|  |  |
| --- | --- |
| **赛题名称:基于WIFI探针的商业大数据分析技术**  **组类（ A ）： A 本科及以上** | |
| 赛题简介：介绍整个赛题的思路和整体要求 | **思路：**       探针的原理是利用智能设备商WIFI模块所发出的无线广播信号进行设备的感知，就像是网站上的Cookie，他会记录你的访问行为和轨迹。不同的是，通过手机MAC地址所采集的是你的线下行为轨迹，比如喜欢逛什么地方，一周逛几次。相同的是，他都无法直接记录你的个人信息，例如你是谁，手机号多少，这都无法直接获得。  通过对采集MAC地址数据的分析与统计，可以把握门店的客流情况，精准监控客流质量，实时展示客流转化情况，从而帮助检测营销效果，发现潜在机会和改进措施，为便捷、高效精细化运营提供全方位数据参考。  **整体要求：**      基于WIFI探针的商业大数据分析技术，通过WIFI探针收集MAC地址、出现时间、出现地点、与探针距离，探针设备会定时(3s)发送数据到**服务端**，通过一段时间数据平台产生大量的用户数据，然后使用大数据分析技术，采用**离线计算和实时计算技术**相结合的方式，能够对商业环境中门店的门前人流量、进店客流量、进入跳出量、新老顾客数量及新增量、在店平均时长进行分析，客流情况更新迅速，便于实时查看线下人群密集程度，有利于总部统一管理并及时调整营销活动方案。新老顾客比例，来访周期，顾客活跃度清晰展示，便于及时掌握入场顾客的构成，为销售策略调整提供参考。通过驻店时长可以快速掌握店招、产品或营销活动的吸引力。 |
| 赛题业务场景：描述赛题相关的真实企业业务背景。从真实场景中，适当简化或者提炼出适合比赛的赛题场景 | 利用探针数据的客流分析打破模式束缚，不仅仅只是提供可信的客流数据分析，同时还利用延伸的标杆管理才能，深刻洞悉并提供有助于推动实际客流量和消费者习惯行为的一系列因素。这种专业才能呈现了经济分析，社交和环境等一些超出你控制范围的因素，却对商业绩效产生主要的冲击力。  首先开发探针设备能够采集客户唯一的定位标识，比如MAC地址，通过数据分析技术，采用离线计算和实时计算结合的方式，为商业环境提供科学的、全面的数据决策依据。不仅对营销能力的评估，也可以对管理上进行优化。 |
| 功能性需求 | 主要是实现三个方面的功能：一是通过探针设备采集可监测范围内的手机MAC地址、地理位置、与探针距离、时间等信息；二是探针采集的数据可以定时发送到服务端保存；三是利用大数据技术对数据进行人流量等指标的分析。系统应具备以下功能：  **（1）WIFI探针（基础分：25）**  探针设备可以在其他渠道（淘宝）购买，也可以自行开发硬件和固件，探针设备可以进行服务端的相关配置（服务端IP、端口、路径、发送数据时间间隔），能够采集MAC地址、地理信息、与探针大概距离、采集时间等信息（不允许采集个人隐私信息，比如手机号等信息），特别针对ANDROID6.0和IOS10版本后的移动终端设备进行测试能否采集到MAC地址。自行开发探针硬件附加10分，自行开发固件附加5分。  **（2）数据采集（基础分：20）**  服务端主要接收探针定时发送的数据，将数据保存到数据分析平台待用，文件系统可以使用HDFS或者其他适合的分布式文件系统。接收数据不能有数据丢失或者数据失真，探针每3秒发送一次数据，数据采集并发量不得低于1000台设备，小于1000扣10分，大于1300附加10分。采集的JSON数据结构样例如下：  {          "id":   "0010f377",    //嗅探器设备id          "mmac": "5e:cf:7f:10:f3:77", //嗅探器设备自身WiFi mac          "rate": "1", //发送频率          "wssid":  "kaituo", //嗅探器设备连接的WIFI的ssid          "wmac": "a8:57:4e:c0:d4:8c", //嗅探器设备连接的WIFI的mac地址        "time": "Sat Jun 04 22:45:28 2016",//时间戳，采集到这些mac的时间        "lat":    "30.748093",  //北半球，纬度        "lon":    "103.973083",        //经度        "addr":    "江苏省南京市玄武大道699-22号", //地址信息      "data": [{                          "mac":  "9a:21:6a:7b:62:6a", //采集到的手机mac地址                          "rssi": "-30",//rssi，手机的信号强度，如rssi=-75dbm  "range": "1.0",//手机距离嗅探器的测距距离字段，单位米                          "ts":   "hello", //目标ssid，手机连接的WIFI的ssid                          "tmc":  "00:01:02:03:04:05", //目标设备的mac地址，手机连接的WIFI的mac地址                          "tc":   "Y", //是否与路由器相连                          "ds":   "N",//手机是否睡眠                          "essid0":  "七天连锁\_wifi"//手机用户9a:21:6a:7b:62:6a曾经连接过的WIFI的SSID                  "essid1"："工商银行"//手机用户9a:21:6a:7b:62:6a曾经连接过的WIFI的SSID                 "essid2":"东方明珠",                "essid3":"home"，                 "essid4":"abcd",                 "essid5":"xiong",               "essid6":"XX会馆"                  }, {                          "mac":  "1c:31:72:5c:83:6b",                          "rssi": "-69"，                                  "range": "14.0"，                          "ts":   "world",                          "tmc":  "00:01:02:03:04:06",                          "tc":   "Y",                          "ds":   "Y",                  "essid0":  "七天连锁\_wifi"//手机用户9a:21:6a:7b:62:6a曾经连接过的WIFI的SSID                  "essid1"："工商银行"//手机用户9a:21:6a:7b:62:6a曾经连接过的WIFI的SSID                  "essid2":"东方明珠",                  "essid3":"home"，                  "essid4":"abcd",                  "essid5":"xiong",                  "essid6":"XX会馆"                  }]  }  **（3）数据分析（基础分：40）**  基本能够分析以下9大指标，但不仅限于以下9点，参赛选手可以自由发挥，只要在采集到的数据（可以使除探针设备以外的其他数据）基础上的分析合理都可附加创新分，每项5分。能够实时展示结果附加10分。   1. 客流量：店铺或区域整体客流及趋势 2. 入店量：进入店铺或区域的客流及趋势 3. 入店率：进⼊店铺或区域的客流占全部客流的比例及趋势 4. 来访周期：进⼊店铺或区域的顾客距离上次来店的间隔 5. 新老顾客：一定时间段内首次/两次以上进⼊店铺的顾客 6. 顾客活跃度：按顾客距离上次来访间隔,划分为不同活跃度（高活跃度、中活跃度、低活跃度、沉睡活跃度） 7. 驻店时长：进⼊店铺的顾客在店内的停留时长 8. 跳出率：进⼊店铺后很快离店的顾客及占比(占总体客流) 9. 深访率：进⼊店铺深度访问的顾客及占⽐(占总体客流)（可以根据定位轨迹或者停留时长判定）   以上数据指标支持环比和历史对比，并且可以从小时、日、周、月多维度分析。数据分析中的关于范围的界定支持自定义阈值。 |
| 非功能性需求 | * 数据分析平台健壮，数据展现界面友好，探针设备配置简单、运行稳定（10分） * 能够对探针设备状态进行监控，并可展现所有接入探针的运行状态（10分） * 能够对探针设备进行在线远程控制，包括关机，重启，参数配置，升级固件，设定自动开关机时间，远程控制开关（例如：短信模块控制电源）（10分） |
| 其他限制条件：开发环境、实验平台、开发语言、数据库、编译器等限制条件（请尽量明确） | * 开发环境：不限 * 开发语言：不限 * 数据库：不限 |
| 测试数据或平台：提供给参赛者的测试环境和测试数据。（可提供电子档） | * 测试环境：数据计算平台使用linux平台，要求离线计算或者实时计算平台必须是分布式环境（可以是多台虚拟机），不能在单机或者伪分布式环境下测试 * 探针数据接收服务端最少支持1000并发，探针设备每3秒传输一次数据 * 决赛时现场提供wifi探针环境 （如果探针设备自己开发自行搭建测试环境） |
| 开发所需设备及设备指标需求说明 | * WIFI探针：可在第三方平台（淘宝）自行购买，也可以自行购买相关模块开发硬件和固件 * 指标说明：探针可以抓取到mac地址、距离、地理位置、时间等数据 |
| 其他要求 | 提交相关文档：   * 需求规格书：详细描述课题的功能与非功能需求； * 系统设计说明书：详细描述软件的架构、设计理念、算法思路等； * 测试文档：描述测试环境的搭建、准备的测试数据、测试方法等； * 程序源代码：指未编译的按照一定的程序设计语言规范书写的程序代码，包括代码结构说明书。   评分标准：  评分以综合方案技术可行性、功能性要求的实现程度和非功能性要求的实现程序三个方面进行。分别占总分值的20%,40%,40%。既要考虑系统解决思路的方向性，也要考虑实现过程中的可实现性以及对技术方案的先进性。 |