



2017 CTRIP BIG DATA PLAN

探索大数据世界

携程度假智能云客服平台

携程度假BU 大数据研发总监 李健

Outlines

携程度假智能云服务平台

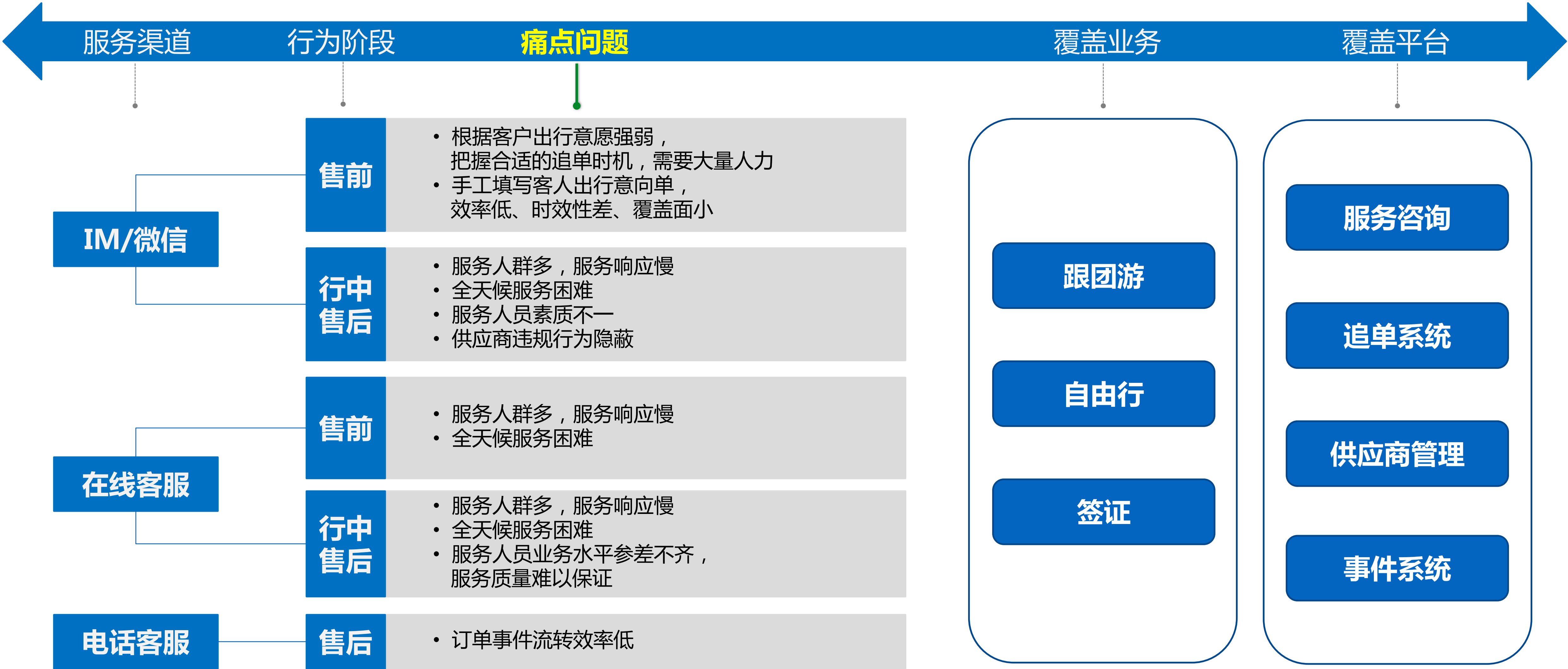
系统架构

算法模型

自助AI平台

未来展望

携程度假智能云客服平台—上线前





服务流转效率

服务响应时间

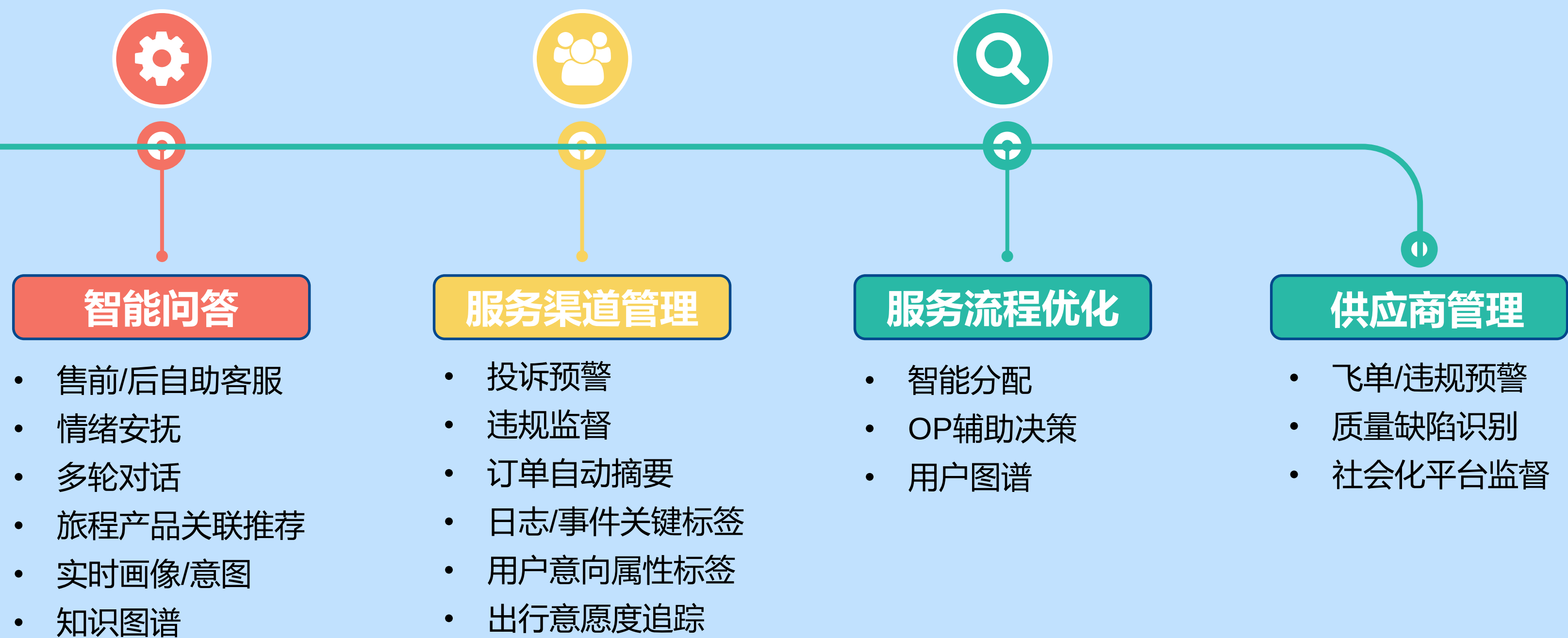
标准化服务

自动化服务

违规管理

流程优化

携程度假智能云客服平台



系统架构

外部应用

自动客服

智能分配

OP辅助

搜索

推荐

产品维护

算法

自动纠错

NLU

多轮对话

内容抽取

图像识别

意图识别

NLG

语言模型

主题模型

语音合成

NER

知识图谱

情绪分析

路径分析

深度推荐

工程

EasyAI

模型管理

语料管理

实时计算

近场计算

任务管理

部署服务

资源调度

并行优化

数据

采集

挖掘

聚合

迁移

清洗

整理

查询

反馈

异构数据源

MySQL

ES

Hive/Hbase

redis

hermes

Spark
streaming

QMQ

生产平台

商品信息

业务日志

VBK

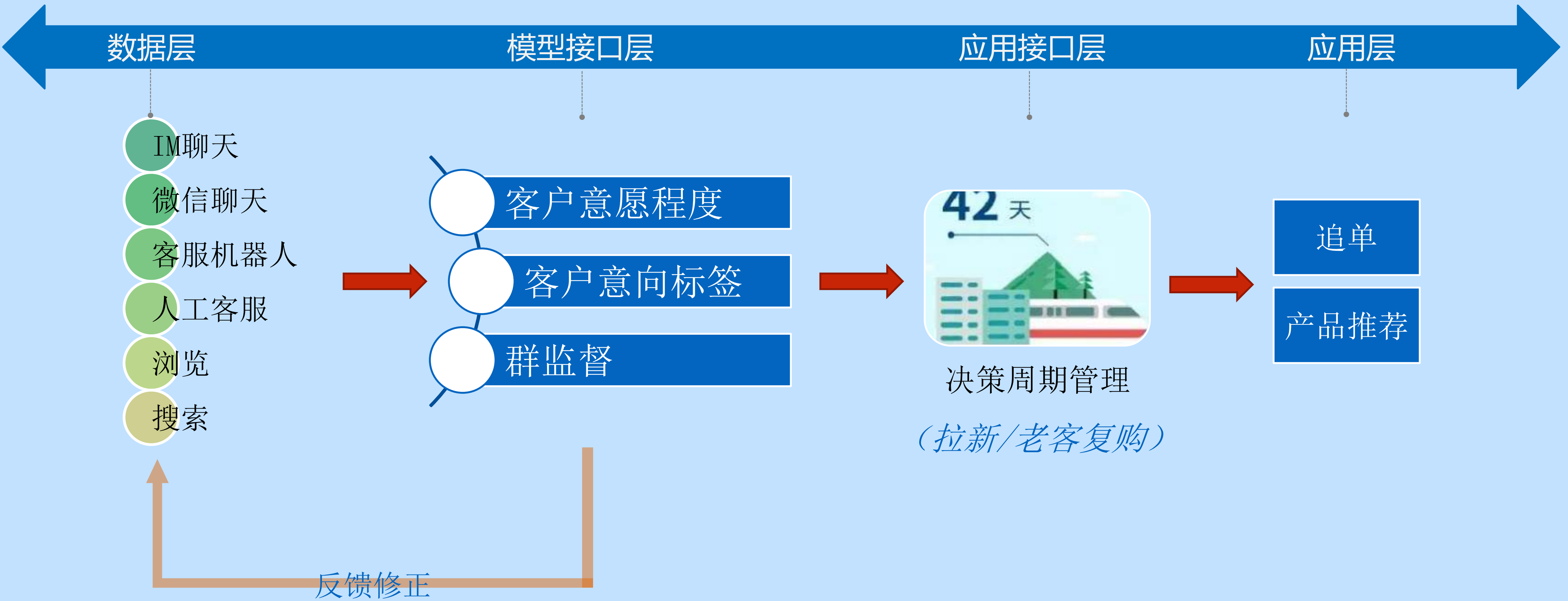
服务日志

用户画像

社会化平台

外部数据

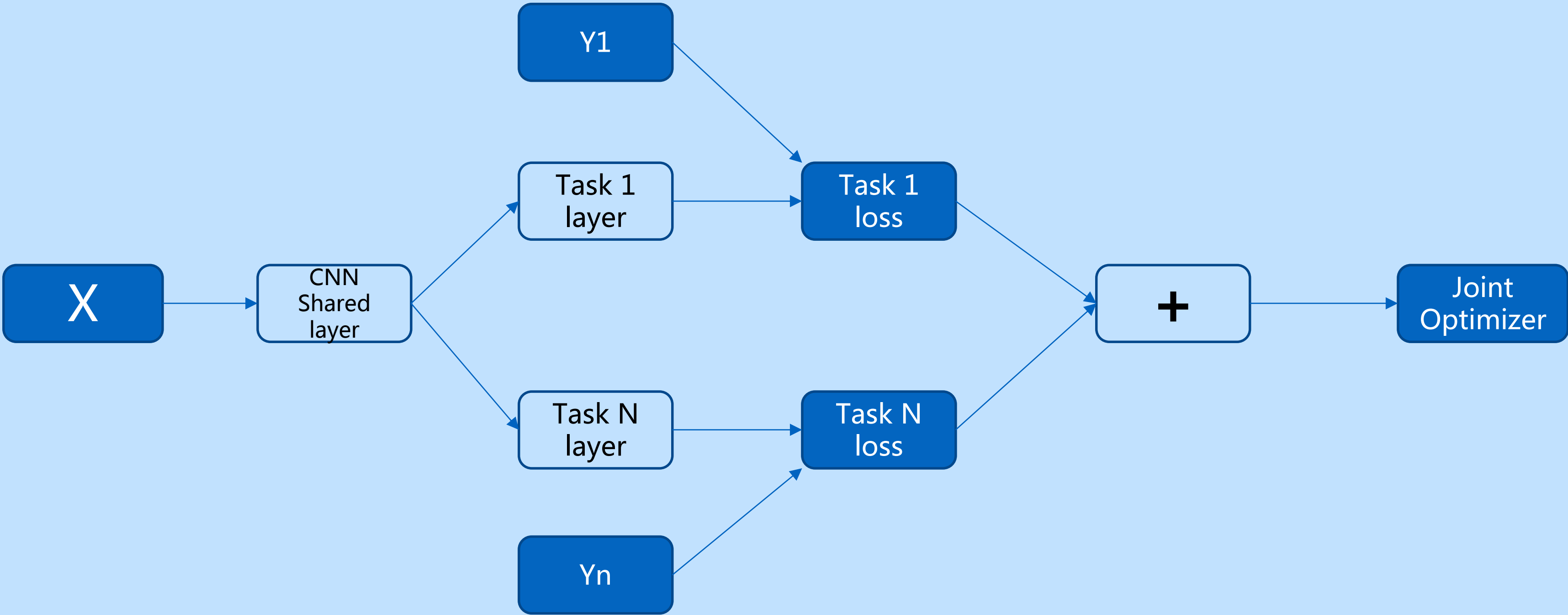
应用架构



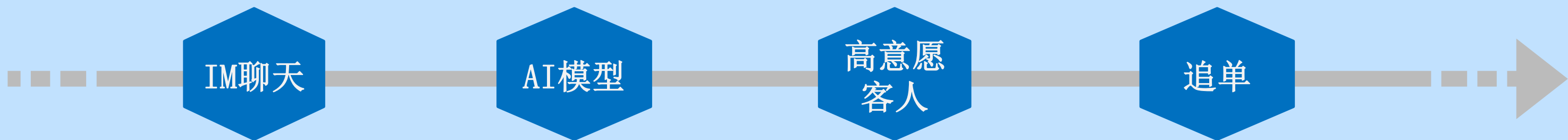
算法模型—以深度学习为主

