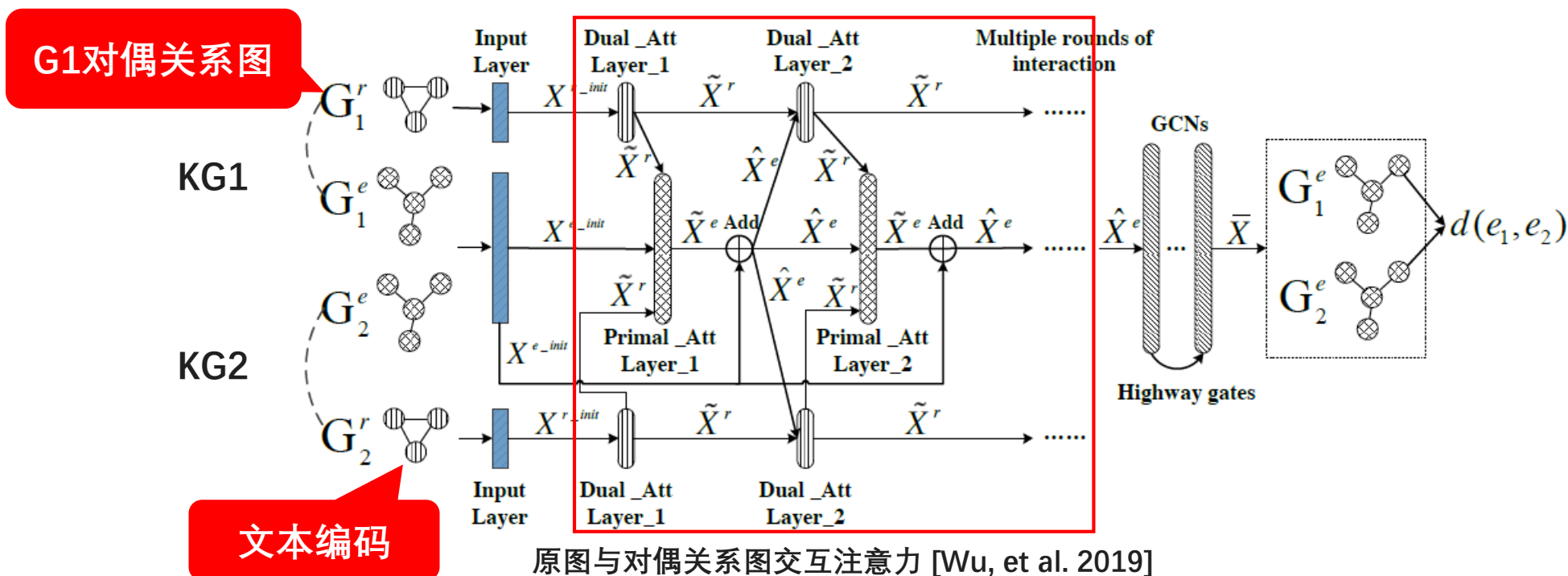


JE [Hao, et al. CCKS2016]

# Entity Alignment

不同知识体系中的实体对齐

- Schema不同、完整程度也不同
  - 属性与属性值的映射 [王雪鹏等. 2017]
  - Trans家族模型
    - TransE[Bordes, et al. 2013], TransR[Lin, et al. 2015], JE[Hao, et al. 2016], MTransE[Chen et al. 2017]
  - 图表示学习
    - 图神经网络: GCN [Wang, et al. 2018], R-GCN [Schlichtkrull, et al. 2017], RD-GCN [Wu, et al. 2019]等



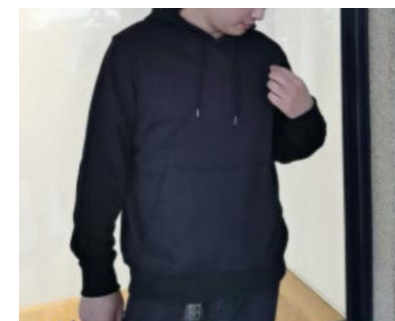
# JDDC 2.0 多模态任务型对话数据集 (<http://jddc.jd.com>)

