

# 需求分析

项目名称：\_\_\_\_\_ 基于大数据的用户画像平台研究与开发 \_\_\_\_\_

项目类别：  
☐ 电子商务  
☐ 移动终端应用  
☒ 大数据分析  
☐ 物联网应用  
☐ 人机交互应用  
☐ 其他(\_\_\_\_\_)

命题企业：\_\_\_\_\_ 浙江创建科技有限公司 \_\_\_\_\_

咨询邮箱：\_\_\_\_\_ wbl.cn@163.com \_\_\_\_\_

2017 年 12 月 1 日

## 项目需求分析

### 一、项目背景

从 1991 年 Tim Berners-Lee 发明了万维网（World Wide Web）开始，到 20 年后 2011 年，互联网真正走向了一个新的里程碑，进入了“大数据时代”。经历了 2012、2013 两年热炒之后，人们逐渐冷静下来，更加聚焦于如何利用大数据挖掘潜在的商业价值，如何在企业中实实在在的应用大数据技术。伴随着大数据应用的讨论、创新，个性化技术成为了一个重要落地点。

相比传统的线下会员管理、问卷调查、购物篮分析，大数据第一次使得企业能够通过互联网便利地获取用户更为广泛的反馈信息，为进一步精准、快速地分析用户行为习惯、消费习惯等重要商业信息，提供了足够的数据基础。伴随着对人的了解逐步深入，一个概念悄然而生：用户画像（User Profile），完美地抽象出一个用户的信息全貌，可以看作企业应用大数据的根基。

用户画像，又称人群画像，是根据用户人口统计学信息、社交关系、偏好习惯和消费行为等信息而抽象出来的标签化画像。构建用户画像的核心工作即是给用户贴“标签”，而标签中部分是根据用户的行为数据直接得到，部分是通过一系列算法或规则挖掘得到，是通过对用户信息分析而来的高度精炼的特征标识。

标签提供了一种便捷的方式，使得计算机能够程序化处理与人相关的信息，甚至通过算法、模型能够“理解”人。当计算机具备这

样的能力后，无论是搜索引擎、推荐引擎、广告投放等各种应用领域，都将能进一步提升精准度，提高信息获取的效率。

## 二、项目概述

本项目对基于大数据的用户画像平台的技术算法进行研究并进行功能开发，通过大数据收集-行为建模-构建画像-数据可视化分析的方式，实现人为规定的高度精炼的用户画像特征标识，在精准营销、用户统计、数据挖掘、效果评估、个性化服务、业务经营分析以及竞争分析等方面发挥决策支持作用。

## 三、项目需求

### （一）功能需求

#### 1. 大数据收集

大数据收集大致分为网络行为数据、服务内行为数据、用户内容偏好数据、用户交易数据这四类。

- 1) 网络行为数据：活跃人数、页面浏览量、访问时长、激活率、外部触点、社交数据等。
- 2) 服务内行为数据：浏览路径、页面停留时间、访问深度、唯一页面浏览次数等。
- 3) 用户内容偏好数据：浏览 / 收藏内容、评论内容、互动内容、生活形态偏好、品牌偏好等。
- 4) 用户交易数据（交易类服务）：贡献率、客单价、连带率、回头率、流失率等。

## 2. 行为建模

该阶段是对上阶段收集到数据的处理，进行行为建模，以抽象出用户的标签，这个阶段注重的应是大概率事件，通过数学算法模型尽可能地排除用户的偶然行为。

这时也要用到机器学习，对用户的行为、偏好进行猜测，好比一个  $Y=kX+b$  的算法， $X$  代表已知信息， $Y$  是用户偏好，通过不断的精确  $k$  和  $b$  来精确  $Y$ 。

在这个阶段，需要用到很多模型来给用户贴标签。

例如：

### 1) 用户汽车模型

根据用户对“汽车”话题的关注或购买相关产品的情況来判断用户是否有车、是否准备买车。

### 2) 用户忠诚度模型

通过判断+聚类算法判断用户的忠诚度。

### 3) 身高体型模型

根据用户购买服装鞋帽等用品判断。

### 4) 文艺青年模型

根据用户发言、评论等行为判断用户是否为文艺青年。

### 5) 用户价值模型

判断用户对于网站的价值，对于提高用户留存率非常有用（电商网站一般使用 RFM 实现）。

还有消费能力、违约概率、流失概率等等诸多模型。

### 3. 构建画像

该阶段可以说是二阶段的一个深入，要把用户的基本属性（年龄、性别、地域）、购买能力、行为特征、兴趣爱好、心理特征、社交网络大致地标签化。

用户画像既应根据变化的基础数据不断修正，又要根据已知数据来抽象出新的标签使用户画像越来越立体。

关于“标签化”，一般采用多级标签、多级分类，比如第一级标签是基本信息（姓名、性别），第二级是消费习惯、用户行为；第一级分类有人口属性，人口属性又有基本信息、地理位置等二级分类，地理位置又分工作地址和家庭地址的三级分类。

### 4. 数据可视化分析

该阶段是用户画像真正利用起来的关键环节，在此步骤中一般是针对群体的分析，比如可以根据用户价值来细分出核心用户、评估某一群体的潜在价值空间，以作出针对性的运营。

## （二）运行环境需求

### 1. 软件环境

服务器操作系统及版本：

Windows Server 2008R2 及以上版本；

### 2. 硬件环境要求

#### 2.1 服务器部署

CPU：4 核 2.5GHz

内存：16G；

外存：硬盘 10T；

输入/输出设备列表：鼠标，键盘，显示器等。

### **3. 网络环境要求**

服务端支持多运营商的同时接入。