用户意向标签

	追单ID	来源	入库时间	处理时间	用户ID	联系人	出发地	用户意向	钻级	产品/目的地	状态	操作		
TD		IM聊天 (IM)	2017-09-18 12:45:03				兰州	80% 		厦门 福建 ( )	未分配	订单	关联单据	
												日志	放弃	



案例：出行意愿度追踪



追单系统

设置筛选条件

查询追单订单

我的预约

追单系统 > 查询结果

输入单据ID

搜索

T团队游 Z自由行 C邮轮 Y游学 M攻略 D疑似重复

来源说明: Online-PC端网页 Offline-电话呼入 Hybrid-携程APP H5-手机网页 You-攻略 Adviser-领队 IM-IM聊天

导出所有单据

	追单ID	来源	入库时间	处理时间	用户ID	联系人	出发地	钻级	产品/目的地	状态	操作
T		IM聊天 (IM)	2017-11-02 08:48:40	2017-11-02 10:58:33			上海	4钻	昆明 云南	已完成 n18164	日志 4 00919410 8
T		IM聊天 (IM)	2017-11-02 08:48:38	2017-11-02 18:39:53		果果	北京		大阪 日本	已完成 n18997	日志 4 01076546 0
T		IM聊天 (IM)	2017-11-01 07:58:47	2017-11-01 14:36:40		李	上海	3钻	厦门 福建	已完成 n28175	日志 4 00551854 5
T		IM聊天 (IM)	2017-11-01 07:58:46	2017-11-01 18:00:25		刘	北京	4钻	伊斯坦布尔 土耳其	已完成 n20468	日志 4 00489368 7

