

# Precision Marketing Based on Hamming Distance Classification Algorithm

Yuan Xie, Xuan Liu, Yani Wen, Yong Xiao\*

Hunan Agricultural University, Changsha Hunan

Email: [xieyuan\\_sss@outlook.com](mailto:xieyuan_sss@outlook.com)

Received: Aug. 26<sup>th</sup>, 2018; accepted: Sep. 7<sup>th</sup>, 2018; published: Sep. 14<sup>th</sup>, 2018

---

## Abstract

It will become a trend that within the absence of large data age, enterprise marketing selling will no longer be limited to the traditional marketing concepts and methods, while relying on modern information means to achieve precision marketing. The paper mainly researches on the purchase data based on a brand mobile phone, targets on the user's personal information and behavior preference, using Hamming distance classification algorithm and BP Neural network to classify the users to achieve the goal of precision marketing.

## Keywords

BP Neural Network, Hamming Distance Classification Algorithm

---

# 基于汉明距离分类算法的精准营销

谢 缘, 刘 璇, 文雅妮, 肖 勇\*

湖南农业大学, 湖南 长沙

Email: [xieyuan\\_sss@outlook.com](mailto:xieyuan_sss@outlook.com)

收稿日期: 2018年8月26日; 录用日期: 2018年9月7日; 发布日期: 2018年9月14日

---

## 摘 要

随着大数据时代的到来, 企业营销将不再局限于传统的营销理念与手段, 依托现代信息手段, 达到精准营销将成为潮流趋势。本文主要研究基于某品牌手机的购买数据, 针对用户的个人信息和行为偏好, 采用汉明距离分类算法和BP神经网络, 对用户进行分类, 从而达到精准营销的目的。

\*通讯作者。

## 关键词

BP神经网络, 汉明距离分类算法

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

精准营销[1]是通过分析海量数据下所隐藏的消费者的行为习惯, 个人偏好, 全方位建立个性化的顾客沟通服务体系, 一对一动态预测消费者的消费需求, 区别于传统营销的无针对性的静态推销, 极大地降低了营销成本。

本文通过探索用户与用户, 用户与商品, 商品与商品之间的关系, 分析影响该品牌手机的购买的因素: 目标用户的基本属性特征、基本行为特征以及个人偏好, 基于“神经网络”的目标用户群潜力值建立潜力挖掘模型, 对不同目标用户群进行不同的营销手段, 从而提高购买率。

## 2. 数据预处理

由于各表数据总数不一致但均为目标用户的基本行为特征以及个人偏好, 因此本文认为数据具有重复值, 按照用户编号进行筛选, 剔除用户编号重复项数据[2], 总共得到 4840 位信息完全的目标用户, 本文在接下来的数据分析中将采用信息完全的目标用户数据; 因此剔除 50 个测试样本, 得到 4790 个训练样本。

## 3. 数据分析与建模

### 3.1. 数据编码

编码与编码之间的距离在信息论中称为汉明距离, 用于度量编码与编码的相似性与差异性[3], 汉明距离越大表明编码与编码之间的差异性越大。

引用此概念, 为研究个人偏好是否对该品牌手机的购买是否产生影响, 在用户个人行为中提取出若干个指标[4], 对用户中发生的行为编码为 1, 未发生的编码为 0。通过对数据的分析, 发现用户数据中“触媒行为”“视频行为”这两组数据信息在一定程度上可体现用户的个人偏好, “电商行为”这组数据可了解用户的主要手机品牌偏好, 因此本文将触媒行为、视频行为、电商行为作为个人偏好指标, 编码规则如图 1。

### 3.2. 汉明距离分类算法(HMCL Algorithm)

对用户行为编码后, 是通过计算编码与编码之间汉明距离的大小来度量用户之间的相似性和差异性, 其值等于码元与码元对应位置的值相减后取绝对值再相加, 汉明距离越大表明编码与编码之间的差异性越大; 在本文中将用户的行为信息进行编码后, 可由用户和用户的汉明距离大小是否属于同一阈值判断两两用户是否属于同一个类别。

得到目标用户个人偏好之间的联系, 从而找到个人偏好相似度高的用户群体, 汉明距离分类算法[5]基本步骤如下:

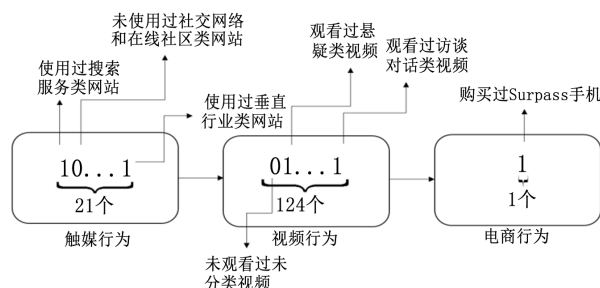


Figure 1. Hamming distance Coding Rules

图 1. 汉明距离编码规则

Step 1: 从集合  $U$  中选出两个最接近的元素  $u_i$  和  $u_j$  作为初始集合  $C = \{u_i, u_j\}$ 。

Step 2: 从剩下的变量任意选取  $u_k$ , 如果  $\{|u_i - u_j| < \epsilon, |u_j - u_k| < \epsilon, \dots\}$  时, 将  $u_k$  归入集合前一个集合  $C = \{u_i, u_j, u_k\}$ 。

