基于…深度学习…的行车安全综合评价系统

摘 要：随着物联网技术的快速发展和应用拓展，车联网也逐渐步入城市交通网络。车联网获取车辆的属性信息，实时监控驾驶员的行为信息以及车辆的运行状态，是防范交通事故发生的有效途径之一。首先，基于K-means算法进行频道时段聚类分析，构建字段比对库，利用PMUC关键字匹配算法构造出了多时段的频道三级标签和节目三级目录；其次，通过分析用户行为内部关联，划分用户层次并进行用户分组，进而基于协同过滤算法实现打包推荐；最后，通过实例分析，刻画用户画像，挖掘用户深层信息，做出用户行为预测与潜在价值分析。

关键词：精准推荐,用户画像,聚类分析,关联分析,关键字匹配

**Marketing recommendations for TV programs**

Abstract: With the rapid development of Internet technology and the expansion of applications, the "three network integration" brings great development opportunities to traditional radio and television media, as well as challenges. In order to enhance enterprise competitiveness and enhance user experience, accurate program recommendation has become the core. Based on the scientificity and accuracy of recommendation and packaging, first of all, carry out clustering analysis, based on K - means algorithm, construct a field comparison library.

Key words: Accurate recommendation, User portrait, Clustering analysis, Correlation analysis, Keyword matching

# 引言

广播电视运营的根本动力来源于对利益的追求，通过各种营销手段吸引用户实现从产品到利润这一过程。然而随着市场经济的不断发展，传统的卖方市场早已经过渡到买方市场，而且随着用户对于产品及服务的个性化需求越来越普遍，整个市场营销环境早已经发生了巨大的变化，从而导致了运营商无法再利用传统的营销方法和手段去满足消费者的个性化需求，进而获得经济效益[1]。如何选择营销策略提高效率、提升效益已经成为了解决的瓶颈问题。

数据在市场营销中的应用早已有之，而在大数据时代，数据在市场营销中的重要性更加突出[2]。运营商可以通过对用户数据的收集和整合，进一步挖掘和筛选具有商用价值的信息，利用这些信息进行市场细分和精准的市场定位，然后面向目标用户采取针对性的营销策略，为其提供差异化的产品和服务，最终获得更多的忠诚用户。

本文一方面从研究的对象及目的出发，通过阅读与大数据挖掘、精准营销等相关的国内外文献，对其进行梳理与总结，掌握本研究领域的一些基础知识，为此次论文写作打下了一定的理论基础。另一方面，通过利用大数据挖掘与精准营销之间的切入点，结合大数据挖掘方法，得出精准营销策略，最后引用一个具体的实例来验证对于基于大数据挖掘的精准营销策略的具体应用[3]。

# 基本思路

问题一的分析

题目要求根据附件1数据，分析用户收视偏好，给出附件2中产品的营销方案。结合数据分析，分三个步骤解决该问题。

（1）将数据分为频道类和节目类数据，以时间、内容纬度描述频道特性，发掘节目特点，通过用户收看频道和节目的记录判断用户收视偏好。拟采用基于K-Means算法的时间聚类划分频道时段，利用网络爬虫获取央广网频道播放内容的参照库，以“频道-时段-内容类别”的形式描述频道特征；通过网络爬虫获取豆瓣电影信息库，构建三级目录体系，利用基于PMUC多模匹配算法确定节目特征。

（2）设置偏好程度评分函数，对用户收视偏好程度进行量化评分。

（3）根据用户收视偏好，结合节目内容和三级目录，利用多模匹配算法从附件2中选取合适的节目并推荐给对应用户。

问题二的分析

题目要求根据附件1~附件3的数据，对相似偏好的用户进行分类，对产品进行分类打包，给出营销推荐方案。参考某电视公司打包定价策略[4],分三个步骤解决该问题。

（1）根据用户收视行为，结合附件3用户套餐情况，综合收视偏好、收视特点和消费能力三个标准，使用协同过滤算法，对用户进行分组,并设置用户标签。

（2）根据问题一发掘的节目特点和用户分组情况，对节目进行组合打包，并设置节目包标签。

（3）通过用户标签和节目包标签的匹配，将打包节目准确地推荐给用户分组本文解决问题的具体思路，如图2-1所示。



图2-1

# 数据预处理

# 在进行驾驶行为数据分析前，必须对采集的数据进巧选取，把数据处理成可用的形式，原始数据中存在的很多问题都必须进行解决比如存在缺失值的数据、字段重复、无效的以及超出一定范围的数据。由于从车载系统采集的驾驶行为数据量较大，存在噪声数据、空缺值数据Ｗ及重复性数据等特点，因此在数据选取阶段需要进行预处理，预处理约占数据挖掘与分析道、体工作量的一半以上，主要包括数据清理、数据变换、数据集成、数据归约等步骤，最终将原始采集的数据集成到一个一致的存储中。数据的预处理工作对数据分析会产生较大的影响，好的数据预处理方法是进行数据分析的基础。

