

# 非结构化场景智能风控实践

曾利彬-算法专家-蚂蚁集团

DataFunSummit # 2023





# 01

## 业务背景介绍

DataFunSummit # 2023



# 业务背景

万里汇 WORLD FIRST



## 全球收款

### 跨境电商收款

选择平台

支持多平台多币种

- 覆盖全球 100+ 主流电商平台, 20+ 支付网关 (独立站)
- 全球常用币种轻松收: 美元、欧元、英镑、人民币、加元、新西兰元、新加坡元、日元、澳元、港币

### 外贸B2B收款

全球多币种收款账户

- 告别传统银行开户流程, 线上秒申请中国香港或新加坡全球多币种账户
- 全球 10+ 大常用币种轻松收国际电汇款项

多币种账户

账户名称 Mark Chen

开户地 HK

收款账号 83

币种

## 全球付款

### 多币种 多场景

- 丰富币种和地区覆盖: 支持付款至 41 个币种以及 200+ 国家和地区 (持续增加中)
- 灵活提现: 阳光结汇、极速提现/转账: 可付款至人民币账户、外币账户或支付宝账号
- 支持付款至供应商 (需提供证明材料)

灵活提现

提现币种: 美元 → 人民币

提现金额: 10,000 USD

提现至: 银行卡 / 支付宝 / 微信

交易记录

提现币种	提现金额	状态
美元	10,000.00 USD	成功
人民币	67,126.70 CNY	成功
港币	-855.47 HKD	成功
新加坡元	+4,851.33 SGD	成功
英镑	+13,008.89 GBP	成功

### 万里付卡

- 易开户, 随时用: 一键开户, 在线申领, 支持万里汇 9 个币种余额转入, 即转即用
- 多用途, 轻松付: 所有支持MasterCard的线上场景都能支付, 如平台店铺费用、广告投放费用、货品采购费用和物流仓储费用等
- 成本低, 权益优: 开卡退卡 0 费用, 转入手续费 1.2% 封顶, 无其他任何隐形收费

