

基于差异化数据集多层分类融合的用户感知反馈算法

贝玉昌^{1,2}, 吴栩欣², 万 海³

(1 华南理工大学 计算机系统研究所, 广东 广州 510000;

2 中国移动通信集团广东有限公司 佛山分公司, 广东 佛山 528000;

3 中山大学 软件学院, 广东 广州 510000)

摘 要: 在用户感知反馈计算的过程中, 传统的反馈算法在多层次用户感知不一致的情况下, 数据的差异会使得感知结果产生分类摇摆, 导致感知反馈结果不准. 为解决这一问题, 提出一种基于差异化数据集分层融合的用户感知反馈算法, 通过 QOE-KQI-KPI 三层体系模型获取用户感知反馈结果并且进行分类, 计算多层系统感知反馈聚类加权的均值, 设定每个层次反馈的加权平均值, 消除多层数据的差异性, 解决由于数据差异过大带来的感知计算不准的问题. 实验证明, 这种方法能够有效融合多层用户感知结果, 取得了满意的结果.

关键词: 用户感知; 加权均值; 数据差异

中图分类号: TM135

文献标识码: A

文章编号: 1000-7180(2012)12-0193-04

Based on the Differential Data Set Multilayer Classified Fusion User Perception Feedback

BEI Yu-chang^{1,2}, WU Xu-xin², WAN Hai³

(1 Institute of Computer System, South China University of Technology, Guangzhou 510000, China;

2 China Mobile Group Guangdong Co., Ltd. Foshan Branch Guangdong, Foshan 528000, China;

3 School of Software, Sun Yat-sen University, Guangzhou, 510000, China)

Abstract: In the user perceived feedback calculation process, the traditional feedback algorithm in multi-level user perception inconsistent cases, the difference of data will make sense the result classification swing, leading to perception feedback results are not allowed. In order to solve this problem, this paper puts forward a data set based on differentiation layered fusion user perception feedback algorithm, through the QOE - KQI - KPI three layer system model for users' perception feedback results and classification, calculation multilayer system perception feedback clustering weighted average, set each level of feedback weighted average, eliminate multiple data differences, solve the data difference too big to bring the perception of the problem of calculation is not allowed. Experiments show that the method can effectively fusion multiple users' perception results, and satisfactory results were obtained.

Key words: users' perception; weighted mean; data difference

1 引言

随着移动通信技术的不断发展, 移动通信设备在人们的日常生活中, 发挥着越来越大的作用. 随着市场竞争的进一步加剧, 对客户体验感觉进行及时的了解, 并且准确的反馈到相关部门, 对出现问题

的部门和服务进行调整, 成为了一项极其重要的课题^[1-2]. 用户的感知对于一个企业的发展、改善、壮大, 有着极其重要的意义, 因此成为售后服务中的重中之重^[3]. 传统的用户感知体验方法是建立在对网元的有效管理基础上, 这种管理方式是通过一些单级反馈系数了解一些新业务的用户感知情况^[4-5].

这种方法的优点是实现简单,但是缺点是无法准确的反应用户的真实感知情况.无法避免多层用户的反馈结果差异过大对用户感知结果的影响,导致用户感知反馈结果误差较大^[6-7],算法无法准确获取真实的用户感知情况,不能满足需求.为了解决这一问题,提出一种基于差异化数据集分层约束融合的宽带用户感知测量算法.

2 差异化用户反馈数据集分层约束融合算法

2.1 基于 QOE-KQI-KPI 的多层用户感知数据获取

设计 QOE-KQI-KPI 三层体系模型对结构中所有用户感知信息进行获取,在所有用户感知信息中放置检测函数,检测出用于建立最终用户感知信息的多层用户反馈信息.根据获取的用户多层感知信息建立一个有效的集合,集合中的样本数目代表用户反馈信息的不同层次和用户感知数量.多层模型中的用户感知结果获取分为主动获取和被动获取两种方式,具体步骤如下所述:

(1) QOE-KQI-KPI 三层反馈信息获取

通过主动方式获取用户感知结果,需要设定一个唯一的检测标准,通过反馈函数中的所有反馈特征表达式获取相关用户反馈信息,进行确认分析,如果符合用户反馈信息的特征标准,那么把这样的信息作为用户感知体验信息,反映到后期的感知分析系统中.

