

如何让配送地址更准确—— 带你走进京东物流大数据算法测试探索之路

主讲人：樊宇

京东物流资深测试开发工程师

一个带配送地址的文本：S1 S2 S3 S4……Sn

S2 一级地址

S5 四级地址

S3 二级地址

S6 姓名

S4 三级地址

S7 电话

目录

Contents

- 01 技术背景
- 02 算法模型的建立
- 03 测试数据的获取
- 04 算法测试执行过程
- 05 算法测试方法论
- 06 未来和展望

01

章 PART

技术背景

京东快递的微信小程序——自动提取关键字算法：

从文本中提取地址信息、姓名、电话的功能。

方便用户复制贴粘配送地址信息，减少人工输入的过程，提高用户体验。

地址自动识别

中国移动 53 B/s 49% 傍晚6:50

< 填写寄件人

请粘贴或者输入文本，点击“识别”完成智能填写，文本如：北京市大兴区科创十一街京东集团总部，京小东，138****0000

识别

寄件人信息 清空

姓名	请添加姓名	
电话	请添加电话	
省市	请选择	
详细地址	请添加详细地址	

设为默认寄件地址

确定

02

章 PART

算法模型的建立

准确率

召回率

概率

余弦相似度

分子??

分母??

自然语言处理

NLP

LBS

地理位置服务

地址

地址编码

坐标

召回率

距离

计算匹配范围
内转换错误的
地址数量

计算正确转换
的地址数量

阈值

03

章 PART

测试数据的获取



如何对**新的地址**进行地址转换正确S的预估判断？

新的配送区域

配送区域内新增的地址

极坐标表示法

地址聚类

簇的质点为圆心

生成极坐标上的点(距离,角度)