尊享返现福利

账户详情

卡号: 1234 5678 9010 1112

可用余额: 4,835.08 USD

万里付卡 WORLD FIRST

## 汇兑服务

### 实时汇率, 0 汇损

#### 汇率优势

- 实时汇率, 0 汇损, 24 小时实时汇率更新
- 全球覆盖, 实时汇率, 实时到账, 实时到账

所有汇率

币种	实时汇率	实时汇率
美元	100 USD	7.13 CNY
欧元	100 EUR	8.97 CNY

实时汇率

万里汇 (WorldFirst) 汇率

灵活收付, 更省心

多币种兑换

- 支持全球 10+ 主流货币实时兑换

多场景高效收付

- 支持全球 10+ 主流货币实时到账
- 支持全球 10+ 主流货币实时到账
- 支持全球 10+ 主流货币实时到账

## 业务风险点:

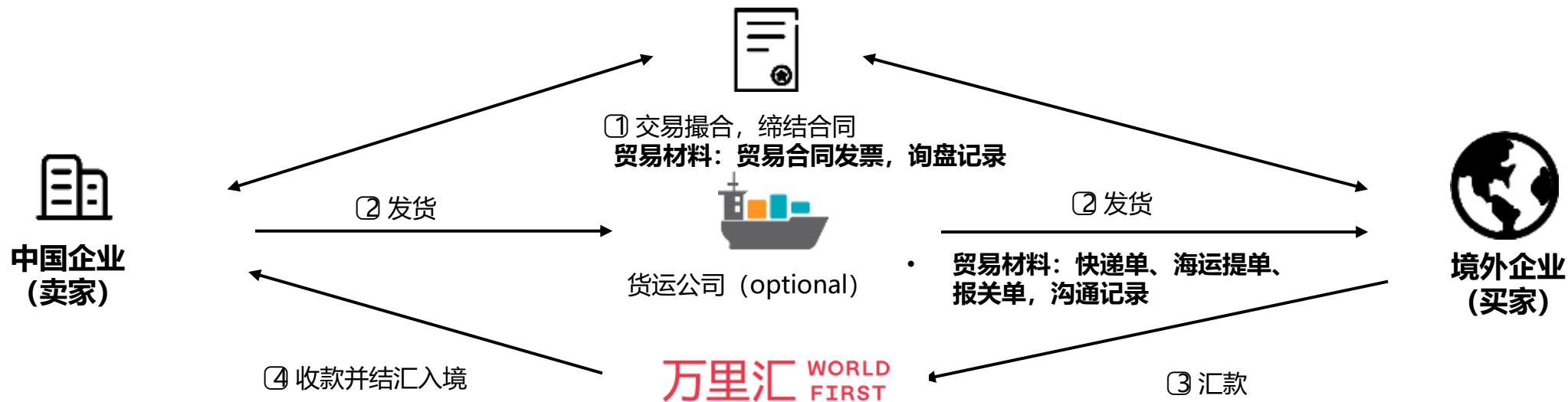
- 商户入驻真实性
- 贸易真实性

- 禁限售
- 信用欺诈

- 收单渠道盗用
- 账户资金安全

# 业务背景-B2B跨境贸易真实性

## B2B跨境贸易一般流程



## 主要风险点

主要风险点为**真实性风险**，非真实贸易收款可能带来合规风险或欺诈风险



# 业务背景-B2B跨境贸易真实性

## 真实性风险案例一

商户提供的材料无法证明其真实贸易背景，资金来源不明，例如提供了无法验证的跨境物流单据

### 货物签收证明

合同名称: XDY(HK) IL TRADE LIMITED  
 合同编号: 453745  
 发货方信息: XDY(HK) 收货方信息: MAI TA BUSINESS LTD  
 发货单位: XDY(HK) 收货单位: MAI TA BUSINESS LTD  
 发货地点: 广州 收货地点: 广州  
 发货时间: 2022.8.10 收货时间: 2022.8.10  
 联系人: 陆 联系人: MJ ITA BUSINESS LTD  
 联系电话: 13100 1017 联系电话: 89013 103

根据公司与 XDY(HK) IL TRADE LIMITED 于 2022 年 8 月 10 日签订的销售合同约定, 现于 XDY(HK) 有限公司已将货物送达双方约定的交货地址, 我公司已签收并接收了如下货物:

### 货物明细

序号	货物名称	型号规格	数量	重量	单价	金额	备注
1	SX663-1		7924		25	266825	
2	SX663-2		7924		25	266850	
3	SX663-3		7924		25	266850	
4	SX663-4		7925		25	266850	



