上海深度挖掘说明

1. 简介

数据业务已成为行业发展的趋势走向，从传统到SOCIAL到数字化营销的必然经过趋势；传统Marketing和Teach公司在转型大数据与AI。传统的数字营销被碎片化截流，用户流量被海量APP截流，传统营销方式覆盖用户压力变大，品牌更加寻求更为有效创新的获客渠道。

因此与联通合作，开展数据业务，运用联通的数据进行深度挖掘，获取有用价值。

1. 数据说明

联通数据主要分为以下几大方向：APP数据、网站使用情况、语音数据、短信数据、位置数据、基本用户属性。

APP数据：用户日常的app使用情况。字段主要为：app名称、时间、当天使用次数等

网站数据：同app数据。

语音数据：语音通话中的两个号码尤其任意一个联通号码的数据。包含通话时间、通话时长、主被叫关系等。

短信数据：同语音数据。

位置数据：用户每天的行为轨迹。

基本用户属性：用户的常用属性，如性别、年龄、电话品牌、话费使用情况等等。

1. 数据挖掘

当前由于绝大多数需求均无法提供以往数据进行分析，只能提供部分资料进行分析，如只提供行业相关的电话号码。

针对此情况，本次采取的挖掘方式为。首先对拨打或者接听过行业电话过的用户进行聚类分析，得到其中一类（该类为有效的真实客户的用户），然后对该用户进行分析，建立模型，得到一套评分卡，最后对所有数据进行评分，获取评分较高的数据。

具体流程图如下所示。



1. 存在问题
2. 无以往用户数据，难以分析用户的特征。
3. 无法灵活使用全国数据。
4. 提取的行业号码语音数据为上海的数据过少。
5. 聚类分析得到的用户不一定是有效用户。
6. 无真实客户数据，难以验证模型有效性。
7. 后续思考
8. 在当前基础上增加特征。
9. 获取有效用户进行模型调优。