Step 3: 当遍历完集合中所有元素, 将得到的集合  $C$  中所有元素归为一类。

对于阈值的选取, 旨在通过用户的各种行为能够将数据清晰的分类开来, 第一层分类以后, 以此类推, 逐层选取阈值再分类

本文通过对数据进行处理、比对、观察后总结得到对于触媒行为阈值选取 1 和 2 可以将其较清晰地划分开来, 以此类推对于视频行为阈值选取 3, 电商行为阈值为购买和未购买。根据以上阈值的选取应用到汉明距离分类算法在触媒行为中将用户分为 3 类, 在此基础上又根据视频行为分为 6 类, 建立在前两次的基础上同样根据电商行为分为 12 类, 这 12 类为最终的类别总数, 对这 12 类依次排序。可对用户按个人偏好进行分类得到不同的目标用户群, 然后对目标用户群的潜力进行排序, 得到如图 2 中每个目标用户群的序列号。

### 3.3. BP 神经网络的应用

BP 神经网络[6] [7]的计算过程由正向计算过程和反向计算过程组成, 正向传播过程, 输入模式从输入层经隐含层逐层处理, 并转向输出层, 每一层神经元的状态只影响下一层神经元的状态。如果在输出层得到的输出值与真实值相差太大, 则转入反向传播, 将误差信号沿原来的连接通路返回, 通过修改各神经元的权值, 使得误差信号最小;

将用户分为 12 类以后, 对于测试样本的分类, 本文应用神经网络对测试样本进行预测, 其中 4790 名用户作为训练样本, 50 名待测用户作为测试样本。4790 位目标用户特征值的  $1 \times 146$  向量, 构成  $146 \times 4790$  矩阵, 将其作为输入数据, 将 12 类目标用户群排序后的对应编号作为输出数据, 使用输入数据以及输出数据建立网络, 为提高模型精确度, 将图 3 中 BP 神经网络中隐含层定为 2, 通过调节节点数来提高模型的精准度。

## 4. 模型结果及结论

待判用户进行指标匹配, 得到个人偏好中的 146 个特征值, 将匹配为空值的用户的个人偏好认为该用户未发生此类行为, 因此可将其特征值赋值为 0; 由此得到  $146 \times 50$  矩阵作为输入数据, 得到输出数据为  $1 \times 50$  的行向量, 输出数值即为该用户所属类别, 通过对 BP 神经网络得到的结果进行分析, 本文认为共有 26 位用户会购买该款手机, 24 位用户不会购买该款手机。

本文通过对每位用户的特征值向量识别, 将其与已知类别中特征值向量组进行匹配, 从而判断其所属类别, 该思路与深度学习较为相似, 均为通过组合低层特征形成更加抽象的高层表示属性类别或特征,

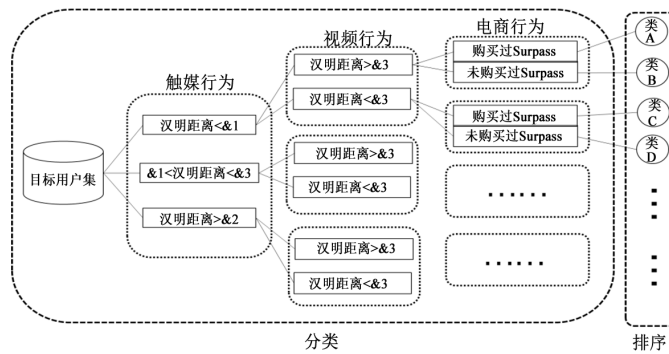


Figure 2. Classification rules  
图 2. 分类规则

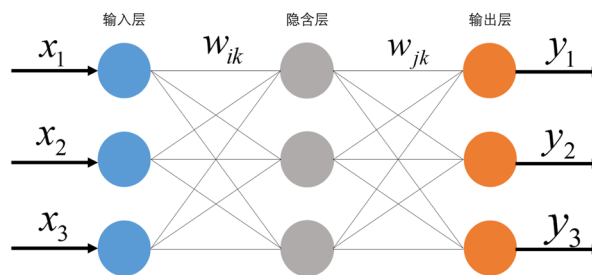


Figure 3. Sketch of BP Neural network  
图 3. BP 神经网络示意图

以发现数据的分布式特征表示，在初步实现深度学习过程中发现自身知识库以及硬件均达不到完整实现深度学习标准，因此本文在后续的研究中采用 BP 神经网络进行模型的构建。

### 5. 营销建议

针对所有的用户群体给出以下建议，作为精准营销的条件：

- 1) 借助微博平台进行品牌推广；
- 2) 制作新颖度高的产品视频投放在各大视频平台以及广告墙背投；
- 3) 相似用户个性化推荐；
- 4) 高频浏览网站广告投放；
- 5) 网络平台合作。

### 参考文献

- [1] 张伟萍. 苏宁客户细分与精准营销研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2017.
- [2] 刘婷婷, 王小丽, 葛明涛. 基于数据挖掘的航空公司客户价值分析[J]. 山东工业技术, 2017(4): 287-288.
- [3] 张焕炯, 王国胜, 钟义信. 基于汉明距离的文本相似度计算[J]. 计算机工程与应用, 2001(19): 21-22.
- [4] 李新欣. 船舶及鲸类声信号特征提取和分类识别研究[D]: [博士学位论文]. 哈尔滨: 哈尔滨工程大学, 2012.
- [5] 申彦. 大规模数据集高效数据挖掘算法研究[D]: [博士学位论文]. 镇江: 江苏大学, 2013.
- [6] 冯婧. 基于 BP 神经网络的个人信用风险评估模型的研究[D]: [硕士学位论文]. 太原: 太原理工大学, 2017.
- [7] 石云. BP 神经网络的 Matlab 实现[J]. 湘南学院学报, 2010, 31(5): 86-88, 111.

**知网检索的两种方式：**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[csa@hanspub.org](mailto:csa@hanspub.org)