

清 华 大 学

# 综 合 论 文 训 练

题目： 基于时域和行为特征的电子  
商务虚假交易的检测与识别

系 别： 自动化系

专 业： 自动化

姓 名： 郭 畅

指导教师： 赵 虹 副教授

2018 年 6 月 22 日

# 关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

**(涉密的学位论文在解密后应遵守此规定)**

签 名：\_\_\_\_\_ 导师签名：\_\_\_\_\_ 日 期：\_\_\_\_\_

## 中文摘要

一个商家，致力于以更优的价格售出更多的商品，以此来获得更多的利润。这个规则同样适用于电子商务领域。为了迎合电商平台的规则，进一步增强竞争优势，不法商家会进行虚假交易。然而，由于电子商务平台数据的商业保密性，目前针对虚假交易领域的研究大多停留在评论及销量层面。本文基于时域以及消费者行为数量，模拟消费者决策过程，提取相应特征进行学习和训练，旨在从新的角度检测与识别虚假交易。

本文基于阿里集团公开的消费者行为日志来建立消费者决策模型，甄别异常消费过程。模型能够较好地分析用户决策细节，为电子商务虚假交易检测领域相关研究提供一定的参考和指导。

**关键词：**电子商务；虚假交易；时域；消费行为

## ABSTRACT

A businessman aims to sell more goods with better prices, in order to obtain more profits. This rule also applies to e-commerce. To cater to e-commerce platform's rules and further enhance its competitive advantage, unscrupulous online retailers will organize fraud transaction. However, due to the commercial confidentiality agreement of e-commerce platform data, most researches in this field focus on reviews and sales. Based on the time domain and consumer behavior, this thesis simulates the consumer decision-making process and extracts corresponding characteristics for learning and training. It aims to detect and identify fraud transaction from a new perspective.

This thesis establishes a consumer decision model based on Alibaba Group's consumer behavior log to identify abnormal consumption processes. The model can analyze the details of user decision making. It provides reference and guidance for relevant research in the field of e-commerce fraud transaction detection.

**Keywords:** electronic-commerce; fraud transactions; time domain; consumer behavior

# 目 录

第 1 章 引言.....	1
1.1 研究背景.....	1
1.1.1 电子商务发展迅猛 .....	1
1.1.2 依附电子商务的虚假交易产业 .....	2
1.1.3 虚假交易引发的问题 .....	3
1.1.4 相关法律文件 .....	3
1.2 研究现状.....	3
1.2.1 基于商家信誉虚增的检测方法 .....	4
1.2.2 基于刷手关系网络的检测方法 .....	4
1.2.3 基于关键词相似度的检测方法 .....	5
1.2.4 基于以往行为模式的检测方法 .....	5
1.2.5 基于虚假评论的检测方法 .....	6
1.2.6 基于自适应神经网络的检测方法 .....	6
1.3 主要工作.....	7
1.3.1 主要内容 .....	7
1.3.2 文章结构 .....	7
第 2 章 虚假交易缘由.....	9
2.1 电商平台的数据监控.....	9
2.2 消费者决策模拟.....	10
2.2.1 引流层面 .....	10
2.2.2 决策层面 .....	11
2.3 虚假交易动机与收益.....	11
2.4 本章小结.....	13
第 3 章 虚假交易运作机制.....	14
3.1 虚假交易角色职能划分.....	14
3.2 虚假交易流程.....	17
3.3 虚假交易特点.....	18
3.4 虚假交易实例.....	19

3.5 本章小结 .....	21
<b>第 4 章 虚假交易判断及识别方法 .....</b>	<b>22</b>
4.1 数据集描述 .....	22
4.2 数据预处理 .....	22
4.2.1 品牌选择 .....	22
4.2.2 数据去重 .....	23
4.2.3 异常处理 .....	23
4.3 特征获取 .....	24
4.3.1 时域特征获取 .....	25
4.3.2 行为特征获取 .....	26
4.3.3 特征统计分布展示 .....	28
4.4 消费行为判定 .....	31
4.4.1 正常消费行为 .....	31
4.4.2 虚假消费行为 .....	33
4.4.3 判定结果展示 .....	33
4.5 特征处理 .....	36
4.5.1 归一化 .....	36
4.5.2 采样 .....	36
4.5.3 特征选择 .....	38
4.5.4 特征提取 .....	38
4.6 模型分类器 .....	41
4.6.1 支持向量机 .....	41
4.6.2 逻辑回归 .....	42
4.6.3 随机森林 .....	43
4.7 本章小结 .....	44
<b>第 5 章 总结与展望 .....</b>	<b>45</b>
5.1 总结 .....	45
5.2 展望 .....	46
插图索引 .....	48
表格索引 .....	49

参考文献 .....	50
致    谢 .....	52
声    明 .....	53
附录 A 外文资料的调研阅读报告或书面翻译 .....	54

## 主要符号对照表

$S$	某一类目下商家构成的集合
$s_i$	序号为 $i$ 的商家
$U_i$	与商家 $i$ 交互过的用户集合
$u_{ij}$	与商家 $i$ 交互过的用户 $j$
$B_i$	购买过商家 $i$ 的用户集合
$b_{ij}$	购买过商家 $i$ 的用户 $j$
$F_k$	特征 $k$ 构成的集合
$\mu_k$	特征 $k$ 的均值
$\sigma_k$	特征 $k$ 的标准差
$P_k$	特征 $k$ 归一化后构成的集合
$N_{\text{fraud}}$	采样过程中负样本的数量