# 业务背景-B2B跨境贸易真实性

## 真实性风险案例二

商户通过电信诈骗等方式骗取境外用户的资金，伪装成B2B跨境贸易背景跨境收款结汇，例如使用虚假的物流单据和网站

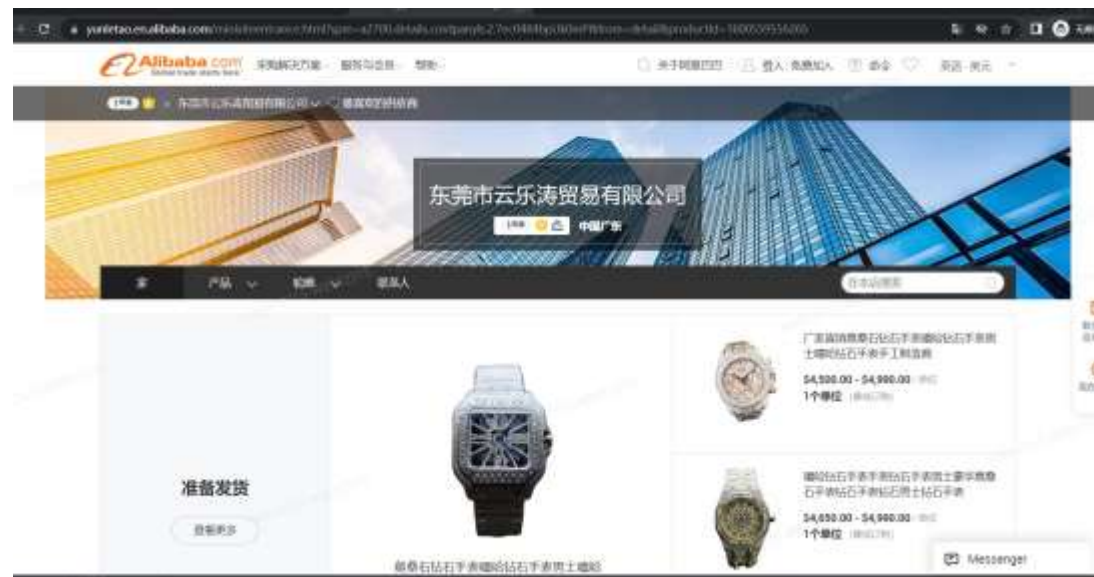
虚假的物流单据：

- 物流状态一直是label created，处于揽收状态，未有实际货物运输



网页主体不匹配：

- 客户提供了alibaba.com的商户主页，网页显示名称与商户报备企业名称不符



# 业务背景-B2B跨境贸易真实性

## 用户提交贸易背景证明材料



## 贸易真实性审核

### 贸易合同发票

跨境贸易合同，标明了买卖双方、商品、交货和付款方式信息



ITEM	PRODUCT/ PART NAME	QTY	UNIT	UNIT PRICE	TOTAL AMOUNT	DATE OF DELIVERY	PAYMENT	AMOUNT
1	2000 P20 Rapid build and parts Part	1	0	10,000.00 C	10,000.00 C	end of 2022	45 days at 100000	
2	2000 P21 Rapid build and parts Part	1	0	10,000.00 C	10,000.00 C	end of 2022	45 days at 100000	
3	2000 P22 Rapid build and parts Part	1	0	10,000.00 C	10,000.00 C	end of 2022	45 days at 100000	
4	2000 P23 Rapid build and parts Part	1	0	10,000.00 C	10,000.00 C	end of 2022	45 days at 100000	
TOTAL								40,000.00 C

### 跨境物流证明

国际快递、海运、空运、铁路运输单



EXPRESS WORLDWIDE WPX DHL

From: Dongguan, Automation Technology Co., Ltd. Origin: DGM

Allen, 708#, Taixin Road, Wanjiaojiao, 523000 DONGGUAN, CHINA, PEOPLES REPUBLIC

To: Parcop S.r.l., Via Le Croci n 20, 80070 NTE DI PROCIDA, ITALY

LEJE IT-NAP-VND

C-PLT

Content: 18pcs industrial inserter

Post/Ship Weight: 15.8/76.6 kg 1 / 6

### 其他材料

- 网址：售卖商品的网址
- 沟通记录：咨询沟通询盘等记录
- 委托证明：代理发货、委托付款等证明

- **贸易材料完整性审核**：资金、合同、货物三方面信息是否一致，例如收款人=合同卖方=发货人，付款人=合同买方=收货人
- **贸易材料真实性审核**：基于第三方数据验证、信息匹配程度等判断其真实性