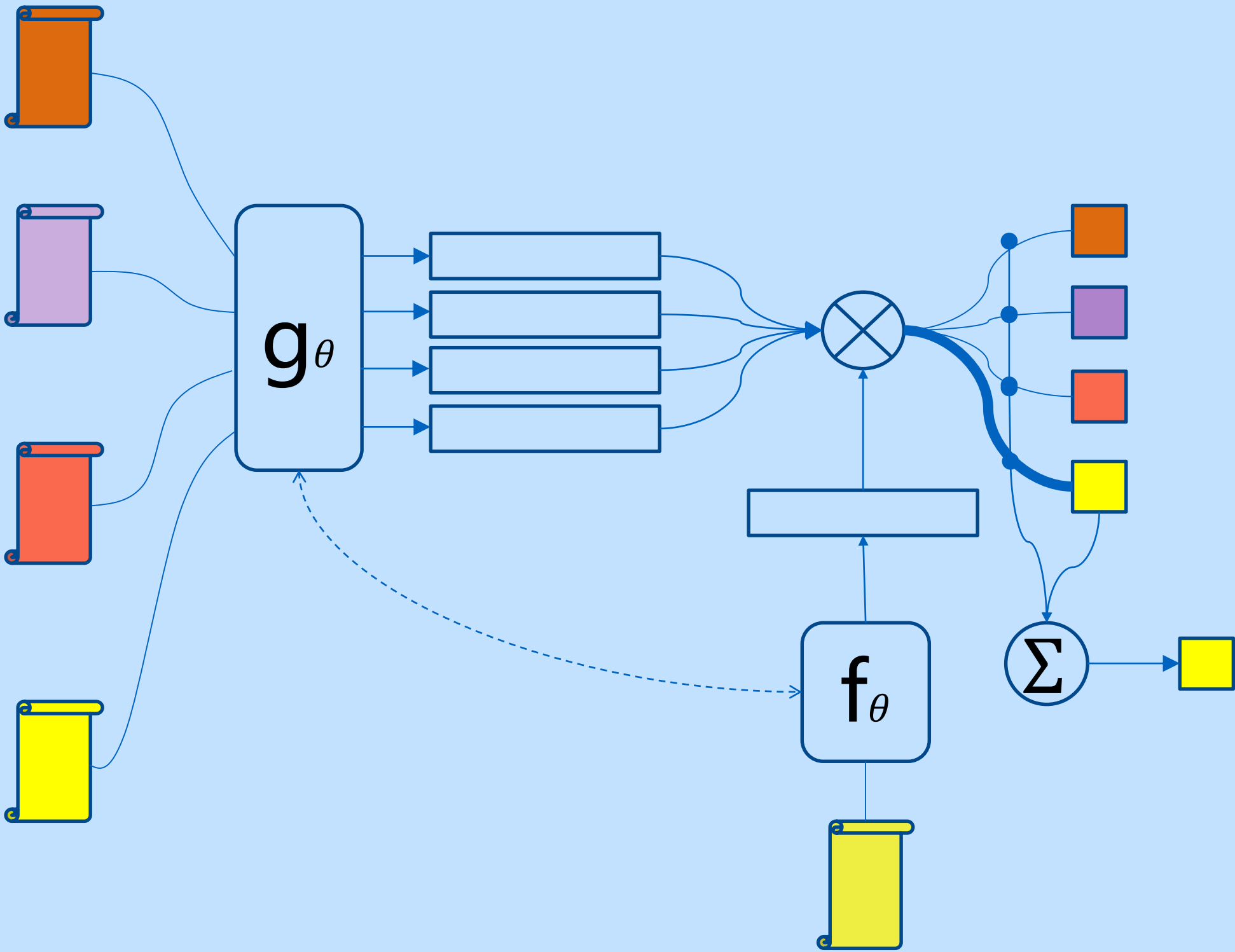
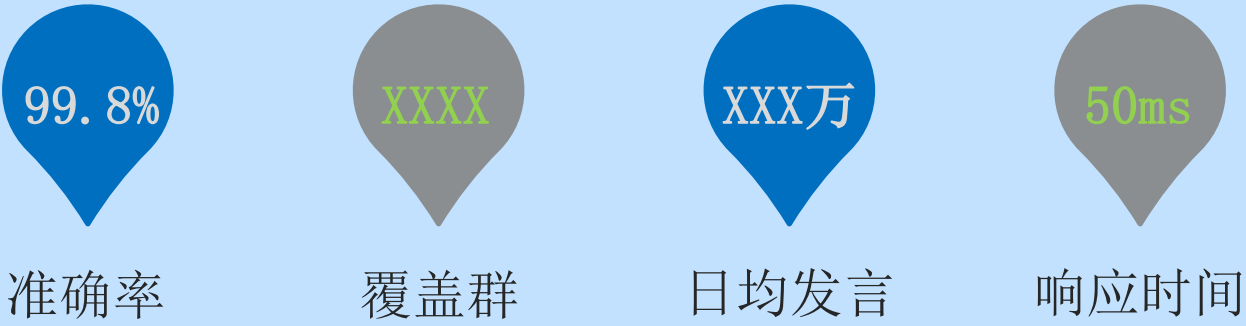
90%
准确率

XX%
转化率

XXX万
月均营收

XX
月均单量

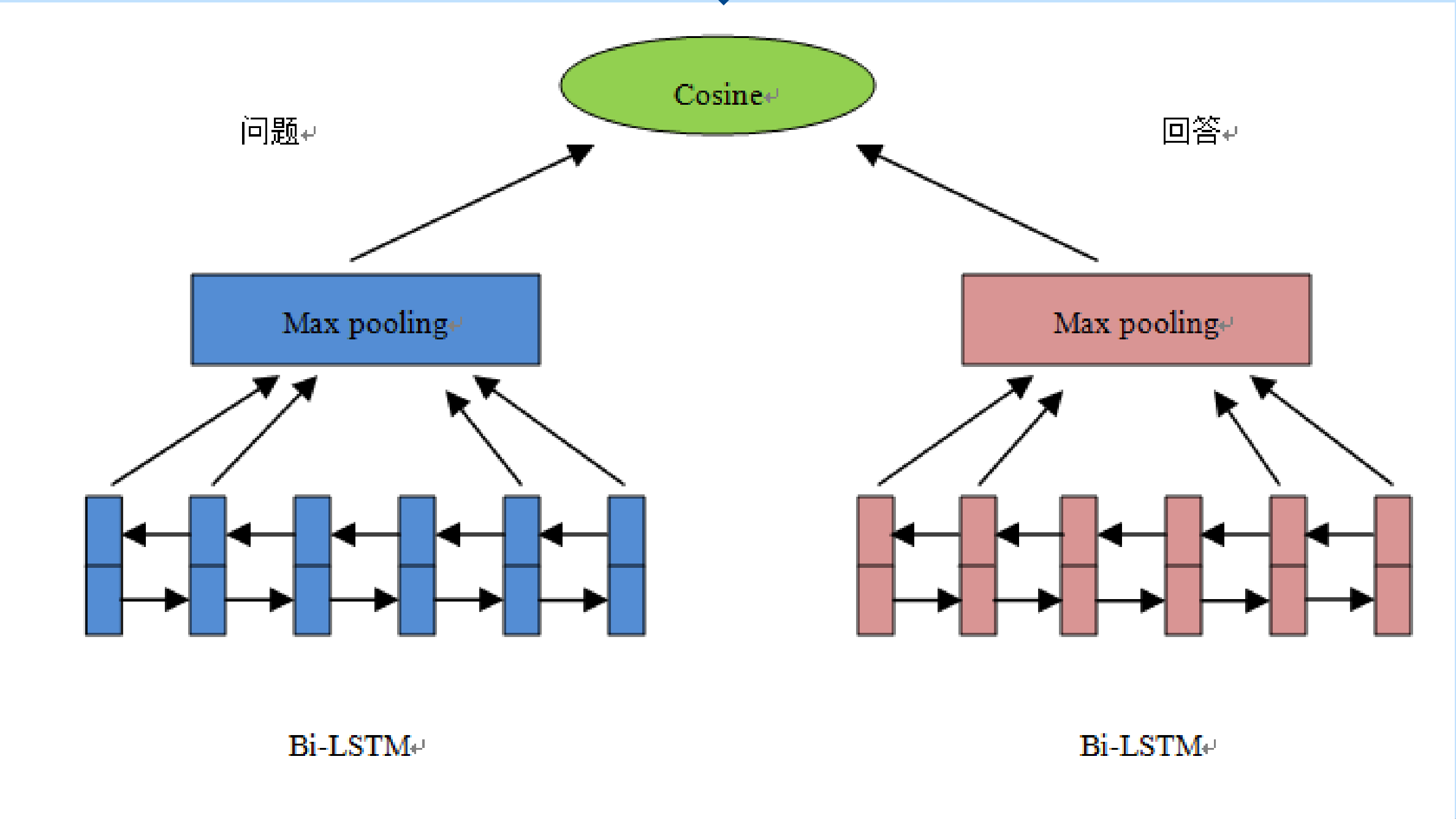
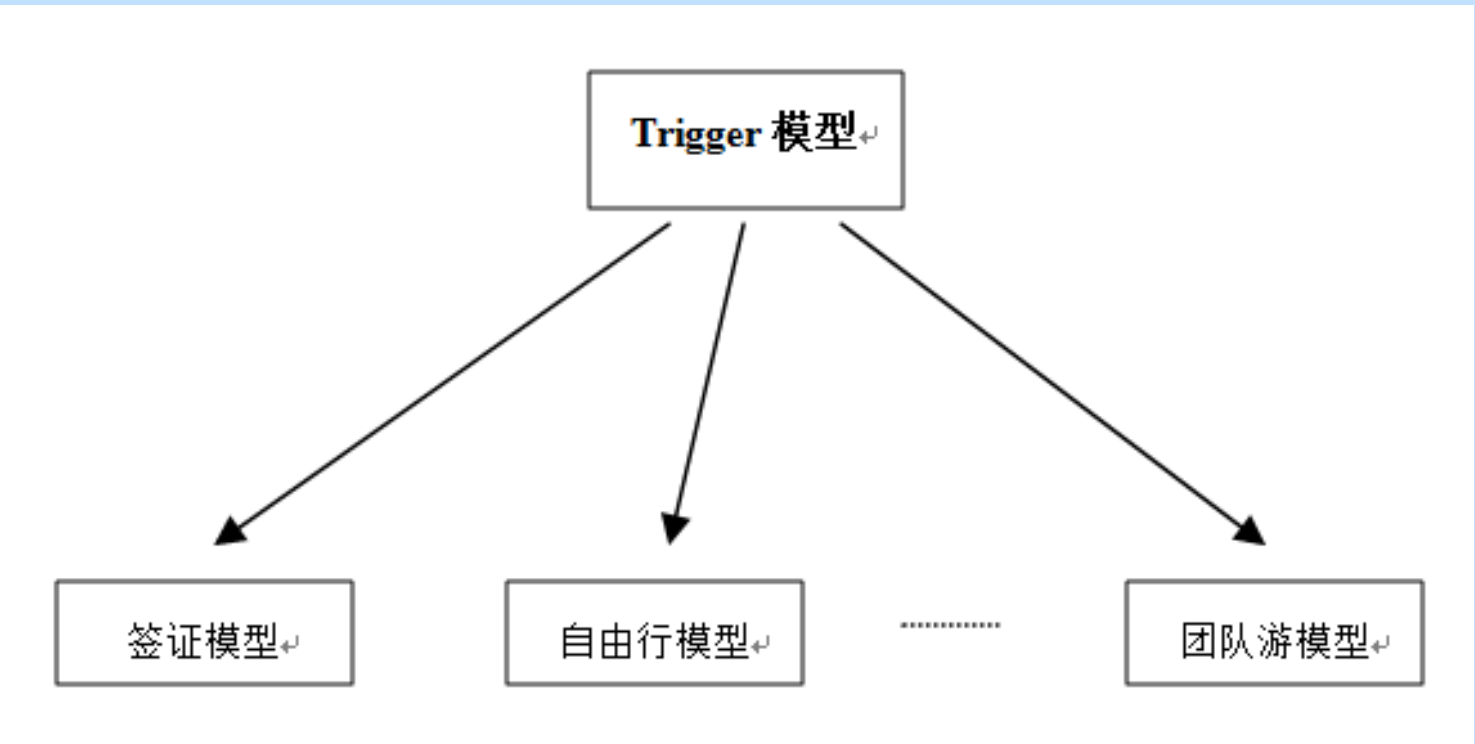
群监督



