# 微视频主题推荐系统设计与实现

## 1引言

在互联网时代，越来越多的传统行业受到了互联网不同程度的影响，当然视频行业也不例外。对于一个视频的评价，最直接的标准就是观看次数。比如，电影的票房、电视剧的收视率等。所以，提高视频的观看次数是视频创作者自始至终都需要关注的问题。随着大数据时代的到来，视频的制作需要将目光更多的转移到用户的需求上，这样才能实现视频点播次数的飞速提升。

大数据技术在视频行业中的成功实践有很多。这些视频在制作前期，都会进行大量的数据挖掘与数据分析。首先，在互联网上获取海量的用户数据。然后，通过对这些数据的分析，得到用户群体感兴趣的题材、演员、歌曲等诸多视频信息。最后，视频制作者只需要按照分析得到的信息来创作视频，就有可能获得巨大的视频观看次数。

本系统通过利用大型的社交平台、视频网站、搜索引擎等途径，获得海量的视频数据。经过系统分析处理，将海量用户数据和公司业务需求相结合，为公司的视频创作推荐更有价值的创作信息，从而使制作出来的视频更加符合用户的需求，进而实现视频观看次数的提升。

## 2 系统设计

以移动互联网、大数据、云存储为代表的新一代信息网络技术的发展，导致了传统的广告微视频创作发生了根本性变化。本系统研究了大数据时代的IT架构设计和系统服务层次划分，对该系统的的结构设计如下图1所示：



在图1中，数据层为系统提供了供系统加工处理的基础数据,并对这些数据进行存储。数据来源种类繁多，主要分为三类：来自社交平台的数据，来自视频网站的数据，来自搜索引擎的数据。

接口层主要针对不同的数据源和数据格式，提供相应的数据采集、导入、加工处理方法。接口层通过标准的数据传输接口，为系统提供数据查询和数据更新接口。

核心层提供支撑整个系统的核心模块：数据分析与挖掘。数据挖掘分析视频、用户的相关数据，是视频主题推荐的关键技术。

服务层以微视频内容为核心，提供以下服务：

（1）热点推荐。该服务可以实现从海量的互联网信息中提取当前比较受欢迎的热点话题。在微视频的创作过程中，创作者可以根据自己需要拍摄的对象，从众多推荐热点中选择最适合表现该对象的若干热点词汇，作为微视频的主题。

（2）社交整合。该服务是将微视频的众多发布平台进行合并。比如，微信、微博、优酷等微视频发布平台。通过将这些发布平台整合到该系统中，就可以实现多平台下的信息同步与更新。

（3）内容聚合。该服务将海量的互联网信息中具有多度相似的内容进行聚合，从而为系统使用者提供更多的参考信息。

（4）评估预测。该服务能够对微视频进行市场效果评估。通过对过去众多微视频的特点及其市场效果的深度挖掘分析，从而实现对未发布的微视频的评估预测。

业务层包括微视频创作和评估，微视频创作平台主要面向微视频的制作团队。该平台可以根据使用者创作对象的不同，进行主题筛选。获得当前最受用户欢迎的符合创作对象内容的主题。微视频制作者只需要根据推荐的相关创作信息进行创作，就可以实现较好的市场效果。微视频的评估平台是面向公司的决策人员开发的。借助于评估系统，决策人员可以更及时的了解所发布微视频的市场反映效果，从而指导决策层进行决策制定。

## 3 关键技术研究

该系统的核心就是对数据进行处理分析，所以数据挖掘技术就显得尤其重要，是整个系统的技术支撑。

内容聚合使用聚类算法：聚类是将物理或抽象的集合分成由类似的对象组成的多个类的过程。内容聚合是将不同来源的、不同结构的、具有高度相似度的数据进行聚类。通过聚类形成一个对微视频统一的外部描述。这种外部描述由于是基于聚类算法获得的，会具有高度的概括性和更广的覆盖性。即能够能好的体现微视频在互联网上的真实反馈。

评估预测使用回归分析：回归分析是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。系统将获得海量数据信息进行整合，利用回归分析，总结出用来表示用户评价与微视频属性之间的一种逻辑关系。当确定了微视频的一些特定属性，就可以通过该逻辑关系来预测该微视频未来的市场走向。

## 4 系统实现

数据爬取

数据装载

数据处理

结果展示