# 02

## 算法技术介绍

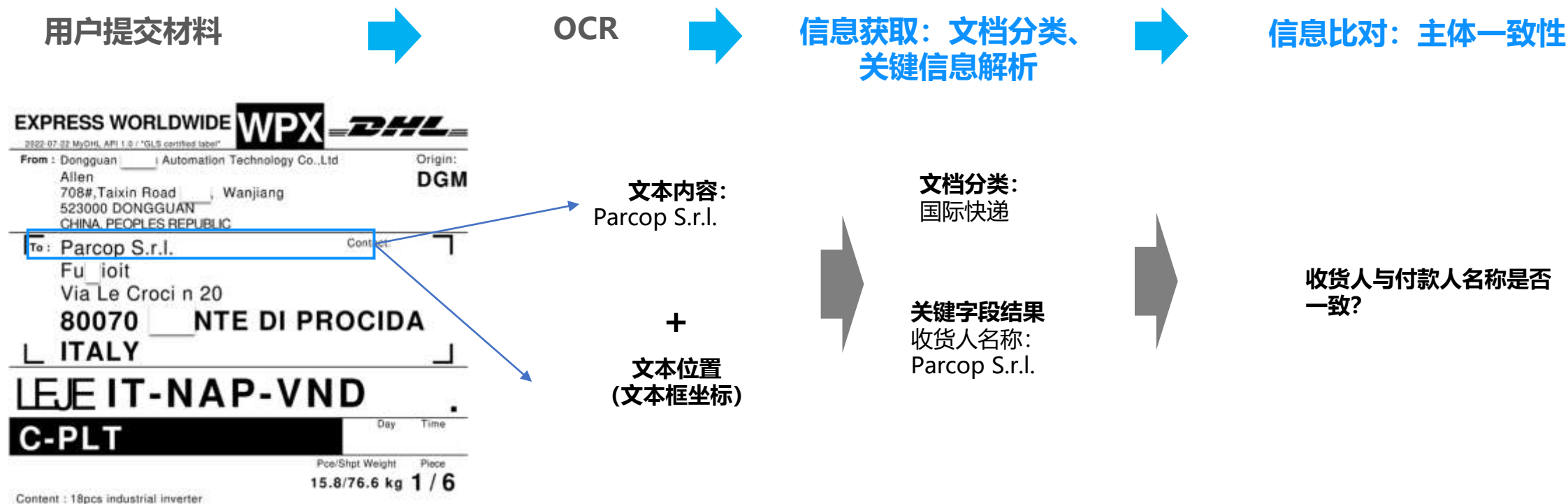
DataFunSummit # 2023





# 非结构化场景算法技术介绍

- **智能风控目标：**借助算法技术手段实现B2B跨境贸易验真场景的自动化验真，实现风险与成本的最优
- **场景特点：**非结构化数据场景，所有信息都需要从非结构化的贸易文档获取
- **算法技术链路：**



# 非结构化场景算法技术介绍

## 技术难点1: 非标文档如何获取信息

每个商家提交的文档风格都不一样，有图像多的，有文字多的，有KV型的，有非KV型的，文档理解是个难题



## 技术难点2: 企业多语种名称如何比对

企业的多语种名称（例如中英文）如何比对是个难题

### 拼音相似

中文名: 广州必一科技有限公司  
英文名: guangzhou biyi technology co., ltd.