(2) QOE-KQI-KPI 反馈信息被动获取方式

这种方式主要是通过被动的获取用户感知信息,完成用户 QOE-KQI-KPI 反馈信息的采集,由于是无选择提取,因此,信息中不会掺杂无用的信息,得到的信息能够直接的代入到后期的系统中.

2.2 多层用户感知信息分类

由于反馈结果中,数据差异较大,反馈结果可分为不同类的数据集和设为 $S = \{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$, 每个计算得到的感知结果数据元组 $X_i (0 \leq i \leq m)$ 用 n 维属性向量 $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ 表示, $y_i \in Y = \{1, 0\}$ 为 X_i 的类标号,即数据元组所属类别.对感知结果的有效分类的前提是对集合的有效划分,差异化的集合分为训练集和检验集,分类过程主要由两个阶段构成:

(1) 用户感知反馈数据中,带有自学习功能的差异化分类器的设计,学习的过程可以用一个函数 $y = f(x)$ 表示,满足学习要求的可以被划分的相应的类别中.

(2) 使用分类器对检验集进行分类,并且对分类后的感知结果进行标号校验,反馈这种数据分类的正确性,计算分类的准确率,对分类结果的验证过程十分重要,它决定着后期分类器的选择,只有在对感知结果的分类准确率达到一定要求时,才有资格参与下一个分类过程,分类原理如图 1 所示.

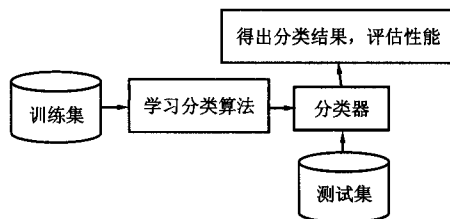


图1 分类原理流程图

由于用户感知的反馈结果存在较大差异,因此需要给出大差异类样本集定义^[7]:

对于差异化的感知数据集 S , 若 $Y^* \subset Y$, $S^* = \{x | x \in S, \text{且 } \text{class}(x) \in Y^*\}$, 对于 ϵ , $\frac{|S| - |S^*|}{|S^*|} \geq \epsilon$ 成立, 则称 S 是大差异类样本集, ϵ 称为差异率. ϵ 越大, 则 S 中数据差异越明显. 其中 Y^* 中的元素称为少数类别, S^* 中的元素成为少数类别样本.

2.3 多层用户感知信息融合反馈

在 QOE-KQI-KPI 用户反馈信息的融合系统中, 假设出现用户的感知反馈信息不一致的情况下, 那么将不一致的用户感知结果进行融合, 从而得到准确率更高的多层用户感知结果. 为了有效整合不同区域的用户感知结果, 采用模糊聚类方法建立 QOE-KQI-KPI 反馈信息整合模块, 对不一致的 QOE-KQI-KPI 反馈信息结果进行整合处理. 通过整合后的多模型信息, 能够得到 QOE-KQI-KPI 反馈信息集合, 然后按照反馈事件中用户感知反馈的相关信息, 确定 QOE-KQI-KPI 反馈信息的初始矩阵.

(1) 初始矩阵如下:

$$\begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1j} \\ Y_{21} & Y_{22} & \dots & Y_{2j} \\ \dots & \dots & \dots & \dots \\ Y_{i1} & Y_{i2} & \dots & Y_{ij} \end{bmatrix} \quad (1)$$

在 QOE-KQI-KPI 反馈信息初始矩阵中, i 代表用户反馈信息集合中样本数量, j 代表 QOE-KQI-KPI 反馈信息数量, Y_{ij} 代表第 i 个 QOE-KQI-KPI 反馈信息点在第 j 层 QOE-KQI-KPI 模型中获取值. 依据获取的 QOE-KQI-KPI 反馈信息通过运算能够得到每层反馈结果的平均值 u 和

多层反馈信息衡量标准 λ , 计算表达式如下:

$$\begin{cases} u=i/\sum_{i=0}^i X_{ij} \\ \lambda=\sqrt{j/\sum_{i=1}^i Y_{ij}^2} \end{cases} \quad (2)$$

根据得到的结果,对不一致的多层 QOE-KQI-KPI 反馈信息结果进行整合,利用模糊聚类方法获取 QOE-KQI-KPI 反馈信息集合在每层模型中的模糊集合:

$$Y'_{ij}=\frac{Y_{ij}}{\lambda} \quad (3)$$

$$s_{ij}=f(Y_i,Y_j) \quad (4)$$

其中, s_{ij} 代表模糊集合近似矩阵. $f(Y_i,Y_j)$ 代表第 i 层 QOE-KQI-KPI 反馈信息和第 j 层 QOE-KQI-KPI 反馈信息的间隔:

$$f(Y_i,Y_j)=\sqrt{Y_i+Y_j} \quad (5)$$

(2)按照多层 QOE-KQI-KPI 反馈模型中每层的用户反馈信息,利用这个模型将 QOE-KQI-KPI 反馈信息之间的关系进行整理. $P=\{P_j|2<j<p\}$ 代表 QOE-KQI-KPI 反馈信息系数集合.