含有丰富多模态信息与知识信息的对话数据集

- ✓ 17家店铺线上金牌客服在线对话
- ✓ 2个品类服饰鞋靴品类和小家电品类
- ✓ 2020年3月期间产生的多模态对话日志

会话session id	商品品类	商品 skuid	对话内容 (文本+图片)	对话角色
e39858	fx	603981	那还麻烦你看一下这个我穿m 20631e463cae3e50ce8565e4dad8	0
e39858	fx	603981	d676.jpg	0
e39858	fx	603981	你这个衣服m是不是也够了	0
e39858	fx	603981	您的身高体重多少呢	1
e39858	fx	603981	不用看我体重	0
e39858	fx	603981	你就看看我发的图片m	0
e39858	fx	603981	这款衣服m码适合身高180 体重 150-160	1
e39858	fx	603981	您可以参考一下的	1
e39858	fx	603981	那我差不多	0

多模态图片



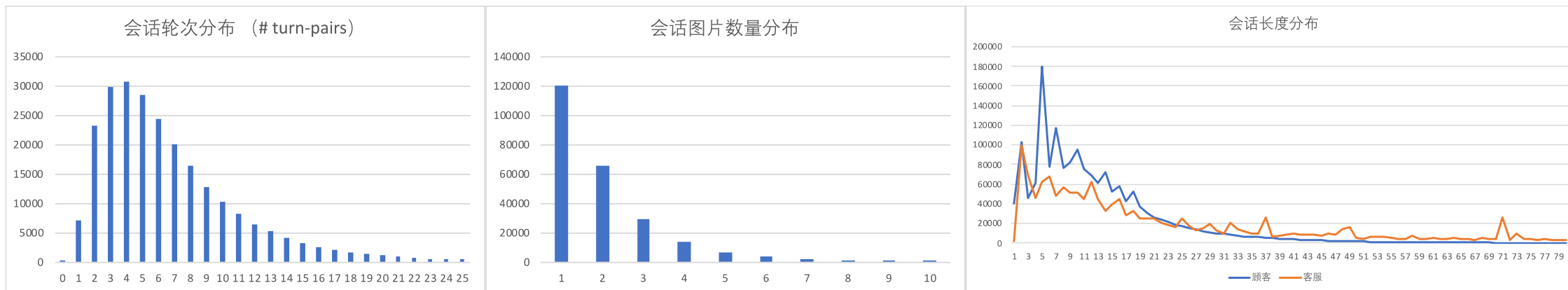
商品知识库

```
{  
  "分类": "卫衣/套头衫",  
  "pid": "603981",  
  "适用人群": "男士",  
  "款式": "套衫"  
}
```