### 读音相似

中文名: 苏州拓浦微实业有限公司  
英文名: suzhou top view industrial co.,ltd



# 非结构化场景算法技术介绍

非标文档中如何获取想要的信息？



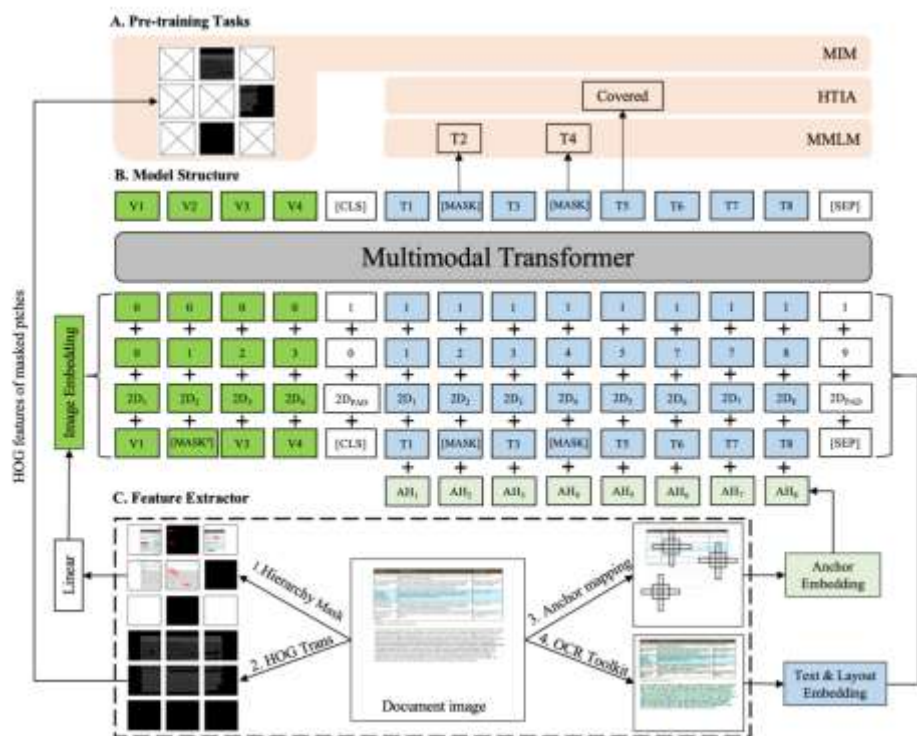
文本+周围图像推测：  
这是一个物流服务商

文本+周围文本推测：  
这是一个发货人名称

1.充分利用文本周围的图文信息  
多模态对齐：图像、文本、布局

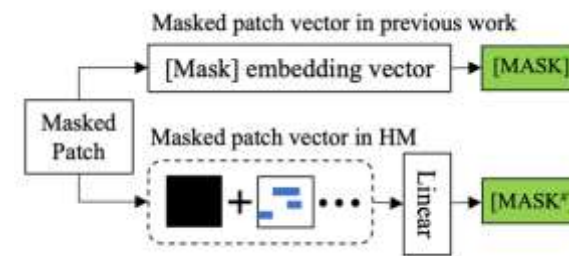
2.模型有足够的先验知识：  
大规模数据预训练

## VeLayoutLM：多模态对齐预训练模型



- **Highlight 1:** 层次化的mask方案，解决多模掩码难题

- **Highlight 2:** 多模对齐的预训练任务，充分利用图像、文本、布局信息



Mask Image Modeling: 通过周围图像预测掩盖图像

Multimodal Mask Language Modeling: 通过周围文本+布局预测掩盖文本，布局和对齐

Hierarchical mask Text-Image: 推测某段文本是否被mask，图像+布局与文本对齐

- **效果：**在公开数据集和业务数据集的多模文档分类、信息解析任务F1均超过90%

Model	公开数据集效果(F1 %)		
	CORD	FUNSD	SROIE
LayoutLM	-	79.27	94.38
LayoutLMv3	96.56	90.29	-
Ve-LayoutLM(ours)	<b>97.12</b>	<b>91.04</b>	<b>97.21</b>



# 非结构化场景算法技术介绍

## 企业的多语言名称对比如何做？

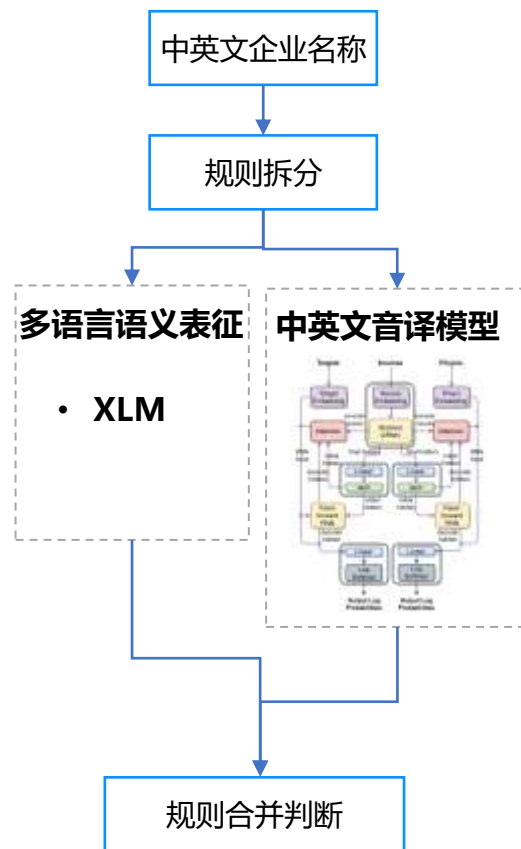
中文名: 广州必一科技有限公司  
英文名: guangzhou biyi technology co., ltd.