(3)每层用户感知反馈的结果代表这层系统获取的结果对于最终定位用户感知准确性的影响,权值与这层用户感知结果对于评估准确性成正比.在多层用户感知结果不同的情况下,使用不同的权值表示每层获得对于获取的用户感知结果对最终结果的影响,能够三层系统获取的用户感知结果不一致的问题.权值主要通过每层系统获取的单个用户感知节点的权值进行平均处理,因为每层获取的用户感知权值对于最终结果确定的重要程度不同,然后确定重要性系数,最后进行归一化处理.得到最终的感知权值确定,如下所示:

$$B=0\begin{bmatrix} b_{11} & b_{12} & \cdots & 0 \\ b_{21} & b_{22} & \cdots & b_{2j} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & b_{j2} & \cdots & b_{jj} \end{bmatrix} \quad (6)$$

通过以上的矩阵中可知 $b_{ij}>0$:

$$X=(\sum_{k=2}^p Y_{1k}+\sum_{k=2}^p Y_{2k}+\cdots+\sum_{k=2}^p Y_{pk})^{-1} \quad (7)$$

$$Q=\frac{p}{\sum_{j=1}^p(\frac{X_j}{Y_j})} \quad (8)$$

通过对 QOE-KQI-KPI 三层体系模型采集的不同用户感知结果,进行整合处理和聚类处理,有效避免了不同层次得到的结果不一致对用户感知结

果最终确定的影响,通过合理的计算权值,保证评估结果准确.

3 实验结果分析

3.1 实验数据集的选取

为了较为真实的模拟用户感知反馈数据的真实性,实验采用的仿真数据集来自于 5 个电信公司的年度公开用户感知反馈数据,不同公司的数据集用不同的编号表示,设为公司 1、2、3、4、5. 5 个数据集除公司 1 外均为两类不平衡数据集,对公司 3 数据集选取类标为“3”的一类作为少数类,这也会自动形成一个差异化的环境. 公司 2 数据集中也自动分类成少数类与多数类的差异化环境. 表 1 为 5 个数据集的基本信息, ϵ 为不平衡率.

表 1 数据集描述

电信公司数据集	1	2	3	4	5
样本总数	235	388	387	356	335
少数类样本数	113	114	13	21	12
多数类样本数	400	378	238	668	448
属性个数	3	5	8	1	2
少数类占比(%)	12.1	22.4	34.0	17.6	22.0
ϵ	2.3	1.7	4.1	3.2	4.0

3.2 数据分类实验标准

仿真实验在 CPU 双核 2.5GHZ,内存为 2G 的 PC 机上进行,将上述数据集分别用随机过抽样方法(random over-sampling ROS)、随机欠抽样方法(random under-sampling RUS)、SMOTE 方法、空间插值法(S-SMOTE)进行处理,以上 4 种方法算法均利用 Eclipse 软件编写 Java 程序实现,使数据集不平衡率 $\epsilon=1$. 在 weka3.5 环境下分别用 J48 决策树算法、NaiveBays 和 BFTree 分类器对 5 个数据集进行学习,并采用十折交叉验证法. 对用户感知数据的分类结果优良的评价可采用几何均值 G-mean 值与少数类的 F-value 值:

$$G\text{-mean}=\sqrt{\left(\frac{TP}{TP+TN}\right)*\left(\frac{TN}{TN+FP}\right)} \quad (9)$$

$$F\text{-value}=\frac{(1+\beta^2)*Recall*Precision}{\beta^2*Recall+Precision} \quad (10)$$

TP 与 TN 分别表示用户感知结果中的大差异和小差异, FP 与 FN 分别表示这两种差异化分类中错误分类的数据数量. G-mean 值中 $TP/(TP+TN)$ 指差异化数据集中的分类正确率, $TN/(TN+$

FP)指小差异的数据集中的分类正确率,两证的对比反应出了,几何均值的变化,为了对大差异数据集的分类性能进行准确的评价,采用了 F-value 值. F-value 值中 $Recall = TP/(TP+FN)$ 与 $Precision = TP/(TP+FP)$ 分别表示小差异数据集中的查全率和查准率,两者的正确变化也间接反应了 F-value 值的变化,因此 F-value 值能正确反应少数类的分类性能.