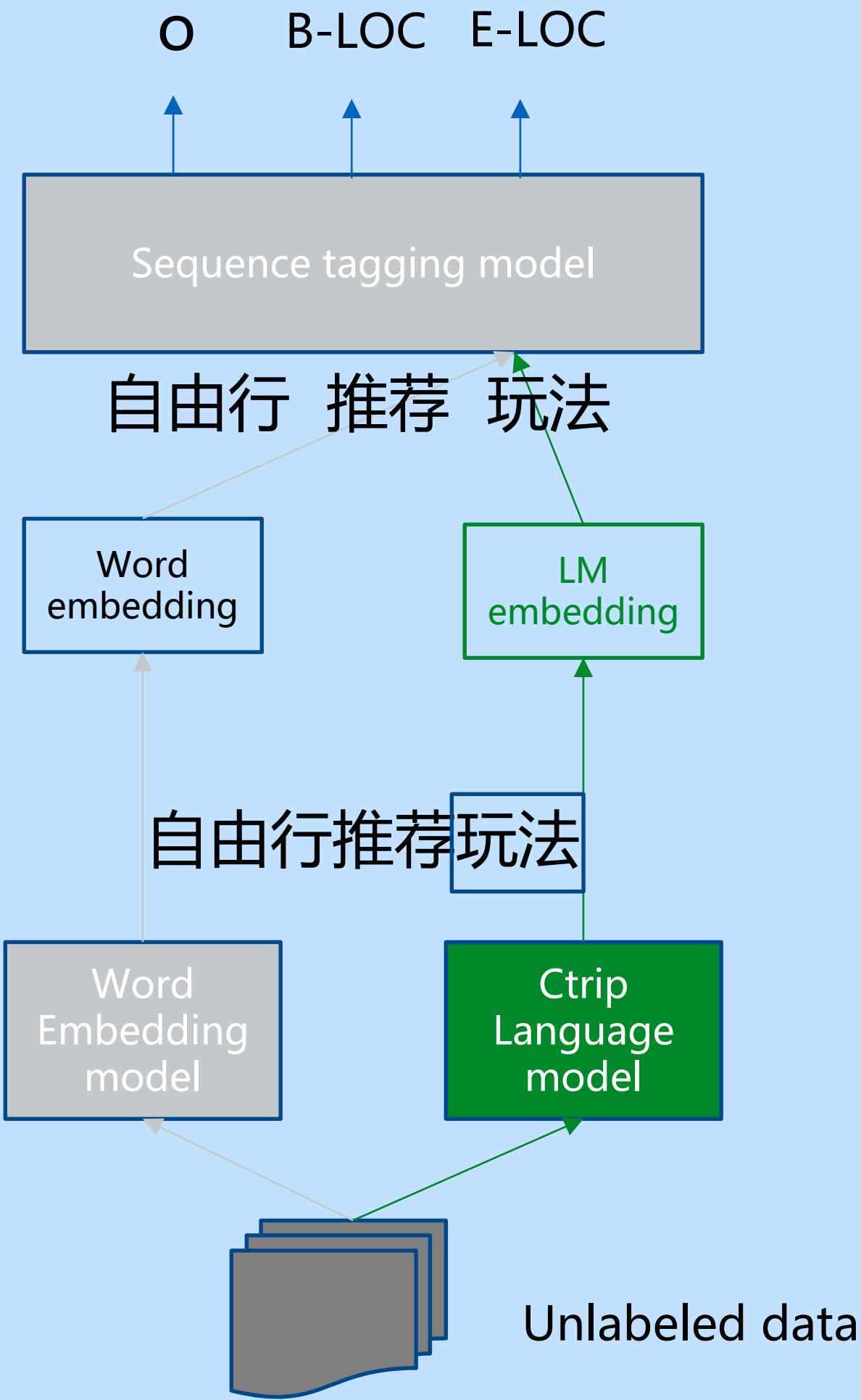
$$\theta = \arg \max_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x, y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

References
Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, Daan Wierstra. Matching Networks for One Shot Learning .2016.

自助问答1.0

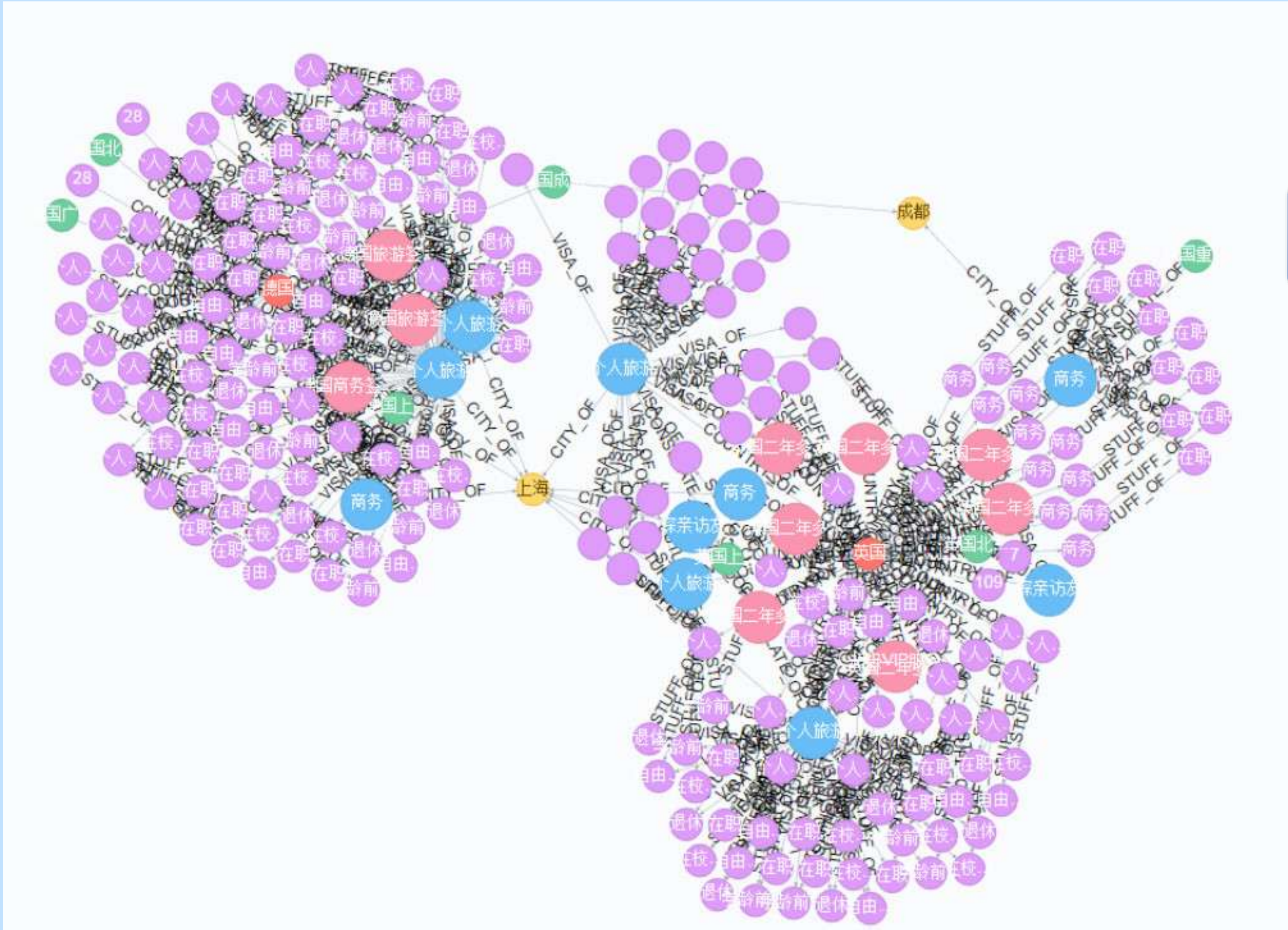


内容抽取 — 半监督



References
 Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, Russell Power. Semi-supervised sequence tagging with bidirectional language models. 2017