逆地址编码生成地址

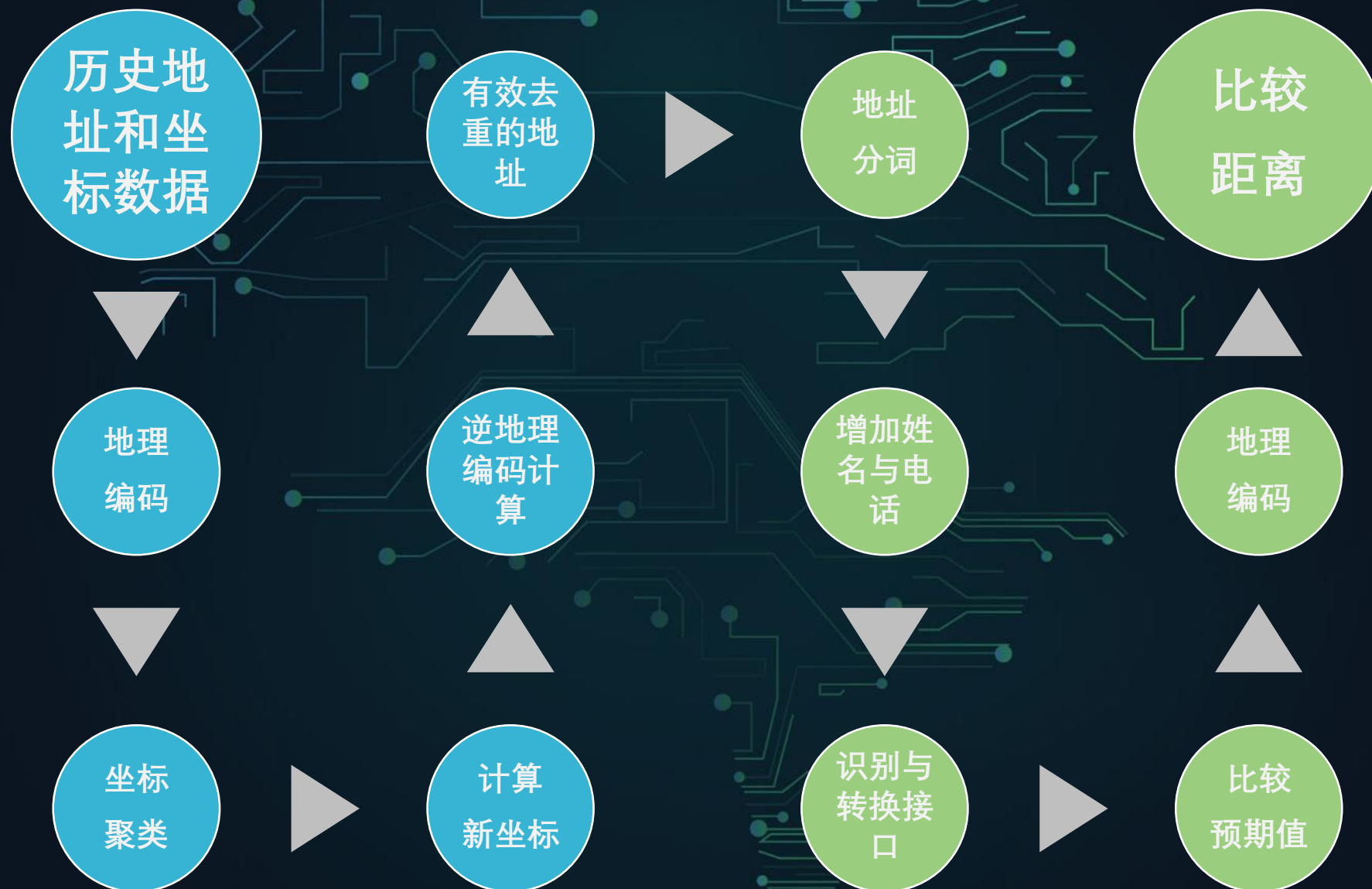
过滤得到可配送地址

簇内可配送地址去重

04

章 PART

算法测试执行过程



经纬度坐标聚类的算法

取配送站坐标为初始质点

计算质点半径范围内的坐标划簇

如果半径内有其它配送站坐标合并两个簇，并重新计算质点

如果簇内元素数量超过预定阈值后终止该簇的聚类

按簇中元素数量从大到排序取前N个簇

按质点生成极坐标

地址识别转换算法

如何提供类似**缺陷列表**的内容？

取历史坐标中最
多的前N个坐标对
应的地址



取距离相距最
大的前M个坐标
对应的地址



在前N个地址中
存在的前M个地
址

生成入参

如何增加**噪点数据**的内容？

北京市丰台区樊羊路XX号

北京市/丰台区/樊羊路/XX号

北京市/丰台区/樊羊路/XX号/樊宇/手机号

北京市/丰台区/樊羊路/樊宇/手机号/XX号

北京/丰台/樊羊路/XX小区x号楼/樊宇/手机号

如何**加速执行**测试数据生成和测试执行的过程？



05

章 PART

算法测试方法论



数学基础

算法

自动化接
口测试

并行计算

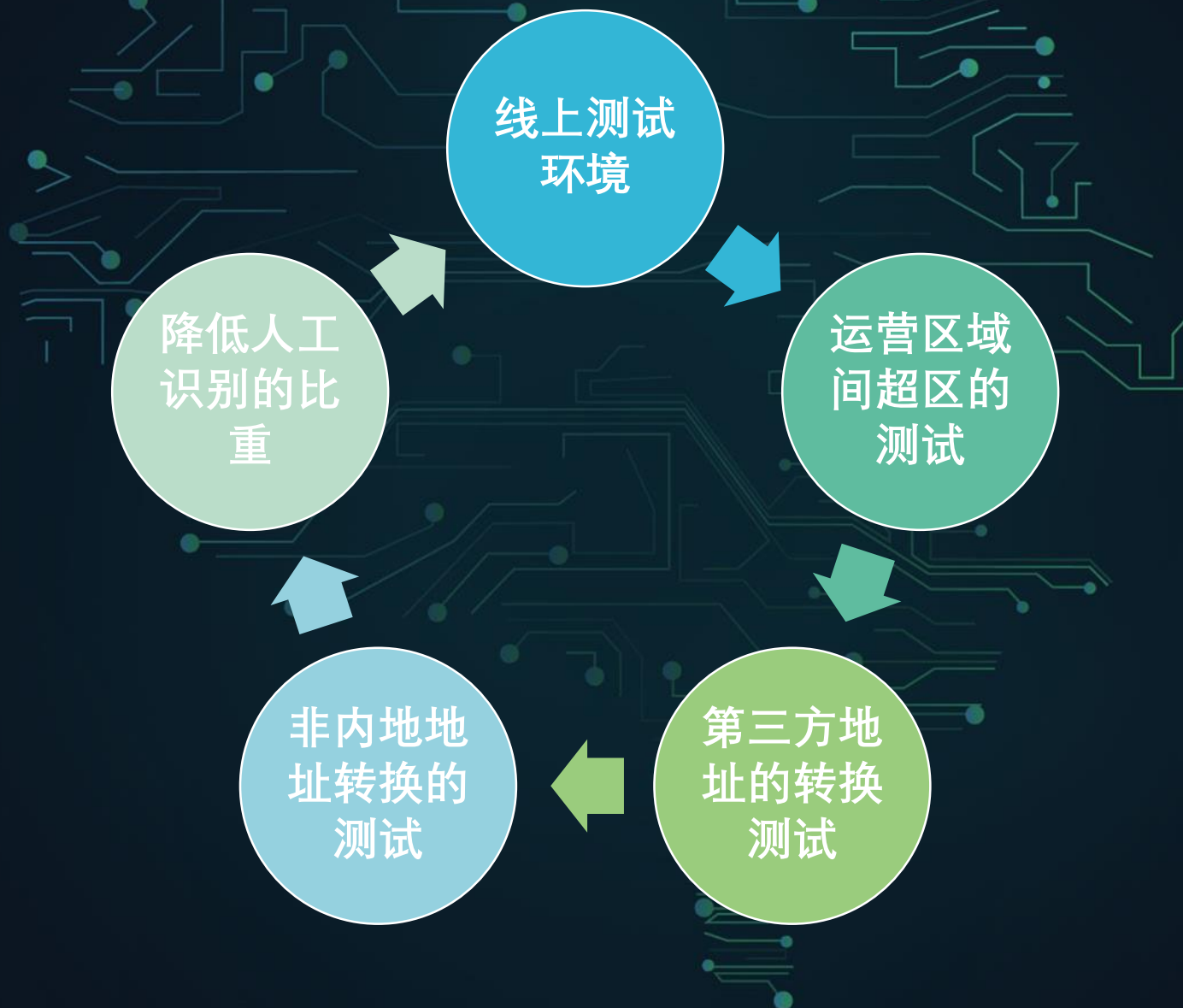
数据挖掘

业务场景



章 PART

未来和展望



京东物流测试技术团队

诚招性能测试、基础架构测试方向、移动测试方向

简历邮箱：weiaisheng@126.com



An abstract graphic of a circuit board with glowing green lines and dots, forming a shape that resembles a map of China, serving as a background for the central text.

THANK YOU

感谢聆听