# JDDC 2.0 多模态任务型对话数据集

含有丰富多模态信息与知识信息的对话数据集

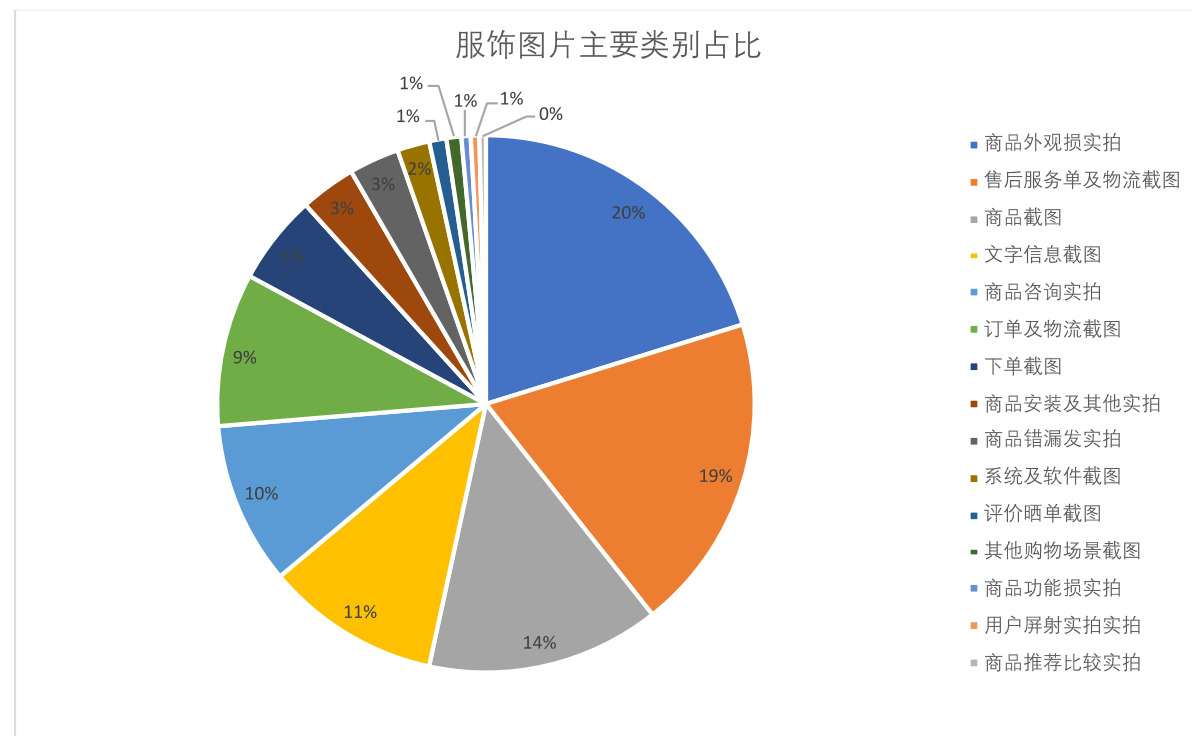
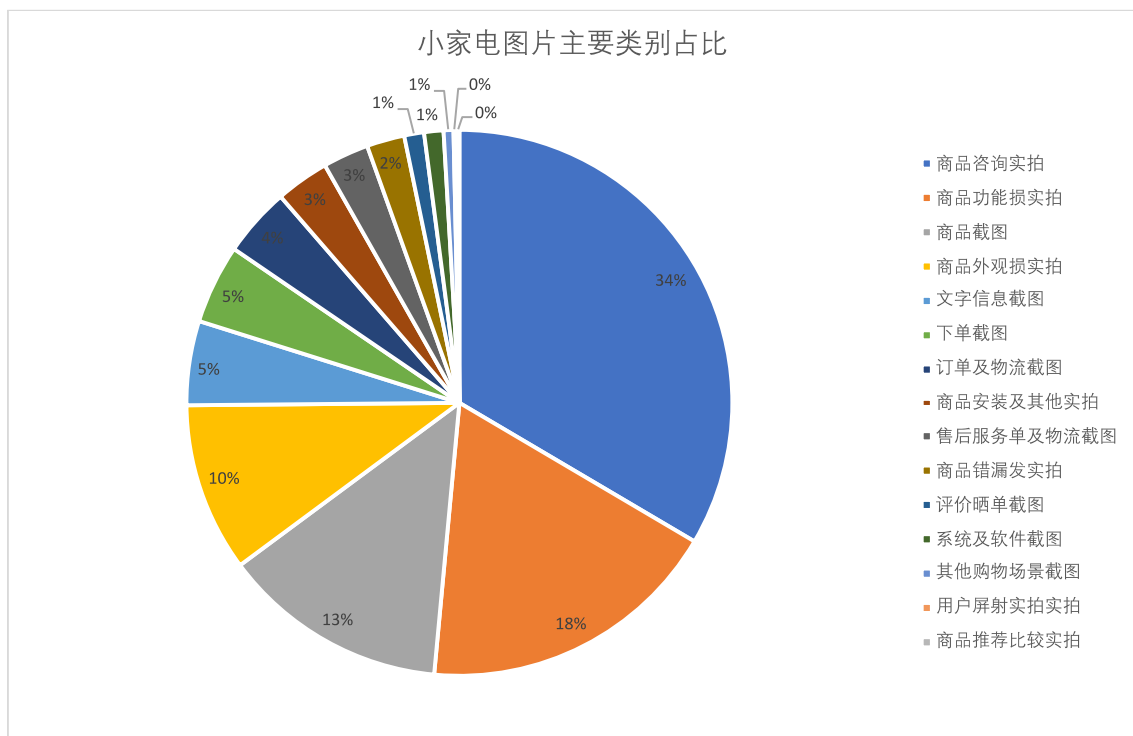
	总计	训练集				开发集			测试集		
		合计	小家电	服饰	合计	小家电	服饰	合计	小家电	服饰	
会话数量	<b>246,153</b>	<b>196,926</b>	103,555	93,371	<b>24,615</b>	12,941	11,674	<b>24,612</b>	12,935	11,674	
会话轮次	<b>3,459,888</b>	<b>2,765,970</b>	1,481,151	1,284,819	<b>348,509</b>	185,879	162,630	<b>345,409</b>	185,197	160,217	
平均会话轮次	<b>14.06</b>	<b>14.05</b>	14.3	13.76	<b>14.16</b>	14.36	13.93	<b>14.03</b>	14.32	13.72	
图片数量	<b>507,678</b>	<b>405,935</b>	210,386	195,549	<b>50,824</b>	26,218	24,696	<b>50,919</b>	26,031	24,888	
平均会话图片数量	<b>2.06</b>	<b>2.06</b>	2.03	2.09	<b>2.06</b>	2.03	2.11	<b>2.07</b>	2.01	2.13	
平均会话长度	<b>27.24</b>	<b>27.25</b>	23.77	31.25	<b>27.21</b>	23.81	31.09	<b>27.2</b>	23.68	31.29	