知识图谱 — 构建



Domain

Schema

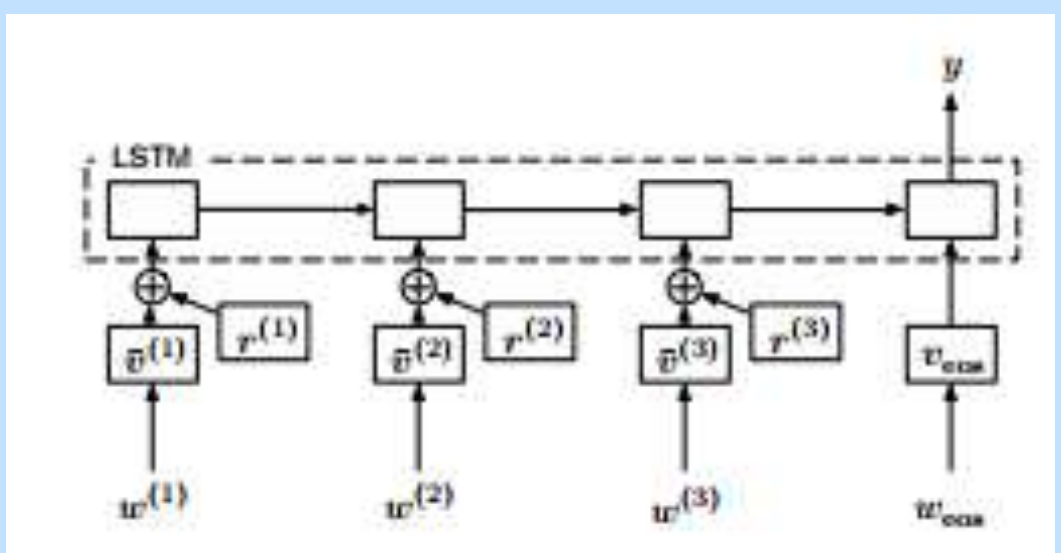
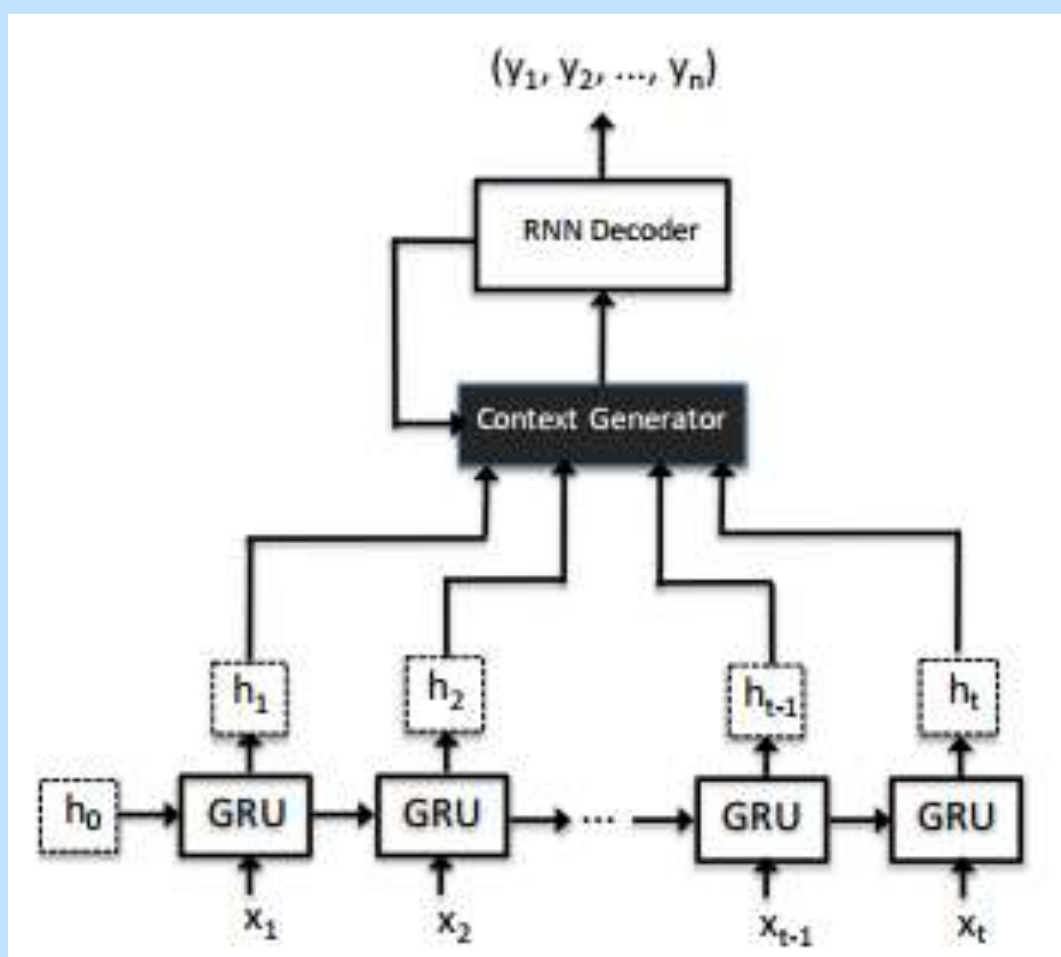
IE

Completion

Fusion

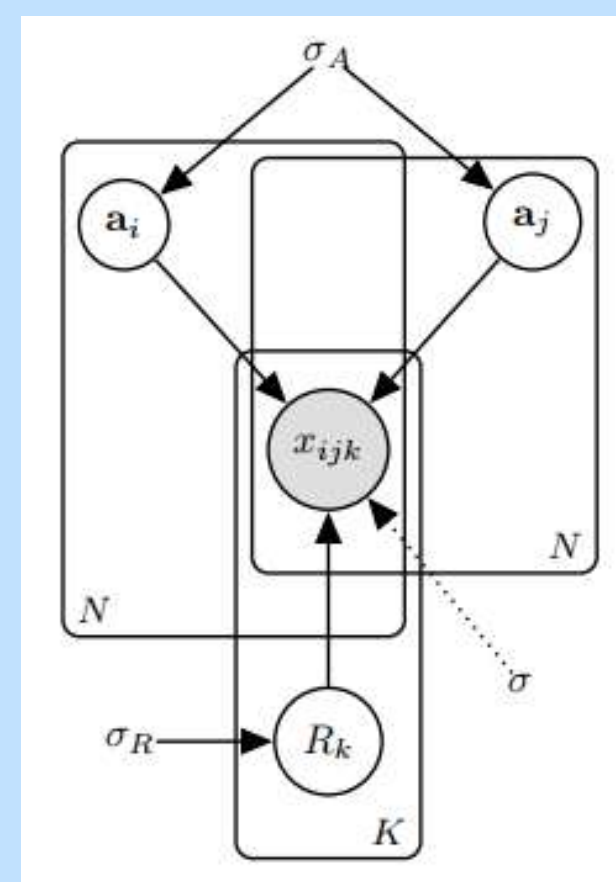
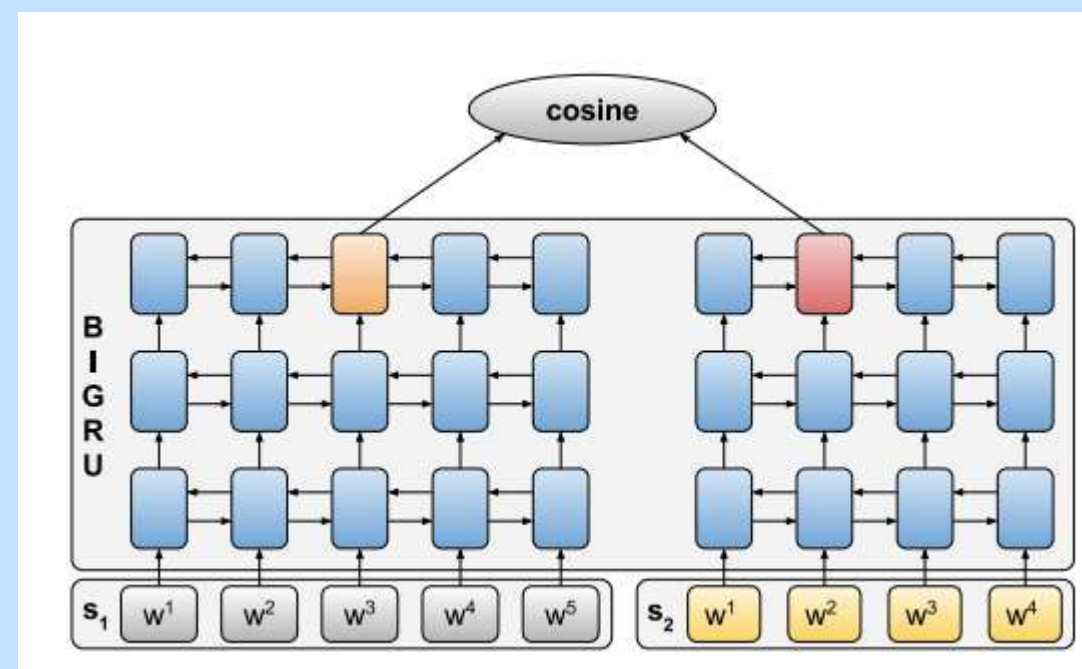
pipeline