3.3 反馈实验结果分析

为了检验多层用户感知反馈结果的准确度,以上述实验环境为基础进行仿真实验,对 5 个大型的移动运营商的某些 SP 业务进行用户感知数据进行分类计算,分别使用传统用户感知数据分类方式和本文提出的算法对用户感知数据进行分类,对分类结果中的几何均值 G-mean 值与少数类的 F-value 值进行比较,比较不同算法下,计算结果误差,误差趋势图如图 2 所示.

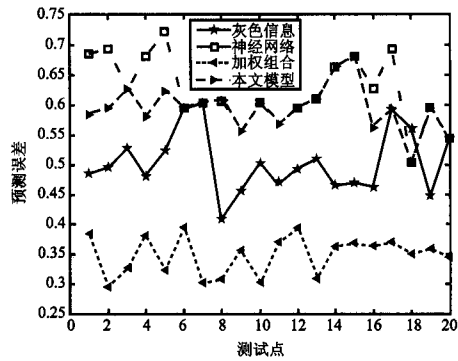


图 2 计算结果趋势图比较

通过图 2 的对比结果可以看出,本文算法下的分类结果误差要明显小于各种传统的分类算法. 优势较为明显,下面对用户感知结果进行计算,运用 CX 的不同取值,代表判断标准下的不同指标权值,选取了 12 个不同标准下的指标权值作为实验的标准取值,结果如表 2 所示.

表 2 各指标权重值

c11	c12	c13	c21	c22	c23
0.16	0.08	0.16	0.06	0.09	0.06
c24	c25	c31	c32	c33	c34
0.045	0.045	0.09	0.09	0.045	0.075

在各个权重指标下,运用不同的方法对用户感知反馈解决进行计算,统计多次试验下,不同算法得到的感知结果与真实结果的误差大小,统计结果如表 3 所示.

表 3 各算法的感知结果统计

评估模型	准确率(%)	错误率(%)
BP 神经网络	90.35	9.65
支持向量机	91.58	8.42
本文模型	97.48	2.52

通过以上的统计结果可以看出,本文提出的用户感知算法在准确获得用户感知方法有着较为明显的优势.

4 结束语

本文提出一种基于感知融合的多层宽带用户感知反馈算法. 通过 QOE-KQI-KPI 三层体系模型获取用户感知反馈结果融合,计算多层系统感知反馈聚类加权的均值,设定每个层次反馈的加权平均值,对多层不同感知结果反馈进行有效反馈. 评估宽带用户上网的感知质量,为运营商进行网络质量监控及网络优化调整提供依据,保证最终结果的准确,解决了用户感知结果不准的问题.

参考文献:

[1] 张光华,仇晶. 认知无线网络中基于信任的合作频谱感知框架[J]. 微电子学与计算机,2011(9):36-37

[2] 戴月明,王金鑫,李彦伟. 基于表决的分布式入侵容忍系统模型及量化分析[J]. 计算机应用研究,2011(6):3-4.

[3] Steven Silverman et al. Miniature thermal emission spectrometer for the Mars Exploration Rover [J]. 2006Avants,laser focus world, 2011(3):67-73.

[4] 刘丽珍,陈蕾,张茂红. 基于贝叶斯网络的大型城市社区消防安全评估模型[J]. 计算机应用研究,2011(01):34-35

[5] Woo-Yong Jang, Maheed M. Hayat, Demonstration of Bias-Controlled Algorithmic Tuning of Quantum Dots in a Well(DWELL)MidIR Detectors[J]. IEEE Journal of quantum electronics, 2009,45(6):56-57.

[6] 黄金国. 改进的遗传神经算法在网络安全检测中的应用[J]. 计算机仿真,2011(6):173-175.

[7] 杨慧中,张文清. 基于特征加权模糊聚类的多模型软测量建模[J]. 控制工程,2011(4):2-3.

作者简介:

贝玉昌 男,(1978-),硕士. 研究方向为数据增值网络的维护工作.

吴翔欣 男,(1977-),研究方向为数据增值网络维护.

万海 男,(1976-),博士. 研究方向为知识表示与推理、逻辑程序设计、程序诊断与修复.