<反欺诈风控模型优化>

方案设计

| 需求编号： | 需求分析人员： |
| --- | --- |
| 文档状态： | 完成日期： |
| 用户确认人员： | 用户确认时间： |

文档相关信息：

| 版本号 | 作者 | 日期 | 备注 |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

相关文档

| 发布日期 | 文档标题 | 版本 |
| --- | --- | --- |
|  |  |  |
|  |  |  |

目录

[1 概述 3](#_Toc520731239)

[**1.1** **背景** 3](#_Toc520731240)

[2 方案设计 4](#_Toc520731241)

[2.1 投保信息相似度 4](#_Toc520731242)

[2.2 费用清单相似度 4](#_Toc520731243)

[2.3 案件诊断相似度 4](#_Toc520731244)

[2.4 账单金额相似度 4](#_Toc520731245)

# **概述**

## **背景**

反欺诈数据模型，根据现有分析维度不满足客户要求，增加模型分析维度。将新增分析维度（即优化因子）加入原有模型框架中。

# **方案设计**

## 投保信息相似度

a 基本属性

b 客户保单所在地

c 是否社保

a, b, c 为该模块下总模型的参数的权重，改变模型参数的重要程度。该模块只有一个模型（是原来本征模型的扩展），只产生一个相似度。基本属性（年龄、性别、婚姻状况、是否有驾照、是否吸烟、案件类型、事故性质）的权重均为，客户保单所在地、是否社保的权重为，其均为参数权重。

## 费用清单相似度

d 项目与诊断相似度

d为项目相似度，与原来不变。

## 案件诊断相似度

e 诊断相似度

f 主要诊断+出险时间

g 次要诊断+出险时间

h 主要诊断+住院费用+住院天数

e 为诊断相似度，与原来不变

f 分诊断（主诊断归类后）下，根据出险时间的长短生成相似度。

生成模型的方法具体方法为：找出出险天数频率曲线的峰值（若有多个峰值，取最右边的）。比峰值对应天数要少的划分为好人，相似度为100。比峰值对应天数多的划分为风险人群，其相似度为其频率值比上峰值的频率值。

计算方法为：分诊断后，根据对应出险时间得出相似度。

g 分次诊断（取时间最长的一个或几个）下，根据出险时间的长短生成相似度。

生成模型时，与f一致。但若有多个次诊断以及多个出险时间（有多个保单），则应在每个出险时间下再分每个次诊断，对所有的情况分别进行频率统计。

计算方法也与f一致，多个次诊断及多个出险时间的情况的处理方法与生成模型时一致。

h 为分诊断下，根据住院费用和住院天数来聚类，按距离算相似度。

## 账单金额相似度

i 理赔险种相似度

j 津贴责任+主诊断+住院天数

k 有无社保+主诊断+住院费用

i 为险种责任相似度，与原来不变。

j 无津贴时，相似度为100。有津贴情况时，分诊断（主诊断归类后）下，根据住院天数的长短生成相似度。

生成模型的方法具体方法为：找出住院天数频率曲线的峰值（若有多个峰值，取最右边的）。比峰值对应天数要少的划分为好人，相似度为100。比峰值对应天数多的划分为风险人群，其相似度为其频率值比上峰值的频率值。

计算方法为：分诊断后，根据对应住院天数得出相似度。

k 与j相似。有社保时，相似度为100。无社保时，分诊断（主诊断归类后）下，根据住院费用的大小生成相似度。先对住院费用进行离散化，以200元为一个单位区间。

生成模型的方法具体方法为：找出住院费用区间频率曲线的峰值（若有多个峰值，取最右边的）。比峰值对应费用要少的划分为好人，相似度为100。比峰值对应费用多的划分为风险人群，其相似度为其频率值比上峰值的频率值。

计算方法为：分诊断后，根据对应住院费用的区间得出相似度。