Deep Learning Enabled Sentence Modeling



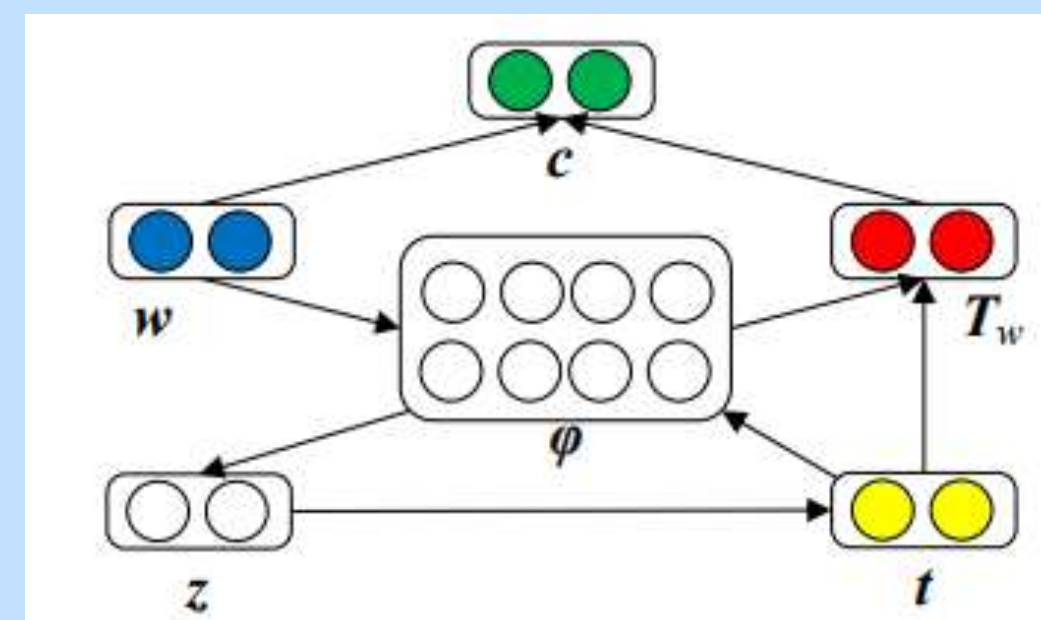
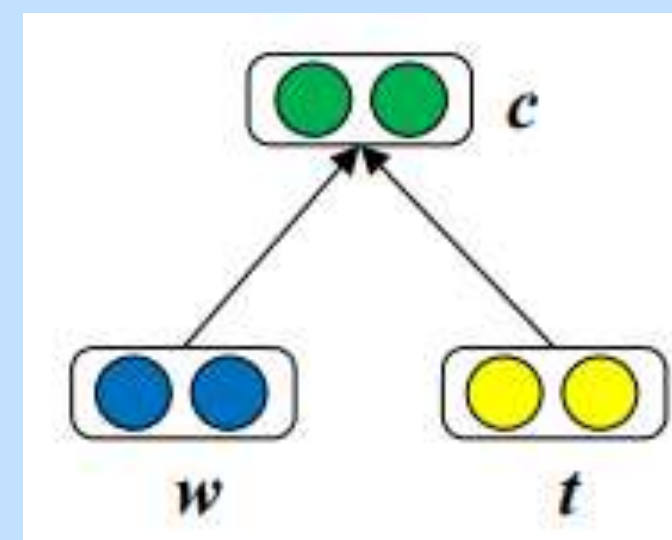
Text Summarization
Sentiment Analysis

Contextual Embeddings For KG Completion



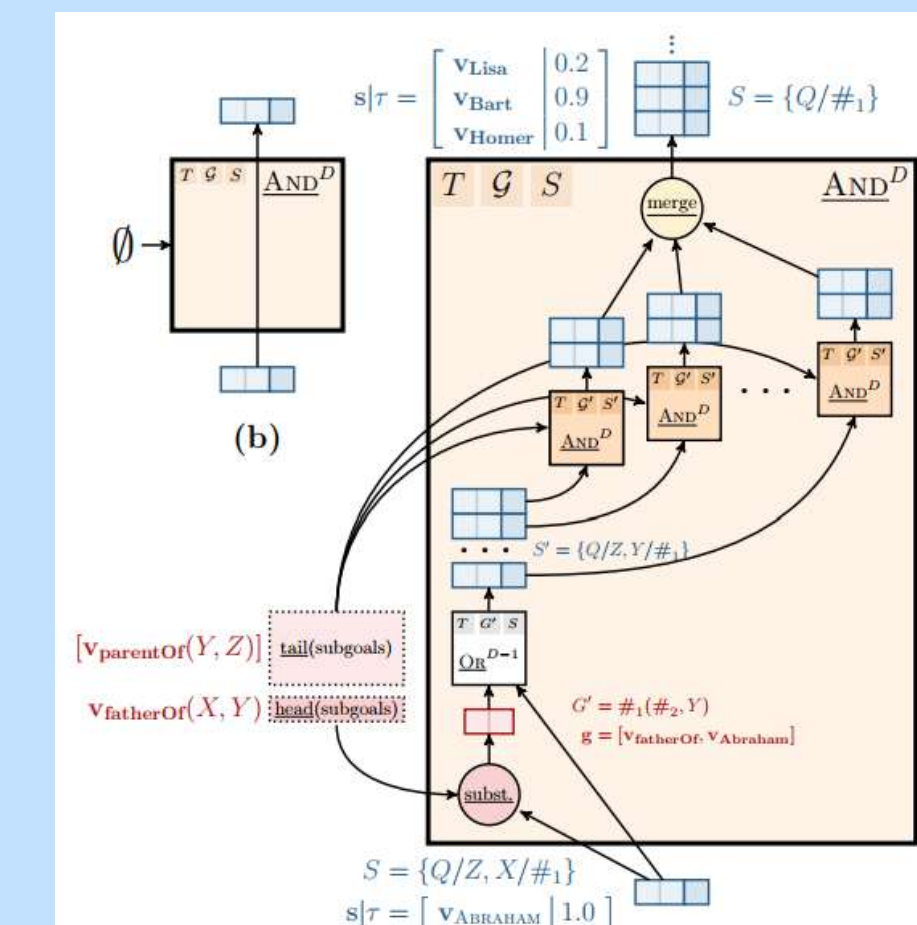
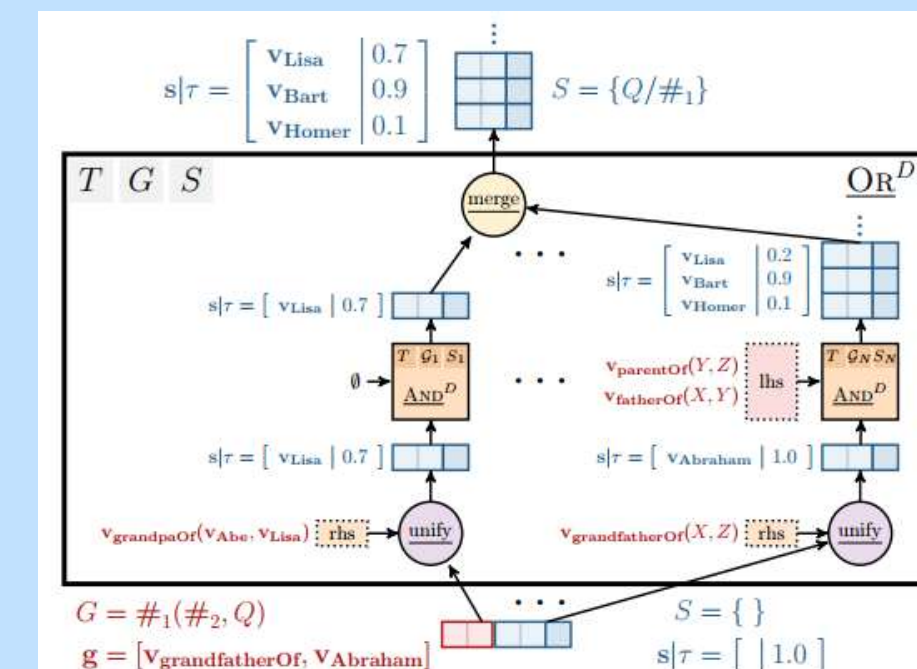
Entity & Relation Type
Prediction