中文名: 苏州拓浦微实业有限公司  
英文名: suzhou top view industrial co., ltd

- 有的词是语义翻译
- 有的词是拼音
- 有的词是读音相似

### • 语义和语音结合比对

## 语义语音分别表征



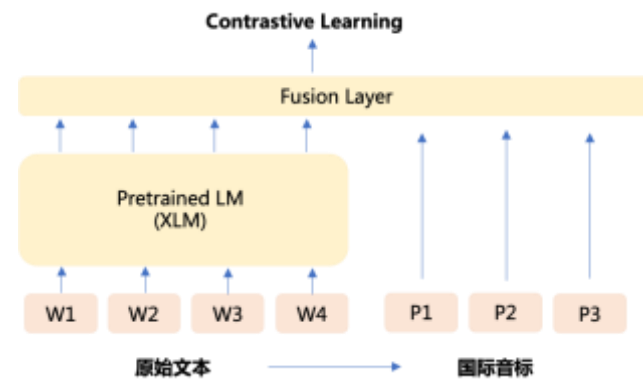
鲁棒性问题: 拆分与合并规则如何设计?

通用性问题: 如何推广到更多实体更多语言?

## 语义语音融合表征

### 语义语音融合的多语言表征

- 参考多模态模型结构, 把音标作为一个模态
- 使用对比学习方式拉近句子级别表征



### 语义+读音的数据增强方案

#### 传统文本增强:

- 同义词替换
- 词序替换
- 翻译替换

中文企业名称  
英文企业名称

#### 读音相似增强:

拼音相似检索  
模糊匹配检索

多语言名称库  
人名: NEWS 2018  
世界人名词典  
企业名称: 业务数据

使用语义语音融合表征+表征距离计算方案, 相比较于初版方案效果更优

# 03

## 智能风控实践

DataFunSummit # 2023

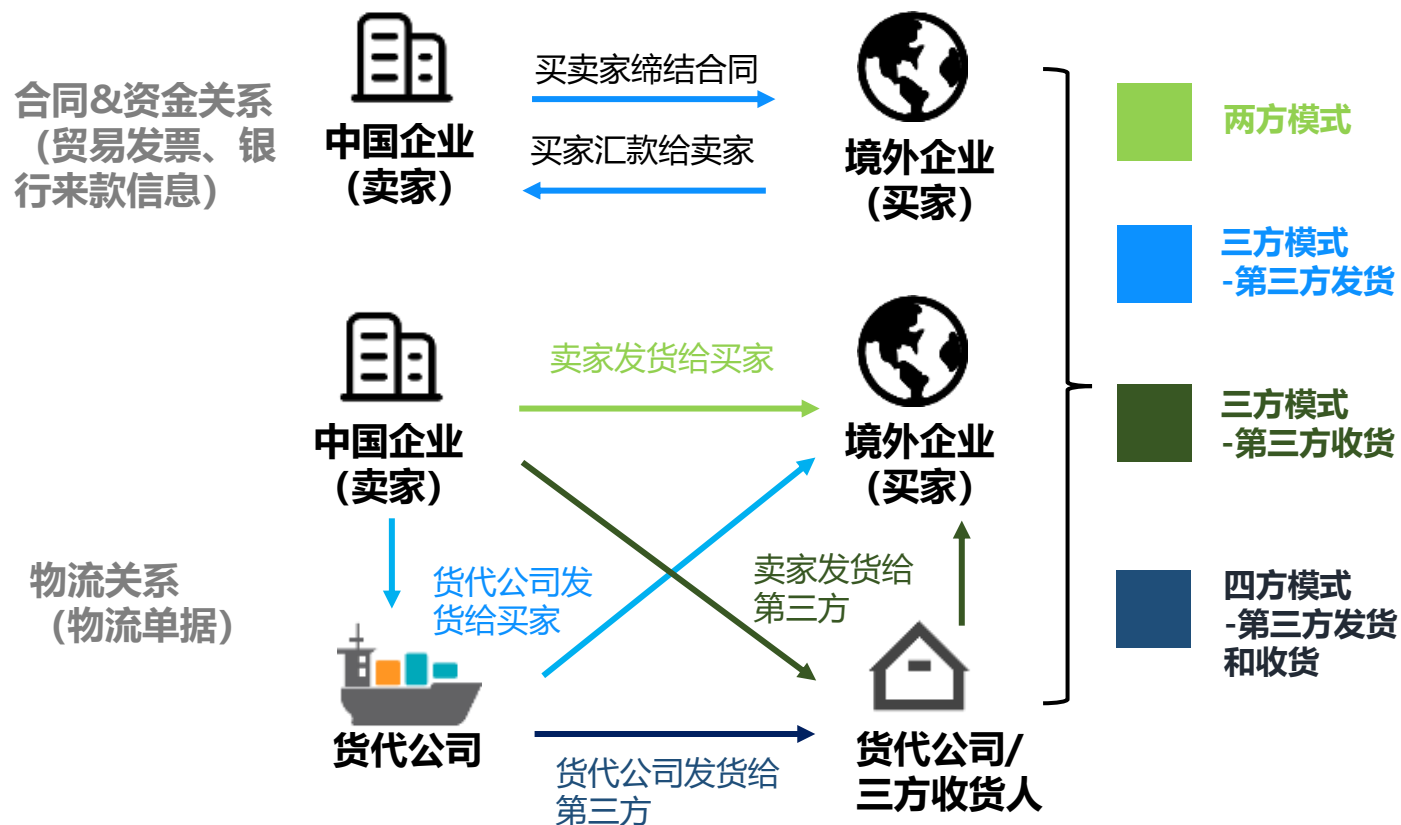


# 非结构化场景智能风控实践

## 业务解决方案-贸易材料完整性如何保障？

多方参与的B2B跨境贸易，天然存在例如发货人与卖家不同的情况，针对每种贸易模式设计针对性的解决方案

### 贸易模式分类



### 贸易模式还原

基于用户上传的贸易材料，匹配合同&资金关系、物流关系，判断其贸易模式

- ✓ 资金收款人 VS 合同卖家
  - ✓ 资金付款人 VS 合同买家
  - ✓ 资金付款人 VS 物流收货人
  - ✗ 资金收款人 VS 物流发货人
- 用户历史有材料证明使用了第三方发货模式

**贸易模式：第三方发货**

✓ **贸易材料完整**



# 非结构化场景智能风控实践

## 业务解决方案-贸易材料真实性如何保障？

三方数据验证和对文档进行篡改检测，提高造假成本

### 三方数据验证

可以进行三方数据验证的点-物流验证和网址验证

#### 物流履约验真

用户上传物流单