# JDDC 2.0 对话数据集 – 图片分类辅助信息

图片细分58种场景分类信息

- ✓ 5000段小家电类对话，涉及8218张图片
- ✓ 5000段服饰类对话，涉及8731张图片



# JDDC 2.0 对话数据集 – 商品知识库辅助信息

详尽描述商品分类、属性、卖点等信息

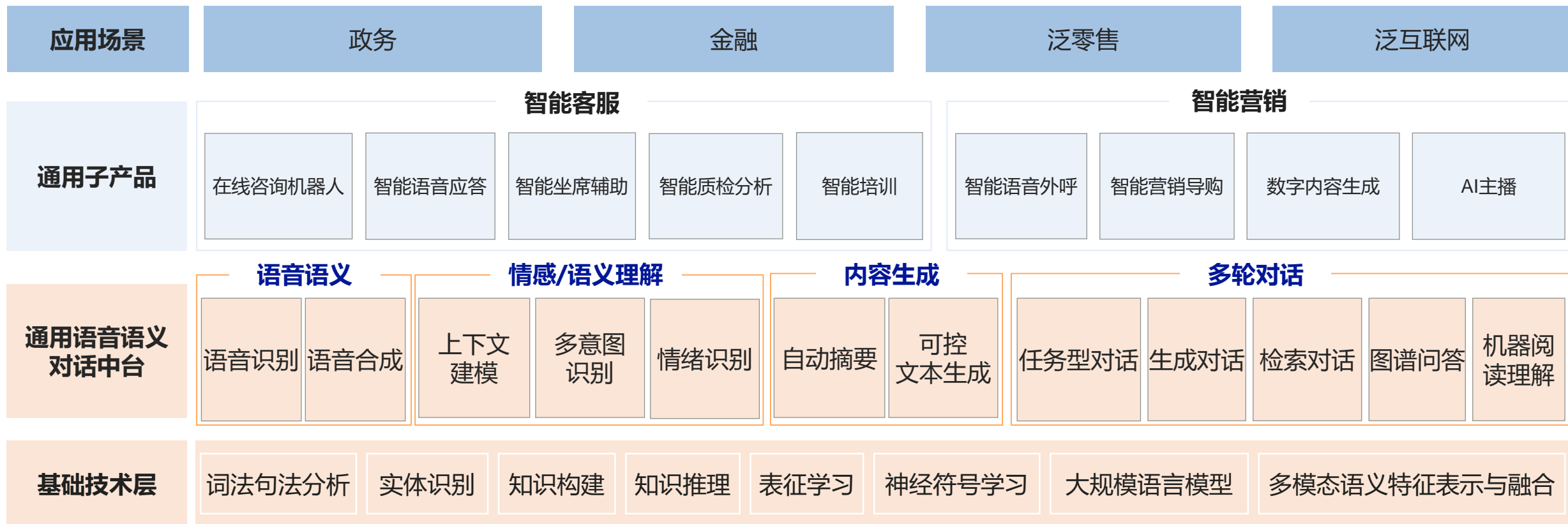
✓ 提供对话数据集涉及30205种商品对应的知识库信息

类别	数量	说明
Entity	30205	对话场景中涉及商品
Entity Type	231	e.g. 养生壶、空气炸锅、篮球鞋、牛仔裤
Relation	759	e.g. 防烧干功能、可拆卸炸篮、鞋面材质、腰型
Triple	219121	详细描述商品属性及商品卖点信息

商品知识库示例:

```
{  
  "分类": "牛仔裤",  
  "sku": "596ec542f79bc157ee57d2d67fada89a",  
  "适用场景": "上班",  
  "弹性": "微弹",  
  "厚度": "常规",  
  "适用人群": "青年",  
  "腰型": "常规",  
  "裤门襟": "拉链",  
  "裤长类型": "长裤",  
  "材质": "棉",  
  "裤型": "直筒裤",  
  "流行元素": "车缝线",  
  "适用季节": "春季",  
  "风格": "基础大众",  
  "工艺": "水洗"  
}
```

# 自然语言处理技术应用全景图



# THANKS

[wuyouzheng1@jd.com](mailto:wuyouzheng1@jd.com)