Building KG Using Topic Modeling



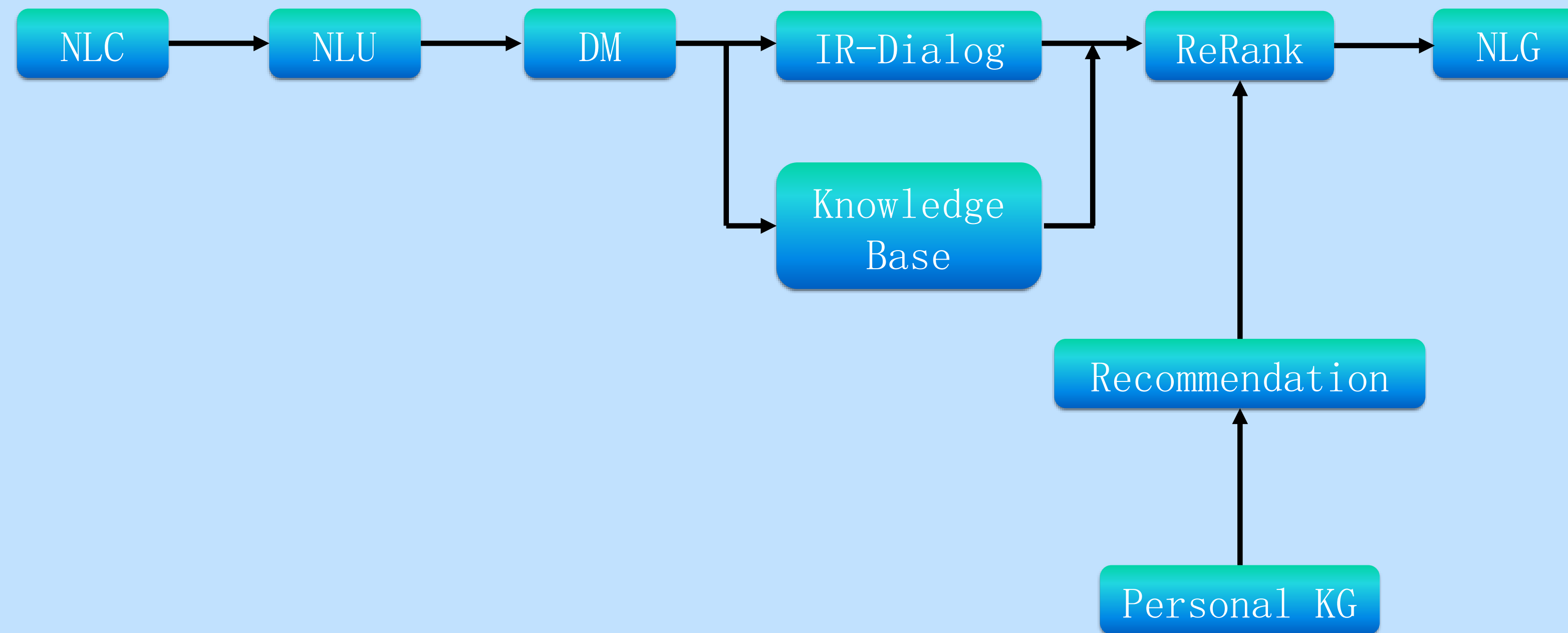
Large corpus reduction and
summarization