物流单分类模型

物流单解析模型

三方数据查询

- 物流状态
- 发货国家
- 收货国家

#### 网址验真

用户网址爬虫



网页元素提取

网页分类模型：  
是否合格网页

信息比对模型：  
名称是否匹配

### 文档篡改检测

适合非标文档的伪造识别方案，核心是“找重复”和“找不同”

#### 找重复：关键要素重复识别

贸易单据解析结果：  
快递单号  
发货人手机号  
发票号码

文本向量化

相似性检索

#### 找不同：文字风格差异识别

预训练backbone

Contrastive Learning

篡改检测：计算贸易单据上文本框与周围文本框的差异性

Positive: 同一张单据相近的文本框

Negative: 不同类型单据的文本框/  
样本生成

# 非结构化场景智能风控总结

## 业务场景实践方案

- **效果：**90%+外贸收款及结汇入中材料可由智能风控系统自动审核完成，相比较于人工审核在风险和成本上均大幅下降，并有效缩短外贸商家整个交易流程耗时

### 材料完整性验证： 贸易模式还原

- 贸易合同发票、物流单OCR和智能文档解析
- 根据文档解析结果还原贸易模式闭环，验证材料完整性

### 物流验真

- 物流单OCR和智能文档解析
- 根据解析物流服务商和单号进行外部查询验真

### 网址验真

- 网页内容智能解析
- 网址与商户名称匹配验证

### 文档篡改检测

- 非标文档篡改检测
- 重复性检测

**用户提交材料：**贸易合同发票、物流单、网址

**资金信息：**资金收付款人、金额

## 算法能力

### 非标文档信息获取

- **文档OCR**
- **贸易文档智能解析**  
商户营业执照/个人证件海关报关单/合同发票/快递单/海运提单/网页截图

### 信息比对

- **主体名称比对**
- **行业比对**

### 外部数据查询

- **工商数据查询**
- **物流数据查询**  
快递/海运/空运/报关单
- **网址查询验证**

### 文档风险识别

- **非标文档篡改识别**
- **重复性检测**
- **黑图库检索**





# 感谢观看