## 3.1 数据理解

## 3.2 数据整合

1. 数据格式的转换

## 3.3 数据降噪

# 基于信息熵的轨迹特征提取

## 轨迹数据分段

## 基于信息熵的轨迹特征提取

对轨迹每点的速度与角度进行排列熵算法的计算,可以得出相应的速度排列熵和角度排列熵的属性,这些属性能够表达出该轨迹的有序程度,能够体现出轨迹内部的细微变化,所以把排列熵作为轨迹的特征值参与轨迹分类。

把轨迹速度排列熵作为轨迹的一种特征值,参与出行模式分类,下面对基于速度排列熵提取的算法进行描述。

(1)获取轨迹轨迹,将轨迹中每个坐标点的速度构成时间序列T= ,其中表示第个点的速度;

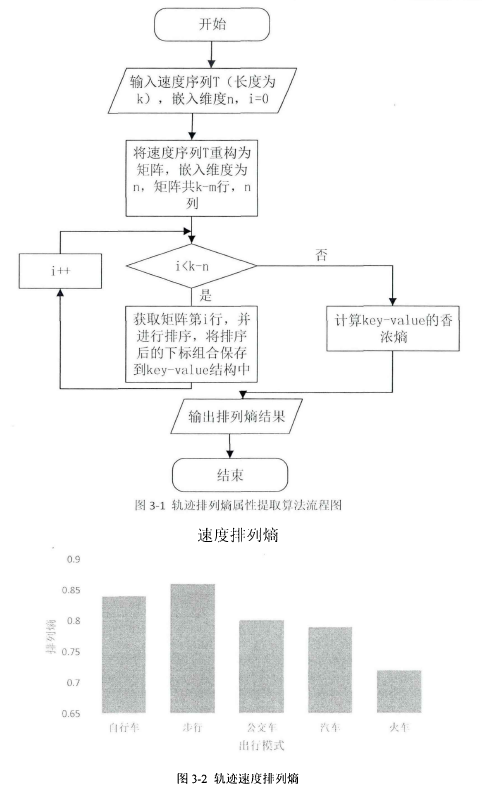
(2)选择合适的维度m，作为求轨迹排列滴的嵌入维度；

(3)将轨迹的速度序列重构为矩阵；

(4)按照升序的顺序对矩阵中的重构分量进行排序；

(5)计算A:个符号向量的香农熵作为Traj的排列熵；

提取轨迹排列熵算法的流程图如图3-1：



## 实验分析：排列熵

使用排列熵以及不使用排列熵的里程、平均速度、加速度更符合期望值。

* + 1. **里程**
    2. **平均速度**

轨迹的平均速度为轨迹包含每段距离的长度总和与每段轨迹所用吋间之和

的比值。公式见3-1:****



* + 1. **加速度**

# 基于BP神经网络的不良驾驶行为模式分析

## 5.1 轨迹分类模型

## 5.2 基于BP神经网络的驾驶行为模式分析

## 5.3 实例分析：不良驾驶行为模式分析

# 6.基于……的评价模型构建

## 6.1 基于……行车安全综合评价模型

## 6.2 基于……综合评价模型构建

## 6.3 实例分析

# 7.结论

# 参考文献

[1]庄薇薇.如何寻找大数据营销的最佳切入点[J].科技资讯,2014,12(30):240.

[2]张楚,冷昕.大数据在电商企业的应用研究——以阿里巴巴为例[J].滁州职业技术学院学报,2017,16(01):60-63.

[3]艾宏.大数据背景下精准营销的探析与思考[J].现代商业,2016(22):34-35.

[4]刘一枞.香港付费电视市场概况及现有付费电视提供商的打包定价策略[J].2012.

[5]龚敏,刘广丹.基于大数据的精准营销应用研究综述[J].市场周刊(理论研究),2016(07):58-60.