Scalable Inference for KG

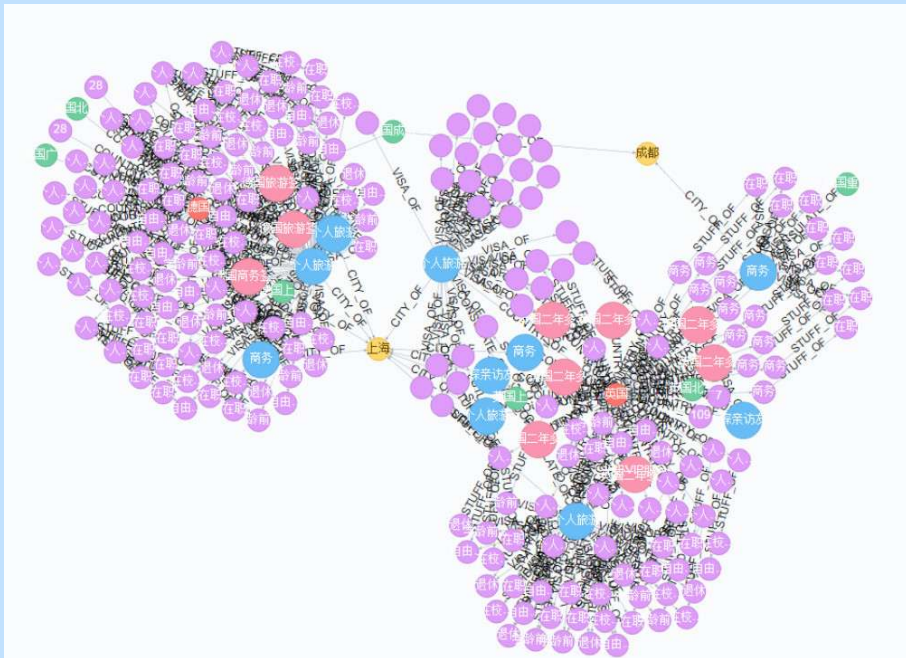


Scalable knowledge and relationship
discovery algorithm

自助问答2.0



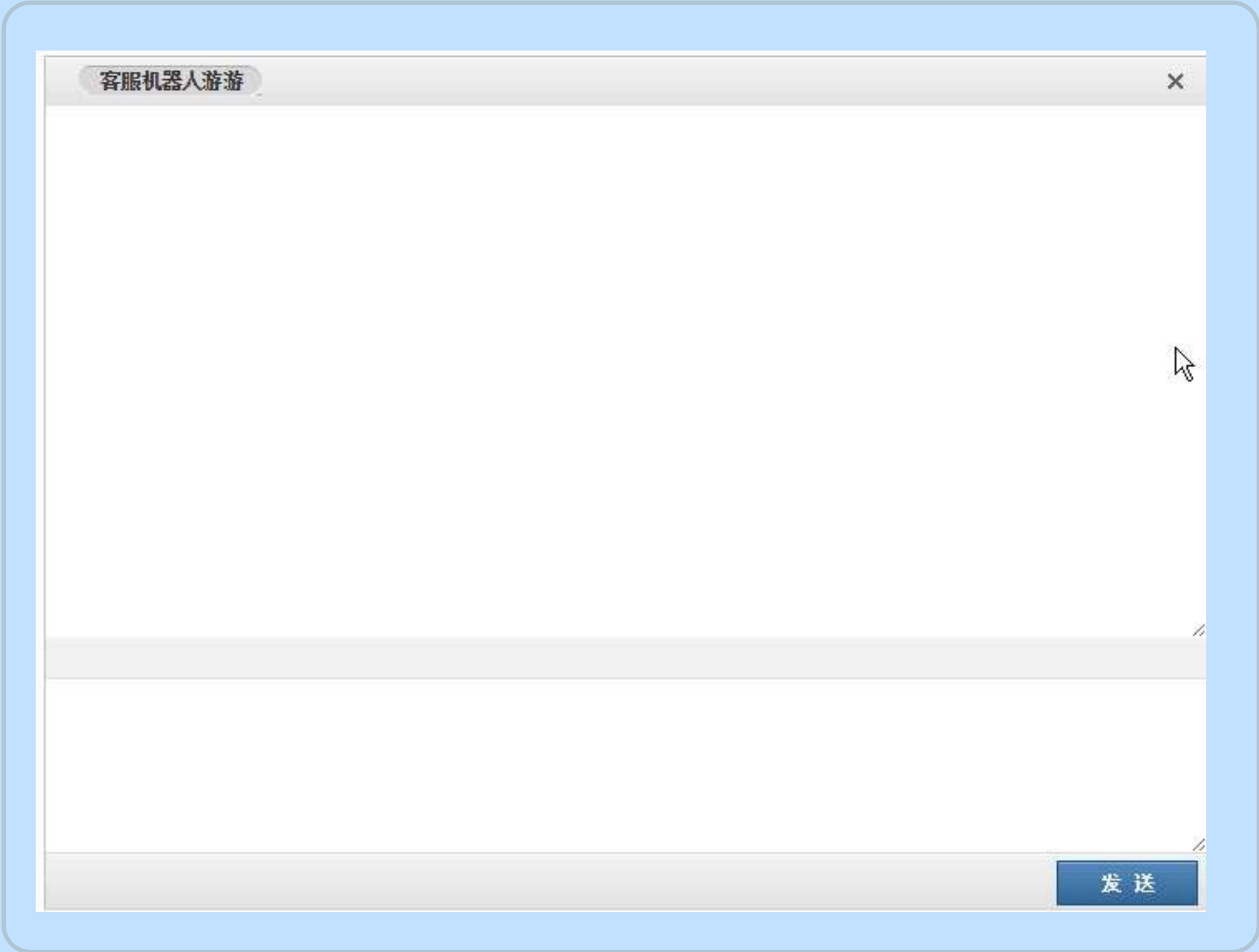
案例：自助问答2.0



知识图谱

多轮对话

细节问答



EasyAI平台 1.0

减少模型的重复建设

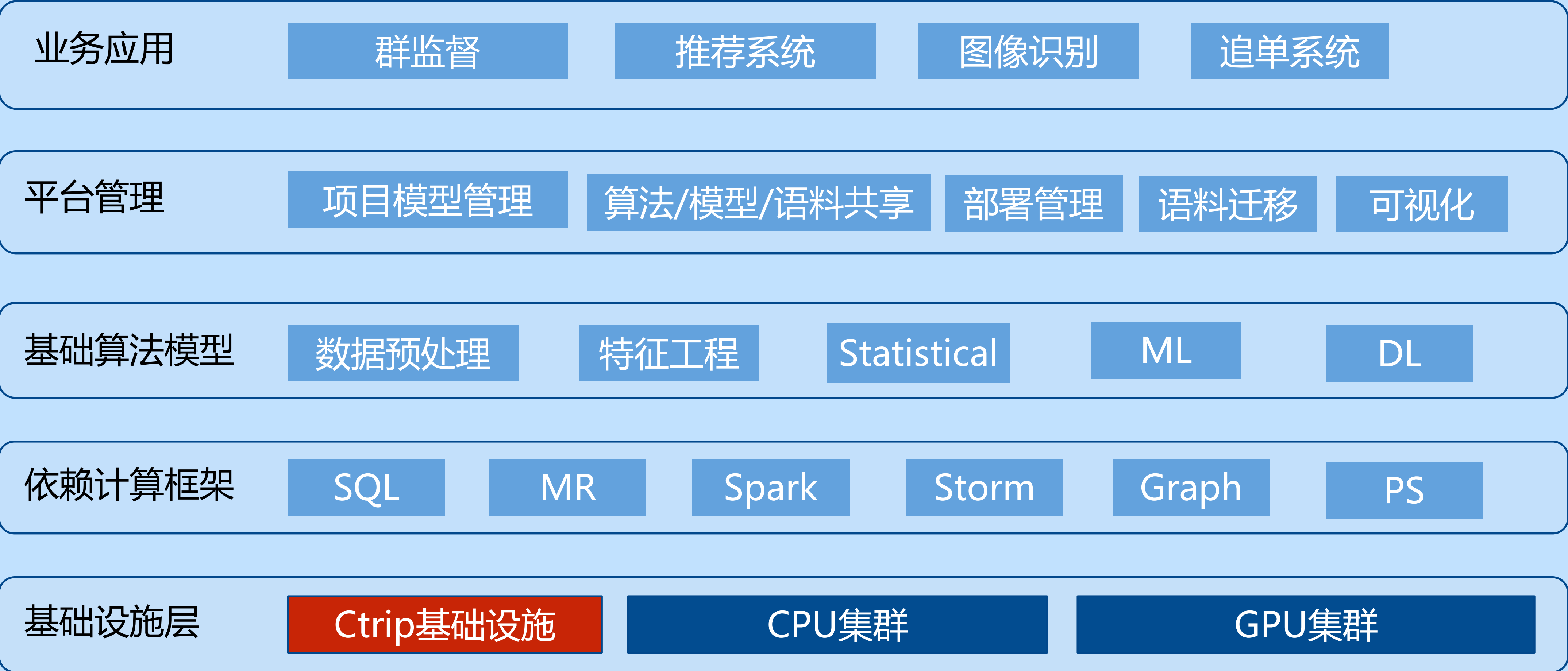
缩短建设周期

标注语料的共享

提高语料标注效率

提高模型训练部署效率

EasyAI 1.0系统架构



EasyAI

首页 用户管理 帮助 yangyz

携程技术中心

IT大咖说

输入模型名称或关键词

MAIN

Dashboard

新建任务

语料

设置

我的任务

TASK实例

客户出行意愿度识别模型

step1

step2

step3

step4

Loss

250

200

150

100

50

0

0

2000

1504

x: 22

Highcharts.com.cn

Evaluate

correct count = 9969, error count = 34, correct ratio = 0.996601

部署

运行

报告

任务属性

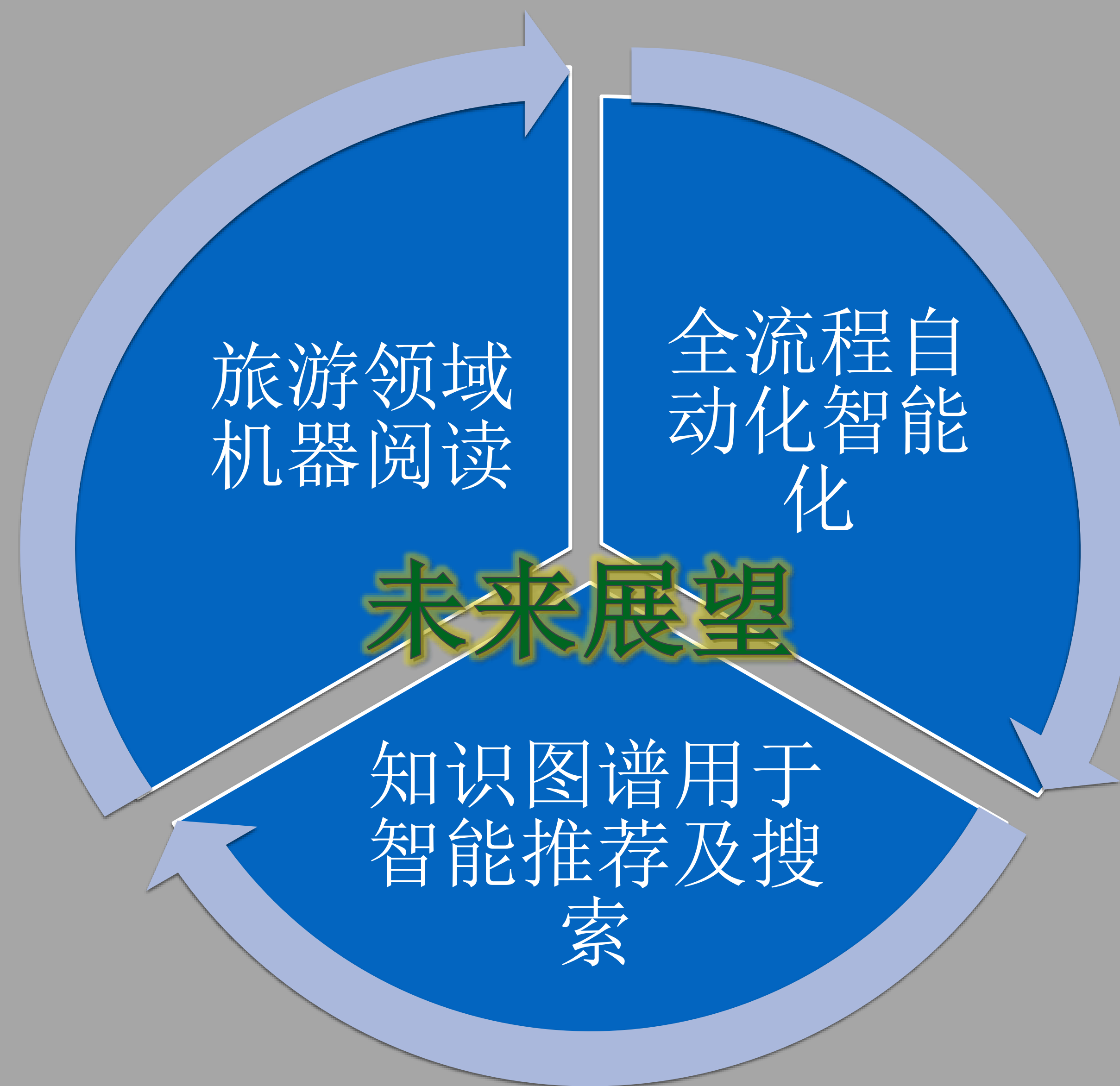
创建时间: 2017-08-12 14:30:21

名称: 客户出行意愿度识别模型

描述:

客户出行意愿度识别, 识别范围包括行程、日期、机票、酒店、证件等。

部署状态: 未部